Philosophie morale et démarche expérimentale, une approche critique

Yves Serra

SORBONNE UNIVERSITÉ

ECOLE DOCTORALE Concepts et Langage

Laboratoire Sciences, Normes et Démocratie

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ SORBONNE UNIVERSITÉ

Discipline : Philosophie

Présentée et soutenue par

Yves SERRA

le 21 octobre 2020

Philosophie morale et démarche expérimentale

Une approche critique

Sous la direction de :

Mme Anouk BARBEROUSSE Professeur, Sorbonne Université

Membres du jury :

M Denis FOREST – Professeur, Université Panthéon Sorbonne

Mme Isabelle PARIENTE-BUTTERLIN – Professeur, Aix Marseille Université

M Pascal LUDWIG – Maître de Conférence, Sorbonne Université

Mme Marta SPRANZI – Maître de Conférence, UVSQ

M Brent STRICKLAND – Maître de Conférence, ENS Ulm

Sommaire

Remerciements **11**

Introduction **13**

1 Morale et philosophie morale **25**

1.1 Le domaine moral 27

1.2 Les théories morales 38

1.2.1 Proposition de définition 39

1.2.2 Les grandes familles de théories morales 44

1.2.2.1 Le déontologisme 44

1.2.2.2 Le conséquentialisme 46

1.2.2.3 L'éthique des vertus 47

1.2.3 La promesse des théories morales est inatteignable 48

1.2.3.1 Mise en doute du discours moral 48

1.2.3.2 La morale, vecteur d'asservissement 50

1.2.3.3 L'équilibre réfléchi 51

1.2.4 Comparer les théories morales : les huit critères de Timmons 53

1.3 Quatre perspectives sur le domaine moral 57

1.3.1 Analyser le domaine moral 57

1.3.2 Perspectives et désaccords moraux 61

1.3.2.1 Les désaccords individuels au sein d'une communauté morale . 62

1.3.2.2 Les désaccords dus à des systèmes moraux différents 65

1.3.2.3 Les désaccords entre philosophes moraux 68

1.3.3 Le réalisme moral de David Enoch 74

1.3.4 Le sentimentalisme constructif de Jesse Prinz 77

1.4 Le domaine moral, point d'étape 80

2 Le mouvement XPhi de philosophie expérimentale 83

2.1 Le mouvement XPhi, premiers pas 85

2.1.1 Aux sources des XPhi : déception et espérance 86

2.1.2 Plusieurs distinctions utiles 90

2.2 Les résultats des XPhi, deux exemples introductifs 94

2.2.1 La connaissance 95

2.2.2 La référence 97

2.2.3 De nombreux programmes de recherche lancés 99

2.3 Les réticences au mouvement XPhi 100

2.3.1 Rien de nouveau 101

2.3.2 Le programme XPhi négatif est faible 104

2.3.3 Philosophie expérimentale et réplication 106

2.4 La philosophie morale expérimentale 108

2.4.1 Des spécificités de la philosophie morale expérimentale. 108

2.4.2 Le cas paradigmatique de l'effet Knobe 111

2.4.3 Les exemples de l'attribution des états mentaux et du déterminisme . . 115

2.5 Un bilan en demi-teinte 118

3 La démarche scientifique expérimentale 121

3.1 Introduction : pourquoi une métaphore ? 121

3.2 La métaphore de l'hélice 124

3.2.1 Une métaphore graphique 124

3.2.2 Les connotations de la métaphore 126

3.2.3 L'hélice, l'empirisme et la vérité 127

3.2.4 Limites de la métaphore 131

3.2.5 La métaphore de l'hélice et la psychologie 132

3.3 Trois temps de l'hélice expérimentale 135

3.3.1 L'opérationnalisation 136

3.3.2 L'objectivation 139

3.3.3 L'interprétation inductive 141

3.4 L'expérimentation dans les sciences de la nature 143

3.4.1 L'expérimentation en appui du développement des théories 144

3.4.2 Le développement autonome du domaine expérimental 146

3.4.3 La spécificité de l'apport de l'expérimentation. 148

3.5 La réplication d'expérience et son paradoxe 151
3.6 La démarche scientifique expérimentale, point d'étape 155

4 Cinq études de cas **157**
4.1 Expérimenter sur l'expérimentation : une mise en abyme 157
4.2 Le tramway, introduction à la méthode 161
4.2.1 Le tramway, entre expérience de pensée et expérimentation 161
4.2.2 Présentation des dilemmes sacrificiels 162
4.2.3 Appliquer nos intuitions morales à de multiples scénarios 164
4.2.4 L'approche expérimentale du dilemme 167
4.2.5 Le tramway, drosophile de la philosophie morale 169
4.2.6 Mais le bilan reste en demi-teinte 172
4.3 La surestimation du nombre de musulmans 174
4.3.1 Le cadre de l'étude 174
4.3.2 Le questionnaire de mai 2017 178
4.3.3 Premiers résultats 181
4.3.4 Les interprétations possibles de la surestimation et les interrogations soulevées 184
4.3.5 Du questionnaire en philosophie expérimentale 187
4.4 La transmission des émotions par les larmes 191
4.4.1 Les larmes communiquent les émotions par l'odorat 191
4.4.2 Les larmes sont spécifiquement humaines 193
4.4.3 La difficile nécessaire collaboration entre chercheurs 194
4.4.4 La diversité, source de conflit et d'opportunités 197
4.5 L'IAT : la réplication et les conflits de valeurs 197
4.5.1 Mesurer l'association implicite entre concepts. 197
4.5.2 Le cas de l'IAT, accords et désaccords 199
4.5.3 Résumé de l'histoire de l'IAT 201
4.5.4 La synthèse minimale 209
4.5.5 Au delà de la synthèse minimale 212
4.5.6 Connaissance et action 215
4.5.7 Conclusion : ce que peut la réplication 218
4.6 L'analyse de l'effet Knobe : instabilité des notions 219
4.6.1 L'effet Knobe 219

4.6.2 Méthode de travail . . . 221
4.6.3 Analyses d'ensemble des 33 articles . . . 223
4.6.4 33 argumentations, et presque autant de distinctions conceptuelles . . . 225
4.6.5 L'intentionnalité : un exemple de cible philosophique ambigüe . . . 232
4.6.6 L'instabilité conceptuelle, point d'étape . . . 240
4.7 Sur-interprétation et sous-opérationnalisation . . . 242
4.7.1 La perspective descriptive . . . 243
4.7.2 La perspective prescriptive . . . 247
4.7.3 La perspective méta-éthique . . . 250
4.7.4 La perspective de l'éthique appliquée . . . 252
4.7.5 Ce que l'expérimentation apporte à la philosophie morale, point d'étape 254

5 L'opérationnalisation 257
5.1 Introduction : observer l'inobservable . . . 257
5.2 Le risque de confusion et l'opérationnalisation . . . 264
5.2.1 Retour sur les 5 études de cas . . . 264
5.2.1.1 Les dilemmes du tramway . . . 264
5.2.1.2 Surestimation du nombre de musulmans . . . 268
5.2.1.3 Opérationnaliser la transmission des émotions par les larmes . 270
5.2.1.4 IAT . . . 272
5.2.1.5 L'effet Knobe . . . 275
5.2.1.6 Pour aller plus loin sur la base de ces cinq études de cas . . . 279
5.2.2 Les problèmes de l'opérationnalisation . . . 285
5.2.2.1 Opérationnaliser, c'est résoudre simultanément trois équations 286
5.2.2.2 L'échelle de l'opérationnalisation. . . . 293
5.2.3 Les enjeux de l'opérationnalisation . . . 295
5.3 Trois stratégies de contournement . . . 296
5.3.1 Le behaviorisme . . . 296
5.3.2 L'opérationnalisme de Bridgman . . . 298
5.3.2.1 Définir les entités théoriques à partir des pratiques expérimentales . . . 298
5.3.2.2 Trois arguments contre l'opérationnalisme . . . 299
5.3.2.3 L'opérationnalisme dans les sciences aujourd'hui . . . 301
5.3.3 Le contournement par la technique . . . 303

5.3.3.1 Un exemple technique : le confort acoustique 303
5.3.3.2 L'opérationnalisation progresse en même temps que les savoir-faire théoriques et expérimentaux. 310
5.3.4 Le contournement entre utilité et risque de stérilisation 311
5.4 Opérationnaliser les entités psychologiques 312
5.4.1 Indicateurs observables 312
5.4.2 L'opérationnalisation en psychologie : une étape nécessaire mais délicate 313
5.4.2.1 L'opérationnalisation dans la méthode 313
5.4.2.2 Triangulation et cohérence 315
5.4.3 Une description insuffisante 318
5.5 L'opérationnalisation, six propositions 319
5.5.1 Six propositions 321
5.5.2 Objections 323
5.5.3 Conséquences pour la philosophie morale expérimentale 324

6 Philosophie morale et approche expérimentale 329
6.1 L'expérimentation morale est-elle impossible ? 330
6.1.1 Quatre types d'arguments 330
6.1.1.1 Une réticence ancienne et qui perdure 330
6.1.1.2 Les arguments des philosophes moraux 334
6.1.2 La complexité et la circularité 336
6.1.2.1 Le problème de la complexité du comportement moral 336
6.1.2.2 Le problème de la circularité 337
6.1.2.3 Les réponses à l'argument de complexité 338
6.1.2.4 La dérive de la définition de la complexité 341
6.1.2.5 L'importance de la complexité pour la démarche empirique . . 341
6.1.2.6 Les réponses à l'argument de la circularité 343
6.1.2.7 Circularité et expérimentation 345
6.1.3 La normativité et le problème de Hume 347
6.1.3.1 Les problèmes liés à la normativité 347
6.1.3.2 Le problème de Hume 348
6.1.3.3 Les réponses au problème de Hume 349
6.1.3.4 Le problème de Hume et l'expérimentation 350
6.1.4 L'impossibilité morale d'expérimenter 352

6.1.4.1 Prétendre expérimenter sur l'homme peut-il être moral ? 352
6.1.4.2 Quelles réponses à l'impossibilité morale ? 353
6.1.4.3 Responsabilité morale et expérimentation 354
6.1.5 La nocivité sociale de l'expérimentation 355
6.1.5.1 Expérimenter c'est socialement faire 355
6.1.5.2 Les normes et leurs liens au comportement moral 356
6.1.5.3 Le contenu moral des normes 357
6.1.5.4 Les normes morales en tant que liant social 360
6.1.5.5 Les réponses au problème de l'engagement social 361
6.1.6 Les difficultés : un bilan complexe 364
6.2 Contourner les théories morales : l'équilibre réfléchi 367
6.2.1 L'équilibre réfléchi 368
6.2.2 Les avantages de la méthode 370
6.2.3 Les points faibles de la méthode 370
6.2.4 Les comités d'éthique en milieu hospitalier 371
6.2.5 L'équilibre réfléchi et la place de l'expérimentation 372
6.3 Les sciences et la question normative 373
6.3.1 La diversité des sciences concernées 374
6.3.2 La neuroéthique 376
6.3.3 La psychologie morale 380
6.3.4 Une science empirique a-normative ? 383
6.4 Les théories évolutionnistes de la morale 384
6.4.1 Présentation des thèses évolutionnistes de la morale 384
6.4.2 Pouvoir explicatif et difficultés de ces thèses 385
6.4.3 L'entrelacs temporel : espèce, groupe, individu 390
6.5 Distinguer existence et contenu d'une norme 392
6.5.1 L'existence des normes pour une espèce 393
6.5.2 Le contenu des normes morales identifient les groupes 396
6.5.3 Le niveau individuel 398
6.5.4 Remarques conclusives, le projet éthique de Philip Kitcher 400

7 Point et perspectives **403**

7.1 Les questions au point de départ de l'enquête 403
7.2 Mes axes de travail 406

7.3 De ces travaux, premiers acquis méthodologiques 411
7.4 Trois propositions 414
7.4.1 L'opérationnalisation est un axe de travail pertinent 414
7.4.2 Considérer trois échelles de temps est fécond 418
7.4.3 Les quatre perspectives morales 420
7.5 Perspectives sur l'organisation de la recherche 424
7.5.1 Le complexe réseau des disciplines académiques 424
7.5.2 Deux organisations ? 426
7.6 Que puis-je espérer ? 427

A Petit lexique des théories morales **429**

B Liste des articles relatifs à l'effet Knobe **435**

Épilogue **441**

Table des figures **445**

Index des Auteurs **446**

Bibliographie **451**

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse, Anouk Barberousse, pour son soutien constant et pour son engagement. Ce travail lui doit beaucoup.

Je voudrai également remercier tous les acteurs du laboratoire SND, Sciences Normes et Démocratie, professeurs, chercheurs et doctorants, qui font de cette unité de recherche un lieu d'échanges enrichissant et agréable.

Je tiens ici à remercier Daniel Andler qui m'a encouragé à m'engager dans ce parcours et sans qui je n'aurai peut-être pas entrepris cette aventure, bien loin des territoires sillonnés au cours de mes décennies d'ingénierie.

Et, enfin, merci à Marie-Françoise pour son support quotidien, sa compréhension, son écoute et ses nombreuses contributions.

Introduction

En écho aux deux premières des trois grandes interrogations philosophiques, « Que puis-je savoir ? Que dois-je faire ? Que m'est-il permis d'espérer ? », il est d'usage de distinguer, au sein de la philosophie moderne, la philosophie de la connaissance, ou épistémologie, de la philosophie morale, ou éthique. Mon enquête se situera à la croisée de ces deux questions en les associant : « Que puis-je savoir de ce que je dois faire ? ». Il n'est pas indifférent pour cette enquête que la première appellation, philosophie de la connaissance, et non la seconde, philosophie morale, comporte avec la copule « de » la marque de la distance entre la connaissance, aujourd'hui très largement associée aux sciences, et la réflexion philosophique sur la connaissance ainsi acquise. Il n'en va pas de même pour la philosophie morale, à la fois philosophie et morale, qui regroupe le vaste domaine de tout ce qui peut contribuer à répondre à l'interrogation du « Que dois-je faire ? », vaste domaine couvrant les éthiques appliquées, les théories morales qu'elles soient religieuses ou non, et enfin la méta-éthique qui, elle, pourrait être qualifiée, par analogie avec la philosophie de la connaissance, de philosophie de la morale.

La distance ainsi concrétisée par ce petit « de » entre la connaissance et la philosophie de la connaissance témoigne de l'éloignement entre la science et la philosophie. Le développement de la science moderne en occident, à partir du 17^{e} siècle, s'est en effet appuyé sur un partage des rôles entre les sciences d'une part, chargées d'éclairer le « comment » des phénomènes, en restant à distance convenable des préceptes religieux, et ainsi des bûchers, et la religion d'autre part, qui se charge de leur « pourquoi » ainsi que de l'articulation entre nos croyances et les structures de la société, en bonne entente avec les pouvoirs temporels. Ce partage des rôles conduisit la science à s'éloigner également de la philosophie qui ne saurait limiter, elle, son questionnement. Il a en contrepartie permis le développement des sciences naturelles au-delà de tout ce qui aurait pu apparaître comme vraisemblable dans les siècles passés.

Mais plusieurs perturbations sont venues mettre en péril ce fragile équilibre. Tout d'abord, et dès le début de la science, il s'est avéré très difficile d'arrêter la curiosité qui l'anime aux

portes de l'église, du parlement ou du lycée. Observer le mouvement des planètes n'est pas sans remettre en cause des dogmes établis, qu'on le recherche ou non. Et inversement, plus récemment, les succès de la science ont conduit les églises, les dirigeants politiques et les philosophes à souhaiter être « scientifiquement informés ». On voit mal comment une religion pourrait aujourd'hui maintenir de façon non allégorique qu'un Dieu a créé un univers géocentrique. Dans son mouvement, la science s'est affranchie dès l'origine des limites des territoires politiques et religieux, et les règles politiques, morales ou religieuses locales sont alors apparues comme telles, locales, et ce au risque de saper leurs prétentions à l'universalité, aux justifications absolues. Sur un plan plus pratique, le développement des sciences modernes consomme aujourd'hui des ressources importantes qui supposent l'engagement affirmé de sociétés entières, elle ne peut être le fait de groupes restreints. Ces multiples perturbations du partage initial des rôles dans notre modernité occidentale ont conduit à une relation de plus en plus intriquée des multiples sciences avec la philosophie, les religions et les pouvoirs temporels, appelant alors la philosophie des sciences à une tâche importante : décrire cette intrication au fur et à mesure que croissent les connaissances scientifiques et, avec elles, les capacités à interpréter le monde puis à fournir des outils pour le changer. Quand les sciences, par leurs succès comme par leurs coûts, par leurs apports comme par les risques de déstabilisation qu'elles induisent, portent des enjeux à l'échelle des sociétés, alors vient, en miroir, une vaste question : en quoi les connaissances acquises à la lumière de la démarche scientifique moderne peuvent-elles contribuer à la conduite des affaires du monde ? Les hommes des lumières ont pensé que « beaucoup » était la réponse enthousiasmante à cette question, les positivistes ont peut-être pensé que « tout » était la seule réponse sensée, mais les humanistes ont insisté sur la nécessaire conscience qui devait conduire, avant tout, les affaires humaines.

Face à cette question, j'ai longtemps eu le parti pris de l'ingénieur que j'étais, de formation et de profession. Ce parti pris, largement partagé en occident aux 19^{e} et 20^{e} siècles, consiste à développer des savoir-faire pratiques, structurés autour des théories scientifiques, et à les appliquer à des domaines très circonscrits en espérant, à supposer que la question se pose, que l'approche rationnelle des détails pourra contribuer, ou en tout cas ne pas nuire, à la rationalité de l'approche d'ensemble. Pendant de nombreuses années cette approche a été facilitée par son inscription dans une période de développement principalement axé sur les biens matériels dont les bénéfices évidents rendaient oiseuse toute interrogation sur le « pourquoi ». Trois grands courants ont contribué à mettre fin à cette période. Le premier est la fin de l'attitude qu'on peut appeler consumériste ou, plus précisément, la fin de l'évidence de la croyance que toute production supplémentaire d'un bien matériel est justement un bien.

Le deuxième est le développement des connaissances en sciences humaines, dont les sciences cognitives, avec leur potentiel d'action, non plus sur des biens matériels, mais directement sur les comportements humains. Et le troisième est la perception de la modestie de la dimension de l'espace disponible en regard du nombre d'humains, la perception d'une clôture à l'horizon du développement matériel. A la lumière de ces trois ruptures, et en particulier en considérant que le développement des connaissances a aujourd'hui atteint de façon significative la sphère du comportement humain, la séparation entre les deux premières grandes questions de la philosophie « Que puis-je savoir ? » et « Que dois-je faire ? » ne peut plus être étanche. En effet, comment séparer éthique et épistémologie si, imbriquant ces deux questions, je peux acquérir des connaissances sur ce que je dois faire ou, plutôt, si je peux savoir comment les humains pensent, comment ils raisonnent et, en particulier, comment ils raisonnent quand ils cherchent ce qu'ils doivent faire ? Et, par extension, pourquoi ne pas appuyer nos règles morales sur ces connaissances acquises par les sciences humaines, comme nous appuyons nos objets techniques sur les connaissances des sciences de la matière ? Ou faut-il, a contrario, développer face à cette tentation une réticence salutaire pour l'avenir de la planète, et considérer, rejoignant les humanistes, que ces connaissances sur les hommes et sur le monde sont secondes en regard de la conscience à développer avant de les mettre en pratique ?

Face à ces questions, les stratégies sont multiples, tant pour caractériser cette conscience qu'il faudrait développer pour éviter que les actions sociales appuyées sur les sciences cognitives ne produisent des scénarios dignes des plus sombres dystopies, que pour évaluer l'état actuel et prévisible des connaissances relatives au comportement humain apportées par la démarche scientifique. Comment choisir son chemin d'approche entre ces stratégies ? Pour la première étape, la caractérisation de la conscience à développer, la longue tradition de la philosophie morale, dans sa recherche millénaire de la définition du Bien et des moyens pour y parvenir est le point d'entrée qui m'a semblé approprié. Pour la seconde, dans l'impossibilité d'évaluer toutes les sciences dans tous les domaines, j'ai choisi de porter le fer en un point particulièrement sensible, une des sciences sur lesquelles la philosophie morale aurait à s'appuyer si elle visait à être « scientifiquement informée » sur le comportement humain : la psychologie expérimentale.

Limiter ainsi la vaste question posée plus haut, « Que peuvent les connaissances acquises à la lumière de la démarche scientifique moderne pour contribuer à la conduite des affaires du monde ? », à cette nouvelle question, plus précise mais plus étroite, « Qu'apporte la psychologie expérimentale à la philosophie morale ? » est bien sûr réducteur. La question aurait à être posée pour l'ensemble des sciences, comme le montrent à l'envi les débats actuels sur

le réchauffement climatique, et se restreindre à la psychologie pourrait laisser croire que le problème épargne les sciences de la nature. Il n'en est rien mais la présente thèse limitera son périmètre, déjà trop ambitieux, aux sciences utiles à la philosophie morale, dont principalement la psychologie. Réduire également l'étude de la conduite des affaires du monde à la seule philosophie morale est bien sûr, là aussi, abusif. La philosophie politique aurait à bon droit pu prétendre à ce rôle. Mais le choix de la philosophie morale présente l'intérêt de porter l'analyse au cœur d'un débat, le débat moral, dont on peut espérer que la clarification du rapport à la science expérimentale, apporterait, si elle était possible, des retombées sur l'ensemble des affaires humaines.

Autre considération, d'opportunité, justifiant du choix de la philosophie morale, des philosophes se sont emparés des progrès importants réalisés par la psychologie scientifique expérimentale pour lancer à la fin du 20^{e} siècle un mouvement de philosophie expérimentale, dit XPhi (pour Experimental Philosophy, en anglais). Ce mouvement a pour ambition d'instruire expérimentalement des résultats intéressants sur le plan philosophique, dont celui de la philosophie morale. Il a donné lieu à de nombreuses publications et, également, à des réponses diverses de la part des philosophes moraux. Pour certains, les méthodes employées ne sont pas pertinentes et les résultats ne sont pas valides, pour d'autres, les résultats sont peut-être valides mais n'ont pas la portée philosophique que prétendent leurs auteurs, pour d'autres enfin, ce sont des résultats philosophiquement importants qui apportent des éléments significatifs, relativement aux problèmes traités bien sûr, mais également, de façon plus générale, sur la façon de philosopher. Ce sont donc tous les types de relations entre les apports expérimentaux et la philosophie morale que l'analyse du mouvement XPhi nous donne l'occasion, et le besoin, de détailler.

Le mouvement XPhi hérite de la longue tradition des philosophes qui voient dans la démarche scientifique une voie d'accès à des connaissances utiles à résoudre de nombreux problèmes considérés comme philosophiques. Mais les nouveaux ingrédients apportés par les ruptures que j'ai mentionnées plus haut, et en particulier le développement des sciences cognitives d'un côté et la mise en évidence de l'étroitesse de l'espace dans lequel se développe l'espèce humaine de l'autre, changent, pour les tenants de ce mouvement XPhi, le poids qu'il convient de donner à leur approche. Précisons ces deux points, qu'on peut voir comme deux mâchoires qui enserrent la philosophie morale et la poussent vers de profondes révisions. Pour éclairer l'incidence des sciences cognitives, faisons une petite expérience de pensée. Supposons qu'existe demain une nouvelle technique d'imagerie cérébrale qui atteigne la résolution spatiale du neurone et la résolution temporelle nécessaire à suivre l'influx nerveux dans

ces neurones, c'est-à-dire une imagerie qui pourrait suivre la dynamique de l'enchaînement de la sollicitation des neurones à l'unité près [1]. Appelons cette nouvelle technique l'Imagerie Neuro Cognitive, l'INC, puisqu'elle permettrait de visualiser les processus de la cognition au niveau des neurones, elle donnerait accès à la dynamique de la pensée [2]. Avec cette INC, il serait possible d'analyser les différents raisonnements moraux que tiennent différentes personnes face à un dilemme moral. Et cela ne manquerait pas d'ouvrir de nombreuses et difficiles questions. Quelles incidences sur les débats moraux auraient de tels résultats ? Et faudrait-il même seulement que les philosophes moraux prennent en compte de telles descriptions neuronales comme pertinentes pour comprendre un raisonnement moral ? Quels dilemmes seraient considérés comme moraux et dignes d'intérêt ? Et, plus immédiatement, s'ils le doivent, comment les philosophes moraux peuvent-ils se préparer dès aujourd'hui à l'arrivée de cette INC ? Pour les philosophes expérimentaux, il est temps de se préoccuper de ces questions car elles se sont déjà concrétisées tout au long du 20^{e} siècle avec le développement de la psychologie scientifique, elles ont été renouvelées avec l'arrivée de l'imagerie à résonance magnétique fonctionnelle (IRMf) qui, sans atteindre les performances imaginaires de l'INC, a déjà apporté de nouveaux éclairages sur le fonctionnement du cerveau. Et à chaque fois, bénéficiant de ces apports, la compréhension des processus cognitifs a fait de grands pas en avant. Les philosophes moraux ont, à retardement, pris en compte ces résultats. Nombreux sont les philosophes moraux qui les ont minimisés, arguant de leur imprécision, de leurs insuffisances face à la complexité du comportement humain. Les sauts qualitatifs et quantitatifs à venir sont beaucoup plus importants, et il est douteux que cette stratégie du déni de pertinence puisse perdurer. Il faut, disent les philosophes expérimentaux, se préparer immédiatement aux résultats qu'apporteront les futures INC de cette expérience de pensée et surtout, plus largement et plus immédiatement, aux résultats qu'apportent et apporteront les sciences cognitives. Et la meilleure façon de se préparer est de participer à ces expérimentations.

Le développement des connaissances relatives au comportement humain constitue donc l'une des pressions qui s'exercent sur les philosophes moraux, la seconde que je veux pointer maintenant est celle de la mondialisation, la perception de l'étroitesse de l'espace disponible pour l'espèce humaine. Les morales locales ont permis par le passé de réguler les comporte-

1. Pour mémoire, l'influx nerveux a une vitesse de l'ordre de 10 à 100 m/s, et un axone a une longueur de l'ordre de 1mm à 1m. Les techniques actuelles sont loin des performances qui seraient nécessaires pour que se réalise cette expérience de pensée, elles ne permettent pas de voir simultanément les neurones individuels et l'ordre dans lequel ils sont activés.

2. Le laboratoire de neuro imagerie cognitive est, lui, une réalité. C'est un laboratoire animé par Stanislas Dehaene et non une expérience de pensée. Voir `http://www.paris-neuroscience.fr/fr/equipe/laboratoire-de-neuro-imagerie-cognitive`.

ments individuels et collectifs au sein des communautés restreintes qui constituaient l'environnement des sociétés humaines. Cette régulation par les règles morales avait pour composante de marquer l'appartenance des individus à un groupe, de resserrer leur solidarité, et ce, pour partie, en les opposant aux individus extérieurs aux groupes, à ceux qui ne respectent pas les mêmes règles. Les conflits moraux étaient alors, de ce fait, principalement internes aux groupes, entre individus faisant partie de la même communauté morale. Les conflits externes, les guerres offensives ou défensives, n'étaient pas à inclure dans le domaine moral. Ils se traduisaient par une dégradation du statut du vaincu, une déshumanisation, pouvant conduire à l'esclavage ou au massacre sans que les lois morales internes au groupe ne soient violées. Naturellement, l'histoire n'est pas faite de scénarios aussi tranchés, et on peut penser qu'il y a toujours eu également des voix humanistes pour reconnaître des droits aux vaincus. Mais ce schéma qui sépare fortement, sous l'angle moral, l'intra de l'inter-groupe a surtout été remis fondamentalement en cause par l'universalisme des lumières au 18^{e} siècle, puis, perception de l'étroitesse de la planète aidant, par l'interpénétration et l'interdépendance généralisées de toutes les populations mondiales au 20^{e} siècle. Aujourd'hui, certes il reste des conflits moraux au sein des différentes communautés morales, mais l'urgence est à imaginer des modes de traitement des conflits moraux entre personnes ou institutions qui ne se reconnaissent pas les mêmes règles morales, les mêmes règles politiques ou religieuses. Or les règles morales existantes se sont, pour des parties importantes, structurées les unes contre les autres, aux antipodes de toute possibilité de conciliation, et la philosophie morale, principalement construite sur la base des théories morales locales, est confrontée à la nécessité d'évoluer pour faire face aux nouveaux défis globaux.

L'urgence qui pèse sur l'évolution du domaine moral a donc une double source, d'une part le constat que les morales en place n'ont pas, ou pas toutes, la capacité à servir la cohésion et la régulation du monde globalisé actuel, et d'autre part, l'interrogation sur la capacité des sciences cognitives à être utiles pour relever le défi posé par la nécessité de cette nouvelle régulation globale sans perdre l'ancienne régulation locale. La philosophie morale expérimentale est actuellement marginale au sein de la philosophie académique, et affirmer qu'elle est en développement rapide serait audacieux. Néanmoins, il me semble qu'elle apporte un éclairage à prendre en compte sur la possibilité d'envisager cette hybridation entre sciences cognitives et philosophie morale que la situation d'urgence morale appelle, et ce tant par ce qu'elle apporte, que par les réticences qu'elle soulève.

La vaste question initiale, « Que peuvent les connaissances acquises à la lumière de la démarche scientifique moderne pour contribuer à la conduite des affaires du monde ? », a donc

au fil de ces réflexions perdu en généralité pour se transformer en « Que peut la psychologie expérimentale pour la philosophie morale ? », mais elle y a gagné une possibilité d'approche, l'examen du mouvement XPhi, de la philosophie morale expérimentale, et surtout une forte pression : le 21e siècle globalisé ne pourra survivre avec les morales du 19e, et dans l'analyse du petit nombre de voies de solutions envisageables s'inscrit l'utilisation des connaissances acquises et à acquérir sur le comportement humain.

Plan de la thèse

Chapitre 1

Le premier chapitre a principalement pour objectif de préciser ce qu'est le phénomène moral, ce que sont les théories morales, et, c'est mon objectif initial dans cette approche, d'avancer ainsi vers une caractérisation des désaccords moraux. Je propose dans ce chapitre plusieurs outils de nature à faciliter la cartographie du domaine moral sous la double contrainte que ces outils ne soient pas circulairement trop dépendants des théories morales qu'ils sont supposés cartographier et que la cartographie obtenue soit ensuite utile pour positionner les éventuels apports des sciences expérimentales à la résolution, ou plus modestement à l'éclaircissement, des désaccords moraux.

Chapitre 2

Le deuxième chapitre a pour objet principal de décrire le mouvement XPhi de philosophie expérimentale, sa constitution, les résultats qu'il a produits et les réticences qu'il a soulevées. Ce mouvement est minoritaire au sein de la philosophie, mais il est important pour mon analyse car il est certainement une des tentatives récentes les plus avancées dans l'objectif d'utiliser les démarches expérimentales importées des sciences naturelles pour traiter de questions philosophiques. Le mouvement a concerné, et concerne, toutes les sous-disciplines philosophiques et, à ce titre, peut être vu comme un regroupement par affinité de méthodes, et non comme un corpus homogène par son sujet. Néanmoins, on peut distinguer deux types de projets au sein de ce mouvement. Le premier, générique, dit programme positif, consiste, de façon très large, à accepter l'expérimentation héritée de la psychologie expérimentale pour outil complémentaire à disposition des philosophes. Le second, dit programme négatif, est plus ciblé. Il vise à montrer que les méthodes habituelles des philosophes analytiques manquent de fiabilité. Les réticences soulevées par le programme négatif, qui s'attaque à l'ensemble de la philosophie analytique, sont très importantes. Le programme positif est plus consensuel sur son objectif général mais pose un problème pratique : les philosophes sont-ils réellement équipés pour mener des expérimentations scientifiques ?

Chapitre 3

Le chapitre trois fait un pas de côté, après la présentation du domaine moral et de la philosophie expérimentale. Il a pour objet de présenter la vision de la démarche scientifique expérimentale que j'adopte pour la présente thèse. Il est naturellement hors d'atteinte de ce travail de traiter, même minimalement, de ce que signifient exactement d'être expérimentale pour une démarche ou d'être qualifiable de scientifique pour une expérience. Je procède donc de façon imagée en proposant une métaphore, l'hélice expérimentale, qui marque le caractère dynamique de la démarche scientifique expérimentale, sans début, sans fin et sans fondement autre que la validation des pairs. Appuyée sur les travaux de Ian Hacking et sa proposition de considérer les idées, les objets des laboratoires et les traces issues des expériences comme les ingrédients de base des travaux scientifiques, la métaphore permet également de mettre en avant trois temps de la démarche, l'opérationnalisation qui va des idées aux expériences, l'objectivation qui va de l'expérience singulière à la trace partagée par la communauté des scientifiques et l'interprétation inductive qui va des traces aux idées, bouclant ainsi un tour de l'hélice expérimentale. Cette métaphore me permet de lever quelques hypothèques prises sur la démarche scientifique au temps du positivisme, hypothèques qui restent bien présentes dans les débats des philosophes moraux. La démarche scientifique, itérative, n'a pas vocation à fournir de réponses absolues, mais seulement la (ou les) bonne réponse disponible. Les pairs, la communauté des scientifiques, établissent ce que sont les bonnes réponses à un certain moment du développement des sciences. De nombreuses considérations métaphysiques sont donc ainsi hors de portée de la démarche scientifique expérimentale métaphoriquement représentée par une hélice, ce qui ne l'empêche pas de faire avancer l'embarcation de la connaissance.

Chapitre 4

Disposant à la fois d'un terrain, la philosophie morale, d'un outil, la démarche scientifique expérimentale, et d'un exemple, le mouvement XPhi, le quatrième chapitre a pour objet d'aborder la question de la possibilité de l'apport de la démarche expérimentale à la philosophie morale par une mise en abyme : expérimenter sur l'expérimentation. C'est avec cinq études de cas qui mobilisent, chacune différemment, les rapports entre approches expérimentales et débats de philosophie morale que je me propose d'éclairer la question posée.

La première étude de cas porte sur les dilemmes sacrificiels, quand sacrifier une personne permet d'en sauver plusieurs. Ces dilemmes ont alimenté la philosophie morale depuis de nombreuses années, et en particulier le dilemme dit du tramway, imaginé dans les années 1960 par Philippa Foot, a donné lieu à des expériences importantes pour le mouvement XPhi avec, en 2001, la première utilisation par Joshua Greene de l'IRMf dans le cadre d'un ar-

gument philosophique. Détailler ce cas est donc important pour préciser les apports de la philosophie expérimentale et les réticences qu'elle a soulevées.

La deuxième étude de cas fait suite à un doute récurrent : quel crédit donner à des études qui ne s'appuient que sur les déclarations de personnes sommairement sollicitées par un questionnaire ? Apprécier ce doute est important car une part très majoritaire des expériences menées par les psychologues et utilisées par les philosophes moraux est de ce type. L'étude a consisté à choisir un cas, un article annonçant que les Européens, et les Français tout particulièrement, surestimaient le nombre de musulmans dans leur pays, et à réaliser un questionnaire reprenant cette problématique, à l'appliquer et à le dépouiller. Bien que l'étude, qui se poursuit encore aujourd'hui avec de nouvelles hypothèses à tester, ait confirmé la surestimation, il apparaît que son interprétation est loin d'être évidente.

La troisième étude de cas porte sur la question des études faisant appel à des techniques sophistiquées comme l'IRMf. Le cas étudié est celui d'une suite d'articles analysant la possibilité que des émotions traduites par des larmes se transmettent par l'olfaction. Il met en jeu deux équipes de chercheurs, l'une spécialisée dans la psychologie des larmes et l'autre dans l'étude de l'olfaction. Les difficultés de la collaboration entre les deux équipes et le constat que, dans le cas étudié de la transmission des émotions par les larmes, les deux équipes se tournent mutuellement le dos, sont analysés en regard de la cohabitation de techniques complexes dont la mise en œuvre nécessite de vastes collaborations de différentes équipes de spécialistes en confiance.

La quatrième étude concerne l'IAT, l'Implicit Association Test, et tente d'éclairer une question préalable à l'utilisation des résultats des scientifiques hors de leur champ de recherche : comment s'appuyer sur des résultats scientifiques si des controverses montrent des divergences entre les équipes de scientifiques ? La psychologie étant un champ notoirement fertile en controverses, comment envisager que les conclusions ainsi débattues puisse être utiles aux philosophes moraux ? Le cas de l'IAT est intéressant à cet égard. Il s'agit d'un test dont l'objectif est de mesurer si une personne a, implicitement, des tendances racistes qu'elle ne se reconnaît pas, explicitement. Les controverses autour de ce test et de son utilisation dans la lutte contre les ségrégations aux États-Unis ont donné lieu à de multiples études, de multiples expérimentations et réplications puis, également, à plusieurs méta-analyses menées tant par les promoteurs du test que par ses détracteurs.

La cinquième étude de cas a pour objet une préoccupation liée à la multiplication des études de philosophie morale expérimentale et à l'impression d'instabilité notionnelle qui s'en dégage. Les notions importantes pour le philosophe moral, comme la responsabilité, l'in-

tentionnalité, le libre arbitre, etc. semblent changer de signification à chaque nouvelle étude, rendant illusoire la construction d'un vocabulaire partagé entre chercheurs. Pour tenter de préciser ce sentiment confus, j'ai sélectionné un des articles les plus commentés du mouvement XPhi, celui de Joshua Knobe présentant l'effet qui porte maintenant son nom. Il s'agit de montrer que l'attribution d'intentionnalité dépend de la valeur morale du résultat obtenu. J'ai ensuite inventorié, avec un objectif d'exhaustivité, tous les articles citant cet effet Knobe et publiés entre deux dates. Avec cette étude de cas, c'est la difficulté pour une science aussi complexe que la psychologie d'entrer dans une démarche expérimentale scientifique, au sens de la métaphore de l'hélice, en partant des concepts qui proviennent de la philosophie qui est analysée.

Ces cinq études de cas n'ont aucune prétention à avoir balayé exhaustivement le terrain des expérimentations en psychologie morale et de leurs relations à la philosophie morale. Elles constituent, en ce sens, une expérimentation sur l'expérimentation et devront être, en application de la métaphore de l'hélice, objectivées pour être partagées, ce que j'ai essayé de faire dans ce chapitre, puis interprétées inductivement pour voir si elles peuvent être utiles à émettre des hypothèses d'un certain niveau de généralité qui suggéreront à leur tour d'autres études de cas. Mais avant d'aller vers ce niveau plus élevé d'interprétation, j'approfondis un point particulier, l'opérationnalisation, au chapitre suivant car de façon récurrente dans les cinq études des cas apparaît comme essentielle cette étape qui permet de passer des entités, propriétés et relations postulées par les théories à des terrains d'expérimentation où des phénomènes en lien avec ces éléments postulés peuvent être détectés.

Chapitre 5

Le chapitre cinq a pour objet l'examen de l'étape d'opérationnalisation. Le problème à résoudre lors de cette étape est, pour le psychologue, de trouver pour chacune des notions qu'il utilise au quotidien, comme des croyances, des intentions, des raisonnements, toutes notions indispensables à la formulation de ses hypothèses et théories, des contre-parties qu'il puisse détecter lors de ses expériences. Il doit trouver des observables face à ces notions qui sont inobservables directement.

En repartant des cinq études de cas, et en précisant pour chacune la place de la problématique de l'opérationnalisation, je montre l'importance de cette étape dans la situation actuelle de la psychologie expérimentale en regard de l'utilisation de ses résultats par les philosophes moraux. Le risque de confusion est en particulier mis en avant, c'est-à-dire le risque de prendre pour observable associé à une notion, un phénomène qui peut être interprété de plusieurs façons, dont certaines conduisent à mettre en doute le lien entre l'observable et

la notion étudiée. Ce risque de confusion se décline en de multiples exemples qui font des expériences de psychologie un chemin particulièrement difficile à pratiquer.

Chapitre 6

Le chapitre six, adoptant le point de vue des philosophes moraux sceptiques quant à l'approche des philosophes expérimentaux, a pour objet de reprendre leurs principales objections sur trois grands axes. Le premier est celui de la complexité, le comportement humain est trop complexe et trop divers pour que des expériences menées sur quelques personnes puissent être significatives et qu'on puisse en déduire des conclusions valides et utiles aux philosophes. Le second est celui de l'impossibilité morale de telles études qui placent celui qui les mène dans une position de surplomb moralement injustifiable. Le troisième est celui de leur nocivité sociale. Les sociétés reposent, pour partie, sur les règles morales qui en assurent la cohérence, remettre ces règles en cause en les soumettant au doute systémique du scientifique, c'est leur enlever toute possibilité d'efficacité. Ce n'est pas expliquer le domaine moral, c'est le détruire.

Les réponses des scientifiques à ces réticences sont de plusieurs types, et le chapitre se donne pour objectif d'en restituer une topologie. Une attention toute particulière est portée aux théories évolutionnistes qui proposent des explications du phénomène moral. Ces théories postulent que l'espèce humaine a des propriétés particulières, dont la sociabilité et la rationalité, qui font qu'un équilibre, difficile, doit être trouvé entre les individus et le groupe. La morale est alors comprise comme une réponse de l'évolution à cette difficulté.

Ces théories ont un pouvoir explicatif important, elles soulèvent néanmoins de nombreuses questions qu'il faudra préciser. Le schéma explicatif implique un entrelacs de trois temps, celui de l'espèce humaine et de l'évolution biologique dans les processus de spéciation, celui des groupes humains et de l'évolution sociale dans les processus historiques, celui des individus et de l'évolution personnelle dans le processus d'individuation. Chacun de ces processus est étudié par des équipes de chercheurs différentes, biologistes, sociologues et psychologues, et il semble donc que l'étude du domaine moral en s'appuyant sur les théories évolutionnistes doive conduire à de vastes équipes multidisciplinaires. Expérimenter sur le domaine moral nécessiterait que ces équipes se mettent d'accord sur les opérationnalisations permettant de traiter les notions en cause, ce qui ne semble pas, à l'issue des cinq études de cas, à portée. En revanche, il apparaît avec ces éclairages, et c'est ce que je détaillerai, qu'on peut accéder à une simplification fructueuse en dissociant d'un côté la nécessité de la règle morale et de l'autre le contenu de la règle. Ainsi l'existence des règles morales serait expliquée par les théories évolutionnistes proches de la biologie, pour partie indépendamment de leur contenu,

et le contenu précis de la règle par les théories sociales, pour partie indépendamment de leur existence. Et, de la même façon, les règles morales existant, la psychologie se donne pour objet l'individu qui se construit par acceptation des unes et refus des autres en s'intégrant dans les groupes qu'il constitue autant qu'il est constitué par eux. Comme toute idéalisation, cette séparation est certainement imparfaite, mais, ce sera ma proposition, peut permettre de structurer des démarches scientifiques expérimentales fécondes, mais locales.

Chapitre 7

Le dernier chapitre fait le point de toutes les propositions faites dans les chapitres précédents et propose un certain nombre de perspectives qui sont ouvertes par ces propositions. Et, enfin, un épilogue présente un cadre programmatique au travers d'une fiction, celle d'Homo Fictionalis, ce primate étrange qui a inventé les histoires, puis inventé des histoires sur les histoires, dont celle qu'existeraient certaines histoires qui disent le vrai, d'autres qui disent le beau, et d'autres encore qui disent le bien. Adopter ce cadre programmatique n'est pas une réponse à la question posée dans cette thèse, « Qu'apporte la psychologie expérimentale à la philosophie morale ? », mais ce pourrait être un moyen de rendre possibles des contributions plus riches des psychologues expérimentaux, des psychologues de l'évolution et des philosophes moraux aux perspectives descriptives et méta-éthiques sur le domaine moral soumis aux deux pressions réformatrices des sciences cognitives et de la mondialisation.

Chapitre 1

Morale et philosophie morale

Ce premier chapitre a pour objectif de caractériser ces questions morales que la démarche scientifique expérimentale pourrait, c'est ce que je cherche à analyser dans la présente thèse, contribuer à instruire. Dans un premier temps, pour contourner la très grande difficulté qu'il y a à s'accorder sur ce qui relève du domaine moral indépendamment d'une théorie morale particulière, je me propose de caractériser le caractère moral d'une situation par la présence des énoncés évaluatifs. Un énoncé évaluatif est une proposition comportant un des termes considérés tels : « Bien », « Mal », ... Le domaine moral est alors constitué de l'ensemble des situations humaines ayant donné lieu, réellement ou potentiellement, à un énoncé évaluatif. Il s'agit, plus précisément, d'un sur-ensemble du domaine moral qui a pour mon travail l'avantage d'embrasser dans le même périmètre tous les penseurs moraux, y compris ceux qui, comme les défenseurs du nihilisme moral, pensent qu'il n'existe pas de faits moraux réels objectifs indépendants de ce que nous en pensons.

La définition de la théorie morale que je propose de retenir ensuite est habituelle, principalement constituée d'une part de la définition du Bien à rechercher et, d'autre part, de règles morales à respecter pour l'atteindre. L'objectif d'une théorie morale est d'aider l'acteur moral à répondre à la question du « que dois-je faire ». La présentation se poursuit par une approche des trois grandes familles de théories morales habituellement reconnues, le déontologisme, le conséquentialisme et l'éthique des vertus. Pour couvrir l'ensemble des positions des penseurs moraux, je complète cette rapide présentation avec des théories selon lesquelles l'objectif des théories morales est inatteignable, comme l'affirment le nihilisme moral et la théorie de l'équilibre réfléchi.

Mon objectif n'est pas ici d'entrer dans l'ensemble des débats que soulève la recherche d'une théorie morale idéale mais, plus schématiquement, de proposer une caractérisation

du domaine moral qui se veut instrumentale dans le cadre de cette thèse. Des philosophes moraux ont proposé des critères de comparaison des théories morales, mais ces critères se révèlent, à l'examen, profondément circulaires : leur évaluation dépend de considérations morales liées à la théorie morale retenue comme cadre de réflexion. J'écarterai donc l'utilisation de ces critères au profit d'outils plus neutres appuyés sur une analyse des différents types de désaccords moraux qui structurent les débats entre personnes morales ou entre philosophes moraux. J'ai retenu une métaphore spatiale [1] pour structurer l'espace des désaccords entre philosophes moraux selon cinq embranchements à partir de la base commune du domaine moral issue des énoncés évaluatifs.

Je propose ensuite de distinguer quatre postures que peut adopter le penseur moral, quatre perspectives qui lui permettent d'aborder les différents aspects de domaine moral. La première est descriptive, ce que les humains font. La deuxième est prescriptive, ce que les humains devraient faire en application d'une théorie morale. La troisième est celle du philosophe adepte de la méta-éthique qui cherche à comprendre le phénomène moral et évaluer les théories morales. Enfin la quatrième est celle du praticien confronté à la décision morale dans un contexte particulier. Ma proposition sera que la clarification apportée par la prise en compte de ces quatre perspectives constitue un outil pertinent pour analyser l'apport potentiel de la démarche expérimentale à la philosophie morale. C'est relativement à chacune de ces postures, de ces perspectives, que, dans les chapitres suivants, j'examinerai cet apport dans les différents cas d'expériences que je présenterai.

Pour donner corps à ce cheminement, je parcours ces embranchements avec deux exemples de théories récentes aux antipodes du monde moral, l'une est le réalisme moral proposé par David Enoch, philosophe défenseur de la réalité des propriétés morales objectives, et l'autre la théorie du sentimentalisme constructif de Jesse Prinz, représentative d'une théorie habituellement considérée comme antiréaliste (c'est-à-dire qu'un fait n'a pas de propriétés morales propres, en dehors de ce que nous en pensons). Conjointement, ces deux théories décrivent deux types de théories morales très différentes qui, si elles peuvent être utilement décrites avec les cinq embranchements proposés, valideront cette approche des désaccords moraux entre philosophes, toujours en privilégiant l'analyse de l'apport potentiel de l'expérimentation.

Mon ambition est que cette description des désaccords moraux et cette proposition des quatre perspectives sur le domaine moral, constituent un outil utile à l'analyse du potentiel apport expérimental à la philosophie morale. La question posée devient ainsi à l'aide de ces

1. Métaphore spatiale pour partie empruntée à (Ravat 2019) [189]

outils :

> Selon quelle perspective et pour aider à instruire quel type de désaccord moral une expérimentation signifiante est-elle possible ?

Mon point de vue n'est pas ici celui d'un spécialiste de la philosophie morale, il est plus modestement de dépeindre le domaine qui sert de toile de fond à mon analyse de l'apport potentiel des démarches expérimentales. Il ne vise donc aucunement à une description exhaustive du domaine moral, ce qui serait de toutes façons hors de portée, mais principalement à proposer quelques distinctions qui permettent d'établir une première cartographie utile à l'examen des apports potentiels de la démarche scientifique expérimentale aux débats moraux.

1.1 Le domaine moral

Un exemple paradigmatique

Un homme passe à côté d'un étang, il entend les appels d'un enfant qui se noie. Que va-t-il faire ? Se précipiter au secours de l'enfant et être alors en retard à son rendez-vous et, peut-être même, abîmer dans l'eau son habit et ses chaussures, ou passer son chemin en abandonnant le petit noyé à son destin ?

Face à une telle situation, nous n'avons guère de doute, l'homme a une obligation morale à se porter au secours de l'enfant. Un retard dans sa journée, la perte d'un habit ou d'une paire de chaussures, ne sont rien en regard du drame de la noyade d'un enfant. Ne pas le faire serait immoral et, dit-on, inhumain. [2]

Définir le domaine moral [3] consiste, pour toute personne intéressée à l'étude du comportement humain, à tenter de reconnaître dans une situation paradigmatique comme celle de l'homme passant à côté de l'étang, à laquelle on accorde une dimension morale, ce qui fait qu'on lui attribue cette dimension.

Une telle définition du domaine moral est difficile à préciser et beaucoup considèrent en pratique ce domaine comme indéterminable [4]. Plusieurs stratégies sont néanmoins possibles.

La première stratégie, que je viens d'employer, consiste à donner un exemple exempt de

2. Cet exemple est utilisé par Peter Singer dans son article de 1972 et régulièrement repris dans les articles de philosophie morale (Singer 1972) [210].

3. En suivant la tradition de la philosophie analytique, je considère les termes « éthique » et « morale » comme synonymes, chaque philosophe ayant introduit des nuances différentes entre ces deux termes qu'il sera nécessaire de spécifier au cas par cas si besoin.

4. Par exemple Christine Tappolet dans (Tappolet 2000 p14) [225] : « Il faut se rendre à l'évidence qu'il n'est pas aisé de déterminer ce qui spécifie le domaine moral » ou Geoff Sayre-McCord dans (Sayre 2017) [202] : « Ceci dit, il est étonnamment difficile de pointer avec précisions ce qui relève ou non du domaine moral. »

toute ambiguïté apparente et d'en déduire l'existence du phénomène moral comme évidente. Ce phénomène est celui de l'altruisme désintéressé qui consiste à entreprendre une action coûteuse sans autre raison que, justement, morale. Profitant de l'élan donné par cet exemple évident, on poursuit en soulignant que l'étude de ce phénomène moral est importante pour comprendre le comportement humain et que c'est cette étude détaillée qui précisera plus tard les contours du domaine moral [5]. L'étude détaillée proposera ensuite une conception de ce qu'est la morale et, par contre-coup, contribuera à définir le domaine moral dans le cadre de cette conception. En somme, point n'est besoin de définition précise pour commencer à étudier le domaine moral, quelques exemples suffisent car le phénomène est omniprésent et nos intuitions sont fortes en la matière. [6]

Ce qui fait d'une situation morale qu'elle est morale

Seconde stratégie, et toujours en s'appuyant sur la force spontanée de quelques exemples évidents d'actes altruistes, il est possible d'analyser ces exemples non ambigus de situations morales selon de multiples facettes et de rechercher la source du caractère moral de la situation dans chacune de ces facettes. On peut ainsi, en reprenant l'exemple de l'homme passant à côté de l'étang, proposer que la dimension morale soit apportée par la présence simultanée d'une victime potentielle, l'enfant, et d'une personne qui apparaît maître de ses actes, personne qui sera considérée moralement responsable car elle peut intervenir ou non. La présence d'une victime et d'un acteur responsable semble d'un poids fort pour caractériser cette situation de « morale ». On peut également, autre dimension, souligner la nécessité de la prise de décision : l'homme a son libre arbitre et, seul, décide, ou non, de se mettre à l'eau et c'est cette liberté face à lui-même qui donne son caractère moral à la situation. Pour montrer qu'il en est bien ainsi, on peut s'appuyer sur une variante de la scène où l'homme n'est plus seul face à l'enfant car de multiples témoins sont présents. L'homme se jetterait alors à l'eau dans l'objectif de ne pas se voir critiqué par les témoins, ou même accusé de non assistance à personne en danger, et son geste nous apparaîtrait peut-être comme moins déterminé par les seules raisons morales. L'homme serait simplement prudent, face au risque de procès.

On peut, autre axe d'analyse, s'intéresser à la personne qui juge moralement la situation, ici le lecteur de la présente thèse, et à l'importance de la réaction émotive que la scène, avec ou sans intervention du passant pour sauver l'enfant, déclenche chez lui. Cette réaction est, dans la formulation que j'ai retenue ci-dessus, obtenue par des effets de langage. Les expressions « les appels d'un enfant qui se noie » ou « abandonner le petit noyé » renvoient à des images

5. Nicolas Baumard (Baumard 2010) [20] ou Vanessa Nurock (Nurock 2011) [173] commencent ainsi leurs ouvrages par un exemple sensé établir un accord sur le sujet traité.

6. C'est par exemple la thèse intuitionniste de Michael Huemer dans (Huemer 2008) [121].

à forte charge émotionnelle, et le lecteur est de ce fait amené à une interprétation morale de la situation. L'examen de cet exemple peut suggérer, sous cet angle, que nous attribuerions la dimension morale du fait, au moins partiellement, de ces émotions que la description a induites et non uniquement du fait des caractéristiques de la situation. Une formulation plus froide de la scène, à l'image par exemple d'un rapport de police, aurait alors pour effet de nous éloigner du domaine proprement moral et de nous orienter vers d'autres réactions, comme par exemple un questionnement plus analytique sur le comportement de cet homme : Est-il sourd ? ou mal voyant ? ou a-t-il une peur irrépressible de l'eau ? (. . .) toutes questions qui n'affleurent pas à l'esprit quand, emportés par nos émotions, nous jugeons spontanément détestable celui qui a refusé de mouiller son costume pour sauver un enfant.

En prenant une certaine distance avec la situation décrite, on pourrait également remarquer que ce type de situations où le héros sauve une personne en danger est d'une très grande fréquence dans de nombreuses fictions, et ce dans toutes les cultures. On peut alors proposer que c'est par un réflexe culturel acquis grâce à l'exposition à ces multiples fictions que nous attribuons la dimension morale à cette situation. La morale serait ainsi un façon de nommer les habitudes culturelles quand elles touchent à des actions que nous sommes en situation de juger en conformité avec cette culture acquise.

Du poids que chacun accordera à chacune de ces facettes, responsabilité, autonomie, émotions induites, contexte culturel, etc. pour considérer que l'exemple est paradigmatique d'une situation morale, résultera une évaluation différente de ce qui compte pour être moral dans la situation et, de proche en proche, une définition différente du phénomène moral. Ainsi, chaque philosophe redéfinit le domaine moral aux couleurs de sa théorie : pour Jonathan Haid, ce qui est moral est ce qui est lié aux émotions (Haidt 2013 p 30) [116], pour Katinka Evers, c'est ce qui sépare « eux » de « nous » (Evers 2009 p 128) [81], pour Michael Huemer, c'est ce que nous savons intuitivement être moral (Huemer 2008) [121], et pour Ruwen Ogien, c'est précisément parce que la morale n'est pas définie que chacun peut lui donner un contenu différent et que les conflits moraux surgissent (Ruwen Ogien, p 201 dans Masala 2013) [159].

Des définitions circulaires

Mais, en supposant que nous puissions choisir une des facettes de la situation comme porteuse de son caractère moral, la difficulté de la définition du domaine moral ne sera pas pour autant résolue. Supposons par exemple que nous prenions la dimension émotionnelle comme importante : ce qui définit le domaine moral est la réaction émotionnelle qu'a le témoin d'une scène dans laquelle un acteur se comporte de façon inadmissible [7]. A l'évidence, cette

7. Par exemple, l'ouvrage de Jesse Prinz « The emotional construction of morals » a pour objet central de lier

définition n'est pas suffisante car il existe des émotions que nous ne qualifions pas de morales. Il nous faut donc maintenant aller plus loin et préciser en quoi telle émotion est morale et telle autre ne l'est pas. Trois voies s'offrent à nous. Soit simplement affirmer qu'il existe des émotions morales et d'autres non morales, et accepter que la définition devienne ainsi circulaire. Soit préciser quelles émotions sont morales et lesquelles ne le sont pas, ce qui est un problème aussi difficile que le problème initial que nous cherchions à résoudre. Une troisième voie serait de proposer que les émotions morales sont le fait d'un « sens moral »[8] qui, à l'instar de la perception, nous permettrait d'accéder directement à la dimension morale de la situation. Chacune de ces trois possibilités est problématique et donne lieu à des débats complexes dont aucun n'est résolu à ce jour.

Ce qui est vrai pour les émotions le sera également pour chaque facette envisagée, la difficulté de la définition ressurgit selon la même dynamique. Si c'est la situation qui est choisie pour définir la moralité : pourquoi une situation serait-elle morale (un enfant se noie) et une autre non (une mouche se noie)? Si c'est la prise de décision qui définit la morale : faut-il distinguer l'homme qui sauve l'enfant spontanément, sans réfléchir, de celui qui pense pouvoir tirer un bénéfice de son action, et alors qu'est-ce qui fait que l'une est morale et l'autre non ? Si c'est le jugement moral du témoin de la scène qui compte, qu'en est-il d'une action sans témoin ? Faut-il alors réduire le domaine moral aux cas où il y a un témoin ?...

Chaque facette choisie comme caractérisant le domaine moral offre ainsi son lot de problèmes en série qui, comme la cascade d'Escher, ramène en boucle vers le problème initial : on ne peut définir le domaine moral en analysant une caractéristique de la situation sans avoir à préciser ce qui est moral dans cette caractéristique, et il faudrait déjà disposer d'une définition de la morale pour cela. Ce constat de l'impossibilité qu'il y a à réduire le domaine moral à la présence d'une caractéristique particulière de la situation qui soit une propriété non morale a été proposé par GE Moore sous le nom d'argument de la question ouverte : même après avoir reconnu une situation comme ayant ces propriétés non morales, il semble toujours possible de se demander « Mais en quoi cette situation est-elle morale ? », et puisque cette question fait sens, c'est qu'il est impossible de ramener ainsi analytiquement une fois pour toutes le domaine moral à la présence de telle ou telle propriété non morale.

La morale comme raison d'agir

Une autre approche classique de la définition du domaine moral est de l'appuyer sur l'analyse des raisons d'agir. Observons donc les nombreuses façons de décrire aux autres et à nous

morale et émotions (Prinz 2009) [185].

8. Pour une défense récente des philosophies du sens moral, voir la thèse de Nicolas Baumard (Baumard 2008) [19].

mêmes comment nous sommes parvenus à décider d'une action et, en particulier, sur quels jugements, moraux ou non puisque c'est la question, nous nous sommes appuyés. Nous pouvons par exemple décliner nos raisons d'agir en raisons pratiques, raisons conventionnelles et raisons morales. Les raisons pratiques sont celles qui nous conduisent à répondre à la question du « que dois-je faire ? » pour que les résultats pratiques soient conformes au résultat attendu. Par exemple, si j'ai soif et si je crois qu'il y a de l'eau au robinet et que boire cette eau va étancher ma soif, alors j'ai une bonne raison pratique de me lever, d'aller au robinet et de boire son eau.

Les raisons conventionnelles nous poussent à décider et agir dans le sens de la conformité aux règles des groupes sociaux auxquels nous appartenons. Par exemple, il n'est pas d'usage d'aller en ville en maillot de bain, même pendant les grosses chaleurs [9] et nous avons donc une raison conventionnelle de nous habiller pour sortir dans la rue.

Au delà de ces raisons pratiques et conventionnelles, qui ne sont pas habituellement considérées comme des raisons relevant du domaine moral [10], la morale s'intéresse aux raisons d'agir qui ont pour but le Bien. Il s'agit alors de faire le Bien [11], de se comporter Bien, de viser à être quelqu'un de Bien.

La définition de la morale que suggère l'approche par les raisons d'agir peut apparaître comme marginale : sont morales les raisons d'agir qui restent quand on a enlevé les raisons pratiques ou conventionnelles. Dans une formulation un peu différente, on pourrait qualifier de morales les décisions difficiles qu'une délibération portant simplement sur les conséquences pratiques ne peut trancher. Plus généralement, chaque philosophe pourra imaginer un ensemble de familles de raisons d'agir, pratiques et conventionnelles mais aussi par exemple religieuses, prudentielles, de convenance, par respect de l'étiquette, ou même simplement par habitude. Il n'est pas très clair de situer alors ce que sont les raisons morales. Ce qui reste quand les autres raisons semblent épuisées ? Ou plutôt une dimension différente à définir pour chaque famille de raisons d'agir ? Une raison religieuse ou une habitude pourraient ainsi être, ou non, qualifiées de morales, mais selon quel critère ? A chaque philosophie de l'action correspondrait alors une nouvelle définition du domaine moral : la définition de la morale serait dépendante des conceptions, plus générales, que l'on peut avoir des différentes

9. J'écris ces lignes en juin 2019, pendant une vague de chaleur qui justifierait le maillot de bain, et je confirme la convention pour la ville de Paris : à l'exception de quelques enfants dans les fontaines et jets d'eau, les humains sont en tenues légères mais ni nus ni en maillot de bain dans les rues.

10. Naturellement, pour certains, sortir nu serait considéré comme immoral et pas seulement inconvenant. Peut-être le cas du maillot de bain serait-il plus discuté ?

11. On peut remarquer que l'adjectif « bien » intervient à la fois dans l'expression « bien faire », qui traduit surtout l'adéquation de ce qui est fait à l'objectif visé, donc une raison pratique, et dans l'expression « faire le Bien » qui, elle, vise la raison morale de l'action. C'est pour lever, au moins partiellement, cette ambiguïté que j'utilise ici le Bien avec une majuscule dans ce second cas.

raisons d'agir, actions obligatoires, interdites ou facultatives, contraintes ou non, raisons individuelles ou sociales, ainsi que des conceptions relatives aux problèmes induits de l'autonomie de l'agent, de sa liberté, de son libre arbitre et, en somme, de sa responsabilité morale. La piste de la définition du domaine moral par les raisons d'agir semble ainsi, comme celle par les propriétés des situations morales, ne pas devoir nous conduire facilement vers une solution dont on puisse être assuré qu'elle ne s'avérera pas finalement circulaire.

La morale est un ensemble de règles

Une autre définition classique de la morale fait une large place aux règles et maximes auxquelles se conformeraient les comportements moraux. Par exemple, le dictionnaire des Trésors de la Langue Française Informatisé (TLFI) définit la morale comme :

> 1. Tout ensemble de règles concernant les actions permises et défendues dans une société, qu'elles soient ou non confirmées par le droit.
>
> 2. Ensemble des règles que chacun adopte dans sa conduite, d'après l'idée qu'il se fait de ses droits et de ses devoirs.

On retrouve ici plusieurs dimensions importantes de la morale naïve. Tout d'abord, le lien définitoire entre morale et ensemble de règles. On peut remarquer que cette définition est soumise à la même dynamique que nous avons décrite précédemment pour les facettes de l'action morale, elle ne fait que repousser d'un temps la difficulté de la définition. En effet, prise à la lettre, cette formulation ferait, par exemple, du respect du code de la route un exemple typique de comportement moral, contrairement à ce qui est généralement admis lorsqu'on distingue la dimension conventionnelle de la dimension proprement morale comme précédemment. Pour que la définition du domaine moral évite la circularité, il faudra ajouter encore ce qui distingue une règle morale d'une règle non morale, et on n'a guère avancé en identifiant morale et ensemble de règles. Ensuite, dans la première proposition, le lien est fait entre morale et société : dans chaque société certaines actions sont permises et d'autres défendues par des règles morales, mais il resterait à savoir lesquelles, pourquoi, et par qui ou quoi elles sont définies comme morales. Le TLFI évoque également dans la première proposition la distinction entre obligations confirmées par une loi ou non, insistant sur la différence de perspective entre le moral et le légal, distinction intéressante qui complète la raison conventionnelle ci-dessus d'une notion de légalité proche mais différente, mais qui ne précise en rien la définition du domaine moral. Remarquons également que la définition indique que les actions sont « permises et défendues dans une société » mais n'ajoute pas « ou obligatoires » ce qui occulte la dimension importante de l'obligation morale. Enfin, dans la se-

conde proposition, le choix individuel de chacun remplace la contrainte externe de la société, marquant l'autonomie morale de l'individu face au groupe. Cette seconde proposition établit un parallèle entre « moral » et « respect des droits et devoirs » qui s'imposent à chacun, mais bien sûr, et à nouveau, sans préciser en quoi et comment des droits et devoirs peuvent générer ainsi des règles morales et si tout droit ou devoir relève du domaine moral, ce qui semble contre-intuitif.

On doit encore souligner ici, outre le risque de circularité déjà mentionné, plusieurs autres difficultés que soulève cette conception d'une morale en tant que « ensemble de règles ». Tout d'abord, la morale ainsi définie semble désarmée face à une nouvelle activité humaine, par exemple les biotechnologies, pour laquelle n'existe pas encore de règle et qui, néanmoins, semble bien soulever des problèmes moraux. L'établissement de ces nouvelles règles est un problème moral en lui-même et qui devra donc, peut-on penser, être soumis à des méta-règles encore plus complexes à définir que la morale elle-même. Et si ces méta-règles morales existent, pourquoi faut-il leur rajouter des règles de premier niveau qui semblent pouvoir être créées à volonté ? La conception par les règles va de même se trouver en difficulté pour expliquer les jugements moraux d'événements exceptionnels uniques qui ne peuvent avoir donné lieu au préalable à des règles morales préétablies. Enfin, la question de l'applicabilité des règles à un cas particulier reste inabordée. Par exemple, dans le cas particulièrement important des dilemmes moraux, que faire quand des règles différentes conduisent à des conclusions opposées pour une situation donnée ? Là encore, il faudra des méta-règles complexes pour établir les nécessaires priorités. Pas plus que les caractéristiques de la situation ou l'analyse par les raisons d'agir, l'aide du dictionnaire ne semble nous apporter ici de voie vers une définition analytique satisfaisante du domaine moral. La définition du TLFI nous éclaire sur les rapports entre morale, ensemble de règles, sociétés et individus mais en laissant dans l'ombre ce qui fait d'un jugement moral qu'il est moral.

Distinguer les usages descriptifs et prescriptifs des règles morales

L'entrée sur la définition de la morale dans la Stanford Encyclopedia of Philosophy (SEP) (Gert et Gert 2017) [98] choisit en outre de mettre en avant la distinction à faire entre un usage descriptif et un usage prescriptif de la morale [12] :

> Le terme « moralité » peut être utilisé sous deux angles : 1. Descriptif, et il se réfère à certains codes de conduite mis en avant par une société ou un groupe (par exemple religieux), ou adoptés par un individu pour son propre comportement.

12. Traduction de l'auteur.

2. Prescriptif[13], et il se réfère à un code de conduite qui, dans certaines conditions, serait mis en avant par toute personne rationnelle.

Les auteurs soulignent ainsi explicitement la nécessité de la distinction entre la dimension descriptive du comportement humain et la dimension prescriptive des maximes morales. Implicitement, et par la position d'observateur en surplomb qu'ils adoptent sur les possibilités d'usage du terme « moralité », ils mettent également en avant la nécessaire distinction entre les théories morales, qu'on les observe sous l'angle descriptif ou prescriptif, et la philosophie de la morale qui consiste à s'interroger sur ce que peuvent bien être cette morale et ces théories morales.

Remarquons que la distinction faite entre descriptif et prescriptif si elle est commune, et certainement utile dans la position d'observateur distancié du philosophe, n'a pas beaucoup de sens sur le terrain pratique : on voit mal comment un « code de conduite mis en avant par une société » pourrait n'être que descriptif pour un membre de cette société quand la décision est à prendre ! Cette remarque ne porte pas sur l'étude des théories morales, par exemple leur architecture et leur rôle dans les sociétés, qui est naturellement une composante de l'étude descriptive du phénomène moral dans son ensemble, mais sur le contenu substantiel de la théorie morale dont l'objectif même est de nous guider dans l'action, et est donc prescriptif.

Les indifférents

Une autre implication de toute définition du domaine moral concerne les situations qui en sont exclues. Première hypothèse, ces situations sont ainsi écartées parce qu'elles ont des propriétés propres qui les rendent moralement indifférentes. Seconde hypothèse, toute situation a potentiellement une dimension morale, et il faut alors expliquer pourquoi nous nous saisissons parfois de cette préoccupation éthique et parfois non. Dans les deux hypothèses, les analyses détaillées, qui dépassent le cadre de la présente étude, ne pourront être indépendantes des théories morales adoptées. [14]

Pour une définition très large, instrumentale

Prenant acte de toutes ces difficultés à définir le domaine moral, je propose deux hypothèses de travail, une négative et une positive. La première, négative, est qu'il serait à ce jour impossible d'établir une définition analytique du domaine moral qui convienne à tous, acteurs et penseurs moraux, et à toutes les situations, quotidiennes ou exceptionnelles, si on entend par définition analytique un ensemble de conditions nécessaires et suffisantes pour qu'une

13. Gert utilise ici le terme de « normatively » opposé à « descriptively », je traduis par « angle prescriptif » plutôt que normatif car dans les deux cas l'auteur parle d'un ensemble de codes, notion proche des normes, qu'il s'agit soit de simplement décrire soit de considérer comme obligatoire pour tout être rationnel.

14. A titre d'illustration de ce point, citons l'analyse détaillée du lien entre le problème des indifférents et la théorie morale de Levinas d'Isabelle Pariente-Butterlin (Pariente-Butterlin 2015) [177]

situation appartienne au domaine moral. La seconde, positive, est que, de toutes façons, une telle définition si elle existe, n'est nullement nécessaire pour mon enquête puisqu'une partie importante des problèmes moraux que j'envisagerai est justement attribuable, au moins pour partie, à des différences de conception de ce qui est moral ou non, et donc à la définition même du domaine moral [15] par chacun, philosophe, psychologue ou simple personne morale.

A titre instrumental, de ballon d'essai, et en acceptant volontiers d'avoir à la réviser après examen, je propose de retenir une définition provisoire permettant de situer le domaine moral et appuyée sur la dernière et la plus simplement vérifiable des facettes citées plus haut : l'existence de phrases exprimant une évaluation morale. J'ai trois arguments pour retenir cette définition. Le premier est que je n'ai trouvé à ce jour aucun psychologue ou philosophe niant qu'il existe des phrases évaluatives, des phrases exprimant un jugement qui est entendu comme moral, alors que pour chacune des autres facettes de la situation que j'ai pu évoquer plus haut, il existe des philosophes ayant nié leur existence même (en tant que pertinentes pour le phénomène moral), cela fait de cette définition une base de départ consensuelle inespérée. Le deuxième argument est que cette définition est certainement trop large et peut me conduire à des propos mal adaptés au domaine moral « vrai », s'il existe et si on peut le connaître, mais, de ce fait, elle englobe probablement toutes les autres définitions envisageables et je ne risque pas l'omission. Le troisième argument est une forme de reconnaissance de dette, c'est le choix de George Edward Moore dans Principia Ethica (Moore 1998, page 9) [165]. Je retiens donc la définition suivante :

— Une proposition est morale si elle comprend un des termes évaluateurs comme « Bien » ou « Mal ». [16]

— Le domaine moral regroupe l'ensemble des situations ayant donné ou pouvant donner lieu à une proposition morale.

Si on accepte cette proposition, l'étude du phénomène moral aura pour point de départ le constat, empirique et trivial, de l'omniprésence des propositions évaluatives dans le discours humain : «Alex dit à Bob que ce que vient de faire Charles est Bien. ». L'étude du domaine moral consiste ensuite à mettre en relation ces propositions et la situation [17] : ce que Charles a fait, ce que Alex en sait, en juge, et pourquoi il dit cela à Bob. Il conviendra ensuite de décrire

15. Pour Ruwen Ogien, c'est même l'essentiel des désaccords moraux qui porte en fait sur ce qui est considéré ou non comme moral (dans (Masala2013, p 201) [159]

16. Entrer dans le détail d'une liste plus large de termes évaluatifs serait ici certainement pertinent du point de vue sémantique, en rajoutant par exemple des termes comparatifs comme Meilleur ou Parfait, ou des termes plus épais comme Courageux ou Généreux mais n'apporterait pas grand-chose à mon propos et l'alourdirait considérablement. J'en reste à Bien et Mal et, souvent, quand la généralisation est claire, au seul « Bien » comme représentant de tous les termes évaluatifs.

17. La situation comporte à la fois tous les éléments de contexte généraux et particuliers ainsi que les éléments relatifs à l'acte faisant l'objet de la proposition évaluative. Savoir si une telle situation peut être décrite et connue est un point de désaccord entre philosophes que j'aborderai plus loin.

ce que différentes situations dites morales ont (ou non) en commun au sein d'un groupe, d'une société ou de l'humanité entière et au delà, et comment les évaluations morales s'articulent (ou non) avec les situations en cause.

Objections à cette définition

Une telle définition partant du point trivial de l'existence des propositions évaluatives peut être critiquée de multiples façons. Première critique, pour tout philosophe qui a une théorie sur ce qu'est la morale, l'expression d'un jugement évaluatif n'est probablement qu'une conséquence lointaine du comportement moral du locuteur et ne saurait en constituer une caractérisation valable, après-tout la personne peut se tromper, ou vouloir tromper, ou utiliser une proposition évaluative dans un duel rhétorique sans lien avec la réalité de la situation. Cette objection est parfaitement recevable et je serai heureux d'abandonner la définition instrumentale ci-dessus, mais à la condition que la définition plus précise de la morale qui serait alors retenue soit acceptée par tous les philosophes moraux. Tant que l'échéance de cet accord, et elle est probablement lointaine, n'est pas atteinte, je propose de maintenir pour le travail d'exploration du domaine moral la caractérisation par les propositions évaluatives.

Une autre façon d'exprimer la même critique est de reprocher à cette définition de par trop élargir le domaine moral, il est en effet évident que de nombreuses propositions évaluatives ne sont pas habituellement considérées comme morales, qu'elles relèvent du domaine pratique, « c'est bien de se couvrir par ce temps » ou conventionnel « il est mal de siffler à table ». Mais, à nouveau, pouvoir distinguer celles qui relèvent ou non du domaine moral suppose le problème posé de la définition de ce domaine moral déjà résolu, ce qu'il n'est pas.

On peut également reprocher à cette définition son caractère trop anthropocentriste. Elle semble faire du domaine moral un phénomène uniquement humain lié au langage et écartant ainsi toute recherche sur les animaux. Je répondrai que ce reproche est partiellement justifié en ce qui concerne la nécessité d'une personne exprimant une proposition évaluative, mais que si on reconnaît aux animaux non humains la capacité à exprimer des jugements évaluatifs, alors le caractère anthropocentriste est réduit d'autant. Par ailleurs, rien dans la définition ne limite ce sur quoi portent les propositions évaluatives, les animaux peuvent être considérés en tant qu'acteurs, ou victimes, de la situation à évaluer et s'insérer ainsi dans le domaine moral.

Chaque théorie morale, et chaque penseur moral, qu'il soit psychologue ou philosophe, est conduit à circonscrire à sa façon le domaine moral qu'il étudie, le sur-ensemble que je propose ici est constitué de l'ensemble des situations qui font l'objet (ou pourraient faire l'objet) d'une proposition évaluative. Pour rendre cette proposition plus facile à utiliser et pour

permettre de distinguer, au moins en première approche, les différentes conceptions des penseurs moraux, je propose une notation générale qui traduise ce point de départ commun à toute préoccupation morale : il existe des énoncés évaluatifs portant un jugement « ceci est Bien » ou « ceci est Mal ».

Je note un tel énoncé, point d'entrée dans le domaine moral, selon la forme générale suivante :

> O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien.

Je vais utiliser cette notation à de nombreuses reprises dans le cours de mon travail et il est donc utile d'en détailler le contenu. Je pose ici O comme observateur externe, c'est-à-dire le philosophe ou le psychologue entreprenant l'étude du phénomène moral. La présence de cet observateur est la marque du caractère réflexif de l'approche que j'entreprends ici : j'explicite que le phénomène moral est observé. En ce sens, mon point de vue, et celui du lecteur de cette thèse, serait d'être un observateur de cet observateur, ce que l'on pourrait peut-être appeler une philosophie de l'observateur du phénomène moral, une méta-éthique.

S'il y a énoncé évaluatif, E, c'est qu'il y a un énonciateur X et je pose également un destinataire de cet énoncé, S. Les lettres différentes correspondent ainsi à des rôles différents, sans écarter la possibilité que les divers rôles soient tenus par une même personne (X = S). La définition du domaine moral que je propose, l'existence d'un énoncé évaluatif, conduit à affirmer que seuls E, X et S sont supposés réels dès le point de départ de l'analyse, et que c'est la position de O en tant qu'observateur externe, et seulement dans cette mesure, qui l'atteste.

Je reprends ensuite la description de l'acte en cause dans le jugement moral : Y fait A à Z dans le contexte C. Notons que l'acte A peut être réalisé, en cours ou futur, la formulation au présent est ici une simplification de forme. Notons également pour mémoire que le contexte peut être compris comme celui de l'acte A, celui du jugement moral par X ou celui de l'énonciation de E vers S ou enfin celui de l'observation par O. Le contexte, en ce sens très large, est un aspect important de la description de chacune des étapes du phénomène moral.

J'appelle « situation » l'ensemble 'XSEAYZC'. Cette définition appelle deux remarques. Première remarque, O, en tant qu'observateur externe, ne fait pas partie de la situation, il est supposé neutre et interchangeable (ce qui sera concrètement à vérifier), en revanche le contexte de cette observation de X par O fait partie de C. Deuxième remarque, on pourrait imaginer une définition de la situation restreinte à 'AYZC' qui centrerait donc la situation sur l'objet du jugement moral, Y faisant A à Z, et écarterait de l'analyse les philosophes

moraux qui axent leurs théories sur le juge moral lui-même, sur le comportement de X. Pour inclure ces philosophes, il est donc préférable d'inclure X, S et E dans la situation définie au sens large comme 'XSEAYZC', ce qui correspond mieux au choix de la définition retenue du domaine moral par les énoncés évaluatifs.

Voici quelques exemples de l'utilisation de cette notation pour expliciter certaines positions morales. Pour certains penseurs, le vrai jugement moral est celui que l'on porte sur soi-même, et on peut qualifier de moralisateur, mais non de moral, celui qu'on porte sur autrui. Dans la notation on a alors (X = Y). Pour d'autres, n'appartiennent au domaine moral que les actions vers autrui, et ce que l'on se fait à soi même ne peut être l'objet d'un jugement moral. Dans la notation on doit alors avoir (Y <> Z). Pour d'autres encore, on ne peut juger un acte isolé mais seulement un individu dans la totalité de ses actions, c'est le bilan de tout ce que chacun fait au cours de sa vie qui, à l'instant final, permet le jugement moral. On a alors dans notre notation X juge Y pour l'ensemble des A : [A]. Enfin, de nombreux désaccords entre philosophes moraux portent sur la définition des ensembles auxquels appartiennent chacun des éléments de la situation 'XSEAYZC'. Par exemple faut-il inclure les animaux (sensibles ou non) parmi les victimes morales Z possibles ? Un enfant peut-il être moralement responsable, être un Y possible ? Comment faut-il définir la relation entre Y et A (causale, intentionnelle, volontaire...) pour que le jugement moral puisse s'exercer ? etc.

Ce qu'on peut appeler le « phénomène moral » est constitué de l'ensemble des comportements humains constitutifs de ces situations du domaine moral caractérisées par un énoncé évaluatif. Remarquons que, en toute généralité, le phénomène moral englobe des rôles différents, ceux de X, S, Y et Z, mais que de nombreux discours moraux oublient X et S pour se centrer sur le comportement de Y qui serait, pour partie, moralement déterminé. L'existence même de ce phénomène moral en tant que causalement efficace, le constat que l'homme qui passe devant un étang où un enfant se noie se sente l'obligation d'intervenir et de se mettre à l'eau pour sauver un inconnu, pose les questions de la justification de ce comportement, du pourquoi de ce sentiment d'obligation qui conduit à l'action. C'est un des buts des théories morales que de tenter de répondre à ces questions et, souvent, d'en souligner l'utilité voire l'indispensabilité sociale.

1.2 Les théories morales

Toute cartographie se doit de concilier deux dimensions, la topographie et la toponymie. La première décrit ce qui constitue, physiquement, les reliefs à représenter, la seconde ce

que les humains ont nommé de sommets et de lieux depuis qu'ils parcourent le territoire. La description du domaine moral ne fait pas exception à cette règle et se doit de concilier d'une part la description du terrain psychologique sur lequel il se développe et d'autre part la façon dont les philosophes ont nommé les objets qu'ils ont cru pouvoir y déceler. Et ce qu'ils ont principalement nommé, ce sont les grandes voies des théories morales que l'on suit en espérant trouver ainsi la réponse à la question du « que dois-je faire ». Dans la présente section, mon objectif est de décrire cette toponymie, les grandes théories morales reconnues, et ensuite de présenter les théories qui proposent d'en revenir à la topographie, l'étude du comportement moral sans le filtre d'une théorie morale particulière. Permettre de comparer les itinéraires est un des objectifs d'une carte réussie, et des philosophes ont ainsi établi des critères visant à comparer les voies des différentes théories morales, mais je constaterai que ces critères sont circulaires : choix du critère et choix d'une théorie morale particulière sont liés. La section suivante aura pour objectif de proposer des perspectives indépendantes de ses voies déjà tracées que sont les théories morales. Pour filer à l'excès cette métaphore cartographique, rappelons ce que la photo aérienne a été à la cartographie : un nouvel outil de grande utilité dont été absente toute toponymie. L'expérimentation en philosophie morale pourrait-elle être au domaine moral ce que la photo aérienne a été à la cartographie ?

1.2.1 Proposition de définition

Il est habituel de définir toute théorie morale par son objectif. Elle serait l'outil principal de l'activité humaine qui, pour les individus vivant au sein d'une communauté morale qui fait sienne cette théorie morale, a pour objet de contribuer à répondre à la grande question du « que dois-je faire ? ». Que la question soit au futur ou au présent, la décision étant prévue ou à prendre immédiatement, ou qu'elle soit au passé, pour juger des décisions prises et des actes accomplis, la théorie morale nous aide, et parfois nous contraint, à juger de la valeur morale de cet acte et, plus largement, de son auteur, soi-même ou un autre. Il convient dès maintenant de bien distinguer les « théories morales », outils permettant à une personne donnée dans une société donnée de savoir ce qu'il doit faire, et la théorie de la morale [18] qui a pour objet d'expliciter et expliquer le phénomène moral.

Cette définition de la théorie morale, outil pour aider au raisonnement moral en vue de la bonne action, semble supposer qu'existent, ou pourraient exister, des régularités au sein du domaine moral que l'on peut exprimer sous forme de règles génériques, règles morales que

18. J'utiliserai de préférence à « théorie de la morale » le terme de « philosophie de la morale » dans le présent travail de façon à bien marquer cette distinction : une théorie morale est située en un lieu et en un temps et a un contenu substantiel qui dit comment agir. La philosophie morale étudie le phénomène moral.

nous pouvons, ou devons en tant que personnes morales, chercher à découvrir, à enseigner et à nous approprier pour, surtout, les mettre en pratique, affiner nos jugements moraux et mettre nos actes en accord avec ces jugements.

On peut toutefois accepter l'existence du domaine moral, au sens de la proposition précédente appuyée sur les énoncés évaluatifs, sans pour autant accepter l'existence des théories morales. On peut par exemple considérer que les phrases évaluatives n'ont simplement aucun sens, ne disent rien des situations réelles qui donnent lieu à ces énoncés, et que les théories morales ne sont que des bavardages sans objet (nihilisme moral). Moins radicalement, on peut considérer que les phrases évaluatives ont bien un sens captant une propriété de chaque situation, mais que chacune est différente et qu'il est vain de rechercher des régularités (situationnisme), et encore plus illusoire d'imaginer qu'existent des règles morales qui pourraient subsumer ces situations et contribuer à l'établissement de nos jugements moraux.

En écartant le nihilisme moral et le situationnisme, on peut rechercher ce qu'ont en commun les énoncés évaluatifs, et sur cette base proposer que la théorie morale soit constituée tout d'abord d'une définition du Bien et du Mal qui permette d'utiliser ces termes dans un énoncé en regard d'une situation donnée. La théorie morale aura ensuite à expliciter comment ce Bien et ce Mal peuvent être reconnus de façon à permettre l'exercice du jugement moral. Et enfin, lorsqu'elle s'élève de l'action particulière vers plus de généralité, la théorie morale comportera également des règles ou des maximes dont le respect nous garantira d'atteindre ce Bien recherché, ou d'éviter le Mal, en application du principe générique : il faut viser le Bien et fuir le Mal. Mais la théorie morale n'aura de sens que si elle est effectivement suivie par les personnes morales et elle comportera donc, outre la définition du Bien et des règles pour l'atteindre, un ensemble de justifications, éventuellement implicites, de ces règles et de pourquoi chacun doit, ou devrait, les suivre. La théorie morale dit aussi pourquoi elle devrait être respectée.

A titre d'illustration, citons quelques unes des règles morales que les philosophes ont à l'esprit en définissant ainsi les théories morales, prenons d'abord quelques règles morales issues des dix commandements (Exode 20 : 1-20) et nous verrons ensuite la formulation plus générique de la règle d'or de la réciprocité :

> Honore ton père et ta mère, afin que tes jours se prolongent dans le pays que l'Éternel, ton Dieu, te donne.
>
> Tu ne tueras point.
>
> Tu ne commettras point d'adultère.
>
> Tu ne déroberas point.

> Tu ne porteras point de faux témoignage contre ton prochain.
>
> Tu ne convoiteras point la maison de ton prochain.
>
> Tu ne convoiteras point la femme de ton prochain, ni son serviteur, ni sa servante, ni son bœuf, ni son âne, ni aucune chose qui appartienne à ton prochain.

Dès la première phrase, sont apportées en une seule formulation l'obligation, honorer ses parents, l'origine divine de l'obligation, la récompense par la vie éternelle, et, en creux, la punition. Rhétoriquement, on peut penser que origine, récompense et punition s'étendent à toutes les obligations suivantes.

La justification de la règle semble ici à première vue transparente : Dieu l'exige. Mais c'est une chose de dire d'où vient la règle et une autre de dire pourquoi cette règle est qualifiable de morale et une troisième de dire pourquoi un être humain libre devrait la suivre. On peut donc découpler la justification de la règle en elle-même, sa qualification de morale et la justification de l'adhésion à cette règle. Si la troisième est explicite dans le premier des dix commandements, tu dois respecter la règle si tu souhaites accéder au pays du Seigneur, la seconde est plus ambiguë et donne lieu en particulier au célèbre dilemme d'Euthyphron : soit la règle est morale parce qu'elle est dictée par Dieu, et elle apparaît comme arbitraire car il aurait pu en choisir une autre, soit Dieu la dicte parce qu'elle est morale et Dieu n'est pas tout puissant puisqu'il doit respecter la loi morale. Malgré cette difficulté, l'essentiel de la théorie morale appuyée sur les commandements divins reste clair : il y a identité entre rechercher le Bien, accepter de suivre les commandements et respecter la volonté divine exprimée par ces commandements.

Les règles morales issues des dix commandements sont précises, détaillées, et leur objet présuppose toute une organisation sociale : le « prochain » est un homme qui a femme, servantes et biens, ce qui suppose l'existence de la famille, de la propriété des biens, du servage, du mariage, ... Un examen rapide de ces commandements montre qu'un préalable à la compréhension de ces règles est une définition précise de nombre d'institutions sociales dont l'existence est consubstantielle à ces règles, elles les définissent autant qu'elles sont définies par elles :

— Dieu et le paradis
— Le mariage et la monogamie
— La propriété privée
— La justice et les témoignages
— La relation de servitude

Nous retrouvons ici la difficulté mentionnée plus haut qu'il y a à séparer des règles conven-

tionnelles liées à un contexte social particulier des règles proprement morales qu'on aimerait moins dépendantes du contexte social. La théorie morale appuyée sur la religion et les commandements divins n'a pas ici cette difficulté : elle ne fait pas de différence entre ces deux niveaux d'obligations. Règles conventionnelles et morales sont une seule et même chose. La difficulté n'est pas pour autant définitivement réglée car elle peut resurgir ensuite lorsque l'évolution des sociétés rend inadmissible ou simplement impossible l'interprétation littérale de la règle. S'ouvre alors la nécessité d'une exégèse qu'il conviendra à nouveau de justifier et moralement et du point de vue théologique, si on souhaite faire dire à Dieu, par exemple, ce qu'il pense de telle nouvelle technologie qui n'existait pas quand les dix commandements ont été gravés dans la pierre.

Une théorie morale peut également être formulée de façon plus générale. Prenons pour second exemple la règle morale générique très largement répandue [19] de la règle d'or de la réciprocité. Elle peut être exprimée négativement :

> « Ne fais pas aux autres ce que tu ne voudrais pas qu'on te fasse »

Ou elle peut être exprimée positivement :

> « Traite les autres comme tu voudrais être traité »

Cette règle d'or de la réciprocité, plus générique, est également plus abstraite que les dix commandements. Mais ces formulations plus génériques cachent néanmoins de façon plus subtile les multiples possibilités de variantes de cette règle qu'il faudra spécifier pour pouvoir l'appliquer. Il faudra en particulier définir qui peut être cet « autre » qu'il faut traiter comme soi même : un homme grec de la même cité et ayant des biens ? Un pauvre, une femme ou un étranger ? Un animal ? Ces multiples possibilités offrent l'occasion d'autant de désaccords moraux à partir d'une même formulation de base. Et une fois défini qui est cet « autre », il conviendra encore de préciser de quels traitements il est question. Faut-il simplement ne pas le tuer ou ne pas lui nuire, ou l'aider quand il est en difficulté, ou encore l'accueillir chez soi comme un membre de sa famille ? La règle d'or de la réciprocité, très générique et abstraite va donner lieu à de multiples interprétations dont dépendront non seulement le comportement de ceux qui les suivent mais également la définition même de ce qui relève ou non du domaine moral.

Contrairement au cas précédent des dix commandements, la règle d'or, dans la formulation simple ci-dessus, n'apporte de justification ni à la règle elle-même ni à la nécessité pour une personne morale d'y adhérer. On pourra pour cela à nouveau évoquer un ordre divin,

19. Pour une analyse détaillée de la règle d'or voir (Gensler 2013) [96]

Dieu exigeant ce respect de l'autre du fait, par exemple, qu'il est une part de la création divine au même titre que l'agent moral lui-même, ou bien, sans faire appel à un Dieu, évoquer une nature humaine faite de solidarité sociale ou d'empathie incarnée qu'il serait immoral, car contraire à la nature humaine, de ne pas respecter.

La méta-éthique soulignera qu'à chaque choix ontologique donnant existence ou prépondérance à l'individu, la famille, le groupe, le pays, à l'humanité ou à tous les vivants, conscients ou non, plantes ou animaux, excluant ou non les virus ou les araignées, correspondront des théories morales différentes appliquant toutes le même principe de réciprocité mais à des « autres » différemment définis. Point important de ce point de vue méta-éthique, ce n'est pas seulement la règle morale adoptée et les êtres moraux concernés qui sont définis par chaque théorie morale, c'est aussi la justification de ces règles et de son périmètre d'application, et, finalement, ce qui fait qu'un être humain supposé « normal » a une tendance immédiate à les respecter.

La grande diversité des théories morales conduit les philosophes moraux à les regrouper en quelques grandes familles qui reprennent les principes de base de nos deux exemples, soit on établit un ensemble de règles auxquelles il convient d'obéir, à l'image des dix commandements, soit on établit un principe général qui s'appuie sur l'évaluation morale de nos actes par leurs conséquences, toute combinaison de ces deux stratégies étant bien sûr également possible. Dans la prochaine section, je commence par présenter les trois grandes familles canoniques qui répondent, au moins en première approche, à la définition des théories morales proposée ici, le déontologisme, le conséquentialisme et l'éthique des vertus, puis je présente différentes théories de penseurs moraux qui considèrent qu'une telle théorie morale en lien avec le « que dois-je faire » n'existe pas, ou que, même si elle existait, elle serait de toutes façons inutilisable pour traiter des cas réels. Je présenterai ensuite une proposition destinée à construire des critères d'évaluation des théories morales et je définirai quatre perspectives, quatre postures, que les penseurs moraux adoptent pour en juger. Je finirai par un examen des désaccords moraux, qu'ils soient ou non le fait de personnes adoptant une même théorie morale particulière. Théorie morale dont, à titre de conclusion de la présente section, je reformule la définition :

> Une théorie morale a pour objectif de nous aider à répondre à la question du « que dois-je faire » et, pour cela, est constituée d'un ensemble de règles qui, ensemble, ont pour effet de donner un contenu aux termes évaluatifs comme « Bien » ou « Mal » et de permettre de reconnaître ce Bien et ce Mal dans les situations auxquelles nous sommes confrontés pour pouvoir établir nos jugements moraux et

améliorer nos comportements.

1.2.2 Les grandes familles de théories morales

Il est habituel de distinguer trois grandes familles de théories morales, le déontologisme dont Kant est le grand philosophe de référence, le conséquentialisme anglo-saxon dont Bentham et Mill sont les auteurs de référence et l'éthique des vertus qui remonte à Aristote mais a connu un renouveau récent avec en particulier l'article phare de Anscombe en 1958 [7]. Cette classification est classiquement celle retenue dans les enseignements de philosophie morale [20]. Pour compléter ces trois familles canoniques, une quatrième famille regroupera toutes les théories qui, plus ou moins radicalement, considèrent soit que le domaine moral n'existe pas, soit que le domaine moral existe mais ne peut être ramené à des règles générales, soit, plus généralement, que le projet même de construire des théories morales qui sont supposées nous aider à répondre à la question du « que dois-je faire » ici et maintenant est illusoire.

Chacune de ces quatre familles regroupe un grand nombre de variantes plus ou moins opposées et plus ou moins convergentes dont il n'est pas question de donner ici plus qu'une idée très lointaine. Le petit lexique en annexe peut donner un avant goût de ce foisonnement, il n'a d'autre but que de partager ce qui a été pour moi un outil de travail et de situer certaines positions les unes par rapport aux autres.

Je vais maintenant présenter les grandes familles qui structurent le marché cognitif et social des théories morales disponibles. Ces grandes voies que les humains ont nommé pour se repérer dans le domaine moral, sa toponymie.

1.2.2.1 Le déontologisme

Le déontologisme, ou morale du devoir, est la théorie morale par excellence, celle qui s'exprime directement dans des règles : « Tu ne dois pas faire ceci » « Tu dois faire cela ». L'existence des règles morales est au centre de ces théories, et ce sera le seul point commun aux penseurs moraux de cette famille. La multitude des variantes commence dès qu'on dépasse ce constat. La liste de ces règles à respecter, leurs conditions d'applications, leur ordre de priorité, leur origine, les raisons qui nous poussent à les respecter, tout est sujet à variante et conduit chaque philosophe moral à sa propre théorie morale. Toutefois, la figure de Kant domine cette famille, et il est impossible de ne pas lui faire allégeance. Tout philosophe pro-

20. Voir par exemple sur le site de l'université du Texas la description canonique traditionnelle : `https://ethicsunwrapped.utexas.edu/glossary/moral-philosophy`.

posant une théorie morale, et surtout si elle est déontologiste, se doit de la situer par rapport à la figure tutélaire. Citons Paul Clavier et son article de 2018 dans l'Encyclopédie philosophique dirigée par M. Kristanek [49].

> Kant entreprend une Fondation de la Métaphysique des Mœurs (1785), qui substitue aux traditionnelles définitions substantialistes du Bien une méthodologie formelle. La norme d'une maxime éthiquement correcte réside désormais dans la manière dont la volonté se détermine (et non plus dans l'objet de la volonté, le Bien, la fin, ou l'action bonne). Au-delà des impératifs techniques (concernant le choix judicieux des moyens en vue d'une fin) ou pragmatiques (les conseils de prudence), Kant intronise l'impératif catégorique, inconditionnel, qui commande d'agir uniquement selon une maxime dont nous puissions vouloir qu'elle devienne également une loi universelle (IV, 402, 420).

Les objections envers le déontologisme sont nombreuses, et on peut évoquer deux angles d'approche qui en regroupent une partie significative. Premier angle d'approche critique, ces théories semblent faire du jugement moral un automatisme qui sous-estime l'importance de tous les détails d'une situation pour n'en retenir que quelques grands traits repris dans les règles morales. Pourtant nos intuitions morales courantes montrent bien plus de dépendance aux circonstances. Ainsi par exemple, et bien que mentir ne soit certainement pas moral en général, il n'en reste pas moins qu'on peut mentir en ayant de bonnes raisons morales de le faire si dire la vérité apporte un plus grand Mal. Second angle d'approche critique, l'observation des sociétés humaines a conduit les ethnologues à privilégier une conception relativiste de la morale, des règles différentes pouvant être adoptées dans des territoires différents ou sur le même territoire à des époques différentes. Il est donc peu plausible d'imaginer des règles catégoriques qui, par définition, s'imposent de manière identique dans tout contexte. Confronté à ces deux familles de critiques, le déontologiste pourra répondre en soulignant que sa théorie dit comment les humains devraient agir et non comment ils agissent, que sa théorie est réformatrice et non descriptive. Et, d'ailleurs, c'est bien parce que les humains doivent être réformés que le travail du penseur moral est utile. Reste qu'il aura également à montrer que le changement de comportement souhaité est possible et que l'exigence kantienne n'est pas, simplement, inaccessible aux humains réels.

1.2.2.2 Le conséquentialisme

Plutôt que de juger un acte selon la conformité à des règles, le conséquentialiste le jugera selon ses conséquences, réelles ou prévues. La règle à suivre s'exprime alors simplement : il faut rechercher l'acte qui apporte dans ses conséquences le plus grand Bien possible. La théorie conséquentialiste est simple, elle n'a qu'une seule règle, et particulièrement facile à comprendre, surtout en regard des autres théories morales, il faut entre deux actions choisir celle qui a les meilleures conséquences. Les deux caractéristiques principales de ces théories qui vont définir chaque variante, et elles sont nombreuses, seront d'une part la définition de ce que peut être ce Bien à rechercher et d'autre part la méthode selon laquelle calculer ce Bien pour qu'il soit possible de le maximiser ou, à tout le moins, de le comparer entre deux actions possibles envisagées.

L'utilitarisme est une variante de conséquentialisme qui définit le bien-être humain en tant que la valeur à rechercher. Cette valeur définit le Bien au travers d'une fonction d'utilité. Même si le calcul précis est difficile, et peut-être impossible, il est toutefois assez facile dans de nombreuses situations de comparer le bien-être obtenu face à une situation concrète après plusieurs actions. C'est en tous cas ce que défendent les promoteurs de cette théorie comme Joshua Greene dans (Greene 2015 p 351) [109][21] :

> Nous pouvons débattre sans fin des droits et de la justice, mais nous sommes liés ensemble par deux constats de base. D'abord celui des vicissitudes de la condition humaine et de notre désir commun d'être heureux. Aucun d'entre nous ne souhaite souffrir. Ensuite nous comprenons tous le principe de la règle d'or de la réciprocité et l'idéal d'impartialité qui la soutient. Mettons ces deux idées ensemble et nous avons cette valeur commune que nous recherchons. (...) Entendons-nous pour agir pour faire ce qui marche le mieux, ce qui nous rend globalement les plus heureux.

Comme pour le déontologisme, les objections envers le conséquentialisme sont nombreuses. Une première famille d'objections est liée à la difficulté pratique de sa mise en œuvre. Comment imaginer en effet qu'à l'heure de la délibération nous puissions envisager toutes les options possibles, pour chacune d'elle toutes les conséquences possibles et pour chacune de ces myriades de situations calculer une valeur d'utilité de façon à pouvoir les comparer ? Un tel calcul semble exiger une froideur analytique qui fait plus penser à une machine ou à un psychopathe qu'à un être empathique et moral. Constatons d'ailleurs, et c'est une se-

21. Traduction de l'auteur

conde famille d'objections, qu'en situation réelle nos jugements moraux sont très rapides et ne donnent pas lieu à de telles délibérations, ce qui se traduit, entre autres, par des cas où nos intuitions morales sont en contradiction avec le principe conséquentialiste. [22] Là encore, le philosophe pourra répondre avec l'argument du caractère réformateur, et non descriptif, de sa théorie et, nous le verrons plus loin avec Joshua Greene, en adoptant le conséquentialisme comme théorie de dernier recours lorsque les désaccords moraux ne peuvent être résolus dans le cadre des conceptions morales d'acteurs moraux en situation.

1.2.2.3 L'éthique des vertus

Anscombe critique en 1958 la stagnation du discours moral appuyé sur la pléthore de théories morales déontologistes et conséquentialistes et propose d'abandonner la recherche de règles morales introuvables pour revenir aux fondements aristotéliciens : la quête de la vertu. Il ne s'agit donc pas premièrement de faire des choses Bien mais, avant tout, de chercher à être quelqu'un de Bien et, pour cela, de se comporter en fonction de cet objectif et viser l'idéal des vertus humaines. On peut en cela s'inspirer des exemples historiques ou fictionnels des êtres moralement exceptionnels qui inspirent une société et lui servent de repère.

Une vertu est un trait de caractère stable qui dispose à faire le Bien, à bien juger moralement des situations. Le vice est la disposition inverse à faire le Mal et à émettre des jugements moraux erronés. Chaque philosophe définira la liste des vertus qu'il privilégie, et la base de référence historique d'une telle liste est celle d'Aristote. Aristote définit quatre vertus cardinales, la justice, la tempérance, la sagesse et le courage puis un principe plus générique, la vertu est le juste milieu entre deux excès.

Citons là encore deux grandes famille d'objections contre l'éthique des vertus. D'abord elle suppose que l'enjeu moral opère pour une personne et non pour un acte particulier or de nombreux résultats expérimentaux vont dans le sens de mettre en doute la constance du caractère d'un acte à l'autre et, conséquemment, l'existence même de dispositions individuelles stables que l'on puisse qualifier de vertus. L'ouvrage de Ruwen Ogien (Ogien 2011) [174] souligne ce constat dans son titre : l'influence des croissants chauds sur la bonté humaine. Il suffit de modifier légèrement les circonstances, par exemple en créant une ambiance agréable avec des odeurs de croissants chauds, pour que le comportement des personnes change significativement. Deuxième famille d'objection, reprise par exemple par Prinz [185], nous ne qualifions pas de « morales » de la même façon chacune des vertus, ainsi le courage et la justice ne nous apparaissent pas de même statut « moral ». Si cette intuition est juste, alors cela signifie

22. Nous en verrons plus loin de façon détaillée un exemple expérimental avec le dilemme du tramway.

qu'il faut à nouveaux frais redéfinir ce qui relève du domaine moral pour chacune des vertus visées, et la théorie des vertus n'est plus suffisante pour établir une éthique.

Confronté à ces objections, le philosophe des vertus pourra souligner, avec Anscombe, que sa théorie offre une meilleure description du domaine moral que les deux autres familles et qu'elle est plus opérationnelle, au sens où elle donne un cadre permettant de définir les réformes souhaitables, celles qui visent à mieux connaître et rechercher les vertus. Il convient d'insister ici, pour clore cette présentation rapide des trois grandes familles de théories morales, qu'aucune n'a à ce jour fourni d'arguments définitifs acceptés par la communauté des théoriciens moraux et que le débat infructueux dénoncé par Anscombe en 1958 se poursuit toujours aujourd'hui, l'éthique des vertus n'étant qu'un protagoniste de plus à la table du débat.

1.2.3 La promesse des théories morales est inatteignable

Changeons de point de vue sur notre cartographie du domaine moral, qu'en serait-il si la toponymie offerte par les grandes théories morales ne correspondait pas à la topographie du terrain psychologique ? Si, par exemple, ces grandes voies se contentaient de suivre des crêtes inhabitées, bien visibles mais inatteignables, alors que les humains se concentrent dans les vallées ... C'est ce qu'ont proposé des penseurs que je me propose de suivre maintenant.

1.2.3.1 Mise en doute du discours moral

Chacun des tenants d'une des trois grandes familles de théories morales a posé de nombreuses objections aux deux autres familles, mais des mises en doute plus profondes du discours moral ont également été formulées par les penseurs moraux. Elles peuvent être de plusieurs types : métaphysiques, épistémiques ou politiques. Reprenons tour à tour ces trois types de mises en doute du discours moral.

Du point de vue métaphysique, il s'agit de nier que quelque chose comme des règles morales ou des faits moraux existe : le discours moral constitue simplement une façon de parler, une branche de la fiction, au même titre que les contes de Noël. Il ne s'agit naturellement pas de nier les discours moraux, pas plus que l'existence des contes de Noël, ni de nier leur puissance causale, pas plus que de celle des contes de Noël démontrée chaque année par l'augmentation des chiffres d'affaires autour du 25 décembre et, plus joyeusement, par les émotions positives induites dans les familles.

Selon ce point de vue, le discours moral fait partie, avec les mythes, les habitudes, les

fictions, et les recettes de cuisine, de ce qui anime les groupes humains, et qui est respecté par les individus en tant que membres de ces groupes. En un mot, le discours moral est une partie de la culture de chaque société. On peut décrire les préceptes moraux adoptés par un groupe humain au même titre et avec les mêmes objectifs anthropologiques que l'on poursuit en décrivant la littérature ou les plats typiques d'une région. Il est donc illusoire de parler de « théorie morale » qui exprimerait ce que serait l'essence du Bien puisque la morale n'existe qu'en tant qu'un air du temps (et du lieu) accepté momentanément, élément utile à la coordination et la reconnaissance au sein d'un groupe.

Cet antiréalisme moral [23] a de nombreux atouts, dont principalement celui de répondre du constat anthropologique de la diversité des règles morales, de la diversité des religions qui les appuient et de prendre ainsi au sérieux l'impossibilité de mettre d'accord les penseurs moraux des différentes sociétés sur ce que pourrait être une morale réaliste universelle. L'absence d'accord sur ce que pourraient être des règles morales qualifiables de réelles est encore plus profonde et plus frappante lorsqu'on considère leur évolution au cours du temps. Bien des écrits du 18^{e} siècle sur l'esclavage ou les noirs seraient aujourd'hui inacceptables, alors qu'ils reflétaient en leur temps une opinion commune. [24]

Malgré cette capacité à expliquer naturellement la diversité des opinions morales, l'antiréalisme moral présente la difficulté importante d'être fortement contre-intuitif. Pour chacun d'entre-nous il semble que maltraiter un enfant ou assassiner son prochain soit objectivement Mal et le serait dans toute société humaine digne de ce nom. Plutôt que de nier la réalité de la morale, il serait peut-être plus conforme avec ces intuitions de conserver un noyau du discours moral faisant droit à une base commune à toute l'humanité et d'en écarter une autre rattachée à des conventions locales. Mais la difficulté sera alors de définir ce noyau pour qu'il soit acceptable aussi bien par un métaphysicien occidental d'aujourd'hui que par un penseur aztèque cannibale.

Du point de vue épistémique, le doute porte non sur l'existence des règles morales ou des faits moraux mais sur notre capacité à les connaître. On retrouve là toutes les difficultés que nous avons rencontrées au moment de définir le domaine moral. Comment faire la part de ce qui est moral et de ce qui est, par exemple, simplement conventionnel ? Et si nous ne pouvons le savoir, comment envisager que les raisons pour lesquelles nous respectons les règles

23. Classiquement, l'expression « réalisme moral » est utilisée pour les philosophes qui pensent que les propriétés morales sont réellement des propriétés des objets du monde, objectives, et l'expression « antiréalisme moral » pour ceux qui pensent que ce sont des propriétés mentales, subjectives. Mais de nombreuses variantes sont possibles à partir de ces deux positions opposées. Comme le rappelle Christine Tappolet dans (Desmons 2019 p 176) [71].

24. A titre d'exemple, l'ouvrage de Kant « Observations sur le sentiment du beau et du sublime » offre ainsi des analyses qui seraient jugées aujourd'hui inacceptables tant elles apparaissent au lecteur moderne misogynes et racistes [129].

morales soient différentes des raisons pour lesquelles nous respectons les conventions ? En bref, les notions morales comme le Bien ou le Mal n'ont pas de référent que nous puissions connaître et, donc les phrases évaluatives n'ont simplement pas de valeur de vérité atteignables.

La distinction entre la mise en doute métaphysique de l'existence même des faits moraux et la mise en doute épistémique de notre capacité à les connaître est importante pour le philosophe mais elle a peu de conséquences pratiques. Dans les deux cas les discours moraux n'ont pas pour référent les phénomènes du domaine moral lui-même, qu'ils existent ou non. Les énoncés moraux semblent alors réduits à des outils de coordination et de reconnaissance au sein d'un groupe humain.

1.2.3.2 La morale, vecteur d'asservissement

Les discours moraux ont également été dénoncés sous l'angle politique. Ainsi, les anarchistes comme Kropotkin [144] dénoncent-ils la morale comme un vecteur d'asservissement au service des puissants, clergé, militaires, possédants. Il est plus expédiant pour le Prince de faire adopter par son peuple et intérioriser par chaque individu une règle morale qui le serve plutôt que d'obtenir le même résultat par la force. Selon Kropotkin, les religieux, experts en la matière, construisent ainsi leur position lucrative auprès du Prince en l'aidant à formater son peuple en sujets obéissant aux règles qu'il a édictées. Constatant ce rôle principalement répressif de la morale dans chaque groupe humain, dans chaque pays, dans chaque société, les anarchistes oscillent entre l'option de récuser toute morale, l'anarchiste, homme libre, ne serait soumis à aucune obligation morale ni conventionnelle que ce soit, et l'option de forger une nouvelle morale universelle et égalitariste hors de portée des dirigeants et des religieux avec leurs croyances aliénantes.

La référence de Kropotkin me permet de souligner un aspect du domaine moral souvent négligé dans les ouvrages de philosophie morale et, en particulier, dans ceux qui défendent telle ou telle théorie morale. La morale est souvent, si ce n'est toujours, source d'exactions qui sont faites en son nom à un moment donné puis jugées immorales ensuite. C'est au nom de leur morale que les djihadistes tuent, que les Américains envahissent l'Irak ou que l'inquisition sévissait. « Gott mit uns » est inscrit sur les médaillons des génocidaires. Au sens défini plus haut de l'ensemble des situations susceptibles d'énoncés évaluatifs, le domaine moral inclut ces cas d'exactions morales. La position anarchiste anti morale est, pour partie, une réaction à ces comportements moraux humains souvent qualifiés de terroristes quand ils sont le faits des gouvernés, et de violence légitime quand ils sont le fait des gouvernants. Sans

aller jusqu'au nihilisme moral, Linda Skitka insiste dans son article « The Psychological Foundations of Moral Conviction (dans Sarkissian 2014 p 161) [200] sur le double potentiel de l'attitude morale [25], d'un côté elle peut pousser au dévouement et aux actions vertueuses envers les personnes qui partagent les mêmes convictions morales, mais d'un autre côté, l'attitude morale rigidifie les comportements dans un refus de tous ceux qui ne partagent pas les mêmes règles, allant jusqu'à des actes qu'une personne moins rigide qualifierait d'immoraux. Trop de moralité conduirait ainsi à l'immoralité.

Les trois interrogations précédentes, métaphysique, épistémique et politique, conduisent au même constat de l'impossibilité d'atteindre l'objectif que se donnent les théories morales, c'est-à-dire aider chacun à répondre à la question du « que dois-je faire ? ». Mais il n'est pas besoin de si profondes remise en causes pour arriver à la même conclusion. Il suffit de constater pragmatiquement, domaine d'activité après domaine d'activité, que nous ne pouvons appuyer les décisions comportant une dimension morale sur l'utilisation de ces théories morales. C'est l'objet de la prochaine section.

1.2.3.3 L'équilibre réfléchi

Les théories morales peuvent échouer à déterminer ce que nous devons faire ici et maintenant. Que ce soit pour la raison simple qu'il n'existe pas de propriétés morales, ou que nous ne pouvons trouver de règles à leur endroit ou que ces règles sont inapplicables lorsqu'elles sont confrontées à la complexité des situations réelles, ou encore qu'elles sont détournées comme instruments d'un pouvoir coercitif. Dans tous les cas nous nous retrouvons sans théorie morale pour faire face au problème de l'action, ici et maintenant. Et pourtant, nous considérons, en tant qu'humains, qu'il y a des actions que nous sentons devoir entreprendre et d'autres non et que s'appuyer sur ce sentiment pour aider à la recherche de la meilleure décision semble une voie ouverte à la construction d'accords entre parties prenantes. [26]

Renonçant à s'appuyer sur des théories morales pour pouvoir évaluer ce qui sera moralement bon dans toutes les situations, les penseurs moraux ont proposé d'adopter non des règles morales qui sont soit trop strictes soit trop difficiles à interpréter, mais une démarche dynamique permettant directement de se mettre d'accord, avec les autres ou avec soi-même, dans un cas particulier.

25. Skitka distingue trois types d'attitudes d'intensité croissante, la simple préférence, affaire de goût subjectif qui appelle la plus grande tolérance, la convention sociale, normative et limitée à une communauté, et le mandat moral, perçu comme absolu, impératif et fortement émotionnel et motivationnel.

26. L'expression « parties prenantes » provient de l'éthique des affaires (Anquetil 2008) [5] et (Anquetil 2011) [6]. Elle désigne l'ensemble des personnes concernées par une action qu'on cherche à évaluer moralement : décideurs, acteurs, collaborateurs, clients, fournisseurs, actionnaires, environnement,...

Cette démarche est inspirée du fonctionnement d'un tribunal. Certes nous avons des lois, mais il est souvent difficile en pratique dans un cas particulier de déterminer si une loi s'applique. La justice instaure donc pour cela une démarche qui donne la parole à chaque partie, le juge entend les différentes versions et, après délibération, le juge dit le droit. La démarche de l'équilibre réfléchi poursuit le même objectif et adapte ce fonctionnement au contexte moral. Il s'agit de rechercher un équilibre entre les différentes règles morales dont pourrait relever un cas particulier selon diverses théories morales, et, surtout, de rechercher l'équilibre entre les parties prenantes quand chacune met l'accent sur la règle morale qui sert son intérêt ou son point de vue. La démarche prend pour cela le temps de la réflexion, de la délibération, pour arriver à un jugement bien pesé : l'équilibre réfléchi.

Le parallèle avec la situation juridique ne doit pas occulter des différences importantes entre les deux situations. Premièrement il n'y a pas d'institution judiciaire qui formalise le processus, la morale est avant tout affaire personnelle et chacun est son propre juge. Deuxièmement, le corpus moral n'a pas le statut structuré et explicite du corpus juridique constitué des lois et de la jurisprudence. Chacun peut adopter un ensemble de règles morales potentiellement applicables différent. Ensuite, le juge ne se prononce que sur des faits passés alors que le jugement moral a cette double vocation d'à la fois contribuer à préparer l'action à venir et à juger moralement des faits passés. Et enfin, la parole du juge est performative, lorsque le droit est dit, le droit est défini. Dans le domaine moral en revanche, tout jugement pourra être remis en cause ultérieurement que ce soit par la même personne ou par d'autres ayant ou non les mêmes conceptions morales.

La théorie morale n'est pas directement définie dans le cadre de l'équilibre réfléchi comme elle l'est dans les cas précédents. Elle est toutefois implicitement esquissée et on peut en formuler quelques traits. D'une part il existe des grands principes suffisamment partagés, comme la règle d'or de réciprocité, pour que la démarche puisse être entreprise sur la base de ce fonds commun. Si un tel fonds n'existait pas, il serait impossible d'établir un objet de désaccord et un lieu de discussion. D'autre part, la théorie de l'équilibre réfléchi relève d'un optimisme quant à la nature humaine qui conduit à penser que, confrontés à une situation complexe et après délibération, les humains seraient capables, au moins assez souvent, de se mettre d'accord sur la meilleure décision sur le plan moral. [27]

Comme pour les autres familles que j'ai évoquées, de multiples variantes s'appuyant sur la démarche de l'équilibre réfléchi sont possibles, mais remarquons que son caractère profondément pragmatique visant à une adaptation dynamique à toutes les situations réelles

27. Je reviendrai plus loin sur l'utilisation de cette démarche dans le domaine de l'éthique médicale.

a moins de chances d'induire des affrontements importants entre les penseurs moraux que les familles de théories morales s'appuyant sur une définition stricte du Bien ou des règles morales. Le repas de famille pourrait, toute chose bien pesée, être plus serein.

1.2.4 Comparer les théories morales : les huit critères de Timmons

Nous disposons maintenant d'une première description, très rapide, des grandes voies de la toponymie adoptée par les philosophes moraux pour décrire le domaine moral. Le philosophe moral souhaitera avoir une vue d'ensemble de ces théories morales et, surtout, des outils visant à les comparer car, de son point de vue, il faut bien entreprendre le voyage en choisissant une de ces voies établies. Je me propose maintenant de rappeler ce que pourraient être des critères permettant d'évaluer ainsi une théorie morale. Je reprends et simplifie cette liste de l'ouvrage introductif de Timmons qui, comme il semble être d'usage aux États Unis, ne fait qu'une place minimale à la possibilité que la morale ne soit qu'une fiction (Timmons 2013) [228][28].

— Consistance : Les principes d'une théorie morale appliqués à des informations factuelles pertinentes devraient conduire à des verdicts moraux consistants.
— Détermination : Les principes déterminent un verdict dans un grand nombre de cas.
— Applicabilité : Les principes sont tels que les humains puissent pratiquement les appliquer pour décider.
— Intuitivité : La théorie donne sens à nos intuitions morales les plus courantes.
— Validité interne : La théorie permet de retrouver logiquement à partir de ses principes un grand nombre de nos jugements moraux réfléchis. Inversement, une théorie qui ne permet pas de retrouver ces jugements est à mettre en doute.
— Pouvoir explicatif : Une théorie morale devrait proposer des principes qui expliquent les jugements moraux et nous aider à comprendre pourquoi ce qui est Bien est Bien.
— Validité externe : Une théorie morale peut bénéficier du support des connaissances non morales provenant en particulier des différentes sciences humaines.
— Publicité : Une théorie morale doit pouvoir être enseignée, les théories ésotériques ne satisfont pas ce critère.

La liste de ces huit critères permet d'anticiper tout ce qui, du point de vue de l'auteur philosophe moral d'un pays occidental, va pouvoir différencier une théorie morale sur le plan

28. L'ouvrage de Timmons a été publié en 2001 puis en 2013, il est intéressant de noter que, pour prendre en compte l'évolution de la philosophie morale pendant ces dix années, l'auteur a rajouté dans la rubrique « à lire » une catégorie d'ouvrages supplémentaire, les travaux empiriques, montrant ainsi l'importance de l'apport des travaux des sciences expérimentales dans ce domaine.

du service rendu, c'est-à-dire sur sa capacité à aider à répondre à la question du « que dois-je faire ». Notons toutefois le caractère particulier du dernier critère, la Publicité, qui a pour objet d'exclure les théories ésotériques qui ne peuvent être publiquement enseignées sans passer par un processus d'intronisation, ce critère est sans lien avec les autres et semble pouvoir être isolé.

Les huit critères qui permettraient de qualifier une théorie morale selon Timmons sont donc : Consistance, Détermination, Applicabilité, Intuitivité, Validité interne, Pouvoir explicatif, Validité externe et Publicité. Cet ensemble de critères peut être utilisé avec deux objectifs différents. Le premier est de comparer les grandes familles de théories morales décrites ci-dessus et d'analyser comment chacune privilégie tel ou tel critère dans cet ensemble, et c'est ce que je vais faire maintenant, dans l'espoir initial, qui sera déçu, que cette comparaison puisse apporter ensuite une lumière sur les potentiels apports des méthodes expérimentales. Le second, et principal objectif pour Timmons, est de comparer entre elles les multiples théories morales qui sont des variantes des trois grandes familles de théories, déontologistes, conséquentialistes et éthique des vertus et de permettre ainsi de construire des arguments en faveur de l'une ou l'autre variante. Dans la mesure où je ne souhaite ici défendre aucune variante particulière de théorie morale, je ne poursuivrai pas ce second objectif.

Déontologisme

Les règles morales des théories déontologistes sont construites, comme nous le montre Kant, en partant de propositions morales intuitives, par exemple « il est mal de mentir», puis en les reconnaissant en tant que règles morales si elles répondent à une possible universalisation. En ce sens, ces théories privilégient naturellement les critères de l'Intuitivité et de la Validité interne puisqu'elle sont construites autour de ces propositions morales intuitives et qu'il est peu probable que les règles morales universelles qui en seront déduites par le processus de construction morale kantienne viennent massivement en contradiction avec leur point de départ.

Chaque variante de déontologisme répondra ensuite différemment aux autres critères. Ainsi le Pouvoir explicatif dépendra du système de justification des règles morales proposées, la Consistance dépendra de l'ensemble de règles précisément envisagées et du mode d'arbitrage entre règles différentes, la Validité externe dépendra du compte rendu que peut faire un psychologue des comportements qui seraient impliqués par chaque théorie. Cet ensemble de critères établi par Timmons prouve ainsi son intérêt en permettant de discriminer toutes les variantes de déontologisme, ce qui est le cœur de la préoccupation de l'auteur de ces critères.

Conséquentialisme

Les théories conséquentialistes semblent avant tout viser à répondre aux critères de Consistance et de Détermination. La consistance est obtenue simplement dès la construction du conséquentialisme puisqu'il ne repose que sur une seule règle, qu'on ne suppose qu'un seul Bien. L'objet de la morale est de maximiser ce Bien en s'appuyant sur l'anticipation du calcul de ce Bien obtenu pour chaque alternative. La Détermination est obtenue dès lors qu'il est très souvent possible de comparer le Bien obtenu pour chacune des options d'action envisagées.

Comme précédemment, le Pouvoir explicatif d'une théorie conséquentialiste dépendra du système de justification de ce Bien qu'elle promeut et de son mode de calcul. Les différentes variantes au sein de cette famille pourront jouer sur l'Applicabilité [29], sur la Validité interne en modifiant la définition du Bien pour tenter de retrouver des jugements moraux généralement acceptés, ou la Validité externe. En revanche, le critère d'Intuitivité n'est pas significatif pour les théories conséquentialiste qui, au contraire, cherchent à substituer à des intuitions morales naïves un calcul moral élaboré sur un principe de maximisation.

Éthique des vertus

L'éthique des vertus n'a pas pour principale visée d'aider à la détermination face à une situation de choix moral. Les critères de Timmons sont, inversement, bâtis sur cette visée. Il n'est donc pas aisé de qualifier l'éthique de la vertu sur la base de ces critères qui, tous, gagneraient à être reformulés pour ce type d'éthique. La Consistance n'a pas beaucoup de sens si on ne juge pas les actes un à un, en revanche un critère de constance marquant l'engagement dans une voie serait pertinent. La Détermination, au sens où la théorie peut produire une réponse en toute situation, est également peu significative si on vise à la vertu, on peut même soutenir inversement qu'une vertu qui a réponse à tout est plus proche de l'arrogance que de la pondération. L'Applicabilité, au sens de la possibilité pour les êtres humains d'accéder aux exigences morales est en revanche un critère que pourrait reprendre l'éthique de la vertu en modifiant toutefois la notion de possibilité en jeu. Pour le conséquentialisme, par exemple, c'est de possibilité pratique dont on parle, alors que pour l'éthique de la vertu, ce serait de la possibilité d'atteindre à la stature morale des grandes personnalités de référence. Les critères qui supposent de comparer des jugements moraux constatés en situation avec les prescriptions de la théorie morale, l'Intuitivité, la Validité interne et le Pouvoir explicatif, n'ont pas d'équivalent si la théorie n'a justement pas cet objectif de prescription en situation.

Nihilisme moral et équilibre réfléchi

29. On peut penser par exemple à l'opposition entre l'utilitarisme de l'acte, qui serait en pratique inaccessible tant il exige de ressources cognitives, et l'utilitarisme de la règle moins gourmand.

Le nihilisme moral fait un usage à rebours des critères de Détermination, d'Applicabilité et de Validité externe en affirmant qu'aucune théorie morale ne permet de déterminer ce qui est moral dans un nombre de cas significatif, qu'aucune théorie morale n'est en réalité appliquée par les humains, et qu'aucune théorie morale n'est appuyée par les autres sciences humaines. L'explication la plus simple de cet échec est, pour le nihiliste moral, que l'objet même de la théorie morale, la morale substantielle, n'existe pas. Quatre autres critères, la Consistance, l'Intuitivité, la Validité interne, et le Pouvoir Explicatif peuvent alors être lus comme reflétant la cohérence entre une théorie morale, bavardage sans référent réel, et nos préjugés culturels à un moment donné, ils ne mesurent pas la validité d'une théorie morale, ce qui n'aurait pas de sens du point de vue du nihiliste moral, mais simplement sa capacité à capter et traduire l'air du temps d'une société particulière.

Enfin, les théories appuyées sur l'équilibre réfléchi ont un statut un peu à part car elles n'envisagent que deux critères, la Détermination, la méthode doit conduire à un verdict dans tous les cas, et l'Applicabilité, la méthode doit être à notre portée d'humains réels. Tous les autres critères sont considérés comme sans signification car s'appuyant sur une prétention illusoire des théories morales : avoir la capacité à traiter de façon générale des cas particuliers.

La tentative d'utilisation des critères de Timmons me conduit au constat que chacune des grandes familles de théories morales répond prioritairement à l'un ou l'autre des critères. On ne peut donc utiliser ces critères transversalement à toutes les théories morales, sur le domaine moral en général, puisque à chaque pondération de ces critères correspondra l'adoption d'une famille de théorie morale particulière. Ces critères ne constituent donc pas une grille d'analyse pertinente pour notre travail qui se veut transversal, et je propose ci-après de remplacer ces critères d'évaluation des théories morales par une caractérisation des postures des philosophes examinant le domaine moral. Je vais multiplier ces points de vue avec en particulier quatre perspectives qui correspondent à des approches du domaine moral classiquement reconnues en philosophie morale qui me permettront ensuite de systématiser l'examen des lieux des désaccords moraux. La perspective descriptive est celle qu'adoptent les philosophes moraux quand il s'agit d'observer comment les humains se comportent en présence d'enjeux moraux, la perspective prescriptive quand il s'agit de dire, au nom d'une théorie morale particulière, comment ils devraient se comporter. Dans la perspective méta-éthique, le philosophe cherche à comprendre le phénomène moral, ce qui le caractérise, pourquoi et comment les communautés humaines se donnent ainsi des règles morales et ce qui fait que les individus tendent à les respecter. Enfin, dans la perspective de l'éthique appliquée, il ne s'agit plus

d'édicter des règles générales mais d'apporter un éclairage concret aux acteurs confrontés à des situations particulières.

1.3 Quatre perspectives sur le domaine moral

La cartographie du domaine moral ne peut s'appuyer sur la seule toponymie apportée par les théories morales, elle ne peut non plus utiliser les critères établis pour différencier les théories morales et leurs variantes, car ces critères sont eux-mêmes, circulairement, chargés des théories morales qui en constituent le cadre de réflexion. Dans cette dernière section du chapitre visant à établir une cartographie du domaine moral en vue d'analyser l'apport potentiel de la démarche scientifique expérimentale, je vais proposer deux outils complémentaires utiles à arpenter le domaine moral. Le premier est constitué d'un ensemble de quatre perspectives sur le domaine moral qui permettent de relativiser chaque proposition morale à l'objectif qui la justifie : décrire, prescrire, expliquer ... Le second s'appuie sur l'analyse des désaccords moraux et, en particulier, sur une topographie des désaccords entre philosophes moraux. Mais proposer ainsi des outils suppose, a minima, de les calibrer pour vérifier qu'ils permettent bien de situer des théories morales connues. J'ai retenu pour cette calibration deux théories récentes aux antipodes du terrain à décrire. Le réalisme moral de David Enoch et le sentimentalisme constructif de Jesse Prinz seront ainsi des points de repère, et de validation. Si la description de ces deux théories morales très différentes avec les deux outils proposés dans cette section est pertinente, alors je considérerai pouvoir l'employer dans la suite de l'étude.

1.3.1 Analyser le domaine moral

La section précédente a montré qu'établir des critères pour comparer les règles morales en regard de leur capacité à jouer le rôle de guide vers l'action, de réponse à la grande question du « que dois-je faire ? », conduit à des listes de critères qui, d'une part, présupposent que ce rôle de guide puisse être joué, et donc écartent les penseurs moraux qui ne sont pas sur cette ligne de pensée, et d'autre part sont fortement dépendantes des théories morales elles-mêmes qu'il est pourtant question de comparer. Ces critères ne sont donc utilisables ni pour aborder l'analyse d'ensemble du domaine moral ni, a fortiori, pour analyser transversalement la possibilité de l'apport empirique aux débats de philosophie morale.

Je vais maintenant me rapprocher de la philosophie morale et chercher dans la structure même de la réflexion éthique qui la constitue les axes d'analyse du domaine moral. On

peut en effet envisager le domaine moral selon quatre perspectives souvent posées à la base de la structure académique de la philosophie morale [30], et que j'entends ci-dessous mobiliser comme quatre postures, quatre perspectives, utiles à tout activité de réflexion sur ce domaine moral, qu'elle soit philosophique, scientifique ou simplement le fait d'une personne morale. Ces quatre perspectives sont l'étude descriptive du domaine moral, la prescription d'une théorie morale substantielle particulière, la méta-éthique et l'éthique appliquée. J'emploierai ci-après à titre de raccourci les quatre termes de description, prescription, méta-éthique et éthique appliquée. J'emploierai également le terme de perspective pour marquer le fait que chaque chercheur peut, tournant autour du domaine moral, adopter tour à tour chacune de ces perspectives, et ce de préférence au terme de posture qui marquerait trop un engagement du chercheur à privilégier un seul point de vue. Les quatre perspectives sont complémentaires (nécessaires, mais peut-être pas suffisantes) pour qui veut avoir une vue d'ensemble non biaisée du domaine moral.

Mon ambition est que ce choix de quatre perspectives soit utile pour analyser ensuite les apports potentiels de la démarche scientifique expérimentale à la résolution des problèmes moraux. Pour commencer à établir cette possibilité, j'appliquerai cette grille d'analyse des quatre perspectives aux désaccords moraux dans la section suivante. Mais je dois au préalable préciser chacune de ces quatre perspectives.

Perspective descriptive

Première perspective, la description, l'étude descriptive du domaine moral, individuel et social. Étude descriptive du phénomène moral qui, au moins pour partie, semble être lié aux actions humaines et aux jugements moraux.

L'étude descriptive du domaine moral se fait pour partie au sein des sciences humaines, en lien avec la psychologie, la sociologie, l'anthropologie, etc. pour analyser les comportements moraux, les théories morales, leur genèse et leurs contenus ainsi que leur mise en œuvre par des individus au sein de leurs environnements sociaux. Notons que les théories morales, en tant qu'elles nous aident à porter un jugement moral, n'ont pas d'obligation à être descriptives, elle ont à dire ce que devrait être notre jugement moral et non ce qu'il est. Néanmoins, les tenants d'une telle théorie réformatrice auront tout de même à mesurer l'écart qu'il conviendra de combler pour accéder aux comportements moraux idéaux qu'ils souhaitent promouvoir et ils devront proposer des moyens pour cela. En ce sens, ils ne peuvent se désintéresser totalement de la description de l'existant.

30. Par exemple Monique Canto Sperber et Ruwen Ogien dans « La philosophie morale » (Canto-Sperber 2017 p 9) [38] proposent de distinguer l'éthique normative, l'éthique pratique et la méta-éthique, auxquelles j'ajoute la dimension descriptive qui est commune aux trois domaines.

Entreprendre l'étude descriptive du domaine moral semble présupposer que l'on puisse définir à l'amont ce qui est moral ou non, et donc adopter un point de vue qui risque lui-même d'être dépendant d'une théorie morale particulière. C'est pour éviter cette circularité, et regrouper l'ensemble très vaste et profondément divisé des penseurs moraux dans une même approche, que j'ai proposé d'adopter à titre instrumental une définition très large du domaine moral appuyée sur l'existence des énoncés évaluatifs.

Perspective prescriptive

Deuxième perspective, la prescription d'une morale substantielle qui a pour objet de définir une théorie morale particulière, descriptive et prescriptive, qui vise à dire ce qu'est le Bien et ce qui est Bien, et, selon diverses approches, à l'expliquer, la justifier et, par tant, à la promouvoir. Cette deuxième perspective, partie intégrante des systèmes moraux, est habituellement celle des enseignements moralisateurs, religieux ou non. Remarquons que cette deuxième perspective est par nature agonistique : plusieurs théories morales définissent le Bien différemment et comme toutes ces définitions ne peuvent être, généralement, simultanément acceptées, l'éthique substantielle est, pour partie, un sport de combat.

Perspective méta-éthique

Troisième perspective, méta-éthique. La philosophie morale dans la posture méta-éthique vise à analyser et éclairer ce qui se joue dans les deux premières perspectives. En particulier, elle recherche les fondements des phénomènes moraux ainsi que des différentes théories morales. Elle cherche, peut-être principalement, les raisons pour lesquelles les sociétés instituent des morales et celles pour lesquelles les individus considèrent que les règles morales doivent être respectées, l'origine de leur force d'obligation. La méta-éthique a donc pour objectif, prenant de la distance par rapport à l'éthique substantielle, de répondre à un ensemble de questions sur la sémantique, la métaphysique et l'épistémologie des règles morales : quels sont leurs réalités, leurs fondements, leurs justifications et comment nous les adoptons et, éventuellement, nous pouvons les connaître et les reconnaître dans nos vies.

Perspective de l'éthique appliquée

Quatrième perspective, l'éthique appliquée. L'éthique appliquée vise à fournir un accompagnement éthique dans les cas pratiques souvent rencontrés, des cas particuliers dans des domaines spécifiques, la santé, les affaires, le sport, etc. L'éthique est alors appliquée dans un double sens, elle est appliquée à un cas particulier de façon à aider à la prise de décision de la personne morale concernée par ce cas particulier, et elle est appliquée au domaine d'activité concerné par cette décision, les considérations morales générales doivent alors être déclinée pour un ensemble limité d'activités et dans un contexte précis. Le besoin de cette perspective

est issu du constat que les théories morales objets de l'éthique substantielle, que ce soit sous forme de règles explicites ou sous la forme de règles plus génériques, sont dans l'incapacité de déterminer la décision moralement préférable dans les conditions propres à chaque situation pratique. Chaque profession a de ce fait développé une approche procédurale plus adaptée à aider aux décisions particulières reliées aux problèmes posés par son exercice et dans son contexte.

Chacune de ces quatre perspectives peut être adoptée successivement par toute personne cherchant à explorer le domaine moral, on peut toutefois associer à chacune de ces perspectives une posture caractéristique, et certainement caricaturée, des différents acteurs académiques. La perspective descriptive serait celle des scientifiques, psychologues, sociologues, neurologues, et, plus largement, des scientifiques s'intéressant au comportement humain. La perspective prescriptive, substantielle, serait celle des philosophes moraux engagés dans une théorie morale particulière. La perspective méta-éthique, serait celle des philosophes prenant de la distance en regard de toute théorie particulière, et cherchant à les comprendre et les comparer. Enfin, la perspective de l'éthique appliquée, serait celle des philosophes engagés dans l'action concrète auprès des acteurs d'un secteur donné, par exemple médical, sportif, universitaire ou entrepreneurial.

J'utiliserai ces quatre perspectives pour éclairer plus loin le rapport qu'entretient l'étude du domaine moral avec la démarche scientifique expérimentale, et il conviendra de garder à l'esprit que ce sont des perspectives, donc des points de vue de quelque part à un certain moment, et non des descriptions précises de la position de tel ou tel acteur réel.

L'ampleur considérable de la philosophie morale

La philosophie morale recouvre un domaine d'une ampleur considérable, et c'est ce qui rend l'approche par de multiples perspectives utile. Un parallèle avec le domaine physique est de nature à éclairer cette ampleur. On peut, pour ce parallèle, mettre en regard de la perspective descriptive sur le domaine moral l'étude des phénomènes physiques, dans toutes les dimensions d'observation et d'expérimentation dont les méthodes feront l'objet d'un chapitre suivant. On peut ensuite mettre en regard de la perspective substantielle sur le domaine moral l'élaboration des théories physiques développées par les sciences physiques, par exemple la physique newtonienne ou la relativité, avec une ontologie, des entités, des propriétés et des relations entre ces propriétés. On peut ensuite mettre en regard de la perspective méta-éthique la recherche des justifications de la connaissance scientifique et des méthodes pour la construire, c'est-à-dire la philosophie des sciences physiques. On peut enfin mettre la perspective de l'éthique appliquée en regard des sciences appliquées, à l'articulation avec les

technologies et techniques de mises en œuvre des connaissances apportées par les théories physiques.

Si on accepte la pertinence de ce parallèle avec le domaine de l'étude de la physique, le champ de la philosophie morale s'étend très largement en incluant les équivalents dans le domaine moral de la métaphysique, des sciences physiques, théoriques et expérimentales, de la philosophie des sciences et de l'ensemble des techniques en lien avec les sciences appliquées. Le spectre est extrêmement large, et l'est tout autant la palette des positionnements des philosophes moraux. Il y a peu en commun entre le philosophe moral travaillant en collaboration avec un hôpital pour aider soignants et patients à délibérer sur leurs désaccords moraux, le philosophe moral qui peaufine l'argument de plus qui lui permettra de montrer que la théorie morale qu'il défend est meilleure que cette autre théorie défendue par son collègue universitaire, et le philosophe moral qui mobilise un spécialiste des sciences cognitives en collaboration avec un spécialiste des primates pour montrer que les chimpanzés ont des comportements qualifiables de moraux.

Pour poursuivre la préparation de ma toile de fond esquissant le domaine moral, je propose dans la section suivante d'envisager la description du domaine moral par le biais des désaccords moraux. Cet outil est ici pertinent pour deux raisons, la première est qu'il est raisonnable de penser que les débats de philosophie morale s'ancrent au moins pour partie dans ces désaccords et que les passer en revue est un bon préalable avant d'analyser si et dans quelle mesure la démarche scientifique expérimentale est susceptible de contribuer à leur résolution. La seconde est qu'il m'est apparu nécessaire de distinguer différents types de désaccords qui sont souvent mis sur le même plan alors que les distinguer clarifie les enjeux. Je pense utile de distinguer en particulier les désaccords entre acteurs moraux partageant les mêmes systèmes moraux de ceux opposant des acteurs aux systèmes moraux différents. Utile également de distinguer les désaccords entre agents moraux en situation, des désaccords entre philosophes plus réflexifs et plus distanciés. Ce sont ces distinctions que je vais proposer ci-dessous en complément de la grille d'analyse des quatre perspectives que je viens d'évoquer.

1.3.2 Perspectives et désaccords moraux

J'ai entrepris ci-dessus de situer le domaine moral, puis les grandes familles de théories morales qu'il me semble utile de distinguer. Mon objectif étant d'analyser les apports potentiels de la démarche scientifique expérimentale aux débats de philosophie morale, je prolonge

cette introduction rapide au domaine moral par une tentative de cartographie des désaccords moraux à la source de ces débats[31]. Cette cartographie sera observée en utilisant la grille d'analyse des quatre perspectives : description, prescription, méta-éthique et éthique appliquée, ce qui me permettra de valider cet outil pour examiner ensuite l'apport de l'expérimentation. Si la validation est concluante, j'aurai remplacé la question générique de l'apport de l'expérimentation à la philosophie morale par une déclinaison plus précise : « Qu'apporte cette expérimentation selon telle perspective à la résolution de tel type de désaccord ? ».

Pour cartographier les désaccords moraux, je commence par discerner les désaccords individuels entre personnes ayant le même type d'éducation morale, participant de la même communauté morale, mais en désaccord sur l'application des règles à une situation concrète. Puis, j'envisage les désaccords entre individus relevant de cultures morales différentes avec en particulier les cas concrets dont des traits relèvent du domaine moral pour l'un et pas pour l'autre. Enfin, à un plus haut degré de généralité, les types de désaccords entre philosophes moraux qu'ils soient substantiels ou qu'ils soient méta-éthiques.

1.3.2.1 Les désaccords individuels au sein d'une communauté morale

Un dilemme moral apparaît quand plusieurs règles morales entrent en concurrence dans une situation particulière ou lorsque les circonstances de la situation sont interprétées différemment par différentes personnes. L'acteur moral ne sait laquelle des règles il convient d'appliquer et laquelle il convient d'enfreindre. Les exemples abondent dans les sujets de société. On peut penser à l'avortement : faut-il privilégier la liberté de choix de la mère ou la vie de l'embryon ? A l'euthanasie : laisser souffrir ou aider à mourir ? A l'engagement militant : abandonner sa famille pour servir la cause ou la patrie ? ... Ces dilemmes constituent le quotidien des problèmes moraux et ils fournissent également aux philosophes les cas d'application utiles à illustrer leurs théories. Ils fournissent par ailleurs à tous les auteurs d'œuvres littéraires la trame montrant la tragédie humaine d'avoir ainsi à se déterminer entre actions toutes moralement inacceptables.

Le dilemme moral peut ne concerner qu'un seul acteur moral qui doit se déterminer en son âme et conscience, il peut également en concerner plusieurs qui, ne privilégiant pas le même aspect de la situation, arrivent à des conclusions opposées. Dans les deux cas, traiter le désaccord suppose d'abord l'identification de la situation de désaccord, puis une délibération puis une décision et on peut approcher ce processus complexe selon les différentes perspectives

31. Cette partie de ma réflexion doit beaucoup à Jerôme Ravat et à son ouvrage sur les désaccords moraux (Ravat 2019) [189] ainsi qu'à Joshua Greene et à son ouvrage sur les tribus morales (Greene 2015) [109].

proposées plus haut, descriptive, substantielle, méta-éthique et appliquée.

Dans la perspective descriptive, on pourra, à titre illustratif, identifier les cas de désaccord fréquemment rencontrés, les processus à l'origine des divergences, les règles morales en cause, les différentes façons de délibérer adoptées, et les résultats obtenus. L'apport de la psychologie morale, avec la description des processus mentaux mobilisés sera un élément important dans cette perspective. Dans l'ouvrage « The moral psychology handbook » (Doris 2012) [76] les contributeurs traitent ainsi de différentes questions liées à la psychologie de la décision et du jugement moraux :

— Le comportement moral relèverait de plusieurs niveaux, un niveau rapide lié aux émotions et un niveau lent de raisonnement moral.
— L'examen du caractère motivationnel du jugement moral.
— La morale comme construite à partir des réactions émotionnelles de dégoût et de colère.
— La question de l'altruisme comme anomalie dans la lutte pour la survie individuelle.
— La question de l'existence d'un sens moral qui nous permet d'accéder directement à la dimension morale d'une situation.
— Le rapport entre règles morales et valeurs.
— Le lien entre valeur morale et responsabilité.

Chacune de ces questions peut permettre d'affiner la description des comportements moraux par des observations de situations en laboratoire ou hors du laboratoire du psychologue, indépendamment de toute théorie morale, mais en application des théories psychologiques, la principale différence étant ici que la théorie morale est substantielle en ce qu'elle définit le Bien alors que la théorie psychologique non.

Dans le cas de dilemme moral où les acteurs partagent une même conception morale, la perspective prescriptive d'une théorie substantielle ne portera pas sur le fait d'écarter telle ou telle règle morale, puisque par hypothèse tous les partagent, mais plutôt sur la pertinence à appliquer telle règle morale dans telle situation ou, si plusieurs s'appliquent, sur la façon de les prioriser pour conclure à l'action moralement préférable. Chaque dilemme moral conduira ainsi à affiner les règles, leurs cas d'application et leurs priorités. Jerôme Ravat insiste particulièrement sur le double constat que de telles délibérations ont à la fois pour conséquence de faire évoluer les règles morales mais aussi, en obligeant l'acteur moral à délibérer et à agir en fonction de sa décision, à le forger en tant que personne morale. L'acteur moral doit en effet reconnaître, délibérer et dépasser les dilemmes moraux auxquels il est confronté, seul ou en relation avec d'autres membres de son groupe, et cette expérience construit son histoire morale qui contribue à le déterminer dans ses choix moraux.

Par hypothèse, les personnes morales en désaccord partagent ici le même système moral. Il semble donc en première analyse que la perspective méta, méta-éthique, qui a pour objet d'analyser les différences entre les théories morales et d'examiner leurs fondements soit peu utile dans ce cas. Cette conclusion est certainement valide pour les théories qui sont indépendantes du traitement des désaccords mais elle est à inverser si on adopte le point de vue retenu par Jérôme Ravat. Pour lui, l'histoire des désaccords moraux et de leur traitement joue un rôle majeur dans la construction à la fois de la personnalité morale individuelle et dans la stabilisation des règles morales au sein du groupe, ce qui fait de l'histoire de ces désaccords un élément central de la justification des systèmes moraux.

On peut par exemple remarquer que la complexité des systèmes moraux réels est telle qu'il est peu probable qu'une seule théorie morale appuyée sur un seul fondement soit une option crédible. Ainsi, si certaines règles morales sont d'inspiration divine et d'autres d'inspiration humaniste au sein d'une même société, comme c'est le cas dans notre société chrétienne sécularisée, la façon dont se règlent les dilemmes moraux les opposant sera importante dans la perspective méta-éthique en donnant la primauté à l'un ou l'autre de ces fondements. Par ailleurs, et même si on ne rejoint pas Jérôme Ravat dans son raisonnement extrême, on peut remarquer que les dilemmes moraux sont inévitables et donc considérer que la capacité d'un système moral à les traiter fait partie des caractéristiques importantes que le méta-éthicien se doit d'examiner.

Enfin, la perspective de l'éthique appliquée sera ici particulièrement pertinente. En effet le point de départ, la base morale commune, est supposée acquise et le problème à résoudre est bien celui de la méthode qui permettra d'aider à l'identification des traits moralement importants de la situation, des différences d'appréciation de ces traits et des règles morales à appliquer, et du principe de délibération qui puisse conduire assez souvent à un assez bon résultat moral. Ces désaccords entre personnes d'une même communauté morale sont donc centraux pour les tenants de l'éthique appliquée.

Ces éléments d'analyse permettent de confirmer que la grille d'analyse des quatre perspectives est bien de nature à mieux cerner la problématique visée dans le cas des désaccords moraux au sein d'une même communauté morale :

> Selon quelle perspective, description, prescription, méta-éthique, éthique appliquée, la démarche expérimentale scientifique est-elle susceptible d'aider à l'instruction de ce type de débats moraux entre personnes d'une même communauté morale ?

En suivant ces éléments, il faudra, pour être utile, que la démarche expérimentale apporte des informations dans de nombreuses directions. Par exemple, identifier que la situation est bien telle que les acteurs en désaccord partagent (suffisamment) un même système moral, décrire les éléments du désaccord de façon suffisante pour pouvoir analyser les positions morales de chaque acteur, analyser le lien (éventuel) avec l'application de telle ou telle règle morale, analyser les modes de résolution du désaccord en situation, ... Le niveau de complexité et de difficulté d'une telle démarche apparaît clairement dès cette première formulation et rapproche la problématique de celle plus large de la psychologie expérimentale telle qu'elle apparaît ci-dessus dans la liste des questions abordées dans le « Moral Psychology Handbook ».

Passons maintenant au cas des désaccords moraux où les acteurs ne partagent pas un même système moral.

1.3.2.2 Les désaccords dus à des systèmes moraux différents

Dans le premier cas, j'ai supposé que les acteurs moraux partageaient un même système moral mais étaient en désaccord sur les traits moraux de la situation. Je vais maintenant passer à un cas plus difficile, celui où les acteurs ne partagent pas le même système moral. Je ne me situe pas ici dans un cas pathologique où une des personnes n'a pas du tout de réactions morales, comme certains psychopathes, mais plus simplement quand des personnes de générations différentes ou de pays différents ne sont pas d'accord sur le périmètre de ce qui est moral ou non ou encore croient à des règles morales contradictoires, ce qui est autorisé ou obligatoire chez l'un étant interdit chez l'autre. On pourra utilement discerner ici les oppositions sur des sujets moraux périphériques qui peuvent faire l'objet de désaccord mineurs entre personnes d'accord sur l'essentiel, des oppositions portant sur le cœur des conceptions morales perçues comme irréductibles. Si le désaccord est mineur et la base morale commune suffisante, on sera alors ramené au cas précédent, je me centrerai donc sur la seconde option, le désaccord porte sur des éléments centraux des systèmes moraux, ou considérés comme tels par les acteurs moraux en désaccord.

Pour beaucoup de penseurs moraux, notre monde serait actuellement marqué par le développement exponentiel de ce type de désaccords à proportion de l'augmentation des déplacements des populations de cultures différentes. Par exemple, Joshua Greene dans son ouvrage Moral Tribes (Greene 2015) [109] analyse que les systèmes moraux traditionnels ont permis le développement des sociétés closes sur elles-mêmes au prix de l'invention et de l'exacerbation des luttes de « nous » contre « eux ». Construits pour faciliter la coopération et développer

l'altruisme au sein de petites communautés, ils ont également servi à délimiter ces communautés en établissant des règles de comportement qui permettent d'identifier rapidement qui en est, ou non, membre. Pour Joshua Greene, ces systèmes moraux coutumiers, avec leurs évolutions religieuses, sont aujourd'hui inadaptés à notre nouveau contexte mondialisé qui exige de composer en permanence avec des cultures étrangères.

Les désaccords moraux dus à des systèmes moraux différents portent donc à la fois sur les règles morales et leurs conditions d'application, comme précédemment, mais ils portent également, et peut-être surtout, un enjeu identitaire pour chacun des individus en cause au sein de sa propre communauté morale et pour la société dans laquelle ils ont à résoudre leur désaccord.

La perspective descriptive sur ce type de désaccord est particulièrement pertinente. Elle est également particulièrement difficile puisqu'elle suppose un point de vue externe sur l'objet du désaccord afin de pouvoir décrire les positions des différentes parties prenantes relevant de systèmes moraux différents. Décrire comment ce qui semble une exaction pour l'un est moralement justifié pour l'autre est néanmoins un préalable nécessaire à l'analyse de ce type de désaccord. L'étape suivante est de décrire également s'il y a, ou non, des éléments communs dans les théories morales en jeu permettant d'entrevoir des approches possibles vers une réduction du désaccord moral. L'étude descriptive de l'institution des différentes théories morales en opposition, de leur diffusion et de leur adoption par les parties prenantes est également un aspect important de l'approche de ces cas de désaccord.

La perspective prescriptive d'une théorie substantielle est fondamentalement interne à chaque système moral. Toute personne qui défend une conception du Bien recherche les arguments qui vont dans son sens et, sauf dans les cas rares de conversion, n'entend les arguments des autres théories morales que pour les combattre. Le traitement du désaccord entre systèmes moraux différents n'est donc pas facile pour qui s'attache ainsi à défendre un des système moraux comme étant le meilleur, si ce n'est le seul. On peut avancer que la perspective substantielle, ou plus précisément, les perspective substantielles depuis chacune des parties du désaccord, ne permettront pas d'éclairer un désaccord entre système moraux en vue de sa résolution. Toutefois, s'il y a désaccord c'est qu'il y a un terrain commun sur lequel une action est à entreprendre. Après-tout, si les acteurs n'étaient pas en interaction, chacun restant sur son pré carré moral, le désaccord pourrait n'être jamais apparu. Le philosophe aura donc ici à élargir sa réflexion pour inventer et promouvoir une nouvelle perspective substantielle traitant de ce nouveau territoire moral partagé entre plusieurs communautés.

La perspective méta-éthique permet, en appui de la perspective descriptive, d'analyser

les sources du désaccord au fondement de chacun des différents systèmes moraux. Au delà de cette analyse, la portée de cette perspective sera vite limitée, qu'elle conduise à un relativisme équanime ou, peut-être plus difficile, qu'elle conduise à exprimer une éventuelle supériorité méta-éthique d'un système moral sur un autre, il serait en effet étonnant que ce constat philosophique puisse convaincre une personne morale engagée dans un système moral particulier. En revanche, la perspective méta-éthique sera certainement utile à la construction de la nouvelle perspective substantielle sur le nouveau territoire partagé évoqué ci-dessus.

Dans son ouvrage, Joshua Greene (Greene 2015) [109] propose que cette perspective commune à construire soit l'utilitarisme : comment en effet ne pas être d'accord sur l'objectif qui consiste à maximiser le bonheur sur terre ? Et quelle meilleure base pour entamer la négociation que cet objectif commun ? On pourra néanmoins lui objecter qu'il faut encore que les parties acceptent de s'asseoir à la table de négociation, ce qui peut être simplement interdit par l'un des systèmes moraux, comme le montrent les nombreux exemples de génocides ou de terrorisme dont les auteurs considèrent que le bonheur sur terre ne peut être maximisé qu'en éliminant le système moral pervers de l'adversaire et pour cela l'éliminer physiquement est la meilleure, voire la seule, solution.

La perspective de l'éthique appliquée semble essentielle pour entrevoir une façon de résoudre le désaccord frontal entre deux système moraux. Pour cela, on peut tenter d'adopter une démarche procédurale dans le même esprit que celle de l'équilibre réfléchi car elle permet de ne pas avoir à comparer substantiellement les systèmes moraux opposés. Il faut alors trouver le noyau commun dont pourrait partir la négociation, mais contrairement à la perspective méta-éthique précédente, ce noyau commun peut être limité aux circonstances concrètes du désaccord, il n'a pas à être exprimé sous forme de règles principielles, ce qui semble plus à portée d'une négociation, même si elle reste à l'évidence difficile.

L'analyse de ce deuxième cas de désaccords moraux entre personnes ne partageant pas le même système moral met en avant de nouvelles capacités qu'il faudra développer pour envisager un apport expérimental. Tout d'abord, l'étude descriptive de l'existant suppose une neutralité de l'expérimentateur et de son environnement en regard des différents systèmes moraux en cause qui sera difficile à atteindre, à supposer qu'elle soit atteignable. La question du voile islamique dans l'espace public fournit un exemple frappant de cette difficulté : comment construire un processus expérimental dont les acteurs n'aient pas eux-mêmes un avis sur la question ? Et si, cas peu probable, ils n'en n'avaient pas, ils appartiendraient certainement à une sous-catégorie très particulière du genre humain, et leur capacité à mener une expérimentation devrait être évaluée à cette aune.

Ensuite, si le premier cas de désaccord entre personnes de la même communauté morale ramenait principalement aux difficultés de la psychologie, ce second cas ramène en plus à celles de la sociologie du fait du lien entre systèmes moraux et identification aux différents groupes sociaux. En effet, accepter d'entrer en négociation sur un désaccord moral à propos d'une règle centrale du système moral revient à accepter de remettre en cause son appartenance au groupe social qui a cette règle pour norme [32]. C'est donc à la sociologie, et à la question des multiples appartenances de chacun à de multiples groupes sociaux, ainsi qu'à la psychologie de la constitution de l'identité personnelle, et en particulier de l'identité morale, au travers de ces multiples appartenances que renvoie l'examen des désaccords moraux entre personnes ne partageant pas un même système moral.

Sans être la seule source de difficultés, la question des désaccords moraux, de leur possibilité, de leur examen et de leur résolution a mobilisé les philosophes confrontés à l'analyse des problèmes moraux. Ils en ont développé des approches différentes conduisant à l'ensemble des positions et théories morales que nous connaissons aujourd'hui. Une autre façon de décrire les désaccords moraux est donc d'analyser les différences qu'ils ont, pour partie, induit entre ces théories, c'est l'objet de la prochaine section.

1.3.2.3 Les désaccords entre philosophes moraux

Les débats entre philosophes moraux peuvent être examinés selon chacune des quatre perspectives définies plus haut. Ils peuvent en effet porter sur la description du phénomène moral tel qu'on peut l'observer dans l'espèce humaine (ou d'autres espèces animales si on adopte un point de vue les incluant partiellement ou non en tant qu'acteurs moraux). Ils peuvent porter également, et c'est une partie essentielle de l'enseignement moral, sur la perspective prescriptive d'une théorie substantielle en élaborant puis défendant telle ou telle théorie morale. Ils peuvent également porter sur des désaccords méta-éthiques, les fondements et les justifications des systèmes moraux, dans une position de prise de recul par rapport aux différentes théories morales et, inversement, porter sur des désaccords pratiques liés à l'éthique appliquée et à la mise en œuvre des critères moraux dans une perspective de cas particuliers dans des contextes particuliers.

J'ai rappelé plus haut l'ampleur de la philosophie morale qui couvre les aspects relevant de la métaphysique, de l'épistémologie, de la définition des théories morales et de leur application dans le cadre des éthiques sectorielles. Les désaccords entre philosophes moraux

32. Voir à nouveau, pour exemple, l'ouvrage « Moral Tribes » de Joshua Greene pour une description précise de la morale en tant que définition de « nous » contre « eux ».

peuvent concerner tous les aspects de ce très large panorama et chaque philosophe moral choisit au sein de ces multiples possibilités son positionnement, son camp et ses adversaires. Pour tenter de faciliter la lecture d'ensemble de cet espace et à titre de résumé de ce qui précède, dans l'objectif qui est le mien de tenter de le cartographier pour faciliter le passage à l'analyse de l'apport de l'expérimentation dans les débats moraux, je propose de distinguer plusieurs embranchements possibles entre philosophes moraux à partir de l'acceptation commune de l'existence des énoncés évaluatifs.

La métaphore spatiale est utilisée intensément par Jérome Ravat pour expliciter les comportements moraux dans son ouvrage sur les désaccords moraux (Ravat 2019) [189]. Chacun aurait une carte mentale de ce qui est permis, autorisé, interdit ou obligatoire sous la forme d'une carte des chemins moralement admis qui viendrait se superposer sur le territoire des actions possibles. A chaque embranchement, l'acteur moral choisit son chemin en suivant ce que sa boussole morale lui dicte et il y a désaccord moral entre acteurs quand ces cartes diffèrent, quand des chemins incompatibles se percutent, sortant les acteurs de leurs chemins confortablement tracés.

Je reprends ici cette métaphore spatiale en la déplaçant pour expliciter les embranchements qui conduisent chaque philosophe moral à opter pour une conception particulière du domaine moral. Reprenons donc à partir du point de départ commun. Pour établir des dissensus, il faut être d'accord sur le sujet du dissensus. Un point de départ commun à toute préoccupation morale peut être proposé : il existe des énoncés évaluatifs portant un jugement « ceci est Bien » ou « ceci est Mal ».

J'ai proposé plus haut de donner à un tel énoncé la forme générale suivante :

> O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien.

J'ai posé ici O comme observateur externe, le philosophe ou le psychologue entreprenant l'étude du domaine moral. E est un énoncé contenant un terme évaluatif, X son énonciateur et S le destinataire. A est l'acte jugé fait par Y à Z. Et j'appelle « situation » l'ensemble 'XSEAYZC'.

A partir de cette base commune à tous les philosophes, il existe des énoncés évaluatifs, je propose cinq embranchements pour caractériser différentes dimensions du phénomène moral vues par les philosophes. Chaque embranchement ouvre vers de multiples possibilités que je caricature en nommant l'embranchement d'après les choix extrêmes, ou très contrastés, qui se présentent. Chaque philosophe moral choisit son orientation à chaque embranchement et construit une combinaison qui structure sa théorie morale. Les choix à chaque embranche-

ment ne sont ni complètement indépendants, car certaines combinaisons seraient irrationnelles, comme par exemple supposer que la morale soit une fiction et qu'elle soit d'origine divine, ni strictement hiérarchiques, car de nombreuses variantes sont possibles.

Premier embranchement : fictionnalisme vs. propriété morale

Certains philosophes considèrent les énoncés E comme fictionnels, c'est-à-dire faisant partie, au même titre que les romans, les mythes, les contes, etc. de ce qui construit la relation entre les humains au sein d'une même culture sans autre lien direct avec la réalité que cette relation (fictionnalisme). D'autres philosophes considèrent que ces énoncés se réfèrent à une propriété réelle de l'un ou l'autre des éléments de la situation et principalement de X, Y ou A dans le contexte C (propriété morale).

Pour les fictionnalistes, la partie « O observe que X dit à S » est la plus importante, puisque le rôle d'une fiction est de communiquer à S quelque chose qui est totalement inclus dans l'énoncé E et dont l'objectif est de mettre S dans un certain état (émotionnel par exemple). Du point de vue du philosophe moral qui pense que E réfère à une propriété morale, c'est la partie E qui est la plus importante car elle porte l'information sur le monde que X souhaite partager, et plus précisément sur la situation morale (au sens ci-dessus). Les embranchements suivants restent pertinents dans les deux cas, mais les significations profondes en sont changées selon qu'on soit fictionnaliste ou attaché à l'existence réelle de propriétés morales. Notons que le qualificatif « réaliste », et l'expression « réalisme moral » peuvent changer de signification selon les philosophes et, comme je l'exposerai plus loin avec l'exemple de Jesse Prinz, il convient d'en préciser le contenu au cas par cas.

Deuxième embranchement : ce qui définit une situation comme « morale »

L'énoncé E portant jugement moral peut être schématisée en : « X dit qu'il juge Bien A, A étant ce que Y fait à Z dans les circonstances C ». Chaque philosophe peut définir le grain et les ensembles des X, A, Y, Z et C qui qualifient ou non le jugement E de « moral ». Cet embranchement donne lieu à de très nombreuses possibilités au-delà de la présence du terme évaluatif dans E. A titre d'illustration :

— Qu'est-ce qui qualifie Z en tant que bénéficiaire ou victime d'un acte à portée morale ?
— Qu'est-ce qui qualifie Y en tant qu'acteur moral ?
— Qu'est-ce qui qualifie A en tant qu'acte à portée morale ?
— Qu'est-ce qui qualifie X en tant que juge moral compétent ?
— Et pour qui X doit-il être compétent, X, S ou O ?
— Un jugement n'est-il moral que s'il est porté par l'acteur ? (X = Y)
— Un acte moral suppose-t-il un tiers victime ? (Y <> Z)

— La qualification de « morale » pour la situation résulte-t-elle du cumul des qualifiants atomiques ci-dessus ou porte-t-elle sur l'ensemble de la situation, tous les traits en étant intimement liés ?

Chacune de ces questions se décline ensuite en de multiple sous questions dont les raffinements remplissent les bibliothèques des philosophes moraux. Un survol rapide du lexique des théories morales en annexe montre quelques unes des très nombreuses options qui ont été explorées à partir de cet embranchement considérant ce qu'il faut ou non inclure dans le domaine moral.

Troisième embranchement : il existe ou non des règles morales

Certains philosophes considèrent qu'il y a des règles morales qu'on peut mettre en relation avec les situations (XSEAYZC). D'autres, comme les situationnistes, considèrent que c'est impossible, que chaque situation est particulière et qu'elles ne peuvent être décrites en termes suffisamment génériques pour faire l'objet de règles.

Cet embranchement ouvre de nombreuses possibilités selon les différentes extensions possibles de ces règles, leurs variations ou non d'un groupe à l'autre et selon les modes de relation entre la règle générique et la situation particulière. L'extension envisagée pour la règle peut être celle d'un groupe, on parle de communauté morale, elle peut être étendue à une société, une culture, ou même à l'humanité entière. L'universaliste considérera que l'extension des règles morales est maximale, par exemple parce qu'elle est liée à la nature humaine, alors que le relativiste culturel considérera que l'extension en est celle des sociétés qui ont développé ces règles. Le lien entre la règle générale et la situation particulière peut également prendre de multiples formes, on peut voir par exemple la règle morale comme une abstraction construite à partir des jugements portés dans différentes situations. On peut, inversement, considérer que la situation est à juger en application de la règle.

Quatrième embranchement : ce qui définit le Bien et le Mal

Après ces trois premiers embranchements, le philosophe a choisi un cadre général pour sa théorie morale mais n'est pas rentré dans le vif substantiel du problème moral : ce qu'est ce Bien qui fait que X dit que l'acte A est Bien et que S comprend cet énoncé. C'est l'objet de ce quatrième embranchement. L'amplitude des réponses possibles est telle qu'il serait illusoire de tenter ici même un simple survol des types de désaccords qu'elles peuvent générer, je renvoie donc à l'esquisse des grandes familles de théories morales proposée plus haut et au petit lexique des théories morales en annexe. Il convient ici de préciser que les philosophes distinguent trois niveaux différents qui interfèrent avec cette définition du Bien : le niveau métaphysique, le niveau épistémique et le niveau sémantique. A la métaphysique

correspond d'analyser ce qu'est ce Bien, s'il est une propriété, et si oui, de quelle entité peut-il être propriété, et plus largement, de quel type de propriété il s'agit. A l'épistémologie correspond d'analyser ce que les humains peuvent en connaître, comment ils peuvent construire des croyances ou des connaissances à propos de ce Bien. Et enfin, à la sémantique de faire le lien entre ce que X sait, ce que X dit en prononçant E, et ce que S en comprend. Une définition satisfaisante du Bien aura à répondre des préoccupations soulevées à chacun de ces trois niveaux.

Cinquième embranchement : ce qui explique, justifie, fonde, les théories morales

Lorsque ces théories morales, conception du Bien et règles morales, existent, et que le philosophe moral veut les expliciter et les analyser, se pose la question de leur origine (d'où viennent-elles ?), et de leur justification (pourquoi cette règle là et pas une autre ?) et, surtout, se pose la question de ce qui leur donne la capacité à mobiliser l'obligation morale chez des individus supposés autonomes. Cet embranchement offre plusieurs voies, je citerai ici les voies religieuses, biologiques, humanistes et culturelles, chacune ayant à nouveau de multiples ramifications.

Nous avons vu avec les dix commandements un exemple de la voie divine. Dans les traditions où le Dieu est créateur, Dieu a créé le monde, les humains et les règles morales que ces humains doivent respecter et le choix est simple : le respect des règles morales ou la damnation.

Une autre voie de justification liée à la création est celle des théories évolutionnistes, que je détaillerai dans un chapitre ultérieur. L'espèce humaine telle qu'elle est à ce jour a évolué pour être telle qu'elle est, et la vie en société avec la morale fait partie de cet état. On peut rechercher les origines et la justification de la morale dans les opportunités et les contraintes évolutives qui ont marqué notre espèce dans son environnement. Selon ces théories, nous respectons les règles morales au même titre que nous marchons sur deux pieds et utilisons nos dix doigts, parce que cela nous est naturel du fait de notre histoire évolutive.

Une autre façon de justifier le respect des règles morales par la nature humaine est de l'appuyer sur la rationalité dont ferait preuve notre espèce. Cet argument humaniste peut être résumé de la façon suivante, et je reprends ici l'argumentation développée par Francis Wolff dans (Wolff 2017) [237]. Le point de départ de l'argument est aristotélicien : le Bien pour un être est d'agir conformément à sa nature. Le développement de l'argument conduit ensuite à définir cette nature de l'humain qu'il convient de respecter : un animal conscient, parlant, social, rationnel et dialogique. A chacun de ces qualificatifs correspond un type de Bien qui s'affine et qui, en retour, contribue à préciser ce que signifient ces qualificatifs :

— En tant que vivant le Bien est de continuer à vivre.
— En tant qu'animal, le Bien est d'assurer un fonctionnement de l'individu et de l'espèce s'adaptant aux modifications du milieu.
— En tant qu'animal conscient, le Bien est le plaisir qu'offre l'expérience phénoménale.
— En tant qu'animal parlant, l'humain peut parler du Bien et jouir de se sentir libre de le respecter, ce qui devient partie de ce Bien.
— En tant qu'animal social dialogique, il peut parler de son Bien aux autres, exprimer la justification de ses croyances et chercher à les partager en créant le discours sur les valeurs morales et les règles pour les maximiser.

C'est donc du fait de sa nature que l'espèce humaine a développé cette capacité à construire des valeurs morales, à les dire et à les défendre. Or, et c'est la dernière étape du raisonnement de Francis Wolff, s'il a accédé à cette capacité, il ne peut qu'agréer l'expérience de pensée inspirée par John Rawls [33] du choix des règles morales par des discutants impartiaux sous le voile d'ignorance. Et de cette expérience, il est trivial de déduire du fait que les discutants ne savent pas le rôle qui leur sera alloué que l'éthique choisie respectera la règle d'or de la réciprocité, qu'elle supposera l'égalité de tous les hommes et que chacun aura un même accès au bonheur. La justification du respect des règles morales est ainsi reliée à la nature humaine et à la nécessité métaphysique de respecter la nature de son être.

J'évoquerai enfin une dernière voie qui consiste à voir la morale comme une construction humaine culturelle au même titre par exemple qu'un artefact sophistiqué, disons le réseau Internet, amorcée initialement par la simple formulation de la solidarité nécessaire au sein d'un groupe d'une dizaine d'humanoïdes chasseurs cueilleurs et qui s'est ensuite développée progressivement comme se sont développés tous les autres aspects de la culture, pour faire face à la croissance des groupes et à leur intégration en sociétés complexes. La proposition de Philip Kitcher de voir la morale comme un projet éthique (Kitcher 2014) [134] est un exemple de cette voie. En ce sens, le phénomène moral est le résultat d'un processus cumulatif de collaboration au sein de communautés humaines dont il a suivi et suit encore les évolutions. La morale contribue à faciliter cette collaboration en instaurant un climat de confiance appuyé sur la prévisibilité des comportements obtenue par le respect des règles morales instaurées dans le groupe. A chaque nouvelle situation de collaboration correspond l'instauration de nouvelles règles morales, ce qui explique que les règles morales accompagnent les évolutions

33. Pour John Rawls une société pourrait être juste si elle suivait des règles choisies par un groupe de discutants représentatif travaillant sous le voile d'ignorance, c'est-à-dire sans connaître le rôle qui serait alloué à chacun dans la société [191]. Bien que conçue pour définir le juste, et non le moral, le thème du voile d'ignorance est fréquemment repris dans ce contexte moral, marquant la proximité des deux notions : dans nos sociétés une règle injuste peut difficilement être qualifiée de morale.

sociétales comme, par exemple, l'abolition de l'esclavage au 19ème siècle ou la libération des mœurs dans les années 1960.

Les philosophes moraux ont ainsi proposé de nombreuses explications du phénomène moral, et en particulier des raisons du respect des règles morales par des humains qu'on voit par ailleurs également capables du plus grand égoïsme. Ces explication vont de la morale divine, la morale naturelle, la nature humaine, aux pures conventions sociales. Remarquons que ces multiples possibilités de réponse à la question du pourquoi du respect des règles qui constitue ce cinquième embranchement a toute son importance pour tout penseur du domaine moral, y compris pour le nihiliste moral dès lors qu'il a accédé au point commun initial en reconnaissant qu'il existe des énoncés évaluatifs et que les humains agissent en ressentant ces énoncés comme induisant des obligations qu'ils qualifient de morales. Il faut que le philosophe moral, même s'il considère que les théories morales et les règles morales ne sont que des bavardages sans objet, propose ce qui explique, justifie, fonde les énoncés évaluatifs utilisés par les humains dans leurs jugements et actions quotidiens.

Rappelons les cinq embranchements que je viens de décrire :

— Le choix entre fictionnalisme et réalité des propriétés morales.

— Ce qui définit une situation comme relevant ou non du domaine moral.

— Il existe ou non des règles morales.

— Ce qui définit substantiellement le Bien et le Mal.

— Ce qui explique, justifie, fonde, les théories morales.

Ma proposition est que le parcours au travers de ces cinq embranchements caractérise un chemin descriptif possible vers les théories morales et permette ainsi de situer les désaccords entre philosophes moraux. Pour évaluer cette proposition, je vais maintenant décrire deux théories morales très différentes, le réalisme moral de David Enoch et le sentimentalisme constructif de Jesse Prinz. J'ai choisi ces deux théories car elles sont récentes, bien documentées et représentatives de deux courants de pensée très importants, le sentimentalisme et le réalisme et que, bien qu'opposées sur l'échiquier des théories morales, elles disposent toutes deux de forts arguments en leur faveur.

1.3.3 Le réalisme moral de David Enoch

Dans son ouvrage « Taking Morality Seriously : A Defense Of Robust Realism » (Enoch 2013) [80], David Enoch décrit un réalisme moral objectiviste : les faits moraux existent en dehors de nous, indépendamment de ce que nous en pensons. En reprenant les notations ci-

dessus, il défend donc que, essentiellement, si X dit à S par l'énoncé E que l'acte A qu'à fait Y à Z est Mal, c'est que l'acte A a bien la propriété objective d'être Mal et que, X peut le savoir et le communiquer à S.

Au premier embranchement, fictionnalisme ou non, David Enoch pose clairement sa préférence pour l'existence des propriétés morales et, de plus, pour le choix de les faire porter par l'acte A. Il offre principalement deux arguments pour étayer son réalisme, le premier est lié aux désaccords moraux, le deuxième est un argument d'indispensabilité. Le premier argument peut être résumé ainsi : si les désaccords moraux existent et si nous défendons parfois avec virulence nos décisions contre des opposants sous l'angle moral, c'est bien que nous pensons que les faits moraux existent et s'imposent aussi bien à nous qu'à eux. Si la morale était subjective, chacun aurait un point de vue qui lui serait propre et également valide et il ne pourrait y avoir de désaccord moral. Le second argument d'indispensabilité est construit sur la délibération morale : lorsque nous délibérons sur les questions morales entre plusieurs options, que ce soit seuls ou avec d'autres acteurs, nous recherchons les arguments en faveur ou en défaveur de chaque option. L'existence des propriétés morales de chacune de ces options est indispensable à notre délibération. Comme la délibération est un élément incontournable de notre expérience morale et comme l'existence de faits moraux objectifs est indispensable à cette délibération, alors le réalisme moral est la solution à retenir. On peut noter ici le parallèle que fait l'auteur avec l'argument d'indispensabilité des entités théoriques des sciences naturelles : bien qu'inobservables directement, comme les propriétés morales, nous en acceptons l'existence car elles sont indispensables à la bonne formulation de nos théories.

Au deuxième embranchement, ce qui relève ou non du domaine moral, la théorie de David Enoch a une réponse ontologique et une réponse épistémique. La réponse ontologique est que les propriétés morales sont réellement des propriétés des objets du monde. La réponse épistémique est que nous pouvons connaître ces propriétés morales. La réponse ontologique découle immédiatement de sa conception objective et réaliste : le domaine moral est défini en dehors de nous par les propriétés morales du monde. En revanche, sa position est plus complexe sur le plan épistémique. En effet, il doit accepter que ces propriétés morales des objets du monde sont sans puissance causale car une telle puissance causale d'une propriété qui n'est pas dans le monde physique serait très peu plausible. Étant sans puissance causale, elle ne peuvent être objet de notre perception. La capacité que nous avons néanmoins à connaître les propriétés morales des objets du monde auxquelles la perception ne nous donne pas accès est alors problématique. Pour David Enoch, elle est un acquis évolutif : c'est parce que reconnaître le Bien et le Mal là où il se trouve constitue un avantage évolutif que l'espèce humaine en est

aujourd'hui doté. Cet argument explique également pourquoi la morale est motivationnelle : elle nous pousse à bien agir et c'est pour cela que l'évolution a pu la favoriser.

Le troisième embranchement, il existe ou non des règles morales, relève pour le réaliste de la même logique que la question de l'existence ou non de lois de la nature. Les différentes options doivent être évaluées à la fois sous l'angle ontologique, quelles sont les relations possibles que nous admettons dans notre ontologie des propriétés morales, et sous l'angle épistémique, quelles sont les relations que nous pouvons connaître qui, soit correspondent à de réelles relations du niveau ontologique, soit sont simplement des relations fréquentes observables qui nous permettent de mener des raisonnements moraux.

Le quatrième embranchement, ce qui définit le Bien et le Mal, est ici particulièrement simple et directe : le Bien et le Mal appartiennent à la réalité des propriétés morales du monde. En revanche le problème de notre capacité à les connaître est, à nouveau, plus problématique et la réponse de David Enoch est qu'elle est le résultat de la capacité évolutive présentée plus haut. La chute de ce raisonnement est que notre intuition en matière morale est fiable car fiabilisée par l'évolution : le Bien et le Mal sont ce que nous pensons intuitivement qu'ils sont.

Enfin, le cinquième embranchement, la justification de la théorie morale, repose principalement sur les deux arguments donnés en faveur du réalisme moral : la possibilité du dissensus et l'argument d'indispensabilité pour la délibération morale. David Enoch présente également des arguments contre les théories concurrentes et en particulier le fictionnalisme, et toute forme de conception subjectiviste de la morale, dont l'émotivisme et le sentimentalisme dont je vais présenter plus loin un représentant avec Jesse Prinz. Pour David Enoch ces théories ne prennent pas en compte « sérieusement » la moralité en ne pouvant établir que, par exemple, il est absolument mal de torturer un enfant, et ce indépendamment de toute subjectivité.

Pour conclure cette présentation rapide du réalisme moral de David Enoch utilisant les cinq embranchements proposées, remarquons que David Enoch privilégie fortement, entre les quatre perspectives (pour mémoire, les perspectives descriptive, prescriptive, méta-éthique et appliquée), la perspective prescriptive qui est logiquement reliée à son réalisme fort. Le Bien étant une propriété réelle du monde, et cette propriété nous étant accessible par notre capacité à l'intuition morale que nous devons à l'évolution, nous devons respecter les règles morales de la société dans laquelle nous vivons et toute approche qui risque d'instruire la décision inverse, qu'elle soit descriptive ou méta-éthique ne peut qu'être nuisible et conduire à ne pas considérer la moralité « sérieusement ».

1.3.4 Le sentimentalisme constructif de Jesse Prinz

Pour illustrer l'usage de ces cinq embranchements dans la description d'une théorie morale très différente du réalisme de David Enoch, j'ai choisi ci-dessous la théorie du « sentimentalisme descriptif » de Jesse Prinz, décrit dans (Prinz 2009) [185] d'une part parce qu'elle est à ce jour une des théories les plus en vue et, d'autre part, parce qu'elle permet de mettre en avant la complexité des réponses face aux cinq embranchements ci-dessus et, ainsi, l'intérêt analytique de les expliciter.

Essentiellement, cette théorie morale est un sentimentalisme au sens où l'énoncé moral E est l'expression d'un sentiment d'approbation ou de rejet de la part de X en regard de l'acte A fait par Y. Je reprends à nouveau ici la notation proposée plus haut pour notre point de départ commun à toutes les théories : « O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien. »

O est donc ici pour nous Jesse Prinz, philosophe observateur de la nature humaine.

Premier embranchement, le sentimentalisme constructif est-il un fictionnalisme ou un réalisme ? En première analyse, il semblerait qu'un tel sentimentalisme soit difficilement compatible avec un réalisme moral puisque le jugement moral dépend essentiellement de la réaction émotionnelle de X à la situation et non de cette seule situation. Mais Prinz défend le contraire avec deux arguments. Premier argument, si on considère comme réel ce qui a une puissance causale, alors bien sûr le sentiment de rejet de X a bien une telle puissance en le faisant agir dans le sens de ce rejet. Il y a bien une réalité du rejet moral. Deuxième argument, si X a ce sentiment, et si on suppose qu'il ne l'a pas à tord, c'est que ce que Y a fait à Z a des propriétés bien réelles telles que le sentiment de rejet de X en a été induit. Ces propriétés sont inductrices du rejet moral. Elles ne sont pas morales en elles-mêmes car le sentiment de rejet moral dépend de X, et une autre personne pourrait ne pas le ressentir face à la même situation. Et enfin, ces propriétés inductrices ne le sont pas aléatoirement. Face à des situations semblables, et s'il est lui-même dans les mêmes dispositions psychologiques d'attention et d'implication, X réagit assez souvent de la même façon. Donc pour Prinz, son sentimentalisme est bien un réalisme moral au sens où le sentiment moral s'ancre dans la réalité à la fois par les propriétés de la situation inductrices de ce sentiment et par les conséquences comportementales de ce sentiment sur X, et donc sur E et S.

Deuxième embranchement, ce qui définit la situation comme « morale ». Dans la théorie de Prinz, et c'est le sens du qualificatif de « constructiviste » qu'il donne à son sentimentalisme, la définition du domaine moral est toute entière contenue dans le fait que X a ce sentiment

de rejet moral à la suite d'une double construction, celle de l'individu X qui au cours de son éducation morale apprend et intègre ce qui doit être considéré comme moral, immoral ou indifférent dans sa communauté morale, et celle de la société en son entier qui construit historiquement des règles morales dans un processus d'auto-régulation. Le domaine moral est le résultat de ce double processus historique individuel et collectif.

Troisième embranchement, les règles morales et leur extension. Pour Prinz, les règles morales existent, et tout son sentimentalisme constructif serait incompréhensible si les jugements moraux de X face à une situation n'étaient pas dépendants des règles qui se sont construites dans sa société, règles qu'il a apprises et intégrées lors de son éducation. Que cette intégration se traduise par une disposition à avoir telle réaction émotionnelle face à telle situation et qu'il n'y ait pas de mise en œuvre d'un raisonnement n'empêche pas que cette règle acquise et intégrée existe en tant que telle. Sa théorie suppose également la notion de communauté morale, le groupe humain au sein duquel se sont développées ces règles et qui sert de matrice à l'éducation de X.

La théorie du sentimentalisme constructif est relativiste : les règles morales sont relatives à ces communautés morales. Néanmoins, les processus historiques qui ont vu la genèse de ces règles dans chaque communauté ne sont certainement pas arbitraires, ils sont tous soumis à un même faisceau de contraintes lié à la nature humaine de ces sociétés. C'est ce faisceau de contraintes partagé qui explique les nombreux points communs entre les diverses communautés morales, comme la présence quasi générale de la règle d'or de la réciprocité. Sans ce faisceau de contraintes partagé, ces similarités seraient inexplicables.

Quatrième embranchement : la définition du Bien. Pour Prinz, il n'y a pas de définition du Bien ou du Mal sans référence à ce qui induit chez X la réaction émotionnelle d'approbation ou de rejet moral. Néanmoins, il existe bien des propriétés inductrices de cette réaction que X perçoit dans la situation, c'est-à-dire dans l'acte A que Y a fait à Z. Ces propriétés inductrices peuvent ainsi avoir une connotation morale indirecte en ce qu'elles sont inductrices pour X, et peut-être pour (presque) tout X ayant subi le même formatage moral, et même pour (presque) toute société lorsqu'il s'agit de règles largement partagées comme la règle d'or de réciprocité.

Cinquième embranchement : Ce qui justifie, fonde la théorie morale. La théorie de Prinz décrit un processus historique en deux temps. Le premier est celui de l'éducation morale de l'individu X qui acquiert et incorpore les réactions émotionnelles qu'il convient d'avoir au sein de sa communauté morale. X se doit de respecter les règles morales de sa communauté au titre de son appartenance à cette communauté. Insistons sur le fait que pour X ce respect n'est pas le résultat d'un raisonnement qui le conduirait à choisir d'appartenir à cette communauté

et à marquer ce choix en adhérant intellectuellement à des règles morales. Ces règles morales sont inscrites dans la disposition qu'a X à réagir émotionnellement à certaines situations, situations qu'il perçoit directement du fait de son éducation morale. Le jugement moral de X, en tant qu'il est moral, est avant tout émotionnel et peut passer secondairement par une étape réflexive qui n'est pas indispensable.

Le second processus historique que présente Prinz est celui de l'évolution culturelle de la société qui construit les règles morales en se construisant elle-même. En ce sens le sentimentalisme constructif est compatible avec les nombreuses théories qui décrivent la morale comme une des composantes des cultures humaines dont la justification est justement de stabiliser ces cultures. Il y a en somme co-création des règles morales et de la communauté morale. Pour Prinz, cette évolution morale et culturelle a pris une telle ampleur en fusionnant avec celle des capacités du langage humain, qu'il est peut-être vain de tenter de rapprocher les phénomènes moraux complexes tels qu'ils nous apparaissent aujourd'hui des conditions des premiers balbutiement moraux issus de la phylogenèse. De même qu'il serait vain de tenter de comprendre l'urbanisme d'une grande ville moderne en observant un abri dans une grotte.

Cet examen selon les cinq embranchements proposées permet de mettre en lumière plusieurs traits de la théorie du « sentimentalisme constructif » de Prinz. Celle ci définit le domaine moral à partir des réactions émotionnelles d'approbation ou de rejet moral qui résultent d'une situation en fonction de l'éducation morale reçue. En utilisant les cinq embranchements proposées nous avons pu en détailler plusieurs caractéristiques, son réalisme tel que défendu par Prinz, son relativisme affirmé mais tempéré par la nature humaine commune, sa justification humaniste et historique par la double formation de la société, en tant que communauté morale et éducative, et de l'individu en tant que personne morale intégrant les règles morales de sa société sous la forme d'une disposition à déclencher des réactions émotionnelles d'approbation ou de rejet moral.

Chacune de ces caractéristiques peut être débattue, contestée, selon chacune des perspectives proposées plus haut, en tant que théorie descriptive du phénomène humain, dans la perspective substantielle des règles morales qui sont privilégiées par le sentimentalisme descriptif, dans la posture méta-éthique et, enfin, dans sa capacité à étayer des éthiques appliquées utiles ici et maintenant dans des domaines spécifiques. La théorie du sentimentalisme constructif de Jesse Prinz se veut néanmoins avant tout descriptive du domaine moral dans son ensemble et privilégie ainsi les perspectives descriptives et méta-éthiques.

Les deux outils d'analyse proposés, les quatre perspectives sur le domaine moral et les

cinq embranchements des désaccords entre philosophes moraux, n'épuisent certainement pas tout ce qu'il conviendrait d'analyser en regard des théories de Jesse Prinz ou de David Enoch, mais je pense avoir montré qu'elles en donnent une première vision de nature à me permettre d'aborder la question centrale de cette thèse, la possibilité d'un apport de la démarche scientifique expérimentale à la résolution de ces désaccords.

1.4 Le domaine moral, point d'étape

En conclusion de cette rapide présentation du domaine moral, je reformule les points que je propose de retenir de cette présentation.

Premier point, si le phénomène moral en tant qu'utilisation de propositions évaluatives est omniprésent et important dans la vie humaine, définir précisément ce domaine moral soulève de multiples difficultés et à chaque théorie morale correspond sa propre définition du domaine moral.

Deuxième point, prenant acte de cette difficulté, j'ai suggéré d'adopter à titre provisoire comme base de départ de l'étude du phénomène moral une définition appuyée sur les énoncés évaluatifs :

« O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien. »

Le domaine moral est ensuite défini comme l'ensemble des situations 'XSEAYZC' qui ont donné lieu, réellement ou potentiellement, à un énoncé évaluatif.

Troisième point, après une évocation rapide des différentes grandes familles de théories morales et des critères qui ont pu être proposés pour les évaluer, j'ai constaté que chacune privilégiait un certain type de critères et que, conséquemment, ces critères ne permettaient pas le travail transversal que je souhaite mener.

Quatrième point, j'ai présenté quatre perspectives et cinq embranchements en tant qu'outils d'exploration de ce domaine moral.

Les quatre perspectives sont :

— L'étude descriptive du domaine moral, individuel et social.

— La prescription d'une morale substantielle, appuyée sur une théorie morale particulière, descriptive et prescriptive, qui vise à dire ce qu'est le Bien et ce qui est Bien.

— La philosophie morale dans la posture méta-éthique de comparaison des différentes théories morales.

— L'éthique appliquée à des domaines spécifiques, la santé, les affaires, le sport, etc. et à

des cas particuliers.

Les cinq embranchements sont :

— Le choix entre fictionnalisme et réalité des propriétés morales.

— Ce qui définit une situation comme relevant ou non du domaine moral.

— Il existe ou non des règles morales.

— Ce qui définit substantiellement le Bien et le Mal.

— Ce qui explique, justifie, fonde, les théories morales.

Mon objectif est que ces outils soient utiles pour faciliter l'analyse des désaccords entre philosophes moraux, comme entre acteurs moraux, de façon à rendre explicites les apports potentiels éventuels de la démarche scientifique expérimentale à leur résolution et les conditions qui rendraient ces apports potentiels possibles selon chacune des perspectives envisagées.

Dans la mesure où ces définitions, perspectives et embranchements sont assez larges pour embrasser l'ensemble des penseurs du domaine moral, on peut certainement les voir comme une proposition méthodologique participant de la perspective descriptive, sans vue substantielle. Elles portent néanmoins, selon la perspective méta-éthique, la marque de la prédominance de la perspective descriptive sur la perspective substantielle, ce qui n'est guère surprenant pour qui s'intéresse à l'apport des démarches expérimentales.

Chapitre 2

Le mouvement XPhi de philosophie expérimentale

Le chapitre précédent m'a permis de dresser une trame de fond esquissant le domaine moral, les morales substantielles, les théories morales, les dissensus entre agents moraux et entre philosophes moraux, en un mot le large domaine de la philosophie morale. C'est en prolongement de cette trame de fond que je vais dans le présent chapitre proposer une description de l'usage de la démarche expérimentale en philosophie morale, et, pour cela, je vais m'appuyer sur le mouvement XPhi de la philosophie expérimentale. Ce mouvement ne recouvre bien sûr pas toutes les possibilités qu'a un philosophe moral d'intégrer, utiliser, critiquer ou dénier les résultats des démarches expérimentales en lien avec son domaine d'étude. Néanmoins, il présente comme nous allons le voir dans ce chapitre, le double intérêt de correspondre à une période récente qui, avec le développement de l'imagerie neuronale d'un côté et d'Internet de l'autre, a vu la multiplication des opportunités expérimentales très médiatisées ouvertes et, par ailleurs, a effectivement donnée lieu à nombre de débats philosophiques, tant sur la qualité des apports expérimentaux que sur leur pertinence en regard de ces débats. Voyons ici le mouvement XPhi comme un échantillon (assez) représentatif des études utilisant l'approche expérimentale au profit de la réflexion philosophique.

Dans une première section je présenterai la jeune histoire de ce mouvement XPhi qui comporte de nombreux programmes de recherche couvrant toutes les branches de la philosophie. Beaucoup de philosophes ont retenu comme caractérisant les XPhi un programme dit « négatif », qui vise à montrer que la méthode des cas utilisée par les philosophes analytiques n'a pas la fiabilité nécessaire à la résolution des questions qu'ils visent à instruire. Ce pro-

gramme « négatif », perçu par certains comme une accusation d'irrationalité, n'est ni le seul ni, peut-être, le plus important pour mon étude. Je présente ensuite des résultats significatifs que le mouvement XPhi a permis d'atteindre et j'ai choisi pour cela de commencer par deux domaines centraux de l'épistémologie, la définition de la connaissance et celle de la référence, puis d'évoquer très rapidement l'ensemble des domaines abordés par les philosophes se réclamant du mouvement XPhi. Ce détour par des questions qui ne relèvent pas, prima facie, de la philosophie morale est nécessaire pour que les analyses suivantes, qui seront centrées sur la philosophie morale, n'occultent pas une caractéristique importante du mouvement XPhi : il s'agit d'expérimenter sur le comportement des humains quand ils raisonnent, et ce dans tous les domaines. Que ce soit celui des philosophes (et c'est le programme « négatif »), ou celui des êtres connaissant (la connaissance et la référence), ou, de façon générale, relativement à tout objet de réflexion, morale ou non, les expériences des XPhi portent principalement sur ce comportement humain qu'on appelle raisonnement. Cette caractéristique se déclinera ensuite en regard de la philosophie morale : le philosophe moral expérimental peut se lire alors comme étant soit au-dessus du philosophe moral, et jugeant de la pertinence de ses outils, soit à côté de lui, et analysant le raisonnement des humains quand ils pensent à des questions de philosophie morale, ou à des questions de morale, ou encore, plus quotidiennement, quand ils agissent en contexte moralement chargé.

J'insisterai ensuite sur les réticences que de nombreux philosophes ont exprimées quant à ces résultats des XPhi, et quant à l'utilisation de la démarche expérimentale telle qu'elle est pratiquée par les tenants de ce mouvement. Les réticences les plus fortes, et donnant lieu à des travaux explicitement orientés contre les résultats du mouvement XPhi, visent surtout le programme « négatif » qui s'attaque à la philosophie analytique au travers de l'usage de l'intuition dans la méthode des cas. Ces réticences sont transversales, concernent toutes la philosophie, et ne sont pas spécifiques à la philosophie morale. Mais d'autres types de réticences sont plus ciblées sur tel ou tel domaine expérimental et, tout particulièrement, sur le domaine moral. J'insisterai sur ces réticences spécifiques au domaine moral en entreprenant la présentation de la philosophie morale expérimentale, car elles justifient, pleinement à mon sens et pour quatre raisons que je détaillerai, de distinguer comme je le fais dans le présent travail, l'analyse des apports expérimentaux en regard de la philosophie morale.

Le cas paradigmatique que je propose ensuite pour présenter génériquement la philosophie morale expérimentale est celui dit de « l'effet Knobe », un des fleurons du mouvement XPhi. Son point de départ en est une expérience montrant un phénomène particulier, une asymétrie expérimentale étonnante, qui n'avait pas été anticipée par les philosophes moraux

et qui les contraint à une révision de leurs théories morales. Il a donné lieu à une littérature abondante que j'utiliserai dans un chapitre ultérieur pour analyser certains aspects de la philosophie morale expérimentale, dont la difficulté à expérimenter en utilisant des concepts issus de la philosophie morale, concepts aux définitions instables en regard des situations particulières mises en œuvre dans les expériences. Deux autres exemples, l'attribution des états mentaux à un tiers et la compatibilité du déterminisme avec le libre arbitre, ont pour objet de compléter cette présentation du mouvement XPhi de façon à montrer ses apports et ses limites, dans la perspective d'une interprétation morale de leurs résultats.

Je soulignerai, pour clore ce chapitre, le bilan en demi teinte du mouvement XPhi, bilan que tirent aujourd'hui tant ses promoteurs que ses détracteurs. D'un côté, l'apport de la psychologie et des neurosciences à de nombreuses questions philosophiques est avéré, et le mouvement XPhi a contribué à faire progresser de nombreux points de philosophie. De l'autre, malgré de grandes espérances initiales, ces apports n'ont pas donné de réponses définitives à ces questions et, ajouteront les détracteurs, ni même des réponses très utiles.

2.1 Le mouvement XPhi, premiers pas

Cette section a pour objectif de présenter différents aspects du mouvement philosophique, dit XPhi, de la philosophie expérimentale. Il ne s'agit pas d'un apport personnel sur le fond, synthèse pour l'essentiel de la littérature disponible, mais d'un préalable à la présentation de mes propres travaux au chapitre suivant. La section comporte quatre temps. Dans le premier je retrace rapidement les premiers pas fondateurs du mouvement XPhi, dans un deuxième temps j'aborde la question des résultats de ce mouvement, à ce jour, sous la forme de plusieurs exemples issus de questions philosophiques qui ne relèvent pas du domaine moral. Je présente ensuite des familles d'objections générales soulevées par l'approche expérimentale en philosophie et enfin, dans un dernier temps, je présente les spécificités de l'approche expérimentale pour la philosophie morale.

La philosophie expérimentale que je mobiliserai ici est un mouvement, dit XPhi par les anglophones, dont l'origine est habituellement située à la fin des années 1990 et qui a connu un important développement dans les années 2000. Il est aisé de définir le mouvement XPhi par les objectifs des philosophes qui ont contribué à sa création ou s'en réclamaient. En effet cette tâche est facilitée par l'existence d'un manifeste émis par Joshua Knobe et Shaun Nichols en 2008 dans un recueil éponyme « Experimental Philosophy » (Knobe et Nichols (eds.) 2008, page 3) [140]. Elle est également facilitée par l'existence des outils de collaboration que

s'est donné ce groupe de philosophes créé à l'ère de l'Internet triomphant, dont un blog de référence et une entrée dans les grandes encyclopédies en ligne [1]. La caractérisation de la philosophie expérimentale peut également s'appuyer sur les nombreux ouvrages auxquels a donné lieu le développement de ce mouvement, dont la liste chronologique ci-dessous donne une première approche [2] :

— « Experimental Philosophy » (Knobe et Nichols (eds.) 2008) [140]

— « La philosophie expérimentale » (Cova 2012) [57]

— « Experimental Philosophy » (Alexander 2012) [1]

— « Advances in experimental moral philosophy » (Sarkissian et Wright 2014) [200]

— « Current controversies in experimental philosophy » (Machery et O'Neill 2014) [175]

— « Theory and practice of experimental philosophy » (Sytsma 2016) [224]

— « A companion to experimental philosophy » (Systma 2016) [223]

— « Philosophy within its proper bounds » (Machery 2017) [151]

— « Experimental philosophy A critical study » (Mukerji 2019) [167]

2.1.1 Aux sources des XPhi : déception et espérance

Le manifeste de ce mouvement XPhi permet d'affiner la définition, vraie mais insuffisante, qui consisterait simplement à proposer que la philosophie expérimentale vise à utiliser des arguments issus de démarches expérimentales dans l'objectif d'éclairer des débats philosophiques. Cette définition est insuffisante car elle semble pouvoir s'appliquer à toute philosophie. En effet, comme le rappelle Timothy Williamson (Williamson 2007, page 7) [235], tous les philosophes ont, au moins partiellement, fait appel aux résultats des sciences expérimentales de leur temps, et plus généralement, à l'observation et aux approches empiriques du monde, qu'elles soient scientifiques ou pré-scientifiques. Le manifeste précise donc cette définition sur trois points. Premier point, historique, la philosophie se serait depuis le milieu du 20^{e} siècle trop focalisée sur les seules méthodes de la philosophie analytique, négligeant de ce fait les questions philosophiques éternelles qui ne peuvent se réduire à des questions de langage ou de concepts. Le mouvement XPhi veut rompre avec cette focalisation et ouvrir le regard sur le monde lui-même et non uniquement sur les concepts que nous utilisons pour le décrire. Deuxième point, méthodologique, parmi ces questions éternelles, il en est

1. Le blog : `https://philosophycommons.typepad.com/xphi/` puis `https://xphiblog.com/`, la Stanford Encyclopedia of Philosophy : `https://plato.stanford.edu/entries/experimental-philosophy/`. Citons également le groupe de chercheurs en psychologie morale expérimentale qui recoupe largement ce que l'on peut également appeler « philosophie morale expérimentale » : « Moral Psychology Research Group » `http://www.moralpsychology.net/`

2. L'essentiel des publications concernant la philosophie expérimentale est en anglais. Pour une revue synthétique en français, voir le numéro spécial de la revue Klesis (Cova 2013).[55]

une particulièrement importante pour le philosophe qui est la compréhension du fonctionnement de son propre esprit. La philosophie analytique en s'appuyant pourtant sur l'intuition du philosophe comme outil principal a oublié de mettre réflexivement le « comment » de ces intuitions sur sa table de travail et c'est ce que se proposent de faire les philosophes expérimentaux. Enfin, troisième point, opportuniste, si les philosophes expérimentaux reviennent vers les questions philosophiques éternelles, c'est aujourd'hui armés des nouveaux outils des sciences expérimentales et, en particulier des sciences cognitives, qui permettent d'espérer des avancées significatives.

On peut donc relever à la source du mouvement XPhi une déception et une espérance. La déception est celle du monde académique philosophique confronté à la faiblesse des résultats de la démarche analytique. Malgré des décennies de travail minutieux et acharné, les concepts philosophiques significatifs résistent à être définis analytiquement. Chaque définition proposée finit par être mise à mal par un contre-exemple qui, selon l'intuition des philosophes, correspond bien aux critères de la définition mais ne relève pas du concept, ou inversement. De définitions de plus en plus complexes en contre-exemples de plus en plus élaborés, le philosophe analytique s'épuise sans fin, et on peut douter de la substantialité de la contribution qu'il apporte ainsi à l'étude des phénomènes caractérisés par ce concept. Ce constat motive la réforme des méthodes de travail que les philosophes expérimentaux souhaitent promouvoir.

Mais la déception seule n'est pas un moteur, il faut aussi une espérance pour se lancer dans une nouvelle aventure. Cette espérance est apportée au mouvement XPhi par le développement des sciences cognitives appuyées sur les nouvelles techniques expérimentales utilisées par les psychologues et les neuroscientifiques. Elles ouvrent à l'orée du siècle de nouvelles perspectives vers une possible compréhension du fonctionnement du cerveau humain. S'il est envisageable de mieux comprendre comment et sous quelles conditions les humains ont tel ou tel comportement, telle ou telle pensée, et finalement telle ou telle intuition, alors on pourra utiliser cette connaissance pour mieux analyser les intuitions des philosophes eux-mêmes et ouvrir de nouvelles pistes de réflexions philosophiques.

On peut concrétiser cette espérance avec l'exemple d'une technique : l'IRMf (Imagerie à Résonance Magnétique fonctionnelle). Cette technique permet de mesurer les flux sanguins dans le cerveau et, ainsi, les zones qui se sont activées pendant un processus mental. L'IRMf a été mise au point au Massachusetts General Hospital de Boston au début des années 1990 lorsque des ordinateurs de plus en plus puissants ont été couplés aux appareils d'IRM [3]. Dès

3. Pour une histoire de l'IRMf voir (Houdé 2010 p 22) [120]. Les scanners d'IRMf peuvent produire quatre images

2001, cet outil a été utilisé dans le cadre d'une recherche menée par un philosophe, Joshua Greene [108]. Greene travaillait sur le problème du tramway (voir 4.2, page 161), exemple classique de dilemme moral où une personne pour sauver plusieurs vies doit accepter de sacrifier une victime. Greene cherchait à comprendre pourquoi les humains jugent différemment cette personne selon que la victime sacrifiée le soit directement et volontairement, ou indirectement et involontairement. Joshua Greene dans son ouvrage « Moral Tribes » (Greene 2015 p118) [109] a décrit l'excitation qu'il a ressentie quand il a réalisé que l'approche neuronale, telle que présentée par Antonio Damasio dans « L'erreur de Descartes » [61], allait peut-être apporter un élément de réponse à sa question : les processus mentaux ne sont peut-être pas les mêmes dans les deux cas de jugement moral. Dans cet ouvrage, Damasio décrit ce célèbre patient, Phineas Gage, qui garde des capacités cognitives mais ne peut plus ressentir d'émotion à la suite d'un accident grave ayant détruit une partie de son cerveau. Citons comment Joshua Greene décrit sa réaction :

> Quand j'ai lu ce passage, j'étais seul dans une chambre d'hôtel. J'étais si excité que je me suis mis debout et que j'ai commencé à sauter sur ce lit d'hôtel. Ce que je venais de comprendre, c'était le lien avec le problème du tramway.

Pour Greene, le lien était fait : les jugement moraux différents font appel à des modes de raisonnement différents qui eux-mêmes sont corrélés avec l'activation de différentes zones du cerveau. Et on peut le confirmer empiriquement grâce aux neurosciences. C'est l'IRMf qui allait permettre quelques mois plus tard, en collaboration avec Jonathan Cohen de Princeton, de montrer que ce raisonnement pressenti avec le cas d'un patient lésé était bien également observable chez des sujets sains. On comprend à la lumière de cet exemple la puissance de l'espérance qu'a soulevé le développement des sciences cognitives pour tous les philosophes analytiques confrontés à la stagnation, réelle ou supposée mais en tous cas ressentie, de leur discipline.

Et les perspectives ouvertes par les sciences cognitives avec le déploiement quantitatif et qualitatif des outils d'analyse du comportement humain en ce début de 21^{e} siècle ne se limitent pas à la seule IRMf. Quantitativement, la généralisation d'Internet et la facilité d'utilisation des outils de sondage comme Amazon Mechanical Turk (AMT) ont permis l'explosion du nombre d'études empiriques faisant appel à des sondages en ligne sur tous les sujets. Il est aujourd'hui facile et peu coûteux de proposer un questionnaire à des milliers de personnes dans le monde entier puis de recueillir et d'exploiter leurs réponses avec des logiciels puissants de façon quasi instantanée. Ainsi, par exemple, à une question sur le relativisme

du cerveau par seconde, ce qui permet de suivre le déplacement de l'activité neuronale au cours d'une tâche complexe.

culturel on peut apporter rapidement une réponse d'ampleur mondiale à un coût raisonnable (à titre d'illustration on peut penser au budget de frais de déplacement qu'auraient nécessité les études de Richard Nisbett sur le relativisme culturel impliquant des équipes sur les 5 continents (Nisbett 2004) [172] avant Internet pour mesurer l'importance de ce changement). Ce changement quantitatif résulte également de la numérisation de nombreux aspects du comportement humain qui offre un immense stock de données potentiellement exploitables. [4]

Qualitativement, l'imagerie à résonance magnétique fonctionnelle (IRMf), et plus largement tous les instruments des neurosciences, ont donné à voir les processus neuronaux corrélés à des comportements observables. Le siècle précédent avait vu l'essor de l'utilisation des temps de réaction comme indicateurs individualisés du niveau de difficulté d'une tâche cognitive [5], mais ce qui se déroulait à l'intérieur de la boîte crânienne restait largement inexploré. Ce sont ces processus neuronaux que l'IRMf donne à voir avec la possibilité de les corréler à l'activité menée par le patient. Naturellement ces instruments ne sont pas parfaits et manquent aujourd'hui de précision et de définition. Le Graal que recherchent les neuroscientifiques est constitué d'outils d'imagerie cérébrale qui permettraient de disposer d'une méthode non-invasive, facile de mise en œuvre, et offrant une résolution spatiale à l'échelle des neurones et des synapses ainsi qu'une résolution temporelle compatible avec la dynamique de la pensée. Bien que ce Graal ne soit pas atteint, les progrès réalisés à ce jour permettent néanmoins des avancées importantes de la compréhension du fonctionnement normal et pathologique du cerveau des animaux humains et non humains.

Les philosophes expérimentaux se sont donnés pour cadre de travail d'exploiter l'ensemble de ces nouvelles possibilités tant qualitatives que quantitatives pour tenter d'apporter de nouveaux éclairages sur de nombreuses questions philosophiques, dont la philosophie morale, en reprenant à leur compte les méthodes de travail des psychologues expérimentaux. Dans la deuxième partie de leur ouvrage « Théorie et pratique de la philosophie expérimentale », Sytsma et Livengood (Sytsma 2016) [224] offrent une introduction à ces méthodes de travail destinées aux philosophes souhaitant s'exercer à l'approche expérimentale. Cette introduction est constituée de présentations sommaires de la méthode inductive, des calculs statistiques, du logiciel R [6], et de la démarche expérimentale en général. Bien que le chapitre s'intitule « Comment conduire une recherche empirique en philosophie » (Sytsma 2016 page

4. Le CNRS a publié une vaste perspective de l'utilisation de ces données dans les sciences dans (Bouzeghoub 2017) [29].

5. Pour une histoire de cette avancée majeure, voir (Jensen 2006) [125]

6. Ce logiciel libre et gratuit permet de disposer de toutes les fonctions statistiques avancées qui permettent d'exploiter les résultats des expérimentations. Il est devenu à ce jour l'outil le plus répandu dans le domaine de la psychologie expérimentale. Il est en particulier une des briques qui facilite la mise en commun des résultats d'expériences dans le cadre des initiatives de « open science » qui supposent que soient publiés les algorithmes et codes statistiques utilisés pour l'analyse des données. Voir `https://www.r-project.org/`

123), il ne comporte aucune indication de ce qui pourrait être spécifique à la philosophie en la matière. Tout le développement du chapitre peut s'appliquer à la psychologie expérimentale dont il est d'ailleurs extrêmement proche [7]. On peut donc légitimement s'interroger sur la raison de ce titre de chapitre comportant le terme « philosophie » alors que son contenu ne fait aucune mention de caractéristiques qui seraient spécifiques à l'expérimentation en tant qu'elle porte sur des sujets philosophiques. Deux interprétations peuvent venir à l'esprit. D'abord, sous l'angle pédagogique, il est bon d'impliquer le lecteur en le citant dans le titre pour capter son attention, cela marque que la présentation sera d'un abord facile malgré le contenu technique des sujets abordés. On peut y voir également un effet de cadrage dû à l'objet même de l'ouvrage qui vise simultanément deux objectifs. Le premier est de convaincre les philosophes de l'intérêt de l'approche expérimentale en s'appuyant sur des exemples de résultats obtenus, c'est la première partie. Le second est de les convaincre également qu'ils ont, en tant que philosophes, la possibilité de mener eux-mêmes ces expérimentations en suivant les méthodes présentées, et c'est la seconde partie de l'ouvrage. Je reviendrai en conclusion sur ce point qui me semble tout à fait discutable à l'heure où toute connaissance scientifique est le résultat de larges coopérations de spécialistes de multiples domaines.

Il serait accusable de naïveté de ne pas citer ici, au moins pour mémoire, la dimension budgétaire de cet enthousiasme. Le nombre de postes et les budgets dans les universités de philosophie n'ont évidemment pas le dynamisme de ceux des départements de neurosciences et des sciences cognitives. Rapprocher son domaine de recherche de la manne dont ont bénéficié ces départements plus en vue n'est pas sans intérêt sous cet angle très concret des ressources disponibles. La philosophie expérimentale a également été l'occasion de nouvelles publications qui ont permis l'éclosion d'une génération de philosophes plus proches des sciences que leurs aînés qui ont, eux, l'avantage de la notoriété pour accéder aux publications plus traditionnelles. On peut incidemment remarquer que c'est par un procédé analogue de lancement de laboratoires permettant l'expérimentation que, dans la deuxième moitié du 19^{e} siècle, la psychologie expérimentale s'est éloignée puis séparée de la philosophie (William Wundt 1879 en Allemagne, William James en 1891 aux USA, Benjamin Boudon 1896 en France) [8].

2.1.2 Plusieurs distinctions utiles

De ces premiers pas portés par la déception en regard de la philosophie analytique et par l'espérance en regard des sciences cognitives, le mouvement XPhi a gardé plusieurs distinc-

7. Voir par exemple (Ghiglione 2007) [101]
8. Voir l'article sur Benjamin Boudon écrit à l'occasion du centenaire de la création du laboratoire de psychologie expérimentale en France (Nicolas 1996) [171]

tions utiles à qui veut comprendre ce qu'il a produit ainsi que les critiques qu'il a soulevées. Une première distinction est ainsi posée entre une conception étroite et une conception large de l'objet de la philosophie expérimentale (Alexander 2012) [1], distinction qui est une conséquence de la mise en question des méthodes de travail des philosophes analytiques. Selon la vision étroite, le mouvement XPhi se concentre sur le problème de l'intuition du philosophe à sa table de travail. Cette intuition, « Moi, philosophe, je pense que P » serait à mettre en question en fonction de ce que chacun pense, exprimé par exemple sous la forme statistique « X % de la population pense que P », ou sous la forme issue de la démarche psychologique, « le processus psychologique Y conduit à penser que P ». Le philosophe expérimental se donnerait pour but d'établir empiriquement ce nouveau type de résultats qui ne dépend plus de la seule introspection du philosophe, puis de travailler sur les différences éventuellement constatées entre les résultats rendus disponibles par ces différentes approches.

Selon la vision large, le mouvement XPhi ne se limite pas au seul problème de l'intuition et étend son approche expérimentale à toute question philosophique sur le monde. Remarquons que, pour être utile, cette approche suppose deux préalables méthodologiques importants. Le premier est qu'il soit possible et intéressant de définir ce que peut être une question que l'on dit philosophique. Il s'agit donc de disposer de critères permettant de distinguer une question philosophique d'une question qui ne l'est pas et d'anticiper que ces critères sont de nature à justifier une approche expérimentale spécifiquement adaptée. En effet s'il n'est pas possible de distinguer (sous l'angle de vue de l'expérimentateur) une classe particulière de questions dites philosophiques, et en quoi elles relèvent de démarches expérimentales particulières, alors la démarche large XPhi se dissout dans l'ensemble des approches plus ou moins empiriques de toutes les connaissances. La spécificité des approches empiriques des questions philosophiques est un préalable à démontrer si on veut affirmer la possibilité d'un mouvement XPhi utile.

Le deuxième préalable, si on souhaite adopter la conception du mouvement XPhi élargie à toute question philosophique, est de préciser ce qui le différencie de la philosophie empiriquement informée (Cova 2012 p 6) [57], c'est-à-dire d'une philosophie qui ne tourne pas le dos aux résultats des sciences expérimentales, dont ceux de la psychologie, mais au contraire les connaît et s'attache à les incorporer à sa réflexion. La spécificité du mouvement XPhi serait ici d'ordre organisationnel. Les philosophes expérimentaux ne se contentent pas de recevoir des résultats venant des autres sciences ou domaines de la connaissance mais conçoivent, réalisent et interprètent leurs propres expériences. Ils quittent le fauteuil pour entrer dans le laboratoire et incitent tous les philosophes à en faire autant. Certains (Mukerji 2019 p11)

[167] (Bickle 2018) [25] ont alors cru devoir questionner leur appartenance : sont-ils encore philosophes ou sont-ils devenus, par exemple, psychologues expérimentaux ou bien encore ont-ils créé une niche spécifique au sein de la philosophie ? L'enjeu de cette question n'est pas épistémique, l'intégration des philosophes expérimentaux dans telle ou telle université ne changera ni leurs outils et méthodes, qui pour l'essentiel sont bien les outils expérimentaux provenant des sciences sociales et humaines, ni les difficultés de l'utilisation de ces outils pour traiter des problèmes conceptuellement complexes. Que ces questions complexes soient vues comme philosophiques ou psychologiques ne change pas grand-chose à l'expérimentation à mener. Néanmoins, la question du rattachement académique est importante dans plusieurs directions. Dans un sens, il semble qu'on puisse attendre du fait que des expériences soient menées par des philosophes une plus grande adéquation des expériences aux particularités des questions philosophiques, à supposer que ces particularités existent (premier préalable ci-dessus). On aurait alors un gain à faire travailler ensemble des philosophes expérimentaux s'intéressant à différents domaines de la philosophie pour qu'ils mettent en commun leurs réflexions sur ces nouvelles méthodes. On peut trouver un exemple d'application de cette approche dans la série « Advances in Experimental Philosophy » dont les 11 numéros recouvrent autant de domaines de la philosophie [9]. Par ailleurs, on peut penser que les questions philosophiques recevront un meilleur traitement et une plus haute priorité dans les programmes de recherche dédiés à la XPhi que dispersées dans différentes entités de la recherche en psychologie.

Dans un autre sens, moins favorable, on peut douter que les structures académiques philosophiques soient bien placées pour gérer des laboratoires et des expérimentations. Elles n'ont ni les budgets, ni les locaux ni, en un mot, le savoir-faire correspondant. De même, on peut douter que les formations philosophiques actuellement délivrées préparent à la réalisation de programmes expérimentaux. Former des philosophes expérimentaux susceptibles de maîtriser toute la chaîne expérimentale semble extrêmement difficile dans le cadre actuel de la spécialisation toujours plus fine vers des techniques toujours plus complexes que connaissent la psychologie et les sciences cognitives d'un côté et la philosophie analytique de l'autre. Il peut alors apparaître préférable que le philosophe se consacre à un rôle amont de pourvoyeur de questions, de recherches problématisées, à la bonne compréhension et à la critique constructive des méthodes et outils disponibles et à un rôle aval d'interprétation philosophique des résultats des expériences en laissant aux psychologues expérimentateurs

9. Site de l'éditeur consulté en octobre 2019 : https://www.bloomsbury.com/us/series/advances-in-experimental-philosophy/?pg=1

chevronnés le soin de mener les expériences. [10] Il s'agit en somme de viser à construire des relations entre philosophie et sciences cognitives analogues à ce qui peut aujourd'hui exister entre philosophie et biologie (ou toute autre science plus anciennement structurée que la psychologie). [11]

Une autre présentation des distinctions à opérer au sein des XPhi a été proposée en 2019 par Mukerji dans (Mukerji 2019) [167]. Il distingue trois programmes relevant de la philosophie expérimentale : un programme cognitif, un programme négatif, et un programme positif. Le programme cognitif a pour objet de décrire comment les humains répondent à une question philosophique. En s'appuyant sur les sciences cognitives, il s'agit de décrire les processus psychologiques activés dans le traitement d'une telle question. Les tenants de la XPhi cognitive considèrent que cette question est elle-même philosophique et, d'une certaine façon, méta-philosophique, puisqu'elle amène à observer comment nous philosophons.

Le programme négatif de la philosophie expérimentale a un objectif défini : montrer que les méthodes habituelles des philosophes analytiques ne sont pas fiables. Il vise en particulier la méthode des cas qui consiste à tester les théories philosophiques en les confrontant à des cas réels ou inventés supposés exercer les jugements philosophiques. Cette méthode ne serait pas fiable et il faut en conséquence soit l'abandonner soit la réformer. Edouard Machery dans son ouvrage « Philosophy within its proper bounds » [151] s'inscrit par exemple pleinement dans ce programme.

Enfin, le programme positif de la philosophie expérimentale regroupe toutes les approches qui ne visent pas à critiquer les méthodes philosophiques habituelles, mais visent à les compléter par l'apport d'expériences menées par des philosophes sur des cas conçus par ces mêmes philosophes pour éclairer tout type de problème philosophique. On retrouve là la vision large de la XPhi présentée plus haut qui constitue l'essentiel des publications du mouvement XPhi aujourd'hui.

Cette vision large du mouvement XPhi est celle que défendent les promoteurs de ce mouvement au travers de leurs outils de communication (voir par exemple le blog `https://xphiblog.com/`). Il est par ailleurs utile de constater que les principales réticences qu'a soulevé ce mouvement sont plutôt à l'encontre de la vision étroite du mouvement XPhi, celle qui vise à mettre en cause les outils habituels des philosophes analytiques, plutôt qu'à l'en-

10. Cette position est par exemple celle de Timothy Williamson dans « The Philosophy of Philosophy » (Williamson 2007) [235]

11. Bickle (Bickle 2018) [25] s'appuie sur l'exemple historique de la « neurophilosophie » qui, partant avec l'ambition de remplacer la philosophie traditionnelle, serait aujourd'hui ramenée à une « philosophie des neurosciences » pour annoncer qu'il en ira de même pour la philosophie expérimentale. Je tenterai de montrer plus loin que si je pense qu'il pourrait bien en être ainsi pour la philosophie en général, la philosophie expérimentale étant ramenée à une philosophie de la psychologie (et en particulier de la psychologie des philosophes), le cas particulier de la philosophie morale doit faire l'objet d'un examen beaucoup plus serré du fait du problème de la normativité.

contre de cette vision large. Je vais maintenant tour à tour adopter les deux points de vue des philosophes tenants de, et réticents à, la démarche des philosophes expérimentaux. Dans un premier temps je décrirai rapidement des exemples d'études menées dans différents domaines de philosophie générale puis les objections d'ensemble faites à l'approche XPhi, et ensuite je reprendrai les éléments plus précisément centrés sur la philosophie morale.

2.2 Les résultats des XPhi, deux exemples introductifs

Le mouvement XPhi a couvert de nombreux domaines de la philosophie et continue à produire de nombreux articles sur de multiples axes de recherche. Si on retient comme prisme les onze volumes de la collection « Advances in Experimental Philosophy » de l'éditeur Bloomsbury, c'est la quasi totalité des branches de la philosophie académique occidentale qui est concernée : philosophie des sciences, philosophie de la logique et des mathématiques, méthodologie de la philosophie expérimentale, métaphysique, esthétique, psychologie morale, philosophie des sciences cognitives, philosophie des religions, philosophie du langage, philosophie de l'esprit, philosophie de la connaissance... [12]. Reprendre tous ces axes de recherche ayant donné lieu à des articles relevant ainsi du mouvement XPhi serait ici hors de propos. Et ce non seulement par l'ampleur des domaines abordés, mais surtout parce qu'il n'est pas certain que puissent être utilement distingués, pour mon propos, les travaux ainsi labellisés comme XPhi de tous les travaux scientifiques, et en particulier psychologiques, dont la proximité aux questions philosophiques est forte. Cette proximité peut être liée à la problématique abordée, une question philosophique étant à l'origine de la recherche expérimentale, ou de signification, l'interprétation philosophique du résultat donnant son importance au résultat expérimental. Dans les deux cas, ces travaux expérimentaux rejoindraient potentiellement le corpus qui m'intéresse ici pour étudier l'apport de la démarche expérimentale à la philosophie morale, en complément des travaux des XPhi dont la spécificité est d'être menés par des philosophes.

Pour présenter les résultats récents obtenus par le mouvement XPhi, j'ai donc retenu à titre d'introduction deux ensembles de travaux qui, bien que non classiquement associés à des questions de philosophie morale, illustrent, par leur sujet central à toute expérimentation sur le raisonnement, le cœur des problèmes de base auxquels se confrontent XPhi et philosophes dans leur collaboration. Il s'agit des travaux sur la connaissance et sur la référence. [13]

12. Voir `https://www.bloomsbury.com/us/series/advances-in-experimental-philosophy/?pg=1`

13. Ces descriptions et cette sélection arbitraire, s'appuient pour partie sur le travail d'analyse critique de Mukerji [167] ainsi que sur les ouvrages de référence du mouvement XPhi mentionnés plus haut dont (Alexander 2012) [1] et (Knobe 2008) [140].

2.2.1 La connaissance

La connaissance est l'objet de recherche philosophique par excellence, il est donc peu surprenant que chaque mouvement philosophique tente d'apporter sa contribution à sa définition. Traditionnellement, depuis Platon, la connaissance est souvent définie comme une « croyance vraie et justifiée » (en anglais, JTB pour Justified True Belief). En 1963, le philosophe Edmund Gettier s'est rendu célèbre en remettant en cause cette définition à l'aide de contre-exemples où une telle croyance vraie et justifiée n'était pas, selon son intuition, à considérer comme une connaissance (Gettier 1963) [100] [14]. Le principe d'une des familles de ces contre-exemples est le suivant :

— Soit P une proposition.
— Une personne X sait que P au temps t1
— Sans que X le sache, il arrive un premier événement qui fait que NonP devient vrai, puis un second événement qui fait que P est à nouveau vrai au temps t2. Ces deux événements sont contingents et fortuits.
— Lorsque à t2, après le second événement, X dit que P, les trois conditions de la définition JTB de la connaissance sont réunies : X en a la croyance, P est vrai et X est justifié à croire que P car, d'une part, il en était déjà certain à t1 et, d'autre part, il ne connaît pas l'existence des 2 événements qui ont eu lieu entre t1 et t2.
— Gettier dit que X n'a pas connaissance de P car P est vrai à t2 « par hasard » et X n'a pas d'accès rationnel à cette vérité. La justification qu'il a de P est inconsistante en regard du déroulement des événements.

Sur cette base, les philosophes de la connaissance ont développé une immense littérature pour, soit améliorer la définition de la connaissance, en introduisant par exemple des critères d'une bonne justification qui exclurait le cas de la vérité fortuite, soit nier qu'une connaissance puisse être considérée comme un type particulier de croyance, soit proposer des interprétations des cas de Gettier moins fatales à la définition JTB, soit mettre en doute l'intuition elle-même que X n'a pas connaissance de P à t2.

Les philosophes expérimentaux se sont emparés de cette question. Ils ont conservé l'ossature des contre-exemples de Gettier et les ont soumis à divers échantillons de personnes en faisant varier les formulations des circonstances du contre-exemple. L'article de 2001 de Weinberg et Stitch [234] présente des résultats de sondages qui montrent que les réponses à la question « X a-t-il connaissance de P? » dépendent de la culture des personnes interro-

14. Des arguments de même nature que ceux de Gettier existaient certainement au préalable (Mukerji 2019 p 62) mais il revient à Gettier d'avoir laissé son nom à ce domaine de recherche et d'avoir déclenché la vague de travaux subséquente en philosophie de la connaissance.

gées (asiatiques vs. occidentaux) et de leur position socio-économique (mesurée par différents indices dont le niveau d'étude). Ces résultats jettent un doute sur la validité de l'intuition du philosophe qui, seul dans son fauteuil, déclare que dans le cas des contre-exemples de Gettier il n'y pas réellement de connaissance. Plus largement, le résultat est utilisé par les auteurs de l'étude pour mettre en doute la méthode analytique appuyée sur la seule intuition du philosophe. Quand le philosophe vise à établir une définition de la connaissance (ou de tout autre concept) commune à toutes les cultures et à toutes les classes sociales, il semble difficile d'accepter qu'il adopte pour cela une démarche de travail qui, elle-même, donne des résultats différents en fonction de ces facteurs.

De nombreuses études ont ensuite exploré les possibilités ouvertes par Weinberg et Stitch. Par exemple Swain, Alexander et Weinberg explorent dans (Swain 2008) [222] les effets liés à l'ordre des questions et l'effet dû au cadrage préalable des participants par un pré-questionnaire. Les auteurs affirment que les réponses sont significativement différentes en fonction de ces facteurs qui sont, pourtant, sans lien avec l'évaluation de la connaissance qu'a ou non X de P. Ces conclusions viennent ainsi encore augmenter le scepticisme quant à la valeur épistémique de la méthode analytique pour définir un concept complexe comme celui de « connaissance ».

Ces résultats ont soulevé à leur tour de nombreuses critiques (voir par exemple (Cullen 2010) [59]) : effectifs sondés insuffisants, formulations ambiguës des questions et des réponses proposées, manque de cohérence des résultats entre les différents cas utilisés, biais divers. En somme, les conséquences tirées de ces études auraient une ampleur immense, rien moins que mettre en doute tous les résultats de la philosophique analytique, sans commune mesure avec la modestie et l'étroitesse de l'approche utilisée : une simple corrélation établie sur une base restreinte de participants et quelques cas opportunément choisis.

Les développements plus récents, poursuivant la démarche initiée par Gettier, sont venus renforcer chacune des deux vues contradictoires sur la fiabilité des intuitions philosophiques et, plus largement, de la démarche analytique. L'ouvrage de Machery « Philosophy within its proper bounds » (Machery 2017) [151] consacre les pages 53 à 56 à l'exemple de la définition de la connaissance pour contribuer à instruire le manque de fiabilité de l'intuition philosophique en s'appuyant sur les expériences qui montrent comment cette intuition varie en fonction de nombreux paramètres.

Mais d'autre part, Boyd et Nagel dans l'article « The reliability of Epistemic Intuitions » (dans O'Neil 2014 p. 109) [175] défendent inversement l'idée que l'intuition est suffisamment fiable suffisamment souvent pour être utile au philosophe comme à l'homme de la rue. Les

différences apparentes sur lesquelles repose l'argumentation sceptique peuvent souvent être justifiées par des processus psychologiques et elles ne doivent pas conduire à sous-estimer le très large accord entre tous les humains sur l'essentiel, comme par exemple ce qu'est la connaissance (Apperly, cité par Boyd et Nagel p.117). Accord qui est d'ailleurs peu surprenant si on considère la grande proximité de constitution et de comportement au sein de l'espèce humaine ainsi que les très grandes similitudes entre toutes les cultures, sans bien sûr nier l'existence des variations locales.

En résumé, l'exemple de la définition analytique de la connaissance me conduit à plusieurs observations. Première observation, l'ampleur des travaux générés par l'article de Gettier puis par l'approche expérimentale de Weinberg montre l'importance qu'a pris l'approche expérimentale dans la vie intellectuelle philosophique. Deuxième observation, le débat n'a nullement été tranché par les expériences menées ni sur le fond, ce qu'est la connaissance, ni sur la méthode, écarter ou conserver la méthode analytique des cas. Si l'approche expérimentale a été féconde, elle n'a pas été déterminante. Troisième observation, Gettier a amorcé en 1963 un mouvement d'améliorations itératives, chaque définition proposée appelant de nouveaux contre-exemples qui, eux mêmes, appellent une nouvelle définition et ainsi de suite. Ce qu'introduit Weinberg en 2001 vient compléter ces itérations en rajoutant une nouvelle étape : les résultats des sondages qui constituent les traces qui devront être prises en compte par une nouvelle proposition de définition. Or, ce n'est pas ce qui se passe ici, l'essentiel de l'énergie des chercheurs n'est pas mobilisée vers la bonne utilisation des traces disponibles mais au contraire vers la mise en doute de leur pertinence.

2.2.2 La référence

Autre question philosophique au long cours : comment nos pensées, nos concepts, et finalement nos mots, se raccordent-ils au monde ? Comment, par exemple, un auditeur comprend-il un locuteur qui emploie un nom propre comme « Gödel » ? Un type de réponses à cette question est dite descriptiviste (Bertrand Russell 1905 "On Denoting", Mind 14, pp. 479–493) quand la réponse met en avant la description des caractéristiques de Gödel pour identifier de qui le locuteur parle : Gödel est le logicien qui a démontré les théorèmes d'incomplétude de l'arithmétique.

En 1966, Connellan a introduit une distinction entre la description et la référence qui lui semble à la source d'une possible confusion dans la proposition de Bertrand Russell (Connellan 1966) [53]. Il utilise pour cela l'exemple d'un dénommé Jones qui a assassiné Smith.

Quand un locuteur dit « L'assassin de Smith est un fou », deux interprétations de cette phrase sont possibles. Le locuteur peut vouloir dire que Smith était un garçon adorable et qu'il faut être fou pour vouloir l'assassiner, et le locuteur est supposé adopter la dimension descriptive du groupe nominal « l'assassin de Smith ». Soit, deuxième interprétation, le locuteur veut dire que c'est l'individu nommé Jones qui est fou, indépendamment du fait qu'il a tué Smith, et le locuteur est supposé adopter une dimension référentielle de ce groupe nominal qui échappe à la définition descriptiviste de Russell.

Saul Kripke a élaboré cette distinction en 1972 [142] en s'appuyant sur un exemple de cas devenu l'emblème du problème de la référence. Gödel avait un ami X qui, excellent logicien, a démontré les théorèmes d'incomplétude juste avant sa mort. Gödel s'est alors emparé des manuscrits et les a publiés sous son nom. Kripke pose alors la question suivante : qui désigne-t-on en employant le nom propre « Gödel », l'individu Gödel, que chacun croit être l'auteur des théorèmes, ou X qui en est le véritable auteur ? Si on accepte la théorie descriptiviste de Russell, il semble que ce soit X, la personne qui a vraiment démontré les théorèmes, mais notre intuition va plutôt dans le sens de désigner ainsi par « Gödel » l'individu nommé Gödel, la personne connue et habituellement désignée par ce nom, et indépendamment du fait qu'il n'a pas, dans le cas proposé par Kripke, démontré les théorèmes. Kripke propose alors une nouvelle théorie de la référence appuyée sur un acte de baptême initial qui associe rigidement un nom propre à un objet du monde puis sur la transmission de proche en proche de cette référence au sein d'une communauté linguistique. La description peut venir en appui mais n'est pas constitutive de ce lien entre le mot « Gödel » et la personne à laquelle il réfère.

Les philosophes expérimentaux se sont emparés de ce débat en demandant aux participants de se prononcer sur qui de X ou de Gödel était désigné par le nom propre « Gödel » compris comme « l'auteur des théorèmes d'incomplétude ». L'argument de Kripke repose principalement, comme celui de Gettier précédemment, sur une intuition. Pour Gettier c'était l'intuition qu'il y a ou non connaissance, pour Kripke c'est l'intuition que le nom « Gödel » ne peut désigner X. Machery a en 2004 [150] comparé cette intuition de Kripke au résultat d'un sondage réalisé sur des étudiants de cultures différentes (asiatiques et occidentaux). Les réponses entre les deux groupes sont significativement différentes, les étudiants occidentaux penchant vers la proposition de Kripke alors que les étudiants asiatiques penchent vers la proposition de Russell.

Les résultats de ce sondage ont été confortés, après 2004, lors d'expérimentations menées en 2014 et 2016 qui sont reprises synthétiquement dans l'ouvrage d'Édouard Machery (Machery 2017 p 48 - 52) [151]. Pour Machery, ces résultats renforcent sa démonstration de

l'inadéquation de la méthode des cas pour traiter de sujets de philosophie du langage tels ceux de la référence. La méthode des cas consiste, rappelons-le, à partir d'un énoncé général que l'on cherche à conforter (ou à mettre en cause), puis à rechercher des cas particuliers, des historiettes comme celles de Gödel ci-dessus, qui, intuitivement, correspondent (ou ne correspondent pas) à l'énoncé général. Le raisonnement de Machery est alors le suivant : la méthode des cas repose essentiellement sur ce jugement intuitif, or ce jugement n'est pas fiable, il dépend de caractéristiques sans lien avec l'énoncé général, donc la méthode des cas ne permet pas de tirer une quelconque conclusion.

L'examen du problème de la référence conforte nos deux premières observations faites sur celui de la définition de la connaissance. D'une part, de nombreux travaux sont venus enrichir notre compréhension du problème de la référence après Kripke et Machery et la démarche expérimentale a contribué à renouveler un débat classique de philosophie du langage. Mais, d'autre part, si l'intuition du philosophe n'est pas un argument définitif dans le débat sur la référence, l'expérimentation n'apporte pas non plus un tel argument définitif. Il semble que l'une et l'autre approche se heurtent à la même difficulté : la trop grande complexité du problème posé et les multiples ramifications qui parcellisent à l'infini son analyse. Le programme de recherche consistant à expérimenter sur des cas décrits dans des vignettes et soumis à questionnaire fait apparaître dans une lumière crue l'impossibilité d'expliquer comment les participants comprennent le nom propre « Gödel » dans chacune des variantes imaginées par les chercheurs. Le cadre trop simple des théories de la référence disponibles n'y suffit pas, et l'apport de la philosophie expérimentale est limité à une contribution à ce constat négatif.

2.2.3 De nombreux programmes de recherche lancés

Il sera impossible ici de lister l'ensemble des programmes de recherche que les philosophes expérimentaux ont exploré. Pour situer quantitativement cette production, une façon simple d'opérer est de comparer le nombre d'articles produits en 2019 et indexés par Google Scholar en « experimental philosophy », soit 17500 articles, au nombre d'articles indexés sur la même période en « moral philosophy » soit 18700 articles [15]. Les productions sont du même ordre de grandeur pour ce nouveau domaine de la philosophie expérimentale et pour le domaine traditionnel de la philosophie morale, ce qui est un indicateur, certes très grossier, de l'ampleur prise par cette nouvelle façon de faire de la philosophie en regard de la philosophie

15. Chiffres relevés sur Google Scholar en février 2020, pour mémoire, sur la même période ce sont 120000 articles qui sont indexés avec le mot clé « philosophy ».

traditionnelle.

La liste des ouvrages que l'éditeur Bloomsbury cité plus haut consacre à la philosophie expérimentale comporte, plus qualitativement, la plupart des branches de la philosophie : philosophie des sciences, philosophie de la logique et des mathématiques, méthodologie de la philosophie expérimentale, métaphysique, esthétique, psychologie morale, philosophie des sciences cognitives, philosophie des religions, philosophie du langage, philosophie de l'esprit, philosophie de la connaissance. Il est également possible de retrouver sur le Blog du mouvement XPhi (`https://xphiblog.com/`) un recensement des multiples projets et travaux en cours. De façon générique, on peut aujourd'hui affirmer que, sauf exception toujours possible, toutes les questions philosophiques ayant une composante psychologique, c'est-à-dire ayant pour partie trait au comportement humain, ont été abordées par les philosophes expérimentaux. De façon beaucoup plus économe, on peut également considérer que les XPhi n'ont, en somme, analysé qu'une seule question : comment les humains raisonnent face à une question ayant, traditionnellement, été qualifiée de philosophique. Et la réponse qu'ils proposent est que, dès les questions fondamentales sur la connaissance ou sur la référence puis dans tous les domaines, le comportement des humains est plus complexe que ce qui est habituellement décrit par les théories philosophiques en place.

2.3 Les réticences au mouvement XPhi

Dans la présente section, je vais à nouveau utiliser le mouvement XPhi comme un révélateur : celui des différentes dimensions de la réticence des philosophes face aux résultats des XPhi et, au moins dans une certaine mesure, plus largement des résultats expérimentaux. La première expression de ces réticences que j'explore est le déni de nouveauté. Les philosophes auraient, toujours et de tous temps, pris en compte les connaissances scientifiques de leur temps, voire contribué à les établir. Il n'est donc que de laisser les scientifiques travailler et les philosophes philosopher. Cette première réticence prend comme cœur de cible la prétention des philosophes du mouvement XPhi à mener eux-mêmes leurs expériences sans avoir recours aux scientifiques rodés à cette tâche. Dans un second temps, je reprendrai les réticences soulevées par le programme négatif des XPhi. Il est bien sûr facile de voir dans ces réticences un plaidoyer pro domo d'une corporation attaquée. Elles éclairent néanmoins sur l'ensemble des difficultés qu'affrontent les expérimentateurs face aux questions philosophiques et, singulièrement, face aux raisonnements des humains quand ils abordent des questions philosophiques. Un point central du débat est celui du manque de fiabilité de ces

études. Ce manque de fiabilité se traduit en particulier par la difficulté des expérimentateurs à reproduire leurs propres résultats ou ceux de leurs confrères, ce qui met ce qu'on a pu appeler la « crise de la réplication » au cœur des justifications de la réticence de chacun, dont des philosophes, à utiliser les résultats expérimentaux de psychologie morale.

2.3.1 Rien de nouveau

Kwame Anthony Appiah ouvre son ouvrage « Experiments in Ethics » (Appiah 2009) [9] en soulignant qu'avant que la science physique moderne ne naisse, la philosophie avait toujours été expérimentale. Il soutient que c'est le départ des sciences naturelles de la philosophie au 17e siècle qui constitue un vrai moment de rupture et non le retour de l'expérimentation comme mode d'argumentation philosophique, comme le prétendent les tenants du mouvement XPhi [16]. Appiah appuie cette thèse sur les nombreux exemples des grands philosophes qui ont pratiqué l'expérimentation, depuis Aristote jusqu'à Descartes, Pascal ou Leibnitz, et qui ont utilisé ces résultats pour étayer leurs thèses philosophiques [17]. Appiah rejoint avec cette remarque les philosophes qui considèrent que l'objet de la philosophie n'est pas seulement constitué de l'ensemble des concepts que nous utilisons, concepts qu'elle se donnerait pour but de clarifier, mais que l'objet de la philosophie est de façon la plus englobante possible le monde lui-même [18]. La philosophie a donc de plein droit accès à tout type d'argument sur le monde dont, naturellement, ceux appuyés sur les démarches expérimentales. En ce sens le mouvement XPhi ne peut prétendre être innovant et doit être compris comme une adaptation de démarches anciennes aux outils d'aujourd'hui.

En appui de cette thèse, on peut remarquer que les grands exemples souvent retenus comme paradigmatiques du mouvement XPhi, comme ceux utilisant les contre-exemples de Gettier à la définition de la connaissance (1963), ou ceux sur le cas Godel de la référence (1966), ou le tramway de Philippa Foot (1967) ou les contre-exemples de Frankfurt portant sur le lien entre responsabilité morale et possibilité de choix (1969) se sont tous structurés dans les années 1960, plusieurs décennies avant que n'apparaisse ce mouvement.

Autre expression de la réticence à considérer la philosophie expérimentale comme une nouvelle discipline à part entière, en 2011, le PGR (Philosophical Gourmet Report), organisme professionnel qui permet aux étudiants et chercheurs en philosophie d'identifier les

16. On peut remarquer qu'un ouvrage du 17e siècle de Margaret Cavendish a pour titre « Observations upon Experimental Philosophy » ce qui illustre qu'on pouvait encore à cette époque associer les deux termes de philosophie et d'expérience pour écrire sur des conceptions du monde qui relèvent aujourd'hui des sciences physiques.

17. Sur ce sujet voir également (Anstey 2016) qui reprend l'histoire de la philosophie expérimentale au 17e siècle et fait le parallèle avec le mouvement XPhi [8]

18. Timothy Williamson dénie aussi que la philosophie puisse être réduite à l'analyse du langage dans (Williamson 2007 p 23) [235]

universités de philosophie anglophones selon leurs activités, a refusé d'inscrire la philosophie expérimentale dans sa liste de spécialités philosophiques au motif qu'il s'agissait d'une méthodologie et non d'un nouveau domaine académique. [19]

On peut, a contrario, argumenter que le mouvement XPhi correspond à des changements importants tant qualitatifs que quantitatifs. Qualitativement, il s'agit principalement des perspectives ouvertes par le développement des connaissances sur la cognition humaine. Il s'agit également des changements intervenus dans la compréhension des connaissances acquises à l'aide des démarches expérimentales. Et je classerai comme changement quantitatif les possibilités ouvertes par les outils de l'Internet pour mener les expérimentations, même si beaucoup de chercheurs pensent que ce changement est d'une telle ampleur qu'il vaut mieux le concevoir comme qualitatif et ne pas le réduire à un changement seulement quantitatif.

Qualitativement tout d'abord, le philosophe expérimental est aujourd'hui armé de tous les outils, théories et concepts développés par la psychologie scientifique et, en particulier, les neurosciences. Si ces outils n'étaient qu'anecdotiques, on comprendrait mal l'ampleur des réactions des philosophes aux articles de Libet sur le libre arbitre, aux articles de Gettier sur la connaissance ou à l'utilisation par Greene de l'IRMf pour comprendre le rôle des émotions dans les jugements moraux. La création d'une vraie psychologie scientifique au cours du vingtième siècle a créé les conditions pour une révision de toutes les activités humaines, dont bien sûr la philosophie, à la lumière de ses constats. Cette nouvelle science s'appuie aujourd'hui sur l'imagerie neuronale, sur la chimie neuroactive, sur l'observation clinique, et sur les sciences de l'éducation. Elle n'est plus limitée aux quelques cas pathologiques qui apparaissent aux fondements de la discipline (on peut penser ici par exemple au cas de Phineas Gage cité par Damasio). Que cette science de la psychologie humaine ne soit pas accomplie, qu'il lui reste un chemin immense à parcourir pour être pleinement la science du comportement humain, c'est une évidence. Mais qu'il soit possible de philosopher sans la prendre en compte apparaît à beaucoup comme simplement aussi impossible que de considérer que nous pourrions aujourd'hui relire les propos de Descartes ou Kant sur le temps et l'espace sans les mettre dans la perspective de ce que nous dit la physique moderne.

On peut remarquer, avec Florian Cova [55], que la réticence des philosophes à utiliser les résultats scientifiques, dont ceux de la psychologie, trouve peut-être une de ses origines dans les excès scientistes et positivistes du 19^{e} siècle, on peut alors suggérer qu'un changement qualitatif important dont bénéficie le mouvement XPhi est la fin de ces excès par leur remise en cause dans la seconde moitié du 20^{e} siècle au sein du mouvement de la philosophie ana-

19. Voir le lien `https://leiterreports.typepad.com/blog/2011/08/some-pgr-news.html#tp`

lytique. La prétention positiviste à construire une vision unifiée du monde autour des seules connaissances scientifiques n'est plus aujourd'hui au cœur de la préoccupation des philosophes pas plus qu'elle n'est portée par les scientifiques qui, de façon beaucoup plus modeste et conforme à leur pratique, adoptent (majoritairement) plutôt des conceptions de la connaissance scientifique beaucoup plus empreintes d'humilité telles que celle que je présente plus loin avec la métaphore de l'hélice scientifique expérimentale (voir 3.1, page 124).

Enfin, quantitativement, le mouvement XPhi bénéficie de l'accroissement exponentiel des réseaux et des capacités du traitement de l'information sous de multiples formes. Principalement, l'utilisation d'Internet a permis de développer des outils comme Amazon Mechanical Turk (AMT), et l'exemple de l'effet Knobe qui sera développé plus loin montre que ce sont près des deux tiers des études qui maintenant passent par ce type d'outil. AMT est un outil commercial, une plateforme Internet, sur laquelle s'inscrivent des internautes du monde entier souhaitant, contre une rémunération unitaire modeste, participer à des sondages. Les chargés d'étude, académiques ou commerciaux, soumettent à AMT des questionnaires et les critères permettant de sélectionner les participants (langue, pays, CSP, âge...). En quelques jours et pour quelques euros, on peut ainsi accéder aux réponses qualifiées (autant que le questionnaire le prévoit) d'un échantillon de personnes. Récupérer ces réponses sur le réseau et les intégrer dans un logiciel d'analyse statistique, par exemple le logiciel gratuit R, la référence dans le monde académique, se fait en quelques clics. Il reste alors à exploiter ces résultats et affiner autant que nécessaire le questionnaire pour poursuivre le programme de recherche. L'ensemble de cette chaîne et des outils qu'elle met en œuvre à chaque étape est détaillé à l'attention des philosophes expérimentaux dans l'ouvrage de référence de Sytsma et Livengood (Sytsma 2016) [224]. Je reprends plus loin un exemple de ce type de démarche avec l'étude à laquelle j'ai contribué sur la surestimation du nombre de musulmans en France.

La grande facilité d'usage de ces outils et la modicité de leur coût ont permis une véritable explosion du nombre d'études comportant des résultats empiriques publiées dans les différents domaines d'intérêt de la philosophie expérimentale. Il est certainement trop tôt pour évaluer l'apport qualitatif de ce saut quantitatif, mais il est néanmoins difficile d'affirmer qu'il sera sans conséquence sur le travail des philosophes. En cela, la philosophie se trouve confrontée aux mêmes questions que tous les autres domaines de la connaissance avec la difficulté à anticiper ce que les nouveaux modes de construction et d'exploitation des grandes bases de données comportementales pourront nous réserver de surprises dans les années qui viennent. Supposons ainsi, expérience de pensée, que l'enregistrement systématique des actions et déplacements de tous les individus d'une certaine ville nous permette d'établir un

modèle où la fréquentation de tel ou tel lieu, quartier ou établissement commercial, soit un prédicteur fiable du passage à l'acte délictuel individuel, quel serait le statut moral d'un tel modèle ? Serait-il considéré comme constituant une théorie morale ? La fréquentation de ces lieux devrait-elle être moralement interdite ? La concrétisation de cette expérience de pensée est bien sûr peu plausible aujourd'hui, mais le travail des philosophes expérimentaux pourrait aussi avoir pour utilité de nous préparer à ce type d'éventualité de par leur immersion dans le monde numérisé.

2.3.2 Le programme XPhi négatif est faible

Le programme négatif de la philosophie expérimentale a pour objectif de montrer que les outils habituels du philosophe analytique, et tout particulièrement la méthode des cas déjà évoquée avec les cas proposés par Gettier pour la connaissance et par Kripke pour la référence, manquent de fiabilité. Ce programme s'appuie schématiquement pour cela sur une démarche que je me propose de détailler. La méthode analytique est tout d'abord décrite en insistant sur quatre étapes itératives :

— Première étape : fournir une définition d'un concept. Par exemple, pour reprendre le concept de « connaissance » : « la connaissance est une croyance vraie et justifiée ».

— Deuxième étape : imaginer des cas, inspirés de situations réelles ou non, qui couvrent différentes acceptions du concept étudié. Gettier a ainsi imaginé ses fameux contre-exemples pour tester la définition du concept de connaissance.

— Troisième étape : examiner pour chaque cas ainsi imaginé la pertinence de la définition proposée. La pertinence est acquise si l'intuition confirme, dans ce cas particulier, la définition envisagée et, inversement, si lorsque l'intuition écarte le cas, la définition ne s'applique pas. Application de la définition et intuition du concept doivent avoir strictement la même extension sur l'ensemble des cas.

— Quatrième étape, si les extensions diffèrent, il faut alors réviser le concept et modifier la définition proposée. Dans les exemples de Gettier cités plus haut si la vérité d'une proposition est due au hasard et que le mode de justification est sans prise sur ce hasard, alors cette proposition ne constitue pas une connaissance, une croyance vraie et justifiée n'est pas dans ce cas qualifiée de connaissance. Il faut revoir la définition de la connaissance qui ne peut être une «croyance vraie et justifiée ».

Au sein de ces étapes, la troisième fait appel, et fait principalement appel pour les philosophes expérimentaux, à l'intuition du philosophe en son fauteuil qui décrète que le concept et

sa définition ont ou non la même extension dans les cas proposés. Or cette intuition n'est pas fiable et on peut le prouver empiriquement de plusieurs façons en s'appuyant sur les réponses d'un échantillon de personnes à la question de savoir si le concept et la définition s'accordent dans un cas donné. Première voie, on fait varier une caractéristique des personnes interrogées qui est sans lien apparent avec le concept étudié, par exemple l'appartenance culturelle. Si des personnes de cultures différentes répondent différemment au questionnaire, alors le concept n'est défini, au mieux, que relativement à une culture donnée. De même, on peut faire varier d'autres caractéristiques des personnes interrogées, leur genre, le niveau de formation, leur catégorie socio-économiques, ou leurs caractéristiques physiologiques (personnes présentant des pathologies particulières). Si un désaccord entre le philosophe en fauteuil et telle ou telle sous-population ayant telle ou telle caractéristique apparaît, alors que les caractéristiques sont sans rapport avec le problème posé, alors le désaccord sera interprété comme problématique pour la méthode des cas qui, traditionnellement, ne sollicite que le seul avis du philosophe en son fauteuil.

Deuxième voie, on fait varier le protocole expérimental, par exemple en inversant l'ordre des questions posées, ou en créant des conditions particulières qui changent l'humeur des sondés, sans lien avec les concepts testés [20]. Si les réponses varient en fonction de ces modalités qui sont pourtant sans lien avec la question posée, c'est donc que le processus psychologique qui conduit à proférer la réponse manque de robustesse. En élargissant ce constat, on infère que l'intuition appliquée à ces questions manque de fiabilité.

Edouard Machery dans son ouvrage « Philosophy within its proper bounds » (Machery 2017) [151] développe de façon extensive cette stratégie pour conclure à l'impossibilité pour le philosophe de s'attaquer de façon utile à des questions qu'il qualifie de philosophiquement immodestes car demandant pour être résolues des capacités que les humains n'ont pas. Un point clé de son argumentation est l'impossibilité pour la méthodes des cas de répondre à ce qui lui est demandé par les philosophes analytiques : garantir la définition des concepts.

Le programme négatif de la philosophie expérimentale a soulevé de nombreuses réticences. Ainsi Max Deutsch consacre son ouvrage (Deusch 2015) [72] à démontrer que le jugement du philosophe n'est pas une simple intuition mais repose au contraire sur une analyse complexe et réflexive objectivée dans des argumentations, et que les simples sondages à la base des critiques de la XPhi n'ont aucune de ces caractéristiques. Ces critiques des XPhi se trompent de cible car le travail des philosophes ne repose pas sur l'intuition.

20. Ruwen Ogien a efficacement énoncé cette démarche dans le titre de son ouvrage « De l'influence de l'odeur des croissants chauds sur la bonté humaine. »

Dans une autre direction, Williamson remarque que le travail des philosophes expérimentaux eux-mêmes s'appuie sur l'intuition au moment de considérer tel ou tel cas comme significatif et au moment d'interpréter les réponses des sondés. Plus largement, tout travail philosophique repose ainsi, pour partie, sur l'intuition et la critique des XPhi s'appliquerait à toute philosophie, y compris à la leur [21]. Enfin, autre point de vue sur le débat de la pertinence du programme négatif de la XPhi, on peut remarquer avec (Mukerji 2019 p. 40) [167] que si la démarche du philosophe analytique expérimental conduit à opposer les intuitions des uns à celles des autres, il est à craindre qu'on arrive dans des impasses, chacun mettant en avant les cas qui servent son propos et les interprétations psychologiques proches de ses positions. On trouve une illustration de cette situation dans certains des débats que je décrirai plus loin.

L'ensemble de ces réticences au mouvement XPhi, qu'elles portent sur la critique de la faiblesse du programme négatif ou sur la critique de sa prétention à la nouveauté, ne font pas particulièrement référence à la philosophie morale mais s'attachent plutôt à l'utilisation des méthodes de ce mouvement XPhi pour l'ensemble de la philosophie. De façon encore plus générale, les critiques des philosophes ont également beaucoup porté sur la qualité des études des XPhi qui serait inférieure à celle des études de psychologie, elles-mêmes inférieures à celles des sciences naturelles en général. Je me propose maintenant de faire le point sur ces critiques méthodologiques en reprenant les éléments disponibles sur la réplication des expériences de philosophie morale.

2.3.3 Philosophie expérimentale et réplication

La question de la réplicabilité des expériences se pose de façon particulièrement aiguë dans le domaine de la psychologie, depuis les constats alarmants de 2015 [50] qui montraient d'une part qu'une portion infime des articles comportant des expérimentations bénéficiaient d'une tentative de réplication publiée et, d'autre part, que quand une telle tentative avait lieu, le taux de réplication était étonnamment faible à moins de 40 % [22]. Les philosophes expérimentaux, conscients de cette difficulté, ont entrepris une campagne de réplication de leurs résultats qui a conduit à la publication d'un constat très positif avec environ 70 % de réplication réussie (Cova 2018) [58].

Ce résultat est positif sur au moins trois points. Premier point, il montre la mobilisa-

21. Pour aller plus loin dans le détail des critiques aux XPhi, voir (Kauppinen 2007) [131], (Sosa 2007) [215], (Horvath 2010) [119], et la discussion récente des arguments et contre arguments relatifs à l'utilisation de l'intuition par les philosophes dans (Cappelen 2014) [39]

22. Sur la réplication dans la démarche scientifique expérimentale, voir 3.5, page 151

tion de la communauté des chercheurs philosophes expérimentaux autour de ce thème de l'exigence de réplication et, conséquemment, pour conforter la qualité épistémique de leurs articles. Deuxième point, il s'agit d'une réponse intéressante aux critiques méthodologiques dont avaient fait l'objet certains des premiers articles de philosophie expérimentale de la fin des années 1990, d'amateurisme expérimental et statistique. Sous cet angle, le haut niveau de réplication montre que les méthodes expérimentales et statistiques employées sont raisonnablement fiables, et au moins au même niveau que celles des psychologues expérimentaux. Enfin, troisième point, les réplications comportant inévitablement quelques variations de contexte, ne serait-ce que par le décalage dans le temps, la réplication montre que les résultats sont robustes face à ces petites variations.

Cette fiabilité méthodologique des publications de philosophie expérimentale a été confirmée en 2018 sous un autre angle, celui de la validité statistique de l'utilisation de la méthode NHST (en anglais, Null hypothesis significance testing) qui consiste à évaluer si un effet existe ou non en calculant la probabilité que les résultats obtenus puissent n'être dus qu'au hasard dans le cas où l'effet n'existe pas. Dans une analyse reprenant 220 articles de XPhi, une équipe de l'université de Tilburg a montré que la qualité statistique en était plutôt meilleure que celle estimée dans les mêmes conditions pour des publications relatives à la psychologie expérimentale (Colombo 2018) [52]. Sur 220 articles, 174 ont été retenus pour une analyse poussée, et sur ces 174 articles, un a été écarté comme comportant des erreurs trop importantes et trop nombreuses, 106 sont satisfaisants, 67 présentent au moins une inconsistance statistique, éventuellement faible, et parmi eux 11 présentent au moins une inconsistance sévère. Bien que restant trop élevés, ces taux d'inconsistance sont inférieurs à ceux observés par les mêmes auteurs avec le même protocole dans les publications de psychologie expérimentale. [23]

La vérification de la réplication et la validité statistique semblent aller dans la même direction : les articles publiés par les philosophes du mouvement XPhi sont au même niveau que ceux relevant de la psychologie expérimentale. En revanche, on peut également remarquer que les réplications ont, de façon générale, suivi des protocoles extrêmement proches des expériences de départ. On est alors conduit à craindre que, dans le cas où le protocole initial comportait des risques de biais, ces biais aient été répliqués en même temps que les expériences. Il est donc vraisemblable que ces réplications n'apporteront guère de changement chez les philosophes qui en critiquent les conclusions non sous l'angle formel de la méthode,

23. La même conclusion est retenue dans (Start 2019) [220] qui procède à une étude sur 365 articles publiés de 1997 à 2017 : à plus de 80 % les articles de XPhi sont statistiquement probants, et le phénomène de p-Hacking rare.

expérimentale ou statistique, mais sur le fond de l'interprétation psychologique ou philosophique de ces résultats. On peut d'ailleurs remarquer que cette campagne de réplication n'a pas ralenti le rythme de nouvelles publications qui, faisant varier de nouveaux paramètres ou prenant en compte d'autres éléments du contexte, ont conduit à remettre en doute le fond des articles précédents. J'illustrerai ce phénomène plus loin avec le cas de l'effet Knobe et de l'analyse de l'attribution d'intentionnalité.

2.4 La philosophie morale expérimentale

Dans cette dernière section, j'en viens au cœur de mon propos, la philosophie morale expérimentale. Dans un premier temps, je souligne les spécificités de la philosophie morale en regard de l'expérimentation et, en particulier, de l'usage qu'en font les philosophes du mouvement XPhi, ainsi que des réticences que cet usage et leurs résultats ont soulevé. Ce point est important pour mon propos, puisqu'il justifie de la pertinence qu'il y a à interroger spécifiquement le rapport de la philosophie morale à l'expérimentation. Dans un second temps, je décris le cas paradigmatique de l'effet Knobe, fleuron du mouvement XPhi, qui illustre particulièrement bien à la fois l'apport de ce mouvement, aucun philosophe n'aurait avant les XPhi prédit cet effet, et les limites de cet apport qui n'a peut-être pas contribué à clarifier notre compréhension de la notion d'intentionnalité. Les deux cas suivants, l'attribution des états mentaux et a question de la compatibilité entre le déterminisme et le libre arbitre viennent compléter cette analyse.

2.4.1 Des spécificités de la philosophie morale expérimentale.

En préalable à cette présentation plus axée sur la philosophie morale expérimentale, je souhaite m'interroger sur la pertinence qu'il y a à séparer ainsi, sous l'angle expérimental, la philosophie morale des autres domaines de l'activité philosophique. En effet, si on voit l'expérimentation comme un simple outil méthodologique à disposition du philosophe [24], au même titre que, par exemple, l'analyse conceptuelle ou l'étude du langage ordinaire, il semble inutile, et peut-être ad hoc, d'opérer cette distinction.

Je pense néanmoins pouvoir avancer aux moins quatre raisons justifiant cette séparation qui ne résultent pas d'un simple effet de cadrage dû au sujet de ma thèse. Première raison, le programme négatif n'a pas pour la philosophie morale la même portée que pour

24. Sytsma et Livengood revendiquent par exemple cette vision de boîte à outil (« toolbox ») dans (Sytsma 2016 p 9) [224]

la sémantique ou l'ontologie. En effet ce programme négatif de la philosophie expérimentale vise, comme nous l'avons vu, à établir empiriquement le manque de fiabilité de l'intuition philosophique utilisé dans la méthode des cas. Or ce qui est profondément en jeu dans le cas « Gödel » ou dans les contre-exemples de Gettier, c'est la compréhension des concepts de « nom propre » ou de « connaissance ». La question n'est pas de savoir si la référence par nom propre ou l'établissement des connaissances fonctionnent ou non, nous partons du constat que ces capacités humaines existent, mais plutôt de mieux établir les concepts qui permettent de comprendre ces capacités humaines. Si les philosophes expérimentaux ont raison, l'intuition du philosophe ne constitue pas un mode d'appréhension fiable de ces concepts et la conséquence est que la compréhension déduite de la méthode des cas est fautive. Par opposition, dans le cas de la philosophie morale, il ne s'agit pas seulement de concepts, il s'agit avant tout d'action, de bien faire, de bien vivre. La question n'est pas principalement de savoir mais de faire. Face à l'action, chacun mobilise les ressources à sa disposition immédiate, instincts, réflexes, intuitions, sentiments, connaissances (et comment pourrait-il en être autrement ?), et que chacune de ces ressources ne soit pas d'une fiabilité absolue est une évidence qu'il n'y a aucune difficulté à établir et qu'il n'y aurait aucune utilité à combattre. C'est justement le problème posé à la philosophie morale de, malgré toutes les limitations humaines, proposer des chemins vers l'action morale. De ce fait, on peut argumenter que le programme négatif de la philosophie expérimentale a peu d'intérêt pour la philosophie morale.

Une deuxième raison pour distinguer, sous l'angle expérimental, la philosophie morale des autres domaines de la philosophie, est appuyée sur la question axiologique. Comme nous l'avons vu les théories morales s'intéressent à ce qui doit être et non à ce qui est. Elles établissent des règles morales qu'il conviendrait de suivre, mais que bien sûr personne ne respecte absolument ici et maintenant. Il semble donc qu'aborder la philosophie morale par l'expérimentation demandera, a minima, une explication particulière sur ce point car la portée de l'expérimentation est limitée au constat de ce qui est, et son intérêt pour aborder ce qui doit être reste à établir. Naturellement, la métaphysique ou l'épistémologie peuvent également s'intéresser à ce qui devrait être, et, par exemple, à comment les humains devraient penser pour prétendre accéder à la connaissance, en un mot elles peuvent avoir une dimension réformatrice et non uniquement descriptive. On peut donc observer que la différence avec la philosophie morale sera une affaire de degré dans la normativité. Proposons que pour la philosophie morale aborder ce qui doit être, et non décrire ce qui est, est au cœur même de ses préoccupations, et que c'est plus périphérique pour les autres domaines philosophiques. On peut avancer (non sans risque comme je me propose de le développer plus loin) que les résul-

tats expérimentaux sont centraux pour qui cherche à comprendre le monde mais simplement utiles pour qui cherche à l'améliorer.

Ensuite, troisième raison, la grande proximité entre la philosophie morale et la psychologie morale n'a pas d'équivalent dans les autres domaines de la connaissance que la séparation des sciences naturelles et de la philosophie a structuré depuis plusieurs siècles. On peut illustrer cette proximité entre philosophie et psychologie par le nombre significatif de chercheurs allant de l'une à l'autre alors que ces mouvements sont rares dans les autres domaines [25]. On peut également analyser en ce sens la composition du groupe Moral Psychology Research Group `http://www.moralpsychology.net/` dont on ne saurait dire, sauf à être un expert de la chose académique américaine, s'il s'agit de philosophes ou de psychologues. L'ouvrage de référence « The Moral Psychology Handbook » (Doris 2012) [76] comporte 20 contributeurs dont 4 appartiennent à des structures académiques de psychologie, 11 à des structures de philosophie et 5 à des structures explicitement mixtes entre philosophie, psychologie et sciences cognitives. Dans cet ouvrage, John Doris rappelle que la psychologie morale est par construction interdisciplinaire puisqu'elle allie la psychologie, c'est sa composante scientifique, à la morale, sa composante philosophique, longtemps rétive à l'approche naturaliste. John Davis considère d'ailleurs qu'on peut dater des années 1960 la naissance de ce mouvement hybride tel qu'on le connaît aujourd'hui avec l'influence grandissante du naturalisme en philosophie, tant en épistémologie qu'en philosophie de l'esprit, et, inversement, l'influence décroissante du behaviorisme en psychologie qui a permis le rapprochement de la psychologie et des sciences expérimentales. Comme je l'ai mentionné plus haut avec l'exemple de l'IRMf, c'est l'apparition des nouvelles possibilités d'investigation apportées par les sciences cognitives, en prolongement des études psychophysiques qui avaient constitué une première approche dans les années 30, qui ont permis l'explosion des études de psychologie morale expérimentale.

Enfin, quatrième et dernière raison que j'avancerai, il existe de très sérieux arguments pour affirmer qu'il serait immoral de ne pas prendre en compte les résultats des sciences cognitives pour l'étude de la moralité humaine. Vanessa Nurock voit ainsi deux principes proprement moraux qui conduisent à ce que le philosophe moral s'empare de cette compréhension de la psychologie humaine pour l'intégrer dans son approche normative (Nurock 2011 p 94) [173]. Le premier, qu'elle nomme principe d'humanité, est qu'une théorie qui viserait à contraindre les humains à des obligations qu'ils sont dans l'impossibilité naturelle de respec-

25. A l'exception des vieux scientifiques qui passent à la philosophie sur le tard, comme l'auteur de cette thèse, mais sans pour autant entremêler les disciplines.

ter serait qualifiable d'immorale. Ainsi, si, pour bien agir, une théorie morale exige qu'il faille embrasser l'ensemble des conséquences potentielles de l'acte envisagé, et si cette tâche est, comme le montrent les psychologues, humainement impossible, alors on peut considérer que cette théorie contraint l'acteur à l'immoralité et, qu'en cela, elle est immorale.

Le second, que Vanessa Nurock nomme principe d'exigence, est qu'il n'est pas du tout équivalent de respecter une obligation parce qu'elle s'impose du fait des caractéristiques de notre constitution psychologique et de la respecter parce que nous avons décidé de le faire. En ce sens, et même s'il y a redoublement apparent de l'obligation, l'approche normative du philosophe moral est autonome en regard de l'approche scientifique. Que le philosophe moral doive, pour partie, prendre en compte les connaissances acquises sur la nature humaine ne signifie en aucune façon qu'il doive s'y limiter.

Chacune de ces quatre raisons, le poids du programme négatif des XPhi, la question axiologique, la proximité entre philosophie morale et psychologie et, finalement, l'obligation morale de prise en compte des connaissances, liée pour Vanessa Nurock aux principes d'humanité et d'exigence, fait de la philosophie morale expérimentale un sujet présentant des spécificités en regard de la philosophie expérimentale en général, ce qui me semble donc justifier l'approche retenue dans la présente thèse.

2.4.2 Le cas paradigmatique de l'effet Knobe

Le philosophe moral, dans une tradition millénaire, construit des théories morales sur ce qu'il est bien ou mal de faire, des théories qui disent comment les humains devraient se comporter. Pour présenter ces théories de façon didactique et explicite, et également les évaluer en regard des autres théories ou du comportement réel des humains, chaque philosophe construit des situations fictionnelles ou inspirées de cas réels sur lesquelles il exerce son jugement moral et invite ses auditeurs à en faire de même. Ces jugements moraux, appuyés sur des intuitions morales partagées à propos de situations réelles ou fictionnelles constituent ainsi son principal matériau pour démontrer, valider et enseigner sa théorie morale. L'enjeu de la philosophie expérimentale est de rompre avec cette tradition du « philosophe en fauteuil » et d'élargir l'approche des questions morales à l'aide des outils expérimentaux utilisés dans les sciences humaines et, en particulier, par les psychologues.

Pour décrire plus précisément les contours de cette approche, prenons l'exemple d'un des articles les plus connus de Joshua Knobe, un des promoteurs du mouvement XPhi. Cet article de 2003 [138] porte sur l'attribution d'intentionnalité d'une action [26]. Il présente le scéna-

26. Cette expérience décrite en détail dans (Knobe 2006) [139] a fait également l'objet d'une présentation vidéo :

rio d'un chef d'entreprise lançant un projet motivé uniquement par sa rentabilité, en ayant marqué son indifférence aux éventuelles conséquences positives ou négatives que pourrait avoir le projet sur l'environnement. Le projet se déroule tel que décidé, est très rentable et a les conséquences prévues, positives ou négatives, sur l'environnement. L'expérience consiste alors à demander à des participants si ce chef d'entreprise a intentionnellement amélioré ou nui à l'environnement, ou, en alternative, si les conséquences sur l'environnement ne peuvent lui être imputées, étant non intentionnelles. Les réponses montrent alors une nette dissymétrie : les participants disent que le chef d'entreprise a intentionnellement causé les dégâts et qu'il faut le blâmer quand les conséquences sont négatives, mais qu'il n'a pas intentionnellement amélioré l'environnement et qu'il n'est pas à féliciter quand les conséquences sont positives.

Cette dissymétrie expérimentalement constatée qui fait dépendre l'attribution d'intentionnalité du jugement moral porté sur les conséquences de l'action décidée n'est pas un phénomène nouveau pour les psychologues, pas plus que la méthode de questionnaire utilisée. L'apport de Joshua Knobe, du point de vue du psychologue, n'est donc ni de fond ni de méthode. Son apport se situe à un autre niveau, il est proprement philosophique : l'auteur considère que cette dissymétrie a de multiples interprétations philosophiques possibles, interprétations qu'il va d'ailleurs lui-même décliner dans plusieurs articles [27]. On peut en particulier défendre un argument qui pose que cette dissymétrie ne pourrait facilement être prise en compte dans le cadre des deux grandes familles de théories morales offertes par les philosophes que sont le conséquentialisme (voir point 1.2.2.2 page 46) et le déontologisme (voir point 1.2.2.1 page 44).

En deux mots, pour le conséquentialiste, une action est morale si les conséquences en sont souhaitables, pour le déontologiste, l'action est morale si elle est conforme à des règles morales impératives. Ni dans un cas, ni dans l'autre, il n'est cohérent que l'intention même de l'action soit attribuée au chef d'entreprise en fonction du caractère positif ou négatif de notre jugement, moral, des conséquences de l'action. Dans le cas du conséquentialisme, il est évidemment déséquilibré de considérer que la conséquence est à prendre en compte dans le jugement moral si elle est négative mais pas si elle est positive. Un conséquentialiste cohérent devrait féliciter le chef d'entreprise en cas de succès autant qu'il le blâme en cas d'échec. Le cas du déontologisme est un peu plus complexe. On peut le résumer ainsi : soit le projet lancé par l'entrepreneur est conforme aux règles morales, soit il ne l'est pas, et tous les élé-

https://www.youtube.com/watch?v=sHoyMfHudaE

27. 15 articles de 2003 à 2009 d'après le site web de Joshua Knobe consulté le 27 septembre 2019, https://campuspress.yale.edu/joshuaknobe/publications/

ments nécessaires au jugement moral lui sont connus au moment de la décision, y compris la prise de risque environnemental (positive ou négative) acceptée par le décideur. Si on juge que le chef d'entreprise est à blâmer (ou à féliciter) sur le plan moral, alors il doit l'être, que la conséquence réelle constatée après le projet soit positive ou non, car la faute (ou la bonne action) morale a été accomplie avant, au moment de la décision. La conséquence environnementale liée au projet advient après l'application, ou la non application, de la règle morale impérative, elle résulte d'événements contingents hors de la capacité d'agir du chef d'entreprise, et ce de façon explicitement acceptée puisqu'il a marqué sa totale indifférence envers ces conséquences environnementales. Un déontologiste cohérent devra féliciter le chef d'entreprise s'il a suivi les normes morales et le blâmer s'il les a enfreintes, indépendamment des conséquences du projet.

Aucune des deux grandes familles de théories morales ne peut rendre compte de l'effet Knobe de façon satisfaisante. On peut donc considérer qu'elles seraient empiriquement mises en difficulté par l'expérience de Joshua Knobe et qu'il faudrait écarter, ou a minima reconsidérer, ces deux familles de théories morales.

En suivant cet exemple de l'effet Knobe, on peut dégager deux caractéristiques de la philosophie morale expérimentale. D'abord, et trivialement, elle s'appuie sur les méthodes et analyses développées par les psychologues pour aborder l'étude du comportement humain. Ensuite, et surtout, la philosophie morale expérimentale repose sur la thèse selon laquelle il serait possible de concevoir et mener des expériences en vue de construire des arguments utiles aux débats proprement philosophiques.

Les controverses

Cette approche expérimentale de la philosophie morale a donné lieu à de nombreux articles dont l'ambition est d'apporter de nouveaux arguments empiriques aux philosophes qui acceptent de sortir de leur fauteuil. Mais cette thèse a également conduit à de nombreuses controverses [28] aussi bien sur la pertinence des analyses psychologiques tirées de ces expériences définies avec une visée philosophique que sur la possibilité d'induire à partir des résultats des expériences des arguments intéressants sur le plan philosophique. Ainsi, dans l'exemple ci-dessus de la dissymétrie de l'effet Knobe, il a été proposé que les participants répondaient bien à la question posée, l'attribution d'intentionnalité, en cas de conséquence positive sur l'environnement mais que, en cas de conséquence négative, les participants souhaitaient punir le chef d'entreprise et donnaient donc la réponse qui va le plus sûrement conduire à cette condamnation. Si on admet que ce processus psychologique offre une bonne

28. Voir une revue de controverses dans (O'Neill et Machery (eds.) 2014) [175].

interprétation du phénomène, alors les participants déplaceraient la question de l'intentionnalité vers celle de la culpabilité et, si c'est le cas, l'interprétation psychologique qui consiste a contrario à considérer que le jugement moral a directement influencé l'attribution d'intentionnalité est insuffisante. L'argument philosophique, appuyé sur l'effet Knobe, supposé mettre en difficulté les deux grandes théories morales perd alors sa pertinence. Mon propos n'est pas ici de prétendre que telle ou telle interprétation psychologique est absolument, ou même relativement, meilleure qu'une autre, ni même de prétendre qu'une telle interprétation psychologiquement valide existe et qu'elle serait (ou non) accessible à l'expérimentateur, il est plus simplement de souligner que l'interprétation psychologique est une condition préalable à la construction d'une argumentation philosophique cohérente.

Retournement de perspective

Il est donc possible, en variant les interprétations psychologiques de l'effet Knobe, de mettre en doute les conclusions philosophiques qu'on peut en tirer car elles dépendraient de résultats provisoires entachés du doute lié à toute approche scientifique inductive et, plus spécifiquement, à la pluralité des approches des psychologues. Mais ce n'est pas la seule arme dont dispose le philosophe moral, s'il est peu enclin à accepter cette intrusion expérimentale. Il peut également nier tout intérêt à cette approche sans avoir à rentrer dans l'argumentation psychologique. Il peut pour cela bloquer l'inférence qui va de l'impossibilité pour les théories morales de rendre compte de l'effet Knobe vers la non pertinence de ces théories en remarquant que les théories morales déontologistes et conséquentialistes ont une visée principalement réformatrice et non descriptive. Ces théories ne disent pas comment les humains jugent mais comment ils devraient juger. Nous avons donc là un effet de changement de perspective, pour reprendre la distinction proposée au premier chapitre, l'effet Knobe appartiendrait à la perspective descriptive alors que la perspective de la philosophie morale substantielle est réformatrice. En ce sens l'effet Knobe peut être intéressant sur le plan psychologique pour pointer un biais qu'auraient les humains lorsqu'ils jugent, mais n'apporte pas grand-chose dans la perspective du philosophe moral engagé dans une théorie substantielle.

Cet exemple me conduit à un constat et à une question. Le constat est que l'expérience menée par Joshua Knobe est, ou n'est pas, jugée philosophiquement importante, y compris dans une perspective descriptive, selon que l'on adopte telle ou telle interprétation psychologique des réponses des participants. La question est alors celle de la généralisation de ce constat : cette fragilité des prolongements philosophiques est-elle un cas isolé dû à la structure particulière de cette expérience ou traduit-elle une difficulté plus profonde liée aux spécificités de la philosophie morale ? C'est pour tenter d'apporter un élément de réponse à cette question que

je vais développer plus loin (point 4.6 page 219) l'analyse d'un corpus de 33 articles portant sur cet exemple de « l'effet Knobe » de façon à préciser les concepts et mécanismes proposés par chacun de ces articles dans le cadre de l'étude de l'intentionnalité. Ces déplacements conceptuels conduisent à mettre en avant des processus psychologiques différents d'un article à l'autre et, en prolongement, à des interprétations philosophiques diverses dépendantes de l'interprétation psychologique des résultats des expérimentations.

2.4.3 Les exemples de l'attribution des états mentaux et du déterminisme

Outre l'attribution d'intentionnalité explorée par l'effet Knobe, les philosophes moraux expérimentaux ont exploré de nombreux aspects du comportement humain. Pour aborder la question de la généralisation potentielle à tous ces comportements des constats portés par la section précédente, je présente ci-dessous deux études portant sur des concepts importants pour la relation entre psychologie morale et philosophie morale. La première concerne le domaine de la conscience, et, plus précisément, la question de l'attribution d'états mentaux conscients aux robots ou au groupes. La seconde concerne le débat entre déterminisme et libre arbitre. Dans chacun de ces cas, l'apport des Xphi se traduit, comme pour l'effet Knobe, par une complexification des questionnements lorsqu'ils ont à prendre en compte les résultats empiriques.

Attribuer des états mentaux

La question de l'attribution d'états mentaux à des robots est aujourd'hui économiquement importante du fait du développement des interactions quotidiennes avec tous types de machines et, en particulier, du fait de l'objectif d'utilisation des robots sociaux en contact direct avec les malades ou les personnes âgées. En 1970, le roboticien japonnais Mori Masahiro a étudié les réactions de personnes mises en contact avec différents types de robots et il a montré empiriquement que ces réactions dépendaient de façon complexe de la possibilité de reconnaître physiquement le robot comme humanoïde (Mori 1970) [166]. Au delà de l'étude des robots, ces résultats sont importants pour la philosophie de l'esprit car ils portent sur un cas limite de l'attribution d'états mentaux et ainsi éclairent la capacité humaine à lire, deviner, et parfois inventer, les états mentaux chez les autres humains, les animaux et les artefacts.

En 2008, Joshua Knobe et Jesse Prinz se sont proposé d'étudier non la conscience en elle-même mais les intuitions qu'ont les personnes sur les questions impliquant l'attribution

d'états mentaux (Knobe 2008) [141]. Ils posent deux hypothèses, la première est que toute personne a, spontanément et sans formation à la philosophie, un concept de conscience phénoménale [29] qu'elle sait distinguer d'autres états mentaux. Ainsi, toute personne ferait une distinction entre « voir un objet rouge » et « savoir qu'il y a un objet rouge dans la pièce », dans le premier cas il y a conscience phénoménale, liée à la perception, dans le second cas il y a conscience, mais pas phénoménale.

La seconde hypothèse est que, toujours en dehors de toute formation philosophique, nous attribuons à un tiers des états mentaux, et en particulier la capacité de conscience phénoménale, différemment selon la constitution de ce tiers, humain ou robot, individu ou société ... Ainsi, nous pouvons dire d'un robot qu'il a une connaissance, « le robot sait faire des soudures », mais cela sonne étrangement si nous lui attribuons une conscience phénoménale « le robot est ému par la vision de ce rouge ». En combinant ces deux hypothèses, on obtient des propositions qui sont accessibles à l'expérimentation. Par exemple, on peut émettre la proposition qu'il ne sera pas attribué de la même façon à une entreprise une croyance (Total pense que la culture de l'huile de palme est un bien) qu'il lui sera attribué un état de conscience phénoménale (Total éprouve de la douleur) alors qu'on attribue les deux, croyance et conscience phénoménale, à un individu physique, la différence étant supposée liée à la constitution de l'entreprise différente de celle d'un individu.

Les auteurs déploient successivement 6 études sur cette base pour instruire leur proposition : l'attribution d'états mentaux (non phénoménaux) à l'interlocuteur est un outil permettant de faciliter l'explication, la prévision, le pilotage de l'action en anticipant les réactions de l'environnement. En revanche l'attribution d'états mentaux phénoménaux est un outil destiné à permettre le jugement moral de l'action envisagée. Il est donc avantageux, selon l'opinion commune mesurée par les études des auteurs, d'attribuer des états mentaux à tout un ensemble d'objets avec lesquels nous sommes en interaction, dans l'objectif de mieux les intégrer à nos projets et, ceci, indépendamment de leur constitution [30]. Et il n'y aurait en revanche attribution de conscience phénoménale que lorsque l'interlocuteur est perçu comme faisant l'objet d'obligations morales, c'est-à-dire ayant la constitution nécessaire pour cela (constitution qui peut varier en fonction de la théorie morale de chacun). Cette conclusion, si on la suit, est importante à la fois pour les études sur la conscience qui devront prendre en compte cette différence dans les modalités d'attribution d'états mentaux à autrui, et pour

29. La conscience phénoménale serait l'état mental dans lequel nous met le fait d'avoir une certaine perception, l'effet que, par exemple, nous fait de voir un objet rouge.

30. On peut penser par exemple à notre comportement vis à vis des pannes de machines avec des expressions comme « Cette imprimante m'en veut ! Elle tombe encore en panne la veille de la remise du document. »

la philosophie morale dont les théories devront intégrer les contraintes imposées ainsi dès la perception de l'état mental d'autrui différenciée selon son statut moral.

Déterminisme et libre arbitre

Autre exemple exploré par les philosophes moraux expérimentaux, le dilemme du déterminisme et du libre arbitre figure parmi les problèmes philosophiques les plus étudiés depuis l'antiquité. Dans sa formulation moderne, après Laplace, il se pose ainsi : si les lois de la nature déterminent l'avenir (et le passé) à partir de l'instant présent, quel sens donner à la liberté que l'homme aime se reconnaître et, en prolongement, à la responsabilité morale ? Soit tout est déterminé, et la liberté et la responsabilité sont illusoires, soit l'homme est libre et responsable et le déterminisme est une théorie fausse. A ce dilemme correspondent deux positions dites l'une « compatibiliste » selon laquelle la responsabilité morale est compatible avec le déterminisme, l'autre « incompatibiliste » selon laquelle il faut alors renoncer soit au déterminisme soit à la responsabilité morale [31].

Ce domaine de recherche a mobilisé les philosophes expérimentaux en prolongement des articles de Libet qui ont conduit à montrer que la prise de conscience d'un mouvement à faire intervenait après que le mouvement ait été initié, suggérant que si le libre arbitre suppose le choix conscient, alors l'action ne résulte pas de ce libre arbitre ainsi conçu (Libet 2002) [146]. Dans des articles de 2005 et 2006, Eddy Nahmias a soutenu qu'on pouvait accéder empiriquement au fait de savoir laquelle des deux options, compatibiliste ou incompatibliste, était la plus intuitive. Il a présenté les résultats d'enquêtes tendant à montrer que, contrairement à l'intuition des philosophes, l'option compatibiliste est dans certains cas la plus adoptée par les sondés (Nahmias 2006) [170]. De nombreux développements ont suivi ces premiers articles et le débat est toujours ouvert comme le montre un article de 2018 qui met en cause les résultats de Nahmias (Lim2018) [147]. Nahmias présente aux participants des petites histoires sous forme de vignettes qui sont censées présenter des mondes parfaitement déterministes, et présenter ensuite les actions qui, dans ce monde, doivent être jugées moralement. Les auteurs mettent en avant les grandes difficultés de compréhension soulevées par ces vignettes présentées par Nahmias et, en particulier, sur ce que signifie pour un monde d'être déterministe. Il n'est donc pas certain que les réponses des participants puissent être interprétées comme réellement compatibilistes.

Je pourrai multiplier les exemples des recherches des XPhi qui ont abordé, comme je l'ai rappelé plus haut, tous les domaines de la philosophie en général et de la philosophie morale en particulier. Mais les deux exemples ci-dessus suffisent à suggérer que, d'une part, il n'est

31. Pour un ouvrage en français sur ce sujet, voir (Appourchaux 2014) [11]

pas de question philosophique, aussi complexe soit-elle, qui ne puisse donner lieu, au moins sous un aspect particulier, à une approche expérimentale. D'autre part, il apparaît peu vraisemblable que, dans chacun de ces deux exemples, les apports qui peuvent être attendus de ces approches expérimentales soient de nature à trancher définitivement les débats. Je vais revenir sur ce bilan en demi-teinte dans la section suivante. Ces deux exemples viennent ainsi conforter ce que nous avons vu avec l'effet Knobe : des questions centrales pour le philosophe supposent de manier des concepts métaphysiquement lourds, comme « liberté », « libre arbitre » ou « conscience » que l'expérimentateur a bien du mal à mettre en œuvre dans une expérience sans, qu'à chaque fois, le comportement réel des participants à l'expérience n'en révèle une nouvelle facette.

2.5 Un bilan en demi-teinte

Pour conclure cette section, et souligner le caractère en demi-teinte du bilan du mouvement XPhi tel qu'il peut être porté en 2019 par un philosophe favorable à ce mouvement, citons l'ouvrage de Nikil Mukerji (Mukerji 2019, page 120) [167] :

> En résumé, soulignons que les philosophes expérimentaux n'ont été capables de fournir de résultats définitifs dans aucun des quatre domaines examinés dans ce chapitre [32]. Mais soulignons également que, dans chacun de ces champs, ils ont été capables d'apporter des contributions intéressantes avec de nouvelles idées et possibilités auxquelles il aurait été difficile de penser depuis son fauteuil. Ces contributions ont aussi permis de lancer de nouvelles questions qui peuvent maintenant être abordées tant avec les méthodes expérimentales qu'avec les outils de la philosophie traditionnelle.

Cette citation permet de revenir sur la question, naïve, initiale : que peut-on attendre de la méthode expérimentale ? Et, plus précisément, que peut-on appeler un « résultat définitif » dans ce constat de Mukerji ? Et peut-on encore préciser ce constat en utilisant l'outil des quatre perspectives proposées au préalable, les perspectives descriptive, prescriptive, méta-éthique et de l'éthique appliquée ?

Si par « résultat définitif », il faut entendre la fin du débat philosophique et la résolution de la question psychologique posée comme on entend la résolution d'une équation mathématique, il est très peu plausible, et 2500 ans de philosophie en attestent, qu'une démarche quelle qu'elle soit, analytique ou expérimentale, puisse le fournir, même si clarifier le débat

32. Il s'agit ici de la définition de la connaissance, de la référence, de l'action intentionnelle et du libre arbitre.

peut être un objectif partiellement atteignable. Il est plus vraisemblable qu'un tel résultat définitif ne puisse être atteint qu'en adoptant une démarche dogmatique qui part de ce résultat définitif sans qu'il soit besoin de se préoccuper de comment on y arrive, car Dieu (ou l'évolution) se sont déjà chargés de le fournir.

Écartant cette attente exagérée, on peut détailler la question selon les quatre perspectives. Sous l'angle descriptif, est-il possible d'avoir un niveau de description, non définitif mais satisfaisant, des phénomènes qui sont à l'œuvre dans les deux exemples de l'attribution d'un état mental ou dans la perception de la liberté (ou de la responsabilité) d'un agent ? Dans cette perspective, l'apport de l'approche XPhi est avéré : on peut construire des expériences permettant de décrire dans quelles conditions les humains déclarent que des phrases comme «x veut ceci » « x aime cela » « x sait que P » sont acceptables en fonction de la nature de x (objet, entreprise, animal,..), de la nature du verbe et de la nature du « ceci, cela ». Naturellement, mener ce type d'enquête est complexe, donne des résultats statistiques sans certitudes, et fait apparaître de multiples complexités, et ce sera un des objectifs des études de cas développés plus loin que de souligner toutes ces limites par les cinq études de cas que j'ai menées. Mais il n'en reste pas moins que, d'une part, ces déclarations des humains sont accessibles et que le philosophe qui étudie ces phénomènes doit en produire une explication. Ainsi si les humains se disent « compatibilistes » et que le philosophe ne l'est pas, il se doit d'expliquer cette (éventuellement apparente) dissonance.

Sous l'angle prescriptif, si un résultat définitif devait être de dire ce que les humains doivent faire, dire ou penser, l'apport des expériences des XPhi à ces débats de philosophie morale apparaît insignifiant et, si on reprend les exemples ci-dessus, c'est plutôt en sens inverse que les apports se font, dans le sens de dévoiler des complexités plus importantes que ce que chacune des théories morales ne semble poser. Le terme souvent rencontré est alors celui de « pluraliste ». Si ni la théorie A ni la théorie B ne semblent pouvoir dire ce qu'il convient de faire dans toute circonstance, alors c'est qu'il faut peut-être prendre A dans certaines circonstances, B dans d'autre, et des croisements de A et B dans d'autres encore. En ce sens, les résultats des XPhi sont donc de nature à alimenter la réflexion dans la perspective méta-éthique dans la mesure où, si le philosophe considère comme important, ou au moins signifiant, que sa théorie puisse être mise en œuvre par des humains dont la psychologie est connue, alors il se doit de considérer cette nouvelle théorie « pluraliste » AB.

Enfin, peu est dit par les XPhi dans la perspective de la philosophie morale appliquée, et je reviendrai plus loin sur ce point mais, avant cela, il me faut maintenant préciser de façon plus détaillée ce que j'entends par l'expression « démarche scientifique expérimentale »

utilisée par les philosophes expérimentaux, et les psychologues moraux expérimentaux, pour qualifier leurs travaux. Ce sera l'objectif du chapitre suivant que de présenter une métaphore graphique de cette démarche. Cette description faite, je pourrai ensuite décrire les cinq études de cas qui viendront concrétiser mon approche de la philosophie morale expérimentale.

Chapitre 3

La démarche scientifique expérimentale

3.1 Introduction : pourquoi une métaphore ?

Il semble de bonne démarche analytique dans une thèse ayant trait à l'étude critique de l'apport potentiel de la psychologie scientifique expérimentale à la philosophie morale de préciser ce qu'il conviendra d'entendre par « démarche scientifique expérimentale » lorsque cette expression sera utilisée dans la suite de la thèse, et elle le sera souvent. C'est l'objet du présent chapitre. Malheureusement cette tâche, si on l'entend comme répondant à une demande de définition de ce qui est ou n'est pas expérimental et de ce qui est ou n'est pas scientifique, est en pratique impossible tant le sujet est vaste et les avis nombreux et variés [1]. Je suis donc amené à me rabattre, à titre opératoire, sur une solution de repli. Je choisis de m'appuyer pour cela sur une métaphore qui capture une partie importante de ce que j'entendrai par l'expression « démarche scientifique expérimentale » et qui, sans prétendre au rang honorifique de définition [2], permet, je crois, d'éviter bon nombre de malentendus. Ce sera l'objectif de la première section de présenter la métaphore de l'hélice, propulseur équilibrant les apports des différentes activités théoriques et expérimentales qui constituent la démarche, de présenter ce que cette métaphore véhicule, et d'évoquer ensuite son apport et ses limites de façon générale puis plus spécifiquement pour les sciences humaines.

La métaphore de l'hélice expérimentale, essentiellement, illustre le caractère dynamique

1. Pour une introduction détaillée à ce vaste domaine voir par exemple (Andler 2002) [4] et (Barberousse 2011) [16]

2. Une définition d'un concept au sens plein serait entendue comme l'ensemble des conditions nécessaires et suffisantes qui permettent d'utiliser le concept.

de la démarche scientifique expérimentale. Je n'entre pas ici dans l'ensemble des débats que soulève la recherche d'une définition de la démarche scientifique mais, plus schématiquement, j'en propose une description qui se veut instrumentale dans le cadre de cette thèse, c'est-à-dire une description permettant d'appuyer la réflexion sur l'apport potentiel de la démarche scientifique expérimentale à la philosophie morale. Cette limitation est ici portée par l'adoption d'une métaphore qui montre, mais ne démontre pas, les traits les plus importants de la démarche expérimentale. Positivement, cette métaphore traduit la dynamique de l'activité scientifique qui, sans qu'il soit besoin ni de fondements absolus, ni de quête de certitude [3], avance par étapes et itérations vers la résolution des questions que l'itération précédente a contribué à construire. Défensivement, cette présentation a pour ambition d'éviter les utilisations non qualifiées de l'adjectif « scientifique » dont les connotations doivent encore beaucoup à l'opposition envers des conceptions positivistes d'un autre siècle, qui, malheureusement, sont encore bien en place dans les travaux sur les rapports entre philosophie morale et démarche scientifique.

Avant d'entrer dans une description détaillée de la métaphore de l'hélice expérimentale dans les sections suivantes, soulignons les objectifs visés. Cette métaphore décrit une démarche essentiellement dynamique, la dynamique épistémique itérative qui, de théories en expériences, d'expériences en traces et de traces en théories, s'appuie sur les connaissances et savoir-faire en place, ainsi que sur les doutes et incertitudes qu'ils soulèvent, pour lancer de nouvelles recherches expérimentales dans un processus itératif sans fin. La métaphore illustre également le refus de trois postures en excès qui conduiraient à l'arrêt de cette dynamique : l'excès de dogmatisme, de relativisme et de pragmatisme. L'excès de dogmatisme conduit à considérer une théorie comme définitivement vraie, l'excès de relativisme consiste à considérer qu'il n'y a rien à généraliser à partir d'une expérience particulière, et l'excès de pragmatisme qu'il n'y a pas lieu de rechercher de théorie lorsqu'une réponse ponctuelle pratique suffit à apporter une réponse, même provisoire, à une question. Chacun de ces excès met un terme à l'interrogation itérative qui, du point de vue que je retiens ici, caractérise la démarche scientifique expérimentale.

Je poursuis ensuite la présentation en exploitant les connotations qu'offre cette métaphore de l'hélice en regard de plusieurs débats qui se trouvent ainsi non pas traités, ce qui serait encore une fois impossible dans le cadre d'une seule thèse, mais simplement momentanément écartés du chemin que je souhaite suivre avec l'analyse de l'apport de l'expérimentation à la philosophie morale pour cible. Il s'agit principalement, par l'adoption de cette métaphore,

3. Pour reprendre le titre de l'ouvrage de John Dewey dont cette métaphore est proche (Dewey 1929) [74].

d'écarter un certain nombre de points annexes qui pourraient venir polluer un aspect central du débat épistémique. Par exemple, la question du réalisme scientifique pourrait interférer avec la question du réalisme moral et cette interférence n'apporterait que confusion à mon propos. Autre exemple de confusion possible à écarter, le type de vérité que peut atteindre la démarche scientifique n'est pas semblable à celui requis par certaines théories morales. Ce sont ainsi plusieurs points importants de métaphysique que la métaphore fait apparaître comme auxiliaires pour la démarche scientifique, ce qui sera d'une certaine importance pour les réflexions sur la philosophie expérimentale.

Ensuite, je reprends dans cette présentation les différents types d'apports qui sont généralement attribués à l'expérimentation dans le cadre de cette démarche scientifique expérimentale. Ce travail permettra d'établir une base de comparaison avec les apports qu'on peut attendre de la démarche expérimentale dans le cadre de l'étude du domaine moral. La métaphore de l'hélice permet en particulier de mettre en évidence trois moments de la démarche scientifique, l'opérationnalisation, l'objectivation et l'interprétation inductive, qui seront au centre d'une partie des questions posées par l'application de cette démarche aux questions de philosophie morale. Ces trois moments seront présentés ici sommairement, éclairés par la métaphore de l'hélice, et seront repris de façon plus détaillée dans les chapitres suivants en regard de plusieurs études de cas d'utilisation de la démarche expérimentale par les philosophes moraux expérimentaux.

Enfin, la métaphore permet également de donner sa pleine dimension à la délicate question de la réplication des expériences, sujet qui a donné lieu à des constats alarmants pour la psychologie expérimentale, et que je présenterai sommairement dans ce chapitre. Si l'absence de réplication devait être interprétée, métaphoriquement, comme des dysfonctionnements de l'hélice, elle remettrait en cause l'effet de cumul attendu de la démarche scientifique et connoté par la métaphore de l'hélice. Il m'est donc apparu important de clarifier les enjeux liés à la réplication afin que cette métaphore puisse être validée pour la démarche scientifique expérimentale.

Je conclurai rapidement ce chapitre par un point d'étape permettant d'en résumer l'essentiel : une formulation de la métaphore de l'hélice expérimentale qui porte ce que j'entends ici par l'expression « démarche scientifique expérimentale ».

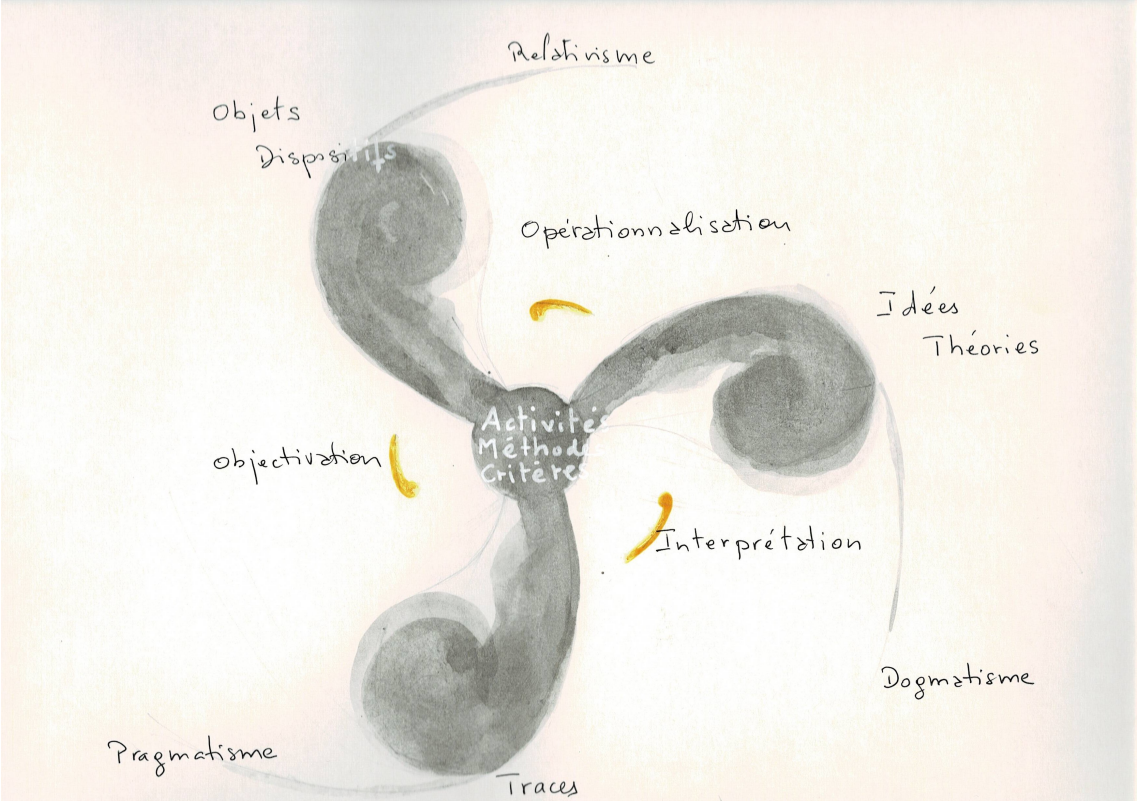

FIGURE 3.1 – L'hélice scientifique expérimentale

3.2 La métaphore de l'hélice

3.2.1 Une métaphore graphique

Pour présenter la métaphore de l'hélice, et en souligner le caractère instrumental, j'en propose ci-jointe une représentation graphique. Cette métaphore de ce que j'entends par « démarche scientifique expérimentale » emprunte à Ian Hacking[4] l'enchaînement des trois dimensions de la démarche scientifique faite d'idées, d'objets et de traces, et j'ai prolongé sur cette base dynamique la métaphore de Ernest Sosa qui proposait de concevoir la connaissance comme un radeau et non comme une pyramide (Sosa 1980) [214].

La métaphore de l'hélice[5] consiste à identifier chacune des pales à une activité particulière. Chaque pale-activité contribue au mouvement de l'ensemble mais chacune n'est efficace que parce qu'elle est reliée aux deux autres par et sur un arbre moteur. Ma proposition est de regrouper métaphoriquement sur chacune des pales de l'hélice une des trois dimensions proposées par Ian Hacking. Les idées sont regroupées en théories qui définissent les entités postulées, leurs propriétés et les relations qui les lient, propriétés et relations qui peuvent être

4. Voir Ian Hacking « The self-vindication of the laboratory sciences » pages 29-64 dans (Pickering (ed.) 1992) [182].
5. Dessin réalisé par Marie-Françoise Serra, voir www.mariefrancoiseserra.fr

simplement qualitatives dans certains cas ou quantifiées et mathématisées dans d'autres cas plus élaborés. Les objets, dispositifs pratiques d'observation ou d'expérimentation comportent à la fois ce qui est étudié, à la fois les instruments d'observation ou de mesure et l'ensemble du laboratoire ainsi que, par extension nécessaire, les savoir-faire pratiques que suppose leur mise en œuvre pour donner à voir les phénomènes étudiés. Ils sont regroupés sur la seconde pale. Les traces, résultats enregistrés et partagés de toutes les observations et mesures antérieures à un moment donné, sont regroupées sur la troisième pale qui symbolise ce qui est appelé les cahiers de paillasse dans certaines traditions scientifiques, et carnet de recherche dans d'autres, qu'on complète aujourd'hui par l'ensemble formé des bases de données et métadonnées, brutes et élaborées, issues d'un programme de recherche.

L'axe de l'hélice symbolise l'ensemble des activités qui font le quotidien des scientifiques au travail, à leur bureau ou à leur clavier, quand ils ne se battent pas corps à corps avec leurs expérimentations en cours. Ces activités sont faites des déductions, des calculs, des abductions, des inductions et généralisations, des simulations numériques, des phases d'invention et de conceptualisation... En bref, de tout ce travail intellectuel, aujourd'hui souvent soutenu par l'informatique, qui permet de faire tourner l'hélice de traces en théories, de théories en projets d'expérimentation et d'expérimentations en traces partagées.

Une pathologie importante des hélices intervient lorsqu'une des pales se fissure, déséquilibrant l'ensemble et entraînant le blocage et la destruction de l'hélice. De même à chaque moment, la démarche scientifique court le risque de se figer, de perdre sa dynamique. L'arrêt peut se faire de façon différente à chacun des moments figurés par les trois pale. L'excès de dogmatisme consiste à prendre une théorie pour vraie, définitive et complète. Le dogmatisme annihile le besoin de nouvelles confrontations à l'expérience et conduit à l'arrêt de l'écoute de ce que disent les traces des expériences passées et en cours. L'excès de relativisme consiste à prendre tout phénomène pour non généralisable, rendant, d'un côté, illusoire la recherche d'une expérience intéressante et, de l'autre côté, inutile la tentative de constituer et conserver les traces pour appuyer des théories générales. L'excès de pragmatisme, enfin, consiste à se contenter d'une trace comme résultat ici et maintenant, à appuyer ses décisions sur cette seule trace acquise sans rechercher de compréhension théorique plus profonde.

Le fonctionnement harmonieux de l'hélice est celui qui solidarise les pales, équilibrant l'apport de chacune à l'axe moteur, repoussant le risque d'arrachement d'une d'entre-elles. De même, le fonctionnement harmonieux de la démarche scientifique est celui qui équilibre les apports théoriques, expérimentaux et des données et repousse les risques d'excès dogmatiques, relativistes ou pragmatiques. Insistons sur la notion d'excès. Bien sûr il faut que le

théoricien soit convaincu de sa théorie pour qu'il s'investisse à la défendre, un peu de dogmatisme peut être utile. Bien sûr il faut que l'expérimentateur soit conscient des spécificités de son expérience de façon à en rechercher les faiblesses, un peu de relativisme ne peut pas nuire. Et bien sûr, il faut exploiter les traces quand elles sont pertinentes et immédiatement utiles, un peu de pragmatisme est raisonnable. Mais aucune de ces postures, conduite à l'excès, ne permet à l'itération épistémologique à la base de la démarche scientifique expérimentale, telle qu'elle est captée par la métaphore de l'hélice, de se mettre en place. [6]

3.2.2 Les connotations de la métaphore

L'intérêt d'une métaphore est de suggérer de nombreuses connotations avec une grande économie de moyens. Insistons sur ces connotations portées par l'hélice scientifique expérimentale.

La connotation sociale est transparente : l'activité scientifique est sociale et tous les scientifiques sont « dans le même bateau ». Les théoriciens, expérimentateurs, experts des données sont à la fois spécialisés et solidaires dans un mouvement auquel tous contribuent. Les trois pales d'une hélice ne sont pas en position hiérarchique, toutes contribuent à avancer, chacune successivement en position haute ou basse. La métaphore suggère que la démarche ne peut se comprendre que portée par plusieurs pales : une hélice à une seule pale ne peut mécaniquement fonctionner sans soumettre l'axe à des efforts impossibles, il en faut au moins deux, trois ou plus pour que l'hélice, équilibrée, joue son rôle. Le génie scientifique isolé, à la fois théoricien, expérimentateur et statisticien, s'il a jamais existé, n'est plus de ce monde.

La métaphore insiste sur le caractère dynamique de la démarche scientifique, toujours en mouvement, sans début ni fin. A chaque tour d'hélice, le bateau avance, partant d'une position pour aller vers une autre, progressant ainsi par appui successif sur chaque pale pour faire avancer les deux autres, et propulser le bateau. Une fois la dynamique lancée, elle ne s'arrête pas, suggérant que la curiosité est sans fin, sans limite et ne saurait être complètement assouvie.

Un physicien sera certainement sensible à un autre aspect de la métaphore : c'est parce que le fluide est visqueux que l'hélice est efficace et, de même, c'est parce que le monde résiste à être connu que la démarche expérimentale fait avancer la science.

La métaphore permet également, par défaut, de faire apparaître ce que la démarche scien-

6. L'excès de pragmatisme donne actuellement lieu à des questions importantes en regard du développement des programmes de Réseaux de Neurones Artificiels appliqués à des bases de données, dites Big Data. Ces programmes contribuent, justement, à bâtir des modèles de prévision sur la base des traces, sans passer par l'étape de la construction des théories. Je n'aborde pas ces questions dans la présente thèse. Pour une analyse détaillée on pourra se référer par exemple à (Krivine 2018) [143].

tifique n'est pas. Particulièrement important est le constat que si l'hélice propulse, elle ne permet pas de diriger le bateau, il faut rajouter un gouvernail pour donner et tenir un cap. De plus, l'hélice propulse le bateau à partir d'une situation donnée qu'elle ne contribue pas à définir. Elle ne peut non plus définir par elle-même quand elle doit s'arrêter, le port étant atteint. On peut alors pousser la métaphore un peu plus loin et imaginer un océan parcouru de multiples embarcations, toutes propulsées par le même type d'hélice, mais parties de positions initiales différentes et sur des caps différents. La diversité des approches expérimentales des différentes sciences spéciales est ainsi suggérée par cette métaphore océanique.

Ce qui caractérise la démarche scientifique, au travers de cette métaphore, n'est ni dans son contenu, ni dans son objet ni dans les critères régissant ses activités, car tout cela change à chaque tour d'hélice, mais dans une dynamique sociale et individuelle maintenue par la curiosité, le plaisir d'avancer, de répondre à des questions, et le refus de sortir de la spirale de la connaissance : refus de la magie, des excès de dogmatisme, de relativisme, ou de pragmatisme qui, tous, empêchent le cycle de se poursuivre.

La métaphore de l'hélice permet également de différencier la démarche scientifique expérimentale, la dynamique, de ce qu'on pourrait appeler « la science des manuels » [7], statique. Cette science des manuels est une photographie instantanée de ce à quoi la démarche a permis d'accéder à un moment donné. Si la science en train de se faire est l'activité humaine qui fait tourner l'hélice, la science des manuels est une représentation à destination du public, principalement enseignants et étudiants, des connaissances, données et savoir-faire acquis à un certain moment sur cet océan ainsi parcouru par une flottille de bateaux.

3.2.3 L'hélice, l'empirisme et la vérité

Avant de rentrer plus précisément dans ce que la métaphore de l'hélice permet de mettre en avant quant au rôle de l'expérimentation, observons-la à l'œuvre sur quelques problèmes philosophiques classiques. Cet examen est important pour la compréhension du type d'enquête que j'ai mené et qui sera présenté au chapitre suivant. Il s'agit de montrer que la démarche scientifique expérimentale ne se donne aucunement pour objectif de s'inscrire dans des débats métaphysiques en visant à y répondre mais, en revanche, les éclaire d'un jour particulier en distinguant ce qu'il est possible de connaître par cette démarche et les limites de cette connaissance. Ces limites ne sont pas anecdotiques et temporaires mais sont à la base de la construction de la démarche qui rejette les excès de dogmatisme, de relativisme et de pragmatisme. Il restera donc toujours à charge des métaphysiciens de reprendre ensuite,

7. Voir par exemple pour l'histoire des manuels en chimie l'ouvrage (Bensaude 2003) [22]

éclairés par ces apports s'ils le souhaitent, ces débats.

Rationalisme vs. Empirisme

Dans (Markie 2017) [156] l'auteur pose la question de l'opposition philosophique entre le rationalisme et l'empirisme :

> Le débat entre le rationalisme et l'empirisme porte sur la question de savoir à quel point nos connaissances dépendent de nos expériences sensorielles. Les rationalistes soutiennent qu'il existe des voies d'accès à la connaissance indépendantes des sens. Les empiristes soutiennent que tous nos concepts et toutes nos connaissances sont ultimement dépendants de notre expérience sensorielle. [8]

Ce que suggère la métaphore de l'hélice est que, loin de s'opposer, ces deux dimensions sont complémentaires dans l'activité scientifique, aspects différents d'un ensemble qui en comporte bien d'autres et que, précisément, c'est le refus de s'arrêter sur un de ces aspects qui est une caractéristique importante de la démarche scientifique expérimentale.

A chaque moment, la démarche s'appuie sur la dynamique acquise dans les itérations précédentes. Les théories en place sont importantes pour la conception et la réalisation des expérimentations, et les traces des expérimentations précédentes importantes pour ces théories, comme les expérimentations précédentes ont été importantes pour la construction de ces traces partagées. Ce que connote, essentiellement, la métaphore de l'hélice expérimentale, c'est cette dynamique de mutuelle construction, vision qui enlève une grande partie de l'intérêt porté aux discussions sur le fondement de la connaissance scientifique qui devrait être exclusivement soit théorique soit expérimentale, soit rationnelle soit empirique, alors qu'elle repose sur l'abandon de ces dichotomies.

Fondationnalisme vs. cohérentisme

Dans le débat entre différentes conceptions de la connaissance, il est philosophiquement utile de distinguer les théories fondationnalistes, les pyramides de Sosa (Sosa 1980) [214], des théories cohérentistes, le radeau de Sosa.

Ce que suggère la métaphore de l'hélice est qu'il est toujours possible de tenter de mieux appréhender un objet d'étude en multipliant les approches privilégiant successivement les idées, les objets et les traces dans le cadre d'une démarche scientifique. Il est alors possible, mais jamais certain, que soit atteinte une « connaissance » (une proposition vraie, crue, justifiée), mais en tout état de cause, et même si une telle proposition était atteinte, il ne saurait être question de la tenir pour telle, et absolument vraie, car ce serait tomber dans le dogmatisme. On est alors confronté à une difficulté : si on considère que pour être valide une

8. Traduction de l'auteur.

connaissance doit porter sur une proposition vraie et qu'on sait être vraie, alors il ne peut y avoir de connaissance établie par la démarche scientifique.

En suivant cette analyse, la « Vérité » au sens absolu et définitif, n'a pas sa place au sein de la démarche scientifique expérimentale. « Vérité scientifique » est un oxymore à remplacer par « proposition scientifiquement établie au moment t ». On doit remarquer ici que cette distance posée entre une proposition suffisamment établie et la Vérité, si elle existe, relève de notre capacité à connaître, de l'épistémologie, et qu'elle est largement indépendante des options métaphysiques quant à la Vérité absolue elle-même. Que la Vérité existe ou n'existe pas ne change pas grand-chose à la démarche scientifique expérimentale qui, de toutes façons, n'a pas cette Vérité pour objectif. [9]

Cette impossibilité de principe dont souffre la démarche scientifique, telle qu'imagée par la métaphore de l'hélice, d'accéder à la vérité absolue, et en supposant qu'une telle vérité existe, n'impressionnera pas particulièrement les philosophes car la métaphore laisse complètement hors du champ deux questions qui leur reviennent. Premièrement, la démarche scientifique expérimentale est-elle la seule visant à établir des propositions vraies ? Et, deuxièmement, s'il en existe d'autres, comment les comparer entre elles et, en particulier, avec la démarche scientifique expérimentale ? La démarche scientifique expérimentale n'est pas dépendante des réponses que chaque philosophe pourra apporter à ces questions qui ont une dimension métaphysique. Toutefois, inversement, chacune des autres démarches visant à établir des propositions vraies pourrait avoir des composantes accessibles à la démarche scientifique qui pourront faire l'objet d'évaluations expérimentales. Par exemple, si le philosophe envisage que les êtres humains ont des connaissances a priori, il sera possible de proposer des phénomènes qui tendraient à illustrer ces connaissances a priori et lancer la démarche scientifique expérimentale sur la recherche de ces phénomènes. La conclusion n'en sera jamais définitive, mais construira des arguments empiriques dont le philosophe attentif à l'avancée des connaissances scientifiques aura à rendre compte, ceci dans la seule mesure où il souhaite insérer son discours dans l'ensemble des résultats scientifiquement acquis.

Réalisme scientifique vs. antiréalisme

Lorsqu'une entité, par exemple un électron, apparaît comme partie constituante d'une théorie scientifique bien qu'il soit impossible d'accéder directement à une telle entité avec les sens, il est philosophiquement classique de se demander si cette entité doit être incluse dans notre ontologie ou s'il est préférable de lui donner un statut d'outil pratique sans épaisseur

9. Un scientifique qui pense que la Vérité existe définira celle-ci comme son horizon, et définira par ailleurs les objectifs de ses recherches. Nous pouvons atteindre un objectif, jamais l'horizon.

ontologique [10]. A titre de simplification, je considère ici que les théories scientifiques postulent l'existence d'entités, postulent également que ces entités ont certaines propriétés, et postulent enfin qu'il existe des relations entre ces propriétés et ces entités. En ce sens, une théorie est définie par l'ensemble de ces trois types de postulats, entités, propriétés et relations. La question du réalisme scientifique peut alors être posée par le métaphysicien : les entités, propriétés ou relations postulées par les théories scientifiques sont-elles réelles ou ne sont-elles qu'instrumentales ?

La métaphore de l'hélice apporte deux types de réponses à ces questions. D'une part, si être réel est pris en un sens absolu, comme pour la vérité dans le point précédent, la démarche expérimentale scientifique ne saurait donner une réponse définitive sur la réalité des entités, propriétés ou relations théoriques, sans risquer de tomber dans un dogmatisme antinomique avec sa démarche. Mais d'autre part, si une entité ou une relation contribue à une théorie satisfaisante, qu'elle donne lieu à des savoir-faire pratiques dans le cadre d'expérimentations qui aboutissent à des observations réussies et qu'elle est en bonne place pour fournir une compréhension des traces laissées par toutes les observations partagées, alors le scientifique est en droit de considérer qu'elles sont aussi réelles qu'une entité ou une relation puissent (scientifiquement) l'être.

Une autre question, proche mais différente, consisterait à se poser la question du réalisme des entités, propriétés ou relations théoriques, stipulées dans le langage ordinaire ou par un discours non scientifique, une religion ou une théorie morale par exemple, et inaccessibles à la démarche scientifique. Posée de cette façon, la question est, par construction, exclue du champ scientifique, mais l'activité scientifique peut néanmoins apporter de fortes présomptions sur l'existence ou non de ces entités ou, plus exactement, sur l'existence (au sens restreint exprimé dans le paragraphe précédent) de la contrepartie accessible à la démarche scientifique de ces entités. [11] Prenons un exemple dans le domaine moral. L'éthique des vertus stipule que chaque individu a des dispositions à bien agir, des vertus, et qu'il se doit de les cultiver. Le psychologue est en droit d'en déduire que, si de telles dispositions existent, alors elles doivent se traduire par une certaine constance dans les comportements, par l'existence d'un « caractère » suffisamment stable et durable qui caractérise l'individu et peut comporter ces vertus. Il peut alors construire une démarche expérimentale pour tenter de trouver des traces significatives de cette constance, de ce caractère. S'il échoue, et que chaque individu semble agir non en fonction de supposées dispositions constantes, ce caractère supposé, mais plutôt en fonction

10. Pour une introduction au réalisme scientifique voir (Chakravartty 2011) [40]
11. L'ouvrage de Dawkins instruisant la non existence du Dieu des religions est un bon exemple de cet exercice (Dawkins 2018) [64]

des circonstances, alors, le psychologue peut émettre un doute sur l'existence du caractère et des vertus et, par voie de conséquence, sur la pertinence de l'éthique des vertus. Mais deux difficultés séparent le constat expérimental de cette conclusion. La première, comme je l'ai déjà signalé, est que la vérité absolue n'est pas atteinte par le psychologue et que son résultat peut toujours donner lieu à des doutes que pourra exploiter le philosophe moral. Par exemple, le défenseur de l'éthique des vertus conclura simplement que les personnes ayant participé à l'étude n'ont pas les vertus en question, ce qui n'est pas très étonnant puisque les saints sont rares, et que, donc, l'expérience ne prouve rien. La seconde porte sur le mode de justification de la théorie, ici l'éthique des vertus, qui relève, pour le philosophe moral, du domaine de la spéculation et non du domaine de ce que l'on peut observer. En effet, l'absence de vertu constatée, et même l'absence de constance dans le comportement, donne pour lui une image de l'ampleur des progrès moraux à accomplir, sans avoir une quelconque incidence sur la validité de sa théorie.

En d'autres termes, la démarche scientifique expérimentale telle qu'elle apparaît au travers de la métaphore de l'hélice ne peut être qu'instrumentale. Une proposition scientifiquement établie à un instant t, elle, peut être dite aussi vraie et réaliste qu'on peut (scientifiquement) l'être. Ceci ne sera probablement pas du goût des philosophes qui cherchent des réponses absolues et définitives [12] mais on peut avancer que c'est assez proche de la position majoritaire chez les scientifiques, comme le défend Arthur Fine dans son article «The Natural Ontological Attitude.» (Boyd 1991, p 261-277) [30].

3.2.4 Limites de la métaphore

Pour poursuivre cette présentation de la métaphore de l'hélice, soulignons-en plusieurs limites importantes. La métaphore laisse inabordées les questions sur la construction du bateau et l'invention de l'hélice. Une analyse plus complète de la démarche scientifique ne pourra faire ainsi l'économie d'expliquer comment débute ce phénomène si important pour l'espèce humaine. Cette ignorance des origines de la démarche scientifique a toutefois peu d'incidences pour l'objet de la présente thèse en relation avec la psychologie scientifique, car celle-ci est apparue bien après que la démarche scientifique se soit structurée et ait porté ses fruits dans de nombreux autres domaines. La psychologie pourra être tiraillée entre la spécificité de ses méthodes, liée au domaine très particulier de l'étude des humains, et l'adoption de méthodes générales utilisées par les autres sciences, mais ne pourra certainement pas ignorer ces dernières qui constituent un point de référence reconnu de la démarche scienti-

12. voir par exemple (Mcarthur 2006) [160]

fique.

La métaphore ne permet pas non plus de rendre compte des évolutions des différentes disciplines qui peuvent se séparer, se rejoindre, dépérir ou fusionner dans un ensemble plus vaste. Elle laisse donc inabordées des questions importantes comme celle du lien entre disciplines proches, celle de la possible réduction de disciplines spéciales à des disciplines dites fondamentales, et celles liées au projet global de convergence vers une connaissance scientifique unifiée. Là encore, cette faiblesse de la métaphore aura peu d'incidences sur mes propos qui se situent à un moment du développement des sciences naturelles en général et des sciences de l'homme en particulier qui est loin de rendre urgente une telle préoccupation.

Sur un plan proche du précédent, la métaphore laisse également de côté des questions ne portant pas directement sur la démarche scientifique expérimentale mais, plus largement sur la Science ou les sciences (avec ou sans majuscule et au singulier ou au pluriel). Une exemple d'une telle question serait de savoir s'il existe des sciences non expérimentales, les mathématiques étant un candidat évident, et quelles peuvent être les démarches de ces sciences non expérimentales et leurs rapports à la démarche expérimentale. J'aurai à revenir sur ce point puisque certains philosophes, comme Timothy Williamson (Williamson 2007) [235] proposent un parallèle entre l'apport des mathématiques aux sciences naturelles et l'apport de la philosophie à ces mêmes sciences.

Enfin, dernière limite que je souhaite souligner, la métaphore reste floue quant à la qualification des différentes activités qui en constituent le noyau. Quelles sont précisément ces activités qui permettent aux scientifiques de faire tourner l'hélice ? Faut-il rejoindre Feyerabend (Feyerabend 1988) [85] et accepter tout ce qui peut faire avancer ? Faut-il donner un rôle particulier aux activités appuyées sur des structures mathématisées ? Ces questions ne sont pas éclairées par la métaphore de l'hélice mais ont une certaine importance pour le développement de la psychologie scientifique, et j'aurai à revenir sur ce point dans les chapitres suivants. Pour l'instant soulignons simplement ce qui qualifie particulièrement les activités expérimentales dans le domaine de la psychologie.

3.2.5 La métaphore de l'hélice et la psychologie

Il est commun de distinguer depuis Claude Bernard trois types d'approches des phénomènes psychologiques. L'observation subjective, ou introspection, est ce que chacun peut exercer sur sa propre personne. L'observation objective des comportements par un tiers est à la base de la psychologie, et satisfait aux contraintes posées par les phénoménologues d'évi-

ter de faire appel aux états mentaux et à l'introspection. Enfin, la méthode expérimentale consiste pour Claude Bernard à faire varier des paramètres de la situation et à observer les changements induits sur le phénomène étudié. On peut, en prolongement, distinguer différents cas dans le cadre de cette méthode expérimentale selon la maîtrise qu'on peut avoir de la situation et de l'apparition du phénomène. Au niveau le plus bas, on aura de simples observations différentielles constatant, sans intervention particulière, ce qui se passe dans différentes circonstances. A un niveau intermédiaire, on pourra avoir des interventions directes fixant certains des paramètres et en explorant les effets. Enfin, au niveau le plus élaboré, le scientifique aura la possibilité de réaliser des expériences hautement contrôlées ne faisant varier qu'un seul des paramètres au sein d'une situation parfaitement maîtrisée et obtenant à volonté les effets désirés. Le scientifique essaye alors de se rapprocher de l'idéal de la réalisation d'une manipulation ciblée sur un seul élément, « toutes choses étant égales par ailleurs » qui est le Graal de l'expérimentateur. [13]

En prolongement de la métaphore de l'hélice, je propose dans ce contexte de la psychologie scientifique d'appeler « démarche expérimentale » et, pour faire court, expérimentation, une situation qui répond à trois conditions :

— 1 – Chercher à observer ou obtenir quelque chose de prédéfini (un phénomène, une mesure, . . .),

— 2 – Constituer un terrain d'expérimentation externe à l'expérimentateur correspondant à la recherche,

— 3 – Obtenir des traces partageables de cette expérimentation.

Cette définition laisse largement ouvertes les possibilités d'expérimentation en les inscrivant toutefois dans la contrainte des trois refus suggérés par la métaphore, les excès de dogmatisme, de relativisme, et de pragmatisme. Bien qu'ouverte, cette définition n'en pose pas moins quelques limites auxquelles j'aurai à me frotter dans les études de cas des chapitres suivants. A titre d'exemples, citons des activités qui sont en délicatesse avec chacune de ces trois conditions, sans pourtant être absolument à exclure. Lancer une expérimentation sans savoir ce que l'on recherche, et sans avoir exprimé au préalable cet objet de la recherche, présente le risque de n'avoir pas préparé le terrain de l'expérimentation en fonction d'une attente particulière et donc de se trouver ensuite dans l'impossibilité de l'interpréter. Mais

13. Par opposition au terme d'expérience, le terme d'expérimentation est souvent réservé dans la littérature de philosophie des sciences à ce niveau plus élevé de maîtrise. Je ne retiendrai pas cette distinction comme très utile en psychologie dans la mesure où une des questions difficiles en psychologie expérimentale est justement d'évaluer pour chaque cas concret le niveau de maîtrise que l'expérimentateur peut avoir sur l'ensemble des éléments de contexte influant sur les phénomènes étudiés. Mener des expériences ou expérimentations de psychologie exige donc toujours de préciser tous les éléments détaillés des protocoles permettant de mettre en doute pas à pas les résultats sans pouvoir renvoyer, comme dans les sciences plus dures, vers des corpus considérés acquis.

l'exploration naïve systématisée est également une étape incontournable quand tout manque, et qu'on n'a ni traces, ni théories, ni expérimentations préalables pour construire son projet de recherche. La deuxième condition, construire un terrain d'expérimentation externe dont le chercheur vise à être distancié, est évidemment violée par l'introspection lorsque le chercheur la pratique sur lui-même. Mais, en revanche, le chercheur peut observer des rapports d'introspection de tiers et en tirer des traces qui prennent alors une dimension objective. Enfin, dernier exemple portant sur la troisième condition, il a souvent été souligné par les philosophes de l'esprit que l'expérience phénoménale (l'effet que cela fait de voir du rouge) ne pouvait être partagée, et qu'il ne peut donc y avoir de trace partageable permettant l'expérimentation dans ce domaine. Si tel est vraiment le cas, l'étude de l'expérience phénoménale serait inaccessible à la démarche scientifique expérimentale [14].

Chercher à intégrer une de ces trois activités dans une démarche scientifique expérimentale supposera à chaque fois une réflexion spécifique pour construire des premières itérations épistémiques qui puissent servir de base de départ à la dynamique scientifique. Les premiers tours d'hélice, comme toute personne ayant démarré un bateau à moteur a pu en faire l'expérience, donnent l'impression de remuer l'eau sans faire avancer le bateau, et ce n'est qu'au bout d'un temps d'incertitude transitoire que le flux s'organise en une certaine dynamique et que le bateau avance. Peut-être la métaphore de l'hélice nous offre-t-elle ici une ressource pour imager ce qui se passe quand des activités humaines comme l'introspection ou l'analyse de la conscience phénoménale s'incorporent progressivement dans ce qui peut devenir une dynamique scientifique expérimentale.

Résumons en trois points l'image de la démarche scientifique expérimentale que porte la métaphore de l'hélice expérimentale :

— La démarche scientifique expérimentale décrit un moyen d'avancer dans l'exploration du monde. La métaphore de l'hélice permet d'en souligner le caractère dynamique.
— La démarche scientifique n'est pas la seule source de nos croyances. Des options métaphysiques sur le vrai ou le réel peuvent (et peut-être doivent) se développer indépendamment d'elle. Elles seront supportées par des arguments qui, s'ils visent à être absolus et définitifs, ne pourront être apportés par la démarche scientifique.
— C'est bien cette démarche-là, épistémologiquement efficace et métaphysiquement limitée, dont je cherche à analyser l'apport aux questions morales.

De façon à préciser ce qui sera particulièrement significatif dans cette démarche dans les

14. Pour une étude détaillée récente de la conscience phénoménale, qui en propose une conception illusionniste qui expliquerait en quoi ce phénomène n'est pas partageable, simplement parce qu'il n'existe pas tel qu'il nous apparaît, voir (Kammerer 2019) [128]

chapitres suivants en regard de la philosophie morale, je vais poursuivre encore l'approfondissement de la métaphore de l'hélice expérimentale en présentant ci-après les trois temps qui la font tourner de théorie en expérimentation, d'expérimentation en traces et de traces en théorie : l'opérationnalisation, l'objectivation et l'interprétation inductive. Je poursuivrai à la section suivante avec l'examen du problème de la réplication, question proche mais différente de celle de l'itération épistémique imagée par l'hélice.

3.3 Trois temps de l'hélice expérimentale

Comment ce que nous disons du monde, ce que nous en pensons, se rattache-t-il aux objets de ce monde ? Ce que propose la démarche scientifique expérimentale, et illustre la métaphore de l'hélice, c'est de donner à cette question une réponse dynamique en trois temps. Nous partons d'une situation réelle, ici et maintenant, et nous tenons effectivement des propos sur cette situation et sur ce monde, mais ces propos ne nous satisfont pas. Ils soulèvent de nombreuses questions et, pour tenter d'y répondre, nous recherchons des situations qui nous donneraient à voir plus, ou mieux, ce monde. Premier temps, c'est de l'opérationnalisation que sont attendues des pistes de situations possibles. Une personne, ayant concrétisé, en tout ou partie, ce qui lui semble être une de ces situations, a recueilli des éléments d'observations. Comment traduire ces résultats ponctuels, circonstanciés, en des traces utilisables par tous en regard des questions posées ? C'est de la phase d'objectivation, deuxième temps, dont on attend cela. Nous disposons ainsi d'une base de résultats, mais quelles conséquences peut-on en tirer sur les propos insatisfaisants que nous tenions. C'est l'étape d'interprétation inductive, troisième temps, qui va nous permettre de reformuler nos propos de façon un peu plus, mais jamais totalement, satisfaisante. Et le cycle pourra alors recommencer. Ces trois temps de l'opérationnalisation, l'objectivation et l'interprétation inductive, suggérés par la métaphore de l'hélice, ne sont certainement pas séparés de façon étanche dans la réalité de l'activité scientifique, il convient de les voir comme un outil d'analyse de cette activité très particulière, ainsi que comme un outil d'analyse de l'inscription de cette activité scientifique expérimentale dans l'ensemble des questions posées par une autre activité, et, pour ce qui me concerne ici, dans les questions que pourraient poser les philosophes moraux aux psychologues.

3.3.1 L'opérationnalisation

La métaphore de l'hélice permet de situer graphiquement trois temps de la démarche scientifique expérimentale que j'utiliserai dans les études de cas des chapitres suivants. L'opérationnalisation est décrite ci-dessous, l'objectivation et l'interprétation inductive feront l'objet des sections suivantes. L'opérationnalisation consiste à rechercher comment mettre en relation une hypothèse théorique avec un dispositif pratique opérationnel dans l'objectif principal de tester expérimentalement cette hypothèse. La théorie postule des entités, des propriétés et des relations entre ces entités et ces propriétés. Les propriétés et relations peuvent être simplement qualitatives ou donner lieu à des quantifications élaborées, qui comportent des mesures et la mathématisation des relations. Une hypothèse théorique peut porter sur chacune de ces postulats et élaborations.

De façon générale, une hypothèse théorique en psychologie sera de la forme « Dans le contexte C, si l'événement S a lieu, alors les effets E auront lieu » [15]. L'objectif de l'opérationnalisation est de définir un dispositif expérimental, ou observationnel, qui mette en relation le contexte théorique C avec un contexte opérationnel Co, des événements théoriques S et E avec des événements observables So et Eo tels que dans le contexte Co, si l'on observe So on peut vérifier que l'on a bien Eo. La métaphore de l'hélice suggère que l'opérationnalisation est l'étape qui permet la transition entre les théories et les objets permettant l'expérimentation.

Par exemple, nous pouvons émettre l'hypothèse psychologique qu'un individu confronté à une autorité qu'il considère légitime, pourrait obéir à un ordre au-delà de ce qu'il ferait hors de ce contexte, allant même jusqu'à mettre en péril la vie d'autrui. Stanley Milgram a imaginé un dispositif expérimental mettant des participants en situation de contribuer à des expériences scientifiques sous l'autorité d'un professeur qui leur demandait de soumettre des comparses à des chocs électriques s'ils se trompaient [16]. Le contexte d'autorité, C, est ici opérationnalisé par la prétendue étude scientifique et l'autorité supposée d'un professeur sur le participant, constituant le contexte opérationnalisé Co, la sollicitation du participant S est mise en scène par le professeur et les comparses avec de prétendues réponses erronées du comparse à des questions, c'est le stimuli opérationnalisé So, et le niveau d'adhésion à l'autorité du participant E est opérationnalisé par l'intensité de la décharge électrique qu'il accepte d'imposer au comparse qui simule la douleur, Eo.

Évaluer la validité d'une telle opérationnalisation, et donc la possibilité pratique de réa-

15. Je reprends ici l'approche de l'opérationnalisation telle qu'elle est formulée dans les manuels de psychologie, voir par exemple (Ghiglione 2007) [101]

16. Un classique de la psychologie du comportement des années 60 : Stanley Milgram, Behavioral study of obedience, Journal of Abnormal and Social Psychology, 1963, Vol. 67, pp. 371–378 [164]

liser une expérimentation qui soit en rapport avec l'hypothèse théorique de départ, est un exercice essentiel pour la démarche scientifique expérimentale mais outrageusement difficile en psychologie. Je souhaite montrer ces deux aspects de l'opérationnalisation dans les chapitres suivants en m'appuyant sur plusieurs études de cas. Je reprendrai ensuite, au chapitre sur l'opérationnalisation, l'analyse détaillée de ce processus dont je pense qu'il abrite une part importante des difficultés que la démarche expérimentale affronte en psychologie en général, et en regard de la philosophie morale en particulier.

Lorsque l'opérationnalisation n'est pas complète, lorsqu'elle n'ouvre pas vers une possibilité pratique de mise en œuvre mais décrit simplement un contexte et des événements habituels qui semblent proches des hypothèses théoriques, on parle d'expérience de pensées [17]. Ces expériences qui n'en sont pas au sens de l'hélice expérimentale, puisque la confrontation au monde n'a pas lieu, sont un outil à disposition des théoriciens pour donner à voir les conséquences que pourrait, en principe, avoir une théorie trop abstraite pour être facile à transmettre. Ainsi, et dès l'antiquité, Achille et sa tortue on mis l'accent sur les difficultés liées aux différentes conceptions théoriques du continu et de l'infini, plus récemment les jumeaux de Langevin ont joué ce rôle pour la relativité, et le chat de Schrödinger l'a joué pour la physique quantique. Les expériences de pensée sont utiles au sens où elles rendent possible la communication de théories abstraites non intuitives, au sens où elles peuvent donner des idées pour de futures expériences et, principalement quand on les observe du point de vue de l'opérationnalisation complète inaboutie, au sens où elles mettent l'accent sur les difficultés qu'il faut affronter pour trouver un chemin vers de possibles mises en pratique d'expériences significatives en regard des théories d'un haut niveau d'abstraction.

L'opérationnalisation est une étape difficile de la démarche scientifique expérimentale et les chercheurs ont eu, à plusieurs reprises, la tentation de passer outre en tentant plusieurs stratégies que je détaillerai au chapitre sur l'opérationnalisation. J'en présenterai rapidement trois. La première sera l'opérationnalisme de Bridgman, la seconde le behaviorisme, et la troisième celle des scientifiques et techniciens confrontés à une entité psychologique. L'opérationnalisme de Bridgman consiste à chercher à définir les entités théoriques non plus par une stipulation théorique, mais en partant des opérations matérielles qui permettent de les manipuler. Il s'agit, par métaphore, d'inverser le sens de rotation de l'hélice expérimentale en la faisant remonter directement des objets vers les théories. Ce serait ainsi le thermomètre, la constitution matérielle d'une graduation portée sur un tube en verre partiellement rem-

17. La thèse de 2016 de Rawad El Skaf [79] propose une analyse détaillée du rôle des expériences de pensée dans la démarche scientifique

pli de mercure, qui serait la meilleure définition de la température. Nous verrons que cette théorie est maintenant abandonnée pour les sciences naturelles mais a encore des échos en psychologie.

Ces échos résonnent, au moins du point de vue de leur période commune de développement dans les années 1920, avec le behaviorisme qui consiste à refuser l'existence des états mentaux et à n'accepter comme valides que la description des comportements observables par un tiers. Cette approche avait pour principale justification d'exclure ainsi toute utilisation de l'introspection, avec son lot d'erreurs et de biais, en tant qu'outil pouvant valider une opérationnalisation et ainsi une expérimentation. Nous verrons les difficultés rencontrées par cette démarche en regard des méthodes qui donnent aujourd'hui à voir l'intérieur des processus intracrâniens.

La troisième stratégie que je présenterai au chapitre sur l'opérationnalisation est celle des scientifiques et techniciens confrontés à un problème de psychologie. Elle consiste à contourner la spécificité des entités théoriques psychologiques dans un premier temps, à s'appuyer pour cela sur des entités techniques qu'ils maîtrisent, et ce en espérant que, peu à peu, se construiront à la fois les entités théoriques psychologiques et leur opérationnalisation.

Enfin, je pense utile d'évoquer ici une autre tentation que je ne détaillerai pas au chapitre opérationnalisation et qui consiste, lorsque les théories sont jugées comme bien établies, à exploiter les possibilités offertes par les simulations numériques pour passer directement des théories aux traces, en passant par dessus l'étape de l'expérimentation et, ainsi, n'avoir pas à se confronter aux difficultés de l'opérationnalisation et aux rigueurs et imprévus de la sanction expérimentale. La simulation numérique fournirait directement les résultats et les traces partageables qu'on pourrait atteindre par expérimentation dans les conditions simulées. Comme le souligne Julie Jebeile dans (Jebeile 2019) [124], une telle simulation présente de nombreux intérêts, dont ceux d'être plus facile à réaliser qu'une expérimentation et, également, celui de permettre d'alimenter les inférences visant à l'interprétation inductive des résultats. Mais la simulation n'a pas le même statut épistémique qu'une expérimentation puisque, comme l'expérience de pensée, elle ne comporte pas de confrontation au monde réel, non virtuel. Il s'agit ici, pour la métaphore de l'hélice, non plus d'une inversion du sens de rotation, comme dans le cas de l'opérationnalisme de Bridgman, mais au contraire d'une accélération qui vise à passer directement des théories aux résultats, au prix d'une baisse de la valeur épistémique de la démarche.

Ces stratégies permettent de ne pas affronter l'étape difficile de l'opérationnalisation, mais aucune ne permet d'avoir à la fois l'apport de la théorie et celui de l'expérimentation

de façon équilibrée, ce que la métaphore de l'hélice expérimentale vise à imager comme étant un des aspects clé de la démarche scientifique expérimentale. Je les écarte pour l'instant pour passer à l'examen du deuxième temps de la démarche scientifique expérimentale telle qu'elle est métaphoriquement représentée par l'hélice : l'objectivation des observations faites lors d'une expérimentation.

3.3.2 L'objectivation

J'appelle ici « objectivation » l'activité qui consiste à élaborer et diffuser les résultats d'une expérience particulière, réalisée par une équipe en un certain lieu en un certain temps, dans le but de les rendre disponibles à la communauté des scientifiques qui contribuent à la démarche scientifique expérimentale. Il s'agit donc principalement de construire les « traces » en explicitant, au mieux, le déroulement de l'expérience, de les préserver, de les communiquer, de les contextualiser pour qu'une équipe souhaitant soit interpréter ces traces soit refaire l'expérience ait, au mieux, tous les éléments pour cela. La métaphore de l'hélice expérimentale nous permet, là encore, d'écarter le débat sur « l'objectivité scientifique » [18] en le remplaçant par cette étape d'objectivation des résultats. Étape qui est menée « au mieux », ce qui signifie ici « en s'appuyant sur toutes les itérations épistémiques antérieures, et dans la limite de ce qu'ont construit ces itérations épistémiques antérieures ».

Cette proposition lie l'étape de l'objectivation à une itération épistémique particulière, ce qui permet en particulier de rendre compte de l'évolution dans le temps de l'évaluation de traces qui pouvaient être satisfaisantes à un moment donné et deviennent insuffisantes ensuite. Un exemple classique souvent repris est celui des expériences sur le cholestérol, des traces établies avant que soient distingués le « bon » du « mauvais » cholestérol pouvaient être satisfaisantes à l'époque de leur établissement, mais sont devenues inutilisables après que cette distinction, indispensable à toute analyse des résultats, soit advenue. Les traces étaient un reflet objectif et utile de l'échantillon avant, mais plus après cette distinction.

L'étape d'objectivation est particulièrement délicate pour la science psychologique et donne lieu à de nombreux débats. Citons-en deux, le débat sur les données d'introspection, dites à la première personne, et le débat sur le situationnisme, sur l'impossibilité de décrire valablement un contexte. Lorsqu'une expérience de psychologie s'appuie sur des états mentaux auxquels le sujet de l'expérience accède par introspection, se pose le problème du statut à donner à cette information dite à la première personne. Dans un premier temps, il semble, qu'au

18. Pour une analyse détaillée de l'évolution de la notion d'objectivité scientifique, voir l'excellent et très esthétique ouvrage de Daston et Galison (Daston 2012) [63]

moins dans certains cas, les philosophes aient défendu l'accès immédiat et certain à cette information : si je ressens « j'ai mal » c'est que j'ai mal, et personne ne peut en douter [19]. Si cet accès direct est retenu, alors il n'est guère partageable avec d'autres car un tiers ne peut dire si l'expression « j'ai mal » formulée est effectivement le reflet d'un mal ressenti ou s'il est tentative de manipulation. De multiples possibilités ont été explorées dans ce domaine, et celle proposée par Piccinini dans (Piccinini 2009) [181] consiste à considérer le ressenti lui-même comme intransmissible, et donc à ce titre hors du champ des traces utiles aux itérations épistémiques. En revanche, la déclaration faite à un tiers est, elle, enregistrable « le patient X a dit « j'ai mal », avec le niveau x sur une échelle de 1 à 10, dans telles conditions ». Cet enregistrement constitue alors une trace et peut servir d'amorce à l'analyse des phénomènes étudiés. Il serait ainsi possible de passer d'une information à la première personne à une information à la troisième personne objectivable dans les traces de l'expérience.

Autre débat, les phénomènes psychologiques seraient trop complexes et trop dépendants de tout un ensemble de détails contextuels pour que les conditions de l'expérience puissent être, utilement, objectivées dans des traces. La connaissance de la situation globale serait, en particulier, indispensable à l'établissement d'un jugement moral pertinent, mais serait impossible à transmettre à un tiers, et donc impossible à traduire dans ces traces qui sont un des éléments de la démarche scientifique expérimentale [20]. S'il en allait ainsi, alors la démarche expérimentale serait impossible à entreprendre dans le domaine moral mais, avec elle, c'est toute affirmation générale sur le domaine moral, et donc toute théorie morale qui serait également sans signification. Peu de philosophes moraux seront prêts à payer ce prix élevé et beaucoup préféreront estimer qu'il est possible de dégager des informations moralement pertinentes de ces situations, certes complexes, et, de ce fait, ils ouvriront également la possibilité d'une première itération de description qui servira d'amorce à la démarche expérimentale. Naturellement, cela ne signifie pas que la démarche puisse aller plus loin que cette amorce, et ne se trouve submergée ensuite par l'ampleur de la complexité du comportement humain, mais, en tous cas, commencer à expérimenter, au sens de la mise en route des premiers tours de l'hélice expérimentale, semble possible.

Remarquons pour finir, et avant de passer à l'interprétation inductive, qu'il y a un peu d'arbitraire à situer une question dans l'opérationnalisation, dans l'objectivation ou dans l'interprétation. De la même façon que toute théorie, toute expérience et toute trace s'appuie sur les itérations épistémiques précédentes et contient donc, pour partie, les théories antérieures,

19. Pour une étude détaillée actuelle de cette perception directe de nos état mentaux, et pour sa remise en cause, voir (Kammerer 2019) [128]

20. Pour une présentation du situationnisme, voir l'article de Jérôme Ravat dans (Cova 2013) [55]

les traces et expériences antérieures, de la même façon, opérationnalisations, objectivations et interprétations s'enchaînent et sont instruites des compétences et savoir-faire accumulés et une question non traitée à une étape se rappelle aux scientifiques à l'étape suivante.

3.3.3 L'interprétation inductive

L'interprétation inductive consiste, en s'appuyant sur les traces objectivées et partagées des résultats d'une expérimentation, à proposer les modifications des théories qui permettraient d'en rendre compte. La métaphore de l'hélice suggère que l'interprétation inductive est ce qui permet la transition entre les activités d'enregistrement des traces laissées par l'expérimentation et les activités de construction des théories. L'analyse statistique, prise dans un sens très large, est un des outils principaux mis en œuvre dans cette étape, mais il n'est pas le seul. Dans le contexte particulier évoqué plus haut des simulations numériques, Julie Jebeile [124] démontre l'importance des représentations graphiques pour, littéralement, donner à voir les traces et permettre aux chercheurs de produire les interprétations et inférences pertinentes à partir de ces traces pour instruire les décisions, qu'elles soient à visée pragmatique ou théorique, quand il s'agit d'explorer les sens dans lesquels faire évoluer les théories. [21]

Dans l'exemple de l'expérience de Milgram ci-dessus, le résultat concret laissé par l'expérimentation est constitué du descriptif écrit du protocole expérimental, de la liste des paramètres que l'expérimentateur a fait varier, des tableaux de chiffres figurant les niveaux d'intensité de la décharge électrique imposée par les participants aux comparses, des comptes-rendus écrits des expérimentateurs et des entretiens avec les participants, et enfin de l'ensemble des objets qui ont pu être conservés, outils, relevés intermédiaires, ... L'interprétation inductive consistera, dans un premier temps, à peser le caractère particulier de ces éléments en regard de leur potentiel de généralisation. La situation d'autorité est-elle atypique, ou au contraire proche des situations communes ? La décision d'infliger le courant électrique est-elle représentative de la soumission à l'autorité ? Le niveau létal du courant est-il bien compris comme tel et interprétable comme significatif de la décision de tuer ? ... Puis, dans un second temps, en s'appuyant sur les éléments supposés généralisables, l'interprétation inductive consistera à rechercher en quoi ces comportements appellent une révision des théories décrivant le comportement humain.

Remarquons ici que la métaphore de l'hélice expérimentale permet de différencier les

21. Pour visualiser l'importance de l'exploitation graphique visuelle des traces, comme Julie Jebeile la propose, deux références importantes : (Bertin 2005) [23] et (Tufte 2001) [229].

traces des faits. Les traces s'inscrivent dans la dynamique scientifique et, à l'idéal, comportent tous les éléments qui permettent d'en suivre la genèse, en partant des questions théoriques à l'origine des interrogations, en poursuivant par l'opérationnalisation effectuée, et jusqu'aux expériences menées, aux résultats et aux rapports des membres des équipes ayant contribué à ces travaux. Les faits, en revanche, ne s'inscrivent pas dans une démarche particulière et leur définition dépend de ce que chaque philosophe fait de ce concept. J'éviterai donc l'appellation « faits scientifiques » qui me semble ambiguë entre ces deux notions de traces et de faits et garderait la terminologie de « trace » pour les traces scientifiques, terminologie qui, étymologiquement, marque mieux leur place : les traces sont ce qui reste suite au passage des scientifiques. [22]

Pour exploiter ces traces laissées par l'expérimentation, l'interprétation va consister à faire le chemin inverse de celui de l'opérationnalisation. Pour chacun des éléments du contexte expérimental, pour chaque résultat partiel et définitif, il convient d'imaginer de quelles parties des énoncés théoriques en jeu on peut le rapprocher. On peut comprendre cette étape comme une tentative de traduction dans le vocabulaire de la théorie de chacun des éléments des traces laissées par l'expérimentation. Il s'agit à la fois d'une interprétation et d'une induction. Une interprétation des traces pour leur donner un sens dans le domaine théorique et une induction car tout énoncé théorique a une vocation générale à s'appliquer au-delà des seuls cas particuliers expérimentés.

La littérature de psychologie scientifique montre à l'envi que cette étape d'interprétation est le lieu principal de mise en cause des résultats présentés comme acquis et je le montrerai là encore sur la base des études de cas. Je défendrai en particulier deux résultats que je pense pouvoir tirer de ces études de cas. Le premier est que l'opérationnalisation et l'interprétation inductive sont deux faces de la même médaille. On peut parler de dualité, au sens où de nombreuses questions peuvent être formulées soit dans le vocabulaire de l'opérationnalisation soit dans celui de l'interprétation inductive de façon superficiellement différente mais profondément semblable. Ce résultat est suggéré par la métaphore de l'hélice et je le défendrai en m'appuyant sur un cas tiré de mon expérience personnelle de chercheur en acoustique.

Le second est qu'il est regrettable que, dans les cas tirés de la psychologie scientifique que je présenterai, autant d'attention soit portée à l'interprétation et aussi peu à l'opérationnalisation. L'explication de cette dissymétrie reste à faire, mais on peut avancer qu'elle est liée

22. La citation souvent reprise d'Henri Poincaré « On fait la science avec des faits, comme on fait une maison avec des pierres, mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison » (La science et l'hypothèse, H Poincaré, 1908) reprend cette distinction entre des faits bruts, assimilables à des pierres qu'on ne peut qu'entasser, et des faits de sciences, que j'ai appelé traces, dont on fait les maisons scientifiques.

à la facilité d'argumentation : si je ne suis pas d'accord avec les conclusions d'un article qui s'appuie sur des résultats expérimentaux, il est plus facile de nier le lien entre les traces et la théorie et d'imaginer pour cela des contre-exemples ad hoc (souvent sous la forme d'une expérience de pensée) que d'analyser en profondeur le dispositif expérimental qui a conduit à ces résultats et de rentrer dans la construction d'un dispositif alternatif.

Après ce bref survol des trois temps que suggère la métaphore de l'hélice expérimentale, je me propose de détailler les attentes que l'utilisation de la démarche scientifique expérimentale a pu contribuer à généraliser en regard de l'expérimentation : les rôles qui lui sont attribués au sein des sciences naturelles ayant adopté cette démarche. Mon ambition dans cette section est de décrire ces rôles de façon à pouvoir les comparer à ce qui pourrait être attendu dans le domaine de la connaissance du comportement humain utile à instruire les questions de philosophie morale.

3.4 L'expérimentation dans les sciences de la nature

L'épistémologie de l'expérimentation dans les sciences de la nature a fait l'objet de développements récents importants avec par exemple les travaux de Ian Hacking , Peter Galison, Bas Van Fraassen et Allan Franklin [23]. Je me propose d'esquisser sur cette base les multiples rôles que l'expérimentation joue dans la démarche des scientifiques faite, pour reprendre la formule de Hacking, d'idées, d'objets et de traces. J'ai retenu dix axes. Les quatre premiers se situent à l'articulation entre la théorie et l'expérimentation, deux sont, sans surprise, de conforter la validité d'une théorie (axe 1) puis, inversement de la réfuter (axe 2), et les deux autres affinent l'apport de l'expérimentation à la théorie en confirmant ou infirmant l'existence d'une entité (axe 3), puis en précisant la forme d'une relation théorique (axe 4). Les trois axes suivants marquent l'autonomie du domaine expérimental en regard de la théorie : développer des savoir-faire opérationnels (axe 5), étendre la recherche (axe 6), préparer la sérendipité (axe 7). Enfin, les trois derniers axes sont centrés sur la contribution méthodologique particulière de l'expérimentation au sein de la démarche scientifique : détecter les sources d'erreurs (axe 8), multiplier les variantes (axe 9), contribuer aux itérations épistémiques (axe 10).

La question du rôle de l'expérimentation dans les sciences ne sera certainement pas épuisée avec ces dix axes et les formulations retenues ne sont que trop grossières en regard des sources citées plus haut. Ces dix axes constituent néanmoins une esquisse utile pour une

23. La liste qui suit est en particulier appuyée sur les références suivantes : (Hacking 1983) [115], (Galison 1987) [94], (Van Fraassen 2010) [231], (Allan Franklin et Perovic 2016) [92], (Allan D. Franklin 1981) [91].

première approche du point de vue de l'épistémologie des sciences naturelles sur le rôle de l'expérimentation et pour servir de base à une comparaison, proposée ci-dessous, avec la philosophie morale expérimentale.

3.4.1 L'expérimentation en appui du développement des théories

Axe 1 : Conforter la confiance dans une théorie

Le premier rôle que l'on peut attendre d'une expérience est de conforter la confiance que l'on accorde à une théorie. Exemple classique, la théorie de la relativité qui prévoyait la déviation de la lumière passant près d'une masse importante a été confortée par la mesure de la déviation de la lumière des étoiles proches du disque solaire à l'occasion d'une éclipse en 1919. Une limitation forte de ce rôle a été soulignée par Duhem puis Quine ; il n'est pas possible qu'une expérience, ou même un ensemble d'expériences, établisse la vérité définitive d'une théorie. Il reste toujours plusieurs théories possibles. Cette thèse dite de sous-détermination des théories par les expériences affirme qu'il est toujours possible de trouver plusieurs théories différentes qui sont satisfaisantes en regard des observations disponibles. Le théoricien qui souhaite néanmoins choisir entre théories empiriquement équivalentes, aura à définir d'autres critères, outre l'adéquation aux résultats expérimentaux, et ce peut être par exemple la théorie la plus simple, celle qui mobilise le moins d'entités théoriques (principe du rasoir d'Ockham) ou la plus explicative (principe d'inférence à la meilleure explication) ou tout autre critère (non empirique naturellement, puisque les théories en compétition sont supposées empiriquement équivalentes).

Axe 2 : Réfuter une théorie

Karl Popper (Popper 1934) [184] a proposé de définir le rôle de l'expérience dans la démarche scientifique comme, essentiellement, d'offrir la possibilité de réfuter une théorie. S'il n'est pas possible qu'une expérience valide définitivement une théorie, il semble qu'elle puisse en revanche l'invalider définitivement si l'observation contredit les prévisions théoriques. Comme pour la validation des théories, la portée de la réfutation a été l'objet de discussions : une expérience dans un contexte particulier ne peut à elle seule réfuter une théorie car, d'une part, les scientifiques n'abandonnent pas aisément une théorie qui fonctionne bien dans un grand nombre de cas et, d'autre part, il est souvent possible de trouver des modifications mineures de la théorie ou de la prise en compte des particularités du contexte qui rendent compte des résultats d'une expérience discordante. La mise en doute des instruments de mesure, soit parce qu'ils dysfonctionnent soit parce qu'ils s'appuient sur des théories antérieures

non pertinentes, est ainsi souvent mobilisée pour minorer la portée d'une expérience aux résultats discordants.

Une autre raison, plus large, qui pousse à relativiser l'importance de la réfutabilité de Karl Popper est liée à l'objectif même que l'analyse de ce philosophe poursuit : donner des critères clairs permettant de distinguer les théories scientifiques de celles qui ne le sont pas. Or on peut défendre aujourd'hui, comme le fait par exemple Daniel Andler dans (Andler 2013),[2], qu'il existe une multiplicité d'approches scientifiques, et qu'un tel critère de démarcation n'existe simplement pas. Ni la réfutation, ni même la réfutabilité, ne peuvent donc être retenus comme critères absolus ni de scientificité ni d'invalidité théorique.

Même s'il convient de relativiser la réfutabilité en tant que critère essentiel de la scientificité d'une théorie, on peut en revanche, et en prolongement des travaux de Karl Popper, affirmer qu'une théorie qui serait par construction non réfutable ne pourrait que très difficilement être qualifiée de scientifique.

Axe 3 : Confirmer ou infirmer l'existence d'une entité théorique

Les deux paragraphes précédents envisagent les théories comme des touts qui sont soit confirmés soit réfutés grâce à l'approche expérimentale. Ce troisième, plus modestement, vise à examiner si l'expérimentation permet de conforter la confiance dans l'existence d'une entité stipulée par la théorie mais non directement observable. Par exemple les électrons ne sont pas directement observables, mais la physique en a postulé l'existence. Des phénomènes liés à ces entités peuvent être décrits par la théorie et l'expérimentateur se donnera pour objectif de les observer, de façon à confirmer ainsi indirectement l'existence des entités prédites. A cet apport de premier ordre de l'expérimentation en tant que vérification de phénomènes donnant confiance en l'existence d'une entité théorique, Ian Hacking (Hacking 1983) [115] a défendu la thèse de second ordre que c'est en manipulant ces phénomènes, en mettant au point des instruments permettant de créer, modifier, supprimer, ces phénomènes, que les scientifiques acquièrent intimement la certitude opérationnelle [24] de l'existence d'entités stipulées par les théories.

Axe 4 : Préciser la forme (souvent mathématique) d'une relation

La théorie peut postuler une relation entre plusieurs entités mais laisser à l'expérimentateur la tâche d'en trouver la forme exacte. Ainsi, par exemple, le psychologue peut proposer que la perception de l'intensité sonore soit fonction de l'énergie acoustique de l'onde sonore, mais la forme exponentielle de cette relation, qui conduit à mesurer le niveau sonore perçu

24. C'est cette certitude opérationnelle qui, indépendamment de toute option métaphysique, leur permet, individuellement et en groupe, de manipuler ces nouveaux objets comme ils le feraient d'objets du quotidien immédiatement accessibles à la perception et à l'action.

par un logarithme de l'énergie en jeu, est le résultat des expériences de psychométrie.

Et la philosophie morale expérimentale ?

Les quatre axes ci-dessus s'appuient sur une démarche en trois temps : une théorie est présupposée, sa capacité prédictive est suffisante pour que puissent en être déduits des phénomènes observables, l'expérimentation consiste à tenter de mettre en œuvre les phénomènes prédits dans des contextes proches des conditions prévues par la théorie et à vérifier si les phénomènes prédits adviennent. Chacun de ces trois temps donne lieu à des questions particulières que doit affronter le philosophe moral expérimental. Les théories morales sont-elles assez puissantes pour être prédictives ? Assez précises pour que des contextes opérationnels puissent être élaborés ? Elles sont certainement plus statistiques que déterministes, mais alors comment fixer les seuils significatifs pour leur validation ? Les philosophes moraux qui se préoccupent de la vie bonne, des critères du bien ou du juste, ou d'autres sujets de ce haut niveau d'abstraction proposent des théories dont une vérification ou une réfutation globale ne serait guère envisageable. A l'extrême, certaines théories morales, comme par exemple le situationnisme [25], sont, par leur formulation même, hors d'atteinte de toute approche empirique. En revanche, l'objectif plus modeste de conforter ou infirmer l'existence d'une entité théorique ou la forme d'une relation est l'objectif même des approches que nous verrons plus loin. Beaucoup visent à corréler des comportements moraux, actes ou jugements, avec des traits psychologiques dont l'existence est postulée par les théories morales. Cette corrélation est alors interprétée comme un indice de validation à la fois de l'existence réelle de ces traits psychologiques et, à la fois, de la pertinence de l'opérationnalisation de ces traits, entités théoriques, par le comportement détecté. Faire la part, provisoire et révisable, des dissensus moraux qui relèvent ou non d'une possible approche empirique de ce type est une tâche difficile qui reste à entreprendre.

3.4.2 Le développement autonome du domaine expérimental

Les quatre premiers axes que j'ai évoqués ci-dessus, confirmation et infirmation d'une théorie, confirmation d'une entité et précision de la forme d'une relation, semblent donner à l'expérience un rôle auxiliaire de la recherche théorique. Hacking a souligné que pour les sciences naturelles, il n'en est rien : l'activité expérimentale est, pour une part importante, autonome. Précisons les trois axes liés à cette autonomie.

Axe 5 : Développer un savoir-faire opérationnel

25. Pour mémoire, le situationnisme est la thèse selon laquelle les jugements moraux dépendent de l'ensemble d'une situation et ne sont pas réductibles à l'application d'une règle morale en partant de quelques traits saillants de cette situation. Tout détail compte.

Pour que l'observation ou l'expérimentation puissent avoir lieu, il est nécessaire de bâtir des savoir-faire opérationnels, techniques, matériels et organisationnels qui évoluent de façon à pouvoir créer (ou observer) les phénomènes, à pouvoir agir sur (ou prendre en compte) les paramètres qui peuvent influer sur ces phénomènes, à améliorer en permanence (précision, spécificité, justesse,. . .) les instruments de détection, de mesure et d'intervention, etc. Un rôle important de l'expérimentation est de donner l'occasion du développement de ces savoir-faire opérationnels qui portent en germe les résultats du futur. Le double sens du mot expérience, à la fois ce qu'on fait pour la première fois et compétence acquise par ce qu'on a déjà fait, trouve ici sa pleine expression.

Axe 6 : Explorer des domaines de paramètres non couverts par les théories et expériences existantes

Un deuxième élément de l'autonomie de l'expérimentation est sa capacité à rechercher systématiquement des phénomènes dans toutes les zones de valeur des paramètres accessibles à l'intervention et à l'instrumentation. Ainsi, par exemple, la physique des particules s'est développée en construisant des appareils de plus en plus gros de façon à explorer des énergies hors des domaines déjà théorisés.

Axe 7 : Préparer la sérendipité

Enfin, lorsque l'expérimentateur sort des sentiers battus et tend son attention vers ce que de nouvelles approches ou de nouveaux instruments vont produire d'inconnu, il arrive qu'il découvre des phénomènes hors de son objectif initial de recherche. C'est ainsi que la radiographie par rayons X ou le four à micro-ondes ont été découverts, par hasard dira-t-on trop rapidement, car ce type d'évènement ne peut advenir qu'après un long travail de préparation construisant le contexte de la découverte.

Et la philosophie morale expérimentale ?

La philosophie expérimentale repose en pratique sur la psychologie expérimentale pour ses méthodes d'investigation. En première analyse, la philosophie expérimentale n'aurait ainsi aucun apport au développement autonome de l'expérimentation, qui serait uniquement du ressort de la psychologie. Cela ferait de la philosophie expérimentale, sur ce plan, un parasite de la psychologie expérimentale. On peut a contrario avancer que, d'une part, les questions posées par la philosophie morale étendent le domaine de la psychologie et, à ce titre, contribuent à son développement y compris sous l'angle expérimental. L'exemple de l'expérience de Libet [26] montre comment des questions philosophiques comme celle du libre

26. Les expériences de Libet ont conduit à constater que l'influx nerveux moteur était émis avant que la personne n'ait conscience du choix à faire, mettant ainsi en cause l'enchaînement rationnel habituel : perception, délibération, décision, action.

arbitre sous-tendent de nombreuses expériences de psychologie en leur fournissant à la fois une justification à l'amont et un immense impact académique à l'aval. On peut également avancer d'autre part, que les interprétations philosophiques croisées des expériences psychologiques sont particulièrement utiles dans des domaines dont la charge idéologique est forte, ce qui peut augmenter le risque de biais idéologique au sein même des équipes de recherche et de leurs processus de financement. Yves Gingras a ainsi souligné les objectifs religieux de certaines études scientifiques financées par la fondation Templeton dans (Gingras 2016) [103], ceci ne constitue bien sûr pas une preuve directe de biais, mais appelle à une vigilance particulière dans l'analyse de la validité de ces études.

3.4.3 La spécificité de l'apport de l'expérimentation.

Outre les rôles d'auxiliaire de la théorie et d'acquisition de savoir-faire opérationnel, l'expérimentation a également un rôle systémique de remise en doute permanente de son propre protocole. En rupture avec l'optimisme indispensable au théoricien qui échafaude des hypothèses, l'expérimentateur a pour objectif d'être plus pessimiste que le plus pessimiste de ses opposants jusqu'à considérer, comme le propose Peter Galison (Galison 1987) [94] qu'une expérience ne soit finie que lorsque toutes les objections imaginables, toutes les interprétations alternatives, ont été empiriquement écartées. [27] Les trois derniers axes vont dans ce sens.

Axe 8 : Détecter et diminuer les sources d'erreur

Une partie essentielle du travail de l'expérimentateur est de détecter les multiples sources des erreurs qui peuvent mettre en cause les résultats de l'expérimentation, que ce soit par l'irruption de paramètres non souhaités (impuretés, imprécision, influences externes, ...), par les erreurs de manipulation, ou par des dysfonctionnements des instruments de détection et de mesure. Je développe plus loin plusieurs études de cas concrétisant ce travail d'exploration des potentielles erreurs du dispositif expérimental dans le domaine de la philosophie expérimentale, et on peut également sur ce sujet rappeler toutes les difficultés de l'éthologie à interpréter les expériences destinées à mesurer l'intelligence des animaux rapportées par Frans de Waal dans (Waal 2016) [67]. Chaque animal a des capacités de perception, d'intelligence et d'action qui diffèrent et pour tester l'intelligence, il faut pouvoir isoler cette capacité centrale de l'intelligence de ses entrées, la perception, et de ses effets, l'action, ce qui n'est guère facile quand on n'observe que le comportement. Ainsi, on a longtemps cru que les éléphants n'avaient pas conscience d'eux mêmes car ils ne réussissaient pas le test qui consiste

27. Dans ce même ouvrage (Galison 1987 p 244) cite un aphorisme attribué à Einstein qui met en lumière cette différence entre théoriciens et expérimentateurs : « Personne ne croît aux théories, sauf leurs auteurs, en revanche tout le monde fait confiance aux expériences, sauf ceux qui les ont conduites. »

pour un individu à se reconnaître dans un miroir avant de s'apercevoir que, en augmentant la taille du miroir, l'éléphant se voyait en entier et alors se reconnaissait. La vue partielle dans un petit miroir ne permettait pas à l'éléphant de se reconnaître. L'échec ne correspondait donc pas à une insuffisance de ses capacités mais à ce détail du dispositif expérimental. L'expérimentateur en traquant ainsi les erreurs du dispositif expérimental est un révélateur des difficultés et des complexités qui ont pu être négligées à la première conception de ce dispositif.

Axe 9 : Définir toutes les interprétations possibles et imaginer des variantes

La thèse de Duhem Quine de la sous-détermination des théories par les expériences a pour contrepartie opérationnelle que le travail de l'expérimentateur doit comprendre d'envisager toutes les interprétations alternatives possibles compatibles avec des résultats et d'imaginer les variantes du dispositif expérimental permettant, au mieux, de donner les arguments empiriques permettant de les évaluer. Là encore, les cas d'études présentés plus loin permettront de souligner l'importance de ce point pour la philosophie expérimentale. Il convient de noter que la thèse de Duhem Quine s'applique de façon particulièrement fréquente dans ce domaine psychologique où manquent les observations intermédiaires précises sur les processus intracrâniens. Avec pour toute observation celles des stimuli à l'amont, celle du comportement à l'aval et quelques indications sur des corrélations avec des mesures physiologiques, de nombreuses théories concurrentes peuvent être empiriquement indiscernables entre-elles.

Axe 10 : Contribuer aux itérations épistémiques

Pour répondre au problème de la coordination [28] entre ce que l'on souhaite mesurer et le dispositif expérimental qui permet d'obtenir cette mesure, Hasok Chang propose ce qu'il dénomme l'itération épistémique. Cette proposition consiste à accepter que nous ne disposons pas de fondement absolu, de vérité première, mais que nous partons de l'état des connaissances existant que nous tentons d'améliorer par un processus itératif portant sur nos théories, expériences et mesures. Par son caractère dynamique, cette notion d'itération épistémique est proche de celle suggérée par la métaphore de l'hélice : une itération épistémique est un tour d'hélice qui a permis de reprendre les traces, les théories et les protocoles et savoir-faire expérimentaux pour affiner la compréhension d'une première mesure. Barwich et Chang ont suggéré en 2015 que ce processus pourrait également être pertinent dans le domaine de la psychologie de la perception (Barwich et Chang 2015) [17]. Le rôle de l'expérimentation est de contribuer à ces itérations épistémiques dans une interaction avec la théorie

28. Comment savoir si une mesure est valide si on ne la compare pas à une mesure déjà faite ? Mais alors, comment cette précédente mesure a-t-elle elle-même été validée ? (Van Fraassen 2010) [231] (p 115)

du phénomène étudié ainsi qu'avec toutes les théories opérationnellement embarquées dans le dispositif expérimental et, en particulier, dans chacun des instruments utilisés pour la détection et la mesure des phénomènes.

Et la philosophie morale expérimentale ?

Les itérations épistémiques et le rôle important de l'expérimentation dans ces itérations, donnent de la science une double image saisie par la métaphore de l'hélice. D'une part celle d'une démarche itérative sans fin et, d'autre part, celle d'un ensemble formé de théories, d'objets et de traces des expériences passées, constitutif de la science à un instant donné, la science des manuels. Cet état de la science réalisé à un moment donné est souvent évoqué avec l'expression « nos meilleures théories scientifiques », mais, en suivant les propositions de ce qui précède, il serait préférable de compléter en « nos meilleures théories scientifiques appuyées sur notre meilleur savoir-faire expérimental et les traces de nos meilleures expériences ». Les itérations permettent à chaque discipline scientifique de construire un consensus au moins partiel sur l'état des connaissances, la science des manuels, ainsi que sur les difficultés soulevées par cet état qui définissent le périmètre des recherches à poursuivre.

La philosophie n'a évidement pas la même structure : aucun sujet n'y fait l'objet de consensus, que ce soit sur l'état de l'art ou même sur l'inventaire des problèmes ou difficultés. Une question qui peut alors caractériser l'épistémologie de l'expérimentation en philosophie serait la suivante : n'est-elle qu'une façon de présenter des arguments dans le débat philosophique en complément des outils conceptuels habituels du philosophe ? Ou, comme pour les sciences de la nature au 17^{e} siècle, est-elle le point de départ d'une nouvelle démarche itérative qui changera la nature de la connaissance philosophique, au moins en partie, pour une nouvelle branche qui s'en détacherait ? On peut reformuler cette question dans le cas de la philosophie morale expérimentale de la façon suivante : quelle part de la morale relève d'une science des comportements humains et, à ce titre, s'appuie sur l'expérimentation pour avancer par itérations épistémiques, et quelle part relève de questions philosophiques éternelles qu'aucun argument empirique ne saurait trancher ?

Cette section m'a permis de proposer dix axes qui résument les rôles qu'il est habituel pour les sciences naturelles d'attendre de l'expérimentation. Elle m'a également permis de souligner, sur chacun de ces axes, les grandes différences entre le contexte de ces sciences naturelles et celui de la philosophie morale expérimentale. Avant de clore ce chapitre sur la démarche scientifique expérimentale, je vais aborder un autre point qui a souvent été évoqué pour marquer la différence entre ces deux contextes, celui de la réplication des expériences qui permet de valider les résultats expérimentaux et qui serait satisfaisante pour les sciences

naturelles mais pas pour les sciences humaines sur lesquelles la philosophie morale expérimentale s'appuie.

3.5 La réplication d'expérience et son paradoxe

Avec les idées et les objets, les traces partageables, description des protocole employés et résultats obtenus, sont une composante importante de la démarche scientifique expérimentale. Disposer de traces partageables permet par exemple à d'autres équipes de rejouer les protocoles et de comparer les résultats lorsque l'expérience est répliquée. Cette réplication d'expériences est alors comprise comme contribuant à montrer que la démarche scientifique expérimentale a été menée de façon à fournir les traces permettant une réplication effective par une autre équipe de chercheurs. La réplication augmente la confiance accordée aux résultats expérimentaux, et la confiance entre scientifiques est un des éléments déterminants pour l'efficacité de cette démarche. Les objectifs portés par la réplication d'expériences sont donc nombreux, et, en particulier, dans le domaine de la psychologie qui a fait l'objet de constats alarmants en la matière. Si les expériences ne sont pas réplicables, alors la psychologie n'est peut-être pas encore assez scientifique, et le philosophe moral serait bien inspiré de ne pas se précipiter sur des résultats incertains avant de leur donner trop de poids. L'analyse de la réplication est porteuse d'un enjeu important pour l'étude de la philosophie morale expérimentale.

A titre d'illustration, citons certains des objectifs assignés à la réplication, repris pour partie d'un document récent de l'académie des sciences des Pays Bas (KNAW 2018) [136] et complétés :

— La demande de réplicabilité posée par les institutions sources des financements conduit les scientifiques à augmenter la qualité et la précision de la description des protocoles qu'ils utilisent.
— La réplication effective est un indice que ces descriptions soient suffisantes.
— Si les résultats sont statistiques, et c'est souvent le cas en psychologie, la réplication augmente la base des mesures et la puissance statistique du résultat.
— La réplication réussie est un indice en faveur de la non dépendance des résultats en regard de facteurs qui ne sont pas définis dans le protocole.
— La réplication est également un entraînement augmentant le savoir-faire des équipes d'expérimentateurs. Certaines expériences sont difficiles et demandent des capacités qui peuvent être atteintes par hasard exceptionnellement mais doivent être profondé-

ment maîtrisées pour être régulièrement répliquées.

— La réplication diminue le risque lié aux biais psychologiques des équipes de scientifiques.

— La réplication diminue le risque de fraude et de résultats faussés ou inventés.

— Enfin, il est important de remarquer que les résultats négatifs constatés lorsque la réplication échoue à obtenir les mêmes résultats alors qu'il semble que les protocoles soient les mêmes, sont aussi utiles que les résultats positifs. Ils mettent en lumière les difficultés pratiques qui soulignent les précautions à prendre avant de confirmer ou infirmer le résultat précédent. Ces difficultés peuvent être de tous ordres, relever de la théorie, des instruments de mesure ou de toute autre composante des protocoles expérimentaux. Enfin, quand les résultats précédents sont entachés d'un doute important, la réplication en échec aura permis d'économiser du temps et de l'argent en évitant d'aller chercher la cause de phénomènes dont l'existence même reste incertaine.

La réplication est un moyen pour construire un accord partagé sur les résultats empiriques au sein d'un domaine de recherche. Les résultats empiriques gagnent en crédibilité quand ils sont répliqués par plusieurs groupes indépendants. Ce gain, lorsqu'il se répète, expérience après expérience, appuie globalement la validité des travaux scientifiques et augmente la confiance envers les connaissances acquises au sein du domaine de recherche ainsi que la confiance envers les équipes de chercheurs. Inversement, l'absence de réplication fait douter d'un résultat et, si l'échec devient récurrent dans un domaine de recherche, c'est le domaine entier qui est mis en doute.

La réplication est importante pour la démarche scientifique expérimentale. Naturellement, elle ne saurait être une condition suffisante et nécessaire pour que la démarche soit dite scientifique. D'abord parce que cela limiterait par trop les approches possibles, et en particulier seraient pour partie exclues toutes les disciplines historiques comme la géologie ou l'astrophysique qui pourtant peuvent prétendre au rang de disciplines scientifiques même si la réplication y est par nature plus difficile. Ensuite parce qu'une bonne première expérience non encore répliquée est évidemment de la science. Et enfin, parce qu'une mauvaise expérience mal conçue et comportant des biais ne devient pas scientifique parce qu'elle a été répliquée, avec ses biais. La métaphore de l'hélice suggère d'ailleurs qu'une réplication à l'identique n'a pas beaucoup d'intérêt et qu'il faut refaire un tour complet sur les traces et les théories en cause pour concevoir une nouvelle expérience qui ne saurait être parfaitement identique, puisque le bateau aura avancé entre temps.

On peut néanmoins avancer sans risque que la réplication fait partie de ce que les scienti-

fiques eux-mêmes considèrent comme une bonne pratique devant être généralisée. L'éthique du chercheur suppose de fournir avec tout résultat l'ensemble des informations utiles à un autre chercheur pour qu'il puisse les confirmer et, symétriquement, il est de bonne méthode d'aborder un nouveau champ de recherche en commençant par répliquer les expériences de base du domaine de façon à en acquérir les compétences théoriques et pratiques.

Pour conclure, et évoquer l'insuffisance de la réplication dans la recherche telle qu'elle est décriée par exemple dans les documents cités plus haut [136], soulignons ce qu'on peut appeler le « paradoxe de la réplication » avec deux injonctions contradictoires. D'un côté, tout chercheur doit publier pour prouver qu'il existe, et travaille, or l'innovation est un critère important pour que soient acceptées les publications qu'il propose aux revues et, plus généralement, pour juger de leur intérêt. Donc il ne doit pas perdre de temps à répliquer des travaux déjà publiés. D'un autre côté, l'absence de réplication nuit à la réputation d'un domaine de recherche et, de ce fait, à tous les chercheurs qui le développent. Donc les chercheurs doivent passer du temps à répliquer et à le faire savoir. Le constat du caractère contradictoire de ces deux injonctions, il faut être innovant, et il faut répliquer pour confirmer, est à la base des plans d'action institutionnels comme celui de l'académie des sciences des Pays-Bas visant à un meilleur équilibre des deux types de publications.

Refaire une expérience déjà faite n'est évidement pas innovant. Le chercheur verra diminuer la probabilité d'acceptation d'un tel article par une revue, et on pourra ajouter que c'est une bonne chose pour le lecteur pressé de faire cette économie de temps si les résultats ne sont que confirmés. Si en revanche les résultats sont infirmés, il faudra alors que le chercheur gère le conflit potentiel avec les auteurs de l'article précédent, ainsi qu'avec tous les chercheurs qui se sont appuyés sur ces premiers résultats [29], et ce sera une grande consommation d'énergie, là encore pour un gain de réputation problématique car négatif, au moins dans un premier temps.

Ce paradoxe de la réplication a pris une certaine ampleur avec la prise de conscience publique de deux problèmes. Le premier est que lorsqu'elle est tentée, la réplication fait apparaître un nombre très important de résultats pas ou mal confirmés. A titre d'exemple illustrant ce doute généralisé, Ionnadis a choisi pour titre de son article de 2005 : « Why Most Published Research Findings Are False ». Il considère ainsi comme suffisamment acquis par les lecteurs qu'une majorité des résultats publiés est fausse, et peut poser ensuite la question de comprendre le pourquoi d'une telle situation (Ioannidis 2005) [122]. Le second problème,

29. Le cas de l'IAT développé plus loin est une illustration de cette difficulté lorsque l'utilisation de résultats potentiellement peu fiables a été généralisée.

plus large, concerne le faible taux d'articles comportant des compte rendus de réplications. Le manifeste de 2017 pour une science reproductible commence d'ailleurs par ce constat et la menace qu'il constitue pour la crédibilité de la science expérimentale dans son ensemble (Munafò et al. 2017) [168], le taux d'articles ayant fait l'objet d'une tentative de réplication serait par exemple de 1,07 % dans le cas de la psychologie (Mukerji 2019 page 71) [167].

Sans être la seule concernée, la psychologie est en première ligne des disciplines visées par le doute induit par le manque de réplication de ses résultats expérimentaux. Un article de 2015 (Collaboration 2015) [50] rend compte d'un effort important ayant consisté à reproduire une centaine d'expériences. Sur les 100 articles publiés répliqués, 97 comportaient initialement des effets significatifs qui, pour 57 d'entre-eux, n'ont pas été confirmés avec la même puissance statistique. Ce taux d'échec important se retrouve dans d'autres disciplines comme la psychologie sociale ou la pharmacie. Il est toutefois particulièrement dommageable à la psychologie, discipline récente dont la scientificité a longtemps été mise en question.

La crise de la réplication issue de ces deux problèmes a appelé des réactions des institutions académiques dans différentes directions. [30]

— Mesurer pour chaque discipline le nombre d'articles comportant des compte-rendus de réplication pour évaluer l'ampleur de l'absence, sans qu'il faille en déduire trop rapidement que la présence de réplication soit un indice fiable de qualité scientifique (KNAW 2018 p 48)
— Promouvoir et financer des campagnes de réplications par disciplines (KNAW 2018 p 49)
— Promouvoir la communication des résultats expérimentaux bruts, dans une optique de « science ouverte », pour toute nouvelle étude financée.
— Promouvoir et financer les dispositifs éditoriaux incluant la communication préalable des protocoles expérimentaux, des méthodes statistiques et algorithmes prévus, et ce avant que l'expérience ne soit menée.
— Lutter contre la réticence des revues scientifiques à publier des travaux de réplication.

Mon objectif n'est pas ici d'évaluer ce type de plan d'action ou d'en proposer de nouveaux, mais de marquer l'importance du problème de la réplication dans le débat public. Pour de nombreux scientifiques, et pour leurs institutions comme le montre l'exemple ci-dessus, la réplication est une pierre angulaire de la démarche scientifique, mais le taux réel de réplications tentées et publiées reste faible et le taux de confirmation quand la réplication est ten-

30. Pour un exemple récent de ce type de plan d'action voir la publication de la Royal Netherlands Academy of Arts and Sciences (KNAW 2018) [136].

tée étonnamment bas. Cette situation accroît le risque d'un scepticisme envers la démarche scientifique dont il faudra analyser l'impact sur l'acceptation de ses résultats par les philosophes moraux. Ce risque a été à la base de l'initiative de réplication dans le domaine de la philosophie expérimentale qui a conduit au constat d'un très haut taux de réplication, 70 %, sur un échantillon de 40 articles de philosophie expérimentale. Les résultats sont analysés dans l'article (Cova 2018) [58].

De plus, dans de nombreuses situations, l'apport de la réplication est problématique quand la multiplication des expériences, des méta-analyses et même des méta-méta-analyses ne semble pas permettre d'avancer. Un exemple familier de ce phénomène est observable dans le débat sur l'efficacité de l'homéopathie, régulièrement repris dans la presse. Des laboratoires reprennent des méta-analyses pour soit montrer que l'homéopathie ne fait pas mieux que l'effet placebo, soit montrer que l'homéopathie apporte un réconfort aux patients. On peut vouloir écarter cet exemple de l'homéopathie en l'attribuant non à la démarche expérimentale elle-même mais à des luttes entre une industrie, les laboratoires du domaine, et une administration cherchant à diminuer les remboursements, ou les scientifiques étudiant l'efficacité des produits, mais cela n'épuise pas le sujet et j'explorerai plus loin le cas de l'IAT (Implicit Association Test) qui montre l'opposition qui peut se construire et perdurer au sein même d'une discipline scientifique, entre équipes du même domaine, au-delà des accords obtenus grâce à la réplication.

Je reviendrai donc plus loin sur la réplication en regard de la philosophie morale expérimentale, mais notons d'ores et déjà qu'il ne saurait y avoir d'équivalent de la « crise de la réplication » dans le domaine de la philosophie traditionnelle, car l'apport empirique n'y a pas le statut de potentiel garant épistémique qu'il a pour l'ensemble des sciences expérimentales.

3.6 La démarche scientifique expérimentale, point d'étape

En conclusion de cette présentation de la métaphore de l'hélice expérimentale, que je tiens pour significative de la démarche scientifique expérimentale, à titre instrumental, dans le cadre de la présente thèse, je souhaite tout d'abord rappeler qu'il ne s'agit ici que de disposer d'une vue utile à la question posée, celle de l'examen de l'apport potentiel de la démarche scientifique expérimentale à la philosophie morale. Il est bien sûr hors de portée du présent travail de justifier en profondeur dans toutes leurs conséquences chacune des connotations portées par la métaphore de l'hélice expérimentale. Mais avant de présenter au chapitre suivant les cinq études de cas qui concrétiseront ce que les philosophie moraux soumettent à

l'enquête expérimentale, soulignons à nouveau que l'objectif de ce chapitre sera atteint si j'ai clarifié la conception de ce que j'entends par « démarche scientifique expérimentale ». Aujourd'hui bien loin des conceptions scientistes du 19e siècle, la démarche scientifique expérimentale est comprise comme un processus complexe qui enchaîne théories, expériences et traces dans des itérations épistémiques bénéficiant de la collaboration d'équipes de théoriciens, d'expérimentateurs, de statisticiens et, avec eux, d'une société entière de scientifiques engagés par leur commun refus de sortir, par excès de dogmatisme, de relativisme ou de pragmatisme, de la ronde de ces itérations.

Cette conception me permet d'une part, positivement, de mettre en perspective ce que la démarche expérimentale a pu, peut et pourra apporter, de façon générale et tout particulièrement sur la connaissance de la psychologie humaine. Et, d'autre part, défensivement, d'expliquer pourquoi je ne placerai pas un poids important dans ce qui suit à combattre certaines oppositions à la démarche scientifique qui sont appuyées non sur cette compréhension de la dynamique scientifique, mais sur une vision positiviste étroite de la science qui n'est plus celle de la science en train de se faire aujourd'hui, vision qui, malheureusement, est encore souvent rencontrée dans les débats entre philosophes moraux. C'est donc armé de cette conception de la démarche expérimentale et de la description du domaine moral proposée dans les chapitres précédents que je vais décrire cinq études de cas de l'emploi de cette démarche pour éclairer des questions mettant en jeu la morale humaine, le domaine moral décrit dans les chapitres précédents.

Chapitre 4

Cinq études de cas

4.1 Expérimenter sur l'expérimentation : une mise en abyme

Un chapitre précédent m'a permis de présenter le mouvement XPhi qui vise à bénéficier des avancées des sciences expérimentales et, en particulier, des sciences cognitives et de la psychologie expérimentale, pour apporter de nouveaux éclairages sur les questions de philosophie. J'ai souligné le bilan en demi-teinte de ce mouvement, reconnu par les acteurs du mouvement comme par les opposants, dont les expériences ont souvent été considérées comme apportant des éclairages nouveaux et intéressants à ces questions mais en aucun cas déterminants. Comment expliquer la faiblesse de ce bilan, et les doutes qu'elle soulève, alors que les succès de la méthode expérimentale dans les sciences naturelles laissait augurer de progrès rapides et déterminants ? Plusieurs voies de réponses sont possibles, et chacun peut construire sa théorie explicative de la difficulté de l'expérimentation à apporter des éléments déterminants aux débats philosophiques. Pour les uns, s'inscrivant dans la métaphore de l'hélice, il s'agit de premières tentatives, de ce fait imparfaites, et en poursuivant la méthode expérimentale, les connaissances issues des expériences gagneront peu à peu en pertinence et en précision, comme pour toutes les sciences naturelles. Pour d'autres, le sujet étudié, le comportement humain, est extraordinairement complexe et les méthodes dont nous disposons ne sont pas à la mesure de cette complexité. Pour d'autres encore, il s'agit moins de difficultés de méthode que de principe : il y aurait une erreur de catégorie à vouloir appliquer les outils des sciences naturelles, profondément marqués par le déterminisme, à des phénomènes en lien avec la liberté humaine.

Comment faire la part entre ces différentes considérations théoriques ? J'ai choisi pour cela de mettre la démarche expérimentale en abyme : d'expérimenter sur l'expérimentation.

J'ai cherché à expérimenter la démarche expérimentale appliquée à la philosophie, ce mouvement hélicoïdal enrichi de la démarche scientifique expérimentale qu'il conviendrait d'instaurer au sein de la philosophie morale. Ce choix de l'expérimentation s'appuie tout d'abord sur une nécessité : cette méthode est certainement la meilleure, et peut-être la seule, façon pour disposer d'une description intime, de l'intérieur, de la pratique expérimentale envisagée pour la philosophie. Ensuite, ce choix permet de varier les points de vue, ce que je fais au travers de cinq études de cas. Chacune des ces études a pour objectif de contribuer à donner un contenu substantiel, dans différents contextes, à une même question : est-ce que je peux m'exercer à une démarche expérimentale, telle qu'elle est reprise aux psychologues scientifiques, qui contribuerait à construire des arguments intéressants dans les débats de philosophie (en général) et surtout de philosophie morale (en particulier)? Chaque cas visera alors à cibler une des composantes postulées par les différentes théories explicatives du bilan en demi-teinte des XPhi.

Le premier cas, incontournable dans toute approche de la philosophie expérimentale, a pour objet l'expérience du tramway fou proposée initialement par Philippa Foot en 1967 (republié dans (Foot 1978) [88]) et qui a donné lieu depuis à de multiples ramifications. L'étude de la tramwaylogie est importante pour mon étude pour plusieurs raisons convergentes. D'abord, et comme pour de nombreux philosophes et psychologues moraux, cette expérience de pensée devenue ensuite expérimentation a été pour moi un point d'entrée riche et durable dans le mode de travail des philosophes expérimentaux. Ce cas donne à voir le passage de la méthode des cas utilisées par les philosophes analytiques « en fauteuil » à la méthode des philosophes expérimentaux appuyée sur des questionnaires inspirés par les méthodes des psychologues. Ensuite, c'est dans ce cadre que Joshua Greene a utilisé pour la première fois en 2001 l'Imagerie à Résonance Magnétique Fonctionnelle (IRMf) pour construire un argument destiné à un débat classique de philosophie morale (Greene 2001) [108] (voir 4.2, page 161). Cet article marque une étape importante dans l'histoire de l'utilisation des résultats scientifiques pour évaluer des théories morales. Utiliser ainsi l'imagerie cérébrale pour observer de l'intérieur un humain philosopher n'est ni la première pierre de cette histoire, ni la dernière, mais il n'est évidemment pas anodin que la naissance du mouvement XPhi en soit contemporaine. Enfin, les multiples variantes du problème du tramway reflètent des préoccupations très diverses permettant d'illustrer très largement tant les apports potentiels de cette démarche de philosophie expérimentale que les réticences soulevées au sein de la communauté philosophique par ces apports et par cette démarche.

Les études suivantes sont issues de travaux que j'ai menés dans le cadre de cette thèse au

sein du séminaire de philosophie expérimentale de l'ENS Ulm animé par Brent Strickland. Pour la deuxième étude de cas, le travail a consisté à réaliser une enquête portant sur l'évaluation par un échantillon d'internautes du nombre de musulmans en France. Ce travail me permettra ici de mettre en avant les caractéristiques de ce type d'enquêtes par questionnaires sur Internet très fréquemment utilisé par les philosophes expérimentaux, ainsi que certaines des multiples difficultés d'interprétation qu'elles soulèvent.

Le troisième cas porte sur un test destiné à mesurer les associations implicites que nous faisons entre des concepts, le test dit IAT pour, en anglais, Implicit Asociation Test. J'ai retenu ce cas d'étude pour plusieurs raisons, dont principalement le fait que l'utilisation généralisée de ce test aux États Unis dans l'objectif de lutter contre la discrimination raciale a donné lieu à des débats polémiques entre équipes de psychologues. Cet exemple me permettra de montrer de façon contrastée que la méthode expérimentale a permis qu'un accord soit établi entre ces équipes, pourtant en opposition, sur une part importante de l'analyse du phénomène IAT sans pour autant que s'éteigne la polémique qui, c'est ce que je vais mettre en relief, peut être analysée comme liée aux poids relatifs différents donnés à des systèmes de valeurs morales qui sont, dans cet exemple, incompatibles. Le cas de l'IAT permet d'illustrer ainsi tant la puissance que les limites de la démarche expérimentale appliquée par des scientifiques à des questions dont la dimension morale est importante.

Le quatrième cas retrace une opposition entre deux équipes, une de psychologues et l'autre de spécialistes de l'olfaction, sur le rôle des larmes humaines. Ce cas me permettra de contraster les méthodes utilisées par les deux équipes de psychologues aux spécialités très éloignées et, en particulier, d'illustrer que la grande complexité des outils utilisés rend en pratique très difficile la collaboration entre ces équipes, aux cultures très différentes, quand elles ne sont pas en confiance, malgré la proximité de leurs domaines d'étude.

Enfin, dans le cinquième et dernier cas, j'explore un autre résultat classique de philosophie morale expérimentale ayant donné lieu à de multiples variantes et prolongations, la dissymétrie d'attribution d'intentionnalité pour des effets indirects selon qu'ils sont moralement positifs ou négatifs, ce qu'à la suite d'un article de 2003 de Joshua Knobe [138] on a pu baptiser « l'effet Knobe ». Le point de départ de mon étude était qu'il me semblait que les notions de responsabilité, d'intentionnalité et de causalité étaient mobilisées de façon légèrement, mais significativement, différente dans chacun des articles cherchant à affiner les processus du jugement moral porté sur les résultats indirects d'une action, et ce flou notionnel générait chez moi un sentiment d'insatisfaction. Ma démarche pour analyser cette insatisfaction a été la suivante. D'abord, pour éviter le biais de sélection, prendre exhaustivement tous les articles

se référant à l'effet Knobe dans une période donnée, ici 33 articles de 2016 à 2018. Ensuite analyser pour chacun de ces articles les notions prises en compte par l'étude expérimentale. Enfin, en prenant la hauteur nécessaire à l'observation simultanée de tous les articles, donner un contenu plus précis au constat d'instabilité notionnelle source de mon insatisfaction de départ.

Prenant du recul en regard de ces cinq études, je propose dans une dernière section trois axes synthétiques principaux pour en subsumer le contenu, pour examiner les traces produites par cette mise en abyme, cette expérimentation sur l'expérimentation. Le premier est, trivialement, que les résultats apportés par la démarche expérimentale sont, en pratique, toujours intéressants. De nombreuses distinctions sont apparues que des démarches du philosophe en fauteuil n'avaient pas fait apparaître. C'est un des objectifs du présent chapitre d'en témoigner.

Le second axe concerne les difficultés opératoires rencontrées pour mettre en œuvre la démarche scientifique expérimentale dans ce cadre de la psychologie morale. En application de la métaphore de l'hélice, je propose de comprendre la faiblesse récurrente des interprétations des résultats des expérimentations comme la marque d'un faiblesse antérieure, celle de l'opérationnalisation, qui elle-même traduit certainement pour partie une autre faiblesse encore antérieure, celle de la définition des notions que les théoriciens du domaine moral emploient. Le chapitre suivant aura pour objet d'approfondir cette étape d'opérationnalisation et sa mise en œuvre dans le domaine moral.

Enfin, troisième axe en synthèse de ces études de cas, chacun des cinq cas a laissé derrière lui un sillage de philosophes moraux réticents à accepter l'importance, l'utilité ou même la validité des conclusions expérimentales. J'interprète cet état de fait comme la conséquence de la très grande primauté donnée à la perspective descriptive dans toutes ces expériences, aux dépends des autres perspectives. Je fais référence ici au premier chapitre qui a marqué l'utilité des quatre perspectives que peut adopter successivement le philosophe moral sur le domaine moral : perspective descriptive, substantielle, méta-éthique et éthique appliquée. Tenter, comme je le fais dans ce chapitre, d'exercer en philosophe expérimental, c'est certainement privilégier la première perspective, descriptive, en plongeant un expérimentateur, l'auteur, dans les méandres des jugements moraux réels en situation et en contexte d'expérimentation. Par une mise en abyme de la démarche du mouvement XPhi, qui me servira lui-même d'expérimentation sur ce que peut l'expérimentation en philosophie, je tente de relever si, et par quels chemins, cette perspective descriptive peut éclairer les trois autres. Ce sera l'objet des chapitres suivants. Mais avant d'envisager cette réflexion comme simplement pos-

sible, il m'est nécessaire d'entrer dans le détail de la pratique car, de la démarche scientifique expérimentale, je retiens l'indispensable construction des savoir-faire expérimentaux comme aussi importante que les constructions conceptuelles et théoriques. Une hélice non équilibrée est vouée aux fissures et à la destruction. Ce sera donc l'apport de ce chapitre d'appliquer aux cinq études de cas décrits la grille d'analyse développée au premier chapitre sous la forme des quatre perspectives de la philosophie morale, descriptive, prescriptive, méta-éthique et appliquée.

4.2 Le tramway, introduction à la méthode

4.2.1 Le tramway, entre expérience de pensée et expérimentation

L'expérience de pensée du tramway est à la philosophie morale expérimentale ce que le plan incliné de Galilée est à la loi de la chute des corps. Un moment initial, une source majeure d'innovations, et une référence incontournable [1]. Incliner le plan sur lequel roule une bille a permis à Galilée de ralentir la chute libre et, ainsi, de rendre quantifiables les temps de chute malgré le manque de sensibilité des instruments de mesure du temps dont il disposait. Galilée a alors pu établir que la vitesse augmentait dans la même proportion à chaque pas de temps, posant une pierre sur le chemin qui conduirait à la loi de la gravitation de Newton. De façon analogue, l'expérience de pensée du tramway a permis aux philosophes moraux d'accéder à l'expérimentation pour approcher le phénomène complexe du jugement moral et à la possibilité d'établir des mesures en exagérant le phénomène par l'utilisation de situations extrêmes. Naturellement, la loi de la gravitation morale n'a pas encore été trouvée (ou, en tous cas, reconnue comme trouvée par la communauté des philosophes moraux), mais les philosophes moraux expérimentaux peuvent espérer être en chemin et cette expérience du tramway est entrée massivement dans la boîte à outils des philosophes et psychologues. [2]

Comme certainement de nombreux chercheurs s'intéressant à la philosophie expérimentale, cette expérience du tramway a été pour moi un point d'entrée dans ce domaine et mon intérêt pour elle a été encore renforcé avec la découverte de l'article de Joshua Greene de 2001 [108] qui est considéré comme le premier article d'un philosophe utilisant l'IRMf pour instruire une question de philosophie morale.

1. Aucun des ouvrages d'introduction à la philosophie expérimentale (ni d'ailleurs les ouvrages plus avancés), cités dans la présente thèse ne fait l'impasse sur cette expérience du tramway [224] [223] [204] [200] [175] [140] [167] [149]

2. Interrogé le 5 novembre 2019, le site Google Scholar indique 1110 entrées pour « Moral Philosophy Trolley » depuis 2019, donc sur 10 mois, et 5830 entrées depuis 2015. Chaque jour ouvrable, plus de 5 articles mentionnant l'expérience du tramway sont publiés.

Après avoir mentionné différents dilemmes sacrificiels, ensemble dans lequel s'inscrit le dilemme du tramway, je résumerai rapidement en quoi consiste cette expérience de pensée, puis montrerai ce qu'a signifié le passage de la philosophie « en fauteuil » vers la philosophie expérimentale dans ce contexte et en quoi une telle expérience est porteuse d'un potentiel de recherches important, tant en psychologie qu'en philosophie.

4.2.2 Présentation des dilemmes sacrificiels

L'expérience du tramway fait partie d'un ensemble d'expériences de pensée qu'il est habituel d'appeler « dilemmes sacrificiels » qui consistent à proposer des situations où le sacrifice d'une personne permet de sauver plusieurs autres personnes. Le philosophe moral s'intéresse à ce qui différencie les cas où le sacrifice nous semble moralement acceptable, voire souhaitable, des cas où il nous apparaît moralement inadmissible. Avant d'entrer dans le détail de la description de l'expérience du tramway, citons rapidement deux autres exemples de tels dilemmes sacrificiels, le dilemme des greffes et le dilemme des naufragés. Dans le dilemme des greffes, cinq personnes sont dans un hôpital atteintes de maladies incurables et mourront si elles ne reçoivent pas rapidement une greffe. Une personne en bonne santé arrive à l'hôpital suite à un accident, ses chances de survie sont faibles mais non nulles. Le chirurgien peut-il négliger ces chances de survie et utiliser les organes de la personne sacrifiée pour sauver les cinq patients en attente de greffe ? Le scénario ainsi présenté permet de faire varier les caractéristiques de la situation depuis des cas où une majorité de personnes interrogées répond positivement, par exemple si la personne sacrifiée est déjà en état de mort cérébrale, jusqu'à des cas où une majorité de personnes répond négativement, par exemple si la personne est simplement un passant en bonne santé [3]. Pourtant dans les deux cas, le bilan peut apparaître comme proche à un conséquentialiste convaincu, une personne sacrifiée pour cinq personnes sauvées. Le second dilemme met en scène des naufragés : six personnes sont sur un radeau qui n'arrivera pas au rivage si elles restent toutes sur l'embarcation, en revanche si une est sacrifiée, les cinq autres seront sauvées. Là encore, le scénario peut être modifié pour étudier quelles caractéristiques le rendent moralement acceptable.

Ces deux dilemmes des greffés et des naufragés, ainsi que d'autres dilemmes sacrificiels, ont été également étudiés, mais le dilemme du tramway a connu un succès particulier et ce serait une étude à mener de comprendre pourquoi les psychologues et philosophes se sont attachés ainsi à développer les variations sur cette expérience ferroviaire. Si l'interrogation

3. Pour une présentation imagée de 7 scénarios construits pour être progressivement de moins en moins acceptables sur cette base du dilemme des greffes voir l'excellente vidéo de Monsieur Phi (Thibaut Giraud) `https://www.youtube.com/watch?v=AZBDMN5wZ-8&t=241s`

« Moral Philosophy Trolley » donne 1100 retours sur Google Scholar depuis début 2019, la même interrogation en remplaçant « Trolley » par « Sacrificial Dilemma » donne 1200 entrées dont les 1100 précédentes. Il semble à l'examen rapide des 20 premiers articles de la sélection que ces autres dilemmes soient utilisés en complément, peut-être pour confirmer des résultats obtenus avec le dilemme du tramway. Entrons dans le détail de ce dilemme du tramway.

Le dilemme proposé par Philippa Foot en 1967 [88] s'appuie sur le scénario suivant :

> Un tramway fou sans chauffeur dévale une pente et va, si rien ne l'arrête, écraser cinq personnes qui travaillent sur les rails [4]. Vous êtes témoin de la scène et vous êtes à proximité d'un aiguillage qui vous permet d'envoyer le tramway vers une voie de garage. Malheureusement une personne est sur cette voie de garage et sera écrasée si vous basculez cet aiguillage. Que faites-vous ?

Dans l'article inaugural de Philippa Foot, ce scénario, ainsi que plusieurs autres, avait pour objectif d'aider à discerner ce qui rend le jugement moral sur l'avortement si complexe avec, en particulier dans certains cas, la nécessité de peser les valeurs relatives des vies de l'enfant et de la mère.

En 1976, Judith Jarvis Thomson [226] a proposé une variante de ce dilemme. Le tramway dévale toujours la pente au péril de la vie de 5 personnes, mais là, plus d'aiguillage, en revanche vous êtes sur un pont avec à votre côté un gros homme. Si vous le poussez sur la voie, alors l'obstacle qu'il constitue arrêtera le tramway, mais le gros homme périra dans le choc.

Le bilan est le même que précédemment, 5 vies sauvées au prix d'une, mais l'attitude majoritaire dans chacun des cas est diamétralement opposée. Dans le premier cas l'intérêt qu'il y a à sauver 5 vies l'emporte et nous voyons la conséquence néfaste comme un mal induit, par contre dans le second cas notre réticence à tuer délibérément le gros homme l'emporte et nous reculons devant cette décision qui, pourtant, entraînerait le même bilan. Judith Jarvis Thomson utilise cet exemple, et d'autres comportant la même structure, pour souligner la complexité du jugement moral qui ne peut se satisfaire des théories trop simples que sont le conséquentialisme, qui conduirait à pousser le gros homme comme on baisse l'aiguillage puisque les bilans sont les mêmes, ou le déontologisme, qui conduirait à ne jamais accepter de tuer une personne, même si c'est le prix à payer pour en sauver cinq.

4. Détail technique : un tramway est toujours sur rail alors qu'un trolley prend son énergie à des caténaires et est, en France, plus fréquemment sur pneu. Il convient donc de retenir le terme de tramway en français pour décrire l'expérience proposée par Philippa Foot. Le terme de trolley peut être utilisé dans le monde anglo-saxon où de nombreux trolleys sont sur rails.

4.2.3 Appliquer nos intuitions morales à de multiples scénarios

Le dilemme du tramway ainsi enrichi d'une variante « aiguillage » et d'une variante « gros homme » permet à chacun d'exercer son intuition morale et de vérifier si cette intuition est en accord ou non dans ce cas particulier avec la théorie morale qu'il privilégie. A ce stade, nous pouvons considérer ce travail comme caractéristique de la philosophie « en fauteuil » telle que je l'ai évoquée plus haut. Ce travail se poursuit sans interruption depuis 1976 et en 2014 Stijn Bruers et Johan Braeckman [33] ont compilé dans un article les nombreuses variations publiées pendant 35 ans ainsi que les aussi nombreuses théories morales proposées pour établir si, et pourquoi, dans chaque cas, il est moralement admissible de tuer une personne pour en sauver 5. Bruers et Braeckman proposent de regrouper en huit familles les scénarios qui ont pu être élaborés :

— 1 Le scénario de l'aiguillage : vous pouvez agir sur un aiguillage en sauvant les 5 personnes mais au prix d'une personne écrasée sur la voie où vous avez envoyé le tramway.
— 2 Le scénario du gros homme poussé du pont : vous poussez le gros homme, il arrête le tramway mais meurt dans le choc.
— 3 Vous agissez sur un aiguillage qui dévie le tramway sur une dérivation qui rejoint la voie avant les 5 travailleurs. Sur cette dérivation travaille un gros homme qui sera percuté et arrêtera le tramway au prix de sa vie.
— 4 Identique au scénario 3 sauf qu'il y a une grosse pierre sur la dérivation derrière le travailleur victime et c'est elle qui va arrêter le tramway.
— 5 Vous pouvez pousser un camion sur la voie du tramway, et c'est son conducteur (ou un passager caché dans le camion) qui est tué dans le choc arrêtant le tramway.
— 6 Identique au scénario 4, et le mort n'est plus un travailleur mais un passant lui-même écrasé par la pierre mise en mouvement.
— 7 Les 5 personnes ne sont plus sur les rails mais sur un wagon à l'arrêt et vous pouvez déplacer ce wagon vers une voie de garage sur laquelle il y a une personne qui sera écrasée dans la manœuvre.
— 8 Dans ce scénario plus complexe, un dispositif de sécurité empêche une grosse pierre de tomber sur la voie du tramway. Ce dispositif est maintenu en place par un surveillant. Vous pouvez tuer ce surveillant et lever ainsi le dispositif, la pierre tombera, arrêtera le tramway et les 5 personnes seront sauvées.

La différence initialement constatée entre les deux premiers scénarios est interprétée par

Bruers et Braeckman comme mettant en évidence notre difficulté à porter un jugement moral quand une même action déclenche deux chaînes causales dont l'une aboutit à une conséquence positive, 5 personnes sont sauvées, et l'autre une conséquence négative, 1 personne est morte. Chacun de ces 8 scénarios recensés par les auteurs fait varier ces deux chaînes causales pour tenter de rendre sensibles les différences que les philosophes moraux ont introduites entre des actions intentionnelles ou non, entre des conséquences facilement prévisibles ou non, qu'elles soient évidemment certaines ou seulement possibles, entre les cas où la mort de la victime est recherchée en tant que moyen et ceux où elle est un sous-produit indirect non recherché.

Le dilemme du tramway permet au philosophe en son fauteuil d'imaginer, pour chaque distinction conceptuelle envisagée, de nouveaux scénarios permettant de concrétiser les deux branches de la distinction. Une fois ces multiples scénarios et ces multiples distinctions conceptuelles établis, le philosophe moral va faire l'inventaire des théories morales disponibles qu'il souhaite évaluer en regard de cet ensemble de scénarios. Dans leur article de 2014, Bruers et Braeckman écartent rapidement les grandes théories morales, le déontologisme et le conséquentialisme, qui ne peuvent répondre de façon satisfaisante et citent trois familles de théories morales qui ont été défendues, et en partie développées, pour répondre aux dilemmes du tramway. La première famille insiste sur l'importance de la différence morale entre une action considérée comme une fin et une action considérée comme un moyen (cette famille s'inspire de l'impératif kantien de ne jamais utiliser autrui comme un simple moyen). La deuxième famille marque l'importance de la distinction à faire entre dévier une menace existante, ce qui serait permis, et créer une nouvelle menace, ce qui serait interdit. Enfin la troisième famille tente de clarifier comment notre jugement moral prend en compte les chaînes causales complètes qui vont de la décision vers, d'un côté, les cinq vies sauvées et, de l'autre côté, la victime sacrifiée. Une des propositions, dans cette troisième famille, consiste à supposer que notre capacité de compréhension des chaînes causales est limitée au cas où tous les événements sont sur une seule et même chaîne d'événements. Notre jugement moral serait du fait de cette limitation sensible à la conséquence négative, le sacrifice, lorsqu'elle est sur la même chaîne d'événement que la conséquence positive (le cas « gros homme »), mais serait insensible à cette conséquence négative quand plusieurs chaînes causales indépendantes ont divergé, comme dans le cas « aiguillage » où, une fois l'aiguillage basculé, ce qui se passe sur la voie principale n'est que gain, indépendamment de ce qui peut se passer sur l'autre voie.

Le résultat de ce travail, principalement en fauteuil, est un tableau comportant en ligne

les scénarios imaginés et en colonne les familles de théories morales retenues avec, à chaque intersection, comment une personne adoptant cette théorie morale jugerait l'action dans ce scénario. Il reste ensuite au philosophe en fauteuil à exercer son intuition morale pour affirmer que l'action est ou non acceptable dans tel ou tel scénario et à comparer cette intuition au résultat déduit de chacune des théories morales.

Table 1 Answers to the trolley dilemmas, according to the three accounts

Dilemma	Mere means account	Same threat account	Causal chain account
1. Switch	+	+	+
2. Bridge	−	−	−(?)
3. Loop	−	+	−(?)
4. Loop and stone	+	+	−(?)
5. Truck	+	−	−(?)
6. Rockslide	+	−	+
7. Platform	+	−	+
8. Loop and avalanche	+	+	−

FIGURE 4.1 – Tramway, scénarios et théories d'après Bruers et Braeckman 2014

Bruers et Braeckman sont des philosophes moraux à la recherche d'une théorie morale satisfaisante, qu'ils comprennent comme un « algorithme » appliqué aux caractéristiques objectives du scénario et ne prenant pas en compte les caractéristiques psychologiques individuelles des participants. Les trois familles de théories morales retenues reflètent cet objectif de recherche. Un philosophe moral moins algorithmique et plus psychologue viserait à mieux comprendre le processus du jugement moral et, pour cela, multiplierait les scénarios en faisant varier des éléments influant sur les émotions ressenties par les participants, par exemple en étudiant la différence de jugement moral qu'apporterait une qualification des 5 personnes à sauver comme des travailleurs ou comme des terroristes en train de poser une bombe, ou en étudiant la différence entre une description abstraite, disons « 5 personnes » sans aucune indication précise, et une description concrète nommant les 5 personnes, Pierre, Paul, Jacques, et précisant leurs âges, situations familiales... et ainsi de suite, à chaque hypothèse sur les processus psychologiques en jeu pourrait correspondre la recherche de scénarios les mettant en évidence.

Une telle approche conduirait donc, en regard de l'article de Bruers et Braeckman, à une grande augmentation du nombre de scénarios et, parallèlement, à une augmentation du nombre de théories morales listant les processus psychologiques envisagés. Après cette inflation du nombre de lignes et de colonnes dans le tableau ci-dessus, le travail restant à faire au philosophe en fauteuil consisterait à évaluer si, dans chaque scénario, l'application de chaque théorie morale examinée donne un résultat conforme ou non à sa propre intuition morale.

Le philosophe (et le psychologue moral) en fauteuil auront ainsi inventorié les multiples théories morales, dont le nombre et la finesse de définition peuvent varier en fonction de la thématique de recherche, ainsi que les multiples scénarios conçus pour donner à voir les différences entre ces théories morales. Le philosophe en fauteuil entreprend alors l'analyse de ce tableau, d'un côté pour chaque scénario et chaque théorie il émet un jugement moral conforme à la théorie et, de l'autre côté, il émet un jugement moral conforme à son intuition. Si les deux jugements diffèrent, il en déduit un argument contre la théorie[5], si les deux jugements sont concordants, il en déduit un argument pour cette théorie. C'est le bilan de tous les scénarios qui lui permet d'évaluer chaque théorie.

Le philosophe en fauteuil considère qu'il peut mener seul, ou en débattant avec ses collègues philosophes, l'étape d'évaluation des scénarios sur la base de sa propre intuition morale. Point clé, cela suppose qu'il considère son propre jugement moral comme moralement valide, ou au moins significativement intéressant, en regard du cas étudié. Les philosophes expérimentaux considèrent que c'est précisément cette étape, donnant ce rôle clé à l'intuition, qui constitue le cœur du problème de la méthode. La méthode n'est globalement valide que si l'intuition du philosophe l'est.

4.2.4 L'approche expérimentale du dilemme

La première étape que franchit le philosophe expérimental en rupture avec la méthode du philosophe en fauteuil est de remplacer l'intuition du philosophe par un sondage sur un échantillon de personnes. Ce changement peut paraître mineur[6] ; il ne s'agit que de poser à plusieurs personnes la question que le philosophe s'administrait à lui-même et apportait à ses collègues philosophes dans une publication spécialisée. Mais, à l'examen, il constitue un premier pas important : il transforme l'affirmation du philosophe « je pense que P » en une affirmation vérifiable par un tiers « n % des humains pensent que P »[7]. Rendre ainsi l'observation vérifiable par un tiers, et construire les traces qui s'inscriront comme résultat de l'expérimentation exploitable par la communauté scientifique, est un préalable à l'approche scientifique expérimentale telle que je l'ai décrite avec la métaphore de l'hélice.

Dans le cas du tramway fou, le résultat du sondage est le suivant : 80 % des personnes

5. L'argument est discutable et discuté, il consiste à déduire de l'existence d'une différence de jugement entre théorie et intuition l'appréciation que la théorie n'est pas intuitive, et qu'une théorie non intuitive est plus faible qu'une théorie qui reproduit bien les jugements intuitifs. Une version moins forte de l'argument consiste à viser simplement à reporter la charge de la preuve sur les défenseurs de la théorie non intuitive. On peut douter de la force de ces deux arguments à la lumière du caractère non intuitif de nombre des résultats scientifiques reconnus, et en particulier de la physique moderne et du caractère non intuitif de ses théories.

6. Voir par exemple Bernard Baertschi p 95 dans (Merrill et Savidan 2017) [162]

7. Pour l'analyse de la validité de l'observation à la première personne voir (Piccinini 2009) [181].

disent agir sur l'aiguillage alors que seulement 30 % des personnes disent pousser le gros homme [8]. Ces expériences ont été reproduites : les résultats ont pu varier dans le détail sans remise en cause du résultat principal. Une large majorité des personnes interrogées, mais pas la totalité, agit sur l'aiguillage et une petite minorité passe à l'action quand il s'agit de pousser le gros homme. Certains scénarios intermédiaires, comme le scénario 2 qui comporte l'envoi du tramway sur une dérivation sur laquelle se trouve un gros homme, conduisent à des résultats plus équilibrés avec 50 % des sondés disant qu'il est moralement correct d'agir sur l'aiguillage et 50 % que c'est incorrect.

Les apports du passage de l'affirmation du philosophe « je pense que P » au résultat d'une enquête « x % des personnes pensent que P » sont nombreux, et, en particulier, permettent d'accéder également à la sensibilité du phénomène mesuré par « x » selon les variations des circonstances présentées ainsi qu'à la quantification des raisons des choix en enrichissant le protocole d'enquête. Pour le cas de l'aiguillage, les raisons exprimées par les participants confirment le poids de l'argument conséquentialiste : il est pertinent de sauver 5 personnes au prix d'une. En revanche pour le second cas, le gros homme, les participants ne mentionnent plus le bilan des conséquences mais évoquent les émotions négatives qu'induit chez eux l'idée de pousser un homme vers une mort certaine. Avec ce recours au sondage, l'interprétation des résultats se complexifie par l'irruption des émotions dans le débat entre conséquentialisme et déontologisme. Mais le sondage ne permet pas pour autant d'analyser comment les émotions et le calcul des conséquences s'articulent pour produire le jugement moral.

On peut évoquer, en première approche, trois pistes d'étude des rôles respectifs des émotions et du calcul des conséquences. La première consiste à proposer que l'idée de pousser le gros homme suscite des émotions qui activent des réactions morales profondément déontologiques. Aucune émotion n'étant induite par l'idée de pousser un aiguillage, l'esprit est libre pour faire des calculs, ce qui conduit à donner le prima au bilan 5 contre 1. En revanche, l'idée de pousser un homme vers la mort nous met dans un état psychologique qui active les règles déontologiques profondément ancrées en nous (par l'éducation ou par nos gènes, peu importe ici) et cela nous interdit d'aller plus loin dans le calcul, le rejet s'impose.

Une deuxième piste consiste à proposer, inversement, qu'enfreindre une règle morale est une source d'émotions qui inhibent l'action. Par exemple dans les théories morales supposant l'existence d'un « sens moral », on affirme que le sens moral est sensible aux conditions de présentation du scénario, il n'est pas activé dans le « cas aiguillage » et activé dans le cas

8. Les pourcentages varient selon les articles de 80 % et 30 % dans les articles initiaux de 2001 à 90 % et 10 % dans l'article de synthèse de 2014. Chaque chercheur aura donc à reconstruire son échantillon témoin pour caler ses propres expériences dans ces fourchettes.

« gros homme ». Lorsque la violation de la règle morale est détectée, le sens moral induit un état psychologique qui inhibe l'action en cause, indépendamment de tout calcul des bénéfices qu'elle peut apporter.

Enfin, troisième piste, on peut également envisager que les réactions émotionnelles et morales ne soient pas causalement liées, comme dans les deux premières pistes, mais combinent leurs effets. Dans ce cas, la décision est conçue comme un processus cognitif complexe multifactoriel qui prend en compte globalement les émotions, les croyances, les jugements moraux et tout élément disponible en les composant, sans qu'une priorité systématique puisse être affirmée.

Les possibilités sont nombreuses et on voit ici en œuvre l'apport de l'approche expérimentale : l'irruption des émotions comme élément important pour les participants oblige le philosophe moral à expliciter comment elles sont prises en compte dans sa théorie morale. A la différence de l'expérience de pensée définie par le philosophe et menée sur lui-même à l'aide de ses propres intuitions, la confrontation à l'expression des participants élargit le débat ; à charge pour le philosophe d'en tirer profit pour l'approfondir.

4.2.5 Le tramway, drosophile de la philosophie morale

Le philosophe expérimental et le psychologue moral se sont emparés du paradigme du tramway fou en tant que base expérimentale particulièrement féconde. Elle est simple : on constate une différence importante, et surtout quantifiée et assez stable, de jugement moral entre les cas « aiguillage » et « gros homme » qui ne s'explique pas sans difficulté avec les grandes théories morales disponibles. Elle est extensible : on peut imaginer de nombreuses variations du scénario qui permettent d'approcher les aspects qui ont une influence sur le jugement moral, en prolongement des scénarios imaginés par les philosophes moraux pour illustrer les distinctions conceptuelles qu'ils promeuvent. Et on peut également faire varier, outre le scénario, les caractéristiques des personnes interrogées et les conditions d'application du questionnaire de façon à approcher l'influence de ces caractéristiques psychologiques sur le jugement moral en regard d'un résultat de base assez stable : dans le cas d'une population témoin aléatoire, 80 % des sondés affirment qu'ils basculeraient l'aiguillage et 20 % non alors que 30 % seulement affirment qu'ils pousseraient le gros homme et 70 % refusent

Le philosophe et le psychologue peuvent maintenant utiliser le paradigme du tramway fou pour explorer l'influence de nombreux éléments de contexte par la méthode des différences : si une population qui a telle caractéristique s'écarte par ses réponses de la population témoin,

c'est que cette caractéristique intervient dans le processus du jugement moral. Parmi les multiples variantes imaginées, remarquons à nouveau que les philosophes moraux ont privilégié, comme je l'ai précisé plus haut avec Bruers et Braeckman, celles qui portent sur la causalité, au sens d'un chemin quasi algorithmique qui prend en entrée les circonstances de l'action et donne en sortie ce que devrait être le jugement moral correct, car elles permettraient selon eux de comparer les théories morales entre-elles, indépendamment de la psychologie des sondés [9]. Les psychologues privilégieront plutôt les variations expérimentales qui portent sur les processus psychologiques individuels des participants sondés de façon à éclairer quels processus psychologiques sont à l'œuvre dans le jugement moral. Ils chercheront à savoir quelle part du jugement moral est lié à un raisonnement logique, ou à une émotion comme la culpabilité, la honte ou la colère ou à tout état mental particulier qu'il serait possible de provoquer en jouant sur les conditions de l'expérience. Citons rapidement certains des paramètres dont l'influence psychologique a été recherchée.

— Les conditions d'anonymat et d'impunité assurés au participant. L'étude vise à analyser si la réponse négative concernant le gros homme est liée à la crainte de pouvoir être ensuite accusé de sa mort. Si c'est le cas, la garantie d'anonymat devrait conduire à diminuer la différence entre les deux cas de l'aiguillage et du gros homme.

— La langue du questionnaire pour des participants bilingues, en s'appuyant sur la relation affective avec la langue maternelle qui véhiculerait mieux les émotions. La différence entre les deux cas « gros homme » et « aiguillage » serait ainsi amoindrie en utilisant une langue apprise, plus neutre sur le plan émotionnel.

— La tonalité humoristique ou non des vignettes utilisées, qui pourrait permettre une certaine prise de distance favorisant l'approche conséquentialiste. Là encore, l'idée est que l'expression humoristique est moins porteuse d'émotions.

— La formation des participants aux théories morales, pour analyser la différence entre population générale et experts de la chose. Cette approche vise à répondre à l'argument de l'expertise, cet argument utilisé par les critiques de la XPhi soutient que les réponses de non professionnels de la philosophie à des questions philosophiques est de peu d'intérêt. Si un sondage fait auprès de philosophes donne le même résultat, la critique se voit émoussée.

— Le remplacement des vignettes et d'une présentation écrite des scénarios par une immersion dans une image de réalité virtuelle en 3D. L'étude vise à analyser si les ré-

9. Ces auteurs proposent de qualifier de « agent neutral », neutres vis à vis de l'agent, les explications qui ne font pas appel à des caractéristiques psychologiques particulières des agents et seraient valides pour tout agent rationnel.

actions morales sont identiques quand l'immersion se fait de façon plus intense, se rapprochant plus, ou en tous cas différemment, des conditions de la vie réelle.

— Faire tomber le gros homme par une trappe activée par un levier de façon à éviter l'effet supposé du rejet du contact physique avec la victime. Si le refus de pousser le gros homme était dû au rejet de l'acte physique de toucher quelqu'un pour le pousser vers la mort, alors l'intermédiation par un dispositif technique de type de la trappe devrait réduire l'écart entre les cas aiguillage et gros homme.

— Formuler les vignettes de façon abstraite (cinq personnes contre une personne) ou de façon concrète en donnant par exemple des noms et des éléments de biographie. L'idée est que le jugement moral dépend de la capacité d'identification aux acteurs et que celle-ci est plus importante avec des formulations concrètes.

— Enfin, citons pour mémoire toutes les études qui sont déclinables quel que soit le sujet de l'expérimentation envisagée. Elles recherchent les différences de jugement moral selon les caractéristiques des sondés, la culture, le genre, les pathologies, ...

Il sera impossible de détailler ici toutes les variantes tant la littérature de « tramwaylogie » est abondante ; j'en ai ci-dessous extrait trois exemples de nature à montrer comment les psychologues et philosophes expérimentaux ont progressivement utilisé cet instrument.

Le premier exemple porte sur la comparaison entre un échantillon constitué de personnes ayant des troubles autistiques et un groupe de contrôle (Gleichgerrcht 2013) [104]. Les résultats montrent qu'il n'y a pas de différence entre les 2 groupes dans le cas « aiguillage » (80 % décident d'agir) mais une différence importante dans le cas « gros homme » : les autistes le poussent plus souvent (presque 40 % des personnes agissent contre 10 % pour le groupe de contrôle). Les auteurs ont fait le lien avec le déficit émotionnel caractéristique des troubles autistiques qui favoriserait la règle conséquentialiste. Les auteurs de cet article mettent en œuvre ici la méthode des différences : on a une population ayant une caractéristique particulière, ici les troubles autistiques, et on compare ses résultats à ceux d'une population témoin n'ayant pas cette caractéristique. La différence étant significative, elle peut être interprétée comme un indice de la corrélation entre cette caractéristique et le phénomène mesuré. Naturellement un tel indice ne porte que sur la corrélation et ne démontre en rien un lien causal, les troubles autistiques sont complexes et multiformes. Il conforte simplement l'idée qu'une influence des émotions sur le processus qui conduit à refuser de pousser le gros homme n'est pas mise en défaut par ce test sur les autistes. [10]

10. Remarquons que cet article fait écho à une analyse prétendant que Jeremy Bentham, un des fondateurs du conséquentialisme, aurait lui-même souffert de troubles autistiques (Sheeran 2006) [205].

En 2001, le développement de l'IRMf a permis que soit menée une nouvelle variante de l'expérience analysant les réponses neuronales des participants lorsqu'on leur soumet les deux scénarios (Greene 2001) [108]. Les auteurs concluent à l'activation des corrélats neuronaux des émotions dans le cas « gros homme » et pas dans le cas « aiguillage », confortant ainsi le résultat des expériences précédentes sur le rôle des émotions. La différence de réponse entre les deux cas est corrélée à l'activation neuronale de zones du cerveau elle-même corrélée avec l'activité émotionnelle. Cette corrélation semble aller dans le sens de l'existence de modes de raisonnements différents pour établir le jugement moral, un mode non émotionnel jauge les conséquences et établit sur cette base si l'action doit être entreprise, un autre mode, émotionnel, entreprend l'action principalement sur la base du ressenti émotionnel face à l'action. Avec toutes les précautions liées au caractère innovant de l'étude en 2001, l'article peut être interprété comme assimilant le conséquentialisme au premier mode de raisonnement moral, calculateur, et le déontologisme au second, émotionnel.

4.2.6 Mais le bilan reste en demi-teinte

Pour illustrer la difficulté à interpréter tous ces résultats expérimentaux, je vais m'appuyer sur un troisième exemple de variante du tramway fou. Dans un article de 2015, deux chercheurs grenoblois ont soumis au dilemme du tramway des participants recrutés dans des bars et ayant des taux d'alcoolémie variables (Duke et Bègue 2015) [78]. L'alcool est connu pour diminuer la rationalité et augmenter les réactions émotives. En suivant les analyses précédentes qui semblent lier le refus de pousser le gros homme aux émotions, on devrait s'attendre à ce que les personnes sous l'emprise alcoolique tendent vers ce refus. Or les relevés vont significativement dans l'autre sens : ces personnes alcoolisées tendent plus souvent que le groupe de contrôle à pousser le gros homme. Qu'en conclure ? Que l'alcool pousse à un raisonnement conséquentialiste ? Ou plutôt que le modèle simpliste opposant conséquentialisme et rationalité d'un côté à déontologisme et émotions de l'autre est inopérant ? Ou qu'il existe de nombreuses émotions dont les rôles moteurs et inhibiteurs peuvent varier ? ...

Sans entrer dans les arguments en faveur de chacune des interprétations possibles, on peut souligner qu'aucune des variantes précédentes du tramway fou de 1967 à 2015 n'avait mentionné le taux d'alcoolémie des participants en tant que critère intéressant. Faut-il en conclure que cette dernière variante compromet les interprétations ayant négligé cet aspect et que la complexité du comportement moral humain ne permet pas ce type de simplification ? Et si l'alcool peut ainsi intervenir dans le jugement moral, n'en sera-t-il pas de même de toute

une série de caractéristiques connues ou à découvrir ?

Les méthodes mises en œuvre dans ces trois expériences répondent à plusieurs critères reconnus comme étant de bonne pratique expérimentale dans les sciences de la nature : réplication, méthode des différences, cohérence et triangulation de l'opérationnalisation [11] des concepts. Reprenons successivement ces trois points. Premier point, les expériences ont été répliquées par plusieurs équipes indépendantes, les réplications ont sollicité des groupes différents de participants, dans des laboratoires de recherche différents et les réponses ont été interprétées par des chercheurs ayant des enjeux différents à le faire. Ces réplications sont donc de nature à augmenter la confiance dans les résultats présentés. Deuxième point, la méthode des différences qui vise à comparer deux sous-populations se distinguant par un seul paramètre, ici l'autisme, toutes les autres caractéristiques étant partagées, est la marque même de la démarche expérimentale telle qu'elle est définie en biologie par Claude Bernard. Enfin, troisième point, lorsqu'un phénomène est difficile à observer, ce qui est le cas des émotions, les chercheurs ont multiplié les approches, ils ont opérationnalisées ces émotions de plusieurs façons, par une déclaration de l'intéressé, par la méthode des différences et par leurs corrélats neuronaux détectés par IRMf, et cette multiplicité des approches, cette triangulation, permet que chaque observation conforte les conclusions des autres approches et augmente la confiance globale dans les résultats.

L'exemple du tramway permet également de construire un autre parallèle avec les sciences de la nature en ce qu'il construit un domaine d'expertise, la tramwaylogie, qu'on peut rapprocher de celui de la mouche du vinaigre (la drosophile) en biologie : une expertise qui permet d'étudier une large catégorie de phénomènes sur la base d'un même dispositif expérimental dont la maîtrise est ainsi partagée entre de multiples équipes. La concentration des expérimentations sur un même dispositif (le tramway en philosophie morale, la drosophile en biologie) facilite la mise en œuvre expérimentale et construit un environnement scientifique partagé permettant d'approfondir plus rapidement les recherches en bénéficiant des apports des autres recherches déjà réalisées. Naturellement, ce gain se paye ensuite par la nécessaire étape de vérification préalable à toute généralisation pour évaluer si les résultats obtenus sont spécifiques au dispositif expérimental choisi ou s'ils sont plus généraux.

On peut inférer de la liste des variantes proposées par les philosophes et psychologues en sa grande variété une méthode pratique pour écrire des articles de tramwaylogie qui ont une bonne chance d'être publiables. Première étape, choisir un article de psychologie qui mette

11. L'opérationnalisation est la construction d'une relation entre un trait psychologique non directement observable, les psychologues disent latent, et un phénomène observable supposé permettre de donner accès à ce trait et sera détaillée plus loin au chapitre 5.

en avant un phénomène intéressant ou inattendu. Exemple qui n'a pas, à ma connaissance, été encore ferroviarisé : les décisions de justice dépendent du petit déjeuner du juge (Danziger 2015) [62]. Deuxième étape, établir un scénario du tramway qui utilise ce phénomène. Exemple : faire venir les participants à jeun et soit leur servir un bon petit déjeuner soit leur donner un verre d'eau. Puis, troisième étape, dérouler le dilemme du tramway « aiguillage » et « gros hommes ». Dans notre exemple, le bon petit déjeuner sera, peut-être, une incitation à être conséquentialiste (ou l'inverse, ou neutre). Et le travail est ensuite de fournir une interprétation du résultat qui mette en rapport la différence de jugement moral et le phénomène observé. De cet exemple, certes un peu trop caustique, on comprend que face au déferlement d'articles de tramwaylogie rendus ainsi possibles, certains philosophes moraux aient eu un mouvement de recul et doutent de l'utilité de cette approche (Kahane 2015) [126]. Pour cet auteur, les chercheurs se sont focalisés sur des dilemmes extrêmes et, paradoxalement, s'appuient sur eux pour faire varier une multitude de paramètres sans que soit assurée la pertinence du paradigme de base. Cette approche ne permet pas de donner d'indication sur ce que serait l'influence de ces paramètres dans un contexte de choix moral effectués dans la vie quotidienne, et il propose de choisir des cas moins invraisemblables, par exemple dans le domaine de la santé, allant dans ce sens

On peut néanmoins retenir de l'exemple du tramway, en synthèse, qu'il permet au philosophe moral de disposer d'un outil connu de tous et facile d'emploi lui permettant de soumettre en parallèle à son intuition morale et à un sondage sur des populations de participants des scénarios significativement différents quant aux caractéristiques du scénario et quant aux conditions d'exercice du jugement moral. Je détaillerai dans une prochaine section, après avoir présenté les autres études de cas de philosophie morale expérimentale, ce que le philosophe pourra en attendre selon les différentes perspectives décrites au chapitre 1.

4.3 La surestimation du nombre de musulmans

4.3.1 Le cadre de l'étude

La philosophie expérimentale s'appuie de façon importante sur la réalisation de questionnaires à base de vignettes, des scénettes suggérées par des dessins ou de courts descriptifs d'une situation et comportant des questions auxquelles répondent les participants. Le philosophe doit ensuite interpréter les réponses à ces questionnaires pour construire des arguments dans son domaine d'intérêt. Cette méthode soulève de multiples difficultés et pour les

appréhender concrètement chaque participant au séminaire de philosophie expérimentale animé par Brent Strickland à l'ENS Ulm, a pour objectif de construire un tel questionnaire, de le dérouler et d'en présenter les résultats.

Dans ce cadre [12], nous nous sommes intéressés à un phénomène présenté en 2016 dans la presse : lorsqu'ils sont interrogés dans un sondage, les européens, et tout particulièrement les Français, surestiment le nombre de musulmans présents dans leur pays [13]. Mon objectif avec cette étude de cas est de montrer d'une part les nombreuses difficultés qui parsèment le chemin de l'utilisation des questionnaires et, d'autre part, de soulever la question non moins difficile de ce qui justifie de s'intéresser à un tel sujet. On peut en effet penser qu'un questionnaire sur un sujet sensible d'actualité crée lui-même pour partie le problème qu'il se donne à résoudre.

Je vais dans un premier temps décrire la problématique posée à partir de l'article du Guardian, puis détailler l'enquête menée en 2017 et ses résultats, et ensuite présenter plusieurs interprétations de ces résultats et marquer mon scepticisme quant à l'utilité scientifique de ce type de travail qui résulte, ce sera ma proposition, d'un ensemble cumulatif de biais de publications appuyé sur un phénomène de base de faible consistance.

Sur le tableau ci-dessous [14], on peut lire que, en moyenne, les Français estiment qu'il y a 31 % de musulmans en France alors que, selon l'organisme Pew Research, le nombre réel était de 7,5 % en 2010. L'estimation de 31 % provient d'un sondage réalisé par Ipsos [15] et les chiffres dits réels, dont le 7,5 % pour la France, d'enquêtes propres à chaque pays [16].

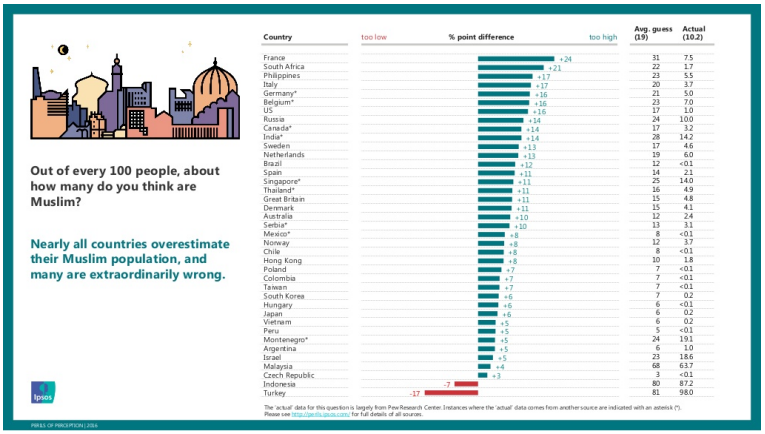

FIGURE 4.2 – La surestimation du nombre de musulmans (source Ipsos 2016)

12. J'ai mené cette étude en collaboration avec Aurelien Fermo au premier semestre 2017. Les données analysées dans cette étude sont à disposition sur `https://dropsu.sorbonne-universite.fr/s/oj2e3Wcarg7Rq8a`. L'étude se poursuit en 2020 en collaboration avec Aurelien Fermo et Antoine Marie
13. L'article relayant ce constat : `https://www.theguardian.com/society/datablog/2016/dec/13/europeans-massively-overestimate-muslim-population-poll-shows`
14. Source : `http://perils.ipsos.com`
15. Selected countries, Ipsos Mori Perils of Perception 2016
16. Pew Research / De Standaard (Belgium) / Statistics Canada...

Le signataire de l'étude est Bobby Duffy (bobby.duffy@ipsos.com). Ce collaborateur d'Ipsos a régulièrement proposé dans la presse ce type d'enquête. La surestimation était déjà présente, mais moins prononcée, dans une enquête sur le nombre d'immigrés (et non de musulmans) menée en France en 2014.

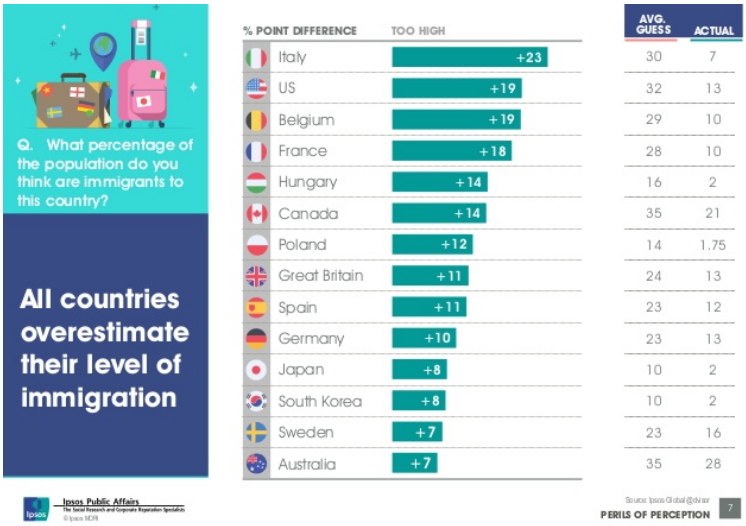

FIGURE 4.3 – La surestimation du nombre d'immigrés (source Ipsos 2014)

La surestimation du nombre de musulmans annoncée dans l'étude d'Ipsos fait passer de 7,5 % de la valeur réelle anticipée à 31 % de valeur moyenne annoncée par les sondés. Si elle est confirmée, on est en droit de se demander quelle est la provenance d'une telle erreur. On peut pour y répondre tester par sondage les multiples explications envisageables suivantes, inspirées de divers biais cognitifs que les psychologues ont identifiés comme pouvant entacher ce type de déclaration.

Le biais de disponibilité nous conduit à considérer comme important un phénomène qui nous est disponible car fréquemment mentionné dans notre environnement. Le cheminement implicite serait alors schématiquement le suivant : « Le sujet de la présence des musulmans est important, car je vois des articles tous les jours dans la presse, c'est donc qu'il y en a beaucoup, disons 30 % »

Le biais pragmatique consiste à donner une réponse non en fonction de ce que l'on sait (ou croit savoir) mais en fonction d'un objectif à atteindre en donnant la réponse. Le raisonnement schématique est alors : « Je n'aime pas les immigrés et il y en a partout, il faut faire quelque chose, et si je dis 30 %, peut-être cela convaincra également mon interlocuteur et conduira les autorités à prendre une décision. »

Le biais idéologique consiste à répondre en fonction d'une idéologie particulière, à donner du poids à des inventions confortant les thèses de cette idéologie. Un biais idéologique qui

se traduit par un raisonnement conformiste : « Je suis militant, je crois que nos problèmes viennent des musulmans, mes amis et moi employons toujours cet argument et j'y crois, ils sont trop nombreux : 30 % »

On peut également penser que, sans qu'il y ait biais à proprement parler, il peut y avoir erreur de jugement. Le sondé est en possession d'une croyance qu'il pense, à tort, justifiée : « J'ai vu quelque part, à une source de confiance, qu'il y avait 30 % de musulmans en France ». Pour mémoire, Jean-Marie Le Pen avait effectivement donné ce chiffre de 30 % dans une déclaration [17] et on peut se souvenir du chiffre sans se souvenir de la source et de sa connotation partisane.

Autre source d'erreur, la sémantique. Les catégories comme « musulman » ne sont pas précisément définies et on peut avoir des acceptions du terme qui correspondent à des estimations très différentes si les réponses à l'enquête et les chiffres réels ne sont pas sur une même base. Qu'est-ce qu'un immigré ? Un nouvel arrivant de première génération, ou de la deuxième génération, ou tous les habitants qui ne sont pas français de souche avec 4 quartiers ? Qu'est-ce qu'un musulman ? Un pratiquant régulier, quelqu'un qui se dit musulman, quelqu'un qui vient d'un pays musulman ?...

Ensuite, autre exemple de voie d'exploration des explications possibles, on peut évoquer le biais de plus forte mémorisation des thèmes négatifs. Les musulmans étant souvent reliés dans la presse à des problèmes générateurs d'anxiété comme l'insécurité, l'extrémisme... les sondés gardent l'importance de ce sujet en mémoire et donnent un chiffre important. Le caractère de négativité renforce le biais de disponibilité.

Enfin, on sait depuis les travaux de Kahneman [127] que les humains ont des capacités cognitives très limitées en statistiques. On peut parler de biais de dyscalculie statistique. L'analyse de la surestimation devient alors très délicate puisqu'elle peut simplement résulter d'une heuristique sans lien précis avec les valeurs numériques en cause. Par exemple « Je n'ai aucune idée sur le nombre de musulmans, mais il faut que je réponde, c'est sûrement moins de 50 car la France est un pays catholique, disons 30 % ». On peut ainsi envisager que cette erreur ne soit pas reliée à la question des immigrés mais soit au contraire un fait général pour toute question sociale. Le même Bobby Duffy auteur de l'article de 2016 suggère d'ailleurs une telle conclusion dans un article antérieur de 2013 « Les Britanniques ont faux sur tout, voici pourquoi » [18]. Et le même analyste, dans un autre article de 2013 [19] suggère

17. "Entre 15 et 20 millions de musulmans vivent en France", a déclaré Jean-Marie Le Pen dans une interview à un journal russe "Komsomolskaïa Pravda" le jeudi 15 janvier 2015
18. http://theconversation.com/british-people-are-wrong-about-everything-heres-why-16018
19. http://theconversation.com/hard-evidence-how-does-the-public-feel-about-immigration-19220

qu'il y aurait bien une corrélation historique globale entre les variations de l'estimation des migrations et les flux réels, mais que les erreurs sont massives à un instant donné en valeur instantanée et, également, relativement à la répartition entre types d'immigration. Ces deux articles penchent donc plutôt vers un certain scepticisme quant à la pertinence des résultats numériques ainsi donnés par les sondés, et donc sur la validité des explications trop entachées des préjugés de chacun sur un phénomène supposé qui rentre trop bien dans le paysage intellectuel de l'interprétant.

Le point de départ de notre étude était donc, sur la base de l'article du Guardian, d'une part de tenter de confirmer la surestimation du nombre de musulmans telle qu'elle apparaît dans cet article, et d'autre part d'explorer quels facteurs influent sur les réponses des sondés pour arriver à une erreur finale d'une telle ampleur.

4.3.2 Le questionnaire de mai 2017

Le problème posé de la compréhension de la surestimation du nombre de musulmans ou d'immigrés est potentiellement porteur d'enjeux importants : comment peut-on prendre démocratiquement de bonnes décisions si la majorité des citoyens a une vue aussi erronée de la réalité ? Mais avant d'élaborer des stratégies de clarification de cet enjeu, il faut vérifier que le phénomène existe et pour cela répliquer les résultats du sondage. L'étude menée dans le cadre du séminaire XPhi a pour premier objectif de vérifier si nous pouvons confirmer cette surestimation. Le second objectif est d'explorer certaines des différentes causes possibles du phénomène s'il existe, et principalement les biais idéologiques liés à l'orientation politique.

Tout d'abord, nous avons recherché les données sur l'immigration facilement accessibles à tout un chacun par une simple requête sur Google. Les résultats [20] confirment les chiffres dits réels publiés par The Guardian : 8 % d'immigrés et 19 % avec les enfants d'immigrés. De même il existe des données sur les religions facilement accessibles [21] : il y aurait en France 4 710 000 musulmans soit 7,5 % de la population. Selon une autre source IFOP, il y aurait en France 64 % de catholiques, 3 % de protestants et 28 % de personnes se déclarant sans religion, ce qui ne laisse que 5 % pour l'ensemble des autres religions dont les musulmans. On sait par ailleurs [22] qu'il y a de l'ordre de 1,6 millions de musulmans régulièrement pratiquants, soit 2,5 % de la population. Les trois chiffres de 2,5 %, 5 % et 7,5 %, tous facilement accessibles au grand public, ne sont pas cohérents entre-eux, mais cela situe le nombre dans une fourchette large, disons de 2 à 10 %, qui reste dans tous les cas très inférieure à l'estima-

20. Wikipedia : Données statistiques sur l'immigration en France
21. Chiffres de PEW 2010 `http://www.globalreligiousfutures.org/religions/muslimsÂă[archive]`
22. Organisation du culte musulman.

tion du sondage Ipsos qui est, rappelons-le, de 31 %.

On peut néanmoins faire deux remarques à partir de ces constats globaux facilement accessibles. Première remarque, la surestimation est moins clairement avérée pour l'article sur l'immigration que pour le nombre de musulmans. En effet, selon la définition que l'on donne de la catégorie des immigrés, et si on compte la seconde génération comme incluse dans cette catégorie « immigrés », un ordre de grandeur de la valeur réelle de 20 % est acceptable et cette valeur n'est plus très éloignée des valeurs spontanément estimées. Il conviendra de garder cette remarque en tête car je montrerai plus loin que les résultats du sondage indiquent qu'une confusion entre les deux notions est possiblement à l'œuvre. Seconde remarque, même si la surestimation du nombre des musulmans est confirmée, elle n'est pas de l'ampleur annoncée dans l'article initial. On peut alors s'interroger sur l'importance relative qu'il y a à multiplier par 6, de 5 à 31 %, ou multiplier à par 3, de 10 à 31 % une estimation. Il semble que cela puisse relever de mécanismes psychologiques différents, hypothèse qu'il faudrait évaluer. Je ne reviendrai pas sur ce point qui n'est pas éclairé par les résultats de nos questionnaires.

Un premier questionnaire a été établi et réalisé en mai 2017 de façon à répliquer l'évaluation du pourcentage de musulmans en France et tenter une analyse de corrélation avec les données politiques. Le sondage a été établi en utilisant le logiciel Qualtrics et en recrutant les sondés sur le site Amazon Mechanical Turk contre une rémunération minime [23]. Le sondage a eu lieu entre les deux tours de l'élection présidentielle qui a vu l'élection du président Macron. La cible de 200 personnes a été rapidement atteinte en 48 heures. 198 questionnaires ont finalement été exploités, les deux réponses écartées étant incomplètes. La structure du questionnaire était la suivante :

— Consentement et vérification de l'attention
— À combien estimez-vous le nombre de musulmans pour 100 personnes en France ?
— Question complétée et alternée aléatoirement avec l'estimation du nombre d'immigrés
— Questions démographiques : Diplôme, Sexe, Age
— Questions politiques : Orientation politique et religieuse, vote 1er tour, intention de vote 2nd tour

Les résultats du sondage ont été exploités de mai à juillet 2017 et présentés au séminaire XPhi de l'ENS Ulm en vue de préparer une seconde étude complémentaire sur 2017-2018. Les principaux résultats de ce premier sondage sont repris ci-dessous. [24]

23. L'étude a été validée par le comité d'éthique de l'ENS et financée par le budget de l'ENS Ulm
24. Un troisième sondage est actuellement en préparation en collaboration avec Aurelien Fermo, Antoine Marie et Brent Strickland et devrait donner lieu à un projet d'article en 2020.

En préalable, les données démographiques ne font pas apparaître de biais dirimant dans la constitution de l'échantillon des sondés, exiger mieux serait incohérent avec la faiblesse de la taille de l'échantillon. Les opinions politiques sont également bien représentées et sans surprise en regard du résultat du second tour de l'élection présidentielle :

— Répartition de l'échantillon par sexe : 84 femmes 114 hommes,

— Répartition de l'échantillon par tranches d'âge : moins de 30 ans : 61 participants / de 31 à 40 ans : 62 participants / de 41 à 50 ans : 41 participants / plus de 50 ans : 34 participants.

— Prévisions pour le 2ème tour : Macron 50 % / Le Pen 13 % / Indecis 7 % / Abstention 29 % / Ne veulent pas répondre 2 %

Par ailleurs, nous avons vérifié que la déclaration de vote au 1er tour et l'intention annoncée pour le 2ème tour étaient politiquement cohérentes. Nous avons également vérifié que les critères de diplôme, de sexe et d'âge n'apportaient pas d'information significative, sans que soit écartée la possibilité d'une influence de ces facteurs qui pourrait être statistiquement détectable avec un échantillon de plus grande taille.

Il est intéressant de remarquer que la question sur la religion nous permet d'avoir une estimation directe du nombre de personnes se disant musulmanes au sein de notre échantillon. Les résultats du sondage sont les suivants :

— Sans religion : 46 %

— Catholiques : 37 %

— Musulmans 9 %.

Cette question sur l'orientation religieuse nous permet d'estimer à nouveau le nombre de musulmans qui reste pour notre échantillon dans la fourchette large de 2 à 10 % proposée plus haut. En revanche, les réponses ne confirment pas du tout l'estimation du nombre de personnes se disant sans religion qui est beaucoup plus élevé ici que le chiffre de 28 % retenu dans les sources visibles du grand public citées plus haut. Comme l'appartenance religieuse est sans conséquence sur les principaux résultats du sondage, ce point n'a pas été creusé plus avant.

Enfin, dernière remarque préalable, les estimations des nombres d'immigrés présentent une répartition très analogue à celle des estimations de nombre de musulmans. Le nuage de points immigrés vs. musulmans ci-dessous montre une certaine symétrie autour de la première bissectrice. Ce schéma général se retrouve dans les moyennes identiques de 19,7 % pour le pourcentage de musulmans et 19,3 % pour celui des immigrés ainsi que dans le fait que ces résultats sont indifférents à l'ordre des questions (musulmans puis immigrés) ou (im-

migrés puis musulmans). On peut inférer de ces résultats que les personnes donnent la même réponse à l'une ou l'autre question. Tout se passe comme si les sondés considéraient les deux termes comme synonymes en regard de la question posée. Cette remarque va dans le sens d'un amalgame entre les deux populations par les sondés, ce qui est bien sûr conceptuellement et statistiquement erroné, mais suggère une piste d'explication de la surestimation du nombre de musulmans par la confusion entre « musulman » et « immigré ».

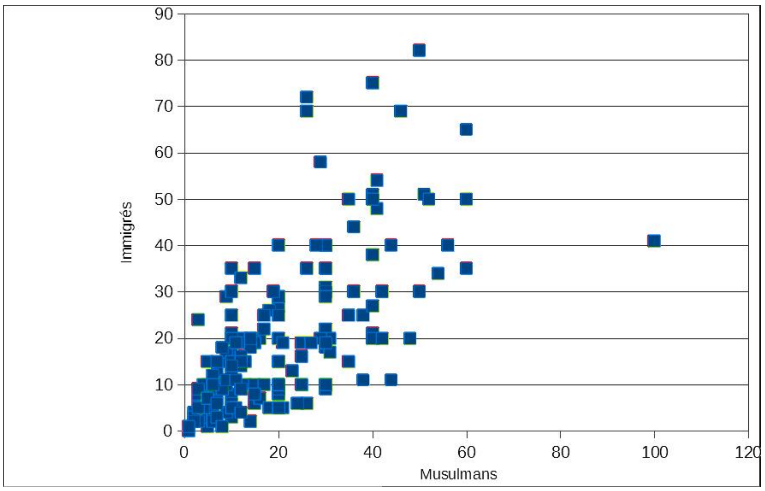

FIGURE 4.4 – Corrélation entre estimations du nombre d'immigrés et de musulmans

Après ces remarques préalables, examinons les premiers résultats que l'on peut obtenir de ce questionnaire.

4.3.3 Premiers résultats

L'exploitation des réponses au questionnaire apporte des premières éléments et soulève de nombreuses questions. Commençons par les éléments de réponse sous la forme de cinq résultats.

Tout d'abord, premier résultat recherché, la surestimation apparente est confirmée. Les sondés répondent en moyenne qu'il y a environ 20 % de musulmans en France, très loin de la fourchette haute de la valeur réelle qui est inférieure à 10 %. Néanmoins, cette évaluation est d'un tiers inférieure aux 31 % de l'article d'Ipsos. Cette différence de 11 points est importante, sans que l'on puisse l'expliquer avec les éléments à disposition [25]. Interrogé sur ce point, l'auteur de l'article de l'Ipsos a indiqué qu'il avait obtenu 31 % par deux fois, mais qu'il ne faisait pas de suivi longitudinal de ce résultat et qu'il ne pouvait donc contribuer à analyser cette différence [26] dont la provenance pourrait être simplement une certaine sensibilité au

25. L'estimation à 20 % a été confirmée en 2018 par une étude de suivi menée par Aurelien Fermo.
26. Mail privé de Bobby Duffy à l'auteur du 21 juin 2017

moment de l'enquête en regard de l'actualité.

Deuxième résultat, la relation de la surestimation à l'orientation politique est avérée. L'extrême droite fait pencher vers la surestimation. Le schéma global est le suivant : quelques personnes, principalement de gauche, donnent une estimation comprise dans la fourchette de la valeur réelle (au plus large, de 2 à 10 %), une très large majorité de sondés donne une estimation autour de 20 % et quelques personnes, principalement d'extrême droite, donnent des estimations outrageusement exagérées, dépassant 50 %. Le diagramme ci-après visualise cette répartition, la ligne supérieure donne la répartition des 198 réponses selon l'intention de vote. Les lignes inférieures redonnent cette répartition par tranche d'estimation de 10 %. Les trois lignes supérieures correspondant à des réponses de plus de 50 % de musulmans sont principalement constituées de personnes disant avoir voté pour Le Pen. Les lignes inférieures, numériquement les plus importantes, reflètent des répartitions par vote conformes à la répartition globale sur l'échantillon.

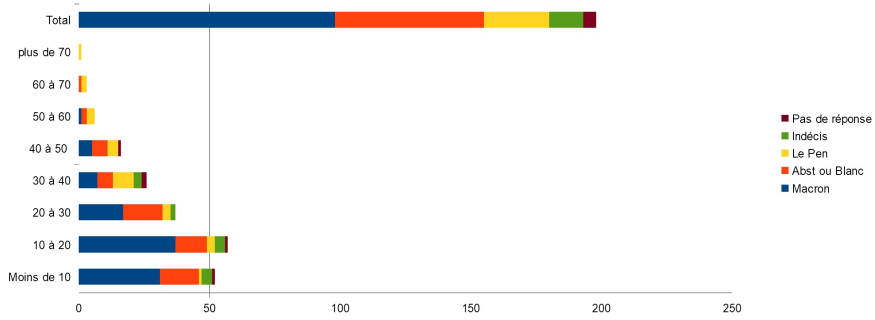

FIGURE 4.5 – Nombre de réponses par tranche d'estimation selon l'intention de vote

Troisième résultat, la surestimation subsiste pratiquement inchangée si on enlève les réponses des partisans de Marine Le Pen. Ceci est lié au fait que si on enlève les quelques réponses extrêmes, les réponses des électeurs de droite ne sont pas significativement différentes de celles des autres électeurs.

Nombre Réponses	Vote	Quartile	Médiane	Quartile
25	Le Pen	20	30	40
173	Hors Le Pen	8	14	25
198	Total	8	15	30

Quatrième résultat, les participants musulmans contribuent à la surestimation avec une moyenne de 20 % proche de celle de l'ensemble de l'échantillon. Ce résultat peut paraître surprenant si on considère que les musulmans devraient mieux estimer le nombre de leurs coreligionnaires. Il l'est moins si on pense que la réponse au sondage est le résultat complexe de multiples biais qui s'appliquent aussi bien aux musulmans qu'aux non-musulmans car ils sont peu dépendants du rapport de chaque sondé à l'objet de la question posée.

Cinquième résultat, les sondés répondent préférentiellement avec des valeurs rondes en dizaines de pourcents. Le diagramme ci dessous donne le nombre de réponses par tranches de 5 %, il fait apparaître une forte oscillation liée au fait que les tranches qui contiennent une valeur ronde de dizaines, donc une sur deux, sont surreprésentées. Deux interprétations très différentes de ce phénomène sont possibles. Première interprétation, la préférence pour les valeurs rondes est un artefact lié à la présentation du questionnaire (échelle glissante), seconde interprétation, la préférence est réelle et marque la réaction du sondé qui, par exemple, ne connaissant pas la réponse exacte juge préférable d'annoncer un ordre de grandeur. Une étude complémentaire serait nécessaire pour lever cette ambiguïté.

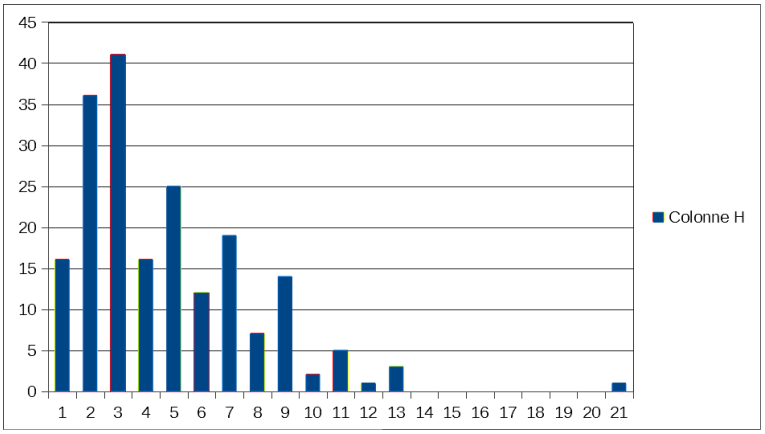

FIGURE 4.6 – Nombre de réponses par tranche de 5 %

Les cinq résultats ci-dessus sont obtenus sans recours à des calculs complexes. Ils résultent modestement de statistiques descriptives de base, des moyennes, des quartiles et des représentations graphiques donnant à voir les données. Ces outils de base sont ici les plus pertinents du fait de l'objectif exploratoire de ce questionnaire, du fait également qu'il est facile d'avoir une vision globale de 198 points sur quelques critères et, enfin, parce qu'un calcul plus sophistiqué de type p-value donnerait une image de précision illusoire alors que les ordres de grandeur même du phénomène observé ne sont pas confirmés, avec une estimation inexpliquée du pourcentage de musulmans multipliée par 1,5 de 20 % pour notre questionnaire à 31 % pour celui d'Ipsos.

Cette très grande différence entre les 31 % de l'étude initiale et notre résultat à 20 % peut être interprétée de plusieurs façons. Première hypothèse, l'estimation serait dépendante de l'actualité. Un attentat récent attribué à un extrémiste musulman ou la proximité des printemps arabes pourraient, par exemple, avoir influé l'évaluation d'Ipsos faite en 2016. Comme ce contexte n'est pas apparent dans les questionnaires, l'hypothèse reste ouverte. Seconde hypothèse, un des biais agit différemment pour une enquête de l'Ipsos et pour une enquête académique. Explorer cette piste supposerait de systématiser des comparaisons, ce que nous n'avons pas réalisé ici. Une différence de formulation serait potentiellement une autre hypothèse, nous n'avons pas eu accès au questionnaire en français utilisé par Ipsos et n'avons pas pu le vérifier, mais cette hypothèse semble peu probable si on retient comme formulation de la question utilisée par Ipsos le libellé qui apparaît dans l'article du journal The Guardian. Enfin, dernière possibilité, les échantillons sont de natures différentes. Là encore nous n'avons pas le détail de la constitution du panel de l'Ipsos pour le comparer à celui de Amazon Mechanical Turk. Intuitivement, la première hypothèse apparaît plus crédible du fait des attentats de 2015 mais les autres hypothèses ne sont pas écartées par les résultats disponibles.

4.3.4 Les interprétations possibles de la surestimation et les interrogations soulevées

En synthèse, l'étude a permis de confirmer le phénomène de la surestimation apparente, même s'il n'a pas, et de beaucoup, l'ampleur annoncée dans l'article initial.

Elle permet également de constater des réponses exagérément surestimées de certains participants d'extrême droite. On peut considérer que ces participants savent qu'il n'y a pas plus de 50 % de musulmans en France et il semble donc approprié de qualifier ces réponses de surestimation volontaire. On peut interpréter cette surestimation volontaire comme un signal de provocation dont le destinataire est le chercheur enquêteur, perçu comme un représentant des institutions. Cette interprétation n'est pas une déduction directe du contenu des réponses au questionnaire, mais il est difficile de trouver une explication alternative à des évaluations de 60, 80 ou 100 %.

En revanche, et pour l'ensemble des personnes, tous bords politiques confondus hors ces extrêmes, la surestimation ne présente pas de tendance claire en fonction de l'orientation politique. Le biais idéologique est donc à écarter comme raison de la surestimation globale qui subsiste quand on enlève l'extrême droite. Et, plus profondément, il n'est pas certain

qu'il s'agisse d'une surestimation volontaire, au sens du signal donné ci-dessus, on devrait plus prudemment parler de mésestimation. En effet, plusieurs indices issus des réponses au questionnaire vont plutôt dans le sens d'une absence de rapport entre ces réponses et la situation réelle :

— L'évaluation est la même pour le nombre d'immigrés et pour le nombre de musulmans, sans lien avec la réalité statistique.

— Les valeurs rondes à la dizaine sont privilégiées, marquant l'absence de précision de la connaissance du pourcentage réel.

— L'évaluation est la même pour les participants se disant musulmans.

— Des questionnaires sans lien avec les musulmans ou les immigrés donnent lieu à des mésestimations analogues, comme l'article de 2013 de Bobby Duffy le rappelle.

— Enfin, la question porte sur une catégorie peu précise, les « musulmans » qui peut être comprise très diversement comme l'ensemble des personnes venant des pays musulmans et leur descendance. Ce flou conceptuel pourrait contribuer à la dispersion des estimations.

En prolongement de cette interrogation sur la réalité d'une surestimation, on peut ouvrir plusieurs pistes d'hypothèses que des questionnaires ultérieurs pourraient tenter d'affiner. Première piste, que les indices ci-dessus favorisent, il n'existe pas de surestimation mais, simplement, une mésestimation qui est un phénomène commun à toute question pour laquelle les sondés n'ont pas d'idée de réponse. Leur réponse, sollicitée par la situation de questionnaire rémunéré, est alors établie selon un processus à étudier, partant par exemple des émotions portées par le sujet de la question, et ce processus conduirait à une valeur numérique donnée qui ne peut donner lieu à aucune comparaison rationnelle avec des valeurs réelles inconnues du participant donc sans incidence dans le processus de choix de la réponse. [27]

Deuxième piste, la surestimation non volontaire, ou la mésestimation, serait néanmoins un indicateur intéressant du ressenti des personnes sur le sujet soumis à l'enquête, la présence des musulmans, et le questionnaire pourrait être enrichi pour évaluer l'ensemble des opinions relatives aux musulmans au delà de la seule orientation politique pour analyser si, indépendamment du nombre réel de musulmans, l'estimation qu'elles en donnent peut être corrélée avec ces opinions. Notre étude a montré l'absence de corrélation avec le positionnement politique électoral mais il se pourrait que la corrélation soit plus forte si on dispose

27. Une étude de suivi a été menée en 2018-2019 par Aurelien Fermo. Une des questions complémentaires posées a été « combien de musulmans pensez vous qu'il serait souhaitable d'avoir en France ? » et les réponses à cette question tendent à conforter l'hypothèse de l'ignorance : certains sondés d'extrême droite ont donné, comme dans notre sondage, des valeur supérieures à 50 % pour l'estimation actuelle et ils ont ensuite donné 25 % pour le nombre souhaitable. Ils ont ainsi divisé par 2 pour donner le signal qu'il y en a trop mais ont en fait proposé d'en augmenter le nombre par rapport au chiffre réel !

d'une question plus directement liée, par exemple, à l'acceptation ou au refus de la diversité culturelle.

Enfin, troisième piste, conservant l'hypothèse qu'il y a bien surestimation éventuellement inconsciente et non volontaire, enrichir le questionnaire en amenant les sondés à affiner progressivement leur estimation. Une piste pourrait être d'annoncer que la rémunération du participant sera plus importante si l'évaluation donnée est plus proche de la valeur réelle. Une autre piste pourrait être de demander d'évaluer la répartition de la population entre les diverses religions, catholiques, sans religion, musulmans et divers, le total étant imposé à 100. L'objectif serait de pousser le sondé à une réflexion rationnelle plus poussée et de comparer cette seconde réponse, tous comptes faits, à une première réponse plus spontanée.

Chacune de ces pistes pourra donner lieu à des questionnaires complémentaires qui viendront enrichir la réflexion et mieux qualifier le processus par lequel les Français interrogés arrivent, en moyenne, à une estimation de 20 % pour un nombre réel de musulmans inférieur à 10 %. Mais, outre l'analyse du questionnaire et des réponses ainsi enrichies, on est également en droit de s'interroger sur la raison d'être d'un tel travail de la part d'Ipsos, enquête qui a conduit à l'article initial dans The Guardian à l'origine de notre propre étude. Il est alors important de souligner que l'étude initiale d'Ipsos a été faite sur fonds propres, elle n'a pas été facturée à un client [28]. L'intérêt objectif de ce type d'étude est pour Ipsos, entreprise à but lucratif, d'obtenir une couverture presse de nature à mettre en valeur sa capacité à mener des études marketing qui seraient, elles, facturées. Il s'agit donc d'une opération croisée lors de laquelle le sondeur bénéficie d'un espace de communication gratuit et, en échange, fournit à l'organe de presse un article de nature à attirer ou satisfaire un lectorat. Disons que l'article doit être intéressant pour le lecteur, pour qualifier ce que la presse et le sondeur, tous deux intéressés au sens économique, en attendent.

Une étude de la surestimation du nombre de chauves en France ou du nombre de voitures bleues circulant la nuit aurait à l'évidence moins de chances d'être publiée car elle n'apporterait pas de bénéfice au journal, bénéfice échangeable contre une publication gratuite. En conséquence, ces études ne seront jamais menées par Ipsos, et nous ne saurons jamais si elles auraient donné lieu aux mêmes phénomènes de mésestimation que les sujets plus intéressants pour le lectorat, et donc plus facile à faire publier, sur les immigrés ou les musulmans.

Pour clarifier ce point par exagération, faisons une expérience de pensée. Supposons qu'une erreur commune fasse qu'une partie de la population pense que 2+2 = 5, et que, n'ayant ja-

28. Communication personnelle avec Bobby Duffy.

mais posé la question de cette façon générale, nous réalisions un sondage pour savoir combien font 2 musulmans + 2 musulmans. Nous aurions alors une partie des sondés qui répondrait 5 musulmans. Dans ce cas que pensez-vous que The Guardian publierait ? Que la question sur les musulmans n'a aucun intérêt, car il s'agit simplement d'une erreur commune appliquée à ce cas particulier, ou plutôt un article au titre alléchant « L'arithmétique est incompatible avec l'Islam » ? La lecture de ce journal ne laisse aucun doute, la seconde hypothèse est la bonne et ce constat alimente les positions sceptiques quant à la pertinence des publications sur les estimations du nombre d'immigrés ou du nombre de musulmans qui n'auraient probablement pas été publiées si le résultat avait été correct : c'est parce qu'il est surestimé qu'il est publié car il attire le lectorat.

Cette argumentation ne prouve naturellement pas que la surestimation du nombre de musulmans par les Français sondés n'existe pas, elle prouve simplement que si l'étude a été réalisée, c'est parce que Ipsos savait qu'elle serait publiée car le rejet des musulmans est un sujet de société intéressant les lecteurs, donc la presse. Ce processus conduit ainsi à jeter un doute sur la signification profonde d'un résultat qui n'existe que parce qu'il est croustillant et non pour sa valeur épistémique.

On peut ensuite aller un pas plus loin et suggérer que les sondés, et en particulier ceux qui ont un point de vue partisan, joueraient, consciemment ou non, de ce processus en comptant sur la publication d'opinions délibérément sans lien avec la réalité, comme le suggèrent les valeurs supérieures à 50 % recueillies dans notre questionnaire. La surestimation, pour la partie venant des positions partisanes des sondés d'extrême droite, serait alors interprétable comme le résultat d'un phénomène de résonance entre d'une part des personnes exprimant un rejet par des chiffres aberrants et, d'autre part, des sondeurs et journalistes exploitant économiquement ces chiffres, dont les chiffres aberrants, pour générer de l'audience. Une telle interprétation n'est bien sûr en aucune façon directement déduite des résultats du sondage que nous avons mené, mais elle est pleinement compatible avec eux et pourrait être prise en compte pour imaginer de nouvelles variantes des sondages à mener dans le futur.

4.3.5 Du questionnaire en philosophie expérimentale

Prenant de la distance par rapport au débat sur les musulmans, cette incursion dans le concret du monde des sondages nous permet de faire plusieurs observations qui, bien qu'issues de ce cas particulier, évoquent des arguments souvent repris pour questionner l'utilisation de cet outil en psychologie expérimentale. Je résume ci-après les enseignements sug-

gérés par l'étude du cas des questionnaires portant sur la surestimation des musulmans. Je constate tout d'abord que le vocabulaire est imprécis et que clarifier l'incidence de cette imprécision est plus que difficile sur la seule base des réponses au questionnaire. L'interprétation des résultats en est rendue délicate. Je souligne ensuite l'ampleur des connaissances du contexte implicitement mobilisées dans une telle étude, en contraste frappant avec une formulation des résultats qui apparaît proche de celle les articles scientifiques. Et enfin, je mets en lumière la logique de publication d'articles qui préside à l'ensemble de la démarche.

Le vocabulaire utilisé dans le sondage est imprécis, le terme « musulman » utilisé dans la question peut être compris de plusieurs façons par les sondés et la réponse ne permet pas au sondeur de choisir parmi ces différentes possibilités celle qui a été retenue par le participant. Le sondeur n'a que le résultat final du processus psychologique qui conduit le sondé à répondre ainsi.

L'interprétation du sondage en tant que « surestimation » suggère l'hypothèse que le sondé connaîtrait une valeur réelle et en donnerait une autre délibérément (ou inconsciemment) exagérée en poursuivant un objectif qu'il faudrait alors débusquer. Pour vérifier cette hypothèse, il faudrait accéder aux deux valeurs, celle que le sondé pense vraie et celle qu'il extériorise. Toutefois, le sondage ne donne pas accès à la première valeur qui serait la valeur vraie supposée connue, elle ne donne accès qu'à l'estimation extériorisée. Le sondeur se trouve alors confronté à plusieurs configurations contradictoires possibles qui, toutes, conduisent à la même valeur exprimée. Le sondé peut croire en une valeur « vraie » plus ou moins erronée, par excès ou par défaut, ou plus ou moins exacte et exprimer une valeur « estimée » avec ou sans altération à la baisse ou à la hausse par rapport à sa croyance. Et encore, le sondé peut n'avoir aucune croyance particulière quant à la valeur « vraie » et construire une réponse « estimée » sans lien avec la situation réelle. Il est très difficile au sondeur de reconstruire simultanément la valeur « vraie », objet de la croyance du sondé, et la stratégie d'altération que le sondé peut adopter dans le cadre de sa réponse, en ne disposant que du résultat final. Seuls quelques cas limites comme ce militant annonçant 100 % de musulmans en France peuvent être analysés avec une relative assurance.

La présentation de notre travail prend la forme habituelle des articles scientifiques en trois temps, le premier exposant un problème avec plusieurs hypothèses possibles, le second avec la description d'un questionnaire suggérant un protocole expérimental et le troisième avec l'exploitation des résultats et une ouverture vers leurs conséquences sociétales. Cette présentation est à première vue justifiée car elle reprend bien le déroulement de l'exercice mené mais elle peut être également jugée abusivement rationalisée dans la mesure où le

phénomène réellement mis en œuvre est un processus très complexe qui n'est qu'effleuré par l'analyse des réponses au questionnaire. Un tel programme de recherche pourrait être décrit de façon globale plus complète, itérative et moins linéaire avec toutes les étapes qui conduisent au choix du thème du sondage, puis à l'établissement du questionnaire, à sa diffusion à un panel d'internautes disponibles car intégrés dans une organisation commerciale, à la compréhension par le sondé de la question posée et à la psychologie de l'élaboration de la réponse qu'il propose, au recueil de cette réponse par le sondeur et, enfin, à la complexe étape de dépouillement et de sélection des résultats puis à l'interprétation en vue d'une formulation des résultats de l'étude. Chacune de ces étapes est en elle-même complexe et peut donner lieu à des biais qu'il conviendrait de prévenir en les documentant très précisément de façon à permettre au lecteur d'évaluer la pertinence des interprétations des résultats statistiques du questionnaire.

J'ai en particulier commencé à aborder un point épineux : qu'est-ce qui fait que la présente étude de cas existe au sein de mon travail de doctorant ? L'enchaînement est ici accessible car relativement simple :

— Ipsos publie régulièrement des articles de société en collaboration avec la presse dans le cadre d'un échange équilibré de bons services : les articles sont intéressants pour le lectorat, donc pour l'organe de presse, et Ipsos obtient une couverture presse gratuite au prix de la réalisation d'une étude sur fonds propres.

— Les institutions académiques, professeurs et étudiants confondus, ont intérêt à suivre ces publications, car elles témoignent de préoccupations sociétales et que, pragmatiquement, elles peuvent donner lieu à des travaux d'étudiants publiables. Les psychologues ne peuvent se désintéresser des débats sur les musulmans parce qu'ils sont importants, naturellement, pour eux-mêmes et pour leurs potentielles conséquences sur les politiques publiques, mais également parce qu'ils donnent lieu à des articles (plus) facilement publiables. Le fait que la France soit classée en tête des mauvais élèves dans l'article du Guardian ne peut que rajouter à la curiosité spontanée soulevée par cet article.

— Parmi différents sujets d'étude, je choisis celui-ci car le contexte m'en est connu par les nombreuses publications disponibles, et qu'il me sera donc possible rapidement de rentrer dans le sujet de la construction et interprétation du questionnaire, sans passer par une longue phase de compréhension du phénomène.

Le biais de publication intervient donc ici à trois reprises. Une première fois dans le cadre de l'échange entre Ipsos et The Guardian, une deuxième fois pour le choix académique de

ce type de sujet dans le cadre du séminaire de Philosophie Expérimentale de l'ENS, et une troisième fois pour mon propre choix de ce sujet. Cette situation a pour effet de multiplier les articles sur quelques sujets phares, soit qu'ils correspondent à des problèmes de société largement relayés par la presse, soit qu'ils correspondent à une culture partagée par les chercheurs d'un domaine particulier, et nous en verrons plusieurs illustrations ci-après avec l'IAT et l'effet Knobe. Le cas de l'IAT montre d'ailleurs que les deux possibilités ne sont pas exclusives.

Insistons, pour finir, sur le constat que ces difficultés et biais ne constituent pas des éléments de preuve de l'inexistence du phénomène étudié, la surestimation du nombre de musulmans en France. Simplement il met en évidence que si ce phénomène de surestimation (ou plutôt de mésestimation) existait également pour d'autres thèmes moins médiatiques, nous ne le saurions simplement pas. Ce que nous retrouvons en sortie de cette étude est ce qui s'y trouvait en entrée : le problème des musulmans est perçu comme important dans notre société par les lecteurs de la presse grand public. On a ainsi contribué à une boucle qui s'auto alimente : c'est parce qu'il est perçu comme important qu'il donne lieu à une audience solvable, qu'il est donc publié et qu'ainsi il est perçu comme important à étudier, études qui elles mêmes pourront donner lieu à publication.

Pour ensuite évaluer l'importance de ces préoccupations sur l'ensemble du domaine moral, il est utile de caractériser les ingrédients qui ont été nécessaires pour que ce phénomène de boucle auto alimentée se mette en place. Trois ingrédients semblent principalement nécessaires. D'une part un thème socialement important permettant l'amorce du phénomène, d'autre part l'existence d'un groupe d'activistes (extrémistes ou non, à l'image des électeurs d'extrême droite annonçant plus de 50 % de musulmans en France) prêts à intervenir dans la presse (et aujourd'hui sur Internet) pour promouvoir leur point de vue, et enfin un système d'amplification par la presse alimenté par l'audience [29]. On doit alors souligner que de nombreux sujets de philosophie morale peuvent facilement proposer les deux premiers ingrédients, sujets sensibles et extrémistes disponibles, et, s'appuyant sur le même système économique, constituer des bases potentielles pour que se construisent des études à base de questionnaires qui donnent lieu à des boucles auto alimentées faiblement reliées aux phénomènes réels et alimentant des polémiques de plus en plus indépendantes de toute réalité autre que polémique.

29. J'écris ces lignes en octobre 2019, alors que le phénomène vient de se reproduire avec la provocation d'un élu d'extrême droite interpelant une accompagnatrice scolaire voilée. L'importance médiatique prise par le sujet pourra, on peut en prendre pari, se prolonger d'études par sondages ad libitum sur le refus du voile dans l'espace public...

4.4 La transmission des émotions par les larmes

Un argument fréquent des non scientifiques, dont les philosophes moraux, contre l'utilisation de l'expertise scientifique s'appuie sur la diversité de leurs avis d'experts et sur les controverses qu'elle génère. J'ai choisi d'explorer deux cas qui mettent en lumière cette situation. Le premier, objet de la présente section, appartient au domaine de la psychologie de la perception humaine et, plus précisément, de l'olfaction. Il présente l'intérêt de montrer une forme de dissymétrie dans l'acceptation des résultats expérimentaux entre une équipe de psychologues spécialiste des émotions, et en particulier des larmes, et une autre équipe spécialiste de l'olfaction qui utilise des équipements techniques élaborés, dont l'IRMf, pour étudier comment les émotions sont sensibles à l'odorat. Ma conclusion sera que la confrontation des équipes est utile à bien mettre en lumière toutes les difficultés de ces expérimentations, comme des théories en jeu, même si dans le cas d'espèce aucun consensus n'a été construit.

4.4.1 Les larmes communiquent les émotions par l'odorat

En 2011 une équipe de l'institut Weizmann de Tel Aviv a publié un article (Getstein, Sobel et al. 2011) [99] cherchant à montrer que les émotions peuvent, en s'exprimant par les larmes, induire des changements de comportement chez des personnes soumises à la seule olfaction de ces larmes. Cet article s'inscrit dans deux axes à fort enjeu scientifique. Tout d'abord, l'olfaction est un sens relativement délaissé par les études cognitives, en particulier par rapport à la vision, et on est en droit d'attendre de son étude de nouveaux éclairages sur la psychologie humaine. Et ensuite, l'odorat étant traditionnellement présenté comme un sens que nous partageons étroitement avec les mammifères, on peut attendre de ces études des éléments d'information sur ce que l'espèce humaine partage avec ces proches cousins et sur ce qui la différencie d'eux.

L'article décrit un protocole expérimental qui comporte, en première étape, le recueil des larmes émises par des femmes émues par des fictions. Puis, dans la deuxième étape, des hommes reçoivent ces larmes sous la forme de patchs appliqués sur la lèvre supérieure. Les tests sont menés en double aveugle en comparant l'effet des larmes à celui de solutions salines de même composition chimique mais sans les composés organiques des larmes. Dans la troisième étape, des visages de femmes sont présentés à ces hommes porteurs des patchs et le protocole consiste à mesurer s'ils réagissent différemment à ces visages selon qu'ils sont soumis aux larmes ou aux solutions salines. Trois modes d'évaluation de l'influence des larmes sur l'attractivité sexuelle éprouvée par les hommes sont mis en œuvre. La première approche

consiste à demander une évaluation subjective aux participants, qui classent les visages selon leur attractivité. La seconde vise à une évaluation directe par des mesures physico-chimiques, dont le taux de testostérone, et la troisième à une observation par IRMf de l'activité du cerveau. Le protocole comporte également un ensemble d'expériences de contrôle pour conforter le résultat global : les hommes soumis aux substances secrétées dans les larmes des femmes sont moins enclins à l'attraction sexuelle.

Ces expériences confortent plusieurs hypothèses relatives à la perception des émotions par l'odorat. Trois sont soulignées dans l'article. La première est que l'influence des larmes via leur olfaction est totalement inconsciente pour le participant qui y est soumis. La deuxième est que les larmes humaines, bien que n'ayant pas d'odeur consciemment reconnue, comportent des phéromones efficaces. Et enfin, que, sous cet angle, les larmes humaines ne sont pas très différentes des larmes des autres mammifères et, en particulier, des rongeurs, sur qui le même type de phénomène a été détecté.

Mettons en lumière l'opérationnalisation menée à chacune des étapes du protocole employé et les interrogations qu'elle soulève. Dans la première étape du dispositif expérimental, l'émotion de tristesse est obtenue par la visualisation d'une fiction, et les larmes recueillies sont à la fois l'indicateur de cette tristesse et le vecteur supposé de transmission efficace de cette émotion. L'opérationnalisation du stimulus générateur de tristesse est ainsi obtenue par la fiction. Il existe des débats sur les similitudes et différences qu'il pourrait y avoir entre les émotions générées par des fictions et celles de la vie réelle [30] ; ce point n'est pas abordé dans l'article. La transmission de l'odeur des larmes est assurée par des patchs appliqués sur la lèvre supérieure des participants.

Cette méthode permet d'isoler l'influence de l'odorat de celle de la vue car dans l'expérience, les hommes ne voient pas les femmes pleurer. On pourrait toutefois opposer à ce dispositif que les hommes voient quelque chose, et ici des photos de femmes, et que l'indice apporté par ces photos pourrait être utilisé pour interpréter les odeurs. On pourrait dans cet esprit envisager la possibilité que les larmes ne transmettent pas réellement une émotion de tristesse mais seulement une mise en alerte sur la potentielle arrivée d'une émotion et que la vue du visage féminin soit le déclencheur de l'émotion de tristesse qui ne serait que préparée par les larmes. Une telle hypothèse pourrait par exemple conduire à un comportement différent avec des images de violence faisant interpréter les larmes comme des larmes de colère. Cette hypothèse n'étant pas explorée dans l'article, je ne la mentionne ici que pour souligner la difficulté de cette étape d'opérationnalisation.

30. Voir par exemple l'analyse des émotions dans la fiction dans (Renauld 2014) [193]

La troisième étape exploite la disposition des hommes supposés avoir une attirance sexuelle qui peut être activée ou diminuée selon les circonstances. Pour opérationnaliser cette attirance, l'étude a utilisé trois méthodes. Les trois partent de la même présentation de photos de visages de femme dont l'effet a été préalablement étudié [31] et diffèrent par la façon de mesurer l'attirance. Dans la première version, le participant donne son avis sur l'attirance qu'il ressent pour la femme dont la photo lui est présentée. Dans la deuxième, des traceurs biochimiques reconnus pour être habituellement liés à l'activité sexuelle sont recherchés, par exemple le taux de testostérone. Enfin, dans la troisième version, le participant est soumis à une IRMf, les zones du cerveau actives lors d'une attirance sexuelle ayant été repérées au préalable pour chacun des participants. Chacune de ces trois méthodes donne des résultats différents mais jugés concordants par les expérimentateurs. Cette approche applique la technique de triangulation qui permet de diminuer le risque de biais de mesure et d'interprétation. Les mesures directes, physico-chimique ou par IRMf, viennent confirmer les déclarations des participants et la cohérence de leurs valeurs renforce la confiance dans chacune des mesures [32]. Naturellement, une équipe plus exigeante dans la concordance de ces trois ensembles de résultats pourrait en induire la nécessité de poursuivre les expérimentations et non, comme le fait l'article, en conclure que les larmes comportent des phéromones efficaces.

4.4.2 Les larmes sont spécifiquement humaines

Cette conclusion a d'ailleurs été contestée, en particulier par une équipe dirigée par Ad Vingerhoets, spécialiste de la psychologie des larmes et auteur d'un ouvrage sur le caractère spécifique des larmes humaines « Why Only Humans Weep : Unravelling the Mysteries of Tears » (Vingerhoets 2013) [232]. Dans cet ouvrage, l'auteur inscrit les larmes humaines dans l'activité complexe de la communication interpersonnelle qui supposerait la conscience de soi et de l'autre et il en fait ainsi un marqueur important, et même spécifique, du comportement humain. Cet ouvrage ne s'inscrit pas dans une approche explicative du contenu physico-chimique des larmes mais s'attache seulement à leurs effets entre personnes.

L'enjeu de l'article de l'institut Weizmann est important pour Vingerhoets : si la fonction des larmes est de communiquer les émotions par des phéromones, et qu'elle se retrouve à l'identique chez tous les mammifères, alors l'argument de la spécificité des larmes humaines apparaît fragilisé. L'équipe de Vingerhoets a donc cherché à reproduire l'expérience en utilisant le protocole de l'équipe de l'institut. Mais sans succès, et ils ont critiqué les conclusions,

31. On retrouve là l'apport de l'autonomie des expérimentateurs : l'attirance sexuelle a déjà fait l'objet de nombreuses études et on bénéficie donc d'un corpus d'images et de leurs effets comme base de départ.
32. La triangulation utilisée en psychologie expérimentale sera présentée plus en détail en 5.4.2.2, page 315

hâtives et non justifiées à leurs yeux, affirmant la transmission des émotions par les seuls phéromones des larmes.

L'institut Weizmann a ensuite répondu aux critiques (Sobel 2017) [212]. Soulignons la dissymétrie entre les arguments des deux équipes. L'équipe de Vingerhoets a tenté de reproduire les effets présentés dans l'article initial et a conclu à la non-réplicabilité de l'expérience et donc à la remise en cause du phénomène de transmission des émotions par les larmes. L'équipe de l'institut Weizmann a, elle, critiqué le dispositif expérimental employé par l'équipe de Vingerhoets pour la tentative de réplication, en insistant sur le fait que cette équipe ne travaille pas dans le cadre d'un laboratoire dédié à l'olfaction. Elle n'a donc pas les compétences nécessaires à ces expériences délicates et il n'est pas étonnant que les chercheurs n'aient pas reproduit le phénomène initial.

Soulignons enfin que les équipes ne se sont pas rapprochées pour établir une expérience coordonnée, comme le proposait l'institut Weizmann. Considérant la non réplication de l'expérience initiale comme preuve suffisante, l'équipe de Vingerhoets a poursuivi ses travaux en écartant ceux de l'équipe de l'institut Weizmann et sans plus les mentionner par la suite. Ils ne sont par exemple pas référencés dans l'article « Why Only Humans Shed Emotional Tears, Evolutionary and Cultural Perspectives » (Gracanin 2018) 2018 : [106].

4.4.3 La difficile nécessaire collaboration entre chercheurs

L'expérimentation est délicate et les enjeux importants

Mon objectif n'est pas ici de trancher entre les visions des deux équipes mais, encore une fois, de souligner la sensibilité de ce type d'étude à l'étape d'opérationnalisation. Cette étape est difficile sous l'angle épistémique et sensible sous l'angle théorique. Difficile sous l'angle épistémique car l'ensemble de la chaîne qui va des émotions aux larmes et des larmes à la modification du comportement est complexe et requiert de nombreuses compétences très pointues pour être constitutive d'une expérience raisonnablement contrôlée. A chaque étape du protocole expérimental, différentes techniques sont mobilisées et personne ne peut avoir une vue à la fois globale permettant d'affirmer la pertinence d'ensemble du protocole en regard du problème posé, et une vue détaillée de chacune des opérations techniques spécialisées à mener. La diversité de ces techniques conduit à une situation où même les critères de qualité et de pertinence de chacun des spécialistes sont incompris par les autres membres de l'équipe, rendant la notion d'une validation globale peu praticable. L'étape d'opérationnalisation est également sensible sous l'angle théorique car l'enjeu idéologique est important pour

les deux équipes alors même que les théoriciens ne peuvent maîtriser l'ensemble de la chaîne expérimentale. Les enjeux idéologiques portent ici sur la spécificité des larmes proprement humaines et, plus généralement, sur l'existence d'une différence importante et objective permettant la distinction entre animaux humains et non humains, sujet qui a alimenté de nombreux débats dans de nombreux domaines.

Alimentés par ces difficultés et par ces enjeux, les biais de confirmation semblent porter un risque particulièrement élevé dans cette étude. Le biais de confirmation pourrait ici être évoqué tant pour l'équipe de l'institut Weizmann, qui la conduirait à disqualifier trop rapidement la critique de Ad Vingerhoets, que pour ce dernier, qui ne réussirait pas à répliquer une expérience parce qu'il ne souhaite pas la voir confirmée.

Et la coopération entre chercheurs devient indispensable

Malgré ces enjeux qui les rendent difficiles, les coopérations sont une voie pour diminuer les risques de biais de confirmation. Les interactions entre équipes d'expérimentateurs et de théoriciens sont des ressources à mobiliser contre les risques liés aux difficultés de l'opérationnalisation, et en particulier, face aux nombreux biais dont l'importance n'est plus à établir. Le cas de ces deux situations en miroir, et sans préjuger de la validité de l'une ou de l'autre, montre tout l'intérêt qu'il y aurait à explorer une autre voie : l'interaction entre les deux équipes qui, si elle aboutit, neutralisera, symétriquement, les deux risques de biais de confirmation.

Le spécialiste de la perception qui mène l'expérience sur l'olfaction n'a probablement qu'une idée assez vague du mode de fonctionnement détaillé des instruments utilisés dans chacune des trois méthodes de mesure employées. Que ce soit sous l'angle de leur description théorique ou sous l'angle de leur mode d'emploi pratique, il ne peut cumuler les connaissances et les savoir-faire nécessaires à la maîtrise de la méthodologie de l'enquête par questionnaires, à l'outillage physique et chimique utilisé pour détecter les traceurs biochimiques et encore moins les techniques sophistiquées déployées pour réaliser une IRMf. L'opérationnalisation des émotions et de leur transmission par les larmes suppose ici une relation de confiance entre la personne qui construit le cadre général de l'expérimentation et chacun des spécialistes mobilisés pour la mise en œuvre de chacune des techniques utilisées.

Cette confiance peut être également comprise comme le résultat d'un transfert de responsabilité du chercheur qui souhaite mesurer l'attirance sexuelle, vers les équipes d'expérimentateurs et de constructeurs de machines qui maîtrisent chacune de ces techniques et leur mise en œuvre. Ces équipes, formelles ou informelles, ont construit et entretenu cette maîtrise, non pour mesurer l'effet des larmes, mais pour toute une catégorie de situations expérimentales

qui en ont, en retour, validé les dispositifs de mesure. C'est ainsi tout un ensemble d'études qui, dans la pratique quotidienne de l'expérimentation, a pu contribuer à l'amélioration permanente de toutes ces techniques et outillages de mesures. Cette amélioration est largement autonome en regard de chacune des théories qui en bénéficient. [33]

Dans le domaine de la psychologie humaine, à ces difficultés de méthode s'ajoutent celles qui sont dues aux enjeux idéologiques qui structurent les opinions des chercheurs, comme dans notre exemple la proximité de l'homme par rapport aux mammifères non humains. Ces enjeux facilitent [34] l'émergence de polémiques parfois appuyées sur la fragilité de l'étape d'opérationnalisation. Le risque de biais de confirmation est ici multiplié par l'enjeu lié au domaine : si les conclusions de l'article de l'institut Weizmann venaient à être acceptées par la communauté des psychologues, une capacité qui a été analysée par d'autres comme spécifiquement humaine rejoindrait l'héritage commun qui nous rapproche des autres mammifères. C'est l'impact de cet enjeu, de ce qui fait d'un animal un humain, sur les équipes de recherche que montre également la polémique autour de la fonction des larmes humaines.

Reconnaître l'importance de l'étape d'opérationnalisation en tant que lieu de la relation de confiance et de négociation entre théoriciens et expérimentateurs pourrait avoir comme bénéfice de diminuer le risque lié au biais de confirmation. L'équipe de l'institut Weizmann utilise un ensemble de techniques diverses pour approcher un grand nombre de questions sur l'olfaction. La difficulté de ce domaine conduit les chercheurs à devoir développer de multiples partenariats et de nombreux savoir-faire qui sont validés par l'ensemble des études faisant intervenir l'olfaction, et non pour la seule expérience particulière portant sur les émotions. En revanche, l'équipe de Vingerhoets n'a pas l'habitude d'expérimenter sur l'olfaction et, bien que son expertise au regard des émotions soit bien établie, elle n'a pas construit le même réseau de savoir-faire autour de la mesure de l'olfaction. De notre point de vue, et pour bénéficier au mieux de l'investissement expérimental envisagé pour l'étude sur la transmission des émotions par les larme, il serait idéal que l'équipe de Vingerhoets se charge de préciser l'hypothèse théorique à tester (contexte, stimuli, effets), que l'équipe de Noam Sobel se charge de l'expérience (opérationnalisation du contexte, des stimuli, des effets) et que la discussion s'instaure d'une part sur la validité de l'opérationnalisation, la pertinence de la chaîne expérimentale en regard du phénomène théorique testé et, d'autre part, à supposer que la pertinence soit établie, sur ce qu'il est possible de déduire des résultats de l'expérience sur l'hypothèse théorique testée.

33. Ce thème est récurrent dans l'œuvre de Ian Hacking, voir par exemple (Hacking 1983) [115]
34. La facilitation peut être liée à l'engagement militant du chercheur ou à l'obtention de budgets d'organismes eux-mêmes engagés à défendre une cause.

4.4.4 La diversité, source de conflit et d'opportunités

L'exemple de la transmission des émotions par les larmes me permet, en conclusion, de souligner un point important. Les deux problèmes symétriques de la difficulté des expériences sur l'olfaction et de la faiblesse de la théorie de la spécificité des larmes humaines n'émergent que parce que les équipes des psychologues théoriciens des émotions et les psychologues praticiens de l'olfaction sont des équipes différentes aux enjeux propres et sans interaction. C'est de la collaboration, éventuellement agonistique, entre ces équipes que, comme le suggère la métaphore de l'hélice expérimentale proposée plus haut, peut naître le mouvement, sans que cette naissance soit le moins du monde certaine. La diversité des équipes constitue donc à la fois un risque, si chacune s'isole dans son domaine de compétence, et, à la fois, des opportunités pour détecter les différences d'approche et pour en rechercher le dépassement dans de nouvelles collaborations.

Je n'entends pas ici en induire par une ample généralisation, certainement abusive, que le philosophe expérimental aurait tort de vouloir réaliser lui-même ses expérimentations. Néanmoins, l'exemple montre que l'entre-soi peut comporter des risques en facilitant l'expression des biais cognitifs. Rappelons, plus simplement, que la démarche scientifique expérimentale est, le plus souvent, répartie sur plusieurs équipes et que cette répartition est à la fois une contrainte, par le haut niveau de collaboration et de confiance qu'elle impose, et un atout par les multiples biais qu'elle permet d'éviter en imposant que se mettent d'accord des chercheurs de différents domaines ayant, potentiellement, des enjeux différents.

4.5 L'IAT : la réplication et les conflits de valeurs

4.5.1 Mesurer l'association implicite entre concepts.

Comme nous l'avons vu dans les chapitres précédents, la réplication d'expériences est une des composantes importantes de la méthode scientifique expérimentale. J'ai souhaité examiner plus avant dans quelle mesure il en était de même pour les expériences de psychologie en m'appuyant sur un exemple actuel, le test d'association implicite (en anglais, IAT). L'IAT a été proposé par des psychologues en 1998 (Greenwald, McGhee et Schwartz 1998) [113] pour mesurer l'intensité des associations que des participants exercent entre deux concepts, éventuellement de façon non consciente. Il convient donc d'obtenir cette information de façon implicite, et sans leur poser des questions explicites auxquelles ils répondraient de façon biaisée. Par exemple, on peut essayer de mesurer la force de l'association entre le concept « fleur »

et le concept « bien être » en observant le comportement des participants mis en présence d'images de fleurs, et ce comportement pourrait apporter des informations sur l'association faite entre « fleurs » et « bien être », informations complémentaires, et peut-être différentes, des réponses explicites à une question directement posée sur ce sujet. L'IAT, lancé en 1998, a été largement utilisé et reproduit par les psychologues au cours des 20 dernières années pour mesurer ces associations implicites entre concepts. En particulier, une version dite « Race IAT » a pour objectif de mesurer l'association implicite entre les concepts de race, noir ou blanc, et les sentiments positifs ou négatifs. Ce Race IAT a été proposé dès 1998 pour évaluer le racisme implicite, quand les personnes se disent, explicitement, non racistes. De ces test IAT et Race IAT, toutes sortes de réplications ont été entreprises : exécuter à nouveau les traitements des algorithmes statistiques à partir des mêmes données brutes, répliquer l'expérience avec exactement le même protocole expérimental, répliquer l'expérience avec des modifications légères ou profondes du protocole. Malgré ces efforts, l'IAT et, plus précisément, sa version Race IAT font encore l'objet de débats animés entre scientifiques, débats relayés dans la presse. [35]

L'IAT s'appuie sur l'idée que l'association implicite d'un concept, par exemple la race, avec un autre, par exemple un sentiment positif, doit se traduire dans le comportement par un temps de réaction plus rapide dans une expérience qui demande d'associer ces concepts que dans une expérience qui demande de les dissocier. J'appelle cette dissymétrie de temps de réaction le « phénomène IAT » et j'appelle « indice IAT » l'index élaboré à partir de plusieurs mesures de ces temps de réaction dans le cadre d'un protocole expérimental normalisé. Les promoteurs de l'IAT affirment que cet indice mesure la force de la relation implicite entre les concepts et permet de quantifier les tendances racistes d'un individu, y compris lorsque celui-ci se dit, explicitement, non raciste.

Au travers d'une analyse de la littérature sur l'IAT, je propose de mettre en évidence qu'en appliquant les principes de la démarche scientifique expérimentale et tout particulièrement la réplication, les différentes équipes de chercheurs sont arrivées à ce que j'appelle une « synthèse minimale » qui comporte les éléments suivants :

— Le phénomène IAT existe et donne accès à la relation implicite entre concepts.
— Le phénomène est facile à mettre en œuvre. Il est facile de bâtir un service sur Internet présentant des images et mesurant les temps de réaction des participants face à ces images.
— Mais, après ce recueil de données de base, la construction statistique pour élaborer un

35. Voir par exemple en décembre 2017 https://qz.com/1144504/the-world-is-relying-on-a-flawed-psychological-test-

indice est complexe et difficile à stabiliser.

— En tant qu'indice individuel, l'IAT est peu fiable, principalement parce qu'il varie d'un test à l'autre pour la même personne ce qui le rend difficile à utiliser, surtout quand il n'est mesuré qu'une seule fois.

Malgré un accord sur cette synthèse minimale, les psychologues sont en profond désaccord sur ce qui peut et doit être fait de cet indice. Le désaccord est à la fois de nature épistémique, comment améliorer notre connaissance des processus psychologiques sous-jacents au phénomène IAT, et de nature pragmatique, peut-on utiliser l'indice Race IAT comme outil dans la lutte contre le racisme. Mon analyse va montrer que si le très grand nombre de réplications a bien permis d'atteindre la synthèse minimale, il n'a pas, et peut-être ne peut pas, résoudre ces désaccords.

Je vais d'abord présenter l'intérêt que présente l'IAT dans le cadre de mon approche, puis je présenterai le protocole visant à établir l'indice IAT et le contexte de son succès académique. J'évoquerai ensuite quelques éléments des débats entre promoteurs et opposants à l'IAT qui ont conduit à l'accord sur la synthèse minimale et j'explorerai enfin les désaccords entre chercheurs appuyés sur des systèmes de valeurs différents.

4.5.2 Le cas de l'IAT, accords et désaccords

Mon objectif étant d'évaluer ce que peut apporter la réplication dans un contexte d'étude psychologique complexe, je prends l'IAT pour cas intéressant à étudier et quatre types d'arguments peuvent être avancés pour justifier ce choix.

Premièrement, l'IAT est un protocole bien connu, facile à comprendre sans faire appel à des technicités particulières, ce qui permet d'éviter de rendre les désaccords incompréhensibles aux non experts. D'ailleurs le lecteur pourra lui-même faire ce test en quelques minutes pour se faire sa propre expérience[36]. Cela ne signifie pas que les désaccords entre psychologues n'ont pas une certaine dimension technique, en particulier statistique, mais plutôt que la dimension non technique est prépondérante et accessible à l'étude philosophique.

Deuxièmement, l'IAT a été massivement répliqué de multiples façons et utilisé de façon généralisée dans le cadre de la lutte contre les discriminations. De nombreuses équipes ont repris les expériences en visant soit à vérifier l'élaboration des conclusions en partant des mêmes données brutes, soit à reproduire les mêmes résultats en conservant le même protocole, ce qui permet à la fois de retrouver le phénomène et d'augmenter la validité statistique des conclusions en augmentant la taille des échantillons, soit enfin à refaire des expériences

36. Voir "Project Implicit" sur le site de Harvard : `https://implicit.harvard.edu/implicit/`

avec des protocoles modifiés pour prendre en compte les cas d'ambiguïtés identifiés dans les interprétations. Certaines des publications relatives à l'IAT ont suivi les recommandations de l'OSF (Open Science Framework) qui visent à faciliter la réplication et il convient de remarquer ici que Brian Nosek, un des promoteurs de la méthode IAT, est également un membre très actif de cette organisation. Cet ensemble de réplications a donné lieu depuis 1998 à de nombreuses publications ainsi qu'à des méta-analyses (voir infra).

Troisièmement, l'IAT est un cas intéressant car récent. La faible réplicabilité des expériences en psychologie était déjà un phénomène connu en 1998 et il n'y a pas de risque d'anachronisme à appliquer au cas de l'IAT des critères épistémiques modernes, ce qui serait le cas si on appliquait ces mêmes critères à des cas d'étude des 17^{e} ou 18^{e} siècles.

Enfin, quatrièmement, le débat académique entre promoteurs de l'utilisation généralisée de la méthode IAT et opposants à cette généralisation a dépassé les frontières des revues spécialisées et s'est largement répandu dans la presse. Il est donc facile d'accès pour toute personne intéressée par ces débats.

Pour toutes ces raisons, l'exemple de l'IAT est pertinent pour mon approche. Il permet d'analyser comment, dans ce cas, opposants et promoteurs ont débattu en utilisant de multiples réplications et, ainsi, de pouvoir évaluer ce que ces réplications ont apporté aux débats. Néanmoins, il est également important de souligner les difficultés d'une telle analyse. D'abord on peut remarquer que le débat sur l'utilisation de l'IAT dans la lutte contre la discrimination s'inscrit dans le contexte politique très particulier des États Unis et que l'aborder sous le seul angle d'approche des publications des psychologues est trop étroit et, par tant, sans intérêt. Mon travail n'est pas ici de philosophie politique, ni de sociologie des sciences pour analyser comment les groupes de promoteurs et d'opposants se sont constitués et se sont développés. Je me limite à l'aspect épistémique de la contribution de la réplication aux débats et espère ainsi montrer que cet apport atteint dans le cas de l'IAT une limite qu'il convient d'expliquer. Deuxième difficulté, la méthode IAT peut s'appliquer à de très nombreux domaines chaque fois que des concepts semblent être implicitement associés pour conduire à une disposition à l'action. On peut citer pour exemple, dans le domaine du marketing, la présentation d'une image d'un produit qui déclenche des associations pouvant induire la décision d'achat. Ou, dans le domaine de la santé, la recherche des concepts qui sont implicitement reliés à la vaccination et peuvent contribuer à expliquer les différents types de réception à son égard. Je me concentre sur le cas de l'IAT utilisé pour l'étude du racisme implicite et mes conclusions ne prétendent pas être directement généralisables à tous les autres domaines de la psychologie ayant utilisé les protocoles expérimentaux de l'IAT. Je vais maintenant décrire plus précisé-

ment le mode de fonctionnement de l'IAT et présenter quelques jalons de son développement depuis 1998.

4.5.3 Résumé de l'histoire de l'IAT

Mon objectif dans la description qui suit n'est pas de rentrer dans tous les détails d'un protocole complexe mais plutôt de donner les éléments utiles à la bonne compréhension des débats entre promoteurs et opposants de l'utilisation généralisée de l'IAT. Je décrirai dans la prochaine section l'accord que ce débat a permis de construire et, dans la section suivante, le désaccord qui subsiste entre promoteurs et opposants. [37]

L'invention de l'Implicit Association Test

Je me propose de décrire le protocole de l'IAT avec un exemple simplifié qui suit les principaux traits de la présentation originale de 1998 (Greenwald 1998) [113] . Prenons deux concepts qui se présentent à nous sous la forme d'un choix à opérer : thé ou café ? Et prenons deux concepts qui qualifient ces boissons : agréable ou désagréable ? Imaginons maintenant plusieurs images ou mots, disons 5, qui sont associés au thé, 5 autres au café, 5 à des situations agréables et enfin 5 à des situations désagréables. Nous avons ainsi 20 images que nous pouvons présenter alternativement à un participant sur un écran d'ordinateur, et le participant réagit à cette image en appuyant sur une touche du clavier soit à gauche soit à droite. Le protocole suit alors les étapes suivantes.

— Les deux premières étapes permettent de vérifier et renforcer le lien entre chaque image et le concept qu'elle est censée évoquer : « Appuyez sur la touche de gauche si l'image évoque le thé. Appuyez à droite si elle évoque le café ». Et de même pour agréable et désagréable.

— Dans l'étape suivante on associe café à agréable et thé à désagréable : « Appuyez sur la touche de gauche si l'image évoque le café ou est agréable. Appuyez à droite si l'image évoque le thé ou est désagréable »

— Et enfin, on inverse l'association : « Appuyez sur la touche de gauche si l'image évoque le café ou est désagréable. Appuyez à droite si l'image évoque le thé ou est agréable »

L'idée centrale des inventeurs de l'IAT est que si le participant associe le café à une sensation agréable, alors il répondra plus rapidement à une question qui associe les deux concepts et il répondra plus lentement quand il doit associer le café à une sensation désagréable. On déroule le protocole plusieurs fois en alternant les images disponibles et en alternant l'ordre

37. Si le lecteur connaît le contexte de l'IAT et de son développement, il peut passer directement à la section suivante.

des questions. Le principal résultat est ensuite élaboré à partir de la différence des temps de réaction entre les cas à association directe et à association inversée.

De façon à pouvoir plus loin comparer l'IAT à d'autres méthodes, je propose d'introduire deux autres acronymes, EAT et BAT. L'EAT (Explicit Association Test) est une mesure explicite de l'association : on demande par exemple au participant si, sur une échelle de 1 à 10, il trouve que le café est une boisson agréable. Ces méthodes explicites supposent de la part du participant un acte d'introspection puis un acte de communication à l'enquêteur, or chacune de ces deux étapes ouvre la possibilité de biais cognitifs qui faussent l'expression du participant. Les biais liés à l'introspection sont identifiés depuis les débuts de la psychologie, il est inutile d'y revenir ici, les biais sociaux sont liés au phénomène d'alignement du comportement du participant sur ce qu'il pense être l'attente de l'enquêteur, ainsi si l'enquête se fait dans un contexte universitaire en SHS (Sciences Humaines et Sociales), le participant peut s'attendre à ce que l'enquêteur soit contre le racisme et modulera sa réponse en fonction de cette attente supposée, alors qu'il modulera sa réponse dans le sens opposé si c'est une enquête de la police américaine. L'IAT est supposé éviter ces deux écueils de l'introspection et de l'attente sociale par l'accès direct aux relations implicites.

Le BAT (Behavior Association Test) est une mesure directe d'un comportement supposé refléter l'association : on présente à un participant qui vient d'entrer dans une salle une table où il peut se servir soit du café soit du thé et on enregistre le choix qu'il fait en faisant varier les conditions de l'expérience. Ce type de test a l'avantage de mesurer directement un comportement lié à la préférence qu'on cherche à identifier. Il soulève toutefois de nouvelles sortes de difficultés. D'une part, cela reste une mesure en laboratoire et l'équivalence entre cette situation et la vie réelle reste à établir. D'autre part, le biais lié à l'attente sociale est également présent : le participant peut choisir une boisson parce qu'il pense que c'est ce qu'attend de lui l'enquêteur. Et enfin, ce type de test n'est possible que pour la frange de comportements pour lesquels il existe une contrepartie en laboratoire crédible. Pour les comportements racistes, par exemple, il n'est pas du tout évident que s'asseoir sur un banc en s'éloignant d'un voisin noir soit équivalent à avoir une disposition à passer à l'acte de la discrimination ou, a fortiori, du crime raciste.

Une partie importante de la littérature sur l'IAT a pour objet de comparer les résultats des trois types de tests, IAT, EAT et BAT. Dans la publication originale de 1998 (Greenwald, McGhee et Schwartz 1998) [113], les auteurs montrent que les résultats des deux types IAT et EAT ne sont pas corrélés (ou plus exactement, sont moins corrélés que différentes mesures EAT entre-elles) ce qui, pour eux, s'interprète comme la preuve que l'IAT donne bien accès à

des associations implicites que les participants ne reconnaissent pas explicitement. Naturellement, les opposants ne retiennent pas cette interprétation et affirment plutôt qu'on peut en déduire que l'IAT est moins fiable que les mesures explicites. Les études de corrélation entre association implicite et comportement, entre IAT et BAT, puis entre associations explicites et comportement, EAT et BAT, ont pour objet d'établir laquelle des deux méthodes, implicite ou explicite, est un meilleur prédicteur du comportement, au moins en laboratoire. L'étape suivante serait de comparer IAT et EAT en regard du comportement dans la vie réelle, ce qui est souvent hors de portée des psychologues expérimentaux pour d'évidentes raisons pratiques et éthiques.

L'IAT, un succès académique, idéologique et commercial

Lancé en 1998, l'IAT s'est rapidement développé dans au moins trois directions. D'une part le succès académique, que j'illustre ci-après avec le nombre d'articles auquel il a donné lieu, ensuite le succès par la diffusion de son idée principale dans de multiple domaines et enfin, le succès commercial avec la multiplication des offres de conseil et de services aux entreprises et administrations pour mieux évaluer la discrimination chez leurs collaborateurs quand, mesurée par une méthode implicite, elle vient contredire des déclarations explicites opposées au racisme.

La figure ci-dessous présente le nombre d'articles parus chaque année selon Google Scholar pour une recherche comportant l'expression « Implicit Association Test ». Le total, 22100 réponses, est imposant ainsi que la croissance continue sur toute la période.

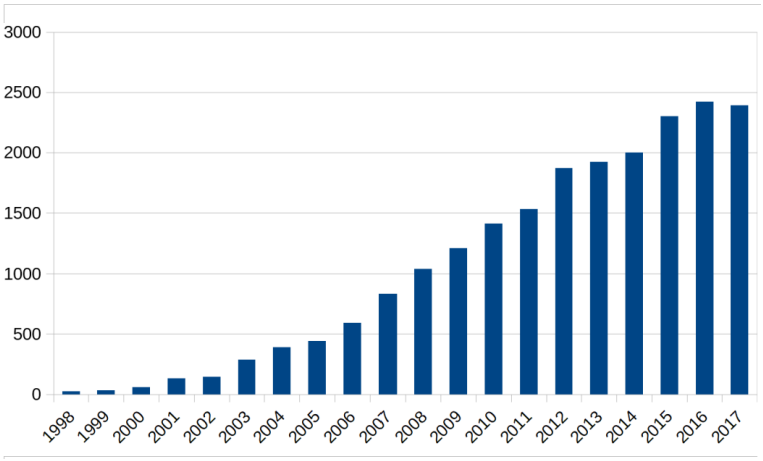

FIGURE 4.7 – IAT Nombre d'articles par année (source Google Scholar)

Ce succès académique traduit la montée en puissance d'une idée : l'association implicite entre concepts est une composante importante de la cognition humaine et peut expliquer certains comportements. La psychologie se doit de prendre en compte cette idée de façon très

générale car de telles associations peuvent s'insérer dans tout processus mental et fausser l'analyse qui en est faite en les négligeant. Le protocole IAT, accédant directement à ces associations implicites, ouvre ainsi la possibilité de proposer des pistes d'explication pour des comportements qui ne sont pas en cohérence avec les déclarations explicites. Le comportement qui peut être vu comme irrationnel si on considère que les associations explicitement formulées sont réellement représentatives de ce qui dirige le comportement des participants, devient explicable si on le considère comme porté par les associations implicites mises à jour par l'IAT. Chaque sous partie de la psychologie, chaque question psychologique, lorsqu'elle repose principalement sur des déclarations explicites, doit alors entreprendre une analyse des nouvelles perspectives ouvertes par l'éventualité d'associations implicites à l'œuvre dans les phénomènes étudiés. Cette remise en question est en résonance avec celle qui a suivi les travaux importants de Kahneman et Tversky (Kahneman, Slovic et Tversky (eds.) 1982) [127] qui ont montré l'importance et la prévalence des biais cognitifs, dont certains pourraient être expliqués en prenant mieux en compte les associations implicites entre concepts.

L'IAT devient, dans cette perspective, une voie d'accès potentielle à la résolution des comportements inexplicables en partant des déclarations explicites des individus. C'est le message central de la publication de 2017 des promoteurs de l'IAT qui présente ce qu'ils appellent la « révolution de l'implicite ». (Greenwald et Banaji 2017) [111].

Mais l'IAT n'est pas seulement un succès académique et théorique, c'est aussi un succès commercial avec de nombreux outils et services disponibles pour toutes les entreprises et les organisations qui souhaitent lutter contre la discrimination. C'est un nouveau marché et les promoteurs académiques de l'IAT se sont positionnés sur ce marché en créant pour cela une organisation, le Project Implicit Organization :

> Project Implicit est une organisation a but non lucratif et une organisation internationale de collaboration entre chercheurs intéressés dans la cognition sociale implicite. (...) Project Implicit a été fondé en 1998 par trois scientifiques, Tony Greenwald (Université de Washington), Mahzarin Banaji (Harvard) et Brian Nosek (Université de Virginie). (...) Project Implicit fournit également des services de conseil, de la formation et des séminaires sur les biais implicites, la diversité et l'inclusion, le leadership, en appliquant la science à la pratique et à l'innovation. [38]

Des sociétés de service, à buts lucratifs ou non, ont développé des offres de logiciels, de service et de conseil pour satisfaire la demande d'outils faciles à utiliser pour mettre en œuvre les méthodes issues de l'IAT et lutter contre la discrimination dans les entreprises privées

38. Traduction de l'auteur, source : `https://www.projectimplicit.net/organization.html`.

ou dans les organisations publiques. La motivation principale de ces prestations n'est pas simplement la mesure de l'association implicite mais surtout la lutte contre la discrimination. Le passage de l'un à l'autre est déjà explicité dans un article de 1995 de Greenwald and Banaji [110] (page 10) qui décrit l'intérêt que présenterait une mesure de l'association implicite si elle existait (elle ne sera présentée que 3 ans plus tard) : si on dit à quelqu'un qu'il a des tendances implicites à être raciste, alors il construira de lui-même un ensemble de mesures permettant de les neutraliser, surtout s'il exprime explicitement des opinions non racistes, il cherchera à mettre en accord son comportement et ces sentiments explicites non racistes.

Les polémiques autour de la pertinence de l'IAT

Mais ce brillant succès de l'IAT n'est pas sans une part d'ombre. Pour faciliter la présentation, je propose de différencier le camps des promoteurs de la méthode du camp des opposants, cette polarisation n'est bien sûr pas souhaitable, mais elle reflète la teneur des articles. Elle est donc pertinente dans une perspective descriptive. A titre d'exemples quelques articles issus du camp des promoteurs :

— 1995 Greenwald et Banaji : Expression du besoin d'un outil de mesure des associations implicites (Greenwald et Banaji 1995) [110]

— 1998 Greenwald et all. : Définition du protocole IAT et lancement du site de Harvard (Greenwald, McGhee et Schwartz 1998) [113]

— 2001 McConnell et Leibold : Premier article visant à établir la corrélation entre IAT et comportement (McConnell et Leibold 2001) [161]

— 2017 Greenwald and Banaji : Article décrivant l'importance de la révolution apportée par le point de vue Implicite (Greenwald et Banaji 2017) [111]

Et, toujours à titre d'exemples, quelques articles issus du camp des détracteurs :

— 2004 Tetlock Arkes : Article mettant en cause la validité du protocole (Arkes et Tetlock 2004) [12]

— 2006 Blanton : Décoder l'IAT et expliciter les difficultés de son élaboration en tant qu'indice mesuré. (Blanton et al. 2006) [27]

— 2006 Fiedler : Critique de I, A et T (Fiedler, Messner et Bluemke 2006) [86]

— 2013 Oswald : Meta Analyse des critères utilisés pour les études relatives à l'IAT (Oswald et al. 2013) [176]

Les deux camps, promoteurs et opposants, n'ont pas eu le même succès académique, comme le prouve une rapide vérification du nombre de citations [39] : 16 666 citations pour les 3 articles les plus cités des promoteurs et 843 citations pour les 3 articles les plus cités des

39. Citations indiquées par Google Scholar et relevées en septembre 2019.

opposants. L'article le plus cité est, sans surprise, l'article de 1998 comportant la définition initiale de l'IAT avec 9580 citations. Le plus cité des articles d'un opposant n'atteint pas 400 citations. Je vais maintenant détailler les critiques des opposants de façon à disposer d'une vue équilibrée des arguments des uns et des autres.

Dans l'article de 2004, (Arkes et Tetlock 2004) [12] qui semble être un des premiers à expliciter les difficultés liées à l'élaboration de l'IAT, les auteurs décrivent trois ambiguïtés qui mettent en cause les interprétations faites des résultats de l'IAT. Tout d'abord, le lien entre les concepts de « noirs » et les sentiments « négatifs » pourraient ne refléter qu'un lien sémantique général sans signification particulière pour les dispositions à l'action raciste d'un participant particulier. Ce lien général expliquerait mieux que l'analyse développée par Greenwald, les résultats généraux très pessimistes faisant d'une grande majorité d'Américains des racistes implicites. Ensuite, le lien implicite entre noir et sentiments négatifs peut s'interpréter de façons opposées selon le positionnement politique de chacun. Pour les racistes, il est naturel d'associer le concept « noir »avec des opinions négatives, puisque les noirs ont effectivement, pour eux, des caractéristiques inférieures. Pour les antiracistes, cette association est également naturelle, mais pour une toute autre raison, les noirs ne sont pas inférieurs mais sont discriminés, et donc le concept « noir » évoque une situation difficile qu'il convient de combattre. Il ne peut donc y avoir, comme le prétendent les promoteurs de l'IAT, de lien immédiat entre un indice IAT négatif et une propension à un comportement raciste. On ne peut nier qu'il est justifié de lier les concepts de noir et les sentiments négatifs en Amérique aujourd'hui compte tenu de la discrimination en place dans la société, et ce lien n'a donc rien à voir avec d'éventuelles attitudes proto racistes de la personne interrogée. En somme les trois ambiguïtés vont dans le même sens ; l'indice IAT mesure peut-être une association entre des concepts, mais il n'y a pas de lien entre ce lien et les comportements. La discrimination est présente aux États Unis de façon flagrante et la mesure de l'IAT n'apporte pas grand chose.

Dans un article de 2006 (Blanton et al. 2006) [27], est développé un argument qui deviendra important pour tous les opposants. Il vise à constater la faiblesse du modèle psychométrique de l'IAT : si l'indice IAT doit être compris comme une mesure, alors il faut disposer d'un modèle causal expliquant pourquoi et comment un changement dans ce qui est supposé mesuré par l'indice Race IAT, la propension à des actes racistes, est causalement lié à un changement de l'indice IAT résultat de la méthode. Les auteurs montrent que ce lien causal peut avoir plusieurs formes et que chacune pose des contraintes sur la structure de l'indice IAT. Les auteurs analysent ensuite ces contraintes en découplant deux étapes, la première est conceptuelle et la seconde observationnelle. Conceptuellement, on évalue une attitude,

positive ou négative, de +1 à -1, 0 correspondant à une attitude neutre, en regard des Noirs (Eb) et des Blancs (Ew). Mais le calcul de l'indice IAT n'utilise que la différence Eb-Ew et ceci pose une contrainte formelle importante : il faut que les évaluations entre Noirs et Blancs ne varient pas ensemble mais toujours de façon opposée car s'il en allait autrement, la différence n'aurait pas de signification. Reprenons notre exemple du thé et du café. Si vous aimez pareillement le thé que le café, alors la différence est à 0 mais peut signifier aussi bien que vous aimez les 2 ou que vous détestez les 2 et que vous préférez en fait le chocolat, choix qui n'était pas proposé. Dans le cas de Race IAT, on peut imaginer un raciste asiatique qui n'aime ni les blancs ni les noirs, il aura alors un indice IAT de 0 qui le classe indûment comme « neutre ». Cette contrainte formelle limite conceptuellement l'utilisation du protocole IAT à des cas où les attitudes sont polarisées, c'est-à-dire qu'elles varient simultanément et en sens opposé.

L'observation du phénomène IAT s'appuie sur le temps de réaction et l'indice IAT proprement dit sur un traitement en plusieurs étapes de ce temps de réaction. L'algorithme de calcul comporte une première étape où sont éliminées les réponses trop rapides ou trop lentes (seuils à 300 et 3000 ms), puis on effectue le calcul du logarithme du temps de réaction (opération classique en psychophysique depuis 1930), puis on opère la différence entre les réactions liant ou opposant les concepts supposés associés (association compatible – association incompatible). Dernière étape, on prend la moyenne des indices différentiels calculés sur plusieurs itérations (on peut alors éliminer les premières itérations qui sont considérées comme nécessaires à la formation des participants et on limite à un nombre d'itérations n'induisant pas une fatigue qui conduirait à une dérive des temps de réaction).

L'IAT : un modèle psychométrique défaillant

Les auteurs, pour décoder l'IAT, proposent de rechercher quels modèles psychométriques seraient compatibles avec ce calcul de l'indice IAT. Un modèle psychométrique consiste en une modélisation causale du phénomène qui permet de clarifier comment une différence dans ce qui est mesuré (ici l'association entre concepts) se traduit dans une différence dans la valeur de l'indice et, inversement, comment une différence de la valeur de l'indice peut être interprétée comme une différence dans le phénomène qu'on vise à mesurer. Le point clé du raisonnement est que chaque modèle psychométrique impose, pour être à la base d'une mesure valide, des contraintes sur le phénomène, et il convient de valider si ces contraintes sont, plausiblement, respectées. Clarifions ce point clé par un exemple. Supposons, pour les besoins de l'exemple, que je cherche à mesurer un trait psychologique, disons un niveau de compétence pour une tâche de discrimination visuelle, dont je suppose, en application d'une certaine théorie, qu'il a une valeur A. Un processus de mesure me conduit à donner à ce ni-

veau de compétence une valeur X calculée sur la base à la fois de la vitesse de réalisation de la tâche et sur le taux d'erreur. Et supposons enfin que, dans ce calcul de X, les termes liés à la vitesse et aux erreurs s'ajoutent. Dans ce cas, le principe même de l'addition implique que l'apport de compétence calculée du à un certain gain de temps de réaction est le même quelque soit le taux d'erreur (et réciproquement). Le modèle psychométrique doit alors montrer qu'il est plausible, du point de vue de la tâche étudiée, que des personnes de niveaux de compétences différents gagnent pareillement en compétence lorsqu'ils gagnent en vitesse, suite, par exemple, à un entraînement (et réciproquement entre vitesse et taux d'erreur). S'il y a des raisons de penser que le phénomène étudié, la discrimination visuelle, ne respecte pas cette contrainte de l'indépendance entre taux d'erreur et temps de réaction, alors le modèle psychométrique est invalide, et la mesure X de la compétence ne peut être cohérente par rapport à la valeur A recherchée. Par extension de ce raisonnement, si aucun modèle envisageable n'est satisfaisant, cela conduit à penser que le phénomène lui-même n'est peut-être pas suffisamment défini pour être mesurable.

Revenons à l'indice IAT. Les auteurs constatent qu'aucun des modèles psychométriques envisageables pour l'indice IAT n'est empiriquement cohérent avec les résultats qu'ils obtiennent dans plusieurs exemples. Le premier exemple tente de mesurer l'inclination des étudiants pour les études artistiques vs. les études mathématiques. Les résultats montrent que les associations compatibles et incompatibles varient selon les individus de façon indépendante. Les étudiants peuvent apprécier les unes ou les autres, les unes et les autres, ou ni les unes ni les autres, il s'en suit que faire la différence des temps de réaction n'a aucune signification. Dans un second exemple, les auteurs proposent une variante méthodologique mesurant indépendamment les associations positives, négatives, Blancs et Noirs en les associant à une condition supposée neutre. Les résultats vont dans le sens d'une grande complexité qui ne peut être captée par le protocole de calcul de l'IAT par différence des temps de réaction. En particulier certains participants n'associent positivement ni les blancs ni les noirs en regard de la condition neutre et d'autres associent positivement et les blancs et les noirs. Il semble donc que l'exigence de polarisation nécessaire à la structure de l'IAT ne soit pas remplie. Autre résultat, il semble qu'il y ait un corrélation entre les déclarations explicites de racisme et les mesures implicites de sentiments négatifs pour les noirs, mais avec aucune des trois autres associations implicites mesurées (noirs positif, blancs négatif, blancs positif). L'ensemble des résultats pointe donc dans la même direction : le protocole d'obtention de l'indice IAT s'appuie sur des suppositions fortes de covariation des préférences qui semblent devoir être mises en doute tant au niveau opérationnel des temps de réaction qu'au

niveau conceptuel.

Les promoteurs de l'IAT ont pris en compte les critiques exprimées et ont proposé un nouveau protocole dans une nouvelle version du protocole de calcul incluse dans un manuel en 2009 (Greenwald et al. 2009) [114]. Ces améliorations n'ont pas substantiellement changé la question du modèle psychométrique, et les promoteurs ont ensuite insisté sur le fait que l'indice Race IAT n'était pas une mesure individuelle mais prenait son sens lorsque appliqué à toute une population il permettait la mise en exergue d'un phénomène collectif.

A ce point de notre description, il est indispensable de préciser que les promoteurs de l'IAT et les opposants sont tous des scientifiques d'institutions reconnues et que les résultats ont été débattus par de nombreux groupes de chercheurs dans de très nombreuses publications (parmi les 21000 articles décomptés ci-dessus) et ont fait l'objet de méta-analyses. Il est naturellement impossible de rentrer ici dans le détail de toutes ces publications, et je vais simplement souligner quelques accords et quelques désaccords auxquels ces débats ont conduit après 20 ans d'expérimentations et, en particulier, de réplications. Ma principale proposition dans la prochaine section sera que ces réplications ont été un facteur clé pour le processus conduisant à une synthèse minimale partagée sur ce qu'est le phénomène IAT.

4.5.4 La synthèse minimale entre opposants et promoteurs de l'IAT

Tout d'abord, reprenons les quatre principales conclusions instruites par l'article initial de 1998 et proposées par les promoteurs de l'IAT. Un, le phénomène IAT existe : pour certains concepts qui donnent lieu à une évaluation positive ou négative, les temps de réaction diffèrent selon que l'on associe chaque concept avec l'évaluation qui lui est implicitement liée ou inversement. Deux, appliqué à des populations, le phénomène IAT permet de montrer que ces populations associent différemment ces concepts. Trois, construire un indice à partir des temps de réaction est une tâche statistique complexe. Quatre, l'indice implicite n'est que faiblement corrélé avec les résultats des méthodes explicites ce qui prouve que, si les unes et les autres sont des méthodes produisant des mesures valides, alors elles ne mesurent pas la même chose.

En parallèle de ces résultats théoriques, et dès 1998, l'IAT était proposé au public sur le site de Harvard, démontrant ainsi de facto qu'il était facile de le mettre techniquement en œuvre. Cette utilisation à grande échelle de l'IAT a été lancée en s'appuyant sur deux suppositions complémentaires qu'il est souhaitable d'expliciter : d'une part, l'indice IAT mesuré pour une personne dans le cadre d'une utilisation unique du site Web est suffisamment valide

pour être communiquée à la personne passant le test. Et d'autre part, plus important, communiquer ainsi un indice IAT est utile car cela peut permettre à cette personne de prendre conscience de son biais raciste implicite et, par la suite d'entreprendre de se corriger. Aucune de ces deux suppositions n'est une conséquence directe de ce que contient l'article de 1998.

Dix ans après l'article initial, les promoteurs de l'IAT ont publié en 2009 (Greenwald et al. 2009) [114] une méta-analyse visant à évaluer la validité de l'IAT en tant qu'outil de prédiction du comportement et du jugement (IAT vs. BAT). L'étude portait sur 122 articles et 184 échantillons pour 14900 sujets. Les articles étaient sélectionnés parce que comprenant des mesures suite à des tests IAT et BAT. Certains articles comportaient également des mesures explicites (156 sur les 184 échantillons) permettant des comparaisons croisées (IAT EAT BAT). Citons les conclusions de cette méta-analyse :

> La présente revue justifie la recommandation de l'utilisation conjointe de l'IAT et de mesures explicites pour prévoir le comportement. Même si le pouvoir prédictif de ces deux types de mesures varie considérablement d'un domaine à l'autre, l'utilisation simultanée des deux types apporte un gain par rapport à l'utilisation d'une seule méthode. La revue montre que le pouvoir prédictif des méthodes explicites est particulièrement bas dans les domaines socialement sensibles et que l'apport des mesures implicites y est relativement plus élevé. Les études portant sur les comportements racistes sont dans cette perspective particulièrement sensibles et dans ces domaines le pouvoir prédictif de l'IAT est significativement supérieur à celui des mesures explicites.

En 2013, les opposants à l'IAT ont mis en doute ces conclusions optimistes. Ils ont conduit une autre méta-analyse visant à évaluer la capacité de l'IAT à prédire un comportement discriminatoire (Oswald et al. 2013) [176]. La méta-analyse couvre 46 articles comportant 308 comparaisons IAT vs. BAT et 278 comparaisons EAT vs. BAT. Citons-en la conclusion beaucoup moins positive que la précédente [40] :

> L'engouement initial de la découverte de l'effet IAT a fait naître l'espoir que l'IAT puisse ouvrir une fenêtre sur les sources inconscientes des comportements discriminatoires. Cet espoir a été entretenu par des études montrant des corrélations statistiquement significatives entre le score au test IAT et certaines mesures de la discrimination ainsi que par la découverte par Greenwald et Poehlman en 2009 que le pouvoir prédictif de l'IAT de comportements racistes envers les afro-

40. Traduction de l'auteur.

> américains et d'autres minorités était supérieur au pouvoir prédictif des mesures explicites. Le présent examen attentif des critères utilisés montre toutefois que l'IAT apporte peu d'éléments sur qui va discriminer qui, et ne donne rien de plus en regard des mesures explicites des biais. L'IAT est une contribution innovante à la quête pluri-décennale d'indicateurs subtils des préjugés, mais les résultats de la présente méta-analyse indiquent que la psychologie sociale doit poursuivre cette quête vers une mesure non-invasive qui permette une prédiction fiable des discriminations. En synthèse, de simples mesures explicites des biais conduisent à des prédictions qui ne sont pas pires que celles de l'IAT. Si les chercheurs s'étaient concentrés sur le développement de ces mesures explicites en visant à diminuer les biais qui les entachent, elles auraient encore pu donner de meilleurs résultats.

Pour répondre à cet article de 2013, les promoteurs de l'IAT ont publié un article en 2015. Cet article de 2015 est central pour mon argumentation (Greenwald, Banaji et Nosek 2015) [112] : les auteurs reconnaissent que les résultats des réplications massives regroupés dans les méta-analyses montrent que l'IAT n'est pas une méthode fiable pour une mesure individuelle visant à un diagnostic de racisme implicite pour une personne donnée :

> Les mesures issues de l'IAT ont deux propriétés qui rendent problématique leur utilisation pour classifier des personnes comme susceptibles de s'engager dans des comportements discriminatoires. Ces deux propriétés sont la modeste fiabilité quand on applique le test plusieurs fois (pour l'IAT, le résultat est typiquement entre r=.5 et r=.6 [41], voir Nosek et al. 2007) et une faible à moyenne validité prédictive du comportement. En conséquence, utiliser ces résultats pour un diagnostic personnel présente de forts risques de classification erronée.

Cette formulation est celle de promoteurs de l'IAT et les opposants retiendraient certainement des valeurs encore plus défavorables quant à la fiabilité et la puissance prédictive de l'IAT ainsi que pour un inventaire des nombreux problèmes posés par ce protocole [42]. Néanmoins, les promoteurs et les opposants arrivent à un accord sur plusieurs points importants. Le phénomène IAT existe et il y a une différence de temps de réaction significative entre les associations compatibles et les associations incompatibles. Et opposants et promoteurs sont d'accord sur le constat que ce phénomène ne permet pas de mesurer la propension à la discri-

41. L'indice r de fidélité test – retest mesure la corrélation entre plusieurs itérations d'une même mesure à des moments différents. L'indice est significatif de la stabilité de la mesure. Il vaut 1 si la mesure est parfaitement stable et 0 si les itérations sont parfaitement indépendantes. Une valeur qualifiable de bonne par des psychologues serait 0,8.

42. Pour un éclairage de la presse sur tous ces problèmes, voir `https://qz.com/1144504/the-world-is-relying-on-a-flawed-psychological-test-to-fight-racism/`

mination au niveau individuel. L'indice RaceIAT n'est pas assez fiable pour classer ainsi une personne particulière.

Ma proposition est ici de considérer que le fait d'accepter la démarche scientifique pour base de discussion avec, en particulier, l'exigence de réplications par des équipes indépendantes a permis l'établissement d'une vision partagée, que je propose d'appeler la synthèse minimale :

— Le phénomène IAT existe et donne un accès à l'association implicite entre concepts.
— Ce phénomène est facile à mettre en œuvre mais difficile à traduire en un indice fiable. Cette élaboration est statistiquement complexe.
— L'indice RaceIAT n'a pas une fiabilité suffisante pour établir un diagnostic au niveau individuel, et encore moins si l'indice ne s'appuie que sur une seule réalisation du test.

L'accord sur cette synthèse minimale est un exemple de ce à quoi permet d'aboutir, même dans un contexte difficile, l'acceptation par toutes les parties de la réplication en tant que composante de la méthode expérimentale. Les différents arguments développés tant par les promoteurs que par les opposants dans les articles de 2013 et 2015 s'appuient principalement sur les méta-analyses des ensembles de données issues des réplications des expériences sur la base du protocole défini en 1998.

L'accord atteint est bien sûr imparfait, de nombreuses controverses ne sont pas résolues, comme par exemple celle portant sur les critères qui conduisent à inclure ou exclure une étude dans une méta-analyse. Soulignons également que l'accord a été long à établir, pratiquement 20 ans (1998 – 2015), mais tout cela ne doit pas occulter le résultat : une synthèse minimale a pu être établie et elle est appuyée sur la réplication et les méta-analyses.

4.5.5 Au delà de la synthèse minimale

A ce point du débat, j'ai proposé deux choses, l'une est l'existence d'une synthèse minimale entre opposants et promoteurs de l'IAT, l'autre est que cette synthèse minimale a été obtenue dans un contexte difficile par l'acceptation de toutes les parties de la réplication comme méthode de travail. Je vais maintenant aborder le point des profonds désaccords qui subsistent au-delà de cette synthèse minimale. On peut regrouper les questions qui conduisent à ces désaccords selon deux points de vue. D'abord celui de la recherche de connaissances fiables sur le phénomène IAT : comment améliorer notre compréhension des associations implicites entre concepts et de leurs liens à nos comportements ? Ensuite celui de la recherche d'outils permettant de lutter contre les discriminations : comment utiliser l'IAT en ce sens ? Ces deux

points de vue diffèrent profondément, le premier vise à la connaissance pour elle-même, le second pour les moyens d'action qu'elle donne. Le premier n'a pas d'échéances sociales ou politiques, le second oui. L'IAT s'analyse du premier point de vue comme un indice manquant de fiabilité, et qu'il faut donc améliorer ou remplacer, mais, du second point de vue, comme un outil, certes imparfait, mais utilisable immédiatement pour participer à l'effort contre les discriminations aux États-Unis.

Les questions relevant du premier point de vue, qu'on peut qualifier d'épistémique, sont par exemple :

— Peut-on améliorer le protocole, ou en inventer d'autres, pour que la mesure de l'association implicite soit plus fiable en vue d'un diagnostic individuel ?

— L'indice IAT n'est pas fiable pour un diagnostic individuel, mais est-il fiable pour mesurer les liens entre concepts au sein d'une population ?

— Même s'il ne constitue pas une mesure fiable, on peut envisager son utilisation en tant que moyen d'action, et comment mesurer alors les réactions d'un participant à qui on a révélé un supposé racisme implicite ? Quel est l'effet sur un groupe si une proportion significative de la population est soumise au test IAT ?

— ...

Les questions relevant du second point de vue, pragmatique, sont par exemple :

— L'IAT est en place sur le site Internet de Harvard, et a été utilisé par des millions de personnes montrant qu'une grande proportion des américains est implicitement raciste. Comment prendre en compte de façon efficace dans le cadre de la lutte contre les discriminations à la fois cette utilisation et le constat du manque de fiabilité ?

— L'IAT est utilisé dans des activités, commerciales ou non, présentées comme des « applications de la science à la pratique et à l'innovation sociale », comment limiter les abus potentiels liés au manque de fiabilité scientifiquement reconnu dans ce contexte tout en poursuivant son utilisation ?

— Si on utilise l'IAT pour lutter contre le racisme, est-ce que les études qui le critiquent ou en précisent les limites peuvent être reprises par des organisations racistes ? Et faut-il donc également lutter contre ces études, même si elles sont empiriquement pertinentes ?

Je ne vais pas ici tenter d'apporter des éléments d'argumentation aux questions relevant du premier point de vue épistémique. Je vais plutôt souligner que les opposants et promoteurs de l'IAT ayant déjà pu arriver à un accord sur une première synthèse minimale partagée, il n'y a pas d'impossibilité de principe à un tel accord et, en prolongeant le travail de recherche,

il est possible que les équipes arrivent à répondre à toutes ces questions épistémiques en s'appuyant sur la méthode scientifique expérimentale, dont sur les réplications comme outil important au sein de cette méthode. Cet accord sera certainement long à obtenir, puisqu'il a fallu 20 ans pour la synthèse minimale, mais je dis simplement qu'il n'est pas impossible d'arriver à un tel accord car cela c'est déjà produit.

Je ne vais pas non plus entrer dans l'argumentation relative aux questions relevant du second point de vue, pragmatique. Mais je veux souligner à quel point ces questions diffèrent des précédentes. D'abord elles sont très controversées et objets de prises de position politiques par voie de presse. Et surtout, elles ont résisté à tout accord jusqu'à présent car il semble que la réalité des caractéristiques empiriques de l'indice IAT soit sans influence possible sur ces questions : il n'y a pas de chemin partant de l'accord sur la synthèse minimale pour aller soit vers l'arrêt du test soit vers sa poursuite malgré son manque de fiabilité et, plus précisément, il n'y a pas de chemin que puissent partager les deux camps. La lutte contre les discriminations est trop importante pour les promoteurs de l'IAT pour pouvoir envisager d'en limiter l'utilisation. Pour les opposants, utiliser ainsi un test non fiable en se prévalant d'une démarche scientifique est tout aussi inacceptable.

Je me propose maintenant de chercher la racine de cette opposition dans le conflit entre deux types de valeurs qui s'avèrent incompatibles dans le cas de l'IAT. D'un côté, première option, l'IAT ne devrait pas être utilisé en tant que test sur une personne puisqu'il est acté qu'il n'est pas fiable. D'un autre côté, seconde option, si en faisant passer le test à toute une population, on a une arme contre le racisme implicite dans cette population, alors il faut l'utiliser. Le débat est principalement politique et donne dans la première option la priorité aux critères épistémiques, et à l'importance de l'intégrité scientifique, et dans la seconde la priorité à l'action sociale et à la lutte contre les discriminations. [43]

Je voudrai ici mettre en lumière que promoteurs et opposants partagent bien les deux objectifs, d'un côté la connaissance du comportement humain, et de l'autre l'espérance que cette connaissance permettra de conduire des politiques efficaces contre la discrimination [44]. Ce double objectif est exprimé dès l'article de 1995 par les promoteurs de l'IAT et régulièrement repris comme une évidence dans les articles des promoteurs comme des opposants. Bien sûr, chacun des deux camps pourra mettre en doute la sincérité des chercheurs de l'autre camp, mais il est plus proche de ce qui est explicité dans ces articles de reconnaître que

43. Cette préoccupation est à la base de la création de la Heterodox Academy qui combat l'abandon des valeurs épistémiques dans l'université américaine avec, en particulier, Jonathan Haidt. Voir `https://heterodoxacademy.org/`

44. Remarquons que la grande majorité des chercheurs en sciences sociales exprime une forte adhésion simultanée à ces deux objectifs. A titre d'exemple d'un tel engagement, voici un témoignage individuel pris au hasard : `http://www.betsylevypaluck.com/research/`

tous se donnent les deux objectifs : et la meilleure connaissance du comportement humain, et la lutte contre les discriminations. En retenant cette similitude d'objectifs globaux, nous sommes conduits à proposer que l'opposition entre les deux camps provient des priorités différentes qu'ils donnent à ces deux objectifs et, concomitamment, aux poids différents donnés à deux systèmes de valeurs. D'un côté les valeurs de l'intégrité scientifique, principalement liées à l'acquisition de connaissances et à la défense de la démarche scientifique comme voie d'accès privilégiée à cette connaissance et, de l'autre côté, les valeurs de l'action sociale, principalement liées à l'amélioration de la condition humaine et à la défense de l'action politique et sociale.

Si cette approche est valide, opposants et promoteurs de l'IAT donnent des priorités différentes à des systèmes de valeurs qui sont ici pour partie incompatibles. La traduction concrète en est que, pour la partie compatible de ces systèmes de valeur, les équipes peuvent s'accorder sur la synthèse minimale, mais pour la partie incompatible, les désaccords sont profonds, perdurent, et ne peuvent être résolus par la réplication des expériences ni, plus largement, par la démarche scientifique expérimentale qui relève du seul point de vue épistémique.

Remarquons enfin que les deux objectifs de connaissance et d'amélioration sociale sont associés tant par les opposants que par les promoteurs de l'IAT, sans que ni les uns ni les autres n'argumentent sur la pertinence qu'il y a à courir ces deux lièvres à la fois. Cette association, ici ironiquement implicite, ne va pas de soi. Par exemple les améliorations sociales peuvent être apportées sans s'appuyer sur une connaissance de type scientifique. Et, inversement, nous pouvons avoir des connaissances scientifiquement établies qui ne conduisent pas pour autant à quelque amélioration que ce soit.

Reconnaître la dissociation profonde entre ces objectifs de connaissance et d'action, ainsi que les systèmes de valeurs différents qui les structurent, permet d'une part de clarifier le débat théorique et d'autre part de clarifier les enjeux pratiques au moment d'allouer des budgets de recherche à l'un ou l'autre camp relativement à l'IAT.

4.5.6 Connaissance et action

Les philosophes, comme tous les êtres rationnels confrontés à des problèmes complexes récurrents, se simplifient la vie dans les cas courants par le recours aux heuristiques [45]. Parmi ces heuristiques, l'une traite des situations dans lesquels des groupes d'experts, tous compétents, n'arrivent pas à se mettre d'accord et poursuivent sans fin d'inutiles arguties. Le

45. Voir Alan Hajek "Philosophical Heuristics and Philosophical Methodology" in (Paul et Kaufman (eds.) 2017) [180]

principe de cette heuristique est de rechercher parmi les bases communes sous-jacentes à ce domaine d'expertise une erreur partagée par les experts de l'un et l'autre camp. Ainsi, par exemple, face à des experts de la vie des anges qui débattraient sans fin pour savoir s'ils ont ou non un sexe, tous les arguments des uns contre les autres détaillés à l'envie ne feront qu'un rideau de fumée ajoutant la confusion à la confusion. En revanche rechercher l'erreur qu'ils partagent donne accès à une réponse beaucoup plus claire : les anges n'existant pas la question de leur sexe ne se pose pas. Je me propose d'appliquer cette heuristique au cas de l'IAT, et au fait que, malgré la synthèse minimale à laquelle sont arrivés promoteurs et opposants, ils n'arrivent pas à tirer de conclusion partagée du fait que le phénomène IAT existe mais que l'indice IAT qui en est tiré n'est pas assez fiable pour diagnostiquer la tendance à la discrimination chez un individu particulier.

Reprenons, très sommairement, ce qui diffère entre opposants et promoteurs. Pour les opposants, l'indice IAT n'est pas une mesure fiable, il n'a pas de pouvoir de prédiction et n'est pas justifié par un modèle psychométrique. Par conséquence, et dans la mesure où il n'est pas justifié, nous ne pouvons l'utiliser comme diagnostic individuel (et les promoteurs en sont d'accord) ni pour l'étude d'une population (et les promoteurs ne sont pas d'accord). L'utilisation répandue de l'IAT au travers des sites de Harvard et de multiples études psychologiques constitue donc plus un problème qu'une solution aux questions soulevées par les associations implicites entre concepts. En particulier, cette généralisation détourne les chercheurs en psychologie des précautions nécessaires à utiliser un indice aussi peu fiable et les détourne également des recherches visant à améliorer les autre mesures tant implicites qu'explicites. Dans ce contexte, les opposants ne peuvent que s'offusquer de la prétention à la scientificité des acteurs commerciaux qui utilisent l'IAT. L'indice IAT n'est pas fiable et son utilisation commerciale ou politique est hors du champ scientifique.

Pour ses promoteurs, l'IAT est efficace pour lutter contre le racisme car si quelqu'un apprend qu'il souffre d'un biais raciste implicite, il aura à cœur, mis en situation, d'en éviter la concrétisation. Les promoteurs reconnaissent le manque de fiabilité de l'IAT, mais ni plus ni moins que d'autres mesures psychologiques, et trouvent là une raison de chercher à l'améliorer mais certainement pas à l'écarter[46]. Les mesures explicites sont également connues

46. Cet aspect n'est pas limité au seul indice Race IAT relatif au racisme, voici un exemple de conclusion des promoteurs de l'IAT dans le domaine du marketing (Brunel, Tietje et Greenwald 2004) [34] : "En conclusion, l'IAT est une mesure valide et intéressante de la connaissance implicite d'un produit par le consommateur. Nous reconnaissons volontiers que cette mesure n'est pas une panacée et que des limitations subsistent. En particulier, la structure différentielle de cet indice le rend impropre à certains contextes. Des études complémentaires sont également nécessaires pour mieux comprendre les bases théoriques et physiologiques du fonctionnement de l'IAT ainsi que, finalement, sur le lien entre l'IAT et le comportement. Néanmoins, il apparaît que les bénéfices et gains potentiels pour la rechercher sur les comportements des consommateurs outrepassent largement ces préoccupations et l'IAT ouvre une possibilité de mesure des penchants implicites avec de solides propriétés psychométriques."

pour souffrir de nombreux biais, et sont toujours malgré tout d'utilisation courante. L'importance des efforts à mener pour lutter contre le racisme, explicite et implicite, est telle qu'il faut continuer à utiliser toutes les ressources possibles, y compris l'IAT. L'indice IAT est donc scientifique (autant qu'une mesure psychologique puisse l'être) et la considérer comme telle appuie l'action publique contre le racisme et la rend plus efficace.

A cette opposition, je propose d'appliquer l'heuristique de l'erreur partagée et, parmi les erreurs possibles (dont l'évidente appartenance à la même espèce humaine et aux biais qui la caractérisent), je souhaite souligner que opposants et promoteurs s'appuient de façon communément exagérée sur le lien qu'ils font entre la connaissance scientifique et la capacité d'action. Bien que souvent pertinent, ce lien est certainement exagéré dans le cas de l'IAT car la connaissance scientifique de la psychologie humaine qui permettrait de lier les temps de réaction, l'association implicite entre concepts et la propension aux comportements racistes n'est ni nécessaire ni suffisante pour structurer des politiques et des plans d'action pour lutter contre le racisme. La croyance inverse en un lien direct entre la connaissance scientifique et la puissance d'action, partagée entre promoteurs et opposants, se traduit de multiples façons. Il est important pour la lutte antiraciste des promoteurs de s'appuyer sur des déclarations de scientificité et, donc, d'être intolérants envers d'autres scientifiques qui mettent en doute leur méthode. Il est important pour les opposants de ne pas risquer de dévaloriser la démarche scientifique par l'utilisation généralisée d'un test non fiable, et donc d'affirmer que le test lui-même n'est pas efficace car il n'est pas scientifique.

Ceci me conduit à suggérer qu'opposants et promoteurs partagent cette surestimation d'un lien implicite, de nature logique entre la connaissance et l'action dans le cas de l'indice Race IAT. Cette suggestion explique bien pourquoi la qualification de mesure scientifique pour l'indice IAT est un enjeu important pour chacun des deux camps.

Si l'on accepte la proposition qu'action et connaissance n'ont qu'un lien faible dans ce domaine (au moins aujourd'hui) alors les promoteurs de l'IAT pourront, pragmatiquement, poursuivre leurs tests en application de choix de politique publique et pourront consacrer du temps à l'estimation des conséquences de son application dans différents contextes. Les opposants pourront de leur côté continuer à développer leurs recherches psychologiques des liens entre comportement, concepts, déclarations explicites, et mesures implicites des associations de concepts en tant que prédicteurs du comportement.

Le fait d'accorder un poids très différent à des systèmes de valeurs tournés l'un vers la connaissance et l'autre vers l'action conduit à un blocage dû à l'erreur partagée de l'existence d'un lien logique entre ces deux systèmes de valeurs. Cette erreur partagée aveugle les deux

camps et les empêche d'accéder à un constat plus économe : les phénomènes sur lesquels ils travaillent sont trop complexes en regard des outils scientifiques dont ils disposent. Ni l'expérimentation, ni a fortiori la réplication ne sont de nature à construire un accord partagé sur les questions qui empiètent sur les deux domaines disjoints de l'approche empirique scientifique et de l'action sociale. Privilégier l'une ou l'autre ne peut se mesurer à la même aune et ne peut être jugé selon un même point de vue.

La suggestion, appuyée sur l'heuristique de la recherche de l'erreur partagée par les experts d'un même domaine, n'épuise certainement pas tout ce qu'il conviendrait d'analyser dans l'exemple passionnant de l'IAT. Je propose néanmoins de considérer qu'assouplir le lien logique entre la connaissance psychologique et l'action sociale et mettre en avant que ces deux aspects mettent en jeu des systèmes de valeurs différents et, parfois, incompatibles, est de nature à sortir des débats stériles de façon plus utile que les attaques ad hominem par voie de presse qui ont joué un rôle trop important dans ces controverses. [47]

4.5.7 Conclusion : ce que peut la réplication

Le cas de l'IAT est complexe et a donné lieu à de brûlantes controverses. Je l'ai ici simplifié à l'extrême en résumant 20 ans d'études qui ont donné lieu à de multiples expériences répliquant le phénomène IAT dans de multiples contextes. Ce grand nombre de réplications, appuyé sur l'acceptation par tous les psychologues des principes de la démarche scientifique expérimentale, qu'ils soient promoteurs ou critiques de l'IAT, a permis de construire une synthèse minimale en trois points. Premier point, l'effet IAT existe et donne accès aux associations implicites entre concepts. Deuxième point, l'effet IAT est facile à mettre en œuvre mais il est difficile d'en tirer un indice fiable. Troisième point, cette fiabilité est en particulier insuffisante pour utiliser l'indice RaceIAT en tant que diagnostic individuel de la propension au comportement raciste.

En prolongement de cette synthèse minimale, on peut espérer que la même démarche de réplications et de méta-analyses produira des réponses consensuelles à des questions aujourd'hui encore disputées : comment améliorer l'IAT et comment le comparer à d'autres mesures implicites ? Jusqu'où les mesures implicites de l'IAT sont-elles corrélées avec les mesures explicites et les comportements ? Et, finalement, quel est l'utilité de l'IAT dans la lutte contre les discriminations ?

Toutes ces questions sont accessibles à la démarche scientifique expérimentale. Elles

47. Voir par exemple l'article de Jesse Singal dans "the cut" : https://www.thecut.com/2017/01/psychologys-racism-measuring-tool-isnt-up-to-the-job.html

peuvent être difficiles et longues à démêler, ce qui est probable au vu des 20 ans qu'il aura fallu pour arriver à la seule synthèse minimale évoquée ci-dessus.

Ce que ne peut pas la réplication

Mais, avant que ces questions ne soient résolues, l'IAT a déjà été utilisé pour des millions de tests et le désaccord entre promoteurs et critiques est profond. D'un côté, « nous ne pouvons attendre que ces questions soient résolues pour lutter contre le racisme ». De l'autre côté « nous ne pouvons, en tant que scientifiques, accepter l'utilisation d'un test non fiable ».

Le choix est, fondamentalement, entre poursuivre la quête scientifique et commencer à agir immédiatement. Ce choix ne peut être instruit par une démarche empirique, et bien sûr répliquer des expériences devient inutile, ou, plus précisément, range cette solution dans le premier camp. Le choix se fait entre deux systèmes de valeurs qui s'avèrent dans ce cas et en ce moment incompatibles.

L'utilisation d'une heuristique consistant à rechercher l'erreur partagée quand une opposition semble insoluble entre experts d'un même domaine m'a amené à proposer que promoteurs et critiques de l'IAT partagent une même surestimation du lien logique entre connaissance scientifique et capacité d'action sociale.

La philosophie morale expérimentale comporte, en son intitulé même, l'amorce du risque d'être confrontée à la même erreur partagée entre être « morale » et être « expérimentale ». Penser qu'un lien logique puissant existe entre ces deux familles d'objectifs risque de conduire au même blocage que celui que l'exemple de l'IAT m'a permis de mettre en avant : les réplications ou méta-analyses n'ont pas permis de construire de vue commune aux deux systèmes de valeurs de l'intégrité scientifique et de l'engagement social contre les discriminations.

4.6 L'analyse de l'effet Knobe : instabilité des notions

4.6.1 L'effet Knobe

En 2003 Joshua Knobe [138] a présenté une expérience mettant en scène un chef d'entreprise qui lance un projet motivé uniquement par sa rentabilité. Le chef d'entreprise marque son indifférence aux éventuelles conséquences sur l'environnement que son adjoint lui présente et insiste sur le fait que seule la rentabilité lui importe. Le projet se déroule. Il s'avère être très rentable et avoir les conséquences sur l'environnement prévues par le collaborateur, qui peuvent être soit positives soit négatives selon les scénarios. L'expérience consiste à demander à des participants si le chef d'entreprise a intentionnellement amélioré ou nui à

l'environnement dans chacun des scénarios, positif ou négatif. Les réponses montrent alors une nette dissymétrie : les participants disent que le chef d'entreprise a intentionnellement causé les dégâts et qu'il faut le blâmer quand les conséquences sont négatives, mais qu'il n'a pas intentionnellement amélioré l'environnement et qu'il n'est pas à féliciter quand les conséquences sont positives.

Cette dissymétrie expérimentalement constatée semble faire dépendre l'attribution d'intentionnalité, le jugement de considérer que le chef d'entreprise a ou non intentionnellement agit sur l'environnement, d'un jugement moral porté sur le résultat, c'est-à dire l'amélioration ou la dégradation de l'environnement. Ce résultat s'inscrit dans la problématique psychologique complexe qui lie intentionnalité, causalité et responsabilité pour conduire au jugement moral et au blâme ou à la louange. Reprenons pour l'appliquer à ce contexte le schéma général de la proposition morale présenté au premier chapitre (voir 1.4 page 80) :

« O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien. »

Dans notre cas, O est le philosophe qui mène l'étude, S est le sondeur administrant le questionnaire, X est le participant répondant au questionnaire, E est sa réponse, et AYZC sont les éléments du scénario présentés ou, plus précisément et selon les chercheurs, ce que X en a compris ou ce que O en pense.

En première lecture ce schéma suppose implicitement d'une part que Y a fait A intentionnellement, d'autre part que c'est bien ce qu'a fait Y qui a causé A et enfin que X est ainsi en droit de juger Y moralement responsable de A. Ce n'est que quand ces conditions sont respectées que X peut juger et exprimer son jugement moral E à S.

Dans les cas habituels, Y fait A intentionnellement par construction du scenario, et X ne peut avoir aucun doute sur ce lien d'intentionnalité. Le scénario du chef d'entreprise est choisi pour supprimer cette certitude en faisant de A un effet indirect non souhaité d'une action qui est l'objectif réel de Y[48]. Le constat expérimental est alors que lorsqu'on interroge X pour savoir s'il considère que Y a intentionnellement causé l'effet indirect A, sa réponse dépend du résultat de l'action A. Si les conséquences de l'acte A sont bonnes, Y n'a pas intentionnellement amélioré l'environnement, si les conséquences sont mauvaises, Y a intentionnellement dégradé l'environnement. [49]

48. Cette configuration est connue des philosophes moraux pour être à la base de la doctrine dite du double effet : on peut faire un Mal si celui-ci est un effet non désiré (ou non prévu, ou inévitable, ou hors de contrôle, ou tout autre critère selon les multiples variantes possibles diminuant l'intentionnalité) d'une action ayant conduit à un plus grand Bien.

49. On peut remarquer que, contrairement aux dilemmes du tramway, la situation fictionnelle décrite est tout à fait réaliste et correspond bien à ce que X peut avoir comme information quand il juge de la responsabilité de tel ou tel responsable sur la base d'un article de journal par exemple.

Les psychologues expérimentaux ont, à la suite de l'article de 2003, multiplié les études pour tenter d'affiner les relations complexes entre intentionnalité, causalité et responsabilité morale dans ce cas particulier de l'effet indirect qui permet d'introduire une différenciation entre les actions causées intentionnellement et les actions causées non intentionnellement [50]. L'effet Knobe constitue un des exemples les plus souvent cités en tant qu'apport de la philosophie expérimentale à la philosophie morale [51]. Les études sur l'intentionnalité de l'action consistent pour partie à rechercher un modèle de relation entre les concepts en jeu qui soit explicatif du jugement moral de X dans les différents scénarios proposés avec leurs multiples variantes jouant sur tous les éléments de la situation 'XESYAZC'.

Un premier examen de l'abondante littérature relevant de ces questions m'a conduit au constat que chaque modèle ainsi construit instituait également une redéfinition des concepts d'intentionnalité, de libre arbitre, de responsabilité morale et de causalité et, simultanément, faisait varier l'interprétation donnée de l'effet Knobe. Ce constat est d'ailleurs partagé par de nombreux auteurs, comme je vais le montrer ci-après, citons pour l'instant l'exemple de (Mukerji 2019 p 93) [167] :

> L'effet Knobe est intéressant si d'une part on accepte la vue méta-philosophique des XPhi et si, d'autre part, on adopte une interprétation particulière de l'effet Knobe.

De nombreux philosophes et psychologues ont accepté ces deux conditions proposées par Mukerji et ont construit des scénarios pour explorer les différents éléments pouvant influer sur l'attribution d'intentionnalité. Mon objectif est maintenant de donner à voir le paysage conceptuel dessiné par ces études, prises dans leur ensemble.

4.6.2 Méthode de travail

Pour mieux apprécier les variations des philosophes et psychologues autour de l'effet Knobe et éviter un biais de sélection d'articles, j'adopte ici une méthode inspirée du ciblage des périmètres de recherche (Arksey 2005) [13] qui a pour objet premier de construire le périmètre d'un domaine d'étude à partir d'un corpus bibliographique. La méthode consiste à définir un ensemble de mots clés assez large, ici ce sera « intentionnalité, responsabilité, philosophie expérimentale », puis sélectionner sur ces mots clés l'exhaustivité des articles récents sur une base bibliographique de référence, ici la base de Sorbonne Université, puis res-

50. En ce sens, l'effet Knobe est connu dans toutes les cours d'école. Il justifie l'utilisation de l'argument « Je l'ai pas fait exprès » qui est censé éloigner la punition.
51. Voir par exemple (Mukerji 2019 p 87) [167]

treindre à une période assez courte pour rendre possible l'analyse exhaustive, ici 18 mois. De cet ensemble d'articles, analyser les résumés et éliminer ceux pour lesquels les mots clés ont conduit à une sélection inappropriée. Enfin, analyser le contenu des articles pour construire le schéma des variations conceptuelles bâti par chacun des articles. Remarquons que cette démarche n'est pas celle d'une étude académique du domaine, au sens en particulier où les références importantes mais anciennes sont écartées. Elle répond empiriquement à une autre question : quel est le paysage conceptuel actuel d'un domaine de recherche ?

Les domaines d'étude de la psychologie de l'intentionnalité et de la philosophie expérimentale liée à l'effet Knobe sont tous deux très majoritairement anglophones. La démarche a donc été menée en anglais. [52]

Sur la base bibliographique de Sorbonne Université, consultée de juin à septembre 2018, j'ai retenu les éléments définitifs obtenus du 20 au 22 septembre 2018. J'ai utilisé les mots clés suivants :

(moral philosophy experimental responsibility) (intentionality)

J'ai ensuite limité la profondeur historique pour descendre en dessous de 1000 articles :

(moral philosophy experimental responsibility) (intentionality) y :[2016-2018]

J'ai ainsi obtenu 563 articles. Un premier examen des résumés a montré qu'une partie quantitativement importante était constituée d'articles théoriques sans lien direct avec une approche empirique ou expérimentale. J'ai donc décidé de limiter l'échantillon en ce sens :

(moral philosophy experimental responsibility) (intentionality) y :[2016-2018] « experiment »

J'ai ainsi réduit la sélection à 395 articles, ce qui restait impossible à analyser avec des moyens manuels [53]. J'ai décidé de limiter la sélection aux articles faisant référence à l'effet Knobe.

(moral philosophy experimental responsibility) (intentionality) y :[2016-2018] « experiment » « Knobe effect »

Le résultat est alors de 58 articles dont j'ai lu exhaustivement les résumés. Parmi ces articles, j'en ai écarté 25 qui me sont apparus comme non pertinents pour la présente étude, ce sont principalement des ouvrages généraux de psychologie, où la partie liée à l'intentionnalité

52. Dans cette partie de la thèse, je ne traduis pas les mots clés utilisés de façon à éviter toute ambiguïté si une partie de ces sélections venait à être répliquée ou prolongée par un tiers.

53. L'utilisation d'outils automatiques d'analyse de texte permet peut-être d'ouvrir de nouvelles perspectives d'approches de corpus de gros volumes. Pour exemple, la construction automatique de l'ontologie formelle de la philosophie entreprise par l'université d'Indiana `https://www.inphoproject.org/`

est trop réduite pour être utile, et des articles sur le problème actuellement très étudié de l'intentionnalité et de l'agentivité des robots.

A l'issue de ce processus il reste 33 articles récents de philosophie ou de psychologie expérimentale faisant référence à l'effet Knobe dans le cadre d'une étude de l'intentionnalité et de la responsabilité morale. La liste de ces 33 articles, avec leurs caractéristiques, est reprise dans un tableau en Annexe.

Après récupération des textes des 33 articles sur Internet (dont la majorité via le site de la bibliothèque de la Sorbonne et une grosse minorité directement sur Internet), j'ai réalisé une première lecture et, à l'issue de cette première lecture, j'ai pu confirmer le constat principal de l'analyse ci-dessous : la multiplicité des directions que peut prendre l'analyse psychologique de l'attribution d'intentionnalité. Cette multiplicité marque la diversité des objectifs et points de vue de chaque chercheur dont il me faut maintenant rendre compte ici en reprenant chacun des 33 articles.

4.6.3 Analyses d'ensemble des 33 articles

Tout d'abord, de façon à donner une vue d'ensemble du contenu de ces 33 articles, je les ai qualifiés selon plusieurs critères de provenance et de contenu (voir le tableau en Annexe pour le détail).

Premier critère, le type de revue d'où provient l'article. Pour une majorité des deux tiers, 23 d'entre-eux, ce sont des revues spécialisées de psychologie, pour 8 d'entre-eux, ce sont des revues de philosophie. Un seul article est issu d'une revue de management et enfin un autre article provient d'une revue scientifique généraliste.

Deuxième critère, le lien à l'expérimentation. La grande majorité des articles, 29 sur 33, présente le résultat d'une expérimentation ou s'inscrit dans la suite immédiate d'une expérimentation. Les 4 autres articles sont plus théoriques ou génériques, tout en faisant référence à la méthode expérimentale et aux expérimentations en psychologie. Je ne les ai pas écartés de l'analyse parce qu'ils ne changent pas significativement les résultats des analyses ci-dessous, leur exclusion ou inclusion ne présente pas d'enjeu de méthode.

Troisième critère, la technique utilisée pour recueillir les données expérimentales. Sur les 29 articles présentant des expérimentations, 28 procèdent par questionnaire et un seul comporte des mesures physiques, ici de type IRM. Cette observation présente une certaine importance face aux déclarations souvent très emphatiques sur l'apport potentiel de ces nouvelles méthodes d'accès à l'activité cérébrale pour la philosophie morale. Leur apport, y compris à

la psychologie expérimentale, reste à ce jour quantitativement très limité.

Quatrième critère observé, le mode de recrutement des participants à ces expérimentations.

— Sur les 29 expérimentations, 15, soit plus de la moitié, sont réalisées en recrutant les sondés sur Amazon Mechanical Turk (AMT). Ce chiffre montre clairement la place immense prise par cette solution commerciale dans les études académiques.
— Il faut ajouter à cela 3 autres études qui utilisent des systèmes sur Internet équivalents à AMT du point de vue technique mais différents du point de vue de la relation commerciale entre la plateforme, le sondeur et le sondé.
— Ensuite, 8 études en restent à la méthode traditionnelle de recruter les participants parmi les étudiants de l'université qui abrite l'étude.
— 2 études concernent des participants particuliers, l'une des enfants dans une école maternelle et l'autre des juges de l'administration judiciaire française. Le mode de recrutement est lié à ces populations particulières étudiées.
— Enfin, dernier cas, une seule étude a recruté les participants directement parmi les passants dans la rue.

Cinquième critère, le nombre et l'ampleur des expériences rapportées dans les articles.

— Au total des 33 articles, 87 expériences ont été rapportées.
— Chacune des expériences a porté en moyenne sur 277 participants, avec un minimum de 23 pour l'expérience avec IRM et un maximum de 1636 participants sur Internet.
— 12 articles ne présentent qu'une ou deux expériences.

Remarquons la double faiblesse du nombre d'expériences, 12 articles, soit un tiers de l'échantillon, n'en présentent qu'une ou deux, et surtout du nombre de participants, en moyenne 277, qui apparaît étonnamment bas en regard du très grand nombre de paramètres à faire varier.

Sixième critère : la position en regard de l'asymétrie de l'effet Knobe.

— 10 articles ont pour objectif de donner une explication de l'effet Knobe
— 20 articles sont neutres, ils ne traitent pas directement de l'incidence de la valence morale sur l'attribution d'intentionnalité.
— 3 articles montrent que l'effet Knobe n'existe pas toujours, et même qu'il peut s'inverser selon le contexte.

Septième et dernier critère global : les thèmes philosophiques explicitement évoqués dans les articles.

De façon implicite ou explicite, tous les articles ont, par construction de la sélection, un lien

avec les thèmes relevant de la philosophie morale, la responsabilité morale et les jugements moraux. Je me concentre ici sur les thèmes philosophiques explicitement évoqués en appui ou en conclusion des articles à titre principal. [54]

— 14 articles se présentent comme purement psychologiques sans thème philosophique explicitement formulé.
— 7 articles font référence à la question du libre arbitre
— 4 articles font référence au dualisme ou au réductionnisme
— 4 articles font référence à l'apport des XPhis à la philosophie morale
— 3 articles font référence au relativisme culturel
— 1 article fait référence à l'implication « ought implies can » ou « il faut pouvoir pour devoir »

Ce sont donc 19 articles sur 33 qui font explicitement référence à un thème philosophique dans leur présentation ou dans leur conclusion alors que, rappelons-le, seuls 8 articles sont publiés dans des revues de philosophie. Ce constat suffirait à montrer empiriquement, s'il en était besoin, le lien étroit qu'entretiennent philosophie et psychologie expérimentale dans le domaine d'étude de l'attribution d'intentionnalité.

4.6.4 33 argumentations, et presque autant de distinctions conceptuelles

Avant d'analyser les distinctions conceptuelles mises en avant par les différents articles, il est nécessaire de rappeler les grandes lignes de l'ossature des conclusions de l'article initial de 2003 de Joshua Knobe [138]. Première conclusion, il existe expérimentalement une asymétrie d'attribution d'intentionnalité. Les participants considèrent qu'un chef d'entreprise qui lance un projet avec des conséquences prévues, auxquelles il ne s'intéresse pas, les a intentionnellement causées si elles sont négatives, mais pas si elles sont positives. Knobe propose, c'est la seconde conclusion de l'article, que le mécanisme suivant est suggéré par cette asymétrie :

— Le jugement moral intervient en premier, très rapidement.
— En conséquence de ce jugement, le participant désire blâmer ou louer le chef d'entreprise, et ces désirs ne sont pas de même intensité, le désir de blâmer est plus fort.
— L'intentionnalité est nécessaire au blâme comme à la louange.
— Pour faire droit à son désir de blâmer, le participant considère que le chef d'entreprise a intentionnellement dégradé l'environnement.

54. Pour simplifier l'analyse et sa présentation, je n'ai retenu ici pour chaque article que le thème philosophique principal, une analyse plus détaillée serait possible en considérant toutes les mentions principales et secondaires.

— Et comme le désir de blâmer est plus fort, l'attribution d'intentionnalité est également plus forte.

Mais ce mécanisme n'est que suggéré par l'article de 2003 et supposerait pour être validé, nous dit l'auteur, de nombreuses études empiriques complémentaires. Knobe lui-même a publié ensuite plusieurs articles exploitant cette asymétrie dans différentes directions (Knobe 2006) [139], (Knobe et Nichols (eds.) 2008) [140].

Je vais maintenant détailler les principales distinctions conceptuelles introduites par chacun des 33 articles pour l'analyse de l'intentionnalité et, concomitamment, pour l'interprétation de l'asymétrie de l'effet Knobe si ce point est explicitement soulevé dans l'article. L'argument principal de chaque article est décrit ci-dessous dans un paragraphe résumant (à l'extrême) l'argumentation proposée. Chacune des affirmations s'entend comme une déclaration de validité empirique en regard de l'expérimentation présentée dans l'article. Mon objectif n'est pas ici de mettre en question cette validité affirmée par les auteurs mais plutôt d'observer le paysage qui est décrit si on accepte l'ensemble de ces affirmations.

Les résumés sont construits pour faire ressortir ce que chaque article apporte de nouvelle distinction conceptuelle, il est donc nécessaire en préalable de rappeler la part importante qu'ils ont en commun afin de ne pas donner l'idée fausse que cette littérature constituerait un champ désordonné sans aucune unité. Tous les articles s'inscrivent, en prolongement de l'article initial de Knobe, dans l'étude de l'intentionnalité, et, plus globalement, de l'étude de la responsabilité humaine dans l'action comme problème psychologique. Une majorité d'article adopte une présentation conforme au standard des revues psychologiques qui sont les vecteurs cibles des auteurs. En une trentaine de pages il s'agit de :

— Premièrement, rappeler la question posée et les paramètres qui ont déjà été étudiés dans la littérature,
— Deuxièmement, annoncer qu'un paramètre important a peut-être été jusqu'à présent négligé, et donc que l'étude a pour objet d'étudier empiriquement si c'est le cas.
— Troisièmement de présenter la méthode de questionnaire utilisée.
— Quatrièmement, présenter les résultats obtenus, et vérifier si la probabilité de l'hypothèse nulle, c'est-à-dire que le paramètre étudié soit sans effet, est suffisamment faible, d'en déduire que le nouveau paramètre importe.
— Enfin, dans une cinquième partie de la présentation type, les auteurs tirent les conséquences psychologiques et philosophiques de la prise en compte de leur nouveau paramètre.

Le numéro d'article indiqué dans la liste ci-dessous renvoie à l'Annexe détaillant les 33

articles. Deux articles qui ne comportent pas de nouvelle distinction conceptuelle par rapport à Knobe 2003 ne sont pas cités. Il s'agit des articles 9 et 15 qui ne comportent pas d'expérimentation, (Fischborn 2018) [87] (Buckwalter 2016) [35]. Ce résultat est en lui-même est important et significatif du mode de travail des chercheurs en psychologie : l'énorme majorité des articles, 31 sur 33 présente une modification du réseau conceptuel selon lequel il convient d'analyser l'intentionnalité, et c'est l'objet principal de l'article.

Quand plusieurs articles s'appuient sur une distinction similaire, ils sont regroupés avec la première occurrence. Ainsi par exemple, les articles 1, 10 et 11 du début de cette liste insistent tous sur la nécessité d'une relation causale affirmée et perçue entre ce que fait l'acteur et les conséquences examinées pour que l'attribution de l'intentionnalité puisse avoir lieu. La grande majorité des articles n'a pas été ainsi regroupée, les distinctions apportées n'étant pas, au moins prima facie, de même nature. Pour permettre la mise en perspective des articles les uns avec les autres, je prendrai pour tous la notation proposée pour les jugements moraux :

« O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien. »

Article 1 – L'attribution par X de l'intention de Y à faire A présuppose que X perçoive une **relation causale** entre ce qu'à décidé de faire Y et les conséquences de A prises en compte par X dans son jugement. Un élément clé pour cela est qu'existe une **alternative**, Y a fait A mais aurait pu faire autre chose, et que X pense que Y a conscience de l'existence de cette alternative. (Hilton, McClure et Moir 2016) [117]

Article 10 – Les intuitions sur les relations causales sont entachées de deux type de biais. Le biais de moralité, qui nous fait surpondérer les relations causales quand elles sont utiles à justifier notre jugement moral, et le biais d'agentivité, qui nous fait surpondérer le besoin qu'il y ait toujours un agent à la source de toute action (quel que soit A, Y existe). Le premier biais justifie l'asymétrie de l'effet Knobe mais, plus important, les deux biais ensemble conduisent à poser que, en particulier dans le domaine moral, **l'évaluation des relations causales manque de fiabilité**. (Rose 2017) [195]

Article 11 – L'attribution de responsabilité suppose à la fois un lien causal entre l'acte et le résultat et, à la fois, le caractère intentionnel de l'acte. Les deux voies sont empiriquement confirmées. La part relative des deux voies varie de façon inattendue : si l'intention est forte, et même si elle vise un effet principal positif, elle aggrave la responsabilité quand A donne lieu à une conséquence négative indirecte. (Martin et Cushman 2016) [157]

Article 2 – Le désir de blâmer est lié à la **violation d'une norme** ou à la négligence

en regard de cette norme. Le respect d'une norme n'est pas suffisant pour être loué. De là provient l'asymétrie de l'effet Knobe. (Hindriks, Douven et Singmann 2016) [118]

Article 16 – La **norme violée** peut relever de la violence ou de l'impur[55]. Dans les deux cas on retrouve l'effet Knobe. Par contre l'obtention de l'effet dépend de l'orientation de l'acte vers un tiers ou vers soi-même (il faut Y<>Z pour que l'effet Knobe apparaisse). (Parkinson et Byrne 2017) [178]

Article 21 – Si le jugement moral s'appuie sur l'estimation par X de l'intention de Y, il suppose que X dispose d'une théorie de l'esprit (en anglais, Theory of Mind ou TOM) pour lire les intentions d'autrui. Si la TOM a une signature neuronale, alors on peut voir le rôle de l'intentionnalité dans le jugement moral à l'aide d'une IRMf. L'étude menée vise à mesurer l'activation des zones du cerveau reconnues comme liées à la TOM dans le cadre de jugement moraux concernant des actes de violence ou d'impureté. Il apparaît que la mesure de la TOM est différente selon le **type de norme violée**. En cas de violence, la différence entre accidentel et intentionnel est activée au cours du jugement moral, pas en cas de violation de règles de pureté. Les processus d'attribution de responsabilité dépendent du type de norme violée (contra article 16 ci-dessus). (Chakroff et al. 2016) [41]

Article 28 – Le lien causal est surpondéré par X quand une **norme est violée** par Y, l'article recherche quel élément contribue à cette surpondération. L'article élimine l'intermédiaire de la culpabilité dont X souhaiterait charger Y, ainsi que le raisonnement contrefactuel que ferait X, « si Y n'avait pas fait A alors le résultat négatif n'aurait pas eu lieu ». L'article propose l'attribution de responsabilité (accountability) comme constituant principal du jugement causal. (Samland et Waldmann 2016) [197]

Article 3 – Le blâme est un point de départ pour un changement de comportement, il permet d'initialiser un processus d'**évolution** et ce comportement de blâme a été sélectionné pour cela. Le blâme est sans lien avec l'intention de Y de faire A, mais seulement en lien avec le fait que la communauté (XY) ayant un intérêt à éliminer A, le blâme de X est un déclencheur du processus conduisant Y (et X) à ne plus faire A. L'asymétrie de l'effet Knobe est liée au fait que l'approbation a moins d'efficacité que le blâme pour amorcer un changement de comportement. (Feldman, Wong et Baumeister 2016) [83]

Article 4 – X peut blâmer Y pour une **action impossible à éviter** car Y est perçu comme pouvant lui-même avoir contribué à créer les conditions C de cette impossibilité. L'attribution d'intentionnalité peut ainsi être complexe portant soit sur A, l'acte, soit sur C, le contexte, soit

55. L'article s'appuie sur les distinctions apportées par Jonathan Haidt qui classe les normes morales en cinq ou six catégories selon les articles, dont la violence faite à autrui et le comportement impur en regard du sacré.

les deux. (Chituc et al. 2016) [45]

Article 5 – Lors de leur développement, les enfant passent d'un stade où ils jugent seulement l'acte à un stade où ils jugent de façon plus complexe à la fois **l'acte et l'acteur**. On doit s'attendre à ce que l'ensemble des jugements moraux et de leurs déterminants dans la situation soit dépendant du niveau de développement de X. (Margoni et Surian 2017) [155]

Article 13 – Un jugement a priori de X sur l'acteur Y influence, mais faiblement, l'évaluation des actes qu'il commet. Il y a bien jugement à la fois de **l'acte et de l'acteur** mais pas déterminant. (Siegel, Crockett et Dolan 2017) [207]

Article 18 – Il est possible de calculer un modèle d'inspiration bayésienne d'influence, partant d'un a priori sur l'acteur, et évoluant par le jugement sur un acte vers à la fois une nouvelle estimation morale de l'acteur et une estimation résultante de l'acte. Il y a jugement à la fois de **l'acte et l'acteur** et l'interdépendance est partiellement déterminante. (Gerstenberg et al. 2018) [97]

Article 6 – Les résultats empiriques relatifs à l'attribution d'intentionnalité sont différents selon qu'on considère une **intentionnalité « proximale » ou une intentionnalité « distale »**. La première est plus immédiate et instrumentale, la seconde est plus sur la durée et motivationnelle. Les occidentaux et les orientaux ne donnent pas le même poids relatifs aux 2 types d'intentionnalité pour établir la responsabilité, non plus que, en occident, les hommes et les femmes. (Plaks et al. 2016) [183]

Article 7 – Il faut distinguer les caractères explicatif et descriptif du lien causal. Une **description abstraite** du scénario présenté (en décrivant Y et Z de façon générique) favorise une interprétation biologique et semble diminuer l'attribution de libre arbitre à Y. Une **description concrète** (en donnant des noms propres à Y et Z) favorise une interprétation psychologique et semble augmenter l'attribution de libre arbitre. (Kim et al. 2017) [133]

Article 8 – La **manipulation de l'acteur Y** par un tiers manipulateur peut entraîner une perception moindre de son agentivité, mais uniquement si le manipulateur a lui-même intentionnellement manipulé. (Murray et Lombrozo 2017) [169]

Article 12- **Blâmer est simple** (automatique et immédiat) mais **louer est complexe** (réflexif et long), c'est ce qui occasionne l'asymétrie de l'effet Knobe. (Clark et al. 2018) [48]

Article 14 – On peut avoir une **vue internaliste de l'intentionnalité** (dépend de l'agent Y) ou une **vue externaliste** (c'est un jugement porté par un tiers X). A cet égard, il semble que l'intentionnalité soit polysémique et complexe, ce qu'ont du mal à refléter les études empiriques du type de celles liées à l'effet Knobe. (Cova 2017) [56]

Article 17 – Y peut simplement croire qu'il doit faire A ou bien il peut le savoir. L'article

propose que si X pense que Y n'a qu'une croyance, l'obligation qu'il attribue à Y de faire A sera moins forte que si X pense que Y en a connaissance. L'obligation est attribuée à un acteur qui a **connaissance des conséquences** de ses actes, plus que s'il s'agit de simples croyances. (Turri, Friedman et Keefner 2017) [230]

Article 19 – Le cerveau se comporterait selon deux modes distincts ; **DMN pour Default Mode Network et TPN pour Task Positive Network**. Ces modes seraient incompatibles et nous basculerions de l'un à l'autre de façon analogue à la perception qui nous fait voir alternativement un lapin ou un canard dans une même image [56]. L'attribution de responsabilité serait ainsi bistable et insoluble car les déterminants qui font pencher X vers la responsabilité de Y sont différents selon que X est dans l'un ou l'autre de ces modes de pensée. (Friedman et Jack 2018) [93]

Article 20 – L'asymétrie de l'effet Knobe est empiriquement confirmée pour des participants recrutés parmi des juges de l'administration judiciaire française. C'est problématique car cela contrevient à un principe fondamental du droit, dit « **mens rea** », qui fait de la responsabilité un caractère indépendant des conséquences de l'acte. La réalité des jugements pourrait ainsi contrevenir quotidiennement à ce principe supposé. (Kneer et Bourgeois-Gironde 2017) [137]

Article 22 – L'acquisition des règles morales s'appuie sur **deux types d'apprentissages**. L'un pour des associations automatiques, qui nous permet des jugements rapides, irrépressibles, ... et l'autre pour l'acquisition de modèles réflexifs appuyant des raisonnements déductifs. L'attribution de responsabilité morale peut relever de l'un et ou l'autre processus de façon complexe dont les déterminants diffèrent en fonction du type de modèle d'apprentissage mobilisé. (Railton 2017) [188]

Article 23 – Le sens commun refuse l'existence d'un lien logique entre **causalité déterministe et inévitabilité fataliste**. Selon la façon dont est présenté un cas, il privilégiera l'une ou l'autre, y compris en se contredisant. (Dhar 2017) [75]

Article 24 – Nos jugements moraux sont entachés du **biais d'optimalité**. Il consiste à surpondérer le fait qu'un acteur agisse toujours de façon optimale en regard de ses intérêts. En conséquence, si un acteur agit de façon non optimale, sa responsabilité est considérée plus forte même s'il ignore qu'il existe de meilleures alternatives. (De Freitas et Johnson 2018) [66]

Article 25 – Il faut distinguer le comportement correspondant à la personnalité profonde,

56. Cette célèbre illusion d'optique a été reprise par les éditions Vrin pour la couverture des ouvrages de la collection « Chemins Philosophiques » `http://www.vrin.fr/collection.php?code=248`

en anglais **deep self**, de celui résultant d'une influence du contexte contingent. L'attribution de responsabilité est plus forte dans le premier cas que dans le second. (Björnsson 2016) [26]

Article 26 – X attribue l'intention de faire A à Y de façon plus systématique s'il pense que l'acteur Y est **conscient des conséquences** (awareness) de A. De plus, pour X, si Y est conscient des conséquences de A, il a l'obligation complémentaire de les prévoir et de ne pas les négliger. (Redford et Ratliff 2016) [192]

Article 27 – L'agentivité est un terme confus dont le sens dépend du contexte (causalité, intentionnalité, responsabilité,...). Un jugement moral négatif peut entraîner soit de considérer l'acteur comme moins qu'humain soit comme un humain méchant. On peut penser que dans le premier cas l'estimation de son agentivité décroît et que dans le second elle croît. L'article penche empiriquement vers la première option, **la déshumanisation**. Si X juge Mal ce que Y a fait, alors il a tendance à dénier à Y la pleine humanité, car celle-ci l'aurait conduit à se comporter autrement (Khamitov, Rotman et Piazza 2016) [132].

Article 29 – Une action peut être dite libre selon de multiples critères (intention préalable, décision consciente, délibération préalable, choix significatif, connaissance des conséquences, décision immédiate spontanée ou mûrement réfléchie, ...). Les sondages ont pour but de vérifier le poids donné à chacun de ses critères par les participants exerçant leur sens commun. Au total, décision consciente et connaissance des conséquences sont les critères principaux et **l'existence de la décision immédiate suffit à définir un acte libre**. Une conséquence de ce sens commun reflété par les sondages est que sont valides les expériences de type Libet, elles opérationnalisent bien le libre arbitre. De plus, et contrairement aux philosophes, le sens commun valide la « liberté d'indifférence » comme une source valide du libre arbitre. (Deutschländer, Pauen et Haynes 2017) [73]

Article 30 – Le sens commun donne un poids important à l'intention « distale » pour attribuer un degré de libre arbitre. L'action A de Y est reconnue comme intentionnelle par X si elle s'inscrit dans un projet affirmé par Y dans la durée [57]. De façon inattendue, l'article montre par ailleurs que l'attribution d'intentionnalité en lien avec le libre arbitre **dépend du genre de l'acteur**, les actes des femmes sont plus souvent considérés comme résultant de leur libre arbitre (effet faible mais certain d'après les auteurs). (Felletti et Paglieri 2016) [84]

Article 31 – L'effet Knobe est inversé pour certaines cultures (Samoa,Vanuatu). Cette **inversion est liée au statut** différent des acteurs. Si on remplace le chef d'entreprise occidental de l'article de 2003 de Knobe par un chef de village dans une société où n'existent pas les

57. Affirmation s'appuyant sur des articles anciens et sur les résultats de cet article, en contradiction avec les conclusions de l'article 29.

entreprises, l'effet Knobe s'inverse car de par son statut, le chef de village ne peut vouloir intentionnellement la mauvaise conséquence de son action. (Robbins, Shepard et Rochat 2017) [194]

Article 32 – Les illusions visuelles ont une influence sur l'évaluation du lien causal. Comme l'évaluation du lien causal influe sur l'évaluation morale, on peut conclure que les illusions visuelles ont une influence sur le jugement moral. On peut se demander si cette influence est directement liée aux inférences de haut niveau faites par le système visuel, ou si elle est indirecte par un processus de raisonnement passant par l'estimation du lien causal, c'est l'objet de l'étude. Conclusion, les **inférences abstraites du système visuel** influencent directement les jugements moraux. (De Freitas et Alvarez 2018) [65]

Article 33 – Premier résultat de cet article, l'effet Knobe pour des décisions d'entreprises est reproduit avec des étudiants allemands et arabes (EAU), et il semble donc que cet effet soit robuste en regard des différences de cultures entre ces étudiants. Deuxième résultat, **le statut social de l'acteur Y** (chef d'entreprise ou employé membre d'un comité de décision) a une incidence sur l'effet Knobe, l'employé est plus loué et moins blâmé, et cette incidence dépend de la culture. Les étudiants arabes blâment significativement moins l'employé que les étudiants allemands. (Kaspar 2016) [130]

Cette description trop sommaire de chacun de ces 33 articles ne rend évidemment pas compte de la complexité argumentative de chacun de ces articles, il n'en restitue que l'ossature principale. Par effet d'accumulation, le lecteur ne peut qu'être pensif face à un mouvement aussi dispersé. Il est d'ailleurs frappant de remarquer que tous ces articles ont été publiés en un peu plus d'un an, montrant la grande ébullition conceptuelle du domaine et l'inflation galopante de publications à laquelle doit faire face le spécialiste de l'intentionnalité. On peut néanmoins proposer un constat qui découle de cette ébullition même : l'intentionnalité est un concept polysémique dont le contenu change avec le positionnement des chercheurs, rendant l'analyse rationnelle de l'ensemble particulièrement difficile.

4.6.5 L'intentionnalité : un exemple de cible philosophique ambigüe

A l'issue de cet inventaire, donc, plusieurs constats s'imposent et principalement celui d'un schéma global complexe au sein duquel chaque article met l'emphase sur un des nombreux traits qui animent le débat de l'attribution d'intentionnalité et de ses conséquences pour le jugement moral. Je vais, dans la présente section, dans un premier temps insister sur cette complexité conceptuelle et montrer qu'elle induit une vision divergente et exponen-

tielle du domaine conceptuel, quand chaque nouveau trait vient démultiplier la difficulté sans qu'apparaisse de possible clarification ou convergence. Face à cette complexité, les doutes se lèvent quant à la capacité de la démarche exemplifiée par ces 33 articles, et ce sera l'objet du deuxièmes temps que d'exprimer ces doutes, qui sont d'ailleurs bien présents au sein même de ces 33 articles. Et enfin, troisième et dernier temps de cette section, je m'appuierai sur un des 33 articles pour tenter de décrire ce que ces articles peuvent dessiner quand aux rôles respectifs envisageables pour la philosophie et la psychologie expérimentale.

La complexité de l'attribution d'intentionnalité

Tentons une représentation simplifiée de la complexité révélée par la juxtaposition des 33 articles relatifs à l'effet Knobe :

— Chacun des aspects de la causalité, de l'intentionnalité et de la responsabilité morale appelle les deux autres lorsqu'on les met à l'épreuve dans le domaine du jugement moral. Ces trois notions interagissent de façon complexe et non linéaire, avec des boucles de rétroactions, ce qui met à mal les tentatives de raisonner par addition ou combinaison de l'influence de différents critères. Tout éclairage partiel n'est alors que lacunaire sans pouvoir être constructif.

— L'intentionnalité peut être comprise comme distale, et réfléchie, ou proximale et immédiate. En donnant du poids à l'une ou l'autre de ces conceptions, les projets à long terme ou les actions immédiates, chaque scénario distord la perception du rôle des agents et redéfinit l'intentionnalité.

— Le processus de jugement moral est divers, dépendant (ou non) du type de norme morale violée ou négligée en relation complexe avec la façon dont elle est violée. Un changement dans la fiction présentée, même mineur (selon certains points de vue), peut changer le jugement moral de X (et de O).

— Le jugement moral peut porter sur l'acte A et, ou sur l'acteur Y. L'interaction entre ces deux axes est à nouveau complexe, et la réponse de X peut varier selon qu'il adopte l'une ou l'autre vue.

— Les caractéristiques d'age, de genre, de statut des personnes jouent, de façon empiriquement démontrée, pour l'acteur Y, mais aussi potentiellement pour la victime Z, le juge X, le sondeur S et l'observateur O.

— La culture des participants X intervient de façon complexe dans les résultats. Il peut sembler une anomalie que la majorité des études qui s'attachent à étudier la sensibilité à la culture montrent son importance mais que celles qui ne sont pas centrées sur cet aspect la négligent presque complètement. En l'état, on peut remarquer que la

majorité des 33 études ne s'appuie que sur un petit échantillon de l'humanité : les jeunes anglophones disposant d'Internet.

— Certaines études insistent sur le fait que nos intuitions dans le domaine de l'intentionnalité sont soumises à de nombreux biais (présentation, moralité, agentivité, optimalité,...) mais les autres études continuent à s'appuyer sur ces intuitions sans mitigation du risque d'erreur engendré.

Aucune des 33 études n'embrasse l'ensemble de toutes ces distinctions complexes au sein d'un même plan d'expérience cohérent. Chaque étude se focalise sur un point, un aspect du domaine, qui apparaît comme nouveau, sans prendre en compte la diversité des cas qui seraient à analyser si tous les paramètres variaient. Cette stratégie de recherche permet de voir apparaître de nouveaux aspects des phénomènes en cause dans l'effet Knobe, à chaque étude proposée, mais au risque de négliger d'autres aspects qui se révéleraient si l'étude faisait varier un paramètre supplémentaire [58]. Prenons pour illustrer cela l'exemple du dernier article, l'article 33. Cet article montre que l'effet Knobe serait globalement indépendant de la culture de X car il est répliqué avec des étudiants allemands et arabes. L'article montre également, de plus, l'influence significative du statut de l'acteur Y. Si la décision vient d'un chef d'entreprise, l'asymétrie selon la valence morale du résultat est conforme à l'article initial de 2003 de Knobe, par contre si la décision provient d'un comité de collaborateurs, l'effet d'asymétrie de Knobe est différent. Cette distinction selon le statut en appelle alors une autre, car cette différence introduite par le statut de Y diffère selon le groupe d'étudiants X, allemands ou arabes. Cette différence culturelle n'est visible que si l'on fait varier à la fois le statut de Y et la culture de X. Elle reste invisible si on ne fait varier que l'un ou l'autre des deux critères. Les critères « culture » et « statut » sont donc combinés, à la lumière de l'article 33, de façon complexe non additive. Et cette complexité est occultée dès qu'on ne prend en compte que l'un ou l'autre critère.

La question posée est alors la suivante : Que faire de ce résultat de l'article 33 si l'on s'intéresse, comme c'est le cas d'une part importante des 32 autres articles, à mettre en lumière un paramètre qui n'est ni la culture de X ni le statut social de Y ? On peut imaginer plusieurs stratégies, chacune ayant de multiples conséquences pratiques sur la façon de mener une

58. On pourrait évoquer ici un biais de sélection des 33 articles qui pourrait être à l'œuvre. En effet, l'article de 2003 de Knobe fait partie des articles qui ont contribué à lancer le mouvement XPhi et sa critique de la démarche traditionnelle philosophique. Il serait donc normal que les articles qui s'y réfèrent s'inscrivent dans ce mouvement critique et soient plus orientés vers la prise en compte critique de nouveaux critères montrant l'insuffisance des méthodes « en fauteuil » que vers la synthèse consensuelle. Je ne peux écarter tout à fait ce biais, puisqu'il faudrait pour cela comparer mes résultats à ceux portant sur une sélection plus vaste écartant le mot clé « knobe effect » ce que je n'ai pas fait systématiquement. Toutefois, la littérature générale de psychologie morale ne m'apparaît pas refléter l'existence d'un tel biais : les psychologues multiplient les complexités sans l'aide des philosophes expérimentaux du mouvement XPhi.

telle étude.

Première stratégie, largement majoritaire au regard du corpus examiné ici, on néglige l'apport de l'article 33. On mène l'étude sans prendre en compte ni le statut de Y ni la culture de X pour la simple raison qu'on s'intéresse à autre chose. Ceci revient à penser que ces deux critères n'interfèrent ni directement ni indirectement, par effet croisé, sur le nouveau résultat recherché. Cette supposition est manifestement fausse de façon générale, puisqu'elle est déjà fausse en particulier dans le seul cadre de l'article 33 où c'est la prise en compte simultanée des deux critères qui conduit à la découverte d'un nouveau phénomène. On peut également espérer que ces critères soient rendus inopérants par l'approche statistique des études, ce qui pourrait être obtenu dans le cas de la culture de X si on avait un échantillon de participants suffisant incluant à la fois, par exemple, des étudiants allemands et arabes, ce qui est rarement le cas, mais sera encore plus difficile à obtenir pour les différences dans le statut de Y, car il faudrait alors multiplier les scénarios.

Deuxième stratégie, en pratique inexistante, on prend en compte les résultats de l'article 33 pour toute nouvelle étude empirique. Il faudrait alors faire varier ces paramètres du statut de Y et de la culture de X dans le plan d'expérience, mais s'imposerait de la même façon la nécessité de prendre en compte tous les paramètres explorés par chacun des 32 autres articles, ce qui serait hors de portée pratique d'une étude, d'autant qu'il n'y a aucune raison de s'arrêter aux 33 articles que j'ai sélectionnés par commodité pour la présente étude et à la dizaine de thèmes qui s'en dégagent [59]. Une tentative d'étude globale conduirait à une complexité telle que la mise au point de l'expérience d'un côté, l'opérationnalisation des critères issus des 33 articles [60], et l'exploitation des résultats de l'autre côté, y compris l'interprétation psychologique des résultats, constitueraient des tâches considérables, certainement très coûteuses et, en fin de compte, probablement très difficiles à fiabiliser.

Troisième stratégie, c'est celle que nous avons vu précédemment à l'œuvre avec l'IAT, multiplier les études partielles pour explorer un domaine aussi complexe, et adopter ensuite une méthode systématisant les méta-analyses appuyées sur ces nombreuses études partielles pour tenter de mettre en lumière des régularités plus générales. On pourrait parler de pointillisme méthodologique : chaque étude n'éclaire qu'une très petite partie du domaine, mais la juxtaposition de toutes ces études construit une vue d'ensemble. Cette stratégie pourra

59. Pour instruire ce point, j'ai reproduit la sélection des articles opérée ici en enlevant la limite imposée à la profondeur de l'historique et en simulant les mêmes filtres de pertinence à l'aval, le résultat serait approximativement de 300 articles pertinents introduisant de nouveaux paramètres (ou modifiant des paramètres déjà proposés) dont il faudrait prendre en compte l'exploration.

60. Notons en appui de ce point que, exemple parmi d'autres, l'article 25 souligne la grande difficulté à opérationnaliser et à interpréter tous les éléments permettant de tester expérimentalement l'argument de la manipulation dans le débat entre compatibilisme et incompatibilisme. Si ces éléments sont déjà impossibles à opérationnaliser dans le cadre d'une étude centrée sur ce débat, il sera d'autant plus impossible de croiser ces critères avec d'autres.

peut-être s'avérer gagnante dans l'avenir, mais aucun des articles de ma sélection ne laisse entrevoir la possible émergence de cette vue d'ensemble. Il est alors difficile d'émettre un jugement sur cette possible émergence qui soit plus qu'une simple reformulation d'un optimisme ou d'un pessimisme préconçus.

La faiblesse des expériences en regard de la complexité exponentielle

Mon point de départ pour la présentation des distinctions apportées par chacun des 33 articles était que j'acceptais leurs conclusions comme fiables en regard du dispositif expérimental. Mais les conclusions même de ces articles font apparaître l'ampleur et la multiplicité des biais qui entachent les jugements moraux et, avec eux, l'attribution d'intentionnalité. Il semble donc naturel de supposer que la fiabilité recherchée ne sera acquise que si un grand nombre d'expériences de contrôle est exigé avant toute affirmation et, en conséquence, avant toute publication. Force est de constater qu'il n'en est rien. Le nombre maximum d'expériences pour un article parmi les 33 est de 7, la moyenne n'est que de 2,6 expériences par article et on peut mettre en doute sur cette base les 12 articles qui ne présentent que 1 ou 2 expériences et n'ont pu concrètement explorer tous ces biais.

Cette conclusion pessimiste est partagée dans plusieurs des 33 articles, pour exemple, en conclusion de l'article 1, les auteurs indiquent :

> Notre résultat qui établit la sensibilité du jugement du poids causal en regard de l'information qu'on peut avoir sur l'implication consciente de l'agent, pourrait également expliquer pourquoi les jugements sur les liens causaux, la responsabilité ou le blâme ne sont pas clairement séparés dans nos expériences ni dans beaucoup d'autres. Il semble que ces trois termes sont ambigus et dépendent du contexte, ils peuvent être utilisés soit pour désigner des facteurs causalement nécessaires soit pour désigner des facteurs moraux ou légaux relatifs à l'événement. Pour cette raison, nous ne suivrons pas Malle (et collègues 2014) dans sa suggestion qu'il faut écarter le concept de responsabilité du fait de son ambiguïté car le même argument devrait alors être appliqué aux concepts de cause et de blâme.

On ne peut mieux dire que, si on simplifie le domaine en négligeant ou écartant un critère, ce sont tous les concepts utilisés qu'il conviendra d'écarter, et le domaine d'étude sera dissout dans l'ambiguïté généralisée.

Et les auteurs de l'article 14 :

> Il a souvent été dit que le terme « intentionnellement » a plusieurs significations et que « l'effet Knobe » est principalement dû aux différents contextes qui

ont pu conduire à ces significations différentes retenues par les participants. Par exemple Cova et al (2012) ont suggéré que « intentionnellement » peut avoir les trois significations suivantes :

1) Un sens positif, selon lequel quelqu'un fait quelque chose intentionnellement quand il le choisit activement, suivant son désir de le faire. 2) Un sens selon lequel quelqu'un fait quelque chose intentionnellement quand il n'est pas forcé de le faire. Intentionnellement est alors opposé à « en y étant contraint » 3) Autre sens encore, quelqu'un fait quelque chose intentionnellement lorsqu'il a la pleine possession des moyens pour le faire. Intentionnellement est alors opposé à « par hasard » ou « par accident ».

Selon Cova et al, les participants seront influencés par le contexte expérimental pour choisir le sens pertinent du terme « intentionnellement » et leurs conclusions différeront en conséquence du sens choisi.

Ou encore, article 24, un exemple de la dévaluation des études précédentes par l'irruption d'une nouvelle distinction (ici le biais d'optimalité ou « efficiency principle ») :

Ces résultats concernent le débat relatif à l'accès par introspection à nos jugements moraux. Les théories morales diffèrent sur la question de savoir si ces jugements sont le résultat de délibérations rationnelles ou d'intuitions inaccessibles. Ces désaccords peuvent en partie être dus au fait que des jugements moraux liés à des principes différents peuvent être différemment accessibles à l'introspection. Par exemple, les gens expriment régulièrement le principe qu'un mal causé par une action est plus blâmable qu'un mal fait par omission, et néanmoins ils expriment rarement le principe qu'un mal est plus blâmable s'il est la conséquence directe de l'action plutôt qu'un effet latéral, alors que les expériences montrent que la majorité des gens en jugent ainsi. Le principe d'optimalité apparaît comme un autre principe que les gens appliquent sans en prendre conscience.

Les difficultés conceptuelles soulevées par ces auteurs et, je l'espère, mises en visibilité par les 33 articles relatifs à l'effet Knobe, peuvent être analysées dans deux directions. La première serait l'épistémologie de la psychologie elle-même, dédiée à l'étude d'un phénomène d'une complexité immense, je ne l'évoquerai pas ici. La seconde, et c'est celle-ci que je vais préciser ci-dessous, concerne le rapport entre la philosophie morale et la psychologie dans la perspective de l'expérimentation tel qu'il apparaît au travers de ces 33 articles.

Quels rôles respectifs pour la philosophie et pour la psychologie expérimentale ?

Dans l'examen détaillé des articles j'ai écarté l'article 9 (Fischborn 2018) [87] qui ne rapporte pas de nouvelle expérience et, en conséquence, n'apporte pas de nouvelle distinction conceptuelle. Cependant, cet article va m'intéresser ici car l'auteur se propose d'y répondre à une question préalable : quels sont les rôles respectifs de la philosophie morale et de la science ?

L'auteur propose trois modèles pour analyser l'apport potentiel de la psychologie expérimentale à la philosophie :

— Le « Modèle minimal » consiste à vérifier que les conditions de la responsabilité supposées par les théories morales sont bien empiriquement valides. Même si ces théories ne sont pas descriptives et ont pour objet le comportement idéal que les humains devraient viser, il est intéressant de savoir si cet idéal n'est pas, au moins, empiriquement impossible pour nous humains.

— Le modèle de l'« Apport de l'intuition de sens commun » consiste à considérer qu'une théorie morale gagne à être intuitive et que l'intuition peut être mesurée par des sondages auprès du public. Ce gain n'est pas de répondre à la question morale posée, car le sens commun peut se tromper. Il serait d'une part de faire porter la charge de la preuve sur les philosophes qui soutiennent une théorie contre-intuitive et, d'autre part, de disposer d'une mesure de l'ampleur du changement à mener pour réformer la perception du sens commun.

— Le « Modèle d'amélioration » consiste à attribuer deux tâches descriptives à l'activité scientifique. La première est la description des pratiques de blâme, louange, punition, ... des causes et effets menant à ces pratiques. La seconde est de décrire ce qui pourrait faire évoluer ces pratiques. Dans ce modèle, la philosophie fait un premier travail de définition conceptuelle, puis soumet ces définitions à l'apport empirique, et ensuite, en possession de la description du sens commun, le philosophe juge s'il faut le réformer et se réfère alors à la science pour trouver la bonne méthode pour mener cette réforme.

Indépendamment de sa validité, le Modèle d'amélioration, défendu par Fischborn et que j'examine ci-dessous, accorde à l'expérimentation, et en particulier aux réponses d'un échantillon de participants aux questionnaires utilisés par les psychologues moraux, des intérêts potentiels qui recoupent pour partie ce que j'ai présenté plus haut pour l'expérimentation en général (voir 3.4.1, page 144) :

— Le questionnaire dit le vrai selon le sens commun.

— Le défenseur d'une théorie psychologique contre-intuitive doit justifier à la fois sa validité et pourquoi le sens commun la rejette.

— Le questionnaire peut contribuer à décrire ce qui est impossible ou possible, et décrire ainsi les limites de ce qui peut être demandé aux humains.
— Les études empiriques indiquent de nouveaux concepts potentiellement en jeu.
— Les études peuvent évaluer l'efficacité d'une méthode de réforme, et mesurer la distance à franchir pour réformer le comportement humain.
— Après l'action réformatrice, les études empiriques permettent de mesurer l'avancement de la réforme morale.

L'auteur défend ce troisième modèle, qu'il appelle « Modèle d'amélioration ». Ce modèle élargit le rôle de la science par rapport au modèle de « l'intuition de sens commun » tout en réservant les aspects normatifs de la réforme du jugement moral à la philosophie. On peut d'ailleurs remarquer que les trois modèles proposés restent conformes à une conception où la philosophie morale précède la science et définit son ordre du jour. Le scientifique a ainsi un rôle subalterne en répondant à des questions posées par le philosophe moral.

A contrario, l'analyse des 33 articles ne suggère pas la validité de ce modèle et on peut avancer trois arguments principaux en ce sens. Premier argument, il n'apparaît pas que les différents philosophes moraux soient à même de se mettre d'accord sur les 2 étapes que le « Modèle d'amélioration » leur confie, définir les concepts moraux et donner la direction dans laquelle il faut les réformer. On pourrait même soutenir que, inversement, ce sont les désaccords entre philosophes moraux qui génèrent des débats appuyés sur la redéfinition permanente des concepts philosophiques et psychologiques. Deuxième argument, de nombreux psychologues choisissent leur sujet en toute autonomie sans se référer aux débats des philosophes. De nombreux articles, 14 sur les 33 de l'échantillon, ne font d'ailleurs pas référence à la philosophie morale pour néanmoins traiter d'intentionnalité. Enfin, on peut constater, c'est le troisième argument, que dans plusieurs cas relatifs au libre arbitre (articles 3, 8, 12, 25, 29, 30, 33), c'est bien le scientifique qui est à l'origine de la découverte de phénomènes que le philosophe aura à reprendre et tenter d'intégrer a posteriori dans ses théories morales. L'apport conceptuel se fait donc, pour une partie importante, dans le sens psychologie expérimentale vers philosophie morale et non l'inverse.

Le modèle qui se dégage des 33 articles est donc plutôt celui d'un double mouvement. D'un côté, la philosophie morale a structuré pendant des siècles des débats, comme celui du libre arbitre, qui imprègnent tous les aspects de la société, et les scientifiques n'en sont pas immunes. Dans ce contexte, chargé des options morales retenues ou en débat, les psychologues choisissent leurs thèmes de travail dans une relative autonomie [61]. Ils fixent eux-mêmes leur

61. L'autonomie n'est que relative dans la mesure où le biais de publication les pousse vers certains thèmes, comme

propre ordre du jour. A charge ensuite pour la philosophie, et c'est le second mouvement, d'en exploiter les résultats et d'en débusquer les préconceptions philosophiques. Ce modèle semble ainsi plus proche de la pratique des scientifiques apparente dans les 33 articles sélectionnés que les trois modèles proposés par Fischborn.

En prolongement de ce modèle à deux mouvements, il conviendrait, mais ce serait une autre thèse, d'approfondir la question suivante. En supposant établie, à l'image des 33 articles, l'instabilité conceptuelle du domaine d'étude des questions philosophiques à fort enjeu de société, qui inclut par exemple les questions du déterminisme, du libre arbitre, de la responsabilité, etc. quelle part de cette instabilité conceptuelle est liée à l'enjeu des questions posées qui conduirait les philosophes à poursuivre une recherche d'arguments dans le but principal de maintenir la possibilité de la victoire rhétorique de théories importantes pour chaque auteur, en tant que porteur d'une idéologie, et quelle part de cette instabilité s'inscrit dans le processus de clarification long et difficile, mais normal pour toute science, de concepts très complexes que l'auteur explore, en tant que scientifique. Formulé autrement, quelle part relève de l'histoire des idéologies et quelle part relève de la démarche scientifique.

Dans les deux cas, quel que soit l'objectif que l'on se donne en allant sur le terrain ou au laboratoire, la possibilité de l'apport empirique passe par la réalisation d'expérimentations insérant dans la pratique expérimentale des comportements difficilement observables liés à ce domaine conceptuellement complexe de l'intentionnalité humaine. L'opérationnalisation à la base de ces expérimentations devra être robuste pour résister aux critiques épistémiques, elle devra relever le défi de la complexité dévoilée par les 33 articles, et elle devra faire face aux critiques des philosophes moraux qui restent centrés sur des théories morales, sur ce qui devrait être et, pensent-ils, ne peut être approché en observant ce qui est.

4.6.6 L'instabilité conceptuelle, point d'étape

L'examen de 33 articles, échantillon exhaustif sur 18 mois des articles mentionnant l'effet Knobe et l'étude expérimentale de l'intentionnalité fait donc apparaître un paysage complexe marqué par plusieurs points saillants que je vais maintenant tenter de résumer en quatre points et une question.

— Premier point : il y a une continuité de préoccupations entre la philosophie morale et la psychologie expérimentale. Une majorité d'articles provient de revues de psychologie et, pourtant, fait explicitement référence à des problèmes spécifiquement philosophiques.

nous l'avons vu avec le cas des études sur les musulmans.

— Deuxième point : une immense majorité d'articles apporte de nouveaux paramètres, démultipliant la complexité du domaine, sans qu'apparaisse un schéma d'ensemble lisible et accepté par tous.

— Troisième point : les notions de « intentionnalité », « causalité », « responsabilité », « culpabilité », « déterminisme », « libre arbitre », … constituent un réseau complexe de concepts qui héritent de multiples interprétations possibles de leur origine philosophique. L'approche expérimentale enrichit ce paysage complexe de la mesure de l'incidence de nombreux paramètres, sans qu'elle permette de trancher entre interprétations contradictoires, ni d'apporter son armature conceptuelle propre issue des résultats expérimentaux.

— Quatrième point : des 33 articles, un seul s'appuie sur une mesure autre que les réponses à un questionnaire. On peut douter que cette méthode ultra-majoritaire, qui importe dans chaque questionnaire les différences d'appréciation de X et de O sur Y, ne soit trop sensible aux préconceptions philosophiques embarquées dans les images et dans les mots des questions et des réponses pour qu'apparaisse une vue expérimentalement partagée de ce qu'est « l'intentionnalité ». C'est-à-dire un ensemble de pratiques expérimentales qui permette de dire quand il y a, ou non, « intentionnalité » de la part de Y. Le prochain chapitre sera entièrement consacré à ce problème de l'opérationnalisation des notions psychologiques.

Et enfin, de cette étude de cas, et dans la perspective de ces quatre points, une question se pose : comment concilier, d'une pat, le fait que la psychologie ne peut ignorer des questions qui alimentent les débats philosophiques depuis des siècles, et, d'autre part, le fait qu'importer ainsi des questions, importe également des concepts et des préconceptions qui, à l'image des conceptions aristotéliciennes venant freiner le développement de la physique moderne, viennent complexifier l'expérimentation en mobilisant contre ses résultats tous les arguments mis au point pendant des siècles par les opposants aux conceptions philosophiques que ces résultats semblent étayer. Il semble que la sur-interprétation philosophique des résultats et la sous-opérationnalisation des notions sur lesquelles portent les expériences s'allient pour construire l'instabilité des notions que l'effet Knobe m'a permis d'illustrer, et qui fait l'objet de la prochaine section.

4.7 Sur-interprétation et sous-opérationnalisation

Des cinq études de cas présentées à la section précédente, il me semble pouvoir soutenir que l'expérimentation, qu'elle soit menée par des philosophes expérimentaux ou par des psychologues, a apporté un éclairage complémentaire que le philosophe en fauteuil reconnaît comme pertinent [62]. On doit néanmoins immédiatement ajouter que, dans aucun des cinq cas, l'expérimentation n'a été déterminante au point de clore un débat de philosophie morale, ni même de construire un ensemble structuré de concepts psychologiques qui puisse servir de base acceptée par tous à ces débats.

La présente section a pour objet de marquer ce constat et d'en souligner les conséquences importantes, en s'appuyant sur les cinq études de cas et en les analysant selon les quatre perspectives que les philosophes moraux adoptent sur leur domaine, perspectives proposées au premier chapitre et que je rappelle : les perspectives descriptives, substantielles, méta-éthique et éthique appliquée. Ces conséquences, particulièrement importantes pour l'économie globale de mon travail, se déclinent selon plusieurs dimensions qui relèvent de l'articulation entre philosophie morale et philosophie des sciences :

— Le faible niveau descriptif des théories morales et des expérimentations de psychologie morale se font écho, en cohérence avec la description de la démarche scientifique expérimentale portée par la métaphore de l'hélice expérimentale.

— Pour le philosophe moral défenseur d'une théorie morale particulière, les difficultés et les risques à exploiter un résultat expérimental, s'il est compris comme faillible et révisable en cohérence avec la métaphore de l'hélice, sont trop importants pour ne pas conduire à une minoration systématique de l'importance de ce résultat, dans cette perspective prescriptive.

— La démarche scientifique expérimentale ne s'intéresse aux cas particuliers que dans des cas contraints (le Big Bang serait un candidat possible pour cela), elle est, normalement, mieux adaptée à l'étude des régularités et, dans de nombreux domaines, ces régularités sont statistiques. Il n'y a alors pas de chemin rationnel entre ces conclusions statistiques et un cas particulier donné. En application de ce constat, l'éthique appliquée aurait peu à attendre de la philosophie expérimentale.

— Enfin, je pense utile de dégager le constat général, et en tant que tel discutable, que les cinq études de cas mettent en exergue : il y a à la fois une sur-interprétation des

62. L'ouvrage de Williamson (Williamson 2007) [235], qui n'est pas particulièrement favorable au mouvement XPhi, est explicite sur ce point : les expériences des philosophes expérimentaux apportent des éléments d'information intéressants sans toutefois pouvoir se substituer à l'analyse philosophique. Il rejoint en cela le bilan proposé par Mukerji cité plus haut, qui, lui, est favorable à ce mouvement [167]

résultats expérimentaux (sous l'angle philosophique) et une sous-opérationnalisation du domaine (il n'y a pas d'acceptation par les pairs de la pertinence des protocoles expérimentaux en regard des entités psychologiques). La concomitance de ces deux traits éclaire nombre des difficultés de la philosophie expérimentale, et la méta-éthique peut s'en trouver fécondée en se positionnant sur les deux fronts ainsi ouverts.

Mais, avant de revenir sur une synthèse des conséquences de ces constats, il me faut les analyser plus avant quant à l'apport potentiel des expériences de psychologie à la philosophie morale et les limites de cet apport. Ayant mis en abyme l'approche expérimentale, ayant expérimenté sur l'expérimentation, il me faut tirer les traits que je pense pouvoir être la base d'interprétations inductives utiles quant à la méthode scientifique à adopter pour que les expériences menées par les psychologues soient, si possible, reconnues utiles à la philosophie morale. Et, pour cela, je reprends la grille d'analyse des quatre perspectives proposées plus haut car je souhaite montrer comment les traits à retenir dépendent de la perspective adoptée.

4.7.1 La perspective descriptive

J'ai caractérisé le domaine moral au premier chapitre comme l'ensemble des situations humaines ayant donné lieu à une proposition évaluatrice. Les cinq cas proposés relèvent bien, en ce sens, du domaine moral. J'ai également proposé de retenir la formulation suivante pour servir de base à l'étude du phénomène moral :

« O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien. »

Les deux questions posées sont, dans la perspective descriptive, les suivantes : en quoi les expériences menées par les philosophes expérimentaux et décrites dans les cinq cas, ainsi que les réponses obtenues des participants sont-elles descriptives ? Et, si elles le sont, de quoi sont-elles descriptives ?

Tout d'abord, précisons ce qui se joue dans de telles expériences en reprenant chaque élément de la formulation ci-dessus.

— L'observateur O est un philosophe ou un psychologue.

— X est le participant au sondage ou, plus largement, à l'expérience.

— Dans la plupart des cas S n'est pas distinct de O (le participant donne sa réponse à l'observateur). [63]

63. On pourrait imaginer que X doive justifier son choix à un tiers S, et évaluer l'influence de la relation entre X et S sur ce que dit X mais, à notre connaissance, ce scénario est peu exploré.

— Il n'y a, généralement, pas d'acte A ni de Y ni de Z ni de C réels. Ce sont des fictions qui sont présentées à X avec différents contenus et sous différentes formes choisies par O. Ces fictions comportent des personnages, une situation, un contexte explicite et implicite général qui sont censés induire chez X des réactions de même type que dans le cadre d'un phénomène réel.
— Selon le scénario, soit X juge un acte décrit dans la fiction, soit X dit ce qu'il ferait s'il était à la place de tel ou tel personnage de la fiction.
— Les résultats de l'expérience, les traces qui en seront conservées et exploitées, sont principalement les réponses de X et, dans quelques rares cas, des résultats d'observations intermédiaires qualifiant le processus suivi par X pour construire son jugement moral (le cas de la transmission des émotions par les larmes étant particulièrement riche en ce sens).

Cette reformulation permet de mettre en lumière plusieurs éléments importants pour mon objet, dans cette perspective descriptive. Principalement, l'apport descriptif de ces expérience au phénomène moral est limité car il ne porte pas sur des événements réels, ni sur des situations réelles, ni sur des relations réelles entre acteurs réels. Cinq éléments sont néanmoins bien réels : O, X, l'énoncé de la fiction, la réponse de X et le contexte expérimental dans lequel X a été placé. Des mesures, elles aussi réelles, peuvent être réalisées sur X pendant qu'il élabore sa réponse, par exemple chronométrage de ses réponses ou IRMf, donnant accès à des informations complémentaires à la réponse explicite apportée par X.

Un premier point à évaluer est la qualité de l'expérimentation en regard de la confrontation de ces cinq éléments aux stipulations théoriques correspondantes. Chacun soulève son lot de questions mettant en cause la validité externe de l'expérimentation. A titre d'exemple, sans aucune autre prétention qu'introductive, un point régulièrement soulevé est celui de la sélection des participants, souvent des internautes ou des étudiants en psychologie, qui fait que X pour être réel n'est pas pour autant intéressant car il est peu représentatif de la population en général.

En supposant cette première évaluation de la validité externe faite et indépendamment de sa validation par les tiers, mais en tant que traces d'une confrontation expérimentale relativement à ces cinq éléments, les milliers d'articles utilisant le paradigme du tramway, le paradigme de l'effet Knobe ou l'IAT nous offrent la possibilité d'accéder à plusieurs descriptions intéressantes :

— Au travers des énoncés fictionnels proposés, on dispose d'une description de ce que l'ensemble des O, philosophes et psychologues, pensent être des préoccupations de phi-

losophie morale.

— Au travers des multiples variantes des scénarios, nous disposons en particulier d'un inventaire des distinctions conceptuelles que l'ensemble des O a souhaité évaluer en tant que pouvant influer potentiellement sur le jugement moral.

— Au travers des réponses des X, et moyennement le traitement statistique adéquat, on peut évaluer les influences relatives sur ces réponses des caractéristiques de la fiction (telles quelles sont perçues ou comprises par X), des caractéristiques de X, et des éléments liés au contexte expérimental, dont O (tel, encore une fois, qu'il est perçu par X).

— De nombreuses variantes se concluent par des pourcentages significatifs de cas difficiles à expliquer par les théories (Par exemple, pour le tramway, pourquoi 80 % des X poussent l'aiguillage et 20 % non ?), cette description des anomalies résiduelles est potentiellement un aliment important pour les théoriciens.

Il est important de souligner ici en quoi les expériences ne sont pas descriptives du phénomène moral dans son ensemble. Aux deux extrémités de l'expérimentation, la fiction présentée d'un côté, et l'action ou le jugement de X de l'autre, l'expérience n'est pas constitutive d'une confrontation à une situation réelle. Le tramway n'existe pas et l'aiguillage non plus, il n'y a pas 5 vies à sauver et une personne qui va mourir. Le chef d'entreprise n'existe pas, non plus que ses projets et leur impact sur l'environnement. La valeur descriptive du résultat de l'expérience dépend donc essentiellement de l'acceptation d'un postulat très lourd : le comportement de X serait identique face à une fiction et face à une situation réelle, identique face à l'action et face à une case à cocher. Ce postulat n'est à l'évidence pas soutenable si on le souhaite absolument vrai. Le décalage entre ce que les humains disent et ce qu'ils font est trop important et permanent pour qu'une correspondance absolue soit plausible entre le jugement moral rapporté dans le cadre d'une expérience et le jugement moral traduit en acte. Il n'est toutefois pas non plus soutenable que ce postulat soit absolument faux et que, a contrario, le comportement de X au cours de l'expérience soit sans lien aucun avec ce qu'il serait dans une situation réelle. Si X a une réaction émotionnelle de rejet face à l'idée de pousser le gros homme, il est envisageable qu'il ait une réaction émotionnelle approchante dans une situation réelle dans des conditions suffisamment similaires. Le philosophe expérimental se trouve ainsi, lui, face à un ensemble de problèmes difficiles : en quoi le comportement expérimental de X est-il descriptif de son comportement en situation réelle ? En quoi est-il proche mais distordu ? En quoi est-il pur artefact ? En quoi une situation réelle est-elle « proche » de la situation décrite dans une fiction ?

Ce constat a pu alimenter certaines des critiques faites de façon générale aux expériences de philosophie expérimentale appuyées sur des scénarios aussi extrêmes que celui du tramway. Les scénarios présentés seraient beaucoup trop éloignés de la réalité quotidienne qui fait le contexte moral habituel des acteurs. A l'évidence, il est impossible de mener l'expérience avec un vrai tramway, de vraies victimes, un vrai décideur, ni même d'imaginer des situations réelles pouvant s'approcher de ces situations extrêmes et peu vraisemblables. Il faut souligner que cette critique ne concerne pas seulement la philosophie morale expérimentale, mais s'étend à toute la philosophie morale qui utilise des expériences de pensée extrêmes, les dilemmes sacrificiels, faites pour donner à voir les différences entre théories morales, sans poids donné à la crédibilité de la situation. [64] L'utilisation systématisée de la mesure des réponses de X peut néanmoins être vue, inversement, comme une façon de montrer que ces situations extrêmes sont signifiantes, dans la mesure où elles donnent lieu à des régularités statistiques objectives. Si les cas extrêmes étaient très loin des préoccupations humaines habituelles, ces régularités seraient difficiles à justifier pour un naturaliste, puisque ne correspondant ni à un acquis empirique ni, et encore moins, à un trait résultant d'une pression évolutive.

Deuxième critique, générique, cette méthode expérimentale s'appuie sur la participation à des expériences en laboratoire et a de ce fait de nombreuses limites. Comme souligné plus haut, le fait qu'il ne s'agisse que de déclarations et non de comportements réels est essentiel. Le passage à l'acte est bien différent d'une déclaration d'intention. Or, sauf à être accusé d'hypocrisie, le philosophe moral ne peut soutenir qu'il considère que le domaine moral est principalement constitué par ces déclarations et non par les actions. Que ne compte moralement que ce qui est dit à l'occasion d'un sondage et non ce qui est fait en dehors du laboratoire. [65]

Enfin, si les expériences de philosophie morale semblent bien répondre à certains critères utilisés dans les sciences de la nature, il est d'autres critères qui, au contraire, semblent inatteignables. Sauf à ne retenir qu'un sens très lâche à ce terme, il n'y a pas ici de mesure à proprement parler, ce qui est obtenu est simplement un recueil de réponses assez informel qui ne construit pas un espace mathématisé des mesures possibles (Stevens 1946) [218]. Ce constat n'est bien sûr pas propre à l'expérimentation en philosophie morale mais est ici frappant : comment simplement décrire ce que pourrait être la grandeur mesurée dans le cadre d'une expérience du « tramway fou » ou d'une expérience sur l'effet Knobe ?

64. La critique s'étend de plus à toute la philosophie analytique qui utilise la méthode des cas en l'appliquant à des circonstances sans lien aux réalités quotidiennes, on peut penser par exemple dans le domaine de la philosophie de la connaissance aux nombreux cas de type Gettier, hautement irréalistes, mais pas impossibles.

65. Voir la critique sur la validité externe des expériences dans (Bauman2014) [18]

Les difficultés d'interprétation des expériences sous l'angle descriptif sont donc nombreuses. Mais, en prolongement des réflexions sur la démarche scientifique expérimentale, constituée d'itérations épistémiques sur l'ensemble des traces, des théories et des expérimentations, il convient d'insister sur le fait que ces difficultés sont la marque d'un état global du domaine concerné à un certain moment de l'histoire de son développement, et ne résultent pas d'une faiblesse des seules expérimentations. Les apports descriptifs de l'expérimentation aux théories morales sont, sous l'angle descriptif, un reflet de ce qu'ont de descriptif ces théories, et ce que montrent les cinq études de cas, c'est que ce niveau est faible, mais non qu'il ne progresse pas et, encore moins, qu'il ne peut progresser. La sur-interprétation philosophique augmente les enjeux alors que la sous-opérationnalisation facilite la critique des expériences. La concomitance des deux résulte en un débat qui semble ne pouvoir avancer sur le plan descriptif, ce qu'illustre particulièrement l'instabilité conceptuelle mise à jour par l'analyse des 33 articles sur l'effet Knobe.

4.7.2 La perspective prescriptive

Plusieurs stratégies d'utilisation des expériences de philosophie morale sont disponibles pour le philosophe qui défend une théorie morale particulière et souhaite la promouvoir : l'utilisation directe, l'utilisation indirecte, ou la dénégation. Examinons tour à tour ces trois stratégies.

Première stratégie, l'utilisation directe, consiste à mettre en avant une distinction conceptuelle moralement importante pour la théorie morale défendue (et la différenciant des théories concurrentes) et tenter de montrer empiriquement sa validité. Il conviendra pour cela de concevoir une variante de scénario expérimental déduit d'un des protocoles communs (tramway, Gettier, Knobe,...) permettant d'identifier cette distinction dans la fiction soumise à X, puis d'exploiter les réponses de X comme des éléments de validation de l'importance de cette distinction. Prenons un exemple, supposons que O soit un conséquentialiste convaincu, déstabilisé par la réponse massive des X qui ne poussent pas le « gros homme » alors que cinq vies sont à sauver. Il peut envisager d'introduire à la marge l'injonction kantienne « tu n'utiliseras jamais autrui comme un simple moyen » et construire une nouvelle théorie morale qui prenne en compte un arbitrage entre les conséquences de l'acte, qui restent pour lui prioritaires, et l'acceptation des moyens, dont certains peuvent, à titre complémentaire, être jugés inadmissibles. Il pourra ensuite vérifier si sa théorie ainsi modifiée fonctionne (au sens de donner un cadre explicatif cohérent) à la fois pour les deux cas « aiguillage » et « gros homme ». Il pourra

également l'appliquer à l'ensemble des scénarios, dans l'esprit du travail de Bruers et Braeckman cité plus haut, un peu comme on teste un algorithme sur différents jeux de données, et en déduire que sa théorie rend bien compte des réponses des participants.

Au mieux, la nouvelle théorie conséquentialiste limitée par les moyens obtient le résultat souhaité et, par exemple, 80 % des participants donnent une réponse conforme à ce que cette nouvelle théorie prévoit. Le philosophe peut alors utiliser directement ce résultat empirique pour appuyer sa théorie. Cette stratégie d'utilisation directe présente plusieurs difficultés et plusieurs risques. Première difficulté, que faire des 20 % qui ne répondent pas de façon conforme ? Est-ce un problème de compréhension, ils n'ont pas compris la fiction proposée comme les autres ? Est-ce un problème de performance, ils se sont trompés en appliquant le raisonnement moral ? Ou un problème de compétence, ils ne connaissent pas la règle morale à appliquer ? Chacune de ces possibilités ouvre une brèche pour la validité de la théorie morale proposée.

Deuxième difficulté, le philosophe moral dans la perspective substantielle dit ce qui doit être, comment les humains devraient se comporter, pas ce qu'ils font déjà actuellement. Elle est réformatrice. Si la théorie est choisie parce qu'elle correspond à ce que les X font déjà, elle gagne en qualité de description mais n'a plus beaucoup de sens réformateur. Dans la perspective substantielle, le gain est faible.

Troisième difficulté, il n'est pas évident qu'un résultat de questionnaire conforte ainsi la validité d'une théorie. Le sondage valide la conformité de la théorie au sens commun, mais celui-ci peut être trompeur. De la même façon, une théorie scientifique, physique par exemple, ne sera pas considérée comme plus pertinente si le résultat d'un sondage lui est favorable, la validité des théories ne se joue pas sur ce terrain des sondages.

Enfin la stratégie d'utilisation directe présente un fort risque car elle peut être mise à mal par une nouvelle version de l'expérience, un nouveau scénario pour lequel les participants ne répondront pas conformément à la théorie. Par exemple, un des scénarios parmi les 8 recensés par Bruers et Braeckman décrits plus haut, le scénario 3 avec une dérivation et un gros homme sur cette dérivation, donne lieu à des résultats en demi-teinte, 50 % des participants disent agir et 50 % non. Le défenseur d'une théorie morale se trouve alors face à deux possibilités toutes deux contre-productives pour lui : soit la théorie morale tranche, et elle est à moitié fausse en regard du résultat empirique, soit elle ne tranche pas, et elle perd en crédibilité en tant que théorie morale construite pour aider à choisir et définir ce qui est Bien dans toute circonstance. La stratégie d'utilisation directe, en tant qu'elle fait dépendre la défense d'une théorie morale substantielle de résultats empiriques toujours complexes et toujours

remis en question, présente donc un risque important dans la perspective substantielle.

Une seconde stratégie, l'utilisation indirecte, est moins ambitieuse que l'utilisation directe, elle consiste non à tenter de montrer la pertinence de sa propre théorie morale mais à tenter de montrer que les autres théories morales sont moins pertinentes. Ainsi, si une théorie morale ne rend pas compte des réponses faites par les participants alors elle est pro tanto moins intuitive et ses défenseurs auront à construire plus de preuves pour la rendre acceptable. Dans cet objectif, le philosophe substantiel mettra en avant les scénarios pour lesquels chacune des théories morales concurrentes présente des faiblesses. Cette stratégie consiste donc à reporter la charge de la preuve sur les philosophes moraux défendant des théories concurrentes. [66]

Enfin, troisième stratégie disponible pour le philosophe, dans une perspective substantielle, tout simplement nier un quelconque intérêt aux expériences du tramway (ou autres, selon le domaine d'activité du philosophe). C'est par exemple la position défendue par Guy Kahane dans (Kahane 2015) [126]. L'argument principal de l'article est le suivant. Souvent les scénarios ne permettent pas de distinguer dans quels cas les personnes répondent en appliquant des règles déontologistes et dans quels cas ils appliquent des règles conséquentialistes, ces scénarios sont alors inutiles en regard de la perspective substantielle. Mais, plus profondément, même lorsque les scénarios permettraient théoriquement de distinguer entre les deux approches, il est erroné d'en déduire que les personnes interrogées raisonnent selon ce schéma. Les expériences montrent plutôt que les participants mettent en œuvre des raisonnements plus simples, par exemple ils choisissent entre deux alternatives celle qui n'a pas d'inconvénient direct et immédiat, sans se livrer à un calcul complexe des conséquences, et le fait que ce choix puisse également être celui du conséquentialiste est purement fortuit. En conclusion, Kahane soutient que les dilemmes sacrificiels du tramway n'apportent rien à la perspective substantielle. Cette conclusion s'étend naturellement à de nombreuses expériences de philosophie morale : même si, en suivant la méthode « algorithmique » de Bruers et Braeckman présentée plus haut nous avons construit une modélisation du processus du jugement moral qui semble donner des résultats proches des jugements observés, il n'est pas pour autant certain que les personnes interrogées suivent effectivement ce processus en leur for intérieur [67]. Sans avoir fait de statistiques précises en la matière, un sentiment qu'il convien-

66. Cette stratégie de report de la charge de la preuve, « burden of proof » en anglais, est fréquente dans les articles de philosophie morale, ce qui ne laisse pas de surprendre un scientifique plus habitué à tenter, positivement, de montrer que sa théorie est valide et à l'améliorer qu'à tenter de se dégager de la charge de la preuve sur d'autres scientifiques.

67. Cet argument perd de son tranchant lorsqu'une nouvelle mesure comme l'IRMf apporte un éclairage complémentaire sur les processus neuronaux, même si cet apport sur le processus intracrânien corrélé est insuffisant à établir une certitude.

drait d'explorer est que cette troisième stratégie, nier l'intérêt des expérimentations comme celle du tramway, est la plus fréquemment suivie par les philosophes moraux souhaitant défendre une morale particulière.

En résumé, pour le philosophe cherchant à défendre une théorie morale substantielle, la stratégie d'utilisation directe des résultats des expérimentations est risquée, car cela revient à soumettre à l'enquête empirique l'adhésion à une définition du Bien, la stratégie d'utilisation indirecte est moins risquée mais son intérêt est limité aux débats entre professionnels de la philosophie morale. La stratégie de la dévaluation de l'utilité de ces expériences est une solution de repli. La sous-opérationnalisation des entités psychologiques dans les expériences est alors un argument disponible en faveur de cette dévaluation.

4.7.3 La perspective méta-éthique

Dans la perspective méta-éthique, le philosophe recherche ce qui donne sa force à l'impératif moral, ce qui permet de justifier une théorie morale, ce qui donne du sens aux énoncés moraux[68]. Il est en droit d'attendre des expériences de philosophie morale expérimentale qu'elles lui apportent des éléments dans plusieurs directions.

Avant tout, il pourra remarquer que, dans le cadre de ces expériences appuyées sur des questionnaires, tous les participants accèdent à la demande : face à un scénario fictionnel comme celui du tramway, il est acceptable que la question du bien agir soit posée. On peut donc confirmer, sans surprise, que le domaine moral s'étend naturellement pour tous les participants à ce genre de situation. L'extrapolation de cette acceptation à tous les humains apparaît peu problématique. L'existence et l'importance du domaine moral s'en trouvent confortés.

Il pourra également constater la régularité des résultats, systématiquement entre 80 % et 90 % des interrogés poussent l'aiguillage. Cette régularité ne peut que très difficilement être le fait du hasard et il est légitime de rechercher la source de cette régularité. Même si la possibilité d'extrapolation de ces résultats aux conditions de la vie réelle n'est pas établie, il n'en demeure pas moins que la régularité du jugement moral porté par les X sur les différents scénarios renseigne sur ce qu'ils considèrent comme devant être rapporté à O sur ce sujet dans un questionnaire. Encore une fois, O, X et les réponses de X à O dans le contexte expérimental sont bien réels, même si X s'exprime à partir d'une fiction. Autre information importante dans la perspective méta-éthique, la permanence des exceptions dans tous les scénarios : 20 % des interrogés ne poussent pas l'aiguillage. Cela permet de souligner que, s'il existe une règle morale, alors son application n'a rien de mécanique. Quelle qu'en soit la

68. Pour une introduction récente à la méta-éthique voir (Desmons 2019) [71]

raison, chaque participant peut faire exception, et certains font effectivement exception de façon statistiquement significative. Ces exceptions peuvent être vues comme non problématiques lorsque la théorie morale est conçue comme réformatrice, puisque ces exceptions ne font que souligner que la situation idéale n'est pas encore atteinte. Elles sont d'ailleurs importantes en tant que marqueurs de la situation de départ que l'ambition réformatrice doit prendre en compte pour établir ses stratégies de changement. Ceci, bien sûr, à condition que les défenseurs de la théorie morale condescendent à ne pas s'occuper uniquement de l'idéal moral, mais s'attachent également à décrire une trajectoire possible vers un monde idéal où la théorie morale serait respectée par tous.

Enfin, on peut imaginer que les régularités et les variations des réponses pourront apporter des indices, certes faibles, quant à la justification, non des théories morales, mais des modes de raisonnement moraux pratiques des participants. Ainsi par exemple si les résultats s'avèrent identiques quelle que soit l'orientation religieuse, athée ou agnostique [69], cela apportera un argument plutôt en défaveur de l'origine religieuse du comportement moral, sans bien sûr que les problèmes métaphysiques en soient tranchés pour autant, ni même le problème de l'origine du comportement moral qui pourrait avoir été religieux à un certain moment puis sécularisé ensuite.

En prolongement de cette perspective méta-éthique on peut reposer la question de la pertinence de ces cas extrêmes sur lesquels reposent les dilemmes sacrificiels. Si les philosophes et psychologues les ont imaginées, c'est, au moins pour partie, pour rendre sensibles les différences entre théories morales. Mais, allant plus loin, il serait envisageable que les théories morales ne se différencient que dans ces cas limites et qu'elles conduisent en pratique au même comportement dans les cas courants. Si une telle constatation devait être confirmée, alors l'intérêt des débats entre philosophes moraux défendant différentes théories morales s'en trouverait fortement relativisé pour la vie quotidienne. Il resterait ensuite à évaluer si ces débats sont plausiblement pertinents en regard des situations extrêmes de vie et de mort, face à une situation de guerre ou d'agression physique, et ce serait une toute autre étude que je n'entreprends pas ici.

Les champs d'expériences comme celui du tramway et de l'effet Knobe constituent donc un outil important pour le méta-éthicien qui peut s'en emparer pour qualifier les théories morales en préalable de sa recherche et cela, même s'il ne peut certainement pas en attendre de réponse définitive à ses interrogations millénaires. L'acceptation par tous les biologistes de la drosophile en tant que terrain d'expérimentation est, indirectement, une mesure de ce

69. Ce qui semble être le cas, voir par exemple (Dawkins 2018 p 267) [64]

que les biologistes partagent quant à leur discipline. De même, l'acceptation par les méta-éthiciens des résultats de la psychologie expérimentale est une mesure de ce que partagent les philosophes moraux et, de ce fait, la psychologie expérimentale est d'un apport important pour le méta-éthicien.

Le méta-éthicien, de par la neutralité qu'il revendique entre les différentes théories morales, peut momentanément écarter la sur-interprétation des résultats et se concentrer sur la sous-opérationnalisation actuelle ainsi que, dynamiquement, sur les expériences qui peuvent étayer de nouveaux accords quant à l'opérationnalisation des entités psychologiques. Supposons ainsi par exemple que les corrélats neuronaux de la TOM (voir l'article 21 dans l'annexe B) soient acceptés comme opérationnalisation pertinente pour la lecture par un participant de l'état d'esprit d'un tiers, alors cette nouvelle opérationnalisation peut venir bonifier l'ensemble des expériences de psychologie morale et, simultanément, clarifier les concepts moraux en jeu.

4.7.4 La perspective de l'éthique appliquée

Dans la perspective de l'éthique appliquée, le philosophe moral se donne pour objectif d'aider des acteurs concrets à prendre des décisions dans des situations concrètes. Là encore, utiliser directement les résultats des expériences de philosophie morale semble hors de propos pour plusieurs raisons complémentaires. D'abord, et heureusement, les décisions morale de la vie réelle ne concernent que très rarement des question dramatiques de vie et de mort, ou des décisions lourdes et irréversibles, que l'on puisse rapprocher des fictions présentées dans le dilemme du tramway ou de l'effet Knobe. Rapprocher une situation réelle de telle ou telle fiction serait d'ailleurs déjà un jugement moral en puissance emportant avec lui les conclusions préjugées. [70]

Ensuite, par construction, les scénarios visent à étudier comment le participant dit qu'il agirait dans chacune des circonstances présentées. Le rôle assigné au participant dans la fiction est souvent celui de l'acteur, celui qui pousse ou non l'aiguillage, et la question posée le concerne directement : est-il moral pour lui d'agir ou non dans cette circonstance. Le jugement moral des six autres acteurs de la fiction du tramway, les personnes sauvées et la victime, est rarement envisagé [71]. De même, dans la fiction de l'effet Knobe, le protocole n'exige pas de se

70. Le choix de la fiction qui est utilisée pour décrire une situation réelle, et qui conduit en particulier au choix du registre de vocabulaire employé, emporte déjà une partie de la conclusion. On peut penser par exemple au cas de la corrida assimilée par les uns à un combat (Francis Wolff, philosophie de la corrida 2007) et les autres à une torture (Michel Onfray, 2017, Miroir brisé de la tauromachie `https://michelonfray.com/conferences/miroir-brise-de-la-tauromachie`). Rapprocher une problématique d'une certaine fiction marque alors plus la fin que le début d'un débat.

71. Avec toutefois des exceptions comme dans cet article qui analyse les intentions de l'acteur vues, en partie, des

mettre à la place des ouvriers du chef d'entreprise, ni d'évaluer l'incidence du jugement moral de l'observateur philosophe (avec là aussi des exceptions, dont l'article de Brent Stickland montrant l'incidence sur le résultat du sondage de l'opinion induite chez les enquêteurs par un cadrage préalable (Strickland 2012) [219]). Dans les situations réelles, chaque acteur peut avoir sa propre autonomie morale et la bonne décision, si une telle bonne décision existe, ne peut se focaliser sur les seules conceptions morales d'un des acteurs, le problème moral ne saurait être résolu sans la prise en compte de toutes les personnes concernées.

Enfin, le philosophe ou psychologue dispose d'une position privilégiée dans l'exploitation des résultats du sondage, il choisit les axes d'analyse qu'il retient et la dimension statistique des résultats conduit à gommer les spécificités de chacun des participants. Dans la perspective de l'éthique appliquée, l'essentiel est d'aider une personne réelle particulière en situation réelle et non d'établir une statistique. On peut ici revenir sur le cas le plus simple, celui de l'aiguillage. A supposer que le philosophe soit à côté d'un acteur moral et doive l'aider alors qu'il entend le tramway dévaler, serait-il utile de lui annoncer qu'à sa place, dans un jeu vidéo, 80 % des personnes interrogées agissent, 20 % n'agissent pas, et que, de plus, notre compréhension des raisons des uns et des autres est extrêmement lacunaire ? On peut en douter.

Une formulation métaphorique peut aider à clarifier ce point important : il n'y a pas de science qui s'intéresse à la chute d'une feuille particulière en automne. La biologie peut expliquer pourquoi, de façon générale, les feuilles de certains arbres tombent quand les jours raccourcissent. La physique peut expliquer pourquoi dans certaines conditions de vent une feuille desséchée tombe plus rapidement qu'une autre. Mais aucune science ne s'intéresse à prévoir quel jour et à quelle heure tombera chaque feuille de chaque arbre de la forêt. De façon analogue, aucune science, biologique, psychologique ou morale, ne s'intéresse à la prévision d'un cas particulier, elle recherche les régularités qui ne peuvent qu'être statistiques, issues de caractéristiques suffisamment partagées pour étayer des théories suffisamment représentatives.

Cette observation, évidente pour la chute des feuilles, transposée dans le domaine de l'action humaine, conduit à proposer que l'action d'un individu particulier dans une situation particulière n'est pas objet de science. Si la psychologie est une science, et si elle se veut une science expérimentale, au sens de la mise en œuvre de la démarche scientifique expérimentale décrite au troisième chapitre, alors elle s'intéresse aux régularités relatives à ces actions et non à l'action particulière. La psychologie morale a pour objet l'étude des régularités ob-

victimes : (Marckyk et Marks, 2014) [154]

servées chez les humains dans des contextes comportant une dimension qualifiée de morale, ce qui suppose d'avoir donné un sens à ce terme dans le cadre de la science expérimentale.

Mais ce même qualificatif de « moral » est également quotidiennement utilisé dans le cadre de l'action particulière, ou du jugement particulier, d'un individu particulier et ce double emploi, scientifique et quotidien, prête à confusion. La confusion conceptuelle se retrouve dans la très grande amplitude de ce que recouvre l'appellation « philosophie morale » allant de la morale substantielle à la méta-éthique, et à l'éthique appliquée. La psychologie morale, et avec elle l'expérimentation, apporte un éclairage scientifique intéressant en ce qu'elle fournit des règles générales, mais n'apporte pas de réponse au cas particulier visé par l'éthique appliquée : la perspective de l'éthique appliquée n'a pas à attendre des expérimentations des philosophes moraux expérimentaux plus que des généralités non précisément déterminantes dans chaque cas particulier.

4.7.5 Ce que l'expérimentation apporte à la philosophie morale, point d'étape

Je souhaite, dans ce point d'étape, insister sur les conclusions importantes des analyses menées. Le domaine moral est vaste, et trop vaste pour que puisse faire sens la question du potentiel de l'expérimentation pour contribuer à la résolution des désaccords moraux de façon générale. En revanche, en distinguant quatre perspectives sur le domaine moral, perspectives qui peuvent être adoptées tour à tour par les penseurs moraux, la question peut être plus utilement abordée et me conduit à proposer que l'expérimentation est d'un intérêt évident dans la perspective descriptive, même si la situation actuelle montre que l'instabilité conceptuelle rend difficile l'exploitation descriptive des résultats d'expérience. Elle est d'un intérêt limité dans la perspective prescriptive du défenseur d'une théorie morale particulière, non du fait de l'expérimentation en elle-même mais du fait des attentes des philosophes moraux engagés dans la défense d'une théorie particulière. Elle est d'un intérêt faible pour l'éthique appliquée, car les résultats expérimentaux ne sont que statistiques et sans possibilité d'éclairer des cas particuliers. Et, surtout, l'expérimentation est d'un intérêt majeur dans la perspective de la méta-éthique. Reprenons ces points.

Les dilemmes du tramway ou de l'effet Knobe ont été imaginés dans l'objectif de construire des arguments ayant la capacité d'arbitrer des désaccords entre philosophes moraux. En prolongement du constat décrit plus haut, 50 ans de philosophie analytique, prolongés par 20 ans de philosophie expérimentale, utilisant ces dilemmes depuis leur création jusqu'à aujour-

d'hui, ont permis d'approfondir les concepts en jeu, de rappeler à la modestie les théories morales descriptives simplistes, et de déployer les nombreux paramètres qu'une théorie psychologique se devrait de prendre en compte pour être descriptive des intuitions morales des participants. Ces résultats sont importants mais l'écart entre les situations réelles et les situations fictionnelles laisse béante la possibilité que les résultats obtenus par ces expérimentations soient peu descriptifs du comportement des humains en dehors du laboratoire des psychologues. L'évidence de l'apport expérimental reste donc à consolider quant à la validité externe, à la représentativité des expériences menées en regard des situations réelles.

Dans la perspective substantielle, et comme les tenants de ces nouvelles approches de philosophie expérimentale le reconnaissent eux-mêmes, le bilan est à ce jour assez faible en terme de résolution de désaccords moraux entre philosophes (Mukerji, 2019) [167]. L'utilisation directe ou indirecte des résultats des expérimentations semble une voie risquée pour qui défend une théorie particulière substantielle avec une définition particulière du Bien. Elle peut même relever d'une erreur de catégorie, ce qui est ne pouvant définir ce qui doit être. Pour le philosophe ainsi engagé, nier ou minimiser la signification des expérimentations décrites ici est la voie la plus simple pour défendre une théorie morale particulière, en l'immunisant contre les résultats passés, présents et futurs des philosophes expérimentaux.

Dans la perspective de l'éthique appliquée, l'expérimentation n'apporte pas grand-chose ni sur le fond de ce que doit être la bonne décision dans une situation réelle particulière, ni sur ce que peut être une méthode satisfaisante pour arriver à un accord sur cette bonne décision. De même que l'éthicien appliqué considère qu'il ne peut décrire une situation réelle donnée de façon à pouvoir la rapprocher d'une règle morale existante, de même il ne pourra rapprocher cette situation réelle des conditions d'une expérience, par construction idéalisée.

J'ai présenté ci-dessus l'apport de l'expérimentation à la philosophie morale dans la perspective méta-éthique comme très important, même s'il n'est pas déterminant. En effet, les résultats des expériences menées confortent l'existence du domaine moral, l'existence des régularités morales et des exceptions à ces régularités. Les résultats permettent également de décrire un vaste système de dépendance des résultats expérimentaux en regard de caractéristiques psychologiques relevant soit du domaine moral, soit de propriétés non morales, par exemple, de la perception des relations causales. Ces dépendances construisent un large faisceau de contraintes qui pourra être mis à profit par le méta-éthicien pour évaluer les théories morales et leurs justifications.

Chapitre 5

L'opérationnalisation

5.1 Introduction : observer l'inobservable

Dans cette section, prenant appui sur le risque de confusion très présent dans le domaine des expériences psychologiques, je me propose d'introduire plus avant l'étape de l'opérationnalisation, étape structurante de la démarche expérimentale qui permet de diminuer ce risque de confusion. Je caractériserai cette étape par six propositions que je vais étayer en reprenant les constats portés par les cinq études de cas présentées précédemment, puis je présenterai le déroulement de ce chapitre consacré à montrer les différentes facettes de l'importance de l'étape d'opérationnalisation pour la psychologie expérimentale.

Le risque de confusion

Une plaisanterie populaire raconte qu'un savant travaillant sur une puce lui dit un jour : « Saute ! ». La puce saute. Il lui coupe ensuite les pattes et redit : « Saute ! ». La puce ne saute pas.

Alors le savant note sur son carnet : « Lorsqu'on coupe les pattes d'une puce, elle devient sourde. »

La plaisanterie cherche à nous faire rire du ridicule de cette conclusion absurde, mais, ce faisant, elle laisse un arrière goût d'inquiétude à tout auditeur qui a eu à mener et présenter les résultats d'une expérimentation. Comment s'assurer qu'il ne se retrouvera pas dans la situation de ce savant ? En effet, la plaisanterie appuie sur un point difficile sur lequel je vais maintenant m'appesantir : comment le savant peut-il, si c'est cela qui l'intéresse, constater que la puce est sourde et être raisonnablement confiant dans son constat ? Plusieurs stratégies qui nous viennent à l'esprit s'agissant d'humains sont ici inutilisables. Que la puce soit ou non sourde, elle ne répondra pas à la question « Es-tu sourde ? ». Et d'ailleurs, pour

un humain, on serait en droit de se demander ce que pourrait signifier la réponse « oui » à cette question... A un autre extrême de la sophistication technologique, il serait certainement difficile de placer une puce dans un appareil à IRMf pour vérifier comment ses neurones réagissent au bruit.

Bref, cette plaisanterie souligne qu'il faut que le savant trouve une solution crédible et, si possible, simple pour éviter le risque de confusion entre la surdité et toute autre cause du comportement observé. Le dispositif expérimental devra convaincre les pairs que l'expérience montre sans ambiguïté que la puce est sourde, alors même que cette propriété n'est pas directement observable. L'objet du présent chapitre est de préciser ce lien entre, d'une part, le souci d'éviter le risque de confusion et, d'autre part, l'étape de recherche d'un procédé pratique révélant (puis mesurant, créant, modifiant...) l'occurrence d'une propriété éventuellement non observable, étape que j'ai appelé « opérationnalisation » en présentant la métaphore de l'hélice scientifique expérimentale (voir 3.1 page 124).

Confusion, démarche expérimentale et opérationnalisation

L'opérationnalisation, avec l'objectivation et l'interprétation inductive, constitue l'un des trois temps de la démarche scientifique expérimentale telle que je l'ai présentée plus haut avec la métaphore de l'hélice. L'opérationnalisation y figure la transition entre les préoccupations théoriques et les expérimentations, l'objectivation y figure la transition depuis une expérience située et réalisée par une équipe donnée vers des traces dont tous les scientifiques du domaine peuvent s'emparer et l'interprétation inductive, la transition depuis ces traces vers des généralisations théoriques. Toutes trois, solidairement, contribuent à faire avancer les itérations épistémiques constitutives de la démarche scientifique expérimentale. Si la vraisemblance de l'opérationnalisation est remise en cause du fait des confusions possibles, alors l'interprétation de l'expérimentation en devient impossible ou, à tout le moins, facilement contestable. Il n'est pas impossible que la puce soit sourde une fois que le savant lui ait coupé les pattes, mais c'est peu crédible, et des interprétations alternatives du résultat de l'expérience sont faciles à imaginer.

Pour que chacun puisse vérifier qu'une expérience permet bien de faire ce que son auteur pense faire, celui-ci explicite les protocoles utilisés et les soumet à la critique de ses pairs. Cette phase de critique des protocoles envisagés pourra bénéficier des nombreuses recherches méthodologiques sur les erreurs à éviter. En particulier, en ce qui concerne la psychologie, une attention toute particulière sera portée aux risques de confusion [1] qui apparaissent quand

1. Les anglophones utilisent le terme de « confound » pour désigner les risques d'interprétation erronée lorsque plusieurs explications psychologiques d'un même comportement sont possibles. Voir par exemple la définition et l'analyse dans (Wunsch 2007) [238]

un même comportement peut donner lieu à plusieurs interprétations divergentes. Ce point étant particulièrement important et fréquent, je me propose de l'expliciter par un exemple significatif pour le domaine psychologique, bien qu'emprunté à un autre domaine plus simple d'approche.

Supposons qu'un appartement dispose à la fois d'un chauffage pour l'hiver et d'une climatisation pour l'été. Le fonctionnement estival normal de la régulation automatique est que, quand la température est trop élevée, la climatisation est augmentée et que, quand elle est trop basse, la climatisation est baissée. De même, en fonctionnement hivernal, la régulation fait que le chauffage est augmenté lorsque la température est trop basse et le chauffage est baissé lorsqu'elle est trop haute. Supposons maintenant qu'à la mi-saison, pour des raisons inconnues, les deux systèmes, chauffage et climatisation, restent simultanément actifs. Alors une élévation de la température extérieure peut être compensée soit par une baisse du chauffage soit par une augmentation de la climatisation et la régulation peut prendre l'une ou l'autre décision de façon non prévue, puisque les deux systèmes ne sont pas conçus pour fonctionner ensemble. Pire, toute température dans la pièce peut être obtenue pour n'importe quel niveau de chauffage, en jouant sur la climatisation, et réciproquement. Supposons maintenant qu'on appelle un spécialiste du chauffage qui ne sait pas que la climatisation fonctionne et qu'il essaie de régler le système, on peut imaginer sa surprise et les difficultés qu'il aura à interpréter les mesures qu'il fera. Ainsi, lorsqu'il augmente le chauffage, la température de la pièce n'augmente pas, et lorsqu'il baisse le chauffage, elle ne descend pas.

Cet exemple simple montre comment une vision insuffisante de la complexité d'un système peut conduire à des interprétations erronées des observations. Il montre particulièrement comment une vision simpliste de la chaîne causale est facilement mise en défaut dès lors qu'il y a une multiplicité de mécanismes contributifs et antagonistes et que, second niveau, chacun de ces mécanismes a lui-même des paramètres qui en augmentent ou en diminuent l'intensité. Cet exemple s'appuie sur un type de régulation qui n'est pas l'exception mais au contraire la règle dans le monde du vivant aux multiples réglages homéostatiques [2]. Pour en venir à la psychologie, un comportement a, généralement, des raisons d'être adopté et des raisons d'être rejeté. Des circonstances favorables à l'adoption ou défavorables au rejet jouent alors dans le même sens, de même, inversement, des circonstances défavorables à l'adoption ou favorables au rejet. Cette complexité est une caractéristique permanente des êtres vivants, dont des êtres humains, qui peut expliquer, pour partie [3], la prévalence importante du risque

2. Claude Bernard caractérisait tout organisme vivant comme dissipatif, loin de l'équilibre thermodynamique, et disposant de régulations homéostatiques de ses propriétés intérieures.

3. Une deuxième caractéristique du comportement humain qui rend le risque de confusion particulièrement im-

de confusion en psychologie.

J'appelle « opérationnalisation » la recherche [4] d'un terrain d'expérimentation, des instruments, des protocoles et des gestes pratiques permettant d'obtenir des résultats empiriques pertinents en regard d'une théorie, c'est-à-dire des pratiques qui tendent à diminuer le risque de confusion. J'ai proposé plus haut que la théorie stipule des entités, des propriétés et des relations, et que l'opérationnalisation consiste donc à rechercher les protocoles, les dispositifs pratiques qui permettent de détecter les phénomènes associables à des occurrences de ces entités, propriétés ou relations, puis, les savoir-faire pratique et théorique augmentant, de les mesurer, de les créer, de les modifier et de les supprimer à volonté au cours d'une expérience.

La généralisation de l'usage des tests de dépistage pour différentes maladies, ou dispositions à les développer, a banalisé les termes de « faux négatifs » et « faux positifs » pour les cas où le test détecte l'absence de maladie là où elle est présente, et inversement sa présence là où elle est absente. Ce problème de la fiabilité des tests de détection, s'il constitue souvent le point de départ des préoccupations liées à l'opérationnalisation, n'en constitue pas le point d'arrivée. En effet, après la maîtrise de la détection du phénomène étudié et l'élimination (ou plutôt la meilleure réduction possible) de ces faux positifs et faux négatifs, l'expérimentateur cherchera à acquérir la maîtrise opérationnelle de la création, de la modification, de la suppression, de la qualification, de la mesure, en un mot de la manipulation, des phénomènes liés à son objet d'étude. L'ambition de l'opérationnalisation est donc beaucoup plus vaste que celle qui préside à la conception d'un bon test de dépistage.

Six propositions à propos de l'opérationnalisation

Des constats précédents sur les cinq études de cas et du bilan en demi-teinte du mouvement des XPhi, je retiens plusieurs propositions qui marquent l'importance qu'il convient d'accorder à l'étape de l'opérationnalisation au sein de la démarche expérimentale quand on veut l'appliquer à la psychologie morale.

Première proposition, il existe un lien fort entre le risque de confusion, les difficultés de l'interprétation inductive et la qualité des opérationnalisations menées. Améliorer l'opérationnalisation serait alors une voie pour diminuer les risques reconnus comme fréquents en psychologie.

Deuxième proposition, l'opérationnalisation est un processus itératif, progressif, s'inscrivant dans la durée nécessaire à ce que progressent à la fois les théories et les savoir-faire

portant lorsque les humains traitent de la psychologie humaine est l'engagement de l'intérêt personnel dans le diagnostic, j'y reviendrai au chapitre suivant en traitant des raisons qui conduisent le philosophe moral à différents types de scepticismes en regard des apports empiriques.

4. Le terme d'opérationnalisation peut correspondre à la fois à l'action de rechercher et au résultat de cette recherche. Je favorise le premier sens, mais peux utiliser le second par économie conceptuelle.

pratiques, dont les équipements et instruments des expérimentateurs.

Troisième proposition, l'opérationnalisation est réussie lorsqu'elle est acceptée par les pairs. Elle est donc révisable dans le temps et dépendante de la communauté des chercheurs concernés. Conséquence de cette nécessaire acceptation par les pairs, le phénomène moral étant pour partie biologique, pour partie social et pour partie psychologique, l'opérationnalisation des entités morales sera dépendante de plusieurs disciplines, travaillant à des échelles de temps et d'objets très différentes. Il est vraisemblable que ces trois disciplines s'orienteront vers des opérationnalisation différentes qu'il ne sera pas facile de faire converger.

Quatrième proposition, ou plutôt constat, l'opérationnalisation acceptée par une communauté de chercheurs résulte souvent d'investissements coûteux et structurants en équipements, en savoir-faire et en organisation de la recherche.

La cinquième proposition est qu'obtenir cet accord des pairs est difficile pour la psychologie, et par contrecoup, pour la philosophie morale expérimentale. Les enjeux idéologiques et politiques portés par la philosophie morale étant importants, comme l'ont montré les cas de l'IAT et de la transmission des émotions par les larmes, la collaboration des pairs en confiance ne va pas de soi.

Enfin, sixième proposition, en ce qui concerne l'étude du domaine moral, l'analyse de l'opérationnalisation est particulièrement importante pour évaluer les apports de l'expérimentation dans la perspective descriptive, et, c'est ce que je soutiendrai, dans la perspective méta-éthique. Elle est moins porteuse d'enseignements dans les deux autres perspectives, prescriptive et appliquée.

Le plan du chapitre

Dans un premier temps, je constaterai à nouveau en m'appuyant sur les cinq études de cas décrites plus haut les multiples confusions potentielles qui alimentent les critiques des études de psychologie, et en particulier de psychologie morale, quand il s'agit d'opérationnaliser des entités théoriques difficiles à cerner et non directement observables comme la conscience, les croyances ou les intentions. Je propose deux axes qui peuvent subsumer une partie significative des directions de ces critiques. Le premier est le schéma causal en jeu dans l'expérience menée. Le second est l'existence même des entités recherchées. La mise en perspective de ces deux axes me permettra en particulier d'éclairer l'apparente contradiction du bilan contrasté que j'ai pu faire des résultats du mouvement XPhi (voir 2.5, page 118). Il semble en effet que les apports des expériences soient avérés pour de nombreuses questions philosophiques mais, à chaque fois, insuffisants à clore le débat philosophique.

Je poursuivrai cette section en présentant les difficultés de l'étape de l'opérationnalisation

de façon générale et, surtout, dans le domaine particulier de la psychologie humaine. J'insisterai sur la voie de réponse qu'apporte la démarche scientifique expérimentale, itérative, au problème de l'opérationnalisation, et sur les limites de cette démarche, que je soulignerai en particulier avec la question de l'échelle de temps des phénomènes sur lesquels portent théories et expériences au sein d'une discipline scientifique.

Face à ces difficultés de l'opérationnalisation en psychologie, il est tentant, et il a été tenté, de contourner l'obstacle. Dans une seconde section, je vais présenter trois stratégies envisageables pour cela. La première stratégie, celle suivie par le behaviorisme, consiste à nier qu'il y ait quelque chose à opérationnaliser : les entités théoriques postulées, les états mentaux, sont sans intérêt, car on ne peut rien en dire de fiable, et, peut-être même, est-il préférable, pour la science, de considérer qu'ils n'existent pas. Il faut, selon le behavioriste, en revenir aux comportements eux-mêmes et, en conséquence, ne s'attacher qu'aux triplets S-O-R, Stimuli - Organisme - Réponse, observables par un tiers.

La seconde stratégie de contournement, l'opérationnalisme de Bridgman, consiste à inverser le rapport entre théories et expériences en posant comme premières les expérimentations et en stipulant que les entités théoriques doivent ensuite être définies par, et seulement par, les opérations qui les constituent ainsi expérimentalement. Les sciences subissent des crises importantes quand, à l'exemple de la physique au début du 20^{e} siècle, les théories évoluent brutalement dans leurs concepts centraux. Face à ces périodes d'instabilité qui voient alors toutes les approches théoriques et expérimentales mises en doute dans leurs fondations, il peut être tentant d'inverser le processus en ne définissant pas les concepts scientifiques au sein des théories mais en les déduisant des protocoles expérimentaux qui les opérationnalisent. C'est ce qu'a tenté P. W. Bridgman dans le cadre d'une théorie, l'opérationnalisme, que je détaillerai. Cette théorie n'est plus considérée dans le domaine de la physique aujourd'hui mais influence encore profondément la réflexion psychologique (Feest 2005) [82].

La troisième stratégie que je présente, celle des scientifiques et techniciens confrontés à un problème de psychologie, consiste à nier la spécificité des entités théoriques psychologiques dans un premier temps, à s'appuyer sur des entités techniques qu'ils maîtrisent, et ce en espérant que, peu à peu, se construiront à la fois les entités théoriques psychologiques et leur opérationnalisation. Je vais pour cet exemple faire appel à mon expérience personnelle de chercheur ingénieur confronté au passage de l'ingénierie à la psychologie dans le cas particulier d'une mesure psychophysique acoustique lorsque j'ai eu à aborder la notion complexe de confort acoustique. Cet exemple est important pour mon expérience personnelle et il a certainement pour partie construit ma conviction que l'opérationnalisation est un point important,

complexe et sous-estimé lorsqu'on aborde des notions psychologiques. Mais au-delà de cette histoire personnelle, il me permettra de décrire cette étape d'opérationnalisation comme participant de l'articulation dynamique au temps long entre des équipes différentes aux savoir-faire complémentaires et évolutifs. De plus, l'acoustique est un domaine de l'ingénierie qui a connu un développement particulier de sa relation à la psychologie dès les années 1930 avec les travaux de psychophysique de Stevens, elle figure donc en bonne place dès qu'est abordé le problème de l'opérationnalisation en psychologie, comme le montre l'article (Feest 2005) [82] qui retrace l'influence de l'opérationnalisme de Bridgman sur la psychologie expérimentale.

Chacune de ces stratégies de contournement permet, à un moment donné de la recherche scientifique, d'éviter l'enlisement dans des débats stériles quand ni les entités théoriques ne sont suffisamment définies, ni les pratiques expérimentales suffisamment stabilisées, pour que l'opérationnalisation des entités théoriques puisse être acceptée par une communauté de chercheurs. Il est difficile, mais peut-être pas impossible, que la démarche scientifique expérimentale puisse naître néanmoins des travaux appuyés sur ces contournements, ou peut-être ne faut-il les voir que comme des travaux permettant de poursuivre des descriptions intéressantes, partielles, en attente des évolutions qui permettront ultérieurement à cette démarche scientifique expérimentale de se mettre en place ?

La troisième section du chapitre sera consacrée à approfondir l'opérationnalisation telle qu'elle est pratiquée par les psychologues pour trouver des indicateurs observables à des entités psychologiques qui ne le sont pas. Je soulignerai la principale difficulté rencontrée, celle de l'existence d'indicateurs incohérents pour une même entité lorsque plusieurs phénomènes psychologiques concurrents sont retenus pour détecter les occurrences d'une même entité théorique, et les dispositions préconisées dans les ouvrages de méthode de psychologie expérimentale pour dépasser cette difficulté. Ces ouvrages peuvent donner une image minimaliste de l'étape de l'opérationnalisation, ramenée à un simple mauvais moment à passer pour trouver comment rendre observable un phénomène psychologique, et ce chapitre, appuyé sur les exemples, a pour objet de montrer que cette image de l'opérationnalisation doit être revisitée en regard de la pratique des scientifiques. Enfin, la conclusion reprendra les six propositions ci dessus, et j'évoquerai plusieurs objections qu'elles pourraient soulever ainsi que les conséquences pour la philosophie morale expérimentale.

5.2 Le risque de confusion et l'opérationnalisation

5.2.1 Retour sur les 5 études de cas

En quoi l'étape de l'opérationnalisation est-elle importante pour la possibilité de l'interprétation des résultats des expériences de philosophie morale expérimentale ? Je me propose d'explorer cette question en m'appuyant sur les cinq études de cas décrites au chapitre précédent et dont, ce sera ma proposition, les interprétations ont été remises en cause, pour partie, au titre justifié d'une opérationnalisation jugée défaillante. Je reprends donc ci-après pour chacun de ces cinq cas ce que les philosophes expérimentaux cherchaient à étudier, comment les expérimentateurs ont opérationnalisé les entités théoriques, les propriétés ou les relations étudiées, les risques de confusion soulevés et, avec la confusion possible, les difficultés d'interprétation inductive des résultats de ces expériences. Cette analyse établit les préalables à la synthèse de la section suivante : l'opérationnalisation en tant que résolution d'un ensemble complexe d'équations sous la contrainte de la faisabilité pratique.

5.2.1.1 Les dilemmes du tramway

La présentation des expériences du tramway dans une conférence donne invariablement lieu à deux phénomènes [5], l'un est une réaction d'intérêt pour ce dilemme dont chacun perçoit l'importance, et, en particulier, le lien avec le sujet d'actualité de la voiture autonome et des capacités morales que l'autonomie semble impliquer devoir lui accorder [6], l'autre, auquel je vais m'intéresser maintenant, est fait de nombreuses questions dubitatives.

> « Est-ce d'une véritable expérience dont il s'agit ici ? N'est-ce pas plutôt une simple expérience de pensée comme les philosophes les pratiquent depuis l'antiquité ? Ce qui est observé est-il en lien avec le comportement réel des humains ? Et, en conséquence, est-ce vraiment quelque chose de nouveau et d'intéressant que cette approche des XPhi ? ».

Mon objectif est d'apporter ici quelques éclairages, obtenus de l'étude de l'opérationnalisation, sur des réponses possibles à ce type de questions. Tout d'abord rappelons la différence entre une expérience de pensée et une expérimentation au sens plein. Comme je l'ai précisé plus haut (3.3.1 page 137), je propose de voir une expérience de pensée comme le résultat d'une opérationnalisation incomplète qui ne s'est pas donné pour contrainte d'aboutir à une

5. Constat simplement établi sur la base de quatre présentations faites par l'auteur en 2019, dont 3 dans le milieu académique et une touchant un public varié, ce constat est sans prétention à valeur statistique.
6. Voir par exemple (Sandberg 2015) pour une analyse de ce lien [198]

expérience possible en pratique mais donne néanmoins à voir, en pensée, un cas particulier qui semble pouvoir servir d'exemple ou de contre-exemple à une proposition théorique. Voyons si cette définition peut s'appliquer aux expériences appuyées sur les dilemmes sacrificiels du tramway, et pour cela, rappelons tout d'abord la formalisation que j'ai proposée pour les énoncés moraux à examiner :

« O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien. »

Dans le cas du dilemme du tramway, les chercheurs visent à instruire des questions relevant du jugement moral du participant X face à des vignettes qui lui sont présentées. Dans le formalisme ci-dessus, cela donne :

« O observe que X dit à O que ce qu'il voit sur la vignette est Bien »

Autrement dit, les expériences du tramway ne visent pas à opérationnaliser A Y et Z, il n'y a ni tramway fou réel, ni aiguillage, ni victime, ni aucune proposition théorique que l'on vise à opérationnaliser d'une façon ou d'une autre avec ces éléments. Dans cette mesure, les dilemmes du tramway sont effectivement à comprendre comme des expériences de pensée, si on les imagine traitant d'un sujet ferroviaire.

En revanche, si on les conçoit comme des expériences de psychologie portant sur X, il y a un stimuli qui est opérationnalisé par la présentation d'une vignette et un protocole qui encadre comment X reçoit ce stimuli, réagit, et rapporte à O son jugement moral. Il y a donc, en faisant varier le stimuli (le contenu des vignettes), les caractéristiques du participant (X) et le protocole par lequel X est soumis au stimuli et rapporte à l'observateur O, construction d'un terrain expérimental au sens plein du terme. Ce n'est plus du tout d'une expérience de pensée dont il s'agit ici mais bien d'une expérience sur la pensée morale de X. Une expérience sur la pensée et non une expérience de pensée.

Les questions évoquées se déplacent maintenant, à la lumière de la problématique de l'opérationnalisation, et je peux les reformuler ainsi : pour quels types de propositions théoriques (portant sur la morale de X) opérationnaliser ainsi le stimuli, la réaction du participant et le rapport qu'il en produit à O, sera-t-il jugé une opérationnalisation satisfaisante par les pairs en regard des interprétations qui en seront faites en vue de ce type de propositions théoriques ?

J'ai rappelé plus haut que le dilemme du tramway avait été inventé dans les années 1960 pour montrer la complexité du jugement moral dans le contexte de l'avortement (voir 4.2, page 161). Face à un jugement définitif tel que « l'avortement est toujours moralement condamnable», l'apport empirique semble avéré sur le plan descriptif pour l'équivalent de ce

jugement définitif dans le contexte de l'expérience du tramway, « tuer un passant est toujours moralement condamnable », ce jugement définitif est empiriquement non valide :

— Les participants répondent à 80 % qu'ils basculent l'aiguillage et à 20 % que non, et ces pourcentages varient à 50 % 50% ou s'inversent à 20 % 80 % en faisant varier des éléments du stimuli qui n'interviennent pas dans la proposition théorique proposée.

— Une proposition théorique qui n'explique pas 20 % du cas général, correspondant au stimuli simple, et qui ne donne aucune piste pour expliquer l'inversion du phénomène avec des variantes du stimuli est à rejeter.

— Il est donc empiriquement à rejeter que le jugement moral soit déterminé par la proposition théorique « tuer un passant est toujours moralement condamnable ».

Cet enchaînement semble purement logique et sa conclusion incontournable. Mais l'interprétation qui peut en être faite dépend essentiellement de l'acceptation de l'opérationnalisation qui est à la base de l'expérimentation. Plus précisément, l'opérationnalisation qui doit être acceptée par les pairs pour que cet apport empirique soit reconnu valide par tous comporte a minima les points suivants, formulés dans le contexte de l'avortement :

1. Le stimuli, la fiction présentée, est une représentation acceptable en regard de l'avortement,
2. La population des X soumise au questionnaire, constitue un échantillon acceptable en regard des humains (au même niveau de généralité que l'on souhaite donner à la validité de la proposition théorique),
3. La façon dont X rapporte son jugement à O, est acceptable en tant que représentant une bonne approche soit de la décision de X d'avorter, soit du jugement moral de X à propos d'un avortement en particulier, soit du jugement de X sur l'avortement en général, ou sur la législation sur l'avortement.

Aucun de ces trois points n'est clairement universellement acquis. Pour le premier, il est tout à fait concevable que des théories morales fassent une distinction forte entre tuer « quelqu'un » et tuer « son enfant en devenir ». Même si des philosophes ont pris en compte ce point en présentant des fictions où Y, la personne qui est à l'aiguillage, est décrite comme un parent de la personne à sacrifier pour sauver les cinq autres, il n'en demeure pas moins que ce n'est pas de l'enfant de X dont il s'agit. Le second point, la représentativité de l'échantillon des X, est le moins controversé dans le cas du tramway, car les réplications en ont été nombreuses et concordantes, supposons-le acquis [7]. Et le troisième point, l'acceptation du lien

7. Prendre en compte les populations, comme les autistes, pour lesquels ce second point n'est pas acquis n'appor-

entre le jugement moral de X et ce qu'il en rapporte à O, est bien sûr le plus problématique tant à chaque théorie morale correspond une réponse différente quant à la validité du protocole de rapport de X à O pour opérationnaliser la question morale posée. Par exemple, si on retient une théorie morale émotiviste, à l'image de celle de Jesse Prinz présentée plus haut (voir 1.3.4, page 77), il est peu plausible que l'émotion de X face à une vignette et une case à cocher soit la même que celle d'une personne face à la situation réelle d'un avortement qu'elle envisage pour elle-même. Du jugement moral de X on ne peut donc rien déduire de ce que pourrait être celui d'une personne concernée. Ce qu'opérationnalise alors (au mieux) l'expérience du tramway est la réponse sociale d'un témoin éloigné, mais pas la décision des personnes concernées. S'en suit un problème difficile pour le philosophe moral, d'un côté le phénomène moral est supposé expliquer nos comportements et le sentiment moral être motivationnel [8], mais de l'autre côté, ce à quoi nous accédons empiriquement n'est justement pas ce qui s'inscrit vers l'action car les personnes concernées sont hors du champ expérimental (constitué, rappelons-le, uniquement de X et O).

Par ailleurs, remarquons que, même en supposant que ces trois points centraux pour l'opérationnalisation soient acceptés par la communauté des philosophes moraux, cela ne validerait que la dimension descriptive de l'expérience du tramway. Les perspectives substantielles et appliquées restent, comme je l'ai détaillé plus haut, largement hors champ. Pour la perspective substantielle, rien ne dit s'il faut ou non pousser l'aiguillage et selon quelles considérations, ou si, l'ayant poussé, il convient de juger comme un bien ou un mal de l'avoir fait. Et pour la perspective appliquée, rien ne permet de dire en quoi deux situations sont moralement équivalentes et, en particulier, en quoi ma situation ici et maintenant est proche de telle ou telle variante du tramway et donc, à supposer qu'il y ait un désaccord moral [9] face à la situation ici et maintenant, comment il est envisageable de le dépasser.

Enfin, dans la perspective méta-éthique, il ne s'agit plus d'opérationnaliser au premier degré les objets d'une proposition morale théorique, mais d'observer comment se comparent différentes théories morales, quel est le sens des propositions morales et ce qui, finalement, explique le phénomène moral en son ensemble. Ce qui est considéré comme opérationnalisé dans l'expérience du tramway, le stimuli fictionnel, X et la réponde de X à O, est alors à compléter plus largement, pour adopter la perspective « méta » en incluant avec l'observateur O, ce qui a conduit O à faire cette expérience, comment il l'a menée et comment, allant au bout

terait pas de façon importante ici, mais serait à considérer plus avant si le sujet d'étude était, précisément, ce type de population.

8. Pour les philosophes internalistes, c'est même une caractéristique sine qua non du sentiment moral que d'être motivationnel.

9. Et les données empiriques vont dans le sens de confirmer l'existence de ces désaccords puisque, dans le meilleur des cas, 20 % des personnes ne suivant pas la majorité et ne poussent pas l'aiguillage.

du processus, il a élaboré puis exploité les résultats. Les questions soulevées par l'opérationnalisation sont alors déplacées vers, par exemple, les thèmes suivants :

— Les principales propositions théoriques de la théorie morale sont elles valablement opérationnalisées pour être différenciées par une de ces situations expérimentales ?
— L'utilisation des cas extrêmes à la base des vignettes du tramway peut-elle être considérée comme constituant une bonne opérationnalisation par des philosophes moraux, alors que la théorie morale peut avoir la prétention à nous aider ici et maintenant ?
— Le mode de traitement (statistique) des réponses est-il signifiant pour le philosophe ?
— ...

On constate ici que le jugement sur l'opérationnalisation dépend de la perspective morale adoptée sur les phénomènes étudiés. Une opérationnalisation visant des questions méta-éthique ne sera pas jugée de la même façon qu'une opérationnalisation visant les mêmes objets mais dans le cadre d'une question de morale substantielle, descriptive ou appliquée. Les propositions méta-éthiques théoriques qui peuvent être reconnues comme bien opérationnalisées par les expériences du tramway seraient alors celles qui sont à la racine de la présente thèse : en quoi l'approche scientifique expérimentale apporte-t-elle des éléments intéressants aux débats de philosophie morale expérimentale ? Je reviendrai sur ce point plus loin, mais je peux par anticipation souligner ici l'intérêt de l'analyse fine de l'opérationnalisation qui permet de faire dialoguer les difficultés d'interprétation des résultats des expérimentations avec une caractéristique plus facile à décrire : comment chaque élément de la proposition théorique en jeu est supposé porté par le protocole expérimental.

5.2.1.2 Surestimation du nombre de musulmans

L'exemple de l'étude de la surestimation du nombre des musulmans illustre l'inversion du raisonnement qui caractérise l'approche journalistique en l'absence d'opérationnalisation. Dans la démarche expérimentale, telle qu'elle est proposée plus haut au travers de la métaphore de l'hélice scientifique expérimentale, le chercheur part d'une situation où, disposant d'une théorie qui a des forces et des faiblesses appuyées sur des traces (imparfaites) issues d'expériences préalables (incomplètes), il élabore une nouvelle proposition théorique et recherche une opérationnalisation susceptible d'éclairer empiriquement cette proposition. Dans le cas de l'estimation du nombre de musulmans, le point de départ est la publication d'un sondage qui ne prétend pas opérationnaliser quoi que ce soit mais constituer, selon la formule consacrée, une « photographie de l'opinion » à destination de la presse grand public britannique.

Pour la présente analyse de l'opérationnalisation, l'exemple de l'étude de la surestimation des musulmans appelle donc surtout un commentaire négatif : en l'absence de proposition théorique préalable et d'opérationnalisation décelable, il est difficile d'envisager de positionner cette enquête comme participant d'une démarche expérimentale scientifique, au sens donné par la métaphore de l'hélice. Naturellement, cette remarque ne répond pas à une question plus générale qui porterait sur l'intérêt de tels sondages pour le philosophe moral qui s'intéresse à l'opinion publique et à ses relais médiatiques. L'utilisation des sondages d'opinion et des media en regard de la philosophie morale n'est pas l'objet du présent travail et je n'irai donc pas plus avant sur ce sujet. On peut néanmoins opérer quelques commentaires complémentaires liés aux différents niveaux de confusion qui peuvent être mis en avant dans cette étude.

La confusion sémantique est avérée. Face à un questionnaire sur Internet auquel il faut rapidement répondre, les sondés répondent en confondant des termes culturellement proches (ici immigrés et musulmans), alors même que la question posée, l'estimation du nombre, exigerait de faire la distinction pour permettre une réponse fiable.

La confusion entre types d'expressions du sondé est également avéré. Certains répondent pour exprimer leurs opinions politiques, d'autres répondent au plus proche de la question posée (le nombre de musulmans). Rendre visible et mesurable cette différence d'état d'esprit du sondé est très difficile et exigerait plusieurs itérations du sondage pour capter cette information, hors les cas caricaturaux des militants d'extrême droite.

Enfin, rien n'a été fait pour vérifier que le phénomène supposé, la surestimation du nombre de musulmans, ne résulte pas d'un phénomène général, la dyscalculie statistique des humains, qui conduirait pour toute question posée sur le nombre de personnes ayant une propriété particulière, quelle qu'elle soit, à une estimation fausse, aléatoirement sur ou sous-estimée [10]. Ce constat de béance méthodologique traduit la différence profonde entre une démarche scientifique expérimentale, qui conduirait à essayer de lever cette importance confusion potentielle, et la démarche journalistique qui s'arrête quand le sujet est publiable avec une audience prévisionnelle satisfaisante, indépendamment de sa valeur épistémique [11].

10. En collaboration avec Brent Strickland et Antoine Marie, une étude est actuellement (février 2020) en cours pour vérifier ce point en comparant l'estimation du nombre de musulmans à l'estimation d'une autre population de même ampleur mais ne mobilisant pas le même contexte politique et social clivant.

11. Le risque pour de nombreuses activités scientifiques étant que la trop célèbre pression à la publication conduise certains scientifiques à adopter la démarche journalistique : arrêter son travail de recherche quand il est publiable.

5.2.1.3 Opérationnaliser la transmission des émotions par les larmes

L'exemple de la transmission des émotions par les larmes m'a permis de montrer le cas de deux équipes de psychologues, l'une spécialiste des émotions, disons l'équipe A, et l'autre des mesures psychophysiques dans le domaine de l'olfaction, l'équipe B, en désaccord quant à l'interprétation de résultats expérimentaux. Pour la première, les larmes sont liées aux émotions et sont une spécificité humaine, pour la seconde les larmes comportent des phéromones dont la perception par l'olfaction influe sur les émotions du récepteur et ce trait rapproche les humains des autres mammifères.

Après la publication des spécialistes de l'olfaction B, les spécialistes des émotions A ont tenté, sans succès, de reproduire l'expérience et, de cet échec, en ont déduit que l'effet n'était pas suffisamment établi. Les points clé du désaccord peuvent être reformulés ainsi :

— L'équipe B a monté une expérience complexe opérationnalisant les stimuli générant la tristesse par la vision d'une fiction. Ils ont capté des larmes sur le visage des participantes. Ils ont opérationnalisé la réception en mettant ces larmes sous le nez de participants. Ils ont ensuite montré à ces participants des visages de femmes, opérationnalisant ainsi un stimuli sexuel, et ils ont enfin approché la réponse sexuelle à ce stimuli par 3 méthodes (déclarative, chimique et IRMf). Ils en ont conclu qu'il existe un effet des larmes sur les émotions.

— L'équipe A a ensuite tenté de reproduire l'expérience et a échoué, elle en a déduit que l'effet n'existe pas.

— L'équipe B a critiqué la démarche de l'équipe A en soulignant la complexité de l'expérience et que, conséquemment, il n'était guère étonnant que l'équipe A qui n'est pas spécialiste de l'olfaction et, plus précisément, dont c'était la première tentative en la matière, n'ait pas réussi à la reproduire.

— L'équipe A n'a pas répondu et considère que l'expérience de B n'a pas de valeur (et n'a pas à être citée dans des articles ultérieurs).

Cet exemple me permet d'insister sur plusieurs caractéristiques de l'opérationnalisation. D'abord l'opérationnalisation ouvre des questions au moins aussi complexes que le problème initialement posé, ici la nécessité d'opérationnaliser l'effet des larmes sur les émotions via l'olfaction ouvre à la fois vers le domaines des émotions, le domaine des stimuli pouvant les provoquer, le domaine des comportements émotionnels, comme les larmes, et le domaine des mesures des éléments relatifs à chacun de ces domaines. Cette complexité part dans de multiples directions, ce sont de multiples techniques qui sont nécessaires pour opérationnaliser

toutes les entités, propriétés et relations théoriques mises en jeu.

Pour certaines parties du protocole, l'opérationnalisation passe par l'utilisation de bases de données existantes calibrées, ici la base de donnée des visages de femme dont l'attraction sexuelle a été mesurée au préalable par différentes méthodes et déjà utilisée dans d'autres études, ainsi que la base de donnée des corrélations neuronales correspondant à des attractions sexuelles. Ces bases de données constituent des ressources indispensables et coûteuses pour pouvoir mener l'étude globale de la transmission des larmes. Soulignons ensuite, autre contrainte importante, que pour certaines des techniques mobilisées, comme l'IRMf, ce sont des équipes spécialisées gérant des équipements également très coûteux qui sont directement mobilisées.

Mener à bien l'expérience nécessite donc un chef d'orchestre, rôle joué ici par l'équipe B, qui en conçoit le schéma général et qui, pour chaque élément à opérationnaliser, sait trouver et faire travailler en bonne entente les équipes rodées à l'obtention des résultats suffisamment fiables pour l'usage qui en est fait dans l'expérience visée. Tout défaut, même local, de cette fiabilité met en péril la crédibilité globale de l'expérience.

Ainsi, pour l'équipe B, c'est parce que l'équipe A n'a pas construit cette compétence de chef d'orchestre, compétence qui doit être reconnue par des pairs du même domaine de recherche après plusieurs expériences sur l'olfaction, qu'il n'est pas étonnant qu'ils n'aient pas réussi à répliquer l'expérience et valider les résultats de leurs travaux.

Et, pour l'équipe A, c'est parce qu'ils n'ont pas construit préalablement de lien de confiance avec l'équipe B qu'ils ne peuvent en intégrer les résultats dans leur recherche et préfèrent tenter une réplication dans leur propre environnement, avec leurs propres équipes, malgré la difficulté de cette expérience.

Le cas de ces deux situations en miroir, et sans préjuger de la validité de l'une ou de l'autre, montre tout l'intérêt d'une troisième voie : l'interaction entre les deux équipes. Les interactions entre équipes d'expérimentateurs et de théoriciens sont possiblement des ressources à mobiliser contre les risques liés aux difficultés de l'opérationnalisation, et en particulier face aux nombreux biais dont l'importance n'est plus à établir. Le biais de confirmation pourrait ici être invoqué tant pour l'équipe B, qui la conduirait à disqualifier trop rapidement la critique de l'équipe A, que pour cette dernière, qui ne réussirait pas à répliquer une expérience qu'elle ne souhaite pas voir confirmée tant elle met en cause son domaine de recherche. On peut, et peut-être doit-on, espérer que l'interaction en confiance, si elle aboutit, neutralisera symétriquement les deux risques de biais de confirmation.

Ce cas d'étude appelle ainsi plusieurs constats importants pour notre débat sur l'opéra-

tionnalisation. Premier constat, la complexité de l'opérationnalisation lorsqu'elle touche à des phénomènes comme le rôle de l'olfaction dans la psychologie humaine. L'opérationnalisation fait alors appel à des équipes différentes et des techniques différentes très spécialisées. Deuxième constat, l'opérationnalisation fait appel à des savoir-faire qui ont demandé des investissements importants qui ne peuvent se justifier pour la seule étude des larmes. Ces investissements, par exemple dans les bases de données de référence, ont une dimension technique liée à la complexité et au coûts engendrés, ils ont aussi une dimension sociale importante car ils sont à la base d'une organisation structurée où certains organismes se voient confier la réalisation de ce type de travaux transversaux qui prennent ainsi une valeur normative pour tous les acteurs du domaine scientifique concerné. Troisième constat, l'étude devant faire appel à différentes équipes, la confiance entre elles est une condition nécessaire à l'établissement de l'opérationnalisation acceptée par toutes les équipes. Enfin, dernier constat, l'échec de partage de l'opérationnalisation conduit les deux équipes symétriquement au refus d'accepter l'apport empirique, et, finalement, à une impasse épistémique où les doutes de chacune des équipes ne peuvent être levés. C'est le blocage de la démarche, le grippage de l'hélice expérimentale, auquel on assiste dans cet exemple.

5.2.1.4 IAT

L'IAT (Implicit Association Test) est issu d'une idée qui est, par elle-même, une idée d'opérationnalisation : le temps de réaction face à une question dont la réponse suppose de rapprocher deux concepts serait plus rapide lorsque nous faisons un lien entre ces deux concepts et plus long dans le cas contraire. Mesurer ainsi directement un temps de réaction donne accès à des attitudes implicites qui ne seraient peut-être pas explicitement formulées par une personne directement interrogée. Cette idée a été prolongée en la formulant sous une forme différentielle : si nous avons une idée positive d'un premier concept et négative d'un second, alors les temps de réaction seront plus rapide si nous avons à réagir en accord avec ce jugement positif pour le premier et négatif pour le second que l'inverse.

J'ai souligné en examinant cet exemple de l'IAT (voir 4.5, page 197) que les chercheurs étaient arrivés à une synthèse minimale commune qui comprend la reconnaissance du phénomène IAT. Le phénomène reconnu est le suivant : dans un protocole présentant successivement des images mobilisant le concept A, puis le concept B, et des images connotées positivement et négativement, la réaction qui consiste à appuyer sur un certain bouton pour (image de A ou positive) et sur un autre bouton pour (image de B ou négative) est plus rapide si le participant considère positivement A (et négativement B) et plus lent s'il doit faire l'as-

sociation inverse (A ou négatif) et (B ou positif). Cet accord signifie que l'opérationnalisation suivante est proposée :

— Le stimuli (les images projetées) opérationnalise la sollicitation d'un concept.

— Le temps de réaction opérationnalise la difficulté d'une tâche.

— Le temps de réaction opérationnalise la facilité de l'accès à un concept.

Bien que moins clairement validée, l'approche différentielle consiste à ajouter à cela que l'on peut comparer entre elles des tâches en comparant les temps de réaction et, plus ambitieux encore, soustraire des temps de réaction pour évaluer la différence de difficulté entre deux tâches.

A partir de ce phénomène IAT, a été élaboré un indice Race IAT qui a pour objet de mesurer le niveau de racisme implicite d'une personne. Les images sont celles d'humains blancs ou noirs que l'expérience demande d'associer à des visions positives ou négatives, l'indice est calculé à l'aide d'une double différence, d'abord en inversant les associations (blanc positif vs. noir négatif), puis (blanc négatif vs. noir positif) et, ensuite, en soustrayant les résultats obtenus dans ces deux configurations.

Une première conclusion, pour les promoteurs de l'IAT, s'en suit : l'indice RaceIAT opérationnalise les attitudes racistes implicites des participants. Une seconde considération est alors ajoutée : si quelqu'un prend connaissance de ses attitudes implicites, et surtout lorsqu'il ne les endosse pas explicitement, alors il corrigera de lui-même ces attitudes. S'il se dit, explicitement, non raciste, mais que le test IAT lui apprend qu'il est, implicitement, raciste, alors il fera plus attention à ce que son comportement ne soit pas raciste. Le constat obtenu avec la généralisation du test IAT sur Internet étant qu'une très large majorité d'américains est implicitement raciste, selon l'indice Race IAT, et bien au-delà de ce qu'ils acceptent de reconnaître explicitement, alors, d'après la deuxième conclusion des promoteurs de l'IAT, diffuser largement le test RaceIAT est une composante intéressante pour la lutte contre le racisme aux Etats-unis.

Les psychologues qui n'ont pas suivi ces conclusions le font, pour partie, relativement à l'opérationnalisation, en bâtissant plusieurs arguments. Premièrement, il est faux que la double différence nécessaire à l'indice IAT opérationnalise quoi que ce soit dans le cas général. Une façon simple de le démontrer est de prendre une situation où trois concepts sont comparés, A, B et C et où on mène un IAT sur 2 d'entre-eux, A et B, l'indice n'a alors aucun sens car une même valeur peut correspondre à des situations diamétralement opposées selon le positionnement relatif du concept C. Or, pour l'indice Race IAT on est précisément dans cette situation puisqu'un asiatique raciste, ou un misanthrope, haïrait aussi bien les blancs

que les noirs et aurait le même indice IAT qu'un humaniste œcuménique qui les aime tous également. On peut donc à la fois reconnaître qu'un temps de réaction est une bonne opérationnalisation de la difficulté d'une tâche et s'opposer à l'élaboration différentielle nécessaire à l'indice IAT.

On peut aussi remarquer que ce qui est opérationnalisé par le temps de réaction est une tâche globale de rapprochement de deux concepts, et qu'il n'est pas possible de qualifier en quoi les concepts sont proches ou non. On peut par exemple supposer que « noir » et « négatif » activent des réseaux de neurones proches (en regard du temps de réaction et dans le contexte du moment) parce que les informations reçues qui ont activé « noir » dans la dernière semaine ont également activé « négatif » car elles sont plus souvent mauvaises que bonnes. Un tel biais de disponibilité ne serait guère surprenant et nous conduirait au constat que cette association ne dit pas grand-chose des attitudes de chacun.

Et, enfin, l'indice IAT ne dit rien sur la prémisse complémentaire, justificatrice de la diffusion du test, proposant que communiquer une attitude implicite à son détenteur l'amène à s'amender. Il faudra là concevoir une nouvelle expérimentation où serait à opérationnaliser l'efficacité d'une telle communication. On peut penser, à un premier stade, le faire en mesurant le changement induit sur l'indice IAT avant et après communication, mais cette opérationnalisation redoublera les critiques ci-dessus relatives à la fiabilité de l'indice. Dans un second stade, on peut envisager de soumettre des populations entières à des campagnes de tests IAT généralisés et évaluer ensuite directement l'évolution du nombre des actes racistes constatés. On quitterait là le laboratoire du psychologue pour rejoindre les approches statistiques des sociologues, et le problème de l'opérationnalisation ne se pose plus dans les mêmes termes, puisque c'est l'action en grandeur nature qui est ainsi envisagée.

Remarquons que si on accepte les mesures implicites pour pertinentes, ce qui semble le cas avec l'accord sur la synthèse minimale relative à l'IAT, alors il semble qu'il faille accepter qu'il existe une différence significative entre les mesures explicites et implicites des attitudes morales. On est alors amené à regarder avec circonspection les méthodes qui s'appuient uniquement sur des déclarations explicites faites à l'expérimentateur. Ceci rajoute donc un argument en regard de l'opérationnalisation faite dans le cadre précédent du tramway : il n'est pas évident que ce que X dit à O révèle effectivement ni ce que X ferait ni même le jugement moral réel de X, et ce, ni bien sûr face à un événement réel, ni même face à la vignette. Le problème posé serait alors de trouver une opérationnalisation de l'attitude de X qui ne passe pas par une déclaration explicite. Une telle méthode implicite permettrait une nouvelle exploitation de la réaction de X au stimuli fictionnel de la présentation d'une vignette dont il conviendra

d'analyser la différence avec l'expression explicite que X en fait à O dans la version classique actuellement généralisée.

Ce que je viens de faire, remettre en cause des conclusions précédentes à la lumière d'une nouvelle expérimentation, est précisément ce que nous avons systématiquement rencontré dans l'étude de l'effet Knobe au travers de 33 articles, je reviens maintenant sur cette dernier étude de cas.

5.2.1.5 L'effet Knobe

Reprenons maintenant l'analyse de l'effet Knobe, l'attribution d'intentionnalité qui dépend, pour un effet de bord non explicitement recherché, du jugement moral porté sur cet effet de bord. De nombreux chercheurs ont refusé de tirer des conclusions philosophiques de cet effet Knobe et ont mis en avant différentes difficultés d'interprétation de l'expérience [12]. Dès les publications initiales de 2003 [138], on a pu par exemple considérer que, lorsqu'il s'agit de féliciter le chef d'entreprise pour une amélioration qu'il n'a pas souhaitée, les participants répondent bien à la question et ne lui attribuent pas d'intentionnalité. En revanche, lorsqu'il s'agit de le blâmer pour une dégradation, alors les participants, parce qu'ils souhaitent le blâmer, répondent en affirmant le caractère intentionnel de l'action car ils pensent que c'est la réponse qui conduit le plus sûrement à la condamnation du chef d'entreprise. En somme, dans le cas positif, les participants répondent à la question posée, celle de l'attribution d'intentionnalité, mais dans le cas négatif, ils répondent à une autre question, celle de la culpabilité du chef d'entreprise.

L'analyse du chapitre précédent, appuyée sur 33 articles récents, montre la permanence de cette dynamique de remise en cause des interprétations précédentes sur la base d'une complexification du modèle interprétatif. Cette complexification est principalement obtenue soit par la prise en compte d'un nouveau paramètre qui influe sur le phénomène (par exemple, le type de norme violée par l'acte A ou le statut social de X) soit par la constitution d'un schéma causal alternatif dans le réseau constitué par les objets mentaux supposés décrire le cheminement mental de X.

Si ma thèse est correcte, ces difficultés d'interprétation trouvent, au moins partiellement, leur contrepartie dans des faiblesses de l'opérationnalisation dans le cadre des expériences liées à l'effet Knobe. C'est ce que je vais m'attacher à défendre maintenant. Je vais d'abord essayer de reconstituer ce que pourrait être une proposition théorique qu'il s'agirait de sou-

12. La discussion, arguments et contre arguments, est reprise par Joshua Knobe lui-même (Knobe 2006) [139], elle est également utilisée dans (Cullen 2010) [59] pour montrer les difficultés de la méthode d'enquête par questionnaires.

mettre à l'enquête empirique, puis observer qu'il n'est pas possible de lister les entités théoriques (entités, propriétés et relations) correspondant à l'effet Knobe et à sa multitude de variantes, et enfin, ne conservant que quelques entités principales, souligner les faiblesses de l'opérationnalisation.

Reprenons le cheminement de Joshua Knobe se proposant d'étudier expérimentalement l'attribution d'intentionnalité en situation de jugement moral. Je reprends également ma notation habituelle pour la proposition morale :

« O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien. »

Dans le cas de l'effet Knobe, X lit une vignette dans laquelle Y dit vouloir faire un projet pour des raisons de rentabilité, ce projet a également un effet de bord A, mais Y est indifférent à cet A. Le rapport de X à O porte initialement non sur le jugement de A mais sur l'appréciation de l'intentionnalité de Y faisant A.

Joshua Knobe n'a pas de proposition théorique définie avant de réaliser l'expérience de 2003. Simplement il étudie le phénomène complexe de l'attribution d'intentionnalité et son lien au phénomène moral. Après l'expérience, il induit une première proposition : « X attribue l'intention de faire A à Y plus souvent si A est moralement jugé négativement et moins souvent si A est jugé positivement. »

Dès ce premier niveau d'interprétation, l'examen des 33 articles montre que l'induction, si on la pense universelle, serait abusive. Dans certaines conditions (par exemple dans les îles Samoa, article 31) cette proposition n'est pas clairement empiriquement valide. Il faudrait en limiter la portée selon le statut de Y, et il n'est pas difficile de penser à d'autres distinctions de même type qui qualifieraient X, Y, A ou le projet souhaité par Y, de façon à invalider l'induction. L'ensemble des éléments de contexte qu'il faut préciser pour que la proposition de Knobe soit empiriquement délimitée est important et, de plus, indéterminé. La vignette mobilise en effet un ensemble de croyances d'arrière-plan sur ce que sont un chef d'entreprise, un adjoint, l'environnement, la rentabilité, la prise de décision, etc. Tous ces éléments du contexte sont autant de facteurs pouvant potentiellement influer sur les réponses des participants et qui, pourtant, disparaissent de l'interprétation des résultats de l'expérience telle que proposée ci-dessus. On pourrait par exemple arguer que le contexte entrepreneurial est connoté négativement dans le monde des acteurs académiques qui mènent l'expérience, et que cet élément de contexte conduit les participants, souhaitant se conformer aux opinions du groupe de chercheurs, à charger le chef d'entreprise et à lui prêter, par défaut, des intentions négatives.

Pour prendre en compte ces complexités, et entamer une conversion méthodologique permettant la clarification des entités théoriques en jeu, il conviendrait d'abord d'établir une proposition théorique mieux spécifiée, ensuite de tenter de lister les entités théoriques en jeu puis de préciser comment l'opérationnalisation de chacune de ces entités est envisagée. Prenons par exemple, à titre de reconstruction d'une proposition théorique, sans prétention de philosophie morale avancée : « La description par X des intentions des acteurs d'une scène dépend du jugement moral porté par X sur les résultats des actions décrites ». Cette proposition porte tout d'abord sur « la description par X » ce qui semble facile à opérationnaliser, classiquement, par une double approche qualitative (quelques X à qui on fait décrire de façon détaillée ce qu'ils ressentent) et une approche quantitative (un questionnaire sur Internet, conformément à la technique utilisée par la majorité des 33 articles)[13]. Il faudra ensuite opérationnaliser les scènes, et, pour cela, on aura à prendre position dans les nombreux débats sur la différence entre fictions et scènes réelles, mais je ne reviens pas sur ce point et j'accepte ici l'opérationnalisation par les vignettes pour me consacrer au problème le plus difficile. Quelles entités théoriques sont mobilisées par la proposition « l'attribution d'intentionnalité par les participants X aux acteurs Y figurant dans la vignette dépendent du jugement moral que le participant exerce sur l'action représentée dans la vignette »? A la lumière des 33 articles, établir cette liste semble inatteignable, et de plus la limitation à 33 n'est plus pertinente ici, il faudrait rajouter les centaines d'articles plus anciens ou, à ce jour, plus récents qui ont introduit d'autres distinctions statistiquement pertinentes.

Une des principales difficultés soulevées par l'examen des 33 articles est liée à la multiplicité des schémas causaux (intracrâniens à X) envisagés par les chercheurs. Ils font intervenir des considérations de morale, de causalité, de responsabilité, de culpabilité, d'agentivité, de connaissance (ou de croyance) des conséquences, de volonté, et de bien d'autres termes qui, de plus, dépendent du contexte linguistique (comme par exemple l'intraduisible « accountability »[14]). On ne peut envisager de façon réaliste de demander à X de se prononcer sur chacune de ces notions (pour lui-même et pour Y), d'une part car elles sont trop spécialisées et donnent lieu à des compréhensions incertaines et probablement différentes, et d'autre part, parce que, devant rapporter à O ce qu'il pense, X filtrerait ses réponses pour qu'elles deviennent cohérentes et rapportables. Et le problème ne s'arrête pas là, puisqu'il faudrait ensuite opérationnaliser l'ensemble des notions retenues et proposer, pour chacune d'elles, une ensemble

13. Les études marketing ont démontré la pertinence de cette double approche, et la lutte contre les biais conduit au même constat de l'utilité d'une triangulation. Pourtant, peu des 33 articles mobilise ainsi une double approche.

14. Le terme « accountability » est souvent traduit par « responsabilité » alors que certains auteurs qui l'utilisent en anglais le font précisément pour distinguer deux notions, d'un côté la responsabilité vis à vis des tiers, qui serait objective, et de l'autre le fait subjectif de revendiquer, à la première personne, de porter cette responsabilité. Cette distinction est difficile à traduire en Français.

d'opérations permettant de les détecter, les mesurer, les modifier, les créer, les supprimer à volonté. Même en réduisant notre ambition au seul fait de les détecter, cela semble à nouveau très difficile. On peut néanmoins s'appuyer sur l'article 21, qui utilise l'IRMf, pour jeter un regard sur ce que serait une opérationnalisation développée sur cette base. La séquence présentée par cet article (aujourd'hui largement hypothétique) serait la suivante :

— Dans un premier temps, on suppose que percevoir ce que pense autrui mobilise une « théorie de l'esprit » (en anglais, Theory Of Mind ou TOM) qui est portée par une activité neuronale particulière.

— Après de multiples expériences d'IRMf, la proposition liant un schéma d'activation neuronal et l'activation de la TOM est confirmée : la communauté de chercheurs accepte de faire le lien entre cette activité neuronale et l'activation de la TOM. On a ainsi, au travers de l'IRMf, une opérationnalisation reconnue de la TOM.

— On peut alors vérifier si X active ou non la TOM dans une variante donnée de l'expérience de Knobe.

— Si une proposition théorique fait intervenir une étape qui donne à penser que X devrait activer sa TOM (par exemple en regard de Y) pour pouvoir l'accomplir, alors nous avons le moyen de réaliser une expérimentation pour explorer cette proposition avec une opérationnalisation reconnue pour cette activation de la TOM.

Cette séquence, encore une fois largement extrapolée par rapport au contenu réel plus modeste de l'article 21, montre que, d'une part, l'opérationnalisation est possible et, d'autre part, qu'elle n'est pas sans conséquence puisqu'elle conduit à concevoir les expériences en bénéficiant des opérationnalisations reconnues dans les limites de validité de cette opérationnalisation.

En l'absence d'une telle démarche, qui permet de construire progressivement des opérationnalisations reconnues pour les entités théoriques en jeu, on perd des deux côtés. D'un côté on perd le gain de définir plus rigoureusement des activités intracrâniennes, qui sont ici liées au comportement moral, et de l'autre, on perd toute possibilité de détecter quoi que ce soit sans prêter le flanc à toutes les critiques sur des interprétations inductives abusives.

Mais il faut également souligner la possible contestation ultérieure de cette opérationnalisation. L'IRMf est une technique récente qui évolue et les chercheurs poursuivent la recherche d'une imagerie qui soit à la fois à l'échelle spatiale du neurone et à l'échelle temporelle des processus mentaux. Si une telle recherche aboutissait, il est possible que des résultats appuyés sur la seule IRMf, avec ses limitations actuelles, soient remis en cause.

En résumé, c'est une chose de montrer empiriquement, comme le font chacun des 33 ar-

ticles, qu'un paramètre intervient (statistiquement) dans un processus défini très généralement, ce serait une chose bien différente que d'essayer de définir puis opérationnaliser des propositions théoriques donnant à voir peu à peu les composantes d'une structure explicative de l'attribution d'intentionnalité, ou même plus modestement, de montrer que ce paramètre a un potentiel explicatif suffisant pour s'inscrire dans une approche expérimentale scientifique d'ensemble plausible.

Le cas des discussions de l'effet Knobe est pour moi paradigmatique en cela qu'il met en lumière la dépendance de la démarche du philosophe expérimental vis-à-vis de l'acceptation de l'opérationnalisation qu'il a menée. L'interprétation des résultats expérimentaux est soumise à la validation de cette étape importante et difficile de l'opérationnalisation. L'opérationnalisation ainsi comprise constitue une étape à fort enjeu pour tout chercheur qui souhaite mettre en œuvre une expérimentation et aura à charge de montrer que la situation expérimentale qu'il envisage est convaincante au regard de la relation qu'il étudie. A défaut de convaincre sur ce point, et comme le montre le cas de l'effet Knobe, l'approche empirique court le risque de n'être plus que d'un intérêt anecdotique, multipliant les faits anodins sans contribuer à la construction d'une vue partagée et, par conséquent, sans possibilité de contribuer significativement au débat philosophique.

5.2.1.6 Pour aller plus loin sur la base de ces cinq études de cas

Dans chacune des cinq études de cas, l'analyse du schéma causal joue le rôle d'un instrument de contrôle de la bonne opérationnalisation. Généralisons ce premier constat. Comme dans le cas du savant qui coupe les pattes de la puce et déclare qu'elle est sourde, il est souvent facile de rechercher, et possible de trouver, des chaînes causales alternatives qui ruinent les conclusions tirées d'une expérimentation. Les philosophes analytiques, qui ont depuis des décennies pratiqué cette recherche au travers de la méthode des cas, sont bien armés pour mettre le doigt sur les faiblesses de l'opérationnalisation. Je vais ci-dessous détailler cette analyse du schéma causal utilisée pour mettre en doute une opérationnalisation.

Un deuxième constat, particulièrement clair avec le cas de la transmission des émotions par les larmes et dans celui de l'effet Knobe et des 33 articles, est la grande difficulté pratique à disposer de solutions pour opérationnaliser l'ensemble des entités théoriques potentiellement en jeu dans une expérience. Cette difficulté pratique peut être liée aux techniques spéciales mobilisées, ou aux bases de données établissant des référentiels, ou, comme dans le cas Knobe, à la multiplication des entités théoriques candidates. Dans tous les cas, cela se traduit par une complexification de l'expérience à mener et, potentiellement, par l'explosion

des ressources à mobiliser, qu'elles soient épistémiques, organisationnelles ou financières, qui conduit à douter de la faisabilité pratique de l'expérience envisagée.

A ce doute de faisabilité pratique succède, troisième constat, un autre doute, plus profond. Si l'opérationnalisation n'était simplement pas possible parce qu'il n'y a rien à opérationnaliser ? Si la question n'était ni pratique, comment faire, ni épistémique, savoir qu'on a bien fait, mais ontologique, car l'entité théorique recherchée n'existe simplement pas ? Détaillons tour à tour ces doutes généralisés liés à l'opérationnalisation.

Première généralisation : le schéma causal

Le risque de confusion n'est pas le seul encouru par l'expérimentateur qui sous-estime l'étape d'opérationnalisation mais ce risque est omniprésent en psychologie et participe grandement de la difficulté qu'il y a à opérationnaliser dans ce domaine. Les nombreux exemples décrits au chapitre 2 avec la présentation des résultats du mouvement XPhi montrent qu'il est pratiquement toujours possible d'imaginer des chemins causaux alternatifs qui jettent le doute sur l'interprétation d'une expérimentation. On peut d'ailleurs suggérer sur cette base une liste de sources d'arguments potentiels à disposition de qui souhaite critiquer un résultat expérimental qui semble prouver que A cause B [15] :

— La cause commune : il existe C tel que C cause A et C cause B. On a donc bien toujours B quand on a A, mais sans lien de causalité entre A et B.

— L'épiphénomène : le contexte (ou un facteur inconnu) cause B, donc A cause B est logiquement vrai, mais sans intérêt pratique.

— Le catalyseur : il existe C, C cause B, mais ce chemin causal n'est efficace dans les conditions de l'expérience qu'en présence de A. Il faut bien A pour avoir B, mais pour autant A ne cause pas B.

— La complexité oubliée : il existe D tel que A cause D et D cause B, dans d'autres contextes on pourrait avoir D sans A et donc B sans A. L'incompréhension du schéma global rend toute extrapolation hasardeuse.

— La complexité de A : A est une conjonction de A1 et A2, il se trouve que A1 cause B et que A2 cause B, donc l'expérience conclut à A cause B mais comme A est complexe, on ne peut pas extrapoler (par exemple A1 cause B' qui ressemble à B mais A2 ne cause pas B', pour évaluer l'extrapolation, l'opérationnalisation qui confond A1 et A2 ne suffit plus).

— La complexité de B : B est une conjonction de B1 et B2, il se trouve que A cause B1

15. Cette liste est librement inspirée des contre-exemples aux différentes théories de la causalité rapportés par Max Kistler dans (Barberousse 2011) [16]

mais pas B2 donc il semble que A cause B mais ce n'est pas le cas quand B est un B2.

— La complexité du lien causal : A cause B par deux voies indépendantes v1 et v2 non discernables dans l'expérience.

— etc.

A chacune de ces configurations potentielles correspondrait la nécessité d'expériences de contrôle complémentaires permettant de diminuer le risque de confusion et d'augmenter la confiance dans la pertinence de l'opérationnalisation de la relation « A cause B ». Comme l'illustre l'examen des 33 articles de l'étude de cas sur l'effet Knobe, le chemin causal est particulièrement difficile à établir dans le domaine psychologique. Une des raisons de cette difficulté, que j'ai évoquée plus haut avec l'exemple de la maison qui est à la fois chauffée et climatisée avec deux automates non reliés, est que les comportements psychologiques sont, en général, complexes. Ils résultent d'équilibres très fins entre des raisons de faire et des raisons de rejeter le comportement et, chacune de ces raisons peut elle-même être favorisée ou défavorisée par de multiples éléments de la situation. Le schéma causal à analyser est ainsi souvent plus proche d'un buisson aux multiples interactions que d'un arbre logique bien ordonné. Une expérience particulière ne donne alors accès qu'à une vue ponctuelle qui ne permet pas de discerner l'important de l'anecdotique, le direct de l'indirect, le local du global.

Deux conclusions opposées, pessimiste ou optimiste, peuvent être tirées de ce constat de la difficulté à décrire le schéma causal en place dans une expérimentation de psychologie morale. La première, pessimiste, consiste à considérer que cette complexité est une raison suffisante pour ne pas espérer que la psychologie puisse un jour apporter plus qu'un éclairage partiel et limité aux problèmes qui intéressent le philosophe moral, et en conséquence, seule la philosophie morale est en mesure de servir de cadre global à l'étude du comportement moral humain en se situant à un niveau d'abstraction qui ne sollicite pas l'apport empirique de façon déterminante. La seconde, plus optimiste, consiste à rechercher au travers de la multiplication des études, certes partielles et locales, des vues d'ensemble qui émergeraient du cumul de ces vues locales, un peu à la manière des tableaux pointillistes de Seurat. La multiplication des méta-analyses illustre la mise en œuvre de cette seconde voie, avec les réussites et les limites que j'ai illustrées par l'étude de cas sur l'IAT.

Deuxième généralisation : le coût de l'expérimentation et les capacités pratiques expérimentales

L'analyse du schéma causal est à la base d'une part importante des critiques conduisant à douter des résultats des expériences en psychologie expérimentale. Face à ces difficultés,

il semblerait conforme à la démarche scientifique expérimentale de prendre acte de cette complexité du schéma causal et de tenter de mener des expériences plus sophistiquées, à l'opérationnalisation plus aboutie, qui permettraient de donner un contenu empiriquement distinguable aux différentes hypothèses causales envisagées. Malheureusement, il semble que des aspects pratiques s'opposent à cette évolution alors que, symétriquement, il est pratiquement de plus en plus facile de mener des expériences qui, a contrario, amplifient le doute.

Les difficultés pratiques sont de plusieurs types qu'il peut être utile de distinguer ici. En premier lieu, l'exemple de l'effet Knobe et des 33 articles montre l'énorme difficulté pratique qu'il y aurait à mener des expériences permettant de distinguer les dizaines d'entités théoriques, plus ou moins proches, postulées et de distinguer également toutes leurs combinaisons. En second lieu, la multiplication des notions conceptuellement très proches conduit chaque philosophe à proposer de les différencier selon ses propres critères (et conformément à ses objectifs de recherche) et il n'est pas du tout évident que ces critères se recoupent suffisamment pour bâtir une opérationnalisation qui conviendrait à tous. C'est d'ailleurs l'inverse qui transparaît par exemple dans le cas de l'IAT : pour les uns l'indice opérationnalise une distance entre concepts sans lien avec des propensions à l'action raciste, pour les autres cet indice opérationnalise un biais cognitif qui a forcement une influence sur tout le comportement de la personne, d'où le risque accru de comportement discriminatoire. S'il serait envisageable de sophistiquer le protocole expérimental pour différencier ces deux interprétations de l'IAT face à une seule (ou supposée principale) distinction à clarifier, un tel projet apparaît impossible si on a simultanément, comme pour l'effet Knobe, des dizaines de notions très proches à distinguer.

Ensuite, et de façon plus terre à terre, le montant du budget nécessaire à l'expérimentation, et donc des dépenses à engager avant d'envisager une publication, est à prendre en compte. Pour pouvoir publier un article de même structure que la majorité de ceux relevés pour l'effet Knobe, il faut un travail analytique approfondi de façon à repérer une distinction conceptuelle potentiellement intéressante pour l'étude de l'attribution d'intentionnalité, puis en faire l'analyse bibliographique, et une fois ce travail fait le coût de l'expérimentation est tout à fait modeste (typiquement quelques centaines d'euros en fonction de l'ampleur de l'échantillon visé) : un (ou quelques) questionnaire sur AMT (Amazon Mechanical Turk) et une exploitation statistique sur R, et on a la base de résultats pour instruire la pertinence empirique de la nouvelle distinction conceptuelle proposée. Si on essaye en revanche de disposer d'éléments de triangulation via d'autres modes d'accès à la psychologie des sondés que

le simple questionnaire, par exemple via l'IRMf ou les temps de réactions ou toute autre méthode, les coûts expérimentaux vont évidemment exploser de plusieurs ordres de grandeur (typiquement quelques dizaines de milliers d'euros selon les protocoles mis en œuvre). Il n'est donc pas étonnant qu'une seule étude exploite l'IRMf parmi les 33 articles relevés sur l'effet Knobe. Maintenant, tentons d'imaginer une expérimentation qui prendrait en compte la totalité des distinctions conceptuelles déjà connues comme significatives dans l'étude de l'attribution d'intentionnalité, du fait de l'abondante littérature existante, et tenterait de développer des études statistiquement significatives et ayant pour chaque paire de concepts à bien différencier au moins deux approches distinctes, disons une approche de type explicite par questionnaire et une approche avec un autre type de mesure (temps de réaction, IRMf, influences des produits psycho-actifs, ...). Il est clair que face à des dizaines de distinctions conceptuelles complexes et à toutes leurs combinaisons, et face à la nécessité de triangulation des approches, le budget nécessaire à une telle expérience (imaginaire) serait encore d'un autre ordre de grandeur (peut-être des millions ou des dizaines de millions d'euros). Et plus encore, une telle étude exigerait pour être menée dans un temps raisonnable de monter une équipe qui évoquerait, par son ampleur, les collaborations menées dans le cadre des sciences naturelles et faisant appel à des modèles d'organisations inconnus des philosophes comme des psychologues.

L'étude des 33 articles de l'effet Knobe me conduit alors à une interrogation. Comment concilier, d'un côté, le constat d'une instabilité conceptuelle, liée à la fois à la complexité du sujet et à la facilité avec laquelle on peut proposer encore de nouvelles distinctions dans un mouvement qui apparaît sans fin, et, de l'autre côté, la stabilité qui serait nécessaire pour que puisse être envisagé le lancement d'un programme de recherche structuré, et structurant, multidisciplinaire et coûteux, visant à opérationnaliser l'ensemble d'un domaine comme l'étude de l'attribution d'intentionnalité qui, pourtant, apparaît comme bien modeste en regard de l'ampleur des questions de philosophie morale.

Troisième généralisation : l'existence même des entités théoriques

Les difficultés à définir des schémas causaux et celles, pratiques, à construire un cadre expérimental, difficultés que je viens de relever, peuvent être analysées pour elles-mêmes, mais elles peuvent aussi être interprétées comme le signe d'un problème plus profond. Si les philosophes et psychologues ont tant de difficultés face à l'opérationnalisation des entités, propriétés et relations qu'ils postulent dans leurs théories, c'est peut-être que certaines de ces entités, propriétés et relations n'existent simplement pas. A l'image du phlogistique qui a pu être postulé mais n'a pas survécu à l'approche empirique développée par Lavoisier, peut-être

les entités postulées dans nos théories psychologiques n'existent-elles pas, ou plutôt, elles n'existent pas telles qu'elles apparaissent dans ces théories. L'exemple des émotions et du tramway peut nous éclairer sur ce point.

L'article de 2001 de Greene [108] que j'ai présenté plus haut, a proposé, en s'appuyant sur l'imagerie fonctionnelle, que le cerveau d'une personne qui répond au dilemme du tramway ne s'active pas dans les mêmes zones selon qu'il s'agit du cas où il faut manipuler un aiguillage pour sauver cinq personnes mais en sacrifier une, ou du cas où il faut pousser un gros homme sur la voie, qui est ainsi sacrifié pour sauver les cinq personnes. Plus précisément, Greene fait un lien entre le premier cas, l'aiguillage, l'activation de zones du cerveau connues pour être celles du raisonnement, et le second cas, le gros homme, et l'activation de zones connues pour être celles des émotions. Acceptons, à titre d'hypothèse, la corrélation entre l'activation d'une certaine zone du cerveau, visualisée par l'IRMf, et le fait que la personne éprouve une émotion, ce que l'on aura asserté par une autre méthode, introspective ou physico-chimique [16]. Dans le raisonnement du spécialiste de l'IRMf, la notion d'émotion est entendue de façon assez large car il n'a pas accès à mieux avec sa technique. Il ne sait simplement pas s'il s'agit de surprise, de colère, de joie, de dégoût, de peur, et cette ignorance est liée tant à la variabilité interpersonnelle des comportements émotifs qu'à la difficulté à distinguer les occurrences de ces émotions une à une et en associations. La corrélation constatée est une statistique sur des événements où varient à la fois stimuli et personnes, et le résultat n'est donc pas interprétable assez finement pour affirmer que telle personne a éprouvé telle émotion.

Observons maintenant l'utilisation de la corrélation qui est faite par Greene dans le cas du tramway. La présentation de cette corrélation est fortement dépendante de l'émotion qui est envisagée, et il n'est pas équivalent que ce soit de la colère, de la joie, du dégoût ou de la peur (ou toute autre émotion de base ou composée que le théoricien souhaitera postuler). Par exemple, si l'émotion ressentie par la personne s'imaginant pousser le gros homme sur les rails vers une mort certaine était de la joie, on aurait à chercher une narration explicative bien différente de celle qui vient à l'esprit si cette émotion est le dégoût, et qui correspond approximativement mieux à la narration explicative standard : nous refusons de pousser un autre être humain car cet acte nous répugne profondément et immédiatement.

La validité du raisonnement de Greene repose donc, implicitement, sur une information, le type d'émotion, qui est inaccessible à l'IRMf. Plus profondément, c'est le regroupement de la joie, la peur, le dégoût, etc. sous un même terme qui pose ici question. Si ces émotions ne

16. Dans un entretien sur Internet de février 2020, Florian Cova met en doute que les spécialistes de l'IRMf d'aujourd'hui accepteraient encore les conclusions de l'article de 2001. Mon argument ne sera que renforcé si on considère cette corrélation elle-même comme fragile.Voir `https://www.youtube.com/watch?v=xc31-WwVlRw`

sont pas sollicitées par les mêmes stimuli, et qu'elles n'ont pas les mêmes conséquences, ne serait-il pas plus expédient de considérer qu'il ne convient pas d'utiliser ce concept trop vague et général d'émotion dans nos théories morales ? Et dans ce cas, la question, vue du spécialiste de l'IRMf, changerait de nature car il s'agirait de comprendre ce qu'ont en commun ces différentes émotions si, et dans la mesure où, elles sont pour partie corrélées avec l'activation d'une même zone du cerveau.

Si on accepte cette analyse que le terme « émotion » correspond à un regroupement de cas très différents quant à leur inscription possible dans les interprétations que l'on peut donner dans le cas de l'expérience du tramway, alors il en résulte que, en regard de ce qui est en jeu dans ce cas, l'entité théorique « émotions » n'existe pas en tant que point d'entrée intéressant à opérationnaliser pour l'expérimentateur, chacune des émotions existe indépendamment des autres et leur regroupement n'a pas de sens expérimental. La formulation contraposée est que, si les expérimentateurs acquièrent des savoir-faire leur donnant accès à une régularité qui n'est pas une émotion précise mais un regroupement de diverses émotions, alors les théoriciens peuvent se demander comment ils vont interpréter cette régularité pour améliorer, et peut-être remplacer, le regroupement théorique qu'ils postulent sous le terme « émotion ».

5.2.2 Les problèmes de l'opérationnalisation

Le risque de confusion, nous l'avons vu, advient lorsque de multiples interprétations semblent compatibles avec les protocoles et les résultats d'une expérience donnée. Une formulation proche, adaptée au cas des tests médicaux permettant de révéler la présence d'une maladie, est celle du risque des « faux positifs » et « faux négatifs », quand le test conclut à trop ou pas assez de malades.

On peut comprendre le terme « d'opérationnalisation » comme regroupant tous les éléments de démarche qui permettent à l'expérimentateur et au théoricien de se mettre d'accord sur les meilleurs protocoles expérimentaux permettant d'éviter ce risque de confusion. Par extension, une première définition de l'opérationnalisation serait en ce sens de rechercher des champs expérimentaux, des protocoles, des instruments, des pratiques, pour détecter, mesurer, modifier, créer, supprimer une occurrence de l'ensemble (entité, propriété, relation) postulé par une théorie en minimisant le risque de confusion.

Plus simplement, l'opérationnalisation est la recherche d'un terrain d'expérimentation et des gestes pratiques permettant d'obtenir des résultats empiriques pertinents en regard d'une théorie, la pertinence étant évaluée sur la base de l'approbation d'une communauté

de chercheurs. Il me faut maintenant être plus précis sur ce que signifient ici les différents termes de cette définition et, en particulier, ce que sont les théories qu'on souhaite conforter empiriquement, et en quoi un terrain d'observation peut être compris comme opérationnalisant de façon pertinente un ensemble d'entités postulées par ces théories.

J'insisterai sur deux caractéristiques importantes de l'opérationnalisation. La première est qu'elle tend à résoudre, au mieux de l'épistémè du moment, un ensemble d'équations complexes optimisant la possibilité d'accord des différentes communautés de chercheurs concernés, sous la contrainte de la faisabilité pratique de l'expérience envisagée. La seconde, déclinant la première dans le domaine de l'étude du comportement humain, est que le résultat de cette optimisation dépend de façon déterminante de l'échelle selon laquelle est abordée l'étude.

5.2.2.1 Opérationnaliser, c'est résoudre simultanément trois équations

J'ai proposé plus haut de caractériser une théorie comme postulant des entités, des propriétés pour ces entités et des relations entre ces entités et propriétés. Opérationnaliser, c'est rechercher des protocoles, des pratiques expérimentales pertinents pour que les résultats empiriques puissent être rapprochés de façon crédibles de ces entités, propriétés et relations postulées par les théories. Illustrons avec un exemple. Une théorie psychologique peut postuler, comme dans le cas de l'IAT évoqué plus haut, qu'existent des concepts (comme celui d'humain Noir ou Blanc, et le sentiment Positif ou Négatif nécessaires à l'indice Race IAT), et qu'existe la propriété de « proximité » de deux concepts (le concept Noir pourrait être alors plus proche de Négatif que de Positif pour certaines personnes).

En ce sens, la proposition théorique « les concepts Noirs et Négatif sont proches pour les individus racistes »[17] signifie donc que, dans cette théorie sont postulés les éléments suivants :

— Il existe une entité « concept », et une entité « catégorie d'individu »
— Il existe une propriété « proximité des concepts »
— Il existe un type de relation qui peut être le cas entre ce type d'entité et ce type de propriété (deux concepts sont plus proches que deux autres) de façon stable pour une « catégorie d'individus ».
— C'est le cas que Noir est un concept, Négatif est un concept, Raciste est une « catégorie d'individus » et que la relation de proximité est avérée entre les concepts de « Noir » et

17. Je simplifie ici le test IAT qui s'appuie sur une double différence entre positif et négatif et entre noirs et blancs. Voir 4.5, page 197 pour le détail de ce test.

« Négatif » pour les individus appartenant à la catégorie des racistes.

Une telle proposition théorique peut faire l'objet de différentes interprétations métaphysiques. Elle peut par exemple être lue comme comportant un engagement ontologique fort : les entités, propriétés et relations postulées existent dans le monde. Cet engagement peut prendre de multiples formes et ouvrir vers de nombreux débats classiques comme celui sur le rapport entre particuliers et universaux. Je n'entrerai pas dans ces débats ici car ils ne sont pas centraux pour mon propos. En effet, la proposition théorique peut également être lue comme simplement utile à une description du monde qui nous facilite la vie, sans engagement fort. Indépendamment de tout engagement ontologique, on peut chercher à conforter empiriquement la proposition théorique en recherchant des situations pratiques que l'on puisse interpréter comme des cas concrets pertinents de l'emploi du mot « concept », du mot « catégorie d'individu », de la propriété « distance entre concepts » et ici plus précisément de l'affirmation « les concepts Noirs et Négatif sont proches pour les individus racistes », et ce sans chercher à préciser l'engagement ontologique porté par cette proposition, au moins dans un premier temps.

Face à cette proposition théorique, envisager la possibilité de l'apport empirique suppose d'avoir résolu simultanément trois équations. Appuyons nous à nouveau pour les expliciter sur l'exemple de l'IAT. L'opérationnalisation envisagée par les expérimentateurs qui ont mis au point le protocole de l'IAT s'appuie sur l'idée que le temps de réaction est court quand il faut rapprocher deux concepts proches et plus long quand il faut rapprocher deux concepts éloignés. Plus précisément, et pour reprendre chacun des éléments ci-dessus, cette idée s'appuie sur un cheminement qui peut être décrit ainsi :

— Le concept est compris comme un état mental activable.
— Un raisonnement s'appuie sur ces états mentaux disponibles, car activés à un moment donné.
— Le raisonnement est plus rapide et plus facile si les concepts sont proches, c'est-à-dire souvent simultanément activés.
— Le temps de réaction permet de voir si deux concepts sont proches.
— Quand deux concepts sont proches, cela caractérise la personne en ce que cette proximité influence la rapidité et la facilité de tout raisonnement les activant simultanément.

L'opérationnalisation consiste en la recherche de cas concrets permettant d'étudier la proposition « les concepts Noirs et Négatif sont proches pour les individus racistes » et de définir ce qui, dans le cas concret envisagé, donne confiance à la communauté des chercheurs dans

la caractérisation de ce cas concret comme pertinent pour juger de la proposition théorique : il s'agit de façon convaincante de concepts, Noir et Négatif, et que ces concepts sont dans la relation de proximité prévue lorsque les personnes appartiennent à la catégorie « racistes ».

Première équation, existentielle

Trois équations sont alors à résoudre simultanément. La première est existentielle, la seconde est épistémique, la troisième pratique. L'équation existentielle est d'arriver à faire correspondre les entités (propriétés ou relations) théoriques avec des observables dont l'extension soit (raisonnablement) identique à ce que prévoit la théorie, en minimisant le risque de confusion, de faux positif et de faux négatif. Ainsi par exemple, si en montrant plusieurs images qui toutes se réfèrent à « un humain Noir », il apparaît que cela n'induit rien de commun qui puisse évoquer qu'un concept a pareillement été activé, alors on ne peut que constater que l'on n'a pas opérationnalisé le concept Noir comme souhaité. De même, si une personne soumise à un test donné supposé opérationnaliser la détection de son caractère raciste n'obtient jamais le même résultat d'un jour à l'autre, on ne peut que constater que l'on n'a pas opérationnalisé l'appartenance à cette catégorie, supposée stable dans le temps, comme souhaité.

Naturellement, si après de multiples essais, l'expérimentateur n'arrive pas à disposer d'une pratique satisfaisante pour définir un observable qui soit dans une relation d'extension cohérente avec une entité théorique postulée, il sera amené à douter de la pertinence de cette entité, et donc de la validité de la théorie, mais l'opérationnalisation elle-même a pour but premier de chercher à résoudre cette équation existentielle, non d'en montrer l'impossibilité.

La nature précise de la relation entre une entité théorique et ce qui l'opérationnalise dans le cadre d'une expérience donnée est sujet à de nombreux débats qui dépendent en grande partie des options métaphysiques du théoricien en regard de sa théorie. Je ne peux rentrer dans ces débats ici et c'est pourquoi je propose ci-dessus une relation minimale très simple, mais nécessaire à la poursuite du travail scientifique : que les expérimentateurs et les théoriciens soient d'accord sur l'acceptabilité de la co-extension entre l'entité théorique postulée et l'observable construit par l'expérimentateur. Ou, autre formulation, que les descriptions de l'expérience, de ses protocoles, de son déroulement et de ses résultats, faites soit par l'expérimentateur, soit par le théoricien, soient acceptées comme possiblement cohérentes. Pour décrire cette relation entre l'entité théorique et son opérationnalisation, on peut également évoquer, en référence au rapport de la carte au terrain, le terme d'homologie qui évoque à la fois la co-extension et une similitude de formes qui donne à penser que l'opérationnalisation est aboutie.

Dans le cadre de l'IAT, le participant doit appuyer sur une certaine touche A du clavier lorsque l'image qui lui est présentée évoque un Noir. La description de l'expérience, partagée par le théoricien et l'expérimentateur, est que l'image active des états mentaux, qui sont des concepts, que ces concepts sont mobilisés par le raisonnement du participant qui suit le protocole proposé et que, en conséquence, il appuie sur la touche A. La co-extension de l'entité théorique « concept Noir » et de cette façon de l'opérationnaliser peut être vérifiée globalement si à toute image évocatrice de Noir, le participant appuie sur A et si à toute image non évocatrice, il n'appuie pas sur A. Naturellement, cette opérationnalisation est sommaire et ne satisferait pas quelqu'un qui souhaiterait savoir ce qu'est exactement, du point de vue neuronal par exemple, ce « concept Noir » dont théoricien et expérimentateur parlent, mais ce n'est pas ici la question posée, l'opérationnalisation doit simplement être reconnue comme pertinente pour permettre d'interpréter l'expérience menée en regard de la proposition théorique « les concepts Noirs et Négatif sont proches pour les individus racistes ». La co-existence du « concept Noir » est considérée comme suffisamment empiriquement établie si, quand l'image est montrée qui évoque un noir, le participant appuie régulièrement sur la touche A.

Le protocole de l'IAT s'affine ensuite en mesurant le temps de réaction, le temps que met le participant entre l'affichage de l'image et le moment où il appuie sur la touche A. L'expérimentateur, sur la base de cette mesure, va montrer qu'il existe statistiquement, pour un même individu, des différences de temps de réaction entre concepts. Expérimentateurs et théoriciens doivent alors évaluer s'il y a une co-extension acceptable entre ces différences de temps de réaction et la facilité d'accès à ces concepts dans le cadre du protocole expérimental. Dans le cadre de l'IAT, l'ensemble des chercheurs a agréé ce point, et considère que l'effet IAT existe : le temps de réaction est accepté comme un indicateur de la facilité à accéder à un concept, ou, plus exactement, qu'un raisonnement appuyé sur des concepts proches va plus vite qu'un raisonnement appuyé sur des concepts éloignés ou contradictoires.

Deuxième équation : épistémique

La deuxième équation, que j'ai qualifiée d'épistémique, consiste à vérifier ce que l'expérimentateur peut connaître des observables et de leurs relations, et en quoi cette connaissance est transposable par le théoricien dans son propre domaine. Résoudre cette équation va mobiliser l'ensemble des théories qui constituent l'environnement intellectuel de l'expérimentateur, de toutes les disciplines mises en œuvre dans l'expérimentation et, plus largement, de la communauté des chercheurs incluant également les théoriciens.

La question épistémique, reprenant l'exemple de l'IAT et du cas simple de « concept », est celle-ci. Comment puis-je être certain que l'emploi que je fais ici et maintenant du mot

«concept» est pertinent si je ne sais déjà ce qu'est un concept ? Et si je reconnais le concept parce qu'il a certaines propriétés, par exemple parce qu'il semble être partagé au sein d'une population, d'où me vient cette connaissance pratique d'une entité théorique ? Si cette connaissance vient de la théorie, alors mon observation est chargée de la théorie que je prétendais étudier, et mon approche souffre d'une circularité vicieuse, mais si elle vient de la pratique, comment puis-je être certain de bien parler de ce « concept » que la théorie postule ?

Définie ainsi, l'équation épistémique semble relever d'une mission impossible tant elle est confrontée aux défis sceptiques. Cette aporie a une structure analogue à celle décrite par Baas Van Fraassen dans (VanFraassen 2010 page 115) [231] sous le nom de problème de la coordination. Il s'agit de se prononcer sur la qualité d'une mesure d'un phénomène physique. Pour dire qu'une mesure est de bonne qualité, et quelle que soit la définition qu'on se donne de cette qualité, il semble qu'il faille la comparer à une autre mesure. Mais alors, comment savoir que cette première mesure est elle-même de bonne qualité ? Soit on ouvre vers une régression infinie et on ne sait comment débuter la première mesure, soit on accepte simplement comme un indice de qualité que les mesures soient proches, et alors il est toujours possible qu'elles soient toutes deux également erronées.

Bref, il semble illusoire de vouloir établir une bonne mesure, et pourtant c'est bien ce à quoi arrivent les scientifiques, au moins pour certains types de grandeurs [18]. Il peut donc apparaître préférable de se tourner vers la méthode qu'ils utilisent concrètement pour contourner ce problème. La solution proposée par Van Fraassen sur cette base consiste à considérer que si l'on ne peut établir une mesure absolument bonne, on peut toujours améliorer une mesure existante. Il rejoint ainsi ce que j'ai présenté avec la métaphore de l'hélice expérimentale et que Hasok Chang appelle l'itération épistémique : à chaque tour de l'hélice nous avons un certain état des théories, des objets du laboratoire et des traces des expériences passées sur lesquels nous appuyer pour bâtir une nouvelle itération qui, en particulier, visera à améliorer les mesures existantes.

Appliquant la même stratégie à l'opérationnalisation, le scientifique s'appuie sur les théories, les opérationnalisations déjà reconnues comme pertinentes, et les résultats des expériences passées pour adopter les critères de pertinence qu'il utilisera dans sa recherche d'une nouvelle opérationnalisation encore plus convaincante. En ce sens l'opérationnalisation n'est donc plus impossible, elle devient possible à condition de s'appuyer sur une communauté de chercheurs qui a construit les savoir-faire théoriques et expérimentaux existants et de se

18. A titre d'exemple extrême, les physiciens pensent pouvoir mesurer le temps avec 18 chiffres significatifs dans les années qui viennent. La définition actuelle de la seconde, 9 192 631 770 périodes de la radiation du césium133, donne, avec 10 chiffres significatifs, une idée de la précision déjà atteinte dans les années 1960.

donner pour objectif non d'établir une opérationnalisation définitive et absolument bonne, ce qui nous reconduirait dans l'impasse ci-dessus, mais, de façon plus modeste et plus réaliste, d'améliorer l'opérationnalisation des entités théoriques que l'on souhaite soumettre à l'enquête empirique. Cette conception de l'opérationnalisation est pour une part importante descriptive de la pratique des scientifiques, comme le montre l'exemple de la température développé par Hasok Chang dans (Chang 2007) [42], elle est pour partie réformatrice lorsqu'on considère qu'elle devrait être mise en œuvre pour que s'améliorent les connaissances dans un domaine donné.

On peut ainsi envisager que l'étape de l'opérationnalisation franchisse à chaque tour de l'hélice expérimentale les phases suivantes. D'abord, pour chaque entité théorique, pour chaque propriété et pour chaque relation de la proposition visée, ou plus largement pour l'ensemble des propositions d'une théorie, relever les protocoles pratiques utilisés dans les expériences passées qui ont permis de détecter, mesurer, modifier, créer, supprimer cette entité, cette propriété ou la relation à étudier, expériences dont on a les traces enregistrées (par exemple dans les publications les relatant). Puis, élaborer une critique de ces protocoles au sein de la communauté des chercheurs [19], et repérer les améliorations à apporter pour écarter les ambiguïtés, améliorer la précision et, en bref, augmenter la confiance de la communauté dans la pertinence de ces protocoles. Enfin, rechercher des nouveaux protocoles expérimentaux susceptibles d'apporter les améliorations requises, les tester, pour pouvoir enfin les soumettre à la communauté des chercheurs.

Soulignons, en commentaire de cette équation épistémique que doit résoudre l'opérationnalisation, ce que l'opérationnalisation ne saurait être. Tout d'abord elle n'est ni métaphysique ni logique, il n'y a pas, au travers de l'opérationnalisation, une volonté d'identification entre un domaine théorique et le laboratoire de l'expérimentateur. Lorsque celui-ci emploie le verbe « être » en conclusion de son expérience, comme dans « C' **est** le concept Noir qui a été activé par l'image présentée », il convient de la compléter par « comme les théoriciens et les expérimentateurs en sont d'accord, le concept Noir est ici bien opérationnalisé par le protocole ». Il n'y a pas identité entre un élément du domaine théorique et une circonstance du domaine empirique, il n'y a qu'un rapprochement validé par les pairs et permettant d'envisager de transférer vers le premier une connaissance acquise sur la seconde.

Autre exclusion à clarifier, l'opérationnalisation n'a qu'un rapport très indirect avec le réductionnisme. Bien sûr, il est possible que le théoricien souhaite, au sein de sa théorie,

19. Je ne rentre pas ici dans l'analyse de comment se constitue cette communauté ni de comment elle agit, voir par exemple comment pourrait se construire un pilotage démocratique des activités scientifiques dans (Kitcher et Ruphy 2010) [135].

réduire l'entité « concept » à une configuration de connexions neuronales, mais ce n'est nullement nécessaire pour que le « concept » soit considéré comme bien opérationnalisé quand on le considère comme activé par une image et mobilisé par un raisonnement dans le protocole de l'IAT. On peut ainsi acquérir une connaissance empirique sur ce « concept » sans pour autant que celui-ci ne soit en cela réduit. La réduction, qui joue entre niveaux théoriques, peut faciliter l'opérationnalisation si les entités des théories de base ont fait l'objet d'opérationnalisations reconnues, mais elle n'est ni nécessaire ni suffisante pour que soit acceptée l'opérationnalisation des entités des théories dont on envisage la réduction.

Troisième équation : pratique

Enfin la troisième équation est pratique : il faut que les opérationnalisations soient réalisables en regard des ressources pratiques, financières, humaines, calculatoires, etc. disponibles à l'équipe de recherche. J'ai souligné l'importance de cette équation pratique pour l'étude de l'attribution d'intentionnalité et de l'effet Knobe : seule 1 étude sur 33 a fait appel à l'IRMf, et la plupart des études sont limitées à une opérationnalisation appuyée sur quelques questionnaires, donc sur la seule introspection et sans triangulation.

La limitation peut être budgétaire, elle est également technologique et profondément dépendante des avancées dans d'autres domaines de la connaissance, comme le montrera le cas de l'étude du confort acoustique présentée plus loin. De façon actuellement très importante, on peut par exemple souligner la non-disponibilité d'une technique d'imagerie cérébrale qui soit à la fois à l'échelle spatiale du neurone, ou de la synapse, et à l'échelle temporelle de la dynamique de la pensée. Cette absence a pour conséquence une forte limitation des opérationnalisations possible quand on recherche des observables neuronaux homologues (ou corrélés) à des indicateurs observables des comportements.

Enfin, si l'équation pratique est budgétaire et technologique, elle est également, et peut-être surtout dans le domaine de la psychologie morale, sociale. Comme l'ont montré les différentes études de cas du chapitre 4, il faut en pratique mettre en communication de nombreuses équipes pour pouvoir espérer mener une expérimentation validée par les pairs. La psychologie morale est par construction fortement pluridisciplinaire[20], et la collaboration entre philosophes, psychologues, neurobiologistes, biologistes, ... ne va pas de soi.

Résoudre simultanément ces trois équations, existentielles, épistémiques et pratiques est donc la tâche à mener pour réussir une opérationnalisation qui minimise le risque de confusion. Ce risque est une préoccupation majeure de la recherche en psychologie. A titre d'illus-

20. Cette pluridisciplinarité est d'ailleurs revendiquée comme la principale caractéristique de la psychologie morale dans (Doris 2012) [76]

tration la recherche sur l'expression complète « confounds in psychological research » sur Google retourne 3660 articles [21], et tous les ouvrages méthodologiques du domaine consacrent une partie importante à ce problème (pour exemple : (Coolican 2014 page 102) [54]).

En résumé, l'opérationnalisation vise deux objectifs qu'elle doit atteindre sous un faisceau de contraintes. Le premier objectif est de trouver des situations qui permettront de mettre en œuvre les protocoles pratiquement exploitables en regard de la proposition théorique à étudier, c'est-à-dire un ensemble de protocoles permettant d'observer, mesurer, détecter, modifier, créer, supprimer, des phénomènes qui seront mis en correspondance des entités, propriétés et relations postulés dans une théorie. Cette recherche s'appuie sur les savoir-faire pratiques et théoriques, sur les expériences, traces et théories déjà en place. Le second objectif est la nécessaire acceptation par la communauté de chercheurs, théoriciens et expérimentateurs, de la pertinence de ces situations en regard de cette proposition théorique. Enfin le faisceau de contraintes pragmatiques est constitué de toutes les limitations liées aux ressources nécessaires pour mener l'étude sur le terrain expérimental retenu (ressource est ici entendu au sens large de ressources financières, humaines, temporelles, d'existence et disponibilité des équipements de laboratoire, des techniques et instruments de détection et mesure, des règles éthiques, etc.).

5.2.2.2 L'échelle de l'opérationnalisation.

Observons la définition de l'opérationnalisation appliquée à l'exemple de l'IAT utilisé plus haut de la proposition théorique « les concepts Noirs et Négatif sont proches pour les individus racistes ». Le psychologue va rechercher comment pratiquement construire une expérimentation où un individu est induit vers un état d'esprit négatif lorsqu'il est mis en présence d'un noir. Le psychologue recherchera comment assurer cette mise en présence et comment évaluer cet état d'esprit négatif de façon convaincante pour d'autres psychologues.

Telle que je viens de la décrire, l'opérationnalisation réussie du psychologue s'appuie implicitement sur l'individu en tant qu'unité étudiée. La formulation de l'expérimentation s'entend dans le contexte de l'individu isolé étudié, du stimuli reçu par l'individu et des effets sur l'individu, rendus observables soit par son comportement soit par des mesures sur l'individu.

Il convient de remarquer que cette échelle de l'objet étudié, l'individu, est arbitraire, il est fondamentalement celui du psychologue, et c'est cette échelle individuelle qui définit sa discipline. On pourrait également envisager une échelle supérieure, par exemple le groupe social ou la société, qui concernerait le sociologue, ou une échelle inférieure, les organes ou

21. Consulté le 29 février 2020.

les neurones, qui concernerait le biologiste. Et à chacune de ces échelles correspondraient des opérationnalisations différentes. Poursuivons pour illustrer ce point l'exemple de l'IAT.

Prise à l'échelle de l'individu, la situation expérimentale semble établie par le test qui établit le lien implicite entre concepts en s'appuyant sur l'analyse des temps de réaction. On pourra, comme on l'a vu, opérationnaliser la présence de Noir par la présentation des images, et la proximité à l'état d'esprit négatif par la mesure du temps de réaction.

Mais, en supposant que cette démarche soit considérée comme pertinente par des psychologues, le serait-elle au niveau social ? On peut imaginer que non, car elle laisse de côté un élément important pour le sociologue : comment expliquer que dans certaines sociétés ce soient les Noirs qui soient ainsi associés aux sentiments négatifs et que, en d'autres lieux ou à d'autres époques, d'autres catégories de personnes aient fait l'objet de phénomènes analogues ? L'intérêt du sociologue pour les conditions d'apparition dans une société donnée d'un phénomène comme l'IAT le conduira à vouloir étudier si, par exemple, dans chaque contexte social il y a toujours (ou souvent) une ou plusieurs catégories de personnes ainsi stigmatisées, si le contenu des sentiments négatifs est le même ou pas, si ce sentiment conduit à certaines discriminations, et si oui lesquelles et sous quelles conditions, etc. Pour conforter empiriquement ces recherches, il aura à bâtir des opérationnalisations qui ont peu en commun avec le protocole IAT car la recherche sera transculturelle et que l'unité de travail est maintenant le groupe social et non plus l'individu du psychologue.

Et à l'échelle intracrânienne, la même question posée à un neurologue appellera des questions portant sur les réseaux de neurones mobilisés dans la perception des visages et comment ils sont en relation avec les réseaux de neurones corrélés avec le concept de catégories de personnes, comme « Noir », et ensuite comment des concepts, en tant que réseaux de neurones, peuvent être « proches » et en quoi cela change le temps de réaction. Là encore, l'opérationnalisation utile à aborder ces questions sera complètement différente et le protocole IAT n'y suffira pas.

L'opérationnalisation acceptable de propositions théoriques proches au premier regard dépend de la communauté de chercheurs concernés, donc de la discipline scientifique concernée, et on peut ici décrire les disciplines comme différant dans l'échelle retenue pour examiner un phénomène particulier aux multiples facettes. En ce sens, l'analyse de l'opérationnalisation permet de révéler la complexité des phénomènes étudiés en regard de propositions théoriques qui visent à expliquer ces phénomènes au sein d'une discipline scientifique particulière. L'échelle choisie par chaque discipline est souvent implicitement portée par l'environnement intellectuel dans lequel se développe la théorie et se traduit concrètement par

l'acceptation du type d'opérationnalisation qui est considéré comme pertinent par ces experts de la discipline concernée à l'échelle des conclusions de sa discipline.

5.2.3 Les enjeux de l'opérationnalisation

Je peux reformuler maintenant le problème posé par l'opérationnalisation en regard de ma recherche. Quand un philosophe moral établit une proposition comme par exemple « Le jugement moral est lié à nos émotion », il inscrit sa proposition dans une théorie qui postule que le jugement moral existe, que les émotions existent, qu'il existe une relation de lien et qu'il est le cas que, dans sa théorie, le jugement moral est lié à nos émotions. Lorsque le philosophe moral expérimental vise à conforter, infirmer ou simplement éclairer empiriquement cette proposition, il doit procéder à l'opérationnalisation nécessaire : rechercher des cas concrets, des terrains d'expérience et des protocoles dont les mises en pratiques soient pragmatiquement possibles et soient acceptées comme pertinentes par sa communauté de chercheurs pour que les résultats puissent être interprétés comme des occurrences d' un « jugement moral », d'une « émotion », et appuyer ainsi l'acceptation de l'existence de ce jugement moral et de cette émotion, et que, enfin, ces résultats puissent être reconnus comme reflétant la relation « le jugement moral est lié à nos émotions » en ayant écarté, autant que faire se peut, les interprétations alternatives de ces résultats expérimentaux.

L'enjeu de l'opérationnalisation est double, selon qu'on l'observe négativement ou positivement. Négativement, si l'opérationnalisation échoue, c'est-à-dire si les pairs ne considèrent pas que les résultats expérimentaux sont pertinents en regard des théories étudiées, parce que, par exemple, les risques de confusion sont trop importants, alors toute interprétation de ces résultats sera entachée de doute et amenée à être remise en cause quand la confusion sera levée. Aucune des entités, propriétés ou relations postulées par les théories ne recevra d'appui empirique face à cette potentielle remise en cause. Ce constat, négatif, porte en lui la conséquence, positive, d'une remise en cause potentielle de la théorie empiriquement réfutée si se multiplient ainsi les échecs expérimentaux.

L'enjeu est également positif, car si l'opérationnalisation réussit, si les pairs acceptent la pertinence du protocole expérimental, alors l'expérimentateur pourra communiquer les traces laissées par ses expérimentations, et les pairs pourront, en confiance, exploiter ces traces pour faire progresser les connaissances du domaine. De plus, cette opérationnalisation réussie permet de disposer d'un accès empirique validé à des entités théoriques, et ces accès pourront être réutilisés pour d'autres expériences portant sur ces mêmes entités, il

s'agit donc d'un investissement utile, une base pour de futurs travaux de recherches. En ce sens, l'enjeu de l'opérationnalisation est de contribuer à « l'effet de cliquet » qui caractérise la démarche scientifique, pour reprendre le titre de l'article d'Anouk Barberousse dans (Silberstein 2018) [209]. L'opérationnalisation, d'un côté, accroît la confiance dans l'existence des entités théoriques dont on a réussi l'opérationnalisation et, d'un autre côté, permet d'obtenir de nouveaux résultats empiriques relatifs à ces entités. Enfin, sur un autre plan, l'opérationnalisation réussie renforce la cohérence des équipes de chercheurs qui adoptent cette opérationnalisation et, avec elle, la façon d'aborder les questions posées et les échelles pertinentes pour les aborder.

Soulignons également un enjeu qui vaut que l'opérationnalisation soit acceptée ou non par les pairs : les débats sur la pertinence de l'opérationnalisation permettent de mettre en évidence les complexités des systèmes étudiés. Que ce soit parce que les communautés de chercheurs diffèrent quant à l'échelle qu'ils privilégient, que ce soit parce que la chaîne causale envisagée a été jugée trop simple en regard du phénomène étudié, que ce soit parce que les ressources disponibles, financières, épistémiques ou technologiques manquent, la volonté d'établir des apports empiriques se heurte à la complexité de l'objectif et, ainsi, permet d'en prendre toutes les dimensions.

Face à cette complexité, il peut alors être tentant, et il a été tenté, de contourner la difficulté de la validation de l'opérationnalisation par des pairs. Dans la section suivante je me propose de décrire trois stratégies qui vont dans ce sens dans le domaine psychologique. La première est le behaviorisme, la seconde est l'opérationnalisme de Bridgman et la troisième l'assimilation d'un phénomène psychologique a un phénomène physique, à titre de première simplification.

5.3 Trois stratégies de contournement

5.3.1 Le behaviorisme

Au début du 20^{e} siècle, des psychologues constatant les dérives du mentalisme avec en particulier les difficultés de l'introspection et désireux de rapprocher leur discipline des sciences expérimentales, ont proposé avec le behaviorisme une nouvelle façon d'aborder l'étude du comportement humain. L'article de John B. Watson « Psychology as the behaviorist views it » (Watson 1913) [233] est généralement considéré comme marquant le début de ce mouvement [22]. Le comportement est analysé comme la réponse d'un organisme à un stimuli externe,

22. Voir l'article (Malone 2014) [152] pour une revue des sources du behaviorisme au 19^{e} siècle, avant Watson.

d'où le nom de modèle S-O-R (stimuli, organisme, réponse) qui est associé au behaviorisme. La réponse peut être innée, et souvent réflexe, ou peut être acquise en conséquence des interactions de l'organisme avec son environnement.

Ce modèle fait ainsi l'économie considérable de ne pas avoir à rentrer dans le détail des processus psychologiques tels qu'ils se déroulent à l'intérieur de l'organisme. L'essentiel du travail est de bien décrire les situations, avec en particulier les stimuli significatifs, puis d'observer les réponses de l'organisme. L'intervention consiste à faire subir des interactions à l'organisme pour voir comment évoluent alors les réponses, à l'image du chien de Pavlov qui, après dressage, salive à l'écoute du stimuli annonçant le repas. Ce programme de recherche s'est développé dans cette direction pendant toute la première partie du 20^{e} siècle, avec en particulier des psychologues comme B F Skinner (1904, 1990) (Skinner 2008) [211].

Dans cette démarche, la difficulté à opérationnaliser les éléments postulés par les théories psychologiques est contournée car seuls restent à observer des éléments extérieurs à l'organisme, d'une part les situations, qu'on peut voir comme une généralisation de stimuli, et d'autre part les comportements observables de l'individu. Mais à partir des années1950, les limites de ce programme de recherche apparaissent clairement. Pour Chomsky, les bases du behaviorisme ne permettent pas de comprendre des phénomènes complexes comme le langage, il est indispensable d'entrer dans la structure interne des compétences des individus pour comprendre comment ils peuvent acquérir le langage alors que, c'est l'argument du déficit de stimuli, il est impossible qu'ils reçoivent dans leur enfance assez d'informations venant de l'environnement linguistique pour structurer toutes les connaissances grammaticales et lexicales qu'ils acquierent pourtant en quelques mois. Donc, le behaviorisme radical de Skinner est insoutenable.

Si contourner la difficulté de l'opération a permis au behaviorisme, et peut encore permettre aujourd'hui, d'étudier des comportements proches des réflexes sans rentrer dans l'activité mentale de l'organisme, de l'individu, il n'apparaît pas, ayant ainsi condamné tout accès à la complexité interne, être en mesure d'apporter un éclairage suffisant pour aborder les phénomènes complexes. Et, si le langage en est un, le comportement moral ne peut manquer d'en être un autre.

5.3.2 L'opérationnalisme de Bridgman

5.3.2.1 Définir les entités théoriques à partir des pratiques expérimentales

La seconde stratégie de contournement que j'envisage ici ne consiste pas, comme le behaviorisme, à nier la question de l'opérationnalisation en psychologie en éliminant les états mentaux du champ des recherches possibles mais, plus largement, et pas dans le seul domaine psychologique, à proposer d'inverser le raisonnement. Il s'agit de la stratégie de l'opérationnalisme qui consiste à partir des pratiques de l'expérimentation pour définir les entités théoriques par, et seulement par, ces opérations pratiques. Pour l'opérationnalisme, le thermomètre ne mesure pas la température, le thermomètre définit ce qu'est la température.

L'opérationnalisme est une théorie développée au début du vingtième siècle sous l'impulsion de Percy William Bridgman [23], qui consiste à définir les concepts physiques par les opérations qui permettent de manipuler et de mesurer [24] les phénomènes associés à ces concepts. Bien qu'aujourd'hui abandonnée sous cette forme extrême pour les concepts fondamentaux de la physique, la théorie opérationnaliste garde une certaine utilité sous la forme modérée de la recherche de l'explicitation de l'opérationnalisation qui permet de construire un chemin entre d'un côté les concepts et les relations théoriques et de l'autre côté la façon pratique d'en envisager l'observation ou l'expérimentation [25].

L'importance de l'opérationnalisation pour les théories physiques est apparue très crûment quand, au début du 20e siècle, tous les concepts de la physique du 19e siècle ont semblé devoir être remis en cause par la relativité générale d'un côté et par la physique quantique de l'autre. Aucune des procédures de mesure du temps et de l'espace utilisées en routine par les scientifiques ne résistait à ces nouvelles théories et, en conséquence, tous les éléments de preuves empiriques appuyés sur ces mesures chancelaient. Cette profonde déstabilisation a conduit Percy William Bridgman à proposer en 1927 une inversion de la façon de penser le lien entre les concepts de la physique théorique et leur mise en œuvre (Bridgman 1927) [31]. L'opérationnalisme [26] qu'il promeut consiste à ne plus voir l'opérationnalisation comme le résultat de la recherche pour rendre mesurable une grandeur physique prédéfinie théoriquement mais, inversement, à considérer les concepts physiques comme étant entièrement

23. La théorie de l'opérationnalisme appliquée à la physique est exposée dans (Bridgman 1927) [31]

24. J'entends ici par manipuler de façon générique l'ensemble des protocoles pratiques qui permettent de détecter, modifier, créer, supprimer, et également évaluer, comparer, mesurer, tout ou partie d'un phénomène ou d'une caractéristique du phénomène

25. Cette filiation est particulièrement soulignée dans (Feest 2005) [82] qui insiste sur l'importance de l'opérationnalisation en psychologie

26. On trouve utilisés dans la littérature les termes d'opérationnalisme et d'opérationnisme, aucun des deux ne figure dans l'ouvrage de Bridgman qui adopte l'expression « operational thinking ». Je retiens ici le premier car il est plus proche de l'opérationnalisation en tant que démarche pour rechercher les opérations pertinentes en regard des entités d'une théorie.

définis par la façon de les mettre en pratique. Citons (Bridgman 1927, page 5 de l'édition de 1960)[27] :

> En général, la signification d'un concept n'est rien de plus qu'un ensemble d'opérations. Le concept est synonyme de cet ensemble d'opérations.

Bridgman propose donc qu'à chaque fois qu'un physicien emploie un terme comme « température » ou « longueur », il doive être en capacité de dire comment le concept dont il parle va être opérationnalisé, selon quels protocoles pratiques il va être mis en œuvre, observé, manipulé, mesuré. A défaut d'une telle définition par les opérations, le concept doit simplement être écarté comme non utilisable par le physicien.

Bridgman mesure parfaitement l'immense impact qu'aurait sa proposition sur le quotidien des chercheurs (Bridgman 1927 p 32 de l'édition de 1960)[28] :

> Penser opérationnellement s'avérera au début une vertu bien asociale. On sera incapable de comprendre la plus simple conversation de ses amis et on se rendra universellement impopulaire en demandant la signification du plus simple des concepts de chaque argument.

Mais, c'est le prix à payer, pense Bridgman, pour ne plus employer que des termes définis par leur opérationnalisation, au plus près des phénomènes réels du laboratoire, et non par des théories qui, comme le montre la physique de son temps, ne résisteront pas à l'arrivée de nouvelles idées. Bien que l'opérationnalisme de Bridgman ne soit plus retenu aujourd'hui comme pertinent pour les sciences physiques, et plusieurs arguments que je vais détailler justifient cet abandon, il reste néanmoins une source importante d'inspiration lorsque la faiblesse des théories pousse à donner le prima à l'opérationnalisation pour définir les concepts. Et cela pourrait être le cas pour la psychologie, comme Feest en débat dans (Feest 2005) [82], mais avant d'approfondir ce point, il convient de détailler les critiques à l'opérationnalisme de Bridgman.

5.3.2.2 Trois arguments contre l'opérationnalisme

L'opérationnalisme de Bridgman a été critiqué dès la sortie de l'ouvrage et je retiendrai ici trois lignes d'arguments, l'inflation du nombre de concepts, le problème du continu, et le problème de la sous-détermination des termes théoriques par les expériences. Premier point, si on suit Bridgman et qu'on définit tout concept par son opérationnalisation, on se trouve

27. Le texte anglais est « In general, we mean by any concept nothing more than a set of operations ; the concept is synonymous with the corresponding set of operations », traduction de l'auteur.
28. Traduction de l'auteur.

confronté à une multiplication des concepts lorsque plusieurs opérationnalisations sont possibles. Par exemple si je peux mesurer une longueur avec un mètre, avec un laser, ou avec un sonar, comment puis-je savoir qu'il s'agit bien de la même grandeur ? Cela semble hors de portée de l'opérationnalisme qui devra admettre qu'il s'agit de trois concepts différents, appuyés sur trois protocoles différents, ce qui n'est guère satisfaisant et ne correspond nullement à la réalité des pratiques des scientifiques.

Second point, lorsqu'on mesure une longueur, on en donne une valeur dans l'ensemble des nombres réels R. Mais celui-ci est continu, et il sera impossible en pratique de faire correspondre à chaque nombre réel un protocole de mesure qui donne précisément une valeur dans R. Au mieux, un protocole de mesure donnera un intervalle dépendant de la précision des instruments utilisés. Il semble donc que l'opérationnalisme seul ne puisse expliquer comment on passe de la longueur précise théorique supposée prise dans un continuum à l'ensemble des nombres réels également continu par l'intermédiaire d'une opération qui, elle, ne peut être ponctuelle. Là encore, cette limitation ne correspond pas à la pratique des scientifiques qui s'appuient sur les mathématiques et en particulier sur R (et pas seulement sur des ensembles plus ou moins bien définis d'intervalles) pour construire leurs théories.

Enfin, troisième point, reprenant la thèse classique de Duhem Quine, l'ensemble des expérimentations sous-détermine les théories et les concepts : de multiples ensembles de concepts théoriques différents sont compatibles avec tout ensemble d'expérimentation. Si le scientifique n'acceptait pour seul point de départ que le résultat de ces expérimentations pour définir ses concepts, la physique n'aurait aucune possibilité de se développer car il lui manquerait les critères complémentaires lui permettant de choisir quelle théorie privilégier parmi toutes les théories opérationnellement compatibles.

Bridgman a reconnu, dans un article publié 30 ans plus tard, (Bridgman 1959) [32] que sa proposition devrait être amendée en prenant en compte ces critiques. Le principal changement concerne, pour le Bridgman de 1959, le fait que l'opérationnalisme semblait proposer une méthode définitive qui construirait des concepts à l'abri de tout changement théorique ultérieur. Ceci n'est pas une hypothèse plausible, à la fois parce que les théories évoluent de façon trop importante et, surtout, parce que l'esprit humain est capable de changements imprévisibles qui ouvriront de multiples possibilités au développement des méthodes scientifiques. Second changement, l'opérationnalisme semble entièrement tourné vers les aspects matériels de la descriptions des protocoles expérimentaux et il faudrait pondérer cet aspect par l'importance de l'analyse conceptuelle dans l'établissement des concepts utiles à la science, pondération qui permettrait en particulier de répondre aux trois critiques mention-

nées ci-dessus portant sur l'inflation du nombre de concepts, le problème du continu et de la sous-détermination.

5.3.2.3 L'opérationnalisme dans les sciences aujourd'hui

L'opérationnalisme n'est plus une théorie retenue aujourd'hui comme intéressante en épistémologie de la physique et on peut analyser cette théorie soit comme trivialement juste, mais sans intérêt, soit comme profondément inexacte pour ce qui concerne les sciences physiques, cadre dans lequel cette théorie est née. En un sens, elle est trivialement juste si on l'interprète comme affirmant que la définition des concepts scientifiques doit prendre en compte les résultats des expérimentations mais ne s'interdit pas d'autres considérations. En ce sens elle ne constitue qu'une reformulation assez lâche de la définition des sciences naturelles expérimentales et n'apporte pas d'éclairage permettant d'affiner l'épistémologie ni d'un point de vue descriptif, comment travaillent concrètement les équipes de scientifiques, ni d'un point de vue normatif, comment ils devraient travailler pour faire de la « bonne » science (à supposer que l'on sache la définir).

En un autre sens, elle est profondément inexacte, si on l'interprète comme affirmant que la définition des concepts scientifiques est exclusivement définie par les opérations pratiques de leur mise en œuvre expérimentale. Comme Bridgman lui-même l'a reconnu en 1959, cette vision n'est bien sûr pas descriptive, puisqu'elle a été bâtie en réaction à la réalité de la physique du début du 20^{e} siècle, mais elle n'est pas non plus une proposition normative considérée aujourd'hui comme plausible pour les trois types de raisons mentionnées plus haut et, surtout, parce qu'il nous est aujourd'hui habituel de considérer que tout protocole de mesure ou, plus largement, tout protocole expérimental, repose lui-même sur l'acceptation de théories qui sont embarquées dans tous les appareils qui peuplent les laboratoires. Il est totalement illusoire d'imaginer décrire un protocole avec des termes dont aucun ne serait chargé préalablement des théories qui ont permis de mettre au point tous les outils utilisés.

On trouvera une vision plus récente et plus proche du travail des scientifiques dans les propositions de Hasok Chang illustrées par l'exemple de l'invention du concept de température dans (Chang2007) [42]. Sans rentrer dans le détail de cette conception, je renvoie à la métaphore de l'hélice scientifique expérimentale présentée plus haut pour en illustrer l'idée principale : en partant d'une conception existante, on affine progressivement les théories (i.e. les entités théoriques, les propriétés, et les relations postulées), leur opérationnalisation dans des dispositifs expérimentaux et les traces laissées par les expérimentations réalisées. Il n'y a pas de fondement absolu à la démarche mais un processus de collaboration entre

équipes de scientifiques permettant qu'à chaque itération épistémique (expression proposée par Hasok Chang), on s'attache à répondre aux questions et incertitudes ressenties par les scientifiques et on s'attache également à en proposer de nouvelles, qu'on espère plus approfondies. Dans cette conception itérative, il est pareillement illusoire de rechercher un point de départ ferme dans la théorie ou de le rechercher, comme le fait Bridgman, dans les protocoles expérimentaux, c'est dans l'incessant va et vient entre ces activités que se construit l'itération épistémique.

Comme le souligne Chang dans son article sur l'opérationnalisme dans la Standford Encyclopedia of Philosophy (Chang 2019) [43], il convient également de conserver les héritages importants de la pensée de Bridgman et en particulier l'attention apportée à la difficulté d'étendre l'utilisation d'un concept au-delà des domaines de validité des procédures pratiques de sa mise en œuvre. L'exemple de la température illustre particulièrement ce point : si on la définit comme ce qu'indique le thermomètre à mercure, que devient le concept de température quand on atteint les points de fusion du verre et d'évaporation du mercure ? Il faudra de nouvelles procédures pour donner un sens opérationnel au concept de température dans ces contextes qu'on atteint, par exemple, dans les fours à porcelaine. Autre point important, Bridgman marque aussi la nécessité de maîtriser l'opérationnalisation de chaque concept particulier, chacun avec ses difficultés de mise en œuvre, et le scientifique ne peut se contenter, en tant que physicien, d'une validation holiste des théories comme celle suggérée par Quine, en tant que philosophe, qui met en regard un ensemble de résultats expérimentaux et un ensemble de concepts et de relations théoriques. Enfin, dernière remarque faite par Hasok Chang pour souligner l'utilité du mode de pensée de Bridgman, l'opérationnalisation d'un concept fait régulièrement surgir des difficultés et celles-ci peuvent être révélatrices d'une plus grande complexité du monde réel que ce que la théorie prenait initialement en compte. En ce sens, l'opérationnalisation est un utile révélateur de complexité.

Bien qu'aujourd'hui pratiquement abandonné dans le cadre de l'épistémologie des sciences physiques, l'opérationnalisme reste une possibilité importante pour la définition théorique des concepts et tout particulièrement en psychologie. Cette possibilité est en particulier reliée historiquement au behaviorisme et au développement des mesures psychophysiques. Les behavioristes, souhaitant définir les objets mentaux uniquement par les comportements dont ils sont causes ou conséquences ont proposé de les définir exclusivement à partir de ces comportements observés. L'analogie entre le behaviorisme et l'opérationnalisme de Bridgman est à la fois de fond, le principe d'une définition par les pratiques observables, et également d'époque, puisque les deux idées ont été développées dans les années 1920.

Le développement des mesures psychophysiques entrepris en particulier par Stanley Smith Stevens en acoustique dans les années 1930 s'appuie sur la même approche : développer des concepts liés à la perception acoustique par la seule considération des effets et causes physiques observables par un tiers. Ainsi le niveau sonore perçu par un humain est entièrement défini par un protocole de mesure conduisant au calcul du décibel A qui embarque en lui à la fois des données physiques, l'intensité de l'énergie sonore par bande de fréquence, et les réactions physiologiques d'un individu moyen, la pondération de la sensibilité de l'oreille humaine par bande de fréquence et logarithmique en intensité. Cette définition correspond aux exigences de Bridgman, elle est exhaustivement [29] définie par un protocole expérimental, et elle correspond aussi aux exigences du behaviorisme, elle ne fait pas appel à des états mentaux intracrâniens.

Les axes de recherche de la psychologie expérimentale ne sont plus aujourd'hui limités à des théories ne faisant pas appel à des processus intracrâniens, et les sciences naturelles ont abandonné l'opérationnalisme, mais il ne serait pas difficile de voir une réminiscence de ces deux mouvements dans les prises de position qui visent à définir des propriétés mentales par la façon de les observer, par leur opérationnalisation. Les débats sur le statut des corrélats neuronaux de tel ou tel phénomène mental seraient une piste probablement féconde dans cette direction, mais je n'irai pas plus loin ici sur ces débats pour revenir à mon cœur de propos. Face à la difficulté qu'il y a à opérationnaliser les entités théoriques postulées par les psychologues, le technicien, pour contourner le problème, peut simplement en faire abstraction et mener son étude « comme si » le problème n'existait pas en supposant cette entité suffisamment prise en compte par des objets techniques qu'il sait manipuler. Je vais présenter un exemple de ce processus dans la section suivante.

5.3.3 Le contournement par la technique

5.3.3.1 Un exemple technique : le confort acoustique

Je présente dans cette section une étude qui contourne la question de l'opérationnalisation et, surtout, me permettra de souligner la richesse de la dynamique qui s'instaure entre les domaines théoriques et expérimentaux dans la démarche scientifique expérimentale. L'exemple est issu de mon expérience personnelle et porte sur l'opérationnalisation de la notion de confort acoustique dans les bâtiments. La recherche, très appliquée, menée au

29. Naturellement, il faudrait réintroduire ici toutes les théories embarquées dans les outils de mesure tels les sonomètres, mais je laisse ce point de côté dans cette argumentation en supposant que le lien est faible entre la physique des sonomètres et la physiologie de l'oreille humaine.

sein du Centre Scientifique et Technique du Bâtiment (CSTB) répond à une demande sociale d'amélioration de l'habitat. L'exemple met en avant la dynamique de la relation entre les notions théoriques et leur opérationnalisation qui modifie, autant qu'elle est modifiée par, les évolutions des hypothèses théoriques et des savoir-faire expérimentaux et permet d'affirmer la fécondité de l'autonomie des équipes d'expérimentateurs par rapport aux équipes de théoriciens.

Je me propose donc d'explorer les développements liés à la notion de « confort acoustique » au tournant des années 1970. A cette époque (1976-1979) je travaillais au Centre Scientifique et Technique du Bâtiment (CSTB) dans la division acoustique. Après l'accent mis sur l'industrialisation du bâtiment, nécessitée par l'effort national de reconstruction des années 1960, c'est à la prise en compte des aspects qualitatifs qu'étaient dédiées les recherches du CSTB à la fin de années 1970 et, entre autres, à la qualité acoustique souvent perçue comme insuffisante par les habitants des immeubles trop rapidement construits. La notion de confort acoustique était présente dans les débats publics et privés et se posait alors au CSTB la question d'établir des règles et des normes de qualité pour améliorer la construction. A l'idéal, ces règles devaient permettre d'anticiper la qualité acoustique sur la base des plans détaillés du bâtiment, d'améliorer la conception des matériaux et les procédés de construction et devaient être, du fait de la compétence normative du CSTB, opposables à des tiers en cas de conflit.

Hypothèse de départ : identité du confort acoustique et du niveau sonore

La notion floue de sentiment d'inconfort acoustique ne permettant pas de définir des seuils précis de qualité, la première étape retenue par les ingénieurs du CSTB a été de se rapprocher des connaissances scientifiques disponibles et d'analyser les processus vibratoires en cause dans le phénomène du bruit. Il s'est agi de :

— Caractériser la source du son, l'émission initiale d'une vibration,

— Décrire les processus de transmission vers un local : réflexions, absorptions, rayonnements, amortissements, transmissions, ...

— Décrire les conditions de réception et de perception du bruit.

Cette approche revenait à privilégier ce qui était déjà connu, à savoir la caractérisation et la transmission des vibrations, et à assimiler le confort acoustique ressenti par un habitant au simple niveau sonore qu'il perçoit dans son logement. C'est ce que je qualifie d'hypothèse triviale : l'identité entre le confort acoustique et le niveau sonore, ce qui permet de contourner les difficultés liées à l'opérationnalisation de l'entité théorique « confort acoustique ressenti ».

L'acousticien disposait d'une panoplie de savoir-faire théoriques et expérimentaux pour débuter ses études sur la transmission des bruits. Sous l'angle théorique, les vibrations sont

bien comprises et ont été largement étudiées pour répondre aux problèmes de stabilité des ouvrages [30]. Les calculateurs analogiques permettent de calculer les transformations de Fourier, et surtout, dès la fin des années 70, les premiers calculateurs numériques ont permis d'entrevoir le développement exponentiel des possibilités de simulation et d'imaginer la faisabilité d'un calcul en contexte réel de la transmission sonore dans un ensemble complexe comme un bâtiment d'habitation. Sous l'angle expérimental, il existe également des travaux disponibles mais beaucoup de choses restent à mettre au point. Deux exemples : il n'existe dans les années 1970 qu'une seule marque de sonomètres de qualité [31] aux capacités simplistes et les différents capteurs, adaptés aux basses fréquences qui sont étudiées pour la stabilité des ouvrages, ne le sont pas aux hautes fréquences nécessaires aux études acoustiques [32].

Franchir le fossé entre ces balbutiements et la réalisation d'un logiciel opérationnel permettant de prévoir sur plans la transmission entre les locaux attenants dans un bâtiment [33] demandera à peu près vingt ans, temps nécessaire pour maîtriser peu à peu l'aspect physique de la transmission du son. L'ampleur du fossé peut être évaluée à l'aide de deux problèmes, un dans le domaine de la théorie physique, l'autre dans le domaine des mesures d'accélération.

De la stabilité à l'acoustique, théories et expérimentations sont remises en cause

La théorie qui prime aux basses fréquences pour l'étude de la stabilité des ouvrages est la résonance. On analyse les modes propres de vibration des structures excitantes et excitées. Si les fréquences sont « assez proches », la transmission est considérée infinie (en première approche), si elles sont « assez éloignées » la transmission est considérée comme négligeable. Il n'est pas nécessaire aux basses fréquences de préciser la signification du terme « assez » car il y a peu de modes de vibration par tranche de fréquences et l'objectif de stabilité exige de rester très loin du phénomène de résonance. En revanche, aux fréquences utiles aux études acoustiques, le nombre de modes propres de vibration par bande de fréquences est de plus en plus grand et la transmission se trouverait toujours dans le cas « infini », ce qui n'a évidement pas de sens physique. La réponse théorique à ce problème a été de changer le mode d'approche pour une théorie dite Statistical Energy Analysis (SEA) qui consiste à comprendre la transmission d'énergie sonore comme dépendant à la fois de mécanismes de couplage mécanique indépendants de la fréquence et de la densité de modes de vibration par tranches de fréquences [34]. Cette approche éloigne du calcul modal analytique classique utilisé en méca-

30. Secousses sismiques et stabilité au vent par exemple.
31. Bruel et Kjaer dit BK qui existe encore à ce jour.
32. La stabilité des ouvrages s'analyse au-dessous de 30 Hz alors que l'essentiel de l'audible est dans la tranche de 300Hz à 3000 Hz et la sensibilité de l'oreille humaine peut aller de 50 Hz à 20 000Hz.
33. Voir par exemple le logiciel AcouBAT du CSTB commercialisé à partir de fin 90, et en filiation directe des études décrites ici.
34. A. Chaumette. Transmission des vibrations à la jonction de deux plaques rectangulaires. Technical Report

nique et donne un poids important à la mesure in situ car les facteurs de couplage ne peuvent être que rarement calculés et doivent en général être directement déduits des mesures pour chaque type d'assemblage utilisé dans la structure étudiée.

Seconde difficulté : pour la mesure des accélérations, les capteurs disponibles avaient un mode de fixation sur la structure qui était adapté aux basses fréquences. Leur utilisation a conduit rapidement à des problèmes importants lorsque l'on augmentait la fréquence. On mesurait alors principalement la réponse du système de fixation, non pertinente. Il a donc été nécessaire de mettre au point de nouveaux modes de fixation utilisant toutes les ressources des nouveaux produits adhésifs de très faible épaisseur. Par ailleurs, l'électronique de la fin des années 1970 n'était pas assez rapide pour analyser le signal délivré, et il a fallu attendre qu'elle progresse pour que puissent être traitées en temps réel les fréquences aiguës dépassant les 10 000 Hz. Ce n'est qu'avec ces améliorations venant de la chimie et de l'électronique que les questions expérimentales ont pu progresser.

Le confort acoustique ne se réduit pas au physique

Outre ces difficultés relatives à la physique du phénomène, et dès le début de cette période, il apparaissait clairement que les calculs de transmission de l'énergie vibratoire ne seraient pas suffisants pour traiter tous les aspects de la notion de « confort acoustique » ni, plus généralement, de l'acceptation du bruit par les habitants [35]. D'abord, la transmission d'énergie n'est pas ce que perçoit l'oreille humaine. L'oreille est sensible à l'énergie de façon logarithmique et non linéaire (d'où la mesure en décibel dB qui est une mesure de niveau de bruit perçu). De plus, la perception du niveau sonore est fonction de la fréquence, ce qui a conduit dès les années 1930 à proposer de pondérer l'énergie en fonction de la sensibilité de l'oreille humaine, pondération concrétisée par la normalisation du décibel A. Cette procédure a permis d'opérationnaliser la perception du niveau sonore, ouvrant la voie à une mesure objective [36].

Remarquons ici que l'opérationnalisation du niveau de bruit perçu par l'oreille humaine, pour permettre une mesure fiable et régulière, implique de remplacer la variété des sensibilités personnelles par une courbe unique moyenne prise pour standard de pondération, en négligeant les écarts entre individus. Insistons sur cette standardisation, qui est nécessaire à l'opérationnalisation. Si on veut mener des expérimentations répétées sur un phénomène qui présente, comme ici la perception par l'oreille humaine, une certaine variabilité individuelle,

2.79.009, Centre Scientifique et Technique du Bâtiment, 1977

35. La psychoacoustique existe depuis le 18e siècle et avait bien avant 1970 établi la complexité de la perception du bruit par l'oreille humaine.

36. Le développement de la distinction entre mesures physiques et perception en acoustique est utilisée dans (Feest 2005) [82] comme un des exemples précurseurs de l'opérationnalisation des concepts psychologiques.

et dégager des conclusions indépendantes de la personne, il faut substituer à l'auditeur réel une fiction, un auditeur moyen, qui sera l'étalon des instruments de mesure. Remarquons que cela conduit alors à séparer d'un côté les études impliquant cet auditeur moyen et de l'autre l'étude de la variabilité des sujets individuels. Cette remarque me permet de souligner que l'opérationnalisation est bien à considérer dans le cadre d'une relation particulière étudiée et non indépendamment de la question posée. On trouve un parallèle à cette situation dans le domaine de la psychologie avec la psychologie systématique qui s'intéresse au comportement normal moyen dans une population et la psychologie différentielle qui s'intéresse aux différences de comportement entre les individus. Il conviendrait certainement d'approfondir la question des différences d'opérationnalisation qui seront nécessaires selon que l'on s'intéresse à l'un ou à l'autre de ces deux aspects.

Deuxièmement, l'homme est sensible à la phase du son [37], ce qui lui permet de percevoir la direction de provenance du son grâce à la différence de phase perçue entre les deux oreilles et, par ailleurs, d'entendre la différence entre un son pur (une vibration sinusoïdale idéale) et un bruit de même fréquence et de même énergie mais dont la phase est hachée.

Il s'agit, dans la recherche qui m'intéresse ici, de donner sa juste place à la complexité de l'oreille humaine tout en se concentrant sur la transmission d'énergie, préoccupation indispensable lorsqu'on s'intéresse au confort acoustique. Le niveau de bruit est bien sûr le premier paramètre perçu par les habitants d'un immeuble et le réduire par une bonne isolation est un objectif qui ne peut être remis en cause par les analyses psychophysiques plus élaborées. Que ce soit entre locaux adjacents ou avec l'extérieur, améliorer l'isolation et, donc, diminuer la transmission d'énergie acoustique, est la première chose qui puisse être réalisée pendant la construction d'un bâtiment. Elle en est pour partie une caractéristique irréversible, ou, a minima, difficilement améliorable après coup. Mais ce caractère structurel ne doit pas être compris comme le point final de la réflexion car, nous allons le voir, les habitants ne perçoivent pas seulement ce qui fait l'objet des mesures physiques.

Trois points vont me permettre d'illustrer les différences de perception du confort acoustique en regard de ce que le physicien mesure. Les deux premiers sont connus et font partie des premiers cours d'acoustique, je ne ferai que les exposer, le troisième exploite une mission menée à Grenoble en 1977 et je détaillerai sommairement les types de mesures faites.

Premièrement, la perception d'inconfort acoustique ressentie peut être très inférieure à ce que la mesure physique fait apparaître. L'exemple des riverains des voies ferrées en est

37. Un son est caractérisé par sa fréquence, grave ou aiguë, son amplitude, forte ou faible, et, comme pour tout phénomène périodique, par sa phase qui indique à quel instant l'amplitude est maximale.

un exemple frappant. Alors que les niveaux sonores atteints sont importants, les riverains disent « ne plus entendre le bruit » des trains et les analyses montrent que ce n'est pas simplement une façon de parler, une justification commode apportée au fait qu'ils n'aient pas eu le courage de déménager : leur comportement montre qu'ils n'entendent effectivement plus ce bruit régulier. Le bruit des trains régulièrement répété est sans connotation agressive et, pour les riverains, rassurant. Simple confirmation de l'état habituel du monde, il est effacé des perceptions à traiter par le système cognitif. C'est même l'absence de ce bruit qui devient une information importante, comme le montre, inversement, l'inquiétude que les riverains ressentent les jours de grève où le silence se fait.

Deuxièmement, la perception peut être sans aucun lien avec ce que la mesure physique peut faire apparaître. L'exemple du bruit blanc, un bruit dont l'énergie est équivalente dans toutes les bandes de fréquences, est bien connu. Il peut être par exemple produit par les vagues un jour de tempête en bord de mer ou par une section d'autoroute en rase campagne. Ces deux bruits sont indiscernables l'un de l'autre par les mesures physiques. Ils sont également indiscernables par l'oreille humaine sur la base d'un enregistrement. En revanche, si l'auditeur identifie l'origine du bruit, alors son jugement sur l'acceptabilité du bruit se transforme, positivement pour le bruit de la mer et négativement pour celui de l'autoroute. Il peut être tentant d'analyser dans ce cas le jugement comme une simple conséquence de la connaissance de la proximité de la mer ou d'une autoroute, sans lien de détermination par le bruit lui-même mais ce serait penser que l'identification de l'origine du bruit n'est pas possible en tant que tel, qu'elle est impossible si l'auditeur n'apprend pas par ailleurs qu'il est proche de la mer ou d'une autoroute, et ce serait oublier que la richesse de la perception directe du bruit n'est pas rendue de façon complète par les différentes mesures physiques possibles ni par l'enregistrement.

Enfin, troisièmement, la perception peut être d'une nature toute différente de ce que la mesure physique fait apparaître. C'est ce que l'on peut retenir d'une mission d'expertise menée à Grenoble par une équipe du CSTB, dont l'auteur, dans un ensemble de logements neufs construits sur les flancs du Vercors dans une zone particulièrement isolée. Les habitants jugeaient très mauvaise la qualité acoustique de leurs logements neufs. Ils ne pouvaient dormir la nuit du fait des bruits de voisinage et ils avaient demandé une expertise de façon à mettre en cause le constructeur. Le CSTB avait accepté la mission qui s'insérait dans les recherches en cours sur le confort acoustique et avait dépêché l'équipe au grand complet pour bénéficier de ce cas d'étude réel.

La mission a consisté à mesurer l'isolation aux bruits aériens internes (des hauts par-

leurs normalisés sont installés dans plusieurs pièces et des mesures faites au sonomètre dans toutes les pièces du bâtiment), aux bruits d'impact (un bâti comportant des marteaux normalisés chutant d'une hauteur constante et animés par un arbre à came reproduit à peu près l'énergie d'une personne qui marche avec des talons aiguilles), et à tester tous les défauts connus (ruptures d'isolation par les aérations, sensibilité au bruit extérieur, bruits des équipements techniques du bâtiment ...). A l'issue de cet ensemble de mesures anormalement exhaustif et précis pour une expertise courante, le verdict était sans appel, l'ensemble immobilier était d'excellente qualité acoustique, les dalles flottantes parfaitement opérationnelles et aucun défaut structurel ne justifiait l'insatisfaction des utilisateurs. En revanche, apparaissait également que lorsque les sources de bruit étaient arrêtées, les sonomètres enregistraient un silence profond, sans aucun bruit de fond. Nous avons alors demandé aux habitants de nous préciser d'où ils venaient et de préciser leurs griefs en terme de bruits de voisinage. Cela a permis de constater que beaucoup venaient du centre de Grenoble, de vieux immeubles mal isolés. La conclusion principale s'imposait alors : le niveau de bruit de fond étant quasi nul, tout signal même faible est intelligible et perçu comme une agression, une rupture de l'intimité.

La mission a permis d'obtenir des résultats sur le plan pratique et sur le plan du programme de recherche. Sur le plan pratique et pour les habitants concernés, les propositions d'amélioration ont été de plusieurs types, d'abord de laisser du temps à l'habituation, de façon à apprivoiser le silence, ensuite de recréer du bruit de fond dans chaque local, une musique de fond pour ne plus entendre les petits bruits de voisinage, et enfin, nous avons proposé de changer certaines bouches d'aération au profit de matériels diminuant l'intelligibilité du son. Sur le plan du programme de recherche, ce cas a conforté l'équipe dans son projet de procéder à des types de recherche complémentaires sur la génération et la transmission du signal acoustique, indépendamment de sa signification, et sur la perception de la signification par les habitants. Ce cas soulignait la nécessité de collaborations entre équipes de physiciens et de psychologues du bâtiment.

Les études menées sur des cas analogues ont permis de montrer que l'analyse psychologique nécessite d'aller beaucoup plus loin que la simple mesure d'énergie par tranche de fréquence pour caractériser physiquement la source du bruit. Il faut, par exemple, ajouter la provenance du bruit interne ou externe au bâtiment, l'origine technique (ascenseurs, ventilations,...) ou non technique (pas, parole, musique, ...) du bruit, et de multiples autres aspects spontanément décrits par les habitants pour marquer leur rejet, ou au contraire leur acceptation, d'un bruit comme acceptable. Ces études ont également montré l'importance de

la nature des transformations du signal sonore dont certaines maintiennent l'intelligibilité de la source malgré un affaiblissement important. De façon symétrique, il s'agit, pour bien préserver l'intimité d'un logement, de faire l'inverse de la téléphonie. Pour le téléphone il faut transmettre la signification avec le minimum d'énergie, pour le bâtiment il faut bloquer la signification sans s'obliger à une diminution trop coûteuse de l'énergie transmise.

5.3.3.2 L'opérationnalisation progresse en même temps que les savoir-faire théoriques et expérimentaux.

Quels enseignements tirer de cet exemple de contournement de l'opérationnalisation d'une notion psychologique complexe, ici le « confort acoustique »? Plusieurs constats s'imposent qui, principalement, soulignent la complexité de l'opérationnalisation d'une telle notion en situation réelle de recherche.

Une première tentative d'intégration de la perception humaine au sein de l'environnement expérimental et théorique de la physique des vibrations a débouché sur la création des dB A et la normalisation des situations de bruit, mais s'est heurtée à de multiples difficultés liées aux hautes fréquences. Ces difficultés ont permis de définir le travail restant à faire par les diverses équipes concernées tant sur le plan théorique que sur le plan expérimental : il devenait nécessaire pour les expérimentateurs de fiabiliser les mesures aux hautes fréquences, pour les théoriciens des vibrations de sortir du calcul modal, pour les modélisateurs de développer des outils assez rapides pour traiter des hautes fréquences et enfin, pour les psychologues, de comprendre comment les habitants d'un immeuble jugent de l'acceptabilité d'un bruit.

Une fois ces travaux lancés, ils ont avancé de façon autonome et ont pu bénéficier, sur chacun des plans, de nouvelles possibilités non directement reliées à la problématique initiale, avec les développements de la chimie des colles, de l'informatique et de la psychologie du bâtiment, en plein développement à cette époque au CSTB.

L'opérationnalisation, telle que l'exemple du confort acoustique la montre, est un processus de longue durée, itératif et peut-être sans fin, qui impose une articulation entre différents domaines de savoir-faire théoriques et expérimentaux, entre plusieurs équipes de chercheurs et, ce faisant, à la fois les remet en cause en vue de leur bonne adaptation au nouveau champ d'étude et bénéficie en retour des avancées largement autonomes de chacun de ces champs. Contourner l'entité complexe du « confort acoustique ressenti » a permis d'avancer ces recherches et de préparer, au moins peut-on l'espérer, de futures recherches la prenant en compte de façon plus réaliste que sa simple assimilation au niveau sonore.

5.3.4 Le contournement entre utilité et risque de stérilisation

Face à la difficulté d'opérationnaliser une entité théorique, dans le cas général et a fortiori dans le cas de la psychologie, chacune des trois stratégies de contournement que je viens d'évoquer, le behaviorisme, l'opérationnalisme de Bridgman et la technique, a présenté des avantages et des limites.

Au titre des avantages, il convient de souligner qu'il peut, à un instant donné, ne pas exister d'alternative au contournement, si l'on désire poursuivre la recherche, et que cette poursuite peut conduire, ultérieurement, à de nouvelles pistes de recherches qui lèveront la difficulté. Illustrons ce point avec l'exemple des conjectures mathématiques. Lorsque le mathématicien n'arrive pas à démontrer un théorème, il peut le proposer en tant que conjecture et développer ensuite toutes les conséquences déductibles de cette conjecture. Il est possible, mais non certain, que parmi ces conséquences apparaisse une piste d'analyse qui permettra de démontrer (ou d'invalider) la conjecture et, ainsi, permettra de la rejeter ou de la transformer en théorème.

Sur un plan différent, il en va de même avec l'exemple de l'acoustique. Bien que la cible finale soit la satisfaction des habitants, et donc la prise en compte de leur ressenti du confort acoustique, contourner la difficulté de mesurer ce ressenti en le remplaçant par des mesures physiques a permis d'avancer à la fois sur les aspects purement physiques (hautes fréquences et simulations informatiques) et, à la fois, sur la compréhension du phénomène psychologique (importance de l'intelligibilité du son).

Enfin, et reprenant la métaphore de l'hélice expérimentale, on peut espérer que les itérations épistémiques bâties sur ce contournement permettront de mieux aborder la question de l'opérationnalisation difficile de l'entité théorique à l'itération suivante.

Au titre des limites de ces stratégies de contournement, on peut citer principalement le risque de l'ankylose. Si l'impossibilité d'accéder aux états mentaux par l'introspection devient un dogme incontournable qui, oublieux de ses origines, s'exprime brutalement en « les états mentaux n'existent pas pour les scientifiques », alors il y a un risque que soient simplement ignorées de nouvelles possibilités qui, le contournement ayant été fructueux par ailleurs, adviennent dans le paysage scientifique. Il est ainsi possible (mais toujours incertain) que les nouveaux outils dont disposent aujourd'hui les psychologues expérimentaux (IRMf, produits neuro-actifs, simulations informatiques,...) soient de nature à apporter des réponses qui déplacent les limites de ce qui peut être opérationnalisé en terme d'états mentaux. Il serait alors nuisible qu'une stratégie de contournement qui a pu être inévitable (et également fructueuse)

à un instant donné se transforme en une rigidité stérilisante.

Je vais dans la section suivante présenter l'opérationnalisation des entités théoriques utiles à la philosophe morale telle qu'elle est prise en compte par la psychologie aujourd'hui et, donc, telle qu'elle est prise en compte implicitement ou explicitement par les philosophes moraux expérimentaux.

5.4 Opérationnaliser les entités psychologiques

Certes, le risque de confusion ne touche pas que la psychologie, et, de même, l'opérationnalisation est un sujet important pour la démarche expérimentale dans tous les domaines de la science bien au-delà des seules sciences humaines. Mais, comme l'ont montré les sections précédentes, le risque de confusion a une prévalence importante dans les cinq études de cas que j'ai détaillées, en lien avec des opérationnalisations défaillantes. L'opérationnalisation est une étape difficile et à fort enjeu. Face à cette difficulté, les psychologues ont été tentés par des stratégies de contournement, comme le behaviorisme. Mais ces stratégies ont montré leurs limites et le problème reste entier : comment expérimenter utilement quand les notions complexes et multiformes de la psychologie humaine sont en jeu ?

Je vais maintenant examiner la définition de l'opérationnalisation telle qu'elle est donnée par les psychologues, à savoir la recherche d'un indicateur observable pour un objet d'étude qui ne l'est pas directement, puis je soulignerai le risque d'arbitraire qui menace le choix de l'indicateur, et les conséquences de cet arbitraire sur la méthode présentée par les psychologues, avec en particulier, la question des écarts entre différents indicateurs observables pour une même notion. J'insisterai sur l'importance de ce problème dans la présentation des méthodes en psychologie et sur les différentes stratégies pour affronter cette difficulté qu'elles proposent.

5.4.1 Indicateurs observables

Les psychologues définissent l'opérationnalisation comme la recherche d'indicateurs observables pour des entités psychologiques étudiées qui ne le sont pas. Entrons dans le détail, cette définition étant attrayante puisque plus simple que celle que j'ai proposée dans les sections précédentes. De façon générale, une hypothèse psychologique posera une relation de la forme « Dans le contexte C, si le stimuli S a lieu, alors les effets E auront lieu » [38]. Par exemple, dans un contexte social sans stress particulier, un enfant soumis à un apprentissage

38. Formulation adoptée par exemple dans le manuel (Ghiglione 2007) p300 [101]

de quelques heures par jour apprendra à compter, mais l'apprentissage sera moins efficace si l'enfant est stressé. Autre exemple : un individu confronté à une autorité qu'il considère légitime pourrait obéir à l'ordre d'accomplir une action qu'il répugnerait à faire hors de ce contexte, allant même jusqu'à mettre en péril la vie d'autrui [39]. L'objectif de l'opérationnalisation est de définir un dispositif expérimental ou observationnel comportant un contexte particulier Co et des événements, stimuli et effets, So et Eo dont l'occurrence soit révélée par des indicateurs observables. Il s'agit de définir comment l'expérimentateur pourra, dans le premier exemple, dire que l'enfant sait compter ou, dans le second, que le participant met en péril la vie d'autrui sans, bien sûr, que ce ne soit réellement le cas. L'interprétation de l'expérience mettra en relation (Co, So, Eo) avec les trois termes de la proposition théorique : le contexte C, les stimuli S et les effets E. Ainsi, dans l'exemple de l'apprentissage, on considérera que les enfants d'une classe supposée représentative, non stressés, soumis à un entraînement dont il faudra dire le contenu, obtiennent à une batterie de tests normalisés des notes supérieures à celles qu'ils auraient obtenues sans l'apprentissage, ou qu'un groupe témoin sans apprentissage a obtenu, et on comparera ce gain à celui obtenu avec des enfants stressés [40]. Préciser le protocole, c'est expliciter comment on a choisi la classe, mesuré le stress, réalisé l'apprentissage, mesuré le gain,... et tout ce qui permet, dans un premier temps, de passer de la relation étudiée par la théorie à un expérimentation et, dans un second temps, d'interpréter les résultats obtenus en explicitant, dans la mesure du possible, toutes les conditions réellement en place pendant l'expérimentation.

L'opérationnalisation réussie a deux dimensions. D'abord, elle permet de disposer d'indicateurs observables pour Eo et So qui peuvent confirmer ou infirmer l'hypothèse dans ce cas particulier. Ensuite, elle permet d'obtenir une acceptation de la validité de l'utilisation de (Co, So, Eo) en tant qu'interprétable dans la perspective de l'hypothèse (C,S,E) étudiée.

5.4.2 L'opérationnalisation en psychologie : une étape nécessaire mais délicate

5.4.2.1 L'opérationnalisation dans la méthode

L'opérationnalisation des notions psychologiques est un thème central de la méthode en psychologie et, par extension, de la philosophie expérimentale [41]. Citons le cours de psycholo-

39. Pour reprendre l'exemple classique de la psychologie du comportement des années 60 : Stanley Milgram, Behavioral study of obedience, Journal of Abnormal and Social Psychology, 1963, Vol. 67, pp. 371–378 [164]

40. Ce qui soulèvera bien sûr de difficiles questions éthiques. Pourquoi le groupe témoin est-il laissé sans apprentissage ? Comment l'expérimentateur fait-il pour avoir des enfants stressés et d'autres non ?

41. Voir (Cullen 2010) [59] pour une revue de tous les articles de philosophie expérimentale renvoyant vers la psychologie pour la dimension méthodologique.Voir également (Sytsma 2016 page 123) [224] pour une description

gie de Ghiglione et Richard [101][42] page 306 :

> Les opinions, les attitudes, les aptitudes, les raisonnements, les attentes, les représentations, les émotions, et on pourrait sans peine allonger la liste, sont des concepts dont la psychologie pourrait difficilement se passer ; pourtant on ne peut pas les observer de façon immédiate. On va alors tenter de les inférer à partir d'observables, verbaux ou non verbaux...

Pour le psychologue, l'opérationnalisation est principalement cette étape qui vise à trouver un indicateur observable qu'il pourra utiliser dans ses expériences et observations. L'indicateur est ou bien verbal, issu de l'expression explicite du participant, ou bien non verbal et déduit de propriétés observables. La dilatation de la pupille de l'œil est ainsi retenue comme un indicateur d'émotions, la rapidité de réponse comme un indicateur de compétence acquise, et ainsi de suite. L'opérationnalisation ouvre la possibilité d'obtenir des observations qui rendent manifeste sur un terrain expérimental particulier une occurrence d'un phénomène relié (par hypothèse de l'opérationnalisation) à une entité non directement observable de la psychologie. Littéralement, l'opérationnalisation bâtit les correspondances entre les entités que le psychologue souhaite étudier et des observables du terrain expérimental. Le psychologue instrumente le terrain expérimental de façon à obtenir des résultats qui seront interprétés comme donnant des indications sur l'hypothèse étudiée, au même titre que le géomètre instrumente un bâtiment en implantant des repères visuels faciles à mesurer dont il interprétera les mouvements comme résultant de la déformation du bâtiment lui-même.

Le choix du meilleur indicateur pour observer ou mesurer un trait psychologique constitue le côté pratique de l'étape d'opérationnalisation. Les psychologues doivent se mettre d'accord sur de nombreuses questions : quelle méthode de recueil adopter ? Quels indicateurs physico-chimiques retenir ? Quels seuils juger significatifs ?[43]... On peut concevoir, et on constate en pratique, que des choix différents mènent à mettre en œuvre des méthodes de travail différentes conduisant à des résultats qui ne coïncideront pas totalement. Le psychologue dispose alors de plusieurs mesures ce qui implique la présence d'écarts qu'il aura à expliquer pour pouvoir exploiter les résultats. Par exemple, les psychologues utilisent classiquement des méthodes qualitatives, où les comportements de quelques participants sont analysés en

des méthodes empiriques en philosophie qui ne mentionne aucune spécificité de la philosophie expérimentale en regard des méthodes de la psychologie expérimentale.

42. Les mêmes constats se retrouvent de façon commune dans cet ouvrage universitaire (Ghiglione 2007) , que je prends ici pour référence, et par exemple dans (Sockeel et Anceaux 2008) [213] ou (Grawitz 2001) [107] ou (Coolican 2014) [54]

43. L'ouvrage (Sockeel et Anceaux 2008) sur la démarche expérimentale en psychologie détaille par exemple en son paragraphe 5.2 un ensemble de difficultés relatives à l'utilisation de la méthode expérimentale et les précautions à prendre.

profondeur, les indicateurs utilisés pouvant être complexes, et, en parallèle, des méthodes quantitatives où les échantillons de participants sont beaucoup plus importants et les indicateurs beaucoup plus simples, construits avec une visée statistique. Si les dissonances entre les indicateurs obtenus par ces différentes méthodes pour un même phénomène sont trop importantes, le psychologue se trouve confronté à la difficulté de choisir entre refuser l'un ou l'autre de ces indicateurs ou les deux. Pour aider le psychologue à affronter toutes ces difficultés liées au passage à l'expérimentation, l'ouvrage de psychologie qui nous sert ici de guide consacre un chapitre entier, le chapitre 3 intitulé « Recueil de l'information verbale », à expliciter toutes les possibilités d'erreurs expérimentales et les méthodes pour les éviter. Ce lourd investissement méthodologique du psychologue traduit concrètement l'importance du problème : la mise en doute de la pertinence du choix d'un indicateur comme représentant satisfaisant de l'objet de l'étude est un des écueils que doit affronter le psychologue expérimental pour que sa démarche empirique aboutisse.

5.4.2.2 Triangulation et cohérence

Pour diminuer ce risque et établir la crédibilité des résultats obtenus, l'ouvrage propose deux types d'approches qui visent l'une et l'autre à tirer bénéfice de la multiplication des opérationnalisations. La première approche s'appuie sur la technique de « triangulation », pour reprendre le terme de Caillaud et Flick dans (Caillaud et Flick 2016) [36], qui consiste à rechercher plusieurs dispositifs expérimentaux ou observationnels pour une même entité théorique. La seconde approche, holiste et cohérentiste, consiste à renforcer la confiance accordée à une expérience en l'intégrant dans un ensemble cohérent d'expériences soutenant globalement un vaste ensemble théorique. Détaillons ces deux approches.

Un des moyens proposés pour diminuer le risque de non pertinence des expériences est de multiplier les approches utilisées. En effet, si l'opérationnalisation multiple peut faire apparaître des dissonances, elle peut aussi montrer des convergences contribuant à lever le doute sur la validité de chacun des indicateurs convergents et c'est pourquoi, pour le psychologue, elle doit être systématisée. Face aux biais de mesure ou aux artefacts expérimentaux, la triangulation de l'opérationnalisation est analysée comme une façon de diminuer le risque en faisant varier les contextes et modes expérimentaux. On en attend deux types d'apports importants. D'une part, sous l'angle expérimental, la triangulation permet de valider les données recueillies elles-mêmes en isolant les biais de mesure qui pourraient venir de chacune des méthodes utilisées [44]. D'autre part, sous l'angle théorique, l'analyse des différences de

44. La triangulation ne peut détecter les biais partagés par plusieurs opérationnalisations. Par exemple, si on

résultats entre méthodes peut mettre en évidence des complexités inattendues demandant un effort théorique complémentaire lorsqu'une variation de contexte ou de mode expérimental conduit à des résultats trop différents au regard des hypothèses testées. Pour illustrer ce point, prenons l'exemple connu qui a conduit à la mise en cause des résultats d'expérience sur la maîtrise du concept de nombre par le enfants [45].

Quand, face à un problème d'évaluation d'un nombre et dans certaines circonstances, on demande à des enfants de produire une réponse verbale, ils donnent un résultat erroné. Mais, lorsqu'on leur demande, face au même problème et dans les mêmes circonstances, de répondre pratiquement, par exemple en choisissant entre plusieurs paquets de bonbons, ils ne se trompent pas ! L'incohérence des résultats entres les réponses verbales et non verbales pointe alors vers une autre caractérisation de la différence entre les expériences, il ne s'agirait pas de la capacité de calcul qui est la même dans les deux cas, mais du processus de l'expression verbale. La triangulation a donc permis dans ce cas de suggérer que l'erreur ne provenait pas d'un manque de maîtrise arithmétique de l'enfant mais du processus d'interrogation par l'expérimentateur, cette suggestion peut ensuite être confirmée et affinée par des expériences complémentaires ne faisant varier que les modalités du processus d'interrogation, verbal ou non verbal. Naturellement, ce résultat serait resté inobservé si, n'opérationnalisant la réponse que verbalement, le psychologue en était resté au constat fallacieux que la réponse erronée de l'enfant semblait impliquer une faiblesse arithmétique.

On peut aller plus loin que la triangulation avec la recherche de cohérence d'un ensemble d'expérimentations à l'appui non plus d'une mais d'un ensemble d'hypothèses. Pour reprendre la formulation proposée plus haut, on considère l'hypothèse qui affirme que dans un certain contexte C, un certain stimulus S aura un certain effet E, et son opérationnalisation qui consiste à rechercher des triplets (Co, So, Eo). Pour être utile, l'opérationnalisation élémentaire ne porte pas sur un seul terme de cette relation mais doit concerner le triplet (C,S,E). La démarche holiste va plus loin en considérant non plus une hypothèse particulière qu'il s'agirait d'étudier isolément, mais l'ensemble des relations et hypothèses constitutives d'un domaine théorique. On a alors d'un côté un ensemble d'hypothèses {(C,S,E)} et de l'autre côté un ensemble d'opérationnalisations {(Co,So,Eo)}. La comparaison entre résultats expérimentaux et hypothèses portera alors globalement sur l'ensemble des notions opérationnalisées et sur l'observation de l'ensemble des résultats expérimentaux au regard des relations pré-

varie l'opérationnalisation en posant différents types de questions à des participants, alors, ces différentes approches partageront d'être toutes également déclaratives et de faire appel de ce fait à l'introspection. Cette triangulation ne pourra donc pas éliminer les biais liés à l'introspection et, plus généralement, à l'approche déclarative.

45. Voir l'analyse de cet exemple dans (Dehaene 2010) p49 [69]

vues dans l'ensemble de la théorie. Le raisonnement est que si la vérification est à la fois locale, pour chaque relation théorique et chaque expérience, et globale pour toutes les relations théoriques du domaine, alors la validation empirique est renforcée et le doute lié à l'opérationnalisation peut être levé.

Observons que la définition initiale de l'opérationnalisation par les psychologues en tant que recherche d'indicateurs observables des notions psychologiques qui ne le sont pas c'est maintenant singulièrement complexifiée puisqu'il convient d'ajouter, pour répondre à l'exigence de fiabilité de ces indicateurs, que les indicateurs doivent faire l'objet d'un ensemble d'expériences complémentaires de contrôle appuyées sur des triangulations et, plus globalement, sur la cohérence d'ensemble de toutes les expériences supposées mettre en jeu les mêmes entités théoriques. Sans rejoindre complètement la définition proposée dans les sections précédentes, ce constat rapproche néanmoins le point de vue des psychologues de la nécessité d'une validation par les pairs non seulement de l'indicateur lui-même mais également de l'ensemble des protocoles utilisés dans les expérimentations.

A l'aide de l'approche par la triangulation et de l'approche cohérentiste qui ont pour objectif de diminuer les risques d'interprétations erronées, le psychologue vise, pendant l'étape de l'opérationnalisation, à répondre à une double question. D'un côté, les indicateurs observables choisis pour opérationnaliser une notion psychologique sont-ils pertinents au regard des hypothèses qui la contiennent ? Et, de l'autre côté, les résultats empiriques obtenus avec ces indicateurs sont-ils bien à interpréter comme représentant les propriétés étudiées et non comme des artefacts produits par le dispositif expérimental retenu ? Le manuel de psychologie, après avoir souligné ces difficultés, précise qu'elles ne sont pas propres à la psychologie et que même la physique est soumise à ce problème (p 309) :

> Le physicien Duhem faisait les mêmes remarques pour la physique et en concluait qu'il ne peut y avoir d'expérience cruciale, décisive : on peut mettre en cause, non l'hypothèse, mais l'opérationnalisation choisie.

L'appel à la physique pourrait être apprécié ici comme simplement rhétorique, les auteurs prenant l'exemple d'une science expérimentale paradigmatique comme point de comparaison de leur propre scientificité. L'ouvrage ne permet pas cette interprétation, il affirme simplement (p 309) l'adhésion à la méthode expérimentale faite de synergies entre théories et expérimentations et de remise en doute permanente à la fois des théories et des résultats expérimentaux.

5.4.3 Une description insuffisante

Nous avons vu que l'opérationnalisation telle qu'elle est présentée dans les manuels de psychologie a pour principal objet de trouver un indicateur observable correspondant à une notion latente importante pour le psychologue. Trouver de tels indicateurs est difficile et de nombreuses possibilités d'erreurs sont mises en évidence par les psychologues expérimentaux. Elles font de l'opérationnalisation, préalable nécessaire à toute approche empirique, une étape à risques multiples. Ces risques peuvent être combattus en multipliant les expériences et en recherchant à la fois des validations locales pour chaque expérience et des résultats cohérents pour l'ensemble des expériences et l'ensemble d'une théorie.

Le point de vue du psychologue présenté ci-dessus appelle plusieurs remarques. Tout d'abord il insiste de façon quasi exclusive sur la détection, alors que l'opérationnalisation, comprise au sens large de la démarche scientifique expérimentale, s'étend à toutes les opérations qui permettent à l'expérimentateur de détecter puis mesurer, et également de modifier, créer, supprimer, ... en un mot de manipuler pratiquement pour obtenir les phénomènes qui sont en lien avec le domaine théorique étudié. Cette limitation que s'impose le psychologue dans la présentation de sa méthodologie ne signifie pas qu'il en soit réellement ainsi dans les études psychologiques, et les exemples montrent le contraire. On peut peut-être interpréter cette auto-limitation comme une marque de la réticence à détailler comment le scientifique psychologue manipule un autre humain au cours d'une expérience. Je reviendrai sur ce point important au chapitre suivant.

Deuxième remarque, la description ne rend pas suffisamment compte de la complexité et de la fécondité de la démarche empirique qui, certes, exige de trouver des observables, mais offre en retour une dynamique de développement qui est au fondement de ses succès épistémiques et pragmatiques en accroissant nos connaissances et nos capacités d'action. Dans cette perspective, l'opérationnalisation n'est pas un processus de simple recherche d'indicateurs observables, comme une lecture superficielle des ouvrages de psychologie expérimentale pourrait le laisser croire, mais un processus itératif au long cours, à la croisée de savoir-faire expérimentaux qui évoluent et de théories qui évoluent également.

Et enfin, troisième et dernière remarque, les entités théoriques de la psychologie sont considérées comme « indispensables » (voir la citation plus haut) et l'opérationnalisation doit donc aboutir. La possibilité que ces entités indispensables n'existent simplement pas n'est pas examinée, alors que l'expérience des sciences naturelles est pavée de tels exemples (le phlogistique était certainement indispensable à certains chimistes), et que les philosophes

également envisagent systématiquement les conséquences ontologiques des choix théoriques pris en compte. Il suffira pour montrer ce point d'évoquer les théories illusionnistes de la conscience, comme celle proposée par François Kammerer (Kammerer 2019) [128] qui sont bâties sur le refus de l'indispensabilité de cette entité théorique qu'est la conscience.

5.5 L'opérationnalisation, six propositions

Je vais, dans cette dernière section du chapitre consacré à l'opérationnalisation, revenir sur six propositions qui me semblent donner à cette étape sa juste importance. J'évoquerai ensuite quelques objections que ces propositions peuvent soulever et les réponses que je peux apporter. Mais au préalable, je me propose de reprendre la définition de l'opérationnalisation revue à l'issue de ce chapitre.

L'opérationnalisation construit une articulation entre le discours théorique et la possibilité de sa mise en pratique expérimentale qui, pour être crédible, doit être validée par les pairs et peut alors donner confiance dans la possibilité d'interpréter les résultats empiriques afin d'en induire de nouvelles propositions théoriques. Détaillons cette conception :

— J'appelle opérationnalisation l'articulation entre les théories et les terrains expérimentaux, entre les théoriciens et les expérimentateurs. Plus précisément, l'opérationnalisation est la démarche qui consiste à rechercher puis à mettre en œuvre pour chaque élément stipulé par une théorie (entité, propriété, relation), les dispositifs de pratique expérimentale (protocoles, observations, mesures, équipements...) qui permettent au théoricien, à l'expérimentateur et à leurs pairs de considérer qu'on a, de façon plausible, détecté, modifié, créé, supprimé, observé, mesuré au cours de l'expérience un phénomène qui correspond à l'occurrence d'un de ces éléments stipulés par la théorie.

— L'opérationnalisation est le résultat, toujours provisoire, d'une dynamique au temps long entre des savoir-faire pratiques et les travaux théoriques, dynamique qui féconde les uns et les autres. Cette dynamique est à la fois sur les savoirs et sur les savoir-faire, elle est à la fois conceptuelle et sociale, portée par les organisations complexes des activités scientifiques et techniques.

— Ensuite, l'opérationnalisation est contextuelle. D'un côté elle dépend de la théorie qu'il s'agit d'étudier et des savoir-faire pratiques disponibles et de l'autre côté sa validation dépend de l'enjeu qu'elle représente pour les pairs, et donc de l'importance potentielle des conséquences théoriques qui seront induites des résultats expérimentaux.

De cette définition, on peut déduire plusieurs conséquences générales, et plusieurs consé-

quences plus directement liées à son application à la philosophie morale. Du point de vue général, il est nécessaire de préciser immédiatement que l'opérationnalisation telle que je la présente dans ce chapitre ne construit pas une propriété statique d'un élément (ou trace) appartenant à la pratique expérimentale qui se trouverait investie à tout jamais de la capacité de jouer le rôle de contrepartie concrétisée d'une entité appartenant à une théorie. L'opérationnalisation est au contraire dynamique, en mouvement permanent entre des hypothèses théoriques qui évoluent et des savoir-faire expérimentaux qui évoluent aussi. Ce caractère dynamique est à la fois un constat, que j'ai étayé par des exemples, et une des raisons de la fécondité de la démarche expérimentale en tant qu'elle bénéficie de deux types d'avancées, théoriques et expérimentales, largement autonomes mais articulées. Cette dynamique épistémique est également profondément sociale, fécondée par la mise en relation d'équipes de chercheurs, théoriciens et expérimentateurs, qui, chacune, a également sa propre dynamique interne.

Plus indirectement, l'opérationnalisation constitue un outil utile à décrire un champ expérimental [46] : quelles entités théoriques y sont opérationnalisées et comment. Et, enfin, grâce à cette description de l'opérationnalisation, on peut entreprendre une analyse des forces et faiblesses de ce champ expérimental en regard des théories que l'on vise à évaluer et développer empiriquement.

En ce qui concerne plus particulièrement la philosophie morale, j'ai constaté au travers des différentes études de cas, que la psychologie morale expérimentale souffre d'une surinterprétation de ses expériences qui n'est pas adossée à une opérationnalisation suffisante et reconnue. De ce fait, les interprétations inductives sont mises en doute car la relation entre tel élément théorique et tel résultat pratique est niée par certains pairs. On peut rechercher les sources de ces difficultés dans la complexité des schémas causaux invoqués par les psychologues, on peut également constater les doutes que soulèvent les expériences sur l'existence même des entités théoriques postulées et on peut enfin souligner que le phénomène moral peut être étudié à plusieurs échelles, sociales, individuelles ou neurobiologiques, échelles portées par des disciplines différentes par des communautés de chercheurs différentes qui, de ce fait, ne regardent pas le même type d'opérationnalisation comme pertinent en regard de leurs sujets d'étude.

De plus, la philosophie morale couvre des sujets à fort enjeu politique, social, et, en un mot, idéologique. On peut donc s'attendre à de longues et difficiles discussions sur la vali-

46. J'entends ici simplement par champ expérimental une expérience et ses variantes comme par exemple ce que j'ai décrit au chapitre précédent avec la tramwaylogie ou l'IAT ou l'effet Knobe.

dité des opérationnalisations proposées. L'ensemble des études de cas présentées au chapitre précédent renforce un même sentiment d'insatisfaction. Certes l'approche expérimentale permet de montrer la très grande variété des caractéristiques des situations qui sont d'une certaine importance pour l'élaboration du jugement moral, certes cette dépendance envers de multiples paramètres, comme par exemple l'odeur des croissants chauds, a mis en évidence que les théories morales disponibles ont une capacité descriptive limitée, et peut-être simplement anecdotique. Mais toutes ces expérimentations, malgré des résultats importants, ne semblent pas construire un corpus de descriptions duquel pourraient émerger des vues méta-éthiques, appliquées et substantielles satisfaisantes. Dans la perspective du philosophe des sciences, mon objectif a été de proposer dans le présent chapitre que l'étude de la sous-opérationnalisation est une voie d'analyse intéressante pour approcher cette insatisfaction.

Néanmoins, l'importance de l'opérationnalisation doit être nuancée selon la perspective que l'on prend pour étudier le domaine moral. Elle est centrale dans la perspective descriptive, mais est certainement plus modeste dans la perspective substantielle de l'apologie d'une morale particulière ou dans la perspective des éthiques appliquées, et intermédiaire dans la perspective méta-éthique. Dans chaque cas, se poser la question de l'opérationnalisation, de comment chaque élément postulé par la théorie est supposé pris en compte par l'expérimentateur dans son dispositif pratique, est un révélateur utile de la complexité réelle des phénomènes étudiés et des angles d'attaque adoptés par les chercheurs face à cette complexité.

5.5.1 Six propositions

La première proposition prend la forme d'un constat : il y a un lien fort entre le risque de confusion que recèlent les analyses des expériences psychologiques, la mise en doute des interprétations inductives construites sur ces analyses et la qualité de l'opérationnalisation menée par les expérimentateurs pour faire correspondre leurs protocoles expérimentaux avec les entités théoriques postulées.

La deuxième proposition est de préciser ce constat en définissant l'opérationnalisation comme la recherche de terrains d'expérience et de protocoles opératoires dont les pairs acceptent la pertinence en regard des éléments postulés par une théorie, entités, propriétés et relations. J'appelle ainsi opérationnalisation l'articulation épistémique et sociale entre les théories et les terrains expérimentaux, entre les théoriciens et les expérimentateurs. Plus précisément, l'opérationnalisation est la démarche qui consiste à rechercher puis à mettre en œuvre pour chaque élément postulé par une théorie (entité, propriété, relation), les pratiques

expérimentales (protocoles, observations, instruments, ...) qui permettent au théoricien, à l'expérimentateur et à leurs pairs de considérer qu'on a de façon plausible détecté, modifié, créé, supprimé, observé, mesuré un de ces éléments postulés par la théorie au cours d'une expérience.

Troisième proposition, l'opérationnalisation est toujours provisoire, elle s'inscrit dans une dynamique au temps long entre des savoir-faire pratiques et les travaux théoriques, dynamique qui féconde les uns et les autres. Il s'agit d'un processus itératif au long cours, à la croisée de savoir-faire expérimentaux qui évoluent et de théories qui évoluent également. L'opérationnalisation n'est pas un processus de simple recherche d'indicateurs observables, comme une lecture superficielle des ouvrages de psychologie expérimentale pourrait le laisser croire. Cette dynamique est à la fois celle des savoirs et celle des savoir-faire, elle est à la fois conceptuelle, technologique et sociale, portée par des organisations complexes où toute innovation locale interagit avec l'ensemble du domaine.

Quatrième proposition, donner de l'épaisseur à cette étape de l'opérationnalisation, lui donner de l'importance, c'est se mettre en capacité de distinguer des domaines théoriques, expérimentaux et technologiques concernés, qui n'évoluent pas selon les mêmes vitesses, qui évoluent au sein de groupes professionnels distincts avec des logiques différentes. L'opérationnalisation est la résolution simultanée de trois équations, existentielles épistémique et pratique. Elle suppose des investissements sur chacun de ces trois axes qui sont coûteux, et de ce fait, structurants. L'opérationnalisation se fait sous la contrainte de la disponibilité des facteurs limitants, y compris financiers ou technologiques.

Cinquième proposition, ainsi définie, l'opérationnalisation est un point faible de la philosophie morale expérimentale. La psychologie morale expérimentale souffre d'une sur-interprétation de ses expériences qui n'est pas adossée à une opérationnalisation suffisante et reconnue [47]. De ce fait, les interprétations inductives sont mises en doute car la relation est niée par certains pairs. L'opérationnalisation est une étape importante de la démarche du psychologue expérimental dans laquelle on peut validement rechercher la source de certaines difficultés d'interprétation inductive des résultats des expériences. L'opérationnalisation, en tant qu'elle peut constituer un objet de discussion privilégié entre équipes différentes, constitue une opportunité pour la diminution des risques de biais.

Enfin, sixième proposition, cette importance de l'opérationnalisation doit néanmoins être relativisée. Elle est centrale dans la perspective descriptive, mais est certainement plus mo-

47. A titre d'illustration du peu de place prise actuellement par la méthodologie, dans (Sarkissian et Wright 2014) [200] sur 12 articles qui font le point sur la psychologie morale expérimentale, un seul traite des méthodes de mesure, et c'est le dernier.

deste dans la perspective substantielle de l'apologie d'une morale particulière ou dans la perspective des éthiques appliquées, et très forte dans la perspective méta-éthique.

Les sections précédentes ayant contribué à établir, positivement, ces propositions illustrées par les cinq études de cas, je me propose maintenant d'envisager certaines objections qu'elles pourraient soulever.

5.5.2 Objections

On pourra opposer à ma proposition principale, l'importance de l'opérationnalisation pour la philosophie morale expérimentale, l'argument suivant : il n'est pas pertinent de souligner l'importance de la seule étape d'opérationnalisation comme maillon fragile de la philosophie morale expérimentale quand c'est l'ensemble de la démarche de la philosophie expérimentale qui est à mettre en doute.

L'argument peut prendre deux formes symétriques extrêmes selon qu'on part des théories ou du terrain, et tout un ensemble de formes intermédiaires. Dans la première forme, l'argument se résume ainsi : les concepts fondamentaux de la philosophie morale sont soit des fictions, soit trop complexes, et les relations théoriques entre eux sont des affirmations dont il est illusoire de rechercher une opérationnalisation satisfaisante [48]. Dans la seconde forme, l'argument serait : on n'a jamais rien mesuré en psychologie [49], tout au plus recherché des classifications approximatives permettant une première approche, ce qui limite fortement la solidité des bases empiriques et la portée de leurs résultats.

Que l'on adopte l'une ou l'autre de ces formes extrêmes de l'argument ou des formes intermédiaires, la fragilité de l'étape d'opérationnalisation n'est plus qu'une conséquence de la fragilité de l'ensemble des démarches de la psychologie expérimentale et de ses théories et il convient de ne lui apporter que peu de poids. Face à cette objection, il me semble que l'on peut défendre la thèse de l'intérêt de l'étude de l'opérationnalisation en tant que point fragile de la philosophie morale expérimentale de deux façons. La première sera un rappel historique sur le développement d'autres sciences, dont la physique, la seconde l'inscription de la thèse dans une dynamique. Premier contre-argument, si on s'intéresse, par exemple, à la température [50], on observe qu'avant le 18^{e} siècle il y avait à la fois une grande confusion conceptuelle entre température, quantité et flux de chaleur, et une grande imprécision des protocoles expérimentaux. Ce sont des avancées sur les deux plans, conceptuels et expé-

48. Dans cette perspective, il vaut mieux s'appuyer par exemple sur des romans dont les situations sollicitent mieux nos intuitions morales, que sur des cas visant à opérationnaliser des entités théoriques.
49. Voir par exemple Stéphane Vautier sur http ://epistemo.hypotheses.org/cours-video et (Michell 2012) [163]
50. Pour une description détaillée du développement de ce domaine des températures voir (Chang 2007) [42]

rimentaux, les unes appuyant les autres, qui ont permis de développer des théories et des mesures plus satisfaisantes de la température. L'opérationnalisation, qui permet justement cette mise en perspective croisée, peut constituer alors un point de vue à privilégier pour l'observation de ces avancées. D'où le premier contre-argument : pourquoi écarter pour la psychologie morale une approche qui a réussi ailleurs ? Sans garantie de résultat, bien sûr, mais avons-nous une meilleure alternative ? Le second contre-argument, en prolongement du premier, est que l'opérationnalisation ne porte pas obligatoirement sur tout un domaine, ici globalement la philosophie morale, mais porte plus localement sur une proposition théorique, ou une entité, particulière. Au cas par cas et progressivement, peut se construire une dynamique de la recherche qui verra se développer à la fois les savoir-faire des expérimentateurs et les concepts des théoriciens ainsi que la relation de confiance entre eux. L'article numéro 21 de l'étude de l'effet Knobe, (Chakroff 2016) [41] utilisant la corrélation entre une configuration neuronale et la théorie de l'esprit pourrait ainsi devenir un exemple de premier pas vers une opérationnalisation acceptée d'une entité théorique potentiellement utilisable dans de nombreux plans d'expérience.

Le deuxième contre-argument nous rapproche naturellement de ma troisième proposition : la durée est une dimension nécessaire à l'opérationnalisation. On pourra opposer à cette proposition qu'elle n'offre pas de critère de fin du processus d'opérationnalisation et qu'ainsi elle ouvre la porte à un scepticisme systématique sur la possibilité d'acquérir empiriquement des vérités définitives. Reconnaissons que si nous recherchons une validation absolue d'une hypothèse, cet argument est valide : l'acceptation de l'opérationnalisation peut toujours être remise en cause par des avancées théoriques ultérieures ou des évolutions des savoir-faire expérimentaux. Mais si on accepte une validation relative aux périmètres expérimentaux testés à un certain moment, alors on peut tout à fait accepter définitivement une hypothèse dont la portée est restreinte à ces périmètres ; ainsi dans l'exemple acoustique, l'opérationnalisation des phénomènes vibratoires aux basses fréquences appuyée sur la théorie modale peut continuer à être acceptée pour les basses fréquences même après qu'elle ait montré des insuffisances aux hautes fréquences.

5.5.3 Conséquences pour la philosophie morale expérimentale

On pourra encore objecter que l'opérationnalisation n'est importante que dans la perspective descriptive de la philosophie morale, qu'elle a peu d'intérêt dans les trois autres perspectives, substantielle, méta-éthique et éthique appliquée. S'intéresser à l'opérationnalisation

n'est donc pas moralement neutre mais résulte d'un penchant naturaliste partisan en ce qu'il est défavorable aux théories morales non naturalistes.

Tout d'abord, concédons que si l'opérationnalisation est centrale pour la perspective descriptive, elle est anecdotique pour la philosophie morale appliquée car, par construction, celle-ci ne vise pas à s'inscrire dans une dynamique générale consistant à bâtir des théories empiriquement renseignées mais à traiter, ici et maintenant, un cas particulier. Ce qui importe est alors de trouver une méthode permettant de résoudre les conflits entre acteurs pour arriver (assez souvent) à une décision (suffisamment) acceptée dans sa dimension morale. Ce n'est que dans cet objectif, et donc indirectement, de recherche de dispositifs de négociation efficaces que des expérimentations pourraient être utilement menées dans la perspective de l'éthique appliquée, et non sur le plan de la morale elle-même.

Pour le défenseur d'une théorie morale particulière, comme nous l'avons vu précédemment, c'est l'option même d'appuyer son apologie sur une approche expérimentale qui est problématique. Cela peut être, bien sûr, totalement impossible lorsque la théorie morale s'appuie sur des entités théoriques non naturelles et sans puissance causale inaccessibles à toute expérimentation (voir par exemple le « réalisme moral robuste » de David Enoch dans (Enoch 2013) [80] et au point 1.3.3, page 74). Mais plus largement, l'acceptation de l'inscription dans la dynamique itérative de la démarche expérimentale, telle qu'elle transparaît dans la métaphore de l'hélice que j'utilise dans cette thèse, ne serait admissible pour aucune des grandes traditions morales, déontologisme, conséquentialisme, éthique des vertus, qui supposent un engagement dans des règles ou des principes ou des maximes (ou tout autre élément définissant la morale) qui ne sauraient être remis en cause suite à une expérience.

La situation est plus complexe, et plus intéressante, pour la perspective méta-éthique. Rappelons que celle-ci a pour objectif de réfléchir sur le phénomène moral dans toutes ses dimensions, sémantique, épistémologique, ontologique et psychologique (Desmons 2019) [71]. Elle a donc à comparer les théories morales sur tous ces axes, comment les phrases évaluatives prennent sens, comment nous en venons à croire (ou à connaître) des vérités morales (et si elles existent), quels engagement ontologiques supposent nos croyances morales, et comment nos croyances en général, et nos croyances morales en particulier, influent sur nos comportements et sur nos jugements. Dans chacune de ces dimensions, la méta-éthique fait appel à tous les outils du philosophe et au premier chef, pour la philosophie analytique constituant l'essentiel de la production académique que j'étudie, à l'analyse conceptuelle. Deux questions apparaissent alors : quelle place donnera le méta-éthicien aux observations empiriques ? Et, seconde question : est-ce que cette place qu'il donnera aux résultats empiriques constitue une

entorse à sa neutralité supposée envers les différentes théories morales ?

Pour tenter d'éclairer cette double question, je propose de relever dans le manuel de méta-éthique déjà cité (Desmons 2019) [71] des exemples de mentions des apports de l'expérimentation à des débats de méta-éthique :

— Descriptivisme vs. expressivisme : l'observation empirique des phrases prononcées dans un contexte de jugement moral montre que, dans toutes les langues, la forme grammaticale descriptive est utilisée pour les propositions évaluatives. L'expressivisme n'est donc pas une théorie morale supportée par les faits de langage. On utilise plus souvent un style descriptif « ceci est mal » qu'un style expressif « je rejette ceci» (Desmons 2019 page 61 contribution de Cain Todd).

— Relativisme moral : le relativisme moral, au sens où certains jugements moraux ne sont pas universellement déterminés mais dépendent du contexte social, est un fait empirique établi. (Desmons 2019 page 91 contribution de Isidora Stojanovic)

— Intuitionnisme : L'intuitionniste pense que nous avons un accès direct aux vérités morales. Les biais portant sur les jugements moraux sont empiriquement bien établis. Si nous avons un accès direct non fiable, et que cette non fiabilité nous est opaque, alors l'intuitionnisme est en difficulté (Desmons 2019 page 131 contribution de Stéphane Lemaire)

— Le constructivisme moral : si ce qui est important n'est pas la règle morale en elle-même mais comment elle s'est construite et, simultanément, comment s'est construite notre disposition à la respecter, alors le constructivisme peut ête vu comme une tentative de conciliation des sciences naturelles et des sciences sociales (Desmons 2019 page 232 contribution de Ophélie Desmons).

— internalisme vs. externalisme moral : L'internaliste motivationnel soutient qu'il existe un lien nécessaire entre le jugement moral et la motivation à l'action. Par une approche expérimentale, les XPhi ont montré que les participants à un sondage ont à 80 % l'intuition inverse : on peut juger moralement sans incidence sur la motivation. (Desmons 2019 page 361 contribution de François Jaquet et Florian Cova)

Comme je l'ai déjà évoqué plus haut (voir 2.5, page 118) on constate ici que les apports empiriques touchent à de nombreux domaines de la méta-éthique mais sans jamais être déterminants. En effet à chacun des arguments, les défenseurs de la théorie morale mise empiriquement en défaut pourront trouver différentes parades pour poursuivre le débat, ce qui est illustré dans chacune des contributions mentionnées ci-dessus. La double question posée plus haut revient alors sous une nouvelle forme. Première question, quel poids donner à

l'argument empiriquement établi en regard des autres arguments développés (souvent analytiques)? Seconde question, le poids fort donné aux résultats empiriques ne va-t-il pas de pair avec un engagement du méta-éthicien en faveur des théories plus amènes aux approches naturalistes?

Les cinq exemples de mentions des méta-éthiciens ci-dessus ne donnent pas de réponse à la première question. Chaque philosophe moral décide du poids qu'il donne aux arguments empiriques, et c'est un choix méta-éthique en soi. Si, à l'image de David Enoch lorsqu'il défend le réalisme moral robuste (voir 1.3.3, page 74), est dénié tout intérêt à l'approche expérimentale au titre du risque qu'elle ferait courir à la prise en compte sérieuse de la morale (Taking Morality Seriously (Enoch 2013) [80]), alors la réponse est immédiate : aucun poids ne peut ni ne doit être donné aux résultats empiriques, si ce n'est marginalement. En revanche, si comme les constructivistes, il est considéré qu'une théorie méta-éthique doit expliquer comment la morale a pu émerger, alors elle ne peut contredire les résultats empiriques venant des sciences naturelles qui lui servent de fondement.

En revanche, pour la seconde question, le reproche de non neutralité, les exemples permettent de proposer qu'il convient de l'inverser : c'est le choix du philosophe moral d'accepter ou non l'apport empirique qui brise la neutralité supposée de la méta-éthique entre théories morales. Ce choix ne peut être neutre car, dans chaque cas ci-dessus, agréer cet apport n'est pas sans incidence morale. La méta-éthique elle-même se doit de scruter tout ce qui touche au phénomène moral et a, comme le montrent également les exemples, la tâche d'inventorier les arguments des uns et des autres qu'ils soient analytiques ou empiriques. Il serait étonnant que, toujours, les arguments empiriques soient déterminants. Il serait étonnant, inversement, que, toujours, les arguments empiriques ne soient qu'accessoires. Accuser de non neutralité la méta-éthique quand elle décide de considérer les arguments empiriques semble donc infondé. Cette accusation est plus à porter envers les théories morales qui conduisent à ne les considérer jamais (comme David Enoch le fait) soit à les privilégier toujours (ce qu'aucun philosophe moral ne fait, à ma connaissance [51]).

Le point de vue que je défends ici, les six propositions qui mettent en avant l'importance de l'opérationnalisation pour la philosophie morale et, tout particulièrement dans les perspectives descriptives et méta-éthiques, peut être jugé comme faisant la part trop belle à l'expérimentation, à l'apport empirique, alors que la philosophie morale doit s'attacher avant tout

51. Même Patricia Churchland dans son dernier ouvrage sur la conscience morale ne va pas jusque là, les théories morales y ont encore une large place en tant qu'elles motivent à un comportement social bénéfique à l'espèce même quand elles recommandent des comportements qui ne sont pas empiriquement universellement constatés.(Churchland 2019) [47]

à être réformatrice, à nous conduire à être meilleurs. Je vais maintenant, dans le chapitre suivant, prendre comme sujet central cette critique des philosophes moraux et voir en quoi et selon quels arguments ils considèrent que la philosophie morale expérimentale n'est que de peu d'apport à leur discipline.

Chapitre 6

Philosophie morale et approche expérimentale

Dans ce chapitre, j'endosse, pour partie, le point de vue du philosophe moral qui, considérant que le rôle essentiel de la morale est d'assurer la cohésion sociale, voit avec méfiance le développement d'une approche expérimentale qui vise à rapprocher les jugements moraux des jugements de faits accessibles à la démarche scientifique. Ce rapprochement risque de saper les justifications qu'il a pu construire avec le temps au sein de sa tradition morale, sans que les nouvelles connaissances puissent se traduire concrètement, ici et maintenant, par des raisons de respecter les règles morales en place. Puis, j'endosse, pour une autre partie, le point de vue des scientifiques qui poursuivent leurs recherches empiétant sur le domaine moral, malgré toutes les difficultés soulevées par les philosophes moraux.

Je vais, dans un premier temps, rappeler que les premiers développements de la psychologie scientifique portaient sur le développement des capacités des enfants et étaient assez éloignées des préoccupations morales, ce n'est que récemment que les développements de cette psychologie sont venus empiéter sur les territoires de la philosophie morale. Je présenterai ensuite quatre types d'arguments qui structurent la réticence profonde des philosophes moraux à prendre en compte les résultats expérimentaux, ils sont d'une part sceptiques, le comportement humain serait trop complexe pour être étudié avec ces outils, et d'autre part moraux, car le scientifique se donne une position de surplomb qui contrevient aux principes de base de la morale exprimés, par exemple, dans la règle d'or de la réciprocité.

J'examinerai ensuite les réponses que peuvent apporter les scientifiques face à ces réticences selon chacune des perspectives que peut adopter le philosophe moral sur son domaine,

la perspective descriptive, prescriptive, la perspective méta-éthique et celle de l'éthique appliquée. Je développerai en particulier l'apport très prometteur des théories évolutionnistes de la morale. Et je conclurai par une proposition : le débat pourrait être clarifié en distinguant, d'un côté, ce qui fait la nécessité de la norme, pourquoi des normes morales sont nécessaires au bon développement de structures sociales au sein de l'espèce humaine et, de l'autre côté, le contenu de ces normes qui est, au moins pour une large partie, dépendant de faits historiques contingents. Cette distinction, que je propose comme une des propositions à l'issue de cette thèse, marquerait ainsi une voie de clarification de la démarcation entre différents domaines du phénomène moral à l'échelle de l'espèce, à l'échelle du groupe social et à l'échelle de l'individu : il serait alors indispensable à celui qui veut aborder expérimentalement le domaine moral dans toute sa complexité de bien distinguer, d'un côté, les expériences tendant à montrer la nécessité qu'il y a pour des êtres sociaux et rationnels à se coordonner et, pour cela, à disposer de règles morales et, d'un autre côté, les expériences tendant à étudier les règles morales instituées dans différentes communautés morales.

6.1 L'expérimentation morale est-elle impossible ?

6.1.1 Quatre types d'arguments

6.1.1.1 Une réticence ancienne et qui perdure

Le philosophe moral s'intéresse aux comportements humains, ce qu'ils sont et ce qu'ils devraient être [1]. A ce titre, il semble qu'il devrait, au moins pour partie, être à l'écoute des sciences de l'homme, et en particulier de la psychologie, de façon à bénéficier des approches empiriques sur lesquelles ces sciences s'appuient. Pourtant, cette affirmation n'a rien d'une évidence et, sans faire ici œuvre d'historien, il sera utile de remarquer qu'on peut distinguer quelques grandes étapes dans l'histoire récente du rapport des philosophes moraux à la psychologie.

Avant le vingtième siècle, on peut considérer que la connaissance du comportement humain était à un stade préscientifique, les philosophes moraux n'avaient alors pas d'intérêt particulier à s'y référer. Cela ne signifie pas que la réalité empirique leur était indifférente mais que, plutôt, ils en avaient un accès direct équivalent à ce à quoi pouvaient prétendre les approches préscientifiques. Le vingtième siècle a vu l'émergence d'un ensemble de disciplines en lien avec le comportement humain, au premier rang desquelles la psychologie. Mais jus-

1. Une version antérieure et réduite de ce chapitre a été publié par l'auteur en juin 2019 (Serra 2019) [204].

qu'à la fin de ce siècle, les philosophes moraux n'ont pas attaché une grande importance à ces nouvelles disciplines et, inversement, la toute récente psychologie s'intéressait surtout à des questions de développement individuel. Elle s'intéressait à comment les humains acquièrent les compétences morales qui sont les leurs à l'âge adulte, en prolongement des travaux de Piaget et Kohlberg qui ont pour principal objet de définir, sur des bases empiriques, différents stades de développement par lesquels passerait tout enfant pour accéder à la pleine et entière personnalité morale.

Les psychologues considèrent que, jusqu'à la fin du 20^{e} siècle, leur discipline s'est centrée sur ces questions de développement et peu sur la nécessité d'apporter des arguments utiles aux débats classiques de philosophie morale. Ce bilan critique est régulièrement repris dans la partie historique des travaux sur la philosophie morale, pour illustrer ce bilan j'ai choisi un texte établi à chaud par Laurent Bègue en 1998 (Bègue 1998) [21] précisément dans l'objectif de proposer une présentation critique de l'état de la psychologie morale au tournant du siècle. Dans cet article de 1998, la théorie de référence est le modèle constructiviste de Kohlberg qui propose un certain nombre d'étapes par lesquelles passe le jeune enfant puis l'adulte pour acquérir la pleine personnalité morale. L'article présente ensuite les critiques de ce modèle, critiques élaborées ou en cours d'élaboration en 1998, et principalement l'importance du « care » proposée par Carol Gilligan (Gilligan 1982) [102] ainsi que les théories liant l'apprentissage moral à la socialisation. Gilligan critique les bases empiriques de Kohlberg qui ne comportent que des participants masculins, ce qui appauvrit l'analyse en mettant l'accent sur la morale du devoir au détriment de la morale de la sollicitude, le « care », dont pourtant le poids est également important lorsque l'échantillon étudié est mixte. Les travaux de Gilligan mettent ainsi l'accent sur le fait que l'apprentissage moral est, pour une part importante, social et, à ce titre, dépend des rôles que la société attribue, ici aux hommes et aux femmes. L'étude du développement telle que Piaget et Kohlberg l'envisagent, c'est-à-dire comme une caractéristique objective et commune à tous les être humains, doit être largement remise en cause pour prendre en compte cette dimension sociale de l'apprentissage moral.

L'article de Laurent Bègue reprend ensuite plusieurs questions transversales, objets d'attentions particulières en cette fin de vingtième siècle. Ces questions sont principalement induites par un constat de carence : aucune des théories psychologique reconnues et étudiées en 1998 ne rend compte de la grande diversité des jugements moraux, que ce soit entre cultures différentes ou, au sein d'une même culture, en fonction des situations et des individus. Pour l'auteur, ce constat mène le psychologue à devoir s'intéresser à de nouveaux problèmes : la contribution de la psychologie à la prédiction du comportement des individus, les variations

socioculturelles du jugement moral, et la dimension stratégique et idéologique du positionnement moral. Aucun des thèmes mis en avant dans cet article de 1998 ne reprend, essentiellement, les grandes questions qui animent les débats entre philosophes moraux, réalisme contre antiréalisme des attributs moraux, périmètre du domaine moral, justification des obligations morales ... En somme, une part très faible de la psychologie morale, dans les termes utilisés comme dans les problèmes abordés en cette fin de vingtième siècle, peut être facilement et directement mise en relation avec les débats de la philosophie morale traditionnelle. Bien qu'abordant le même sujet, le comportement moral humain, les ordres du jour des deux disciplines apparaissent comme profondément indépendants.

La deuxième moitié du vingtième siècle et le début du vingt-et-unième ont vu un important virage lié d'une part au développement exponentiel des connaissances scientifiques relatives au comportement humain en général, et plus particulièrement à la psychologie morale expérimentale (Sarkissian et Wright 2014) [200] ainsi que le développement de l'initiative de la philosophie expérimentale (XPhi) qui a mis en lumière la possibilité pour la philosophie en général, et pour la philosophie morale en particulier, de faire appel aux méthodes expérimentales utilisées par les scientifiques.

Après une phase de fortes polémiques entre les philosophes expérimentaux et les philosophes en fauteuil, le débat est aujourd'hui moins emporté. D'un côté, et au moins en façade, les philosophes marquent leur intérêt pour les avancées scientifiques relatives au comportement humain. On peut citer ici Max Deutsch qui, bien qu'opposé aux conclusions de la philosophie expérimentale, déclare dans (Deutsch 2015 page 157) [72] qu'il est faux d'affirmer que la philosophie en fauteuil ne s'intéresse pas aux résultats empiriques :

> The characterization (of analytic philosophy by XPhi) wrongly suggests that armchair philosophy is unscientific, or unconcerned with empirical results related to its subject matter .
>
> Les philosophes expérimentaux suggèrent de façon erronée que la philosophie en fauteuil ne serait pas scientifique, ou serait indifférente aux résultats scientifiques dans son domaine d'étude.

Mais, au delà de cet accord de façade, les philosophes ont développé de nombreux arguments [2] tendant à faire douter de l'intérêt de cet apport empirique à la philosophie morale et à en évoquer les limites.

Avant de rentrer dans le détail de ces arguments visant à disqualifier l'approche expérimentale, citons à titre d'illustration un article récent de Emilio Martinez Navarro, professeur

2. Pour une revue de ces arguments à charge voir par exemple (Kauppinen 2007) [131]

de philosophie morale à l'université de Murcie (Martínez Navarro 2017)[158] dont l'objectif est précisément d'évaluer la démarche des philosophes moraux expérimentaux [3] [4] :

> 1) Il est impossible de découvrir les fondements de la conduite morale et de mener le type d'expériences approprié pour les découvrir si le concept de ce que nous entendons par «moralité» n'a pas été déterminé philosophiquement au préalable. Par exemple, il n'est pas évident que, dans l'expérience de Greene [5] que nous avons mentionnée précédemment, les jugements moraux soient considérés comme identiques aux jugements prudentiels qu'un sujet élabore devant une certaine situation dans laquelle d'autres personnes ont besoin d'aide.
>
> 2) Ceux qui effectuent des expériences liées à la moralité ne peuvent se passer des théories philosophiques pour interpréter le sens des données scientifiques en relation avec la vie morale, car la connaissance morale ne peut se contenter d'un agrégat de contributions fragmentaires, elle doit également fournir un cadre général permettant d'interpréter les données scientifiques. Il existe une interaction entre le cadre herméneutique et ce que l'on peut qualifier de "découvertes pertinentes".
>
> 3) Les sciences empiriques ne peuvent se contenter d'exposer les fondements biologiques de la conduite morale mais doivent également découvrir le fondement de ces obligations morales et, pour ce faire, une collaboration avec la philosophie morale est nécessaire. Pour qu'un sujet soit considéré comme un agent moral, il est essentiel qu'il soit vivant, en bonne santé, etc. mais cela ne suffit pas encore à lui reconnaître des obligations morales envers lui-même et envers les autres, et encore moins si l'on envisage des obligations morales universalistes, c'est-à-dire, non seulement vis-à-vis de ses proches, mais aussi de toute l'humanité. Nous avons besoin d'arguments philosophiques dans l'objectif d'offrir une justification de la moralité au niveau post-conventionnel au sens de Kohlberg, [6] ce qui est le seul type de moralité qui conviendrait à la conscience morale de notre temps.

Dans cette citation, on trouve à la fois les critiques de méthode et de fond. Sur la méthode,

3. Traduction de l'auteur

4. De nombreux philosophes moraux ont des positions proches, voir par exemple (Larmore 1993 p 43) [145] « Ce qu'il faut c'est abandonner cet attachement naïf et pieux au naturalisme. Il nous faut ouvrir l'esprit à la possibilité réelle que le monde soit quelque chose de beaucoup plus complexe. ».

5. Il s'agit ici des expériences sur les dilemmes du tramway de 2001 appuyées sur les techniques d'imagerie IRMf et décrites en 4.2 page 161

6. Rappelons que dans la théorie du développement moral de Kohlberg, le niveau post-conventionnel est le niveau le plus élevé de la moralité qui consiste non seulement à suivre les conventions, éventuellement à ses dépens, mais également à en comprendre les justifications et à pouvoir les amender et les étendre à de nouvelles situations sociales qui l'exigeraient.

les philosophes expérimentaux n'apprécieraient pas assez le risque de circularité qu'il y a à interpréter leurs résultats en utilisant des concepts qui sont déjà chargés de théories morales et le manque de fiabilité de ces études rendrait leur interprétation discutable. Sur le fond, l'observation ne peut en aucune façon nous offrir la justification des règles morales dont le monde a besoin. Elle ne peut, qu'au mieux, en constater l'existence.

Le point de vue de l'auteur est ici représentatif de la conception qui consiste à considérer que le philosophe moral n'est pas un simple observateur mais qu'il a, au moins pour partie, un objectif réformateur. Il a pour but d'améliorer le fonctionnement du monde. A ce titre il a besoin de la morale en tant qu'outil utile face aux désordres du monde, et celle-ci a besoin de justifications fortes pour être respectée. L'observation des corrélats biologiques du comportement moral n'apporte pas de justification. Elle peut satisfaire la curiosité sur le « comment » des comportements humains mais n'instruit pas leur « pourquoi », leur justification. Elle est donc au mieux insignifiante au regard de la responsabilité réformatrice du philosophe moral, au pire nuisible car elle sape les justifications des règles morales sociales en les soumettant au doute scientifique.

6.1.1.2 Les arguments des philosophes moraux

Je me propose dans les sections suivantes d'expliciter les arguments des philosophes moraux visant à relativiser l'intérêt de l'approche empirique des questions morales, de proposer les réponses qui peuvent être apportées à ces arguments et, finalement, de préciser quelles reformulations de ces arguments permettent au mieux d'en mesurer l'importance en regard de la démarche empirique. Je ne reprendrai pas ici les cas extrêmes les plus défavorables aux expérimentateurs, et supposerai l'absence de fraude [7] ou de manipulation douteuse des traces laissées par une observation ou une expérimentation, les critiques ne porteront donc pas sur la sincérité des résultats obtenus mais sur la possibilité de leur interprétation en appui, ou en défaveur, d'une proposition morale. Plusieurs possibilités de regroupement de ces arguments critiques sont envisageables et, pour partie, chaque proposition de regroupement traduirait la posture adoptée selon qu'on est un scientifique impliqué dans une des disciplines étudiant le comportement humain, ou un philosophe des sciences tentant d'analyser cette activité ou, et ce sera le point de vue que je vais favoriser dans ce chapitre, un philosophe moral directement concerné. Un premier axe possible distinguerait les arguments selon leur niveau d'abstraction, allant de points de pratique expérimentale pour s'élever en abstraction vers

7. La fraude scientifique n'est pas particulière au domaine de la psychologie morale, et nous renvoyons sur ce thème aux études centrées sur ce sujet. (Nicolas Chevassus-au-Louis 2017) [44]

les questions morales. On pourra alors mettre en cause tout d'abord la qualité des travaux eux-mêmes, puis les méthodes utilisées, puis la pertinence de l'approche expérimentale pour traiter des problèmes moraux, et enfin la pertinence morale de ces approches. On peut, différemment, proposer un axe distinguant les arguments sceptiques génériques qui porteraient sur toute connaissance, puis les arguments portant sur toute connaissance psychologique, et enfin ne portant que sur la seule connaissance spécifique au domaine moral. On rejoindrait ainsi les préoccupations épistémiques des philosophes des sciences.

Un autre axe serait de développer les arguments sur une trame de présentation reflétant les débats animés par les philosophes moraux et repris par les psychologues. On aurait alors, à titre d'exemple, des arguments relatifs à l'expérimentation sur la distinction entre normes morales et conventions, entre réalisme et fictionnalisme des propriétés morales, sur le libre arbitre, ou sur les différents cas d'attribution d'intentionnalité, ... Cet axe est celui adopté dans les descriptions des scientifiques (ou des philosophes engagés dans le mouvement XPhi) qui font le point sur les avancées et difficultés de l'expérimentation dans chacun de ces domaines. On a deux bons exemples de cette approche avec (Doris (ed.) 2012) [76] et (Mukerji 2019) [167].

Je propose ci-après de rester au plus proche des argumentations qui, comme celle d'Emilio Martinez Navarro, reflètent la position des philosophes moraux qui ne sont ni opposés systématiquement à la démarche expérimentale ni particulièrement promoteurs de cette démarche. La présentation s'appuiera sur quatre types d'arguments, le premier porte sur la complexité du comportement humain liée, pour partie, à la circularité de l'homme s'étudiant lui-même. Avec le deuxième argument, le problème de Hume, on entre dans le domaine spécifiquement moral, le troisième argument, toujours présent dans les discussions sur l'éthique du chercheur, porte sur la possibilité morale d'expérimenter sur la morale, et enfin le quatrième sur le risque social qui serait, pour certains philosophes, lié à l'approche empirique des problèmes moraux. Le premier type d'argument est pratique et épistémique, la complexité du sujet n'est pas à notre portée, tout particulièrement quand il s'agit de nous connaître nous-mêmes, le deuxième est d'ordre logique, l'impossibilité de déduire ce qui doit être de ce qui est, et enfin, les deux derniers sont des impossibilités morales, soit parce qu'expérimenter enfreint des règles morales, soit parce que chercher à expliquer empiriquement les règles morales diminue le poids de leur justification. Je vais maintenant détailler les critiques relatives à la complexité, la circularité et la normativité, les réponses qui peuvent être apportées à ces critiques et, finalement, l'importance qu'elles ont en regard de la démarche empirique.

6.1.2 La complexité et la circularité

6.1.2.1 Le problème de la complexité du comportement moral

Les arguments liés à la complexité s'appuient sur les caractéristiques qui feraient du comportement moral un domaine à part, hors d'atteinte de l'approche par les sciences expérimentales. Tout d'abord, le comportement moral est affaire de responsabilité et de choix individuels. Il est donc propre à chacun des sept milliards d'individus. Le comportement moral n'est pas qu'individuel, il est également dépendant du contexte social dans lequel chacun a grandi et, à tout moment, exerce ses choix de vie. Il s'inscrit donc dans le réseau complexe des millions de cultures qui cohabitent dans l'espace social. Enfin, le comportement moral a une dimension cognitive et émotive qui repose sur la structure du cerveau humain, organe à part souvent qualifié de plus complexe organe qui soit avec ses centaines de milliards de neurones chacun relié aux autres par plus de dix milles synapses par neurone créant des réseaux neuronaux qui évoluent dynamiquement.

Du fait de cette complexité, on est en droit d'être perplexe quant à la possibilité pour une observation menée par un agent humain ayant son propre référentiel moral, à un certain instant en un certain endroit avec un certain protocole sur un certain échantillon de personnes de pouvoir être fortement signifiante pour le comportement humain en général. Une telle faiblesse est, au moins pour partie, en contradiction avec l'ambition de la psychologie morale qui est de rechercher des régularités morales générales permettant de mieux comprendre le phénomène moral.

Cette perplexité se décline selon deux modalités, d'une part sur le plan des méthodes d'investigation des philosophes expérimentaux et d'autre part sur la faisabilité du principe même de leur projet. Dans le premier cas, on mettra en avant l'insuffisance de la puissance statistique, les biais liés à des échantillon trop souvent limités à un type de population particulier (les étudiants de psychologie en début de cycle), les biais linguistiques, les contenus implicites surdéterminant le contenu moral des cas présentés... Dans le second cas l'impossibilité d'accéder à des cultures n'existant plus ou n'existant pas encore, le problème de la définition des termes moraux dans les différentes langues et, plus intimement, l'impossibilité pour les participants à accéder par introspection à une connaissance non biaisée de leur propre comportement moral...

Toutes ces difficultés sont évidemment connues des psychologues. Elles conduisent a minima à s'imposer une grande rigueur méthodologique et une saine humilité dans l'interpré-

tation des résultats. Certains philosophes s'appuyant sur ces ensembles de considérations [8] franchissent le pas de la perplexité vers des conclusions sceptiques plus définitives : l'expérimentation ne pourrait pas, ou très peu, contribuer à instruire les grandes questions de philosophie morale qui portent sur ce que font ou devraient faire les humains en général.

6.1.2.2 Le problème de la circularité

Les arguments liés à la circularité s'appuient sur la question de la réflexivité : comment l'homme pourrait-il se penser lui-même ? Cet argument a des racines vénérables dès les origines de la philosophie : dans le mythe de la caverne de Platon, les hommes enchaînés ne perçoivent du monde que ce que la lumière du feu projette au fond de la caverne et il faudrait qu'ils puissent se libérer de leurs chaînes pour pouvoir se rendre compte de l'étroitesse de leur point de vue. De la même façon, la connaissance que nous pouvons atteindre, grâce à notre entendement, sur le monde en général et sur notre entendement en particulier est limitée par ce que peut connaître cet entendement. On peut résumer d'une formule cette limitation : si le cerveau a le défaut de ne pas voir ses défauts, alors il ne peut voir le défaut de ne pas les voir.

L'argument de la circularité peut être décliné en deux variantes. Une variante sceptique, illustrée par H. Putnam avec l'image du « cerveau dans la cuve » (Putnam 1981) [187] et popularisée par le cinéma avec Matrix : imaginons un cerveau dans une cuve dont toutes les relations au monde extérieur seraient remplacées par des connexions à une machine. Il vivrait alors dans un monde virtuel créé par cette machine et serait incapable d'accéder au monde réel. Tout ce qu'il ressent, pense, croit, sait, ne serait qu'illusion construite par la machine sans aucun rapport au monde réel externe. L'indiscernabilité fondamentale supposée entre la situation de ce cerveau dans une cuve et celle du même cerveau dans un monde réel est à la racine de l'argument sceptique de la circularité.

Une seconde variante est issue du naturalisme [9] : tous les systèmes naturels que nous observons autour de nous sont finis et ont des capacités limitées. Le cerveau est un système naturel limité (poids, taille, énergie,...) issu de l'évolution [10]. Il est donc fortement probable qu'il soit également limité dans ses capacités [11]. Si ces limites existent, ce que le naturaliste

8. Voir (Kahane 2015) [126] pour un bon exemple d'article niant tout intérêt aux études de philosophie expérimentale sur les cas du tramway

9. Le naturalisme consiste en deux thèses : d'une part la nature est tout ce qui existe et d'autre part les sciences naturelles nous offrent le meilleur accès possible à la connaissance de cette nature (Andler 2016) [3]

10. Notons que si le lecteur préfère une autre théorie à celle de l'évolution, l'argument est inchangé : tout ce que nous avons autour de nous a des capacités limitées et envisager qu'il en aille autrement pour le corps ou l'esprit humain demandera un très haut niveau de justification.

11. Voir la formulation que donne Noam Chomsky de cette limitation dans (Chomsky et Calvé 2016) [46].

semble devoir accepter, nous ne les connaissons pas en totalité [12] et peut-être ne pouvons-nous pas les connaître. C'est cette possibilité d'objets qui nous seraient inaccessibles que le philosophe se doit de soulever, comme le propose Timothy Williamson [13]. Lorsque la philosophie propose des dilemmes moraux dont les paradoxes semblent échapper à notre entendement, c'est peut-être l'indice qu'elle se rapproche de ces zones d'ignorance [14].

Cette variante naturaliste de l'argument de circularité, moins extrême que la variante sceptique, a néanmoins des conséquences drastiques sur la portée de toute approche empirique qui serait, par construction, limitée aux notions qui nous sont accessibles. Elle présente toutefois l'opportunité d'imaginer des approches empiriques, observations ou expérimentations, qui pourraient fournir des indices de ce qui est accessible ou non aux capacités humaines.

6.1.2.3 Les réponses à l'argument de complexité

Face aux doutes liés à la complexité du comportement moral humain, le philosophe moral expérimental pourra développer plusieurs contre-arguments. Citons-en trois. Premier et principal contre-argument, celui de la banalité du constat de la complexité. Certes, le monde est complexe et l'homme en fait partie. Mais de nombreux aspects particuliers de ce monde complexe ont commencé à être étudiés en s'appuyant sur la démarche des sciences expérimentales. Il serait difficile de soutenir que c'est sans résultat, et encore plus difficile de soutenir que ces résultats, bien qu'imparfaits, incomplets et révisables, sont sans intérêt.

Il a fallu deux millénaires pour passer du niveau de connaissance des premières philosophies naturelles à la physique d'aujourd'hui. Celle-ci est certes incomplète, peut-être même incohérente [15], mais elle permet un niveau de compréhension des phénomènes observés et de création de nouveaux phénomènes qui ne pourrait s'expliquer si la démarche expérimentale ne permettait pas à notre entendement individuel et collectif de mieux et plus profondément décrire le monde qui nous entoure [16].

Pour éviter la contagion du scepticisme depuis les questions morales vers un scepticisme

12. Les travaux sur les biais des raisonnements humains en donnent une première approche (Kahneman, Slovic et Tversky (eds.) 1982) [127].

13. Dans (Williamson 2007 p 17) [235], l'auteur évoque les « elusive objects » qui seraient inaccessibles à notre perception comme à notre entendement.

14. Il est significatif à ce propos qu'on ait pu proposer de définir la philosophie analytique comme l'étude des paradoxes (Franceschi 2005) [90].

15. L'incompatibilité actuelle entre les théories quantiques et la relativité générale est un marronnier inévitable des propos sur la science. Elle peut en elle-même appuyer aussi bien des positions métaphysiques anti-réalistes que des positions positivistes affirmant la nécessité d'accroître l'effort scientifique avec des budgets supplémentaires pour découvrir la prochaine théorie. Je ne l'utilise ici que sous l'angle du constat épistémique : même les domaines de la connaissance empirique qui semblent les plus avancés n'atteignent pas à la perfection.

16. Je reprends ici l'argument du non-miracle détaillé par exemple par Sokal et Bricmont dans (Parris, Sokal et Bricmont 1999) [179]

généralisé à la démarche expérimentale elle-même, on pourra arguer que le niveau de complexité de ce qui a été découvert à ce jour dans les sciences naturelles est bien inférieur à celui qu'il conviendrait de maîtriser pour aborder le comportement moral humain. Nous pouvons accepter ce point et maintenir néanmoins qu'il ne s'agit que d'une question de degré et, peut-être, de temps : donnons nous un ou deux millénaires de plus pour avancer et la complexité seule ne sera peut-être plus un obstacle. Par ailleurs, et plus immédiatement, fixer une limite supérieure à la complexité que peuvent aborder les sciences expérimentales est un pari négatif qu'il semble bien difficile d'argumenter. En ce sens, le constat de complexité est un appel à l'humilité, la patience, la modestie et, certainement, à la prudence mais, en aucune façon, une raison de ne pas entreprendre et de se priver pour l'étude du comportement moral humain de ce qui a plutôt bien fonctionné ailleurs.

Le contre-argument de la banalité de la complexité commune à tous les domaines de la connaissance n'instruit aucunement la possibilité de répondre de façon définitive et absolument fondée à toutes les questions posées par la philosophie morale. A l'identique des autres domaines de la connaissance susceptibles de mise en œuvre de la démarche scientifique, cet argument stipule simplement que, comme dans ces autres domaines, il n'y a pas de limite connue a priori au niveau de complexité que cette démarche peut aborder et que, par ailleurs, nous ne pouvons affirmer, également a priori, que les connaissances qui seront acquises seront sans intérêt [17].

Le deuxième contre-argument est une déclinaison du premier dans le contexte de la psychologie. Affirmer que la complexité du comportement moral met son étude hors d'atteinte des méthodes expérimentales ouvre une question délicate quant à l'extension de ce constat à l'ensemble de la psychologie avec trois possibilités. Première possibilité, la conclusion sceptique s'étend à toute la psychologie et il faudrait renoncer à tous les acquis de cette discipline, passés et à venir, ce que peu de philosophes (et aucun scientifique) seraient certainement prêts à accepter [18]. Le doute porterait en effet alors sur la possibilité de toute science humaine et sociale et non sur la seule philosophie morale. Il rejoindrait ainsi la famille des doutes sceptiques qui, par leur trop grande généralité s'autodétruisent [19].

Deuxième possibilité, la conclusion sceptique ne s'étend pas à la psychologie en général mais reste cloisonnée au comportement moral car il serait particulièrement complexe en regard de l'ensemble des comportements humains. Cette alternative ne semble guère promet-

17. Et ce quelle que soit la définition de l'intérêt que chacun pourra adopter.

18. Ce point pourrait conduire à ouvrir la question de la possibilité même d'une science de la psychologie. Nous n'entrerons pas dans ce débat ici.

19. Suivons en cela (Williamson 2007) [235] où Thimothy Williamson suggère que le philosophe gagnerait à ne s'attarder que sur les arguments spécifiques et ne perde pas de temps et d'énergie à combattre des arguments sceptiques qui mettent en doute trop largement toute possibilité de connaissance.

teuse car on ne voit pas très bien en quoi les comportements justement dits de haut niveau comme la cognition, la planification à long terme ou la création artistique seraient de complexité moindre que le comportement moral. A minima, cette alternative supposerait de la part des philosophes qui la défendraient un investissement important pour instruire cette différence de complexité sans, naturellement, qu'ils puissent inclure dans leurs prémisses une différence de nature du domaine moral qui rendrait le raisonnement circulaire.

Enfin, troisième et dernière possibilité, le comportement moral ne serait pas plus complexe mais aurait d'autres attributs qui en rendraient l'étude plus difficile en regard de la psychologie générale. J'irai plus loin sur cette voie avec l'analyse de la circularité et de la normativité dans les paragraphes suivants, mais cette troisième voie me conduit également pour l'instant à relativiser le poids de l'argument de la complexité qui, à elle seule, ne construit pas de doute particulier en regard de la possibilité de soumettre le comportement moral à l'enquête empirique.

Illustrons le troisième contre-argument par un exemple : le cas très médiatisé des expériences de Libet [20]. En 1973 ce chercheur a montré que la prise de conscience d'un stimulus sensoriel n'avait lieu que plusieurs centaines de millisecondes après que les neurones moteurs se soient activés. Sur le plan psychologique, le résultat est d'importance, mais il l'est encore plus sur le plan philosophique : l'expérience a été interprétée comme pouvant conduire potentiellement à la remise en question du libre arbitre. Cet exemple illustre le troisième contre-argument opposé par les philosophes expérimentaux à l'attentisme que semble induire l'argument de la complexité : plutôt que d'attendre que des résultats comme ceux de Libet ne fassent irruption sur le terrain philosophique de façon impromptue, autant participer dès aujourd'hui à ces études en proposant et menant des expériences utiles aux débats philosophiques [21]. Ce contre-argument repose ainsi sur un constat quotidien pragmatique : toute avancée dans les sciences humaines, psychologie ou sciences cognitives, est reprise et commentée en fonction des conséquences philosophiques qui semblent pouvoir en être tirées. Il serait donc illusoire de la part du philosophe de se retrancher derrière l'argument de la complexité alors que ce mur est quotidiennement renversé par les interprétations des résultats d'expériences. Si l'argument de la complexité est en pratique impuissant à empêcher chacun d'interpréter les résultats scientifiques, alors il est certainement préférable que le philosophe s'implique dans toutes les étapes de la construction de ces résultats et, en particulier comme nous l'avons vu plus haut, dans l'importante analyse de l'opérationnalisation des concepts et

20. Voir le récapitulatif de ses travaux dans son ouvrage de 2005 (Libet 2005) [146]
21. Je reprends ici pour partie l'argumentation de (Cova et al. 2012) [57]

relations théoriques mis en œuvre.

6.1.2.4 La dérive de la définition de la complexité

Enfin soulignons qu'il est difficile de définir précisément ce qu'est la complexité et, tout particulièrement, quand les enjeux philosophiques sont importants. Prenons l'exemple de la complexité des jeux. Dans les années 70, lorsque l'ordinateur peinait à battre les bons joueurs de dame, le jeu d'échec était considéré d'une complexité inatteignable, exemple paradigmatique de la supériorité des capacités humaines. Quand l'ordinateur a battu le meilleur joueur d'échec dans les années 90, on a soutenu que c'était le jeu de go qui était la bonne référence pour un haut niveau de complexité. Aujourd'hui, cette étape également franchie, l'argument de la complexité demeure mais, comme nous n'avons plus de jeu plus complexe à soumettre à l'ordinateur, d'aucuns placent la limite entre jeu et vie réelle, sans pour autant que soient analysées les caractéristiques qui font que certains aspects de la vie réelle seraient simples et d'autres complexes. On peut avancer sans risque qu'il en ira de même dans tous les domaines à fort enjeu philosophique : la complexité devient par définition la caractéristique de ce qui n'est pas encore traité, et ainsi la phrase « la complexité ne sera jamais traitée » est analytiquement vraie.

Ce constat peut, paradoxalement, servir deux conclusions pratiques opposées. D'un côté, puisque la pression de l'enjeu philosophique n'a aucun apport au débat scientifique et, au contraire, peut conduire à diminuer les ressources allouées aux chercheurs en minimisant l'intérêt des résultats acquis ou à acquérir, le philosophe aura intérêt à laisser travailler les psychologues expérimentaux (ou les informaticiens dans l'exemple ci-dessus) et éviter toute intervention dans le processus scientifique. Inversement, dans une perspective agonistique du débat philosophique, puisque les philosophes acquis à l'argument de la complexité auront toujours la possibilité de la redéfinir dynamiquement, il est important pour les philosophes acquis à l'intérêt de la démarche scientifique de s'impliquer dans les expérimentations de façon à mettre en lumière les avancées obtenues. On retrouve ici deux options prises par les critiques ou les partisans de la philosophie morale expérimentale.

6.1.2.5 L'importance de la complexité pour la démarche empirique

La complexité du comportement moral humain est une évidence acceptée par tous. Nous avons illustré les conséquences de ce constat pour la philosophie expérimentale avec le cas de l'effet Knobe dans les chapitres précédents. L'attribution d'intentionnalité est un phénomène dont la complexité n'a pas permis, à ce jour, de produire une interprétation cohérente avec

tous les résultats empiriques partiels accumulés. Le psychologue se trouve ainsi confronté, du fait de cette complexité non résolue, à la double difficulté de l'interprétation inductive des conclusions de ses expériences et de la bonne opérationnalisation des concepts qu'il pense manier dans ces expériences.

Reprenons plusieurs constatations faites sur les articles récents évoquant l'effet Knobe. Principalement, l'intentionnalité est un concept complexe multiforme en relation fine avec tout un champ conceptuel incluant la responsabilité, la culpabilité, le hasard, la volonté, le libre arbitre, etc. Chaque expérience met en mouvement, explicitement ou implicitement, la totalité de ce réseau conceptuel mais les résultats ne font intervenir que quelques variables (dites à tort indépendantes) et quelques résultats statistiques plus ou moins significatifs, il est alors en pratique très difficile de conclure quoi que ce soit de constructif sur le phénomène dans sa généralité. On retrouve ici le lien entre opérationnalisation défaillante et interprétation inductive difficile.

Secondairement, et malgré sa difficulté, l'interprétation est entreprise par les chercheurs, psychologues et philosophes, pour de nombreuses raisons liées aux enjeux sociétaux, aux enjeux académiques et aux enjeux personnels. A l'échelle de la société, l'attribution d'intentionnalité est mobilisée dans de nombreux cas largement relayés dans la presse [22] et tout résultat partiel peut arriver de ce fait dans la lumière médiatique en tant qu'argument rhétorique d'un débat politique ou social. A l'échelle académique, la pression à publier contribue à la sur-interprétation des résultats dont le chercheur peut penser qu'elle va faciliter la sélection de son article. Enfin, à l'échelle individuelle, il est naturel que chacun souhaite voir dans ses travaux une contribution profonde aux conséquences morales importantes. La sur-interprétation morale des résultats d'expériences de psychologie semble donc être une tendance incontournable que les psychologues doivent affronter.

Les psychologues ont développé leur discipline en pleine conscience de ces difficultés, et la philosophie morale expérimentale peut bénéficier des avancées ainsi obtenues. En prolongement du chapitre sur l'opérationnalisation, j'insisterai ici sur l'apport que peut avoir dans ce domaine le concept d'opérationnalisation appliqué au domaine du comportement moral humain. Il s'agira concrètement de construire des savoir-faire pratiques permettant, par exemple dans le cas de l'effet Knobe, de capitaliser sur des protocoles transversaux à toutes les expériences portant sur l'attribution d'intentionnalité. Ces protocoles sont susceptibles de rendre progressivement plus cohérents les résultats sur chacune des variables envisagées en

22. On peut penser ici au fameux « responsable mais pas coupable » du ministre de la santé pendant l'épisode du sang contaminé.

se substituant aux méthodes ad-hoc réinventées à chaque expérience. Ce caractère ad-hoc ne permet pas, faute d'opérationnalisation réussie, d'établir les itérations épistémiques d'amélioration progressive des théories, avec leurs concepts et relations, et de l'expérimentation, avec ses pratiques et instrumentations. Comme nous l'avons vu avec la métaphore de l'hélice, la démarche scientifique est bonifiée lorsque le domaine expérimental prend une certaine indépendance, temporelle et de méthode, avec chacun des domaines théoriques auxquels la pratique apporte sa contribution empirique.

L'argument de la complexité n'a donc rien de particulièrement spécifique à l'étude du comportement moral humain. Il est, et aucun scientifique ne le niera, justifié de considérer tous les comportements humains de haut niveau comme particulièrement complexes, et le comportement moral en fait partie au même titre que la création artistique, la planification à long terme ou l'inventivité mathématique. Pour aborder tous ces sujets et respecter le constat de complexité, chaque discipline, dont au premier rang la psychologie [23], doit adapter la démarche empirique à son domaine. On peut souligner que cette préoccupation comportera naturellement la nécessité d'être patient et modeste à l'heure d'interpréter les résultats empiriques en terme moraux. Comme je l'ai montré précédemment avec l'exemple de l'attribution d'intentionnalité et l'effet Knobe, l'opérationnalisation de ces champs conceptuels est très difficile et, par conséquent, l'interprétation inductive des résultats problématique.

Ce que montrent les arguments et contre-arguments ci-dessus est que ces difficultés, réelles et quotidiennes, liées à la complexité ne sont néanmoins pas de nature à rendre impossibles les apports empiriques. Ils appellent plutôt à la plus grande prudence dans leur interprétation. Je passe maintenant à un autre type de critiques liées à la circularité de l'homme se pensant lui-même qui semble soulever des questions plus difficiles.

6.1.2.6 Les réponses à l'argument de la circularité

L'argument de la circularité, dans sa formulation sceptique extrême, celle du cerveau dans la cuve, est logiquement imparable mais sa portée s'étend de façon très large à toute connaissance sur le comportement humain et même au-delà, à toute connaissance humaine. On peut donc lui opposer, comme je l'ai fait plus haut pour celui de la complexité, sa non spécificité et, par tant, comme le suggère Timothy Williamson, ne pas lui accorder trop de poids. Il est très affaibli par son caractère extrême.

Une autre ligne de défense est de souligner que l'argument du cerveau dans la cuve est

23. Et on peut citer ici toutes les sciences qui ont un sujet en lien avec le comportement humain, l'archéologie, l'histoire, la géographie, les neurosciences, la biologie, la linguistique...

purement théorique, c'est une expérience de pensée. Aucune expérience de ce type n'est réalisée ni, selon notre conception actuelle de la genèse et du fonctionnement de la cognition humaine, réalisable car elle oblitère la dimension corporelle indispensable à la structuration et à la vie de notre système nerveux central. Cette ligne d'argument permet d'un côté de diminuer l'importance allouée à l'argument sceptique de la circularité et conduit d'un autre côté à introduire un aspect constructif qui détruit la circularité même : un système dédié à la connaissance ne peut se construire qu'en regard de ce qu'il est censé connaître et l'explication la plus simple à ses éventuels succès est qu'il soit en relation de confirmation avec le monde extérieur. Il y a une antinomie entre la capacité d'apprentissage, qui suppose cet accès au monde, et l'argument de la circularité dans ses formes les plus extrêmes, qui suppose l'isolement. Le sceptique pourra toujours rajouter que notre connaissance de la façon d'acquérir la connaissance est elle-même le résultat d'une illusion créée par le malin génie (ou l'ordinateur) qui manipule le cerveau dans sa cuve et que cette antinomie n'est donc pas plus réelle que l'accès au monde qu'elle prétend prouver. Arrêtons ce débat ici, qui ne peut être tranché par des arguments de logique, et retenons que le scepticisme extrême ne peut être utilisé contre la seule connaissance de la psychologie humaine, pour être efficace, il doit s'étendre à toute connaissance et, en conséquence, est d'une pertinence limitée pour mon propos.

L'argument sceptique de la circularité dans sa variante naturaliste a bénéficié d'un appui très fort au travers de l'étude des biais cognitifs qui a connu un développement important à partir des années 1970, avec en particulier les études de Daniel Kanheman : la cognition humaine est loin d'être parfaite et de nombreux biais conduisent à des raisonnements entachés d'erreurs systématiques, et ce tout particulièrement lorsque la perception que nous avons de nous-même est sollicitée [24].

Ironiquement, ce sont ces mêmes études qui donnent la meilleure réponse contre l'argument de la circularité : puisque les études empiriques ont permis de mettre à jour tous ces biais, et permis également de commencer à construire des stratégies d'évitement de leurs conséquences néfastes, pourquoi cette stratégie gagnante ne donnerait-elle plus dans le futur d'autres résultats significatifs en regard des problèmes posés par l'étude de l'homme par l'homme ?

24. Le Cognitive Bias Codex, diagramme de présentation de ces biais est accessible sur `https://upload.wikimedia.org/wikipedia/commons/a/a4/The_Cognitive_Bias_Codex`. Il montre le foisonnement de ces études en un graphique efficace.

6.1.2.7 Circularité et expérimentation

Dans les études psychologiques bien menées, la circularité est rompue de plusieurs façons qu'il est utile de rappeler. Premièrement, le passage à la troisième personne : l'homme ne se connaît plus lui-même mais le chercheur observe le participant se connaissant lui-même. Le chercheur devient ainsi, au moins en principe, interchangeable avec un autre chercheur ayant les compétences requises et le participant est assimilé à un instrument de mesure de ses propres états mentaux, instrument qu'il convient de calibrer et mettre en doute si nécessaire [25].

Deuxièmement, la réplication par des équipes différentes en contextes psychologiques variés, tant des participants que des expérimentateurs. La réplication a de multiples avantages que nous avons vu plus haut avec le cas du « Implicit Association Test ». Elle a ici, en regard du risque de circularité, celui d'enrayer les risques liés à l'entre-soi d'une équipe de chercheurs publiant pour la première fois une découverte.

Troisièmement, les analyses trans-culturelles. Un risque important lié à la circularité est la possible confusion de prendre pour un phénomène psychologique général ce qui est en réalité un phénomène culturel local. Pour éviter ce risque, les chercheurs répliquent l'expérimentation au sein de plusieurs cultures [26].

Quatrièmement les études trans-espèces. Dans la mesure où le chercheur souhaite, par exemple, tester la validité d'une théorie évolutionniste pour un trait particulier, il aura à analyser si on observe ce trait sous sa forme complète ou sous une forme approchante dans les lignées phylogénétiquement proches de notre espèce, par exemple le chimpanzé. Ce dispositif rompt la circularité de l'homme s'étudiant lui-même avec l'homme étudiant d'autres espèces. On peut toutefois remarquer que beaucoup de chercheurs postulent la spécificité du comportement moral humain, ce qui rend cette approche plus difficile à mettre en œuvre pour la philosophie morale que pour, par exemple, l'étude de la perception où les chercheurs postulent plutôt la continuité du trait.

Cinquièmement, les multiples corrélats biologiques et psycho-physiques objectivables des états mentaux sont une composante importante de la psychologie moderne (IRMf, temps de réaction, dilatation de la pupille, comportements, études cliniques des lésions du cerveau, des troubles psychiques...). Ces dispositifs ont pour effet d'interposer entre le chercheur et

25. Sur l'importance épistémologique du passage de l'observation de la première à la troisième personne voir (Piccinini 2009) [181] et (Ludwig et Michel 2019) [148].

26. Il est habituel de répliquer l'étude dans le monde occidental et dans le monde oriental. Il est non moins classique de reprocher à ce dispositif d'oublier les territoires non couverts par des équipes académiques dans les pays économiquement moins favorisés, de sous-estimer que ni « occidental » ni « oriental » ne sont des catégories bien définies et enfin d'oublier que la culture commune au monde développé actuel est très loin des cultures passées, dont celles qui prévalaient aux débuts du développement de notre espèce.

l'homme étudié un ensemble complexe de données, de techniques, de documents, de traces pour reprendre le terme de Ian Hacking, qui permet une distanciation propice à la décircularisation de l'étude de l'homme par l'homme.

L'ensemble de ces stratégies, menées par des humains, ne saurait offrir une garantie absolue de résultat. On peut en revanche considérer que, menées efficacement et conjointement, elles ramènent le problème de la circularité à celui de la complexité et nous conduisent ainsi aux mêmes conclusions. Les arguments sceptiques pourront toujours être reformulés dans le contexte de chaque étape de la démarche scientifique et aucune réponse absolue ne sera recherchée. Mais, inversement, on voit mal pourquoi il serait pertinent de fixer a priori une limite à la connaissance des faits humains par l'espèce humaine.

Soulignons à nouveau que les deux premières lignes d'argumentation, la complexité et la circularité, concernent potentiellement toute activité humaine visant à connaître le comportement humain. Elles ne sont propres ni à la philosophie morale expérimentale ni à la philosophie morale analytique et vont, si on les poursuit, conduire à douter aussi bien de l'ensemble des SHS, de la philosophie en fauteuil ainsi que de toute réflexion sur l'humain, a priori et a posteriori. Il est ici utile de rappeler aux philosophes détracteurs de la philosophie morale expérimentale que certains de leurs arguments peuvent s'avérer trop puissants et emporter avec eux toute la philosophie morale. [27]

Quelles conclusions utiles à notre propos sur l'expérimentation tirer de cette analyse des deux premiers types d'arguments en défaveur de l'étude expérimentale du comportement moral humain ? Principale conclusion, peu controversée, le sujet est complexe et appelle humilité et patience pour éviter toute interprétation hâtive de résultats empiriques dans la sphère philosophique. Ensuite, et ce sera plus débattu par les philosophes moraux, les dispositions à prendre en regard de la complexité du comportement moral n'ont rien de particulier en regard des autres comportements objets habituels de la psychologie. Enfin, et ce sera plus débattu par les philosophes expérimentaux, les dispositions à prendre en regard du risque de circularité sont celles, connues des psychologues, qui mettent de l'intermédiation et de la diversité au sein du processus expérimental en multipliant les réplications d'expérimentations au sein et entre les équipes, intra et inter-culturelles, intra et inter-techniques et intra et inter-espèces. Je reviendrai sur ce point ultérieurement, mais soulignons ici que cette nécessaire diversité des approches conduit à la nécessaire diversité des méthodes et, donc, des intervenants pour des raisons pratiques d'organisation des compétences et des formations ainsi que pour des raisons académiques du fait de la diversité des disciplines impliquées. Comme nous l'avons

27. Le doute sceptique est un acide fort, comme le rappelle Claudine Tiercelin (Tiercelin 2005) [227]

vu précédemment avec l'exemple de l'étude des larmes, complexité et circularité militent pour un découpage des compétences et des équipes pour que l'opérationnalisation soit le résultat d'une capitalisation transverse des compétences acquises.

Éviter les écueils liés à la complexité et à la circularité passerait donc non pas par une unique meilleure méthode, un unique meilleur intervenant ni même une unique solution d'organisation de la recherche, mais supposerait de multiplier les acteurs et les méthodes. Dans cet esprit, la participation des philosophes au dispositif expérimental n'est ni à exclure, ni à privilégier, mais permet d'accroître cette diversité des organisation. Simultanément et pour la même raison, que les philosophes soient systématiquement seuls réalisateurs des expériences qu'ils contribuent à imaginer ne pourrait conduire qu'à une augmentation des risques de biais.

6.1.3 La normativité et le problème de Hume

6.1.3.1 Les problèmes liés à la normativité

Comme nous l'avons vu, les arguments précédents liés à la complexité et à la circularité ne sont pas spécifiquement liés au domaine moral. Ils concernent la connaissance du comportement humain en général. Je vais aborder maintenant les arguments appuyés sur la dimension normative qui serait par nature différente de la dimension descriptive qu'a naturellement l'enquête empirique sur les comportements humains. La dimension normative me semble plus importante que la complexité ou la circularité car plus spécifique à la philosophie morale expérimentale. La première difficulté que j'examinerai dans la présente section est d'ordre logique, ce qu'il est convenu d'appeler le problème de Hume. Dans la section suivante, j'examinerai la possibilité pour un chercheur de mettre un humain sous son microscope et, dans la troisième section, si le fait de rechercher une description empirique des règles morales peut nuire à leur justification, et porter ainsi un risque social.

Les deux difficultés, logiques et morales, liées à la normativité se combinent pour faire de la question normative un tout difficile à aborder. Il semble que nous ayons d'un côté un monde abstrait fait de règles et de valeurs et d'un autre côté le monde concret dans lequel l'acteur moral agit sans qu'il puisse déduire simplement ce qu'est le Bien concret à partir des règles et valeurs qui l'animent. Et, inversement, l'observation des agents et de leurs actions dans le monde concret ne constituerait pas une base utile à décrire le monde des normes et des règles morales. Entrons plus avant dans les termes de ce débat, tout d'abord avec le problème de Hume, le passage impossible du descriptif au prescriptif, et ensuite je

poursuivrai avec l'engagement moral du chercheur puis, après un détour par un examen de la notion de norme, avec l'engagement social de la recherche en psychologie morale, tels que je viens de les aborder.

Commençons par constater que la définition de la normativité, des normes et des valeurs qui les sous-tendent dans leur relation au comportement moral n'est pas tâche facile et je la reprendrai plus loin. Proposons pour l'instant une définition naïve pour entrer dans le débat : une théorie morale est constituée d'un ensemble des règles générales sur le Bien et le Mal qui aident chacun à instruire des jugements dans les cas particuliers auxquels il est exposé.

Cette définition minimale, dont chaque terme pourrait être, et a été, débattu par les philosophes moraux suffit à faire apparaître deux difficultés pour qui souhaiterait s'appuyer sur l'observation ou l'expérimentation pour étudier ces règles morales. La première est d'ordre logique, et on la dénomme habituellement le problème de Hume [28] : « De ce qui est on ne peut déduire ce qui doit être ». Plus précisément, Hume affirme que pour que la conclusion d'un raisonnement logique puisse être prescriptive, il faut qu'une des prémisses le soit également, ou formulé inversement, si toutes les prémisses d'un raisonnement logique sont descriptives, alors la conclusion ne peut être prescriptive. Reformulons cette règle avec un exemple : l'observation répétée de l'homicide ne nous dit rien des règles morales qui l'interdisent. On peut considérer le sophisme naturaliste (naturalistic fallacy) dénoncé par G.E. Moore dans ses Principia Ethica comme une autre formulation de la même difficulté (Moore, Gouverneur et Ogien 1998) [165].

6.1.3.2 Le problème de Hume

L'impossibilité de déduire ce qui doit être du point de vue moral de ce qui est a tout d'une évidence logique. L'observation, l'expérimentation donnent un accès à ce qui est, et ce que le jugement moral en soit positif ou négatif. L'observation du Bien et du Mal, tous deux également présents dans le monde, ne peut instruire le jugement moral qu'il convient de porter sur chacun. Le passage du descriptif vers le prescriptif apparaît comporter un obstacle conceptuel infranchissable.

Un traitement détaillé du problème de Hume est un passage obligé de toute réflexion méta-éthique [29]. L'argument porte sur plusieurs plans. Pour Hume il est inconcevable de passer d'une phrase disant ce qui est, par exemple « Dieu existe » à une phrase sur ce qui doit

28. Pour une remise en perspective de cet argument avec la philosophie de Hume voir (Nurock 2011).[173], malgré la mise en doute portée par cette perspective, je garde ici conformément à la littérature l'appellation « problème de Hume » pour le passage logiquement problématique du descriptif au prescriptif.

29. Voir par exemple celui proposé par Vanessa Nurock dans (Nurock 2011) [173] ainsi que dans (Aubé 2017) [14] ou (Kitcher 2014) [134].

être « Tu dois aimer Dieu », ce glissement souvent utilisé mais rarement explicité fait passer d'un certain type logique à un autre type logique et n'est pas justifié. De façon proche, Moore considère qu'il est impossible de définir le bien par des caractéristiques non morales. Pour le démontrer, il propose de convenir qu'une telle caractéristique C existe, il suffit alors de remarquer que la question « telle chose qui est un C est-elle vraiment bonne ? » est parfaitement compréhensible ce qui prouve que « être un C » et « être bonne » ne sont pas synonymes.

Le problème de Hume semble important pour notre sujet : si aucune conclusion prescriptive ne pouvait sortir d'une démarche empirique, alors le terme « philosophie morale expérimentale » serait un oxymore. [30]

6.1.3.3 Les réponses au problème de Hume

Plusieurs arguments ont été développés pour tenter de montrer que cette impossibilité peut être contournée dans certain cas. Ainsi, John Searle a proposé que l'expression d'une promesse, ce qui est bien un fait observable, crée pour celui qui l'a prononcée une obligation morale de l'honorer qui est bien un fait normatif (Searle 1964) [203]. Cet argument est toutefois assez faible puisque le caractère normatif est déjà présent dans la notion de promesse avant même que quelqu'un ne la prononce, l'argument est ainsi qualifiable de circulaire : le lien entre les dimensions descriptives et normatives existe dans l'exemple de Searle s'il existe, et parce qu'il existe, déjà dans un des termes prononcés.

D'autres philosophes ont insisté sur la difficulté à formuler précisément la loi de Hume (G. E. M. Anscombe 1958) [7], et cette difficulté pourrait être liée à la circularité qu'elle renferme : pour savoir ce qui est « prescriptif », il faut déjà avoir une définition de la morale et du comportement humain de façon à pouvoir approcher ce qui peut être autorisé ou interdit, mais d'où proviendraient ces définitions si aucune approche descriptive ne peut apporter d'élément ? (Aube 2017 page 9) [14].

On peut par ailleurs remarquer que le problème de Hume ne se pose pas dans les mêmes termes selon notre conception morale. Si on suppose que la morale est simplement une façon de parler, sans lien substantiel avec ce qui est jugé, alors le problème de Hume devient une évidence : les relations au sein d'une société d'individus en interaction qui échangent des jugements moraux ne sont pas déductibles d'une observation tierce des choses jugées. Un certain réalisme moral est donc un préalable nécessaire pour que le problème de Hume puisse être utilement posé.

30. Pour une discussion récente des arguments déniant la signification morale des résultats des expérimentations voir (Sauer 2018) [201]

Un argument empirique peut également être utilisé contre la thèse de Hume. Il est fréquent qu'une institution ou une coutume soit dite « bonne » si elle conduit à des conséquences observables positives. Dans ce cas on a bien une description, celle des conséquences de la disposition adoptée par l'institution, qui conduit à une prescription, celle de la mise en œuvre de la disposition. Hume répondrait certainement qu'il s'agit d'une généralisation abusive et que l'observation du succès ici et maintenant n'est nullement garantie de succès là-bas et demain, et donc que le passage au prescriptif est bien une erreur de logique.

Mais, acceptons ici que le problème de fond reste entier, qu'il semble bien que l'apport empirique ne puisse permettre à lui seul de définir ce qui doit être. Il n'en reste pas moins que, inversement, aucune affirmation prescriptive ne peut être portée sans que soient définis les objets sur lesquels elle porte : acte, acteur, circonstances, ... Cette affirmation comporte donc une part importante qui ne peut résulter que de la description du monde qui nous entoure.

Ce double constat me semble pouvoir être utilement éclairé par le rapprochement avec la thèse de Duhem-Quine de la sous-détermination des théories par les données empiriques. Proposons pour ce rapprochement une analogie entre vrai et bien, en tant qu'adjectifs, et entre la Vérité et le Bien en tant que substantifs. Dans les deux cas, nous sommes confrontés au problème de l'induction de résultats empiriques, et donc limités, vers des conclusions absolues illimitées. Dans les deux cas nous ne pouvons accéder au Bien et à la Vérité par des voies empiriques seules, et dans les deux cas, une vision déflationniste considère que ces substantifs sont abusifs et que seuls les adjectifs sont significatifs. La qualification de bien ou de vrai correspond, dans cette vision pragmatique, à un accord que des interlocuteurs peuvent établir au mieux de leurs capacités et de leurs ressources quand elles ont des conséquences observables. Avec une telle vue, le problème de Hume et la thèse de Duhem-Quine se rapprochent et perdent de leur importance au profit d'un examen des conditions, empiriques, nécessaires aux interlocuteurs pour arriver à un accord utile. Les deux problèmes sont alors perçus comme des effets de langage liés au passage abusif de l'adjectif au substantif[31].

6.1.3.4 Le problème de Hume et l'expérimentation

Si, malgré les réponses que je viens d'évoquer, nous acceptons le point de logique, il ne concerne que l'impossibilité de dériver ce qui doit être de ce qui est. Cette restriction est loin d'épuiser tous les apports possibles de l'approche empirique à l'étude du comportement moral.

Tout d'abord, l'expérimentation et l'observation permettent de cerner le problème posé.

31. Le biais consiste ici à confondre un niveau ontologique et un niveau grammatical, le terme de substantif, avec sa racine « substance », montre le chemin de ce biais.

C'est l'observation des crimes, des forfaits, des trahisons qui conduisent à poser des questions morales sur l'interdiction de tels actes. Ainsi le plus vieux code connu, le code d'Hammurabi [32], procède par une liste de cas décrits et portant condamnation. Si un problème ne se pose pas dans la réalité d'une société, il serait de peu d'utilité qu'une règle morale lui soit appliquée, si ce n'est à titre métaphorique ou poétique (on peut penser par exemple à une règle interdisant l'inceste chez les anges).

Une approche empirique de ce qui est nécessaire ou contingent, possible ou impossible, permettrait également de limiter la charge de travail des philosophes moraux. On peut avancer qu'une règle morale ne pourra ainsi préconiser l'impossible. Ce que le proverbe traduit par « à l'impossible nul n'est tenu » et que le philosophe transcrit en « devoir implique pouvoir » [33]. On peut également s'interroger sur une l'utilité d'une règle morale qui préconise le nécessaire, bien que, selon la théorie morale qu'on adopte, il puisse être moralement différent d'entreprendre une action sans autre raison que l'impossibilité de l'éviter ou parce qu'elle est, en plus, moralement bonne.

La description des comportements actuels en regard d'une règle morale peut également permettre de mesurer la distance entre la situation réelle existante et l'idéal que définirait cette règle. On peut alors envisager l'effort global demandé pour des actions réformatrices. On peut simuler de telles actions et en mesurer l'efficacité ainsi que les éventuels effets pervers, sur le plan empirique comme sur le plan moral. De même, l'expérimentation pourra permettre de savoir si une règle morale est connue ou non, acquise ou non, facile ou difficile à appliquer pour un individu dans une culture donnée et, ainsi, évaluer l'effort que demande la généralisation de son respect.

En méta-éthique, la comparaison des différentes cultures permettra d'évaluer les différentes théories sur les fondements ou sur la genèse des règles morales, sans bien sûr avoir à trancher sur le fond de leur validité, ce qui supposerait de disposer d'une justification de niveau moral à cette justification, et impliquerait un raisonnement circulaire. En ce sens, le problème de Hume a prise sur l'éthique substantielle mais pas sur la méta-éthique. Notons par ailleurs, et j'y reviendrai plus loin, que le problème de Hume n'a également pas prise sur l'éthique appliquée, celle-ci n'ayant pas à définir des règles générales mais seulement ce qu'il convient de faire ici et maintenant.

Les résultats expérimentaux potentiels, limités par le critère de Hume, peuvent paraître bien insuffisants en regard des problèmes posés par la philosophie morale substantielle. Ils le

32. Code Babylonien daté de 1750 BC gravé sur une pierre conservée au musée du Louvre
33. Cette impossibilité d'attribuer une responsabilité morale n'est pas équivalente à une impossibilité de blâmer. Voir par exemple (Chituc et al. 2016) [45]

sont effectivement, même si savoir ce qui est possible et savoir ce que coûte de changer ne sont pas des résultats négligeables. Ils sont en revanche plus prometteurs pour la méta-éthique, surtout en ce qui concerne l'évaluation des scénarios ayant conduit à la genèse des différentes règles morales. Je développerai l'analyse des justifications évolutionniste de l'éthique dans une section suivante.

6.1.4 L'impossibilité morale d'expérimenter

6.1.4.1 Prétendre expérimenter sur l'homme peut-il être moral ?

La difficulté que je propose d'appeler le problème de l'engagement, est d'ordre pratique : expérimenter c'est déjà faire. L'engagement comporte deux dimensions selon que l'on examine une expérience particulière, l'engagement moral d'un chercheur, ou que l'on examine la société à plus grande échelle, l'engagement social d'un programme de recherche en psychologie morale, que j'aborderai à la section suivante.

Les expérimentations de psychologie morale peuvent par exemple consister à mettre des participants en situation de choisir des actions bonnes ou mauvaises, d'observer ce qu'ils décident, et de tenter de comprendre comment et pourquoi ils font ce choix. Un humain, l'expérimentateur, s'octroie ainsi le droit moral de mettre un autre humain sous son microscope. On pourra ici rappeler les nombreuses règles morales que cette dissymétrie viole, ou du moins pourrait être accusée de violer, citons-en deux :

— La règle d'or de la réciprocité : négativement, « ne fais jamais à autrui ce que tu ne souhaites pas que l'on te fasse » ou positivement « fais à autrui ce que tu voudrais que l'on te fasse » [34].

— La règle de la dignité humaine [35] :« ne considère jamais autrui comme un moyen mais toujours comme une fin ».

Ce problème est tout particulièrement apparent quand on observe que plusieurs expériences classiques de psychologie morale réalisées dans la première moitié du vingtième siècle comportaient des risques significatifs et ne pourraient plus être menées aujourd'hui, indépendamment de leur intérêt scientifique [36]. Le risque est en pratique, et institutionnellement, reconnu puisque des comités d'éthique ont été créés dans chaque organisation de recherche. Ils

34. Anecdotiquement, lors d'une réunion de doctorants en philosophie, j'ai mis en scène une pseudo expérience de philosophie morale en posant des questions sur un sujet donné et exploité les réponses sur un autre trait psychologique supposé ainsi caractériser les présents, et c'est ce caractère de tromperie systématique lié aux enquêtes psychologiques qui a été retenu comme significatif de la démarche et, à ce titre, mis en question.

35. Cette règle ouvre la difficile question que je n'aborderai pas ici de l'extension à donner à l'ensemble des être ayant partiellement ou totalement cette dignité : genre, statut, espèce,... et des critères à prendre en compte pour cela.

36. On peut penser ici aux expériences de Milgram sur la soumission à l'autorité ou à l'expérience controversée de Stanford sur le monde carcéral

sont chargés d'évaluer et d'autoriser ou non les expériences. L'existence même de ces comités d'éthique montre que la résolution des conflits de valeur entre recherche de la connaissance et respect des règles morales ne va pas de soi. La recherche de connaissance sur les règles morales peut ainsi s'en trouver entravée pour des raisons de conformité aux règles morales.

Avec l'instauration des comités d'éthique, la question se déplace de la moralité du chercheur vers la question du statut moral de ces comités. Quelles peuvent être les bases morales des décisions qu'ils prennent ? Si elles sont expérimentales, elles résultent d'expériences autorisées et s'insinue un risque de circularité. Si elles proviennent de théories morales, alors elles sont le reflet des choix théoriques des membres et leur validité est limitée. Je reviendrai plus loin sur la démarche de l'équilibre réfléchi mise en œuvre, en pratique, dans ces comités.

Le chercheur pourra arguer que sa visée n'est pas simplement de connaître le comportement humain mais également de le soigner et même de l'améliorer. Il lui sera alors rappelé le mythe de Prométhée : l'humain n'est pas à sa place quand il veut voler aux dieux leurs capacités (Sandel 2002)[199].

6.1.4.2 Quelles réponses à l'impossibilité morale ?

Lorsqu'on aborde l'examen du domaine moral, et que l'approche en soit philosophique ou qu'elle soit scientifique, il semble de bonne méthode de commencer par décrire les comportements qu'on cherche à étudier. Toute réflexion morale semble supposer de chercher à savoir quelle est la situation de départ et donc de chercher à décrire les comportements humains. Toute science expérimentale aura la même évidente nécessité. En ce sens, l'impossibilité morale qu'il y aurait à mettre des humains sous son microscope semble devoir être fortement modulée. L'observation de prime abord ne peut qu'être autorisée, et si difficulté ou interdiction il y a, ce serait pour des interventions violant la vie privée ou l'autonomie des participants.

Malgré cette apparente convergence, on peut craindre que cet accord de prime abord ne s'effrite rapidement lorsque, concrètement, il faudra définir par quelles questions commencer l'examen de l'existant moral. Il pourrait par exemple être indifférent au philosophe de l'impératif catégorique de savoir combien d'homicides ont lieu chaque année et dans quelles circonstances car, par construction de sa théorie, un seul serait déjà de trop et quantifier ou qualifier les homicides n'aura aucun intérêt moral [37]. Autre exemple, l'inventaire de la diversité des règles morales exprimées par les agents de différentes cultures sera certainement un point de départ important pour le scientifique pour bien cerner son domaine d'étude. Ce sera

37. Cet argument est régulièrement utilisé contre l'évaluation morale des crimes de guerre en fonction de la taille des charniers.

une question délicate qu'il faudrait probablement mieux différer pour les philosophes moraux substantiels qui auraient à expliquer pourquoi ce qui est bien ici est mal là bas.

En somme, et dès le premier moment de l'enquête empirique, celle-ci dépendra dans ses objectifs et dans son contenu des théories morales en place au sein des institutions de recherche. En ce sens, il ne saurait y avoir de réponse au problème de l'engagement moral du chercheur qui soit générale et indépendante de la théorie morale des institutions dans lesquelles se développent ces recherches.

L'argument de la nécessité de connaître le comportement humain est une réponse qui semble aller de soi, mais à l'examen, elle met en avant une valeur épistémique qui serait antérieure à toute préoccupation morale. Lancer une recherche pour acquérir de la connaissance suppose que cette connaissance a une valeur en soi qui justifie de ne se poser les questions morales que dans un second temps. Il s'agit ainsi d'un choix moral implicite. La démarche des comités d'éthique que j'ai évoquée plus haut revient à expliciter et mettre en débat ce choix implicite et, pragmatiquement, d'arbitrer au cas par cas les programmes de recherche et les expériences qu'il convient de lancer. Naturellement, cette solution ne saurait satisfaire aux exigences de toutes les théories morales et, en particulier, celles qui ne donneraient qu'une valeur nulle ou minime à la connaissance d'origine empirique et on voit se dessiner trois possibilités :

— Des théories morales extrêmes pour lesquelles l'apport empirique ne saurait justifier qu'un programme de recherche soit entrepris. L'humain ne peut être traité en cobaye.
— Des théories morales modérées pour lesquelles valeurs épistémiques et valeurs morales sont complémentaires et leurs apports relatifs doivent être arbitrés en cas de conflit.
— Des théories morales minimales mettant en avant les valeurs épistémiques, comme peuvent le faire certains scientifiques, et n'acceptant pas qu'un programme de recherche soit interrompu pour des raisons morales.

6.1.4.3 Responsabilité morale et expérimentation

A l'échelle particulière d'une expérience, le chercheur met un autre humain sous son microscope, instaurant une indéniable dissymétrie. Le choix pour le philosophe moral est alors entre considérer que des impératifs moraux catégoriques (dont la dignité humaine : être une fin et jamais un moyen) sont au dessus de tout objectif épistémique et que ces expériences de l'homme sur l'homme sont moralement condamnables, ou considérer que les circonstances priment : les priorités entre différentes règles morales et différents enjeux de connaissance

doivent être établies au cas par cas. Un comité d'éthique ou toute instance analogue aura la charge de cette décision. Ou enfin, troisième voie, considérer que les objectifs épistémiques priment et que toute expérience est justifiée si elle a pour objectif la connaissance.

Le choix entre ces trois options n'est évidement pas du seul ressort des scientifiques, ni même des philosophes, à l'intérieur de leurs périmètres respectifs mais relève de choix globaux de société. Notons néanmoins qu'il y a des affinités entre chacune des options et les grandes conceptions de philosophie morale. L'interdit catégorique de l'approche expérimentale suppose que la morale ait un fondement autre que naturaliste, par exemple surnaturel, mais inversement un fondement surnaturel n'implique pas logiquement l'interdit. Une morale orientée vers le naturalisme aura besoin de l'approche expérimentale pour exister et ne pourra donc l'interdire absolument, sans pour autant être amenée à ne pas l'encadrer par des règles circonstanciées. La troisième option, la prééminence de la recherche sur toute autre considération, nous ramène, d'une part, aux limites que la société se doit d'imposer aux démarches scientifiques et, d'autre part, à la réflexion sur le lien de l'expérimentation abusive avec la déshumanisation, ou la sous-humanisation, d'une partie de l'humanité sur laquelle faire des expériences serait alors hors du domaine moral. La réflexion morale sur les expériences abusives pourra, et certainement devra, s'étendre aux expérimentations dans le monde animal, au-delà du seul domaine humain.

6.1.5 La nocivité sociale de l'expérimentation

6.1.5.1 Expérimenter c'est socialement faire

Outre la dissymétrie morale entre le chercheur et le participant, objet de la section précédente, on peut craindre également les conséquences des études empiriques sur l'évolution morale d'une société par l'image qu'elles donnent du domaine moral. C'est alors l'engagement social de la recherche qui est en jeu. Comme le suggère la citation que j'ai mise en tête de ce chapitre, pour les philosophes moraux qui, comme Emilio Martinez Navarro, suivent la tradition des philosophes prescripteurs d'une certaine morale, il est insignifiant, et en fait inopportun, voire dangereux, de donner une description du comportement moral qui ne s'appuie pas essentiellement sur une justification de l'obligation de chacun à se conformer aux règles morales. Objection à laquelle on peut ajouter qu'une telle recherche s'accompagne souvent de l'explicitation des alternatives, or montrer que ces alternatives existent peut s'avérer dangereux pour qui ne défend qu'un seul livre [38]. Expliquer empiriquement ne pourrait, de leur

38. J'emprunte ici l'expression à Manuel Vazquez Montalban dans l'épilogue du Capitan Alatriste :
Desconfíen siempre vuestras mercedes de quien es lector de un solo libro.

point de vue, que nuire à la transcendance de l'obligation. Je reviendrai plus loin sur ce point : le phénomène moral n'existe que si les individus respectent les règles, et donc, pouvons-nous penser, s'ils croient qu'il faut les respecter et qu'elles sont pour eux justifiées. L'approche empirique, en privilégiant la description sur la justification, détruirait l'objet qu'elle cherche à décrire.

Les sciences physiques nous ont habitués à percevoir la recherche de connaissance comme sans effet sur la chose à connaître, le pendule oscille de la même façon que je l'observe ou non, et même si la physique quantique vient aujourd'hui mettre cette hypothèse en doute, il n'en reste pas moins que nous faisons souvent comme si observer un phénomène ne l'altérait pas. Or cette illusion de neutralité de l'observateur est la source de grandes difficultés dans nos sociétés qui exploitent la connaissance produite par cette démarche scientifique dans de très nombreux domaines d'activité. Ce constat se traduit par l'appel de nombreux philosophes à un contrôle démocratique, exogène à la démarche scientifique, des domaines de recherche : (Kitcher et Ruphy 2010) [135]. Ce problème est particulièrement important pour l'étude des règles morales. Mener une recherche en psychologie morale est déjà une action et change le monde qu'on aimerait simplement observer. A titre d'illustration extrême, prenons l'exemple des périodes sombres de l'inquisition, les règles morales sont alors affirmées d'origine divine et les mettre en cause relève du bûcher : il sera alors bien difficile de monter une chaire de philosophie expérimentale avec pour ambition de les confirmer ou de les infirmer empiriquement. [39]

Pour explorer cet argument, je dois préciser ce que sont les normes [40], et le lien complexe qu'elles entretiennent avec le comportement moral dans un contexte social donné.

6.1.5.2 Les normes et leurs liens au comportement moral

Abordons donc la notion de norme pour tenter d'affiner le rapport entre normes et domaine moral. De façon générique, la norme peut être descriptive, évaluative, prescriptive ou impérative. La norme est simplement descriptive quand elle exprime des habitudes, des tendances, par exemple vestimentaires, largement répandues ou, au contraire, minoritaires au sein d'une population. Elle devient évaluative quand sur la base de certains de ces détails vestimentaires on infère des conclusions sur la personnalité de celui ou celle qui les porte. La

39. On peut d'ailleurs remarquer que c'est pour cette raison que les grandes traditions morales religieuses ont souvent interdit ce type d'études de psychologie morale de peur qu'elles mettent en cause leur prérogative à dire le Bien.

40. Je parlerai ici de normes en simplifiant l'analyse qui aurait à inclure également, pour plus de profondeur, la notion de valeur. La norme serait alors la concrétisation dans un contexte donné et pour un type de situation de la notion plus abstraite de valeur.

norme est prescriptive quand elle conseille ou déconseille tel détail vestimentaire dans telle circonstance et peut devenir impérative quand elle l'interdit ou le rend obligatoire.

On distingue également plusieurs types de normes en fonction des attributs qu'elles évaluent, le beau, le vrai, le juste, le bien, ... On parle ainsi de normes esthétiques, de normes épistémiques ou de normes éthiques [41]. Je me concentre principalement dans le cadre de cette thèse sur les normes épistémique liées à la démarche expérimentale et sur les normes éthiques qu'étudie la philosophie morale [42].

Tenter de comprendre ce phénomène normatif pose de redoutables questions :

— Pourquoi sont-elles aussi omniprésentes ?

— D'où viennent-elles, comment se constituent-elles ?

— Pourquoi un groupe adopte-t-il une norme plutôt qu'une autre ?

— Pourquoi un individu accepte-t-il de s'y conformer ?

— Comment arbitrer entre normes contradictoires ? [43]

Mon objectif n'est pas ici de faire le point de l'ensemble de ces questions mais de reprendre tout d'abord le lien entre norme et comportement moral pour, ensuite, adopter l'angle de vue de la possibilité de l'étude scientifique et de l'expérimentation en philosophie morale.

6.1.5.3 Le contenu moral des normes

Une norme peut n'avoir aucun contenu moral. En particulier une norme simplement descriptive n'emporte pas obligatoirement de considération de bien ou de mal. Dans ce sens, tout objet de jugement peut ainsi être normal au sens descriptif d'un phénomène courant et néanmoins et simultanément contraire aux règles morales [44].

De même une norme évaluative ne sera morale que si la caractéristique évaluée l'est elle-même. La norme évaluative est significative de deux phénomènes psychologiques importants qui constituent conjointement un pilier de notre vie sociale. Le premier phénomène est la catégorisation, le second l'essentialisation [45]. Appiah appelle catégorisation notre tendance à créer des catégories à partir d'indices qui sont efficaces en cela qu'ils sont suffisamment conformes à une norme descriptive, et Appiah appelle essentialisation notre tendance à considérer que les individus qui sont regroupés par ces indices communs sont proches sur tous les

41. Pour une présentation du domaine normatif, voir (Bicchieri, Muldoon et Sontuoso 2018) [24]

42. Pour mémoire, cette multiplicité des normes ouvre des perspectives qu'il serait intéressant de reprendre à la lumière des résultats de la présente thèse, en particulier pour les normes esthétiques.

43. L'exemple de l'Implicit Association Test présenté plus haut m'a permis de montrer comment différents types de normes peuvent entrer en concurrence et comment, finalement, des normes de plus haut niveau sont nécessaires pour permettre à chaque groupe d'arbitrer en donnant la priorité à la dimension épistémique ou à la dimension éthique de chaque situation.

44. On peut penser à nouveau à l'homicide, événement très courant mais exclu par les règles morales.

45. Pour une analyse complète de ce point de vue voir (Appiah 2018)[10].

autres plans car ils partagent une essence commune.

Tant que l'appartenance à ces catégories n'est significativement importante ni pour l'individu ni pour les membres des groupes dans lesquels il se reconnaît, ni pour les membres des autres groupes, cette catégorisation et cette essentialisation peuvent être comprises simplement comme des facilités permettant une grande économie de mémoire. Elles sont à la base de tout un ensemble d'heuristiques qui nous facilitent la vie au quotidien, au prix d'une généralisation porteuse d'erreurs systémiques.

Mais dès lors qu'appartenir ou non à une catégorie prend une signification forte pour l'individu, le groupe ou le hors-groupe, c'est-à-dire dès que nous sommes prêts à prononcer des phrases comme « je suis de telle catégorie » et à considérer que cette phrase a une signification et comporte un engagement qui dépasse le constat statistique, engagement envers les autres et envers soi-même, alors Appiah considère que l'on entre dans le domaine moral. Cette proposition peut donc être vue comme une définition du domaine moral : le domaine moral est constitué des engagements qui donnent sens aux catégories auxquelles chaque individu déclare, aux autres et à lui-même, appartenir [46].

La norme devient évaluative en un sens fort quand ce qui est évalué emporte une considération importante d'appartenance au groupe : l'individu est un « nous » s'il respecte les normes morales ou est un « autre » s'il ne les respecte pas. On comprend alors l'âpreté des désaccords moraux : il s'agit non de traiter d'un désaccord sur un fait moral externe objectif mais, essentiellement, de marquer avec quelle implication chacun défend les conceptions morales sur lesquelles il s'est engagé en tant qu'appartenant à un groupe social dont les règles en jeu dans le désaccord sont des composantes constitutives.

Lorsque la norme, d'évaluative, devient prescriptive, et comme précédemment, la teneur morale dépend de ce sur quoi porte la prescription, des raisons pour lesquelles elle est établie et de pourquoi les individus auraient tendance à s'y conformer. On reconnaît habituellement des prescriptions d'ordre pragmatique, lorsque les conséquences pratiques envisageables sont clairement en faveur de ce qui est prescrit, d'ordre conventionnel, celles que les conventions sociales communément admises préconisent, et enfin d'ordre moral quand ne pas s'y conformer peut emporter une désapprobation forte du groupe et de l'individu lui-même sur sa propre action. Il n'y a pas de limites bien définies entre ces types de prescriptions qui construisent plutôt un continuum.

En passant de la prescription à l'obligation, enfreindre une norme impérative peut se tra-

46. Cette définition n'emporte pas de choix ontologique fort par lui-même. Le domaine moral peut être second par rapport aux groupes et aux individus ou, au contraire, premier et un constituant de leur individuation.

duire par un risque de condamnation pénale lorsque la norme enfreinte est d'ordre conventionnel, ou se traduire par une désapprobation morale quand la norme impérative est d'ordre moral. Dans ce second cas elle a pour conséquence une possible exclusion des groupes qui adoptent cette norme en tant que règle morale et, du point de vue individuel, négativement, la honte et la mésestime de soi ou, positivement, la décision de changer la configuration des groupes par la conversion ou la dissidence. [47]

Reprenons les différents points que je viens d'aborder :

— Les normes en vigueur dans une société peuvent être descriptives, évaluatives, puis prescriptives ou impératives.
— Une norme en ce qu'elle a de descriptif n'a pas nécessairement de composante morale.
— Les normes deviennent rapidement évaluatives sous l'influence de nos tendances à catégoriser et essentialiser les individus avec qui nous sommes en contact.
— Une norme évaluative prend une dimension morale si les catégories qu'elle sous-tend sont importantes pour le groupe et pour l'individu, au point d'être un indice d'appartenance.
— Une norme prescriptive (ou impérative) peut être d'ordre pratique ou conventionnel, et sans contenu moral significatif, ou être d'ordre moral et donner lieu à une règle morale. L'enfreindre est alors porteur du risque d'exclusion du groupe et d'auto-dévaluation ou de dissidence de l'individu.

Je suis parti d'une définition naïve de la moralité : la théorie morale est l'ensemble des règles générales sur le Bien et le Mal qui aident chacun à instruire des jugements dans les cas particuliers auxquels il est exposé. Je dois maintenant, après ce détour par les normes, la modifier pour prendre en compte les observations que je viens de résumer.

La morale est l'ensemble des règles générales sur le Bien et le Mal qui aident chacun à instruire des jugements dans les cas particuliers auxquels il est exposé. Le respect de ces règles est un indice significatif de la volonté des individus d'appartenir aux groupes qui ont pour norme l'acceptation de ces règles.

Cette définition est encore certainement imprécise et incomplète mais elle a pour objectif ici de me permettre d'avancer la perspective normative et morale ainsi esquissée. Je peux maintenant proposer un ensemble de réflexions en lien avec la possibilité d'aborder le domaine moral par la mise en œuvre des outils des sciences expérimentales.

47. Pour une description détaillée des désaccords moraux, de leur géométrie et de leurs conséquences voir (Ravat 2019)[189]

6.1.5.4 Les normes morales en tant que liant social

Chaque individu souhaite, par naissance, par éducation, par inertie, ou par choix au cours de sa vie, faire partie de différents groupes sociaux dont, pour montrer son attachement à ces groupes, il adopte les normes morales. Symétriquement, chaque groupe se constitue autour de normes morales que les individus qui le composent ont adopté à la création du groupe ou aux différentes étapes de la vie du groupe.

Les normes morales sont donc à la fois des constituants des groupes et des indices d'appartenance à ces groupes. Ce processus dynamique a les caractéristiques d'un double phénomène social et psychologique complexe largement étudié dans la littérature tant sociologique que de psychologie morale [48].

Les normes morales d'un groupe s'imposent à tout individu pour qui appartenir à ce groupe est, pour lui et au moins pour partie, constitutif de son identité. Si ce schéma est accepté, alors la norme n'est pas seulement un lien social entre individus différents mais est également un constituant de l'individualisation des individus, autant qu'il est un constituant de l'individualisation des groupes sociaux. L'appropriation des normes par les individus comme, d'une façon différente, par les groupes, est une étape nécessaire de cette double individuation. Ce processus se fait à la fois positivement, par le lien qu'il crée entre les personnes ayant fait les mêmes choix, et négativement, par l'opposition commune qu'il crée avec les personnes et groupes ayant fait des choix différents.

Cette relation entre adoption des normes morales et appartenance à un groupe est opaque pour l'individu. L'individu ne se voit que très rarement comme respectant une règle morale parce qu'il souhaite appartenir à un groupe ou ne pas en être rejeté [49]. Il se voit plutôt comme choisissant un groupe qui a moralement raison. Cette opacité permet ainsi un double bénéfice, d'une part l'existence du groupe est justifiée, c'est le groupe des gens biens qui ont moralement raison, et d'autre part le respect des règles est justifié, il est important de les respecter car elles sont la vraie loi morale. On doit ici insister sur un point important : cette opacité n'est nullement fortuite, elle est indispensable à l'efficacité opérationnelle des règles morales. C'est parce que les personnes croient que les règles morales sont vraies que, d'une part, ils les respectent et que, d'autre part, ils affirment leur identité au travers des groupes qui les ont choisies. Le titre de l'ouvrage de Kwane Anthony Appiah, « The lies that bind, rethinking identity » est un résumé particulièrement pertinent de cette thèse : nos identités sont consti-

48. Les émotions peuvent jouer le rôle d'intermédiaire entre les deux dimensions de la morale, individuelle et psychologique d'un côté et partagée et sociologique de l'autre, comme le propose Prinz dans (Prinz 2009) [185]

49. Ce cas est d'ailleurs à la base des nombreuses tragédies où les héros sont confrontés à des dilemmes liés à leur double appartenance à des groupes différents exigeant des comportements impossibles à concilier.

tuées des (multiples) groupes auxquels nous adhérons, les identités des groupes comportent, intimement, l'adoption commune de règles morales (et le rejet d'autres règles adoptées par les groupes rivaux), nous adoptons ces règles morales pour montrer notre adhésion mais, contre toutes les évidences empiriques, nous prétendons par aveuglement (psychologiquement efficace pour notre processus d'individuation) que ces règles sont vraies, que nous les avons choisies pour cela, et que donc nous adhérons aux groupes qui les défendent.

On comprend mieux avec cet éclairage les réticences posées à l'étude empirique des règles, normes ou valeurs morales :

— Soumettre les théories morales à l'enquête empirique, c'est accepter le doute systémique nécessaire à cette enquête sur les questions morales et la possibilité que ce doute soit levé en faveur, ou en défaveur des choix constitutifs de l'individu (ou du groupe). Ce doute met alors en péril, par sa simple possibilité, l'identité de l'individu et la pérennité du groupe.

— Accepter la position tierce du chercheur expérimental, c'est accepter les normes (épistémiques d'abord, morales ensuite comme l'a montré l'exemple de l'IAT) en place dans les groupes auxquels appartient ce chercheur. Si ces normes ne sont pas celles des groupes identitaires, alors il est expédient et efficace de rejeter les travaux de ce chercheur.

— Tout groupe développe des arguments justifiant son existence (les mythes de la genèse) qui permettent de justifier les normes qui le caractérisent. Cette justification peut elle-même devenir un élément important de l'individuation du groupe, en particulier quand il devient religieux. La remettre en doute en acceptant l'irruption du doute systémique nécessaire à l'approche empirique est alors un acte impensable dans le cadre des normes du groupe.

6.1.5.5 Les réponses au problème de l'engagement social

L'argument de la nocivité sociale des études empiriques du comportement moral humain peut être schématisé ainsi avec quatre prémisses et une conclusion :

1. Le bon fonctionnement de la société est dépendant du fait que les humains respectent des règles comportementales.
2. Les règles morales jouent le rôle de liant social en tant que règles partagées.
3. Les règles morales sont respectées parce que les individus croient qu'elles sont indubitables.
4. Une étude empirique, cherchant à les décrire, les affaiblit.

5. Conclusion : les études empiriques du domaine moral sont nuisibles au bon fonctionnement de la société.

Bien que chaque étape de ce raisonnement puisse être attaquée et faire l'objet de contre-arguments, ce que je vais maintenant développer, il convient de remarquer l'importance pratique actuelle de cet argument, comme l'illustre plus haut la citation d'Emilio Martinez Navarro, et de plus, son omniprésence dans l'utilisation par tous les dogmatiques plus ou moins extrémistes et plus ou moins religieux.

Première prémisse, la société a besoin de règles comportementales respectées. Ce point est évidement exact mais inopérant si on ne met pas en parallèle ce besoin de règles morales de tous les autres besoins naturels des groupes et individus humains pour survivre, vivre et vivre pleinement. A titre d'illustration prenons les besoins de liberté et de vérité. Si la liberté de chacun s'arrête là ou commence celle des autres, comment pouvons nous vérifier que les règles morales ne sont ni trop invasives, ni trop laxistes par rapport à cette exigence de liberté ? Comment pouvons-nous, individuellement et en groupe, arbitrer entre trop d'ordre ou trop de liberté ? Accepter ces deux questions comme pertinentes revient à accepter qu'il existe une limite à l'application des règles morales lorsqu'elles deviennent abusives en regard d'une exigences de liberté qui, de façon habituelle, n'appartient pas au domaine moral.

Autre aspiration naturelle, celle à la vérité. Supposons qu'une règle morale s'appuie sur un mensonge, et on peut penser ici aux interdictions morales appuyées sur la thèse souvent empiriquement démentie de la « pente glissante » : il faut interdire un comportement non parce qu'il est en lui-même nuisible mais parce qu'il conduirait à faciliter ensuite des comportements réellement nuisibles. Mais les expériences montrent que ce glissement n'est, en pratique, jamais observé. Est-il alors préférable de respecter aveuglément une telle règle appuyée sur un mensonge, ou plutôt de viser à connaître et diffuser la vérité qui en sape les fondements ? Là encore, comment arbitrer entre trop d'ordre moral et pas assez de vérité ?

Les contre-arguments à la première prémisse sont ainsi nombreux et divers, autant que le sont les besoins naturels des groupes et des individus. La priorité du domaine moral ne peut être absolue en toute circonstance et face à toute autre considération.

La deuxième prémisse est empiriquement bien supportée. Les règles morales constituent un liant social. Toutefois, comme nous l'avons vu plus haut, elles ont un triple rôle social. Premièrement, elles permettent à chacun d'affirmer son identité. Deuxièmement, elles permettent de lier entre elles les personnes d'un même groupe. Et, troisièmement, elles identifient, et souvent opposent, les personnes des groupes différents. Or, chaque individu a une personnalité complexe faite d'appartenance à de nombreux groupes, quelques exemples en

sont le genre, la nationalité, la religion, la profession, l'université d'origine, le club de football préféré, etc. Il n'est nullement évident que tous ces groupes, et les règles morales correspondantes, soient nécessaires ou même utiles au bon fonctionnement de la société. Ce point est particulièrement illustré par le cas des religions dans notre société française laïque et multiconfessionnelle : les règles morales diffèrent, servent à la fois de lien au sein de la communauté et de repoussoir entre communautés. On peut argumenter que l'ordre social ne s'en trouve pas amélioré mais, au contraire, dégradé.

La troisième prémisse, l'opacité épistémique des règles morales, a été présentée plus haut comme une caractéristique importante du domaine moral : chacun s'engage à respecter des règles morales et, simultanément, les pense vraies. Un engagement envers des règles qu'on pense erronées peut donner lieu à des dilemmes momentanés et à des délibérations, mais une telle situation nous apparaît moralement inconcevable dans la durée. Cette prémisse me semble plus solide que les précédentes, et je remarquerai simplement que, si elle est vraie, alors mettre en doute la vérité d'une règle morale est une attaque forte envers l'individu, car celui-ci aura ensuite à se reconstruire une personnalité et adopter d'autres règles morales qu'il aura alors à croire vraies, ou en tous cas plus vraies que les précédentes. Insistons, il ne s'agit pas ici de remplacer une croyance par une impossible lucidité, mais il s'agira de substituer une règle morale que l'on croit maintenant vraie à une autre que l'on a cru vraie par le passé mais que l'on croit maintenant erronée.

Enfin, la quatrième prémisse porte le constat de la nocivité : étudier empiriquement n'est possible que si on se pose des questions, et se poser des questions c'est prendre de la distance avec l'obligation de respecter les règles morales. Remarquons tout d'abord que ce raisonnement induit une distinction forte entre deux populations, celle qui établit les règles morales, que ce soit par l'exercice d'un compétence peu répandue ou en tant que propagateurs d'une vérité révélée, et ceux qui les suivent et doivent être tenus dans l'ignorance des coulisses. Cette distinction est bien sûr difficile à accepter dans le cadre d'une société posant l'égalité de droit de tous les individus.[50] Par ailleurs, on peut soutenir que le doute n'entraîne pas obligatoirement la faiblesse du sentiment moral. On peut penser ici à Descartes se proposant, dans le doute, de suivre les mœurs de sa nourrice et de son pays, ou, sur un plan différent, les religieux qui plaident pour le doute en tant que passage indispensable de mise à l'épreuve de

50. Cette distinction entre philosophes experts et population générale est à la racine de l'argument de l'expertise développé pour relativiser les conclusions des XPhi. Les réponses de non spécialistes à des questionnaires comportant des termes techniques seraient sans intérêt. Machery expose dans (Machery 2017) [151] les principaux arguments montrant que cette distinction n'est pas pertinente. Elle est empiriquement inopérante : les réponses apportées sur des questions morales par les spécialistes sont les mêmes. Elle est insuffisante : il y a de nombreuses spécialités et aucune expertise n'englobe tout le domaine moral. Enfin, aucun critère précis ne permet d'objectiver cette expertise, le risque est donc fort d'une auto proclamation.

la foi.

Pour cette quatrième prémisse, et comme précédemment, on ne peut que constater que son effectivité dépend de la théorie morale de celui qui définit, finance, promeut ou évalue les études empiriques du domaine moral. A un extrême, s'il agit dans un contexte religieux appuyant les règles morales et la nécessité de leur respect sur la divinité, la prémisse de la nocivité sociale des études empiriques a une certaine force. A un autre extrême, s'il agit dans un milieu rationaliste appuyant les règles morales sur le raisonnement et donnant la priorité à la connaissance, alors cette prémisse est sans poids.

L'acceptation de la possibilité des études, préalable à celle de la recevabilité des conclusions, est ainsi dépendante des conceptions morales en cours dans la société. Ceci est naturellement particulièrement vrai pour l'éthique substantielle qui vise à dire le Bien, mais c'est également vrai pour la méta-éthique qui vise à comprendre comment les théories morales se construisent, se justifient et se comparent. Notons enfin que, dans sa forme la plus extrême, la prémisse de la nocivité sociale conduit à rejeter toute approche rationnelle du domaine moral non limitée par les choix moraux en place. Le rejet ne concerne pas que les approches empiriques ou expérimentales mais couvre également toute approche sémantique ou analytique.

6.1.6 Les difficultés : un bilan complexe

Arrivé à ce point de mon travail, j'ai décrit les axes qui regroupent les arguments développés par des philosophes moraux pour montrer l'impossibilité, ou la très grande difficulté, qu'a la démarche empirique à produire des arguments intéressants pour les débats de philosophie morale. J'ai évoqué la complexité, la circularité et la normativité comme sources de ces difficultés. Pour la complexité, j'ai soutenu que l'argument n'était pas spécifique à la philosophie morale mais était opposable à toute étude des comportement humains dits supérieurs. Pour la circularité de l'homme se pensant lui-même, j'ai souligné la pertinence de cet argument appuyé par la force et la permanence des biais cognitifs que développent les humains quand ils s'observent eux-mêmes. Enfin, pour la normativité, j'ai également reconnu l'ampleur du problème et insisté sur trois facettes, la première est le problème de Hume, l'impossibilité logique d'établir empiriquement des règles morales. La seconde est l'impossibilité, selon certaines théories morales, pour un chercheur d'entreprendre les études empiriques qui le conduisent à s'octroyer une position de surplomb moralement inacceptable envers un autre humain. Et enfin j'ai évoqué la difficulté sociale, plus large, invoquée par les théories morales

qui conditionnent la recherche de connaissance à la pertinence morale : elles considèrent que la mise en doute nécessaire à l'enquête empirique porte un risque de mettre en porte à faux la justification des règles morales, par exemple transcendantes, indispensables à la cohésion sociale.

La difficulté liée à la complexité est d'ordre pratique. Bien que réelle elle ne constitue pas une impossibilité de principe mais nous pousse à l'humilité, la patience et la prudence dans l'exploitation des résultats empiriques à des fins morales. En prolongement des propositions du chapitre sur l'opérationnalisation, ces difficultés d'interprétation sont à mettre en parallèle de la faible pertinence de la mise en pratique des objets et relations que manient les théories morales. L'exemple de l'effet Knobe et de la grande difficulté à décrire l'attribution d'intentionnalité est une illustration de ce point.

Certes le comportement humain est complexe à étudier, mais les connaissances acquises par les sciences cognitives montrent qu'un chemin, partiel et progressif, est possible en utilisant les meilleures approches naturalistes dont nous disposons [51]. Mon analyse ne m'a permis de dégager aucun élément lié à la complexité qui me conduirait à considérer comme raisonnable de renoncer à l'expérimentation en philosophie morale. En revanche, cette complexité plaide pour une grande humilité et une certaine patience quant à l'obtention de résultats ayant un niveau de validité propre à convaincre les différents philosophes moraux qui peuvent l'être. Sur-interpréter philosophiquement les résultats expérimentaux et, de façon équivalente, sous-opérationnaliser les entités théoriques étudiées, voilà le principal risque lié à la complexité.

La difficulté liée à la circularité de l'homme se pensant lui-même est pour partie d'ordre également pratique, et en ce sens, rejoint la complexité. Mais cette difficulté est surtout épistémique et donne lieu à certains des célèbres paradoxes sceptiques. J'ai souligné que l'argument sceptique était logiquement fort mais rhétoriquement faible du fait de sa nature extrême qui impose le silence à celui qui l'adopte. La pratique des psychologues confrontés à cette difficulté s'appuie sur la triangulation des approches de multiples façons, favorisées par la transversalité et l'interdisciplinarité. L'isolement du chercheur ou d'une discipline est, à ce titre, un indicateur fort de risque de circularité.

L'argument de la circularité induit un doute fort quant à notre capacité à nous penser nous-mêmes, ce qui nous intime d'être conscients des risques de biais liés à ces limites et de redoubler de rigueur. J'ai argumenté que les outils permettant de faire face à ces biais pour les questions de philosophie morale n'avaient aucune raison d'être différents de ceux

51. Sur cette dimension du naturalisme voir (Andler 2016) [3] et (Collins, Andler et Tallon-Baudry 2018).[51]

utilisés pour toutes les questions abordées par les psychologues. Là encore, l'humilité et la patience sont certainement à développer, mais j'ai souligné l'importance des stratégies nous permettant de sortir de la circularité et en particulier l'importance qu'il y a à varier les points de vue sur un phénomène en changeant les équipes, les techniques, les cultures, et même les espèces étudiées. Sur-interpréter les résultats expérimentaux et sous-évaluer l'importance d'une opérationnalisation transversale qui permette de sortir du piège de la circularité, voilà les principaux risques liés à la complexité et à la circularité.

La difficulté logique liée au problème de Hume, là aussi logiquement bien étayée, si elle interdit l'établissement empirique de règles morales substantielles, laisse une large place à l'expérimentation pour connaître la situation morale initiale d'une population, définir les comportements problématiques que la théorie morale se propose de réguler, définir le champ de ce qui est possible ou impossible et ne peut donc être moralement imposé, d'évaluer les éventuelles actions réformatrices, de détecter les contradictions entre règles morales dans le comportement observé . . . Enfin, j'ai remarqué que le problème de la dérivation empirique de ce qui doit être ne se pose pas dans les mêmes termes si nous cherchons à établir ou justifier une règle morale, ce qui relève de l'éthique substantielle, ou si, prenant la position de recul de la méta-éthique, nous comparons les théories morales entre elles.

La difficulté morale du chercheur s'octroyant le droit d'expérimenter sur un autre humain amène à la nécessité de poser la question du contexte de la recherche, de qui la définit, l'autorise, la finance, l'évalue, et des règles morales en vigueur dans ce contexte. J'ai pris l'exemple extrême d'un milieu religieux adoptant des règles morales divines incluant la règle absolue de la dignité humaine qui interdit toute instrumentalisation de l'être humain. Dans ce contexte il est impossible au chercheur psychologue de travailler et l'expérimentation est, par construction théologique, impossible.

Dans un contexte plus modéré, l'expérimentation peut avoir lieu mais, comme je l'ai montré avec l'exemple de l'IAT ou comme l'actualité de l'expérimentation animale le suggère, apparaîtront des conflits entre la recherche de connaissance et le respect des règles morales en vigueur dans la société. La constitution de comités d'éthique montre à la fois la reconnaissance de l'existence de ce problème et la tentative de construire un outil pour le circonvenir par une pratique ancrée dans un raisonnement pragmatique : serait moral le résultat de la délibération d'un comité. Je reviendrai sur ce point dans la section sur l'éthique appliquée. On retrouve là par une autre approche le constat de Philip Kitcher (Kitcher et Ruphy 2010) [135] sur la nécessité de soumettre la démarche scientifique aux règles démocratiques dans les sociétés qui le sont ou visent à l'être.

La difficulté sociale à entreprendre l'étude empirique des comportements moraux humains m'a conduit à un détour par l'analyse des normes. Une norme peut être descriptive, évaluative, prescriptive ou impérative. Adopter une même norme relie les membres d'un groupe entre eux, ce qui structure l'identité du groupe, et chaque individu adopte les normes des groupes dans l'appartenance auxquels il ancre son identité de personne. Les individus sont justifiés à suivre les normes, et en particulier les normes morales, en tant qu'ils souhaitent appartenir aux groupes sociaux qui les ont adoptées. Cette justification est opaque pour l'individu : celui-ci ne se perçoit pas comme respectant la règle parce qu'il souhaite appartenir à un groupe, il se perçoit comme convaincu que ce groupe a raison de dire que là est le Bien et que la règle s'impose moralement.

Dans cette optique, remettre en cause la norme morale devient un acte de sédition envers le groupe, quelle que soit la raison de cette remise en cause, y compris pour des raisons épistémiques liée à la méthode scientifique. L'étude empirique des règles morales est alors qualifiable de sulfureuse. Le risque en est le délitement des justifications des règles morales et, en conséquence, des structures sociales.

Et enfin, le réseau intime qui lie l'individu, observateur ou observé, aux groupes dans lesquels il se reconnaît au travers des règles morales nous interdit l'option facile mais sans issue d'une démarche scientifique qui se voudrait « a-normative ». Il nous contraint à prendre au sérieux le défi de l'intégration des trois axes de la réflexion philosophique, ontologie, épistémologie et axiologie dans l'urgence de l'action.

Je vais maintenant aborder différentes réponses apportées aux difficultés évoquées ci-dessus et, tout d'abord la proposition pragmatique de l'éthique appliquée. J'envisagerai ensuite différentes approches scientifiques qui traitent du comportement moral humain et doivent donc répondre à ces difficultés. Ensuite je ferai une analyse plus approfondie des propositions évolutionnistes qui me semblent détenir un potentiel explicatif important pour le domaine moral. Toutefois, mon insatisfaction à l'égard de ces approches évolutionnistes me conduira à faire une proposition de distinction qui me semble de nature à améliorer ces dernières : la distinction entre le besoin de l'existence d'une norme et le contenu de la norme.

6.2 Contourner les théories morales : l'équilibre réfléchi

Être confronté à la nécessité d'une décision morale ici et maintenant, que ce soit une décision professionnelle ou personnelle, est le lot commun de l'humanité. Ce lot prend des contenus dramatiques dans les cas auxquels fait face le monde médical avec les situations

de fin de vie, d'interruptions de grossesses ou d'euthanasie. Les grandes théories morales apparaissent souvent dans ces cas lointaines et incertaines, et il faut alors s'appuyer sur des ressources limitées, mais enracinées dans la situation concrète, dans l'objectif de trouver, tout bien pesé, une solution acceptable, autant que faire se peut, par toutes les parties prenantes. C'est l'objectif des théories de l'équilibre réfléchi objet de cette section.

6.2.1 L'équilibre réfléchi

Les situations concrètes exigeant une prise de position éthique ici et maintenant sont nombreuses. A chaque décision que nous devons prendre, nous nous demandons ce que nous devons ou devrions faire pour être nous-mêmes tels que nous sommes, tels que nous souhaitons être et être reconnus. Mais chacune de ces situations concrètes comporte des spécificités complexes souvent contradictoires en regard de règles morales qui pêchent par leur trop grande généralité. Et nous nous retrouvons seuls et sans appui face à ces décisions à prendre. Ce constat de l'impuissance des théories morales à nous aider se décline dans de multiples directions :

— La réalité de chaque situation morale est faite d'une multitude de circonstances qu'il est impossible de ramener à quelques caractéristiques qui permettraient de juger des situations comme semblables du point de vue moral : aucune description n'épuise cette complexité et, au mieux, seules les personnes impliquées peuvent l'approcher.

— Chaque situation s'inscrit dans un lieu et un moment de la vie des individus et des sociétés. Ces conditions sont souvent implicites et impossibles à reproduire ou à rapprocher d'autres situations.

— Chaque décision peut être instruite de multiples points de vue, lequel privilégier à l'heure d'agir ? Si c'est celui d'une autorité extérieure, laquelle et pourquoi ? N'est-ce pas plutôt vers une convergence des avis des différentes parties prenantes que nous pouvons, au mieux, nous tourner ?

— Les théories morales qui ont un fondement religieux ne peuvent plus être mises en œuvre dans un monde pluriel qui doit voir cohabiter en chaque lieu et à chaque instant plusieurs confessions difficiles, ou impossibles, à concilier.

— Les théories morales qui exigent, comme l'utilitarisme de l'acte, des calculs hors de portée en situation réelle sont pratiquement inutiles.

— Les spécificités morales de chaque domaine d'activité, médical, affaires, sport, ... sont importantes et ne peuvent permettre l'établissement de règles générales communes.

> La multiplication des éthiques appliquées, chacune ayant ses considération propres, en est la conséquence opérationnelle.

Les philosophes moraux pragmatiques ayant ainsi constaté que les règles morales générales semblaient de peu d'utilité à l'heure des choix réels ont proposé de les remplacer par une démarche qu'il soit possible de mettre en œuvre : l'équilibre réfléchi [52].

Cette démarche comporte schématiquement cinq étapes [53] :

Étape 1 : Établir la connaissance morale courante dans un domaine d'activité particulier. La connaissance morale courante s'appuie sur les cas connus, ceux qui ne donnent pas lieu à doute sur le jugement moral. On peut y voir, par une analogie du bien au juste, un équivalent de la jurisprudence. Elle comporte également des principes généralement acceptés, par exemple, la règle d'or de la réciprocité et le « primum non nocere » des médecins...

Étape 2 : Proposer des règles issues d'une recherche d'équilibre entre abstraction et justification des cas connus. La connaissance morale courante donne lieu à l'établissement de règles morales prima facie abstraites à partir des cas connus et pouvant servir à les justifier.

Étape 3 : Examiner les situations réelles. Le cas réel est décrit en mettant en avant les analogies avec les cas connus et en appliquant les règles morales prima facie. Lorsque toutes les parties prenantes sont d'accord sur les décisions appuyées sur ces analogies, les décisions sont acceptées comme moralement satisfaisantes. En revanche, lorsqu'elles sont en désaccord, la situation de conflit est constatée.

Étape 4 : Expliciter le désaccord puis, itérer sur les étapes 2 et 3 pour arbitrer les conflits. Chaque partie prenante explicite son propre point de vue et s'engage à entendre ceux des autres avec, en particulier, mention des analogies rapprochant des cas connus retenus, et des différents principes moraux considérés comme pertinents dans ce cas particulier. Le débat entre parties prenantes vise ensuite à tenter de réduire progressivement les écarts entre les positions et de les dissoudre. Pour cela trois voies sont explorées. Soit on met en doute les analogies entre cas réels et cas connus et on change l'analogie conductrice du raisonnement. Soit on modifie les principes moraux à utiliser dans ce cas particulier. Soit on réévalue les cas connus eux-mêmes. Notons que, pour cette étape, aucun des éléments issus des étapes précédentes n'est gravé dans le marbre. Au titre de la spécificité de chaque cas réel, tous peuvent être révisés.

Étape 5 : Une fois les cas particuliers traités, évaluer plus largement les principes moraux

52. Procédure proposée par John Rawls dans (Rawls 1951) [190]

53. Pour cette description et outre John Rawls, le fondateur, on peut également se référer dans la littérature secondaire à des opposants à la philosophie expérimentale : (Suikkanen et Kauppinen 2018) [221] et à l'ouvrage de Marta Spranzi sur l'éthique en milieu hospitalier (Spranzi 2018) [216]

abstraits de ces cas particuliers dans leurs conséquences sociétales. L'objectif est de les inscrire plus largement dans des théories de la justice, du droit, de la politique, . . . mais, toujours et méthodiquement, sans viser à établir des principes moraux qui seraient applicables sans réitérer 2 3 et 4 à chaque cas particulier.

6.2.2 Les avantages de la méthode

La méthode de l'équilibre réfléchi présente de multiples avantages. Elle est claire et correspond bien à la fois à nos intuitions morales les plus courantes, prises en compte lors de la construction de l'étape 1, et au constat de la difficulté d'en déduire des positions dans les cas particuliers, pris en compte par les itérations nécessaires à la résolution des conflits entre parties prenantes.

Son caractère quasi algorithmique rassure par son systématisme. Il semble augurer qu'on pourra ainsi sortir des conflits moraux sans fin dans un délai raisonnable, ce qui est bien la contrainte qui s'impose aux praticiens des éthiques appliquées.

Enfin, ayant donné lieu à de nombreuses mises en œuvre, la démarche semble bien rodée et bénéficier du savoir faire de nombreuses personnes dans chaque profession concernée.

6.2.3 Les points faibles de la méthode

Néanmoins, des critiques subsistent, principalement du point de vue méta-éthique.

Premièrement, les conclusions de l'enquête semblent dépendre des a priori de l'étape 1, sans que ceux-ci n'aient à être justifiés. Naturellement, les défenseurs de la méthode répondront en demandant comment il pourrait en être autrement. Il faut bien un point de départ, et une démarche qui partirait de conceptions contre-intuitives serait encore plus difficile à accepter.

Deuxièmement, la méthode ne dit rien sur la façon de procéder aux choix nécessaires à l'arbitrage des conflits. Quand et sur quel critère faut-il considérer que le cas particulier à l'étude se rapproche ou non de tel cas connu de référence ? Quand faut-il réviser ces cas connus généralement acceptés ? Quand faut-il mettre en cause les règles générales ? . . .

Troisièmement, la méthode donne beaucoup de poids aux parties prenantes, mais ne fixe pas de critère pour définir qui est ou qui n'est pas partie prenante, et où il faut s'arrêter dans la prise en considération de conséquences de plus en plus lointaines et indirectes, et des personnes de plus en plus nombreuses ainsi concernées.

Quatrièmement, il semble possible que les parties prenantes d'un cas particulier tombent

d'accord sur un compromis sans que celui-ci soit particulièrement « moral ». Il peut être localement satisfaisant pour résoudre le conflit mais ne correspondre ni aux règles en place ni, encore moins, contribuer à en établir de meilleures.

Enfin, cinquièmement, comme le font remarquer Suikkanen et Kauppinnen [221], les philosophes éthiciens n'utilisent pas cette méthode dans leurs propres recherches et continuent leurs travaux de méta-éthique avec les outils habituels de l'analyse conceptuelle des philosophes. Pour ces auteurs, comme il y a des éthiques différentes, il peut y avoir des méthodes, ou plutôt des stratégies argumentatives, différentes et travailler dans chaque cas sur la meilleure méthode à adopter est une nécessité.

Dépassant ces critiques, il convient d'insister sur la large utilisation de cette méthode et nous prendrons ci-après pour exemple les comités d'éthique dans le monde hospitalier tels que décrits dans l'ouvrage de Marta Spranzi [216].

6.2.4 Les comités d'éthique en milieu hospitalier

La méthode de l'équilibre réfléchi est un outil largement utilisé en milieu médical au sein des comités d'éthique. Les décisions à prendre y sont souvent difficiles et concernent la fin de vie, l'arrêt des soins, la procréation assistée, et l'ensemble des décisions à prendre entre le patient, ses proches, et le corps médical [54]. Le constat d'impuissance des théories morales à contribuer à ces décisions quotidiennes et parfois urgentes est patent, comme le rappelle Marta Spranzi (Spranzi 2018 page 14) [216]) :

> Il n'y a pas à espérer des spécialistes de la normativité de solutions toutes faites : ils ne peuvent faire plus que proposer une palette de perspectives largement divergentes.

S'impose alors la mise en œuvre d'une démarche pragmatique supervisée par des comités d'éthique, eux-mêmes en relation avec l'autorité de référence en la matière qu'est le CCNE, le Comité Consultatif National d'Éthique.

On retrouve les étapes décrites plus haut. L'arrière plan de l'enquête éthique est constitué des cas connus et des exemples qui ont fait l'objet de décisions acceptées. Quatre principes de base structurent les débats (Spranzi 2018 page 71) [216] :

— L'autonomie du patient, la nécessité de l'informer et, si possible, d'obtenir son consentement éclairé.

54. Dans son ouvrage de référence (Sicard 2011) [206] Didier Sicard propose un état de toutes ces questions : La procréation, l'avortement, l'euthanasie, le don d'organes, le rapport aux pays du sud, le principe de précaution, le lien à la morale religieuse, le secret médical, le dépistage systématique consenti ou non, le lien à l'économie de la santé, le développement post ou sur humain, le lien au juridique et les considérations du relativisme culturel.

— La bienfaisance : toute action doit être orientée vers un gain pour le patient et pour ses proches.
— La non malfaisance : le «primum non nocere » du serment d'Hippocrate.
— L'équité : chaque patient doit être traité équitablement, sans considération de ses revenus ou de son statut social.

Se rajoutent à ces quatre principes de base les contraintes réglementaires et légales que l'équipe médicale doit respecter. Ces contraintes peuvent porter sur des définitions précises comme la durée maximale de grossesse pendant laquelle l'avortement est autorisé, et constituer ainsi un cadre précis pour la décision à prendre.

Les parties prenantes, les patients, leur famille, les équipes soignantes ont alors à définir et décrire les différentes solutions envisageables de façon à ce que chacun puisse en comprendre la teneur et, au moins en première approche, les conséquences. Si l'accord se fait, le rôle du comité d'éthique est principalement de faciliter la prise en compte de la dimension éthique dans la décision. En revanche, en cas de désaccord, il peut être sollicité par une partie prenante pour intervenir dans la décision.

Il est important de souligner que, dans ce cas de conflit, le philosophe éthicien n'intervient pas comme un sachant qui vient donner la solution à des personnes qui ne sauraient pas trouver la bonne décision. Son intervention s'apparente plutôt à celle d'un maïeuticien qui va aider à rendre visibles les points d'accord et de désaccord et à proposer des démarches permettant que chacun, par concessions successives mutuelles, se rapproche d'une solution possible.

Remarquons enfin que la méthode est appuyée sur un pari : que les parties prenantes ont plus à gagner à se mettre d'accord, et résoudre le conflit, qu'à rester accrochées à des désaccords moraux interdisant toute convergence. Au moment où nous écrivons ces lignes (juin 2019), la famille de Vincent Lambert se déchire encore dans des procédures juridiques interminables qui montrent que ce pari n'est pas toujours gagné.

6.2.5 L'équilibre réfléchi et la place de l'expérimentation

Revenons à notre point de vue expérimental. Une hypothèse forte de la méthode de l'équilibre réfléchi est qu'il est impossible de décrire les cas réels de façon à en épuiser la complexité. Il est toujours nécessaire d'en revenir à des itérations de mise au point des décisions éthiques qui permettent les ajustements indispensables à tout cas particulier. Il est donc fortement douteux qu'on puisse ni concevoir une expérimentation qui soit représentative d'un

ensemble de situations réelles ni interpréter des résultats expérimentaux pour les généraliser. En somme, l'approche expérimentale est au mieux inutile dans cette perspective.

La méthode de l'équilibre réfléchi est également indépendante de toute philosophie morale dans la mesure où ce qui est visé n'est pas de savoir ce qui est « bien » mais uniquement de définir pratiquement ce qui est le mieux pour les parties prenantes dans un cas particulier, ici et maintenant.

L'abstraction, qu'elle soit morale ou scientifique, joue un rôle second, subsidiaire, dans cette méthode de l'équilibre réfléchi et les règles morales sont au mieux des heuristiques qui permettent de gagner du temps en début de processus décisionnel. La méthode est profondément pragmatique.

On peut néanmoins remarquer que, pragmatiquement, l'adepte de l'équilibre réfléchi s'intéressera à l'efficacité de sa démarche. S'il ne recherchera pas au travers de l'expérimentation à conforter des règles morales, il pourra néanmoins chercher à améliorer sa propre démarche : tester de nouvelles façons de travailler avec les parties prenantes, de les mettre en situation de rechercher un compromis, de présenter les cas connus, . . . Il pourra ainsi conforter la plausibilité de sa démarche, préciser les conditions nécessaires à ce qu'elle atteigne ses objectifs dans chaque cas particulier en expérimentant les différentes méthodes ou heuristiques pour atteindre l'équilibre réfléchi. Lorsque le critère pour une « meilleure méthode » est d'efficacité, par exemple le nombre de cas particuliers qui ont abouti à un compromis, l'expérimentation peut être utile dans cette perspective de l'équilibre réfléchi, mais le critère ne pourra être la recherche d'une « méthode plus morale » car il est impossible d'évaluer, dans cette perspective, si le compromis atteint est plus ou moins moral.

Par construction de la démarche de l'équilibre réfléchi, il n'y a pas de chemin interprétatif allant des cas particuliers vers des règles morales abstraites. Il n'y a que des cas particuliers de complexité infinie. La principale mesure de la réussite de la démarche est le nombre des accords qu'elle permet de trouver entre les parties prenantes d'un cas particulier et l'expérimentation peut y contribuer sous cet angle de la construction méthodologique.

6.3 Les sciences et la question normative

Dans les sections précédentes, j'ai souligné les multiples arguments des philosophes moraux qui doutent de la possibilité même d'une approche scientifique des problème moraux. Dans la présente section, et les sections suivantes, je vais maintenant présenter les réponses que proposent les scientifiques concernés par ces doutes. Plus que les réponses, qu'il serait

illusoire d'espérer définitives, il s'agit de montrer ici la diversité des méthodes scientifiques qui ont abordé l'étude du domaine moral et les résultats, partiels et toujours faillibles, auxquels elles ont contribué, sans occulter les difficultés soulevées qui peuvent, ou non, refléter les doutes émis par les philosophes moraux. Je détaillerai en particulier un des domaines d'étude les plus féconds, celui des approches évolutionnistes de la morale, et les perspectives qu'elles permettent d'ouvrir. Je conclurai ce chapitre par une proposition issue de cette étude : la diversité des sciences et la multiplicité des échelles de temps concernées, au niveau de l'espèce, du groupe social et de l'individu, suggèrent tout l'intérêt qu'il y a à distinguer la nécessité de l'existence d'une norme morale, au sein des groupes humains, du caractère largement contingent du contenu de cette norme pour chaque groupe particulier.

6.3.1 La diversité des sciences concernées

Comme je l'ai évoqué plus haut, plusieurs disciplines scientifiques ont pour objet le comportement humain et en particulier son comportement moral. On peut citer au premier rang la psychologie, et tout particulièrement la psychologie morale dont c'est l'objet central. L'anthropologie étudie l'humain dans son ensemble et ne peut ignorer le phénomène moral en tant que composante importante des sociétés humaines. La sociologie s'intéresse à la façon dont les normes morales contribuent à cimenter ou au contraire à disjoindre les liens sociaux. Les médecins peuvent être amenés à étudier les psychopathes qui semblent ignorer toute règle morale et les neuroscientifiques recherchent les corrélats neuronaux de tous les comportements humains, dont naturellement les comportements moraux.

Chacun de ces scientifiques, à des degrés divers et sous les diverses formes qu'adopte sa discipline, fait allégeance à la méthode scientifique expérimentale de façon plus ou moins proche à ce que j'ai évoqué dans les chapitres précédents. Je reprends ci-dessous les difficultés rencontrées, que j'ai déjà évoquées, lorsqu'on vise à expérimenter sur les comportements moraux en adoptant le point de vue de cette démarche scientifique. J'envisagerai pour cela les cas particuliers de la neuroéthique et de la psychologie morale qui, l'une et l'autre, ont pour objet central le comportement moral. J'aborderai à la section suivante les études des origines de la morale appuyées sur la théorie de l'évolution.

Auparavant, reprenons rapidement les arguments visant à dévaluer l'apport de la démarche scientifique expérimentale à la philosophie morale en prenant le point de vue du scientifique qui entreprend, avec son objectif propre, l'étude du comportement moral humain.

En regard de la complexité, j'ai défendu plus haut que le comportement moral n'était pas

significativement différent des autres comportements dits supérieurs. Comment un scientifique peut-il aborder cette complexité ? Première solution, en écartant cette étude de son champ opérationnel et en laissant d'autres approches s'en emparer. On peut considérer, en reprenant le constat proposé plus haut sur le bilan de la psychologie morale en 1998 (Bègue 1998), que cette stratégie d'évitement prudent a été majoritaire pendant une grande partie du vingtième siècle.

Deuxième solution, en prenant en compte cette difficulté par des méthodes adaptées à la difficulté du sujet, c'est l'objet même des disciplines des sciences humaines que de proposer ce type de méthode. Il ne saurait alors y avoir d'impossibilité de principe mais seulement des difficultés à surmonter par des travaux de méthodologie.

Et troisième solution, en considérant simplement que la complexité est au cœur du métier de chercheur scientifique et qu'il n'y a pas lieu de répondre autrement qu'en avançant les recherches à cet argument. La voie scientifique est de toutes façons, pour lui, la plus à même d'avancer, peut-être lentement, sur les sujets complexes.

Le risque de circularité est vu par le scientifique comme une source parmi d'autres d'erreurs qu'il convient de réduire progressivement au même titre que toutes les autres. Les biais à combattre sont ici très résistants, car dépendants de la nature humaine des chercheurs, et qu'ils ont contribué à structurer les institutions de recherche, mais c'est une question de degré et non de principe. Par construction, il ne saurait y avoir une rationalité différente selon le sujet étudié ou selon qui l'étudie, et le comportement humain ne fait pas exception.

En première intention, la démarche scientifique étudie ce qui est et n'a que faire de ce qui devrait être, et donc du problème de Hume. Depuis la révolution du 17e siècle, en se concentrant sur le comment et en évitant le pourquoi, elle semble s'être définitivement éloignée du domaine des normes et valeurs. Mais ce qui a fait ainsi la force de la démarche en physique ou en chimie, puis en biologie, se heurte avec la psychologie aux difficultés qu'il y a à ignorer ces mêmes valeurs lorsqu'elles importent pour le comportement humain à étudier, c'est-à-dire, en pratique, en permanence. Je reviendrai sur ce point ultérieurement, après avoir présenté la distinction entre l'étude de l'origine d'une norme et l'étude de son contenu substantiel.

L'impossibilité morale pour un chercheur de mettre un humain sous son microscope est ancienne, on peut citer ici l'interdiction de la médecine par la religion au nom d'un dieu jaloux de ses prérogatives de vie et de mort sur les humains [55]. Le scientifique en fait abstraction soit en adoptant une théorie morale qui autorise l'exercice de son métier [56] soit en visant une

55. L'exemple des témoins de Jéhovah refusant encore aujourd'hui la transfusion montre que cette préoccupation reste d'actualité.

56. On peut faire le parallèle avec les 80 % d'athées présents chez les philosophes : pour passer sa vie à étudier un

conciliation entre deux modes de connaissances d'ordres différents [57].

Enfin, le risque social lié à l'engagement d'un programme de recherche est certainement le plus contraire aux présupposés implicites de la démarche scientifique : comment pourrait-il être moralement préférable de ne pas savoir ? Cette problématique ne peut qu'être écartée au travers d'un discours reportant la responsabilité de lancer et financer des recherches sur un processus externe à la démarche scientifique elle-même, comme par exemple celui préconisé dans (Kitcher et Ruphy 2010) [135].

Les préoccupations éthiques sont de plus en plus présentes dans la démarche scientifique sous l'angle de la pratique d'une recherche intègre et responsable, titre du guide 2017 du comité d'éthique COMETS du CNRS [58]. Ce guide présente principalement les engagements du chercheur en tant que porteur d'une grande responsabilité pour la reconnaissance sociétale de la connaissance scientifique. Il doit donc refuser le plagiat, la fraude, (...) et s'attacher à l'intégrité de la démarche scientifique et, également, à la proportionnalité de ses conclusions en regard des preuves théoriques ou empiriques produites. Néanmoins, et pour en revenir à l'étude scientifique des sujets à forte connotation morale, il est tout à fait notable que ce sujet ne soit présent dans le guide du CNRS que sous la forme d'une incidente qui dit mieux que tout long discours que la préoccupation morale relative au contenu de la science, et non à sa méthode, n'est pas le problème du chercheur et de son institution :

> Les chercheurs et leurs institutions ne peuvent éviter les questionnements de nature scientifique qui préoccupent les citoyens et se doivent de les éclairer en mobilisant leurs connaissances.

.

6.3.2 La neuroéthique

La neuroéthique est une discipline récente à l'intersection de la philosophie morale et des sciences du cerveau appuyées sur les neurosciences (Roskies 2016) [196]. Dans une optique inclusive, elle peut viser également l'ensemble des sciences cognitives qui contribuent à la connaissance du comportement cognitif humain, sans se limiter aux seules neurosciences qui ont, par le développement important qu'elles ont connu au tournant du siècle, justifié de sa création.

Elle comporte deux dimensions éthiques très différentes : l'éthique des neurosciences, i.e.

sujet il vaut mieux ne pas avoir de réponse dogmatique a priori (Bourget et Chalmers 2014) [28]

57. Pour exemple, l'encyclique papale de 1998 Fides et Ratio : « La foi et la raison sont comme deux ailes qui permettent à l'esprit humain de s'élever vers la contemplation de la vérité. »

58. Accessible sur Internet à l'adresse : http ://www.cnrs.fr/comets/spip.php ?article181 consulté le 17 juin 2019.

tous les problèmes moraux soulevés par les recherches sur le cerveau et son fonctionnement, et la neuroscience de l'éthique, i.e. la recherche des bases neurales et cognitives des comportements moraux (Evers 2009) [81]. La première dimension la rapproche de la philosophie morale appliquée à l'activité scientifique et, en particulier, de la bioéthique, alors que la seconde la rapproche des sciences humaines, dont la psychologie morale. Je n'approfondirai pas ici la question débattue de la pertinence de créer ainsi une nouvelle discipline réunissant des dimensions a priori très dissemblables [59]. Cet exemple de la neuroéthique permet ici de mettre en lumière les différents cadres dans lesquels s'inscrivent l'utilisation des sciences pour l'étude du domaine moral ainsi que, inversement, les problèmes moraux soulevés par les recherches scientifiques.

L'éthique des neurosciences

L'éthique des neurosciences vise à répondre aux questions des domaines que j'ai appelés plus haut l'engagement moral et l'engagement social du chercheur en neurosciences. Quelques exemples illustreront l'ampleur et la diversité des questions posées :

— Des techniques sont mises au point pour soigner le cerveau, pour rétablir un fonctionnement déficient. Peuvent-elles être utilisées pour améliorer un organisme et, tout particulièrement, pour améliorer les performances cognitives ou comportementales humaines ?

— La vie privée, la liberté et l'autonomie de l'individu sont-elles en cause lorsque le chercheur peut lire le contenu des états mentaux, le modifier momentanément ou de façon plus durable ? Comment évaluer ce risque et quelles institutions décideront de la limite acceptable ?

— La découverte de relations plus fortes que de simples corrélations entre des comportements et des états mentaux physiquement déterminés est-elle juridiquement, ou moralement, utilisable pour diminuer la responsabilité de l'agent ?

— La découverte de nouvelles techniques pour améliorer les performances va accroître les inégalités car elles seront réservées aux populations riches des pays développés du fait de leur coût et des compétences nécessaires à leur mise en œuvre. Est-ce admissible ?

— L'avidité médiatique pour les résultats neuroscientifiques pousse à une vulgarisation prématurée de résultats, et concomitamment, à une surinterprétation des résultats en termes moraux pour le grand public. Comment diminuer ou pondérer ce risque ?

— Certaines recherches, comme celles sur les biais implicites, tendent à détecter ou mesurer des traits psychologiques que l'individu lui-même ignore et pour lesquels il ne

59. Voir le détail de la discussion dans (Roskies 2016) [196].

donne pas toujours le consentement de recherche. [60]

— Le chercheur peut tenter de lire le contenu des états mentaux contre la volonté de l'individu, par exemple dans le cas de la détection du mensonge. La méthode peut d'abord être écartée pour son manque de fiabilité, mais cela laisse entier le problème moral et juridique. Tout d'abord est-il acceptable de lire ainsi dans l'esprit d'autrui contre sa volonté ? Et, à supposer ce principe accepté, qui fixe le taux d'erreur qui rend la méthode praticable et comment ce taux est-il mesuré concrètement ?

— Certaines interventions lourdes sur le cerveau peuvent modifier profondément le comportement. Comment faire place à la notion d'intégrité de l'individu dans ce contexte ? Et, par exemple, si le patient donne son consentement dans un certain état mental (optimiste) mais le refuse après la modification (dépressif), faut-il prendre en compte le premier ou le second avis ?

— L'utilisation des techniques dans des objectifs commerciaux, neuromarketing ou neuroéconomie, est-elle acceptable et, si oui, comment faire le départ entre la mise au point de nouveaux produits ou services et la manipulation abusive ?

Cette liste ne fait qu'effleurer l'ensemble des sujets à dimension éthique soulevés par le développement de nos capacités à comprendre le fonctionnement du cerveau humain et à agir sur lui. Il est significatif de constater que, alors que ma réflexion dans la présente thèse porte sur le potentiel apport des sciences expérimentales aux questions de philosophie morale, nous avons ici le constat d'une incidence d'une autre nature. C'est l'importance de la réflexion éthique à élaborer pour pouvoir mener à bien les recherches envisagées liées à la connaissance du cerveau. Le philosophe qui venait chercher des voies de réponse trouve d'abord une multiplication des questions !

En adoptant le point de vue de la neuroéthique, il ne peut y avoir d'impossibilité de principe à la recherche d'une telle connaissance du fonctionnement cognitif. La connaissance est un bien, implicitement ou explicitement, mis au dessus de toute autre valeur et la démarche scientifique n'a donc pas à être justifiée moralement. En revanche, elle doit être encadrée et les abus combattus en mettant en place des institutions qui vont les contrôler. C'est donc par des points de méthode que la difficulté éthique sera résolue avec, par exemple, le développement des comités d'éthique qui seront dépositaires du droit de vie sur les recherches et du contrôle de leur bon déroulement. Remarquons que cette proposition revient à adopter pour les décisions liées à la recherche scientifique la solution de l'éthique appliquée évoquée plus

60. Le chercheur est souvent contraint à cette tromperie sur l'objet de l'enquête. Si le participant devait donner son consentement avant l'expérience, alors il en connaîtrait l'objet et dans de nombreux cas, cela créerait un biais qui rendrait les résultats inutilisables.

haut. Remarquons également que cette solution est politique, au sens où elle permet d'afficher un mode de débat et de résolution de conflits lisible par les médias, les élus et les organes de financement dont dépend la recherche. Remarquons enfin que les conclusions portées plus haut pour l'éthique appliquée s'imposent : la résolution de conflit sera actée quand il y aura accord entre les parties prenantes, ici les chercheurs, les membres des comités d'éthique et les différents corps intermédiaires qui auront pris part au débat, sans que les questions morales substantielles ne soient pour autant nécessairement résolues ni même abordées.

Remarquons enfin le risque de circularité de la démarche. L'existence même des comités d'éthique présuppose qu'un compromis soit possible, et que la recherche de ce compromis soit nécessaire pour répondre aux questions qui lui sont soumises. Les membres du comité d'éthique sont choisis pour être compétents et concernés, leur mission est principalement de limiter les dégâts et de ramener l'ambition épistémique à un niveau admissible par la société et les élus qui décident de financer la recherche. On peut donc remarquer qu'il est peu probable que des personnes opposées pour des raisons morales à la démarche scientifique appliquée au comportement humain y participent[61] et, donc, les décisions iront circulairement dans le sens d'un compromis possible entre science et morale. Cet argument de la circularité pourra être ensuite utilisé par les opposants farouches pour critiquer la méthode et ses résultats.

Les neurosciences de l'éthique

La seconde dimension de la neuroéthique, les neurosciences de l'éthique, regroupe toutes les utilisations des technologies en vue d'analyser les corrélats neuronaux des comportements moraux. On peut utilement la rapprocher du mouvement de la psychologie morale que j'étudierai plus loin dont elle ne se distingue que par les outils utilisés. En regard de ces outils, et principalement de l'imagerie fonctionnelle qui permet de relier comportements, états mentaux et signatures physiques de l'activité neuronale, plusieurs remarques sont nécessaires. Première remarque, la complexité, la nouveauté, le faible ratio signal-bruit, la sophistication des techniques utilisées induit une certaine fragilité des résultats. Il convient de mettre en regard de cette fragilité l'importance des enjeux moraux et sociétaux auxquels elles prétendent apporter des arguments. La prudence s'impose mais la pression médiatique conduit à l'imprudence.

Deuxième remarque, le lien est déjà complexe, ténu et fragile entre le comportement et ce que mesurent les neurosciences, on peut donc affirmer que, au moins à ce jour, les caté-

61. A contrario, des opposants farouches pourraient décider d'y participer dans le but d'en empêcher le fonctionnement fluide, une étude expérimentale serait à mener pour analyser les différentes stratégies en corrélation avec les théories morales et religieuses justifiant du refus de la démarche scientifique.

gories morales n'ont simplement pas de contreparties neuronales. A titre d'illustration, si on constate une certaine corrélation entre certaines zones du cerveau et des réactions émotives, on ne sait pas de quelle émotion il s'agit et l'interprétation morale est très difficile [62]. Voir également le constat de cette même complexité dans (Houdé, Mazoyer et Tzourio-Mazoyer 2010 page 606) [120] « Il est impossible d'établir une correspondance biunivoque entre les aires cérébrales et les modules de traitement cognitif (. . .). ».

Troisième remarque, les corrélations fragiles observées par les neuroscientifiques doivent faire l'objet d'inductions pour pouvoir être interprétées en termes de concepts et relations des différentes théories morales. Or en prolongement des considérations sur l'opérationnalisation dans un chapitre précédent, cette interprétation supposerait que l'on ait une opérationnalisation satisfaisante de ces notions morales en terme de comportements observables et une opérationnalisation satisfaisante de ces comportements en signaux neuronaux. Aucune de ces deux étapes n'étant franchie, l'interprétation morale des résultats expérimentaux ne peut qu'être abusive.

Enfin, Kathinka Evers nous propose dans son ouvrage sur la Neuroéthique[81] d'adhérer à une conception métaphysique particulière, le « matérialisme éclairé », de façon à pouvoir ensuite entreprendre des recherches incluant des interprétations morales des résultats [63]. Indépendamment de l'accord ou du désaccord de chacun avec cette thèse métaphysique, il est ici clair que ce n'est pas l'approche scientifique qui donne un contenu moral aux résultats mais bien les a priori métaphysiques du chercheur qui permettent l'interprétation des résultats comme ayant une tonalité morale. On rejoint ici un des points importants de notre analyse : chaque position métaphysique donnera lieu à des interprétations différentes des résultats expérimentaux parce que la base ontologique est aujourd'hui insuffisante pour permettre une opérationnalisation, et donc une interprétation inductive, même partielle et falsifiable, des résultats.

6.3.3 La psychologie morale

L'appellation « psychologie morale » peut sembler étonnante. La psychologie s'intéresse à toutes les dimensions de l'action humaine et la morale n'en est qu'une particulière qu'il semble étrange de distinguer ainsi au premier abord pour constituer un domaine à part. On

62. Voir par exemple la thèse de W Aubé pour la complexité des réseaux neuronaux mis en œuvre dans chaque type de réaction émotive (Aubé 2014) [15]

63. Le matérialisme éclairé s'oppose au réductionnisme naïf, au dualisme et au behaviorisme. Dans la théorie proposée par Kathinka Evers, tous les processus mentaux sont supportés par des mécanismes physico-chimiques et la conscience est une fonction biologique résultant de l'évolution. Le cerveau est un organe « projectif, variable et actif de manière autonome, dans lequel les émotions et les valeurs sont incorporées en tant que contraintes nécessaires » (p 74)

imagine mal, pour forcer le trait avec une analogie, une « psychologie de la perception du rouge » qui pourrait s'autonomiser de l'étude de la perception en général et de la perception de la couleur en particulier. Cette appellation de psychologie morale ne peut être comprise que dans le contexte actuel de la psychologie, en rupture avec les conceptions du siècle passé qui tenaient les théories morales éloignées du champ de la science psychologique naissante. Revenons sommairement sur ces étapes avant de préciser ce que visent actuellement les psychologues qui se sont regroupés sous ce thème.

La psychologie scientifique moderne est habituellement dite débuter dans la deuxième moitié du 19e siècle avec William James (1842-1910). Le lien à l'expérimentation est d'emblée important avec le développement de la psychophysique avec Gustav Fechner (1801-1887). Toutefois, deux grands mouvements, le behaviorisme et le fonctionnalisme, vont faire que, pendant la majeure partie du vingtième siècle, la psychologie va se tenir à l'écart de l'approche expérimentale directe concernant les fonctions supérieures de la cognition humaine (Houdé 2010 page 18) [120], soit parce qu'elle est considérée comme illusoire du fait des biais de l'introspection, soit parce qu'elle est inutilement complexe, du fait de la multi-réalisabilité [64]. La dimension expérimentale se concentrera alors sur les questions de développement avec par exemple les travaux de Piaget (1932), puis Kholberg, de pathologies, d'études de haut niveau des fonctions cognitives, mais sans rentrer dans le fonctionnement interne du cerveau lui-même.

Les années 1980 1990, avec le développement de l'imagerie neuronale fonctionnelle, ont créé les conditions pour que soit à nouveau possible d'envisager des approches expérimentales directes du comportement humain, dont son comportement moral, appuyées sur une observation à la troisième personne des processus cérébraux. L'article de Joshua Greene de 2001 est un marqueur de ce moment : il a utilisé l'IRMf dans le cadre de l'étude du comportement face à un dilemme moral (Greene 2001) [108]. L'ouvrage de 2010, « The moral psychology Handbook » offre un panorama de ce nouveau champ de recherche à l'intersection de la philosophie et des sciences humaines expérimentales (John M. Doris (ed.) 2012) [76].

Cet ouvrage, dont les auteurs sont pour partie les mêmes que ceux de l'entrée sur la psychologie morale dans la Stanford Encyclopedia of Philosophy (John Doris et al. 2017) [77], établit un lien étroit entre psychologie morale et expérimentation. Il décrit ainsi la discipline académique que l'on doit considérer si on répond positivement à la question de l'apport possible de l'expérimentation à la philosophie morale. Si j'atteins mon objectif, la présente thèse

64. Pour les fonctionnalistes, l'important est l'étude des fonctions cognitives elles-mêmes car celles-ci peuvent être réalisés de façons différentes dans différents contextes matériels, par exemple le cerveau ou un ordinateur, le détail de cette réalisation important peu.

devrait pouvoir être lue comme l'étude du rapport entre la philosophe morale et la psychologie morale définie par John Doris et collègues.

Pour préciser un peu plus avant ce point, reprenons la démarche que proposent les psychologues moraux expérimentaux. Tout d'abord, l'objectif général est bien de réaliser des études de psychologie expérimentale permettant d'apporter des arguments empiriques significatifs dans les débats de philosophie morale. Ensuite, pour cela, il convient d'établir pour chaque théorie morale des conséquences comportementales qui soient assez concrètes et précises pour être soumises à l'enquête empirique. La mobilisation multidisciplinaire des psychologues, des anthropologues, des historiens, des neuroscientifiques, (...) peut alors permettre de construire les études empiriques nécessaires. Pour les auteurs, il est clair que cette démarche ne concerne pas tous les problèmes moraux, certains ne peuvent en effet être ainsi spécifiés pour pouvoir être testés. De plus, même pour ceux pouvant l'être, la dimension empirique ne saurait épuiser le sujet éthique. Néanmoins, les auteurs soulignent qu'une théorie morale qui ne serait pas en adéquation avec les résultats empiriques serait plus difficile à justifier[65].

Remarquons pour finir que ces descriptions méthodologiques de la psychologie morale ne s'appesantissent ni sur le problème de l'engagement moral du chercheur ni sur celui de l'engagement social d'une telle recherche. A l'image des scientifiques cités plus haut, ils reportent ces problèmes à l'extérieur de leur périmètre de préoccupation vers les philosophes moraux et, plus largement, vers ceux qui auront à utiliser leurs résultats empiriques. Pour montrer l'ambiguïté de cette position, on peut imaginer une société structurée autour d'une certaine morale empiriquement faible mais socialement forte. La démonstration de son inadéquation empirique risque de nous apporter une connaissance, mais au risque d'apporter le chaos social. Que faut-il privilégier ?

65. Reprenons ici la formulation de John Doris :

> The plausibility of its associated moral psychology is not, of course, the only dimension on which an ethical theory may be evaluated; equally important are normative questions having to do with how well a theory fares when compared to important convictions about such things as justice, fairness, and the good life. Such questions have been, and will continue to be, of central importance for philosophical ethics. Nonetheless, it is commonly supposed that an ethical theory committed to an impoverished or inaccurate conception of moral psychology is at a serious competitive disadvantage.

La plausibilité psychologique n'est bien sûr pas la seule dimension selon laquelle évaluer une théorie morale. Les questions normatives sont également importantes, en regard des implications de cette théorie en matière de justice, d'équité, et de la vie bonne. Ces questions ont toujours été et resteront centrales pour l'éthique philosophique. Néanmoins, une théorie morale dépendante de conceptions psychologiques faibles ou imprécises a un handicap sérieux face aux théories concurrentes.

6.3.4 Une science empirique a-normative ?

Les préoccupations éthiques sont aujourd'hui extrêmement présentes dans les institutions scientifiques. Mais, comme l'illustre le guide du COMETS, elles mettent beaucoup l'accent sur les problèmes de méthode et de respect de la déontologie du chercheur en tant que garant de la confiance dans la démarche scientifique et peu sur le problème, peut-être très particulier et restreint, de l'adéquation de la démarche scientifique expérimentale en regard de la dimension éthique des comportements humains.

La science expérimentale peut, en extrapolant cette dissymétrie, avoir la tentation de se libérer du problème de la normativité en se limitant à la connotation descriptive de sa recherche. Une telle science « a-normative » permettrait de décrire le monde sans aborder les aspects évaluatifs. Mais ce choix, bien qu'attrayant par la simplification qu'il apporte et les succès qui lui sont liés dans les sciences physiques, est voué à l'échec pour les sciences humaines pour au moins trois types de raisons. Premièrement, on voit mal comment le comportement humain pourrait être décrit en faisant totalement abstraction des évaluations portées en permanence par ces humains sur les situations qu'ils vivent. Deuxièmement, la science est également une activité humaine et sociale, à ce titre, elle comporte nécessairement des normes et, en particulier celles permettant l'évaluation de ses propres critères de pertinence. Et enfin, troisièmement, lancer une étude scientifique sur un sujet donné, c'est déjà choisir un sujet et en écarter d'autres, ce qui est un choix reflétant, de fait, évaluations et préférences.

La possibilité d'une science « a-normative » apparaît donc comme bien compromise de façon générale pour l'ensemble des domaines scientifiques et, plus particulièrement pour ce qui m'intéresse ici, la psychologie et son lien à l'étude du phénomène moral. Dans cette perspective, je propose de faire une place à part à la théorie de l'évolution. Cette théorie ayant l'ambition de proposer un cadre explicatif valide pour tout le monde vivant, elle ne peut, comme les autres sciences, renvoyer vers l'extérieur d'elle-même le problème de la normativité. La théorie de l'évolution doit assumer de prendre en charge la normativité pour que son cadre puisse acquérir une certaine plausibilité explicative de la nature humaine. Et, effectivement, la théorie de l'évolution a donné lieu à de nombreuses propositions récentes pour analyser et expliquer le comportement moral en tant que trait résultant de l'évolution de notre espèce. Je détaillerai ci-après en particulier la thèse de Nicolas Baumard qui propose une théorie naturaliste et mutualiste de la morale. De plus, l'approche évolutive me conduira à distinguer différentes échelles de temps et, en prolongement de cette distinction, à proposer de distinguer entre l'existence des normes morales, qui relève des temps longs de l'évolution

biologique des espèces, du contenu précis de ces normes, qui relève des temps beaucoup plus courts de l'évolution sociologique des groupes humains.

6.4 Les théories évolutionnistes de la morale

6.4.1 Présentation des thèses évolutionnistes de la morale

La thèse de Nicolas Baumard [66] propose une théorie naturaliste et mutualiste de la morale. Elle consiste principalement à appliquer les outils de la théorie de l'évolution à la coopération mutuelle entre individus d'un même groupe pour expliquer le caractère autonome spécifique et inné du sens moral. Cette thèse s'inscrit en prolongement des nombreux travaux qui visent à expliquer l'apparition de la morale en s'appuyant sur les avantages évolutifs qu'elle procure pour la coordination de grands groupes d'individus humains. [67]

Reprenons ici schématiquement les conclusions de cette thèse :

— Les philosophies du sens moral [68] décrivent de façon convaincante, et mieux que les autres théories morales, les phénomènes liés au comportement moral.

— La nécessité de la collaboration au sein de l'espèce humaine conduit chacun à participer à un marché de l'entraide.

— La réussite sur ce marché passe par le développement du comportement moral en tant que garant a priori du respect d'un contrat d'entraide.

— Les personnes ayant un tel comportement moral ont plus de facilités pour trouver des partenaires, survivre et se reproduire.

— La thèse propose ainsi d'élargir la philosophie du contrat de la justice (Rawls 2009) [191] vers la morale. [69]

— Cette théorie naturaliste mutualiste décrit mieux la réalité des comportements que les théories utilitaristes.

— En particulier, la punition en cas de manquement moral s'interprète comme un rétablissement de l'équilibre mutuel indispensable à la crédibilité du contrat d'entraide.

66. (Baumard 2008) [19]

67. Pour une version récente extrême de cette proposition, voir « From Bacteria to Bach and back » de Daniel Dennett (Dennett 2018) [70]. Pour une revue détaillée, voir l'article « Evolution of Morality » d'Edouard Machery et Ron Mallon dans (Doris (ed.) 2012)

68. Les humains seraient dotés d'un sens moral leur permettant de percevoir de façon automatique et innée, comme on perçoit une couleur, le caractère moral ou non d'une action. Cette capacité peut ensuite être développée et éduquée pour mener à l'excellence morale. Pour une introduction détaillée voir (Jaffro (ed.) 2000) [123] et (Wilson 1993) [236]

69. La définition de Rawls, la justice comme équité, s'appuie sur des principes qui seraient choisis par des agents rationnels dans une situation originelle placée sous le signe du « voile d'ignorance », chacun ignore quelle sera sa place dans la société qu'ils conviennent d'organiser au mieux. Pour Nicolas Baumard, cette même situation contribuerait également à définir les meilleures règles morales possibles.

— La morale est (probablement) propre à l'homme et, si c'est le cas, cela pourrait être dû à la fois au très haut niveau de collaboration au sein de l'espèce et à la richesse des interactions liées au langage qui font de la prévisibilité du comportement un atout décisif.

La thèse développe de façon convaincante la proposition que l'évolution naturelle d'une espèce sociale à très haute intensité d'interaction collaborative entre individus conduit à favoriser les individus avec des comportements moraux, garantie pour les autres individus du respect des engagements, et donc de la fiabilité du potentiel partenaire. Les arguments développés en faveur de cette thèse sont principalement liés aux conséquences qu'on peut en tirer et qui semblent bien expliquer les caractéristiques empiriques du comportement moral. Caractéristiques qui sont plus difficiles à expliquer dans le cadre des autres théories morales. De plus, l'histoire racontée par Nicolas Baumard est, prima facie, crédible car elle est supportée par les analogies avec notre vie sociale quotidienne : l'équipe de football dont les joueurs ne collaborent pas ne gagne jamais.

La thèse de Stéphane Debove [68], ultérieure à celle de Nicolas Baumard, développe un autre type d'arguments s'appuyant sur les simulations d'agents virtuels. Elle confirme la pertinence de la thèse qui associe comportement mutualiste, développement du sens moral et succès évolutif d'une population ainsi que des vertus morales au sein de cette population.

L'ouvrage de Michel Puech sur l'évolution de l'éthique dans un monde caractérisé par l'ampleur de l'usage des technologies (Puech, 2016) [186] montre que l'apport du raisonnement évolutif ne s'arrête pas à offrir une explication des origines du phénomène moral,il montre également comment, en adoptant le point de vue d'une triple coévolution des espèces non humaines, de l'espèce humaine et des technologies, ce même raisonnement permet d'expliquer ce qu'est, et devrait être, l'évolution de l'éthique aujourd'hui.

6.4.2 Pouvoir explicatif et difficultés de ces thèses

J'ai pris la thèse de Nicolas Baumard comme exemple de l'utilisation de la théorie de l'évolution pour expliquer la genèse du sentiment moral. D'autres théories, également appuyées sur l'évolution, mettent l'accent sur différents mécanismes précurseurs de la morale. Pour Jesse Prinz, ou Jonathan Haidt les origines de la morale sont à rechercher dans les différentes émotions, le dégoût, l'empathie, la peur, ... qui se seraient développées en comportements socialisés que nous appelons moraux. [70]

70. On peut toutefois remarquer que l'introduction par Jonathan Haidt de plusieurs (5 ou 6 selon les articles) modules moraux différents pose la question de la pertinence de cette proposition d'une même origine pour les différents modules.

Ces thèses diffèrent dans le détail des mécanismes mais ont de fortes similarités. Tout d'abord, à la base de ces théories, il y a un même paradoxe à résoudre : être moral semble être plus coûteux que d'être égoïste et, dans un contexte de compétition entre individus, contre-intuitif. Il faut donc proposer un bénéfice, susceptible de compenser ce coût, recherché et obtenu par la personne morale.

L'humain est un animal social et les groupes d'humains s'organisent avec un haut niveau de collaboration. La sélection naturelle a privilégié des traits favorisant cette collaboration ainsi que, par exaptation [71], des comportements bénéficiant de cette collaboration.

Le développement du comportement moral est à la fois un outil pour fiabiliser la collaboration au sein des groupes et une conséquence de la transformation d'émotions individuelles en phénomènes de groupe. Cette articulation qui se retrouve dans les différentes théories évolutionnistes, permet d'expliquer leur pouvoir explicatif. Le haut niveau de coopération et de cognition qui caractérise les groupes humains suggère que soient collectivisées au niveau du groupe les règles de comportement nécessaires au maintien du groupe, c'est la base du comportement moral. Les arguments en faveur de cette thèse peuvent être de deux types, cohérence avec les comportements observés et confirmation par des outils de simulation.

A supposer que ces arguments soient acceptés, ils instruisent la possibilité du développement historique du domaine moral. En revanche, il est très difficile à partir de ces seuls éléments de bâtir un lien causal précis entre les contraintes liées à l'évolution et l'apparition du phénomène moral du fait, entre autres, de ne disposer que d'une seule observation de l'apparition du phénomène réel humain. Naturellement, cette absence de mécanisme causal précis n'empêche pas de considérer que, raisonnement par abduction, nous n'avons pas de meilleure alternative pour expliquer l'évolution de l'espèce humaine vers le comportement moral que nous observons aujourd'hui.

Les difficultés des théories évolutionnistes

Les théories évolutionnistes de l'apparition de la morale ont néanmoins des difficultés explicatives sur plusieurs plans. Tout d'abord, les explications sont plus ou moins psychologiquement acceptables par les individus selon les conceptions morales en place dans chaque société. En particulier ces théories ne font sens que s'il y a un paradoxe au comportement moral coûteux. Si l'homme est sur terre non parce qu'il a réussi à évoluer mais parce qu'il a échoué à rester au paradis, la première pierre de ces théories disparaît. Les théories évolutionnistes de la morale auront plus de facilité à convaincre les philosophes déjà acquis à la

71. L'exaptation est un phénomène proposé par Jay Gould (Stephen Jay Gould et Elisabeth S. Vrba 1982) [217] consistant à justifier d'un trait non par son utilité directe qui conduirait à sa sélection mais par le fait qu'il est en lien, par exemple structurel, avec un autre trait qui a été sélectionné de façon indépendante.

réalité de l'évolution.

Acceptons l'évolution pour plus qu'une hypothèse. On doit alors constater que tous les groupes humains qui existent aujourd'hui ont évolué puisqu'ils existent. Chacun a ses règles morales et le constat de leur diversité s'impose. L'évolution ne peut s'appliquer pour les règles morale à l'échelle des espèces, car s'il en était ainsi, comment l'évolution pourrait-elle justifier les parties des règles morales qui diffèrent au sein d'une même espèce ? Mais, inversement, si on suppose que l'évolution a lieu au niveau du groupe, de façon à justifier ces différences de règles morales, comment justifier de la très grande homogénéité entre les groupes, que ce soit sur le plan du comportement moral ou sur tout autre critère ?

Les théories évolutionnistes de la morale auront pour répondre à ces questions plusieurs stratégies possibles. D'abord nier la diversité, et montrer que sous une diversité apparente il y a une grande uniformité des règles morales humaines [72]. Ou bien accepter la diversité mais complexifier le scenario avec un premier temps où l'évolution est au niveau de l'espèce et un second où l'évolution est au niveau de chaque territoire, au prix de devoir alors revoir tous les arguments qui fonctionnent avec une évolution au sens biologique au niveau de l'espèce. C'est cette seconde voie que j'explore plus loin.

Autre difficulté des théories évolutionnistes, chaque règle morale est, empiriquement, souvent et en permanence, violée par les humains et pourtant maintenue en tant que règle. La théorie devra justifier à la fois les deux phénomènes. La théorie selon laquelle la fiabilité du partenaire aurait pour bon indicateur son comportement moral semble ici être en concurrence avec de nombreux autres comportements qu'il conviendrait de détailler pour justifier par exemple de la solidarité au sein des mafias ou dans le cadre du phénomène du bouc émissaire.

En élargissant la question du comportement moral aux autres espèces, il semble que nous soyons confrontés à une difficulté : d'un côté nous savons qu'il y a de la coopération dans d'autres espèces sociales et, d'un autre côté, nous avons tendance à penser que le phénomène moral est proprement humain. L'analyse avancée de cette question reste à faire et ne pourra être résolue simplement par les théories évolutionnistes telles que je les ai évoquées. La réponse de Nicolas Baumard qui positionne la différence dans le haut niveau cognitif des humains et dans le langage conduirait à proposer que c'est moins la morale, en tant que guide comportemental, que son expression dans une formalisation parlée et partagée qui est le propre de l'humain.

Enfin, il convient de remarquer les faiblesses récurrentes des raisonnements appuyés sur

72. Pour une telle approche voir l'article récent d'anthropologie (Curry, Mullins et Whitehouse 2019) [60]

la sélection naturelle : un trait existant aurait nécessairement des caractéristiques ayant conduit à ce qu'il soit sélectionné. Ce raisonnement est faible car il occulte au moins deux autres possibilités. D'une part, un trait peut exister simplement par hasard et parce qu'il n'a pas de conséquence suffisamment pénalisante pour que, dans le contexte écologique où l'espèce s'est développée et a vécu, il n'ait pas survécu. D'autre part, le trait peut résulter d'une exaptation : il est, par exemple, structurellement lié à un autre trait qui a été sélectionné tout en n'apportant aucun avantage évolutif en lui-même.

Dans cet esprit, il conviendrait de constater d'abord une évidence : le comportement moral est un trait qui, dans les différents contextes dans lesquels c'est développé homo sapiens, n'a pas constitué un handicap tel que l'espèce ait disparu. On rechercherait ensuite les autres caractéristiques humaines, par exemple cognitives et sociales pour reprendre l'argumentation ci-dessus, qui semblent structurellement liées au comportement moral, sans préjuger d'un lien causal. Et on aurait alors à construire tous les scénarios logiquement possibles pouvant potentiellement mener à la situation observée et, ensuite, tenter d'obtenir les informations empiriques ou analytiques tendant à les départager. Pour simplifier la présentation de cette démarche, prenons deux exemples de scénarios construits pour être très contrastés.

Premier scenario, celui implicitement ou explicitement retenu dans les théories évolutionnistes. Le sens moral est apparu lorsque les primates ont évolué pour donner naissance à l'espèce humaine. Les groupes et individus ayant un comportement moral ont alors eu un avantage compétitif tel qu'ils ont pris le dessus sur les autres groupes dans la course à la survie de l'espèce. Sur ce sens moral servant de base c'est ensuite bâti le domaine moral grâce aux grandes capacités cognitives, langagières et sociales des humains.

Deuxième scenario, un hasard génétique ou épigénétique a conduit à une désynchronisation de la croissance de certains tissus, dont le cortex. Cela s'est traduit par une croissance exagérée du cortex qui, n'ayant plus de place dans la boîte crânienne, s'est replié sur lui-même, mettant ainsi en proximité de nombreuses zones du cerveau [73]. Pendant une longue période cette modification n'a eu aucun effet fonctionnel, ni positif ni négatif, et la modification s'est déployée simplement par un effet de dérive spontanée. A un certain moment, suite à des modifications brutales, rapides et répétées de l'environnement liées à des changements climatiques et à des migrations, les individus dotés de la modification ont pu réagir plus efficacement grâce à une capacité d'abstraction et d'apprentissage plus rapide. Ils se sont alors regroupés entre eux, et ont établi les bases d'une nouvelle organisation utilisant toutes les

73. Ce phénomène est connu des scientifiques sous le nom d'hypertélie. Le parallèle est ici entre le cou de la girafe et le cortex d'homo sapiens, deux organes au développement démesuré.

capacités offertes par la richesse plus grande de la connectivité de leur cerveau. On peut par exemple penser qu'est alors apparue une organisation sociale avec une plus grande spécialisation des tâches et les outils intellectuels de coordination qu'elle suppose, comme le langage articulé. On peut ajouter, enfin, que la morale est ainsi apparue comme extension de l'empathie, une abstraction bâtie à partir de l'empathie. Dans ce second scénario, la morale est un trait peu distinctif, résultat d'une exaptation au rôle sélectif négligeable, sans plus de prérogative que les jeux du cirque ou la fiction des contes qui sont apparus au même moment. [74]

Outre ces deux scénarios, on pourrait certainement imaginer de nombreux autres hypothèses, puis, ayant ainsi ouvert un vaste champ des possibles mettant en œuvre tous les dosages entre le pur hasard, la sélection adaptative directe et l'exaptation, on pourrait ensuite réévaluer les différents arguments en faveur d'une théorie évolutionniste de la morale. Mon objectif n'est pas ici de défendre un scenario plutôt qu'un autre, et encore moins d'affirmer que ces scenarios sont plus ou moins historiquement vrais, mais simplement de souligner la difficulté de la tâche, sous l'angle épistémique. Premièrement, ces deux scénarios, et de nombreux autres, ne sont pas invraisemblables, et il est extrêmement difficile de les départager empiriquement compte-tenu, d'une part du fait que nous n'avons qu'un unique exemple historique à notre disposition, celui de l'espèce humaine, et, d'autre part, que les événements décrits n'ont laissé que peu de traces exploitables par les scientifiques des différentes disciplines concernées. Rechercher des indices dans ces traces laissées par l'activité humaine est difficile et, de plus, très difficile à interpréter. Pour illustrer cette difficulté, prenons l'exemple du début de l'écriture. Il semble que de nombreux traits se soient mis simultanément en place il y a environ 7000 ans : les débuts des civilisations à populations numériquement importantes, la sédentarisation, l'écriture, la religion, les nombres,... poser qu'un trait, moral ou non, est cause ou conséquence d'un autre apparaît comme peut-être indécidable.

Deuxièmement, les arguments philosophiques en place pour défendre les théories évolutionnistes de la morale actuelles s'appuient sur une vision par trop simplifiée de l'évolution qui facilite de multiples utilisations. On peut d'ailleurs constater avec Beaudoin Aubé (Aubé 2017) [14] que la théorie de l'évolution peut être adaptée à de nombreux objectifs et servir à peu près toutes les théories morales.

Troisièmement, il n'est pas indifférent de noter que le scénario qui est actuellement le plus mentionné soit celui qui donne le plus beau rôle à la spécificité humaine : il est bien tentant d'y voir un effet de ce que j'ai appelé plus haut le risque de circularité. L'homme, considérant

74. Je reprends pour partie ici le raisonnement de Philip Kitcher dans (Kitcher 2014) [134] mais je force le contraste entre les deux scénarios en ne le rejoignant pas dans la conclusion humaniste qui le conduit à donner un sens transcendantal au projet éthique humain.

la morale comme un sujet important, aura beaucoup de mal à considérer avec bienveillance les scénarios dans lesquels elle ne joue que les seconds rôles. Remplacer Dieu en tant que fondement de la morale par une modification génétique est déjà un pas difficile, comme l'a montré l'histoire de la réception de la théorie de l'évolution, mais considérer que la roue du paon et le comportement moral humain sont de même nature et de même niveau ontologique semblera à beaucoup viser plus à détruire qu'à expliquer le domaine moral.

Malgré ces nombreuses, et sérieuses, difficultés, les théories évolutionnistes fournissent un cadre dont la pertinence pour l'étude expérimentale du domaine moral apparaît prometteuse. Si, effectivement, le phénomène moral résulte de l'évolution, et quel que soit le sens précis donné à cette expression, il devrait être possible de bâtir des hypothèses sur les processus qui ont conduit à cette situation et, en prolongement de ces hypothèses, imaginer des expérimentations permettant d'en évaluer la plausibilité. Et des conséquences peuvent en être attendues par les philosophes moraux. Pour paraphraser la formulation de John Doris en regard de la psychologie morale : même si la compatibilité avec la théorie de l'évolution n'est évidemment pas la seule qualité que doive viser une théorie morale, celle qui ne saurait résulter de processus évolutifs plausibles aurait un handicap sérieux face aux théories concurrentes.

6.4.3 L'entrelacs temporel : espèce, groupe, individu

En présentant les théories évolutionnistes de la morale, j'ai été conduit à constater des difficultés qui, en somme, ne sont que le reflet de toutes les caractéristiques que j'ai exposées au préalable, la complexité, le risque de circularité et le problème de la normativité qui font du domaine moral un sujet particulièrement difficile à aborder. Les théories évolutionnistes mettent de plus en lumière une caractéristique particulière que je vais aborder maintenant. Le terme d'évolution s'utilise dans trois espaces conceptuels et disciplinaires différents que je me propose de dissocier : l'évolution biologique de l'espèce, l'évolution sociétale des groupes humains, et l'évolution de l'individu. Un même mot, mais des contenus et des mécanismes différents qui s'entrelacent de façon complexe.

L'évolution de l'espèce est du domaine de la biologie. Elle suppose des descriptions en termes d'apparition ou de modification d'organes, ou de fonctions de ces organes, et, dans le paradigme génétique, une description en terme de proximités et de modifications génétiques entre espèces phylogénétiquement proches ou ayant des ancêtres communs. Le temps se mesure alors en millions d'années. La théorie de l'évolution pose le principe que ces modifications

génétiques se font au hasard, que de nombreuses variations sont sans lendemain, produisant des individus non viables ou non reproductifs, et que, des multiples variations restant simultanément présentes, celles qui produisent un avantage sélectif dans certaines conditions prennent le dessus au fil des générations lorsque ces conditions apparaissent. Ce principe suffit à expliquer la diversité du vivant (Gould et Blanc 2006) [105] et la phylogenèse.

L'évolution des groupes humains semble d'une toute autre nature. On a d'abord des groupes intimement liés à la biologie, les groupes familiaux et les groupes tribaux qui sont les structures de base des animaux sociaux. Nous partageons ces structures de base avec nos cousins primates. Puis, dans un temps qui resterait à spécifier, de nombreux autres groupes se sont développés à différentes échelles, relations entre groupes tribaux de base se coordonnant sur un même territoire, relations linguistiques, relations économiques, ... Ces relations ont construit de multiples possibilités de coopération que homo sapiens a mis à profit pour parvenir à la situation historiquement connue : de nombreux groupes coexistant depuis la cellule familiale jusqu'à l'humanité entière en passant par les villages, les entreprises, les professions, les états, les communautés linguistiques, les clubs de football...

Cette évolution des groupes a probablement été contrainte par certaines caractéristiques biologiques, par exemple la distance parcourue à pied en une journée peut définir un territoire de relations [75], mais l'évidence de la diversité des situations actuelles plaide pour une large indépendance entre cette évolution des groupes humains et l'évolution de l'espèce. Le processus d'évolution des groupes que je dénommerai la sociogenèse est largement indépendant du processus de phylogenèse. Le temps se mesure alors en milliers ou centaines d'années pour les structures de fond et en dizaines d'années pour les structures plus superficielles.

Enfin chaque individu doit faire son chemin dans cet ensemble complexe des groupes humains qui lui sont accessibles. Il construit son identité par ses choix, plus ou moins contraints par son milieu social, de profession, de genre, de nationalité, de club de football ... [76]. Les groupes eux-mêmes peuvent aussi être créés, être développés ou être délaissés par les individus et les choix personnels ont ainsi une répercussion dynamique sur l'existence de ces groupes. On peut voir ici une analogie, à cette nouvelle échelle, avec le phénomène de la sélection naturelle des espèces, mais avec une différence importante : il s'agit maintenant de choix faits par les individus eux-mêmes en tant que constituants des groupes et non de processus biologiques de survie exogènes à l'espèce. C'est un des objets de la psychologie que de décrire ces processus d'ontogenèse qui voient l'individu choisir les groupes qui lui importent

75. De manière analogue, Napoléon a défini le département comme ce qui était à un jour de cheval du chef lieu.
76. Pour une description détaillée de ce point de vue, voir (Appiah 2018) [10]

et, simultanément, définir son identité.

Reformulons maintenant une difficulté à laquelle nous confronte la théorie de l'évolution face à la complexité du domaine moral : l'entrelacement des temps de la phylogenèse, de la sociogenèse et de l'ontogenèse. J'ai proposé de définir la théorie morale comme l'ensemble des règles générales sur le bien et le mal qui aident chacun à instruire des jugements dans les cas particuliers auxquels il est exposé. Le respect de ces règles est un indice significatif de la volonté des individus d'appartenir aux groupes qui sont construits autour de l'acceptation de ces règles. On peut donc s'attendre à ce que l'étude de la morale entrelace des préoccupations relevant des trois domaines. La phylogenèse, comme l'a montré Nicolas Baumard, rapproche la morale des caractéristiques de l'espèce humaine qui en font une espèce sociale à haute capacité de coordination et de cognition. La sociogenèse étudie la structure des groupes et le rôle des normes, dont les normes morales, pour leur cohésion et, en fait, pour leur existence même. Et enfin l'ontogenèse de l'individu qui construit son identité en choisissant les groupes qui lui importent, et en adoptant les normes qui les définissent.

Ce constat me permet maintenant de préciser en quoi la contrainte normative pèse sur toute possibilité d'approche expérimentale du phénomène moral : outre sa complexité et sa potentielle circularité qui m'a déjà conduit plus haut à formuler la nécessité de l'humilité, de la patience, et du choix d'une organisation facilitant la multiplication des points de vue, la contrainte de passer du « être » au « doit être » se décline aux trois niveaux d'être que sont l'espèce, le groupe, et l'individu et qu'il est hautement difficile de concevoir une approche expérimentale permettant de mettre ensemble biologistes, sociologues et psychologues dans cet exercice.

Il est possible que cet objectif soit inatteignable, mais, comme je l'ai évoqué plus haut à plusieurs reprises, fixer ainsi a priori une limite à ce que l'homme peut connaître de l'homme est, a minima, extrêmement difficile à justifier. Pour ma part, je me contenterai d'un résultat plus modeste : compte tenu des résultats précédents, proposer une distinction conceptuelle qui soit de nature à mieux structurer les études empiriques du domaine moral dans le prolongement des théories évolutionnistes : distinguer l'étude de l'existence des normes de l'existence de leur contenu, c'est l'objet de la section suivante.

6.5 Distinguer existence et contenu d'une norme

La section précédente a présenté le pouvoir explicatif des théories évolutionnistes. Malgré les difficultés soulevées par cette approche évolutionniste, les apports de ces théories

irriguent une partie chaque jour plus importante de la philosophie morale, tout particulièrement dans les perspectives descriptives et méta-éthiques. Je vais maintenant porter une attention toute particulière à une conséquence de ces théories que je pense importante pour l'étude expérimentale du phénomène moral. L'entrelacs complexe entre phylogenèse, sociogenèse et ontogenèse se traduit, comme je viens de le souligner, par la nécessité de faire collaborer des scientifiques d'horizons très différents, il se traduit également par un point qui n'a pas été assez souligné, à ma connaissance, dans la littérature morale : la nécessité qu'existe une norme, et le contenu précis de la norme retenue au sein d'un groupe particulier, relèvent de domaines différents tant sur le plan temporel que, par conséquence, sur le plan des disciplines concernées. La présente section a pour objectif d'insister sur ce point et d'en tirer les conséquences sur le plan de l'expérimentation.

6.5.1 L'existence des normes pour une espèce

Les théories évolutionnistes de la morale instituent un lien entre l'existence d'une espèce à haut niveau cognitif et de coordination entre individus interdépendants au sein de groupes de grande taille et l'existence de règles morales que se donnent ces individus pour induire mutuellement confiance dans la collaboration. L'argument évolutif est qu'une espèce qui aurait ces caractéristiques mais pas de comportement moral soit éclaterait d'elle-même sous l'action des égoïsmes, soit serait supplantée par une espèce concurrente pouvant s'appuyer sur un comportement moral pour améliorer son fonctionnement interne.

La question qui est maintenant la notre est de voir dans quelle mesure une approche empirique peut apporter des éléments pertinents en regard de ces théories. Et, à supposer qu'une telle approche soit possible, quelles précautions mettre en œuvre pour ne pas tomber dans les multiples écueils que j'ai évoqués plus haut. Tout d'abord, il convient de remarquer que nous n'avons qu'un seul cas attesté d'espèce ayant un comportement moral reconnu par tous : les humains. Trois solutions sont a priori possibles pour une approche empirique dans ce contexte. La première est d'en rester à l'étude anthropologique et historique de l'espèce humaine. La deuxième d'accepter un modèle animal proche et d'analyser les points communs et les différences de comportements présumés moraux. La troisième est de sortir du monde animal et de transposer les questions par une analogie dans un autre univers. Les univers virtuels des simulations informatiques par agents autonomes sont de tels univers, relativement faciles à construire et à moduler en fonction des analogies développées.

Dans le premier cas, l'étude de l'évolution humaine, nous disposons d'une variable indé-

pendante possible, les différents territoires qui ont vu l'apparition des divers groupes humains. Cette variable disparaît au niveau de l'espèce si on considère, avec la plupart des scientifiques, que l'espèce humaine est issue d'une seule spéciation, elle reste intéressante au niveau des divers groupes humains qui ont différentes règles morales. En supposant que l'archéologie nous fournisse des descriptions satisfaisantes de l'histoire des groupes humains sur les différents territoires, nous pourrons chercher à atteindre un résultat empirique important : ou bien il a existé de tels groupes sans développement moral, ou bien, inversement, tous les groupes s'appuyaient sur une morale partagée au sein du groupe. Arriver à une telle conclusion de façon définitive n'est certainement pas tâche facile, ne serait-ce que parce qu'on pourra toujours arguer qu'un groupe pour lequel on n'a pas de trace archéologique pourrait être l'exception recherchée, ou, plus simplement, parce que la morale laisse peu de traces repérables. Pourtant, à ce jour, il semble que nombre d'archéologues et, en ce qui nous concerne, tous les philosophes, considèrent ce point pour acquis et, avec lui, que la nécessité de l'existence de la norme morale est étayée : tous les groupes humains existant ou ayant existé auraient une morale. Cette position doit, empiriquement, beaucoup au fait que les anthropologues considèrent qu'elle est vraie pour tous les groupes qu'ils ont pu examiner, ce qu'on estimera relever du biais de confirmation si on souhaite défendre l'hypothèse contraire.

On pourrait aller plus loin en observant qu'une autre variable indépendante intéressante pourrait être apportée par les différentes espèces d'hominidés que homo sapiens a supplanté. Là encore il faudrait rechercher des traces permettant de décrire la part de comportement moral dans chacun des groupes et viser un résultat empirique qui pourrait être qu'homo sapiens avait, ou non, un comportement moral jouant un rôle pour une meilleure coordination au sein des groupes que les hommes de Néandertal ou de la grotte de Denisova, deux familles proches d'homo sapiens. Compte-tenu du peu de traces que nous avons pour étayer une telle conclusion, il est peu probable que cette voie soit fructueuse.

Le deuxième cas, le modèle animal, ne pourra par construction pas répondre à la question de la spécificité du comportement moral humain. Il pourrait néanmoins permettre d'étayer des hypothèses générales sur, par exemple, une corrélation entre le niveau d'organisation des groupes sociaux et le niveau d'altruisme chez les primates, ou sur la grande proximité de comportement entre les hommes et les chimpanzés [77]. Toutefois, et dans la mesure ou le fait moral s'appuie, comme je l'ai proposé plus haut, sur l'existence généralisée de discours comportant des termes normatifs et que seuls les humains ont la parole, on pourra toujours soutenir que cette capacité langagière est également la marque d'une capacité conceptuelle qui serait

77. Les travaux de Frans de Waal vont dans ce sens : (Waal 2016) [67]

la vraie marque du domaine moral. On pourra également soutenir l'inverse, à l'image des anti-spécistes comme Peter Singer (Singer et Rousselle 2018) [208].

Pour qu'un argument empirique puisse apporter dans ce débat, il conviendrait, en nous appuyant sur les conclusions des chapitres précédents sur l'opérationnalisation, d'imaginer des pratiques expérimentales et d'observation reflétant le comportement d'une espèce sociale évoluée avec ou sans normes morales pour lier les groupes. Ces pratiques devant être suffisamment discriminantes pour pouvoir interpréter ces normes comme morales. Compte tenu des débats sur le caractère moral ou non de certaines normes chez les humains [78], il semble très difficile qu'une telle opérationnalisation puisse être obtenue pour les animaux non humains.

Faute de pouvoir disposer d'approches empiriques de la question de la possibilité d'une espèce à haut niveau de coordination sans règles morales, on a envisagé d'utiliser pour cette étude les ressources de l'analogie avec des sociétés d'agents virtuels que nous pouvons caractériser à volonté. Cette approche est développée dans la thèse de Stéphane Debove (Debove 2015) [68]. Elle consiste à construire des modèles informatiques de populations d'agents virtuels indépendants dont on peut faire varier les caractéristiques, par exemple leur propension à collaborer, et on observe l'évolution des populations en fonction d'une situation de départ, souvent aléatoire, et en fonction des caractéristiques étudiées. Ces simulations visent ainsi à établir des liens statistiques virtuels entre les caractéristiques individuelles [79] et le comportement global des populations. A ce jour, et la thèse de Stéphane Debove en est un maillon, ces approches tendent à conforter la proposition que la coordination au sein de populations importantes d'agents indépendants est facilitée lorsqu'ils respectent des règles morales connues de tous. La validité externe de ces simulations, la possibilité de les extrapoler vers le monde réel, reste une question difficile que je ne détaillerai pas ici [80].

Toutes les approches que je viens d'évoquer étudient la convergence vers une proposition : l'existence des normes morales est un atout pour une espèce qui, comme l'espèce humaine, établit des réseaux de collaboration importants entre individus indépendants à haut niveau cognitif. Cette proposition n'est pas facilement vérifiable empiriquement, mais les approches tentées vont dans le sens de cette convergence.

Mais, et c'est le point essentiel sur lequel je veux insister, rien dans ces approches ne

78. Le débat sur le minimalisme moral dont un point récent est fait dans (Merrill et Savidan 2017) [162] oppose les tenants d'une morale réduite au minimum, c'est-à-dire ne concernant que des actes avérés ayant nui à un tiers non consentant, aux tenants d'une morale élargie aux actes seulement envisagés et aux actes envers soi-même. Ce débat complexe est non résolu pour les humains, et les avis de chacun dépendent, circulairement, des théories morales adoptées.

79. Les caractéristiques des individus peuvent être uniformes, ou être définies aléatoirement, et on comparera alors les résultats correspondant à différentes répartitions statistiques.

80. Voir par exemple (Manzo 2014) [153]

permet d'accéder au contenu substantiel des normes. Que les humains se coordonnent est assez pour que les groupes fonctionnent, et peu importe (au moins en première approche) sur quelle norme ils se sont alignés. Aucun des dilemmes moraux actuels ou passés, l'avortement, l'excision, la fin de vie, la procréation, l'homosexualité, la peine de mort... qui donnent lieu à des réponses différentes selon les époques et les lieux ne peut être abordé avec le seul outil de la nécessité de la coordination. Je propose maintenant de prolonger ce constat : ces approches évolutionnistes qui visent à faire un lien entre l'existence de la morale et la confiance en autrui nécessaire dans une société à haut niveau de coordination fonctionnent à l'identique quel que soit le contenu substantiel des règles morales pour peu qu'elles répondent aux règles spontanées très générales représentées, par exemple, par la règle d'or de la réciprocité. Il leur est donc impossible d'éclairer les règles morales substantielles adoptées par chaque groupe humain.

Cette impossibilité, qui n'est plus seulement pratique mais également de principe, signifie pour moi que nous gagnerons à dissocier deux types d'enquêtes empiriques. Les premières viseront à établir comment les normes viennent à existence, et répondent à des caractéristiques très générales qui en font des normes que l'ont peut dire proto-morales, et j'ai esquissé ci-dessus quelques pistes pour l'espèce humaine. Les secondes viseront à étudier les différents contenus possibles de ces normes et ces dernières que je développe ci-après ne peuvent relever de la théorie de l'évolution au sens biologique.

En regard de l'approche empirique, je propose donc que nous gagnerions à considérer la proposition suivante comme une hypothèse de travail pertinente [81] :

> Le contenu de la norme reste indéterminé au niveau de l'espèce, même si l'existence de la norme est établie comme nécessaire pour l'existence de l'espèce.

6.5.2 Le contenu des normes morales identifient les groupes

Dans son ouvrage sur l'identité des groupes, Anthony Appiah (Appiah 2018 page 29) [10] reprend l'expérience célèbre de 1953 dite « The Robbers Cave Experiment » dans laquelle 24 adolescents de même origine sont séparés aléatoirement en deux groupes à l'occasion d'une colonie de deux semaines [82]. Pendant la première semaine, chaque groupe ignore l'existence de l'autre et sont menées des activités en commun destinées à renforcer les liens au sein de chacun des groupes. La première semaine permet ainsi d'établir les habitudes et normes de

81. Je remarque qu'Edouard Machery arrive à une conclusion proche par des voies différentes en analysant ce qui peut être déduit de la proposition « la morale a évolué ». Voir son exposé « Did Morality Really Evolve ? » de 2013 sur https ://www.youtube.com/watch ?v=EeqqS4ktQno

82. Cet exemple est également repris par Jérôme Ravat dans (Ravat 2019) [189]

chaque groupe. Pendant la seconde semaine, les groupes sont mis en compétition. L'objectif initial de l'étude était de montrer que l'hostilité entre groupes survient rapidement quand des groupes sont en compétition pour des ressources rares, et l'étude a atteint cet objectif. Appiah en prolonge l'analyse en montrant comment dans cette expérience, les groupes ont changé de comportement à partir du moment où ils ont été mis face à un autre groupe. Tout d'abord, ils se donnent des noms, le label ainsi créé renforçant l'appartenance, puis rapidement, il y a essentialisation et attribution de caractéristiques propres à chaque groupe et, en parallèle, développement du comportement agonistique avec polarisation des différences.

Cette expérimentation s'appuie sur une opérationnalisation directe de la génération d'un groupe et, en parallèle, de la génération des normes qui lui sont propres. Elle donne un support supplémentaire à l'argument du caractère indissociable de l'existence du groupe et de l'existence des normes au sein de ce groupe. En revanche, et à nouveau, le contenu des normes apparaît comme arbitraire pour chacun des groupes : plus exactement, le choix des caractéristiques des groupes, et des normes, se fait pour optimiser les positionnement par rapport aux autres groupes. La seule contrainte est d'obtenir une différenciation.

La psychologie sociale représentée par l'expérience que je viens de présenter a mauvaise presse car elle souffre d'une très faible réplicabilité (Collaboration 2015) [50]. De plus, l'opérationnalisation naïve mise en œuvre dans cette expérience conduit à passer sous silence tout le contexte de ces garçons et de ces deux semaines, et, faute de cette réflexion préalable, l'interprétation inductive est fortement limitée par le doute sur l'influence de tous ces éléments de contexte sur les résultats de l'expérience. J'en retiendrai néanmoins la leçon : la genèse du contenu des normes peut être étudiée en tant que coproduit de la genèse des groupes et elle ne peut être comprise que différentiellement.

Ce constat relie le développement, l'évolution et la fin de vie de chaque groupe au développement, à l'évolution et à la fin des normes morales qu'il adopte, et réciproquement. C'est également le message que nous propose la fable des « mangeurs d'œufs par le petit bout et par le gros bout » de Jonathan Swift dans les Voyages de Gulliver : peu importe le sujet de désaccord pourvu qu'il permette aux hommes de se reconnaître, de se différencier et, finalement, de s'unir contre un ennemi commun. Comme le souligne Appiah, les règles morales lient les humains entre eux, positivement, au sein du groupe. Elles les lient entre eux, négativement, contre les membres des autres groupes. Et, de plus, pour être efficaces, elles les aveuglent sur le caractère arbitraire des différences dont l'importance n'est liée qu'à leur rôle dans ces constructions sociales.

Résumons-nous, l'existence des normes est l'objet, au niveau de l'espèce, d'approches empi-

riques d'inspiration biologique tendant à montrer que sont liées existence des normes et existence d'une espèce ayant des caractéristiques proches de celles des humains. Ces approches ne peuvent permettre d'accéder au contenu des normes, sauf à un niveau de très grande généralité. Les constats empiriques de la sociologie (et de la psychologie sociale) tendent à lier existence des groupes humains et existence de normes différenciées entre groupes. Ces approches permettent d'accéder non au contenu, qui reste arbitraire, mais à des processus de création qui portent des faisceaux de contraintes sur ces contenus. Et enfin, en suivant Appiah, les approches empiriques tendent à montrer que les individus sont aveugles au caractère arbitraire des règles morales qu'ils adoptent, en même temps qu'ils construisent leur identité en se déclarant appartenir aux groupes dont elles sont la marque. [83]

Les approches empiriques d'inspiration sociale ou psycho-sociale viseront à étudier ces trois points : la nature du lien dynamique entre groupes et normes, les faisceaux de contrainte qui en résultent pour les contenus des normes, et les caractéristiques des individus qui, tel l'aveuglement au caractère arbitraire du contenu de la norme, sont nécessaires pour que la morale joue le rôle régulateur que les modèles lui donnent. Ce sont donc ces processus complexes que l'enquête empirique portant sur la dynamique des normes devra aborder. Comme souligné plus haut, il ne s'agit plus de biologie mais de sociologie. [84]

L'analyse menée précédemment m'a permis de souligner l'importance de l'entrelacs des temps dans la réflexion sur l'évolution et j'en trouve ici à nouveau une illustration. Chaque groupe a une durée de vie et un empan territorial propre qui s'entrelacent avec ceux des autres groupes, concurrents ou non. Cet entrelacs a une première conséquence, la très grande complexité du sujet et la difficulté à concevoir des approches empiriques qui en discriminent les composantes. Elle a une autre conséquence importante : chaque individu ayant de multiples appartenances doit faire son chemin dans ce maquis de normes plus ou moins compatibles, plus ou moins contradictoires pour pouvoir construire sa personnalité morale. Ce sera l'objet du point suivant.

6.5.3 Le niveau individuel

Dans les approches empiriques d'inspiration biologique ou sociologique présentées ci-dessus, l'individu n'apparaît que comme appartenant à un groupe, appartenant à une espèce. Les

83. Remarquons que les philosophes expérimentaux ont confirmé empiriquement l'expression de cet aveuglement. Par exemple, dans l'expérience de l'inceste entre adultes consentants, les personnes qui le condamnent avancent des arguments rationnels en contradiction avec le cas présenté et, quand on leur souligne cette incohérence, la réponse est dilatoire : ils ne peuvent expliquer pourquoi ils jugent cet acte condamnable.

84. Je rejoins ici l'analyse de Georges Canguilhem (Canguilhem 2007 page 189) [37] qui propose dans « du social au vital » de faire un parallèle entre la société et l'organisme. Les normes morales joueraient pour la première, mais de façon moins rigide, le rôle que jouent les contraintes métaboliques pour le second.

procédures statistiques et la dynamique des groupes effacent les spécificités individuelles au profit de la vue collective. Les questions morales ne sauraient pourtant être complètement abordées sans que soit également développée l'étude de l'individu en tant que tel, et non en tant que membre d'un groupe ou d'une espèce.

Les questions qui relèvent principalement de cette dimension individuelle sont multiples :

— Comment un individu choisit-il ses différents groupes d'appartenance et les différentes normes associées ?

— Pourquoi suit-il les règles morales, quand il les suit, et pourquoi les viole-t-il, quand il le fait ?

— Comment peut-il arbitrer entre des règles contradictoires portées par des groupes différents (ou par le même groupe à des moments différents ou non) ?

— Comment aborder les phénomènes de désobéissance d'un individu, à l'intérieur d'un groupe, et les phénomènes de dissidence, quand un individu engage la création d'un nouveau groupe ?

Et, pour regrouper tout ceci d'une seule phrase, quand l'individu renonce-t-il à sa liberté au profit de son appartenance identitaire, exprimée par les choix moraux ?

Dans le schéma sous-tendu par ces questions, et pour que l'approche empirique puisse en être pertinente, l'individu est envisagé comme le résultat d'un processus d'individuation sans fin au cours duquel il choisit son chemin entre les différents groupes présents dans son champ social et, concomitamment, adopte les normes de ces groupes, dont les normes morales [85]. Les approches empiriques d'inspiration psychologiques (et psycho-sociales) viseront les nombreuses expériences possibles opérationnalisant les conditions de ce triple mouvement entrelacé traduit dans le comportement : individuation, appartenances, normes.

Remarquons encore une fois que chaque individu a des appartenances multiples : genre, pays, profession, région, couleur de peau, religion, famille, ... Le processus d'individuation va le conduire à mettre en avant certaines et à en gommer d'autres. La question des règles morales liées à chacune de ces appartenances, de leur incompatibilité et des dilemmes existentiels qu'elle peut générer, des choix à arbitrer en permanence est centrale dans cette perspective individuelle. [86]

Pour l'individu, l'existence des normes au niveau de l'espèce et le lien entre normes et

85. Dans son ouvrage « Éthique et polémique », Jérôme Ravat propose que l'expérience du désaccord moral constitue l'occasion pour l'individu, et concomitamment pour les groupes, où se forgent les identités. (Ravat 2019) [189]

86. A titre d'exemple très quotidien, en m'appuyant sur plusieurs dizaines d'années d'expérience dans le domaine de l'informatique, je remarque qu'il est impossible de comprendre les décisions des entreprises dans ce domaine sans prendre en compte la double appartenance des informaticiens, à l'entreprise, et au monde informatique. Prosaïquement, leur présent est dans une entreprise, mais leur futur est, en tant qu'informaticien, dans une autre entreprise et ils auront alors besoin des autres informaticiens, fournisseurs, clients ou partenaires, avec qui ils auront tissé des liens.

groupes est une donnée de contexte qui s'impose. Elle s'impose sur le fond, l'individu humain ne peut survivre sans être dans son espèce et dans sa société. Elle s'impose aussi comme élément de toute opérationnalisation en vue d'une démarche empirique, la dimension morale du comportement ne peut être abordée sans prendre en compte les caractéristiques générales des normes morales qui sont dépendantes des contraintes apparues nécessaires pour qu'elles puissent jouer leur rôle biologique et social. L'approche empirique de l'individu choisissant parmi les chemins moraux qui lui sont disponibles pourra permettre de mettre en évidence un autre niveau de contrainte : il faut que les normes (et les groupes associés) lui apparaissent comme préférables à d'autres. Là encore la séparation entre l'existence de la norme et son contenu montre sa pertinence en regard de la démarche empirique, le choix individuel doit être fait, et ne peut pas ne pas être fait, c'est la nécessité de l'existence de la norme, mais le choix de l'individu est libre entre contenus qui sont arbitraires sous la contrainte d'être potentiellement choisis par les individus (et concomitamment par les groupes).

Remarquons enfin que mon analyse me conduit à proposer que le contenu des normes ne soit que faiblement contraint, et de façon très générale, par les données biologiques. Le champ des normes possibles est de façon plus importante restreint par les faisceaux de contraintes venant du rôle social qu'elles doivent jouer et ces contraintes sont encore resserrées par la nécessité qu'elles soient préférées par des individus libres au sein de sociétés pluralistes. Le contenu reste, dans cette perspective, partiellement arbitraire mais devant respecter ces faisceaux de contrainte, et la démarche empirique a beaucoup à apporter pour définir ces contraintes et cette marge de liberté.

6.5.4 Remarques conclusives, le projet éthique de Philip Kitcher

Chercher à construire des approches empiriques qui dissocient l'étude de l'existence des normes de celle de leur contenu me semble une contribution très puissante, même si elle reste partielle et modeste, face à l'énorme difficulté qu'il y a à aborder expérimentalement le domaine moral. Cette approche conduit à proposer, à l'image des poupées gigognes, trois domaines d'études qui, sans bien sûr être indépendants, peuvent être largement découplés, et donc facilités, en appliquant cette distinction.

A la biologie, et plus largement aux approches évolutionnistes présentées plus haut, la charge de montrer empiriquement que pour l'espèce humaine, avec ses caractéristiques, le comportement moral est un atout. A charge également de définir ce que signifie « comportement moral » dans ce contexte, et j'ai proposé que cette définition serait certainement assez

générale, laissant le contenu des règles morales largement indéterminé.

A la sociologie, et à la psychologie sociale, la charge de montrer et qualifier empiriquement le lien entre l'existence des groupes humains et l'existence des normes comportementales, dont morales. A charge également de redéfinir à cette échelle ce que signifie « moral » dans ce contexte, et j'ai suggéré qu'un faisceau de contraintes va s'en trouver généré qui restreint les contenus possibles pour les règles morales sans les déterminer complètement. Enfin à la psychologie le niveau individuel qui entrelace dans un processus complexe l'individuation, les appartenances à des groupes sociaux, et les normes morales de ces groupes. Là encore, il conviendra de redéfinir ce que signifie « moral » dans ce contexte et comment cette définition s'articule avec les définitions déduites des rôles biologiques et sociaux.

Ce cheminement évoque un processus complexe de construction progressive, sociale et psychologique, cheminement qui fait du domaine moral le résultat d'un processus où les trois niveaux biologique sociologique et psychologique s'entrelacent dans un projet sans fin. Pour conclure ce chapitre, soulignons que des qualifications morales opposées peuvent être données à ce cheminement. Les nihilistes moraux y verront la preuve que la morale dans ses règles particulières n'est que contingente, et que, dans ses règles générales, elle répond à la nécessité de coordination des groupes, comme la faim répond à la nécessité de s'alimenter. Mais pour d'autres, ce cheminement a un sens, celui de la construction elle-même, et que c'est ce sens qui donne la meilleure description de ce qu'est la morale : un projet sans fin par lequel les humains se construisent et se définissent. Philip Kitcher adopte ce second point de vue qu'il propose d'appeler le Projet Éthique (Kitcher 2014) [134].

Pour Philip Kitcher, les règles morales ne sauraient provenir d'un commandement divin, ni d'un fondement moral qu'il conviendrait de rechercher. Il propose de considérer, en se rapprochant de la philosophie pragmatique de John Dewey, que les humains en tant qu'être sociaux partagent des émotions avec leurs semblables. Ils disposent ainsi d'une empathie et d'un altruisme naturel qui ont pu être suffisants lorsque les groupes étaient restreints mais ont du être prolongés grâce à l'invention de la « technologie sociale » lorsque les groupes sont devenus des sociétés importantes. C'est cette technologie sociale qui a permis d'élaborer les morales, sous la forme particulière des religions, comme outil de cohésion sociale à l'échelle de ces sociétés devenues complexes. Le Projet Éthique qui a ainsi permis de passer d'un comportement guidé par un altruisme naturel, issu de l'évolution biologique, à une morale construite est aujourd'hui confronté à un nouveau changement de même ampleur : construire une nouvelle morale qui satisfasse aux exigences d'une humanité mondialisée et responsable de l'avenir de la planète. Philip Kitcher propose ainsi de voir les règles morales comme une

construction sociale toujours en cours, élément clé de ce qu'il nomme l'Humanisme Laïque (Secular Humanism).

Mon projet n'est pas ici de défendre la conception de Philip Kitcher face à celle des nihilistes moraux. Toutes deux sont des qualifications, certes opposées, du processus moral tel qu'il est aujourd'hui décrit de façon incomplète mais prometteuse par les théories évolutionnistes de la morale. Ces qualifications diverses montrent simplement que, même en acceptant ces théories, le travail du méta-éthicien ne sera pas clos pour autant, l'approche expérimentale ne pourra trancher entre ces visions opposées du cheminement moral. Mais ma recherche avait un autre but, celui d'étudier comment l'approche empirique peut apporter à l'étude du domaine moral. Et le constat sur lequel je veux insister ici est que de distinguer l'étude de la nécessité de l'existence des normes pour la régulation des groupes humains, de l'étude du contenu particulier de la norme dans chaque groupe, ouvre une nouvelle perspective à l'approche empirique. Le schéma dessiné par ces différents niveaux, biologie, sociologie et psychologie, chaque niveau donnant à voir des mondes possibles et des faisceaux de contraintes, sera empiriquement fécond s'il facilite la collaboration entre les spécialistes des nombreuses disciplines scientifiques dans l'objectif d'établir des opérationnalisations satisfaisantes des entités, propriétés et relations utiles aux théories en relation avec le comportement humain.

Chapitre 7

Point et perspectives

Ce dernier chapitre aurait pu, et peut-être aurait du, s'intituler « Conclusion ». Mais il n'en va pas ainsi car, d'une part, les questions posées au début de cette enquête ne sont pas de celles auxquelles on apporte des réponses conclusives. Même en s'attachant pendant de longs mois à tenter de les restreindre, les délimiter, les borner, ces questions s'échappent du cadre où on veut les contraindre. Le domaine moral est trop vaste, trop complexe, et son implication sociétale trop lourde, pour que toute proposition ne se heurte immédiatement à de nombreuses, et irréductibles, oppositions. Et, d'autre part, un chapitre de conclusion marquerait une fin, en contradiction avec un des objectifs à la base de ce travail : se préparer aux prochains développements des sciences expérimentales, aux prochaines générations d'imagerie cognitive dont nous ne savons ni ce qu'elles seront ni ce qu'elles apporteront exactement. Mais nous pouvons anticiper, et c'est un des résultats de ce travail en prolongement de l'expérience du mouvement XPhi de la philosophie expérimentale, que ces développements ne seront ni insignifiants, ni déterminants, pour les questions de philosophie morale. Ce chapitre ne sera donc pas une conclusion mais visera à faire le point du travail réalisé, présentera ensuite trois propositions pour des axes de travail à venir et, enfin, des observations et perspectives sur des futurs possibles pour l'articulation entre la philosophie morale et les démarches scientifiques expérimentales.

7.1 Les questions au point de départ de l'enquête

La psychologie scientifique se développe rapidement. Les sciences cognitives se développent rapidement [1]. Il était donc naturel que des philosophes s'emparent de ces dévelop-

1. Pour un état des lieux récent, voir (Collins 2018) [51].

pements et tentent d'en faire bénéficier leurs disciplines, c'est ce que s'est donné pour objectif le mouvement de philosophie expérimentale, dit XPhi, à la fin du siècle passé et surtout au début de celui-ci. Ce mouvement a soulevé de nombreuses réticences et, en particulier dans le domaine moral, il a été soutenu que les expérimentations apportaient peu aux débats moraux. Ces réticences, et les questions qu'elles soulèvent sont à la base de mon travail :

— Que peuvent apporter à la philosophie morale les connaissances empiriques acquises sur le comportement humain ?

— Qu'apporte que cette connaissance empirique soit établie selon une démarche scientifique expérimentale ?

— Et, de façon plus prospective, le développement des sciences cognitives se poursuivant avec de nouvelles méthodes d'investigation du comportement humain, que peut-on anticiper des incidences qu'elles auront sur les théories morales qui structurent le domaine moral ?

De façon moins générique, on peut reformuler ce questionnement sur la base de l'article de 2001 de Joshua Greene qui utilise les outils d'imagerie du monde médical dans le cadre de l'expérience du tramway (Greene 2001) [108], quand un participant se trouve confronté à la décision de sacrifier un passant pour sauver cinq vies. Cet article exploite les résultats d'une IRMf pour proposer un lien fort entre le contenu du raisonnement moral et les réseaux de neurones activés : les raisonnements moraux utilitaristes font appel à des zones du cerveau connues pour être le siège du raisonnement rationnel alors que les participants privilégiant des raisonnements moraux déontologistes font appel à des zones du cerveau connues pour être en relation avec les émotions. La question posée est alors celle du poids que l'on peut donner à cet article de 2001. Est-il un événement anecdotique sans conséquence, du type de ceux qui se produisent à chaque fois qu'un scientifique croit pouvoir extrapoler ses résultats au domaine moral avant que l'inanité de la démarche ne soit dénoncée et reconnue ? Ou, au contraire, cet article est-il la manifestation d'une bifurcation qui voit des pans de la philosophie morale quitter la philosophie (spéculative) pour rejoindre les sciences naturelles (expérimentales), et alors quels sont ces pans ?

En prolongement de l'utilisation de l'IRMf par Joshua Greene, imaginons qu'une nouvelle technologie, l'Imagerie Neuro-Cognitive[2], parvienne à multiplier par 10, 1000 ou 100 000 la précision et la définition tant spatiale que temporelle des cartes de l'activité du cerveau. On pourrait ainsi disposer d'une vision neurone par neurone, synapse par synapse, de la

2. J'emprunte cette appellation au laboratoire éponyme dirigé par Stanislas Dehaene `http://www.paris-neuroscience.fr/fr/equipe/laboratoire-de-neuro-imagerie-cognitive`

dynamique de la pensée, du cheminement qui mène de la perception de la situation à la décision prise et à l'action. Que saurait-on alors du comportement moral de l'être humain, ce qu'il est, ce qu'il peut et ce qu'il doit être, et quelles retombées aurait une telle connaissance sur la philosophie morale ? [3]

Avant de présenter la synthèse des travaux que j'ai menés pour tenter d'avancer des pistes de recherche face à ce questionnement, il est utile de préciser quelques éléments du contexte général dans lequel je situe ces pistes. Tout d'abord, si l'expérience est bien une source importante de connaissance sur le monde, et peut-être la principale, mon questionnement ne se donne pas pour contrainte d'affirmer que c'est la seule. Chacun peut avoir des positions différentes sur la possibilité des connaissances non empiriquement étayées, et cela aura peu d'importance pour mon travail dans la mesure où il sera admis que, au moins pour partie, il existe également des connaissances, au moins pour partie, empiriques.

Deuxième élément de contexte, les humains appartiennent à ce monde qu'il s'agit de connaître, ainsi que tous les phénomènes le concernant, dont le phénomène moral et les philosophes moraux. J'exclue donc la position de surplomb, ainsi que la transcendance systématisée, ce qui, encore une fois, est de peu d'influence, car il me suffit pour avancer que, a minima, certains comportements soient, au moins pour partie, considérés comme accessibles aux approches empiriques de ce monde observable depuis l'intérieur de ce monde. Mon étude s'inscrit alors dans la question plus générale des limites de la démarche expérimentale : quelle part des connaissances morales peut-on atteindre avec les démarches et méthodes des sciences naturelles ? Cette question est souvent appelée le problème de la « naturalisation de la morale ». Je n'adhère que modérément à ce vocable car il induit, logiquement, que la morale n'est pas, tant qu'on n'a pas entrepris cette étude, naturelle. Imaginerait-on un programme de naturalisation des arbres ? Et, de plus, cette expression semble donner le problème pour clairement posé : il serait possible de penser qu'on pourra répondre par « oui » ou « non » à ce problème. Mais il n'en est rien. Le naturalisme comporte deux thèses, la première est que tout est naturel, ce que j'accepte ici pour le domaine moral à titre d'hypothèse de travail, mais il me suffit, si on souhaite préserver la possibilité d'entités non naturelles, d'accepter qu'il existe une part de, pour partie, naturel dans le comportement humain pour que mon analyse soit valide. La seconde thèse du naturalisme est que les méthodes des sciences naturelles offrent le seul (ou le meilleur) outil pour acquérir de la connaissance, et cette deuxième thèse, que je mets en perspective, fait partie de ce que je vise à éclairer en regard de la connaissance du

3. L'ouvrage « Neuroscepticisme » de Denis Forest (Forest 2015) [89] construit un dialogue avec les neurosceptiques pour analyser sur la base de leurs réticences les différentes composantes de ce questionnement.

phénomène moral. Viser à répondre par « oui » ou « non » au problème de la naturalisation de la morale, c'est se condamner à ne pouvoir aborder le phénomène moral dans son ensemble.

7.2 Mes axes de travail

Le domaine moral

Mon premier axe de travail a été de décrire le domaine moral de façon à situer ces problèmes moraux pour lesquels je souhaite étudier l'apport potentiel des démarches expérimentales de la psychologie scientifique. Le phénomène moral, cette tendance des humains à s'engager dans des actions coûteuses sans bénéfice autre que moral, est omniprésent et important sur le plan des connaissances, pour la compréhension du comportement humain, et sur le plan pratique, pour le bon fonctionnement des sociétés humaines. Pourtant, il est difficile à définir, tant il semble dépendant de ce que chaque individu, chaque société, considère comme « moral ». J'ai retenu de prendre comme hypothèse de travail une définition, très large, du domaine moral : l'ensemble des situations qui donnent lieu, réellement ou potentiellement, à un énoncé évaluatif. J'ai proposé de schématiser ces situations ainsi :

> O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien.

Ce formalisme m'a permis d'unifier la description des différentes expériences de philosophie morale expérimentale de façon à en analyser et comparer les résultats. Il m'a également permis de proposer une cartographie de l'espace des désaccords entre philosophes moraux en cinq embranchements, j'y reviens ci-dessous, qui partent de ce point de départ commun : pour tous les penseurs moraux les énoncés évaluatifs existent et ont un rôle psychologique et social.

Avec ces cinq embranchements se dessine l'ampleur du champ des affrontements possibles entre philosophes moraux. Pour tenter d'évaluer les apports de la démarche scientifique expérimentale dans ces débats, il n'est pas efficace, ni peut-être même possible, d'en rester à une conception globale de la philosophie morale. J'ai proposé de distinguer quatre perspectives que peut adopter le philosophe moral et qui permettent d'avancer dans cette évaluation. Je reviens également plus loin sur ces quatre perspectives.

J'ai ensuite présenté le mouvement XPhi de philosophie expérimentale, ses résultats et les réticences qu'il a soulevées. Utilisant les démarches des psychologues expérimentaux, il constitue pour mon enquête un exemple grandeur nature de la tentative de transposer cette

démarche à des questions philosophiques. Le bilan en demi-teinte qu'on peut en tirer est riche d'enseignements.

La démarche scientifique expérimentale

La caractérisation du domaine moral par les énoncés évaluatifs, les cinq embranchements qui s'offrent aux penseurs moraux et les quatre perspectives qu'ils peuvent adopter, complétées par l'histoire récente du mouvement de philosophie expérimentale XPhi, construisent la toile de fond de mon enquête sur le domaine moral en vue d'évaluer les apports potentiels de la démarche scientifique expérimentale. Avant d'aller sur le terrain de ces expérimentations, j'ai voulu préciser ce que j'entends par cette expression « démarche scientifique expérimentale », en prolongement de ce que sont les pratiques des scientifiques aujourd'hui, et de façon à écarter les débats liés à des conceptions scientistes qui ne correspondent plus à ces pratiques, telles qu'elles sont détaillées dans les analyses des philosophes des sciences aujourd'hui.

J'ai proposé pour cela la métaphore de l'hélice expérimentale qui souligne le caractère dynamique, itératif d'une démarche scientifique expérimentale qui articule des théories, des objets, et des traces en évitant les écueils des excès de dogmatisme, de relativisme et de pragmatisme. Le mouvement de l'hélice suggère trois temps différents, un temps qui va des théories vers l'expérimentation, et que je propose d'appeler le temps de l'opérationnalisation, un temps qui va de l'expérience aux traces partagées par une communauté scientifique, que j'appelle objectivation, et le temps qui boucle le mouvement des traces vers les théories, le temps de l'interprétation inductive. Cette démarche itérative, cette dynamique scientifique, a permis le développement extraordinaire des sciences naturelles modernes et, à la fin du 20^{e} siècle, apparaît comme dominant, au moins rhétoriquement, le domaine de la psychologie. Cette démarche expérimentale n'a pas vocation à atteindre des Vérités absolues et majuscules, des certitudes [4], mais à construire pour chaque domaine exploré les meilleures théories possibles, appuyées sur les savoir-faire expérimentaux du moment et sur les traces des expériences passées.

Expérimenter sur l'expérimentation

Ayant ainsi clarifié successivement ce que je comprends par « domaine moral » et par « démarche scientifique expérimentale », j'ai choisi de traiter la question de leur articulation en procédant par une mise en abyme : expérimenter sur l'expérimentation. J'ai retenu cinq champs d'investigation, cinq études de cas, pour décrire les entrelacs complexes entre les activités scientifiques et les activités de philosophie morale, entre l'analyse descriptive des comportements et les jugements de valeur intimement associés.

4. Pour reprendre la proposition de John Dewey dans (Dewey 1929) citeDeweyquetecertitudeetude2014

Les cinq études de cas construisent un panorama limité mais concret de ce que sont les rapports entre la philosophie morale et la psychologie expérimentale aujourd'hui. Le premier cas est celui du tramway, il éclaire sur la façon dont est rendu compte des débats relatifs à la philosophie expérimentale. Le deuxième, avec l'enquête sur la surestimation du nombre de musulmans, traite du doute que soulèvent les études comportementales simplement appuyées sur des questionnaires sommairement administrés. Ce doute, déjà présent lorsqu'il s'agit d'évaluer les résultats de la psychologie expérimentale, est encore renforcé lorsque les sujets traités, philosophiques, font appel à des concepts abstraits. Le troisième cas, la transmission des émotions par les larmes, montre comment la démarche expérimentale se complexifie avec l'irruption de techniques comme l'IRMf qui ne peuvent être maîtrisées que par des équipes spécialisées à la suite d'investissements importants. Une conséquence structurelle importante de ce changement de niveau de technicité est la nécessaire confiance qui doit régner entre les équipes de chercheurs pour que puisse se développer la collaboration. Le quatrième cas, l'IAT (Implicit Association Test), donne à voir une controverse entre scientifiques appuyée sur des conflits de valeurs, et enfin le cinquième cas, relatif à l'effet Knobe et l'attribution d'intentionnalité, met l'accent sur l'instabilité des notions que l'expérimentateur importe de la philosophie au prix d'une adaptation mouvante selon les besoins de son étude.

L'opérationnalisation

Ces cinq études de cas ont, après analyse et pour mon enquête, un point commun, celui de révéler l'importance et l'intérêt de l'étape d'opérationnalisation : comment sait-on qu'on peut interpréter un résultat d'expérience comme pertinent en regard d'une entité psychologique ? J'ai détaillé cette analyse en montrant comment les critiques des interprétations inductives des expériences de psychologie s'appuyaient sur la possibilité d'interprétations alternatives conduisant à un risque de confusion. Les résultats de l'expérience ne sont considérés comme valides que si ce risque de confusion est évité, et ce risque n'est évité que si l'opérationnalisation est satisfaisante. La démarche qui semble la plus à même d'atteindre l'objectif, une opérationnalisation validée par les pairs, peut-être schématisée de la façon suivante :

— Une théorie postule l'existence d'entités, de propriétés et de relations.

— Pour chacune, et dans le cadre d'un champ expérimental donné, expérimentateurs et théoriciens se mettent d'accord sur des pratiques permettant de manipuler certains phénomènes.

— Et l'opérationnalisation répond à une triple équation : existentielle (les phénomènes et les éléments de la théorie apparaissent comme ayant la même extension), épistémique (raisonner dans le domaine théorique ou dans le domaine pratique n'apporte

pas de dissonances) et pragmatique (les techniques sont disponibles et les ressources nécessaires admissibles).

Réticences des philosophes moraux et réponses des scientifiques

Pour poursuivre mon enquête, il importait de prendre en compte le point de vue des philosophes moraux réticents à la démarche des XPhi. De façon générique, on peut évoquer le problème de Hume, le problème de l'impossibilité de déduire ce qui doit être de ce qui est, pour justifier de ces réticences. L'expérimentation n'a accès qu'à ce qui est et ne nous dit donc rien de ce qui doit être. De façon plus précise, ces réticences peuvent être regroupées selon deux axes. Le premier axe est sceptique, et s'appuie sur la complexité du comportement humain et sur la circularité à laquelle le chercheur est confronté quand il étudie le comportement de sa propre espèce. Le second axe est moral et peut se situer au niveau de la responsabilité individuelle, le chercheur se donne le droit moral exorbitant d'expérimenter sur autrui et la prise en compte des contraintes éthiques devrait conduire à l'arrêt de ces recherches, et au niveau de la responsabilité sociale, la morale est indispensable à la stabilité de nos sociétés. Analyser la morale comme on dissèque une grenouille revient à lui enlever toute autorité, donc à la détruire et non à l'étudier.

Face aux arguments sur la complexité et la circularité liées à l'étude du comportement humain par les humains, les scientifiques peuvent faire remarquer que, d'une part, d'autres sujets complexes sont abordés avec succès par les différentes sciences et que, d'autre part, il n'y aucune raison de poser une limite supérieure à la complexité des sujets qui pourraient être abordés avec la démarche scientifique expérimentale. Pas plus qu'il n'y a de raison, inversement, pour affirmer que toute question pourra être réglée de façon satisfaisante par cette démarche. C'est donc en avançant que se construisent ces territoires scientifiquement explorés.

Les problèmes posés par le caractère normatif du champ moral sont plus complexes à intégrer dans le champ scientifique. Les réponses à la question de la responsabilité morale du chercheur se donnant le droit d'expérimenter sur autrui sont appuyées sur des règles éthiques adoptées par les institutions de recherche, mais peuvent-elles vraiment être satisfaisantes aux yeux d'un philosophe moral déontologiste ? Et, quant au risque porté par la nocivité sociale de la recherche dans le domaine moral, il est le plus souvent simplement ignoré, tant il est difficile pour un scientifique de considérer que la recherche de connaissance puisse, en soi, être nuisible. L'exemple de la neuroéthique est particulièrement intéressant à ce sujet (voir 6.3.2, page 376). Il montre la distinction faite entre l'éthique des neurosciences et les neurosciences de l'éthique. Or, si la première correspond à ce que j'ai appelé ci-dessus la

responsabilité morale du chercheur, la seconde correspond à l'étude des processus neuronaux mis en œuvre dans un raisonnement moral, et rien n'est dit, ou si peu, quant à la possible nocivité sociale de cette recherche.

La possibilité d'une science a-normative, d'une science qui, ne s'engageant pas dans le domaine moral, n'aurait pas à traiter des sujets difficiles posés par les philosophes moraux, est souvent caressée par les scientifiques. Ils pourraient alors poursuivre leurs recherches sans se préoccuper de questions morales, à charge pour quelqu'un d'autre, la société en général ou les politiques, ou toute personne qui le souhaiterait, de transformer les connaissances acquises en actions ayant des conséquences pratiques, dont une incidence morale. Cette possibilité de neutralité morale de l'acquisition de connaissances n'existe probablement pas, de façon générale, pour toute science, et elle n'a même aucun sens pour la psychologie dont le sujet même, l'étude du comportement humain, est indissociablement lié au domaine moral.

Ce lien indissoluble entre humain et domaine moral a fait l'objet ces dernières années d'hypothèses intéressantes appuyées sur la théorie de l'évolution. L'espèce humaine étant à la fois sociale, rationnelle, projective et dialogique, pour reprendre la formulation de Francis Wolff, a, pour survivre, du se doter d'un mécanisme équilibrant l'altruisme nécessaire à l'intégrité du groupe social et l'égoïsme naturel à un être rationnel soucieux de son intérêt. La morale serait née de ce nécessaire équilibre. Le développement de cette hypothèse conduit à une structure à trois échelles de temps, le temps de la biologie et de la spéciation, le temps de la sociologie et de la création des groupes et de leurs règles morales, le temps de la psychologie et du processus d'individualisation des humains au sein d'une société. Plusieurs considérations s'ensuivent. Tout d'abord ces trois échelles de temps sont entremêlées, ce qui conduit à un domaine moral particulièrement complexe à décrire. Ensuite, à chaque articulation, il est opportun de distinguer ce qui est lié à l'une ou l'autre échelle de temps. Ainsi, le besoin de morale est découplé en, d'une part, la nécessité qu'existent des règles morales, nécessité qui serait d'ordre biologique pour que puisse exister l'espèce humaine telle qu'elle est, et d'autre part, le contenu de ces règles qui résulterait d'un processus historique, d'ordre sociologique, liant la création de ces règles à la création des groupes sociaux. De la même façon, il convient de découpler le caractère social des règles morales des différents groupes, qui établissent un cadre dans lequel l'individu a la nécessité de s'intégrer, et le caractère individuel du choix des groupes auxquels chacun adhère dans un processus d'individuation d'ordre psychologique. L'individuation conduit alors à intérioriser les règles morales des groupes auxquels on décide de s'intégrer. Enfin, dernière considération importante pour notre sujet, le domaine moral apparaît ainsi extraordinairement complexe et mobilisant potentiellement de multiples ap-

proches scientifiques dont la biologie, la sociologie, l'anthropologie, l'histoire, la psychologie. Selon cet éclairage, il est donc hautement improbable que l'apport de la démarche scientifique expérimentale à l'étude du domaine moral puisse relever d'une science unique, elle nécessitera certainement une vaste et complexe interdisciplinarité.

7.3 De ces travaux, premiers acquis méthodologiques

Des travaux précédents, je retiens plusieurs enseignements d'ordre méthodologique quant à la façon dont il est possible d'aborder le problème posé, celui de l'apport de la démarche expérimentale à la philosophie morale. Le premier est qu'il est probablement impossible de qualifier la philosophie morale en son ensemble en ce sens, et qu'il est donc indispensable de procéder à un découpage que j'envisage selon quatre perspectives. Des études complémentaires seraient certainement utiles à affiner ces perspectives, mais au travers de leur utilisation, je constate qu'elles sont en tout état de cause pertinentes en première approche. Le deuxième est qu'il est possible de décrire le domaine moral en partant des énoncés évaluatifs puis en proposant une cartographie des embranchements qui permettent de décrire les différentes positions des penseurs moraux. Il n'est pas à portée de ce travail de vérifier que toutes les positions possibles seraient ainsi prises en compte, mais les tests menés avec deux cas extrêmes (le réalisme moral de David Enoch et l'émotivisme constructif de Jesse Prinz) ainsi que les différentes utilisations faites au cours du présent travail valident, au moins partiellement, cet outil. Le troisième est l'importance qu'il me semble nécessaire d'apporter à l'analyse de l'étape d'opérationnalisation en regard des entités et propriétés proposées par les philosophes moraux, et reprises par les philosophes expérimentaux comme susceptibles d'être objet d'expériences. Je retiens six points pour qualifier cette importance de l'opérationnalisation.

Les quatre perspectives sur le domaine moral

Rappelons rapidement ces quatre perspectives. La première est descriptive, il s'agit de décrire les comportements humains, les énoncés évaluatifs auxquels ils donnent lieu, et tenter ainsi de mieux cerner ce phénomène moral complexe et omniprésent tel qu'il nous apparaît. La seconde perspective est prescriptive, il s'agit non plus de décrire ce que les humains font mais de dire ce qu'ils devraient faire. Pour cela, le philosophe adopte une théorie morale, ce que sont le Bien et le Mal et les moyens d'atteindre le premier et d'éviter le second, et entreprend de convaincre chacun que sa voie est la bonne, celle qui conduit au vrai Bien. La troisième perspective est celle de la méta-éthique. Il s'agit alors d'analyser l'ontologie

du domaine moral, la structure et la sémantique des énoncés moraux, de comparer les différentes théories morales entre-elles et d'analyser leurs justifications, leurs fondements, et leur inscription dans les sociétés humaines. Enfin la quatrième perspective est celle du philosophe engagé dans une activité particulière pour aider les personnes impliquées dans des cas particuliers. C'est la perspective de l'éthique dite appliquée. Je reviendrai plus loin sur des propositions appuyées sur ces quatre perspectives.

Cinq embranchements pour parcourir les positions des philosophes moraux

Les énoncés évaluatifs sont ceux qui comportent un jugement marqué par l'utilisation d'un adjectif comme Bien ou Mal. Aucun penseur moral ne conteste que, d'une part, ces énoncés évaluatifs existent, et d'autre part, qu'ils ont une certaine importance en regard des comportements humains. J'ai proposé de formaliser ces énoncés sous la forme :

> O observe que X dit à S dans un énoncé E que l'acte A fait par Y à Z dans le contexte C est Bien.

A partir de ce point de départ commun j'ai proposé de distinguer cinq embranchements qui permettent de distinguer les grandes familles de philosophies morales. Le premier oppose les tenants du fictionnalisme aux réalistes moraux. Pour les fictionnalistes, la morale ne serait qu'un type particulier de fiction, au même titre que la poésie ou le roman, dont la fonction première est que X et S se perçoivent comme partageant la culture commune portée par l'énoncé E et, en conséquence, tendent à reproduire l'acte A s'il est jugé Bien. Pour les réalistes moraux, il existe réellement des propriétés de Y, A, Z et C qui font que l'énoncé E peut être vrai ou faux indépendamment de ce que X et S en disent ou pensent. Le deuxième embranchement qu'affrontent les penseurs moraux concerne les multiples caractérisations possibles de ce qui relève ou non du domaine moral. Pour certains, restreignant à l'extrême le domaine moral, n'est réellement morale que la délibération envers ses propres actions et lorsque ces actions ont des conséquences sur d'autres personnes humaines. Ils diront alors « moraliste » celui qui juge des actions d'autrui, ils diront « maximaliste » celui qui inclut les actions envers soi-même dans le domaine moral, ils diront « non humaniste » celui qui juge sur le même plan moral les actions envers les choses inanimées, envers les animaux ou envers les humains. Le troisième embranchement concerne l'existence des règles morales. Pour certains penseurs, il existe des règles générales selon lesquelles un acte ou un acteur peuvent être moralement évalués. Ces règles peuvent être connues et les humains doivent donc se donner pour but de les connaître et de les appliquer pour se comporter Bien. Pour d'autres penseurs, même s'il existe des actes ou des acteurs moralement qualifiables, le juge-

ment moral est trop dépendant des situations, de l'ensemble « XESYAZC », trop complexe et circonstancié, pour que puissent être établies des règles générales. Il n'existe donc pas, pour eux, de règles morales. Pour d'autre penseurs, même si ces règles existaient elles ne pourraient être connues et il est donc illusoire de les rechercher. Le quatrième embranchement concerne la définition de ce que sont le Bien et le Mal. Certains identifient le Bien à l'utile, d'autres à ce qui est désiré, ou encore à ce qui est dans notre nature humaine. Là encore chaque penseur moral a pu choisir son chemin parmi un très grand nombre d'orientations possibles. Enfin, dernier embranchement, une fois choisis tous ces éléments qui font la substance des théories morales, il faut expliciter ce qui les explique, les justifie, les fonde. En un mot, pourquoi les sociétés ont-elles des règles morales ? Pourquoi telle règle plutôt que telle autre dans telle société, et surtout, pourquoi un individu libre se sent-il l'obligation de les suivre ? Ces cinq embranchements offrent, conjointement, un cadre possible aux désaccords entre philosophes moraux utile à situer les apports potentiels des résultats de la philosophie morale expérimentale dans les quatre perspectives proposées sur le domaine moral.

Six points sur l'opérationnalisation

J'ai tenté de résumer l'importance de l'étape d'opérationnalisation qui ressort des travaux réalisés en six points (voir 5.5.1, page 321). Premier point, il est pertinent de relier, dans le cadre de la psychologie expérimentale, le risque de confusion, les difficultés de l'interprétation inductive et la qualité insuffisante des opérationnalisations menées. Et, de ces trois aspects du même problème de fond qui consiste à chercher à mettre en correspondance un terrain expérimental et des théories, privilégier d'agir pour améliorer l'opérationnalisation semble être une stratégie féconde. Deuxième point, l'opérationnalisation est un processus itératif, progressif, qui évolue avec les théories et les savoir-faire pratiques, comme le suggère la métaphore de l'hélice expérimentale. Troisième point, la validation d'une opérationnalisation est un phénomène largement social, la validation se fait par les pairs, théoriciens et expérimentateurs, dans un contexte donné à un moment donné. De ces points résulte le quatrième, l'opérationnalisation est coûteuse en investissements et structurante pour les disciplines. Le cinquième point souligne la spécificité du domaine moral dont les enjeux idéologiques, politiques et religieux sont tels qu'il est difficile d'obtenir cet accord des pairs, et ainsi, l'accord sur une opérationnalisation, ce qui se concrétise dans les multiples débats sur l'interprétation des expériences de psychologie morale. Et, enfin, sixième point, l'importance de l'opérationnalisation est à pondérer selon la perspective prise sur le domaine moral. Elle est centrale dans la perspective descriptive et dans la perspective méta-éthique, moins dans les perspectives prescriptives et appliquées.

7.4 Trois propositions

Au-delà des enseignements méthodologiques présentés à la section précédente, les travaux menés m'ont conduit à trois propositions plus substantielles. La première, en continuité des six points donnant sa juste importance à l'étape d'opérationnalisation, est que l'opérationnalisation des entités de la psychologie est un point nodal de la problématique de la psychologie expérimentale et que la question posée ici de l'apport expérimental à la philosophie morale peut, au moins partiellement, être reformulée en questionnant la pertinence qu'il y a à importer des concepts expérimentalement instables de la philosophie morale dans la démarche expérimentale scientifique. La deuxième proposition est que, prenant au sérieux la théorie de l'évolution et les pistes qu'elle suggère pour les origines du phénomène moral, il est fécond de distinguer les trois échelles temporelles de la spéciation, de la création des groupes sociaux et de l'individuation de façon à traiter du phénomène moral au sein des démarches expérimentales de chacune des sciences concernées, et elles sont nombreuses. La troisième proposition, en prolongement du constat du très large domaine de la philosophie morale est que rien d'utile ne peut être dit de la philosophie morale en général en regard de la démarche expérimentale scientifique, et qu'en revanche, des analyses utiles peuvent être proposées en s'appuyant sur les quatre perspectives descriptives, prescriptives, méta-éthique et des éthiques appliquées. Détaillons ces trois propositions.

7.4.1 L'opérationnalisation est un axe de travail pertinent

La première proposition que je souhaite mettre en avant à l'issue des cinq études de cas et de leur analyse, est la pertinence qu'il y a à considérer, en regard de la problématique particulière de la philosophie morale, l'opérationnalisation comme une étape importante de la démarche scientifique expérimentale dont l'auscultation est riche d'enseignements. Plusieurs éléments viennent appuyer cette proposition. Tout d'abord, les critiques des expériences de philosophie expérimentale sont généralement centrées sur l'interprétation inductive de ces expériences qui serait abusive ou non pertinente car des facteurs de confusion ont été négligés. J'ai souligné que l'interprétation inductive et l'opérationnalisation sont, en cela, les deux faces d'une même pièce. Ma proposition est qu'il est plus éclairant de travailler sur l'opérationnalisation car cela contraint à devoir expliciter les théories, avec les entités, les propriétés et les relations qu'elles postulent et, pour chacun de ces éléments, expliciter en quoi expérimentateurs et théoriciens sont en position d'être confiants quant à la mise en œuvre pratique des dispositifs expérimentaux en tant que fournissant de bons indices des instances de ces

entités, propriétés et relations, et, en somme, d'accorder leur confiance aux résultats des expériences.

L'étude de l'opérationnalisation permet d'éclairer les dispositions prises pour lutter contre les confusions, pour vérifier que le chercheur fait bien ce qu'il pense faire au cours d'une observation ou d'une expérience. L'opérationnalisation réussie va plus loin, elle conduit à construire les savoir-faire pratiques qui permettent à l'expérimentateur d'acquérir peu à peu la capacité à manipuler, c'est-à-dire à créer, modifier, supprimer, détecter, mesurer, ... les phénomènes qui sont reconnus par lui et par la communauté des chercheurs, comme reflétant des instances des entités, propriétés et relations stipulées par les théories. Et c'est par l'acquisition de ce savoir-faire que, théoriciens et expérimentateurs acquièrent également la conviction qu'il est pertinent d'inscrire ces entités, propriétés et relations, dans leur ontologie. « L'émotion » existera empiriquement si et seulement si les expérimentateurs savent la détecter, la mesurer, la créer, la supprimer, la modifier,...

Opérationnalisation, instabilité des notions abstraites et pari des XPhi

J'ai souligné que la philosophie expérimentale, le mouvement XPhi, utilisait les méthodes et outils de la psychologie expérimentale, en ce sens, elle est dépendante de la bonne opérationnalisation des entités qu'elle manipule au même titre que la psychologie. Néanmoins deux considérations conduisent à considérer que la situation actuelle est moins défavorable à la psychologie qu'elle ne l'est à la philosophie expérimentale, sous cet angle de l'opérationnalisation. La première considération est liée à l'histoire récente de la psychologie et à l'importance qu'a eu le behaviorisme au siècle passé. Ce mouvement, et la recherche de scientificité associée, a conduit les psychologues à développer tout un ensemble d'outils mesurant les stimuli et les comportements, et ces outils constituent une base pour construire peu à peu des opérationnalisations d'entités plus complexes qui peuvent être abordées progressivement. Le philosophe n'a pas cette base lorsqu'il prend un sujet par le haut, comme je l'ai montré avec l'exemple de l'effet Knobe et le concept d'« intentionnalité » qui mobilise en réseau d'autres concepts comme la responsabilité, la culpabilité, le libre arbitre, sans que l'expérimentateur ne puisse construire les savoir-faire qui permettraient de distinguer toutes les nuances liées à ces concepts dans son dispositif expérimental. Importer ainsi dans une démarche expérimentale des concepts abstraits philosophiques qu'on ne peut maîtriser expérimentalement, c'est faire le pari que l'expérimentation elle-même permettra, progressivement, de mieux les définir.[5] Ce pari, au vu des résultats de nos études des cas, n'est pas en passe d'être gagné

5. Joshua Knobe arrive à cette même conclusion en regard de la notion d'attribution de responsabilité dans (Doris 2012 p 346) [76] mais ne conclut pas par une question mais par l'affirmation que tel est le programme de recherche à mener.

rapidement.

Deuxième considération, les questions traitées par les psychologues peuvent être perçues, pour certaines d'entre elles, comme relativement techniques et loin du terrain politique, alors que les débats philosophiques présentent fréquemment une sensibilité plus forte, qu'elle soit religieuse ou idéologique, avec des conséquences politiques et sociales fortes, comme l'exemple de l'IAT et de la lutte contre le racisme l'a montré. Du fait de ces deux considérations, difficultés à expérimenter sur des concepts abstraits philosophiques et sensibilité des sujets, l'interprétation inductive est toujours soumise à forte critique en philosophie morale et, sans garantie de résultat bien sûr, on peut espérer que mettre l'accent sur l'opérationnalisation permettrait de progresser vers plus de pertinence des résultats expérimentaux. On peut s'inspirer en cela de l'exemple de l'IAT déjà cité qui a détaillé l'atteinte d'un constat partagé, certes partiel, obtenu grâce à la réplication des expériences et aux méta-analyses. Non que les résultats expérimentaux deviennent alors concluants sur le plan philosophique, mais l'opérationnalisation réussie, et donc acceptée contradictoirement, permet de mieux percevoir ce qui relève d'un consensus possible expérimentalement instruit et ce qui relève des conflits de valeurs pour lesquels il est vain d'attendre un tel consensus.

Opérationnalisation et organisation de la recherche

L'importance de l'opérationnalisation se joue également en regard de l'organisation de la recherche expérimentale. Comme nous l'avons-vu dans la présentation du mouvement XPhi, les promoteurs de ce mouvement défendent la pertinence qu'il y aurait pour les philosophes à descendre eux-mêmes dans l'arène expérimentale, à quitter leur fauteuil pour rejoindre le laboratoire. Je voudrais examiner maintenant cette suggestion à la lumière des caractéristiques de l'opérationnalisation que je viens de présenter et procéder pour cela à quelques observations préalables.

Première observation préalable, l'opérationnalisation représente un lourd investissement qui est justifié par le réemploi de la même opérationnalisation (i.e. de la même relation entité théorique – pratique expérimentale) dans plusieurs programmes de recherche. Il faut comprendre ici le terme de justifié au double sens d'épistémiquement justifié, car c'est par cette triangulation que l'opérationnalisation peut être validée, et économiquement justifié, car le coût de l'investissement est amorti sur un domaine plus large. L'opérationnalisation vise deux objectifs, définir des protocoles expérimentaux et les faire reconnaître comme pertinents par la communauté de chercheurs, et ces deux objectifs sont à atteindre sous la contrainte des ressources disponibles. De telles entités théoriques, si elles sont intéressantes, apparaissent certainement dans de nombreuses propositions théoriques et, donc, dans de nombreuses études

empiriques. Il semble alors peu performant d'envisager l'opérationnalisation de ces entités comme une question propre à chaque proposition étudiée. Globalement, un domaine d'étude gagnera à structurer un catalogue de protocoles opérationnalisant les principales entités de ses théories.

On peut pour illustrer ce point prendre l'exemple du physicien qui ne réinvente pas le thermomètre à chaque expérience de thermodynamique qu'il envisage, sauf s'il y est contraint. Au contraire, il s'attachera à se rapprocher des protocoles existants pour le plus grand nombres possible des entités théoriques en relation avec son expérience. Ce partage des protocoles expérimentaux permet du point de vue économique d'amortir les recherches sur la meilleure façon d'opérationnaliser une entité théorique (encore une fois, observer, détecter, mesurer, modifier, créer, supprimer un phénomène en lien avec une entité ou une propriété ou une relation théoriques) et du point de vue épistémique d'augmenter la confiance sur le fait que les différentes expériences d'un côté et les différentes propositions théoriques de l'autre côté emploient bien les vocabulaires de la même façon que dans le cadre des expériences déjà réalisées et déjà interprétées.

Deuxième observation préalable : puisqu'une opération réussie, comme c'est le cas pour la température par exemple, représente un lourd investissement à la fois théorique et expérimental, elle conduit inévitablement à une certaine rigidification des pratiques tant théoriques qu'expérimentales. Il faut des raisons sérieuses pour remettre en cause une opérationnalisation coûteuse qui a fait ses preuves. Elle contribue ainsi à constituer des corps de doctrine distincts, d'un côté comment on mesure une température en pratique et de l'autre côté, comment on inclut la température mesurée dans différentes théories, mais articulés par une commune acceptation de l'opérationnalisation de la température comme ce qui est mesuré par le thermomètre.

De ces deux observations préalables découle une remarque : dans les domaines où de nombreuses entités ou propriétés ou relations théoriques importantes ont déjà été opérationnalisées, on peut s'attendre à ce que l'organisation de la recherche expérimentale sépare nettement les équipes à ce point d'articulation qu'est l'opérationnalisation, les théoriciens d'un côté, les expérimentateurs de l'autre. Inversement, lorsque le niveau de l'opérationnalisation est faible, ou qu'elle est remise en cause à chaque expérience, on peut s'attendre à ce que l'organisation de la recherche expérimentale soit indistinctement incluse dans la recherche théorique. Si cette dernière remarque générale est valide, alors elle va entraîner d'autres remarques plus spécifiques à propos de la psychologie expérimentale et, par extension, de la philosophie expérimentale et de toutes les disciplines qui ont, au moins pour partie, un intérêt

à l'étude du domaine moral. J'ouvrirai plus loin cette réflexion sur les incidences que pourrait avoir cet accent mis sur l'opérationnalisation des entités psychologiques en regard du réseau complexe des disciplines académiques, mais avant, il convient de rappeler, en conclusion, la base empirique de ce constat à la lumière des études de cas que j'ai menées.

La réutilisation systématisée de protocoles n'est pas ce que les cas décrits dans la présente thèse ont donné à voir dans le domaine de la philosophie morale expérimentale. Au contraire, dans le cas de l'effet Knobe, l'examen des 33 articles récents a conduit au constat que l'opérationnalisation des notions comme « intentionnalité » « causalité » ou « responsabilité » change à chaque expérience et, conséquemment, est remise en question lors de l'interprétation sans que se dessine un processus de convergence qui rendrait possible l'apparition d'un protocole commun reconnu comme pertinent pour tout un programme de recherche. Le principe que ces 33 articles donnent à voir est inverse : chaque article s'attache par une nouvelle expérience à redéfinir ces notions de base, rendant ainsi caduques les opérationnalisations précédentes et, avec elles les traces laissées par les expériences correspondantes. Le jugement moral étudié avec les dilemmes sacrificiels fait apparaître la même instabilité conceptuelle de ce qui pourrait expliquer les différentes dissymétries rencontrées d'une variante à l'autre de l'expérience du tramway. On peut donc légitimement s'interroger sur la pertinence qu'il y a à importer les concepts de la philosophie, expérimentalement instables, au sein de la psychologie expérimentale si n'est pas menée au préalable une étape d'opérationnalisation à la hauteur du défi que cet import constitue.

7.4.2 Considérer trois échelles de temps est fécond

La deuxième proposition est issue du constat, descriptif, du rôle de la morale dans la structure des groupes sociaux au sein des sociétés humaines : la morale a pour triple effet de faciliter la coordination au sein des groupes, de permettre d'identifier les personnes du même groupe que soi, et de permettre d'identifier les personnes étrangères au groupe. La morale contribue ainsi à identifier, individualiser, construire les groupes sociaux et les individus qui leurs appartiennent et les constituent. En ce sens, le phénomène moral mobilise trois niveaux d'organisation en interaction et interdépendance. Il s'agit de l'espèce humaine (sociale, rationnelle, projective et dialogique, pour reprendre la terminologie de Francis Wolff), des groupes sociaux, cimentés par leurs règles morales historiques, et enfin des individus qui, pour partie, construisent leurs identités complexes en adoptant les règles morales de ces groupes

Or chacun de ces trois niveaux correspond à des sciences différentes, à des communautés de chercheurs différentes, à des échelles de temps différentes, à des méthodes expérimentales différentes et, également, à des rapports au politique et aux idéologies, dont morales, différents. Le rapport de la morale à la démarche scientifique expérimentale devrait donc (à l'idéal) faire l'objet d'une analyse détaillée croisée selon les quatre perspectives (descriptive, prescriptive, méta-éthique et éthiques appliquées) et les trois échelles d'organisation proposées. Ma deuxième proposition est qu'accepter la complexité de cette analyse est fécond pour mieux comprendre l'apport potentiel de la démarche expérimentale à la philosophie morale.

La théorie de l'évolution joue un rôle particulier dans l'établissement de cette deuxième proposition. Cet outil intellectuel clé qui permet de comprendre « ce qui est » en tant que « ce qui a pu advenir et perdurer» construit un petit pas, partiel, en direction de « ce qui doit être ». Si les humains existent aujourd'hui au sein de sociétés marquées par le phénomène moral, c'est qu'il est possible qu'une espèce humaine (sociale, rationnelle, projective et dialogique) apparaisse et perdure dans son environnement, avec son éthologie comportant des structures sociales cimentées par des règles morales, et regroupant des individus éduqués à ces règles au sein de ces structures. Le constat du phénomène moral, le constat de ce qui est, implique logiquement que l'espèce humaine ainsi instanciée est possible, elle peut apparaître et survivre. Cela n'implique pas que ce soit « ce qui doit être » et, ceci, à plusieurs niveaux. Tout d'abord, il pourrait simplement ne pas exister d'espèce sociale, rationnelle et dialogique. L'espèce humaine aurait pu ne pas apparaître, et d'ailleurs cela a été le cas pendant la plus grande partie de l'existence de notre planète. Ensuite, et c'est une autre question, à supposer qu'une espèce sociale, rationnelle et dialogique apparaisse serait-il possible qu'elle ne soit pas équipée du phénomène moral ? Nous ne pourrions répondre positivement à cette question, du point de vue empirique, que si existait une autre espèce sociale, rationnelle et dialogique dont les groupes sociaux n'auraient pas le phénomène moral pour ciment. Une réponse empirique négative est bien sûr logiquement impossible puisque le constat d'inexistence marque autant notre ignorance qu'un état de fait. Un constat d'inexistence serait un indice, mais pas une preuve, qu'il pourrait être impossible qu'advienne une espèce proche des humains dont les groupes seraient régulés sans qu'existe une morale en leur sein. Une autre formulation de l'utilité de la morale est de dire qu'elle constitue un avantage évolutif. On entend généralement par là que si deux espèces concurrentes, toutes les deux sociales rationnelles et dialogiques, existent sur un même territoire, et que l'une a développé des règles morales et pas l'autre, alors la première a un avantage qui conduira à sa survie alors que l'autre sera

dominée[6]. Et, enfin, troisième niveau de remarque sur la vue évolutionniste de la morale, cette analyse est largement indépendante du contenu précis des règles morales. L'exemple des interdits alimentaires, très fréquents mais tous différents, le montre à l'évidence.

La complexité liée aux trois niveaux d'organisation nécessaires au déploiement de la morale, l'espèce, le groupe social et l'individu, me conduit à considérer comme important de bien distinguer d'un côté la norme (par exemple : tout groupe se définit et se constitue pour partie par des interdits normatifs) et de l'autre côté le contenu de la norme (par exemple : l'interdit alimentaire du porc dans tel groupe). En effet l'existence de la norme relève, du point de vue que je viens de détailler en relation avec la théorie de l'évolution, d'un certain niveau d'organisation (dans l'exemple, l'évolution des espèces) alors que le contenu de la norme relève d'un autre niveau (dans l'exemple, la sociologie et l'histoire de la constitution des groupes). Et cette distinction, entre le caractère constituant de la norme et le caractère historique et contingent du contenu de la norme, est à prendre en compte de façon détaillée à chacune des étapes du raisonnement qui va de l'environnement à l'espèce, de l'espèce au groupe social, puis du groupe social à l'individu. La construction d'un plan d'expérimentation qui ne prend pas en compte ces distinctions entre des niveaux biologiques, sociologiques et psychologiques, avec leur incidence sur la distinction entre nécessité de la norme et contingence du contenu de la norme s'expose à deux types de difficultés. La première porte sur la difficulté à obtenir l'accord des différentes communautés scientifiques sur les opérationnalisations menées. La portée des conclusions sera alors limitée en regard de l'ampleur du phénomène moral. La seconde porte sur les risques de confusion induits, prendre pour général ce qui n'est que contingent ou, inversement, prendre pour relevant du cas particulier ce qui n'est que l'instanciation locale d'une règle générale.

7.4.3 Les quatre perspectives de la philosophie morale en regard de la démarche expérimentale

Troisième proposition, en prolongement de l'acquis méthodologique des quatre perspectives, rien d'utile sur l'apport potentiel de la démarche scientifique expérimentale à la philosophie morale ne peut être dit qui soit intéressant pour la « philosophie morale » en général. La philosophie morale recouvre un périmètre très large qui, pour les sciences de la nature, serait l'équivalent de l'ensemble constitué par la métaphysique, la physique (et les autres sciences naturelles), la philosophie des sciences et l'ensemble des sciences et techniques ap-

6. Cet argument peut être mobilisé pour tenter d'expliquer pourquoi Homo Sapiens a supplanté ses cousins hominines.

pliquées. Ce périmètre a été pour partie remis en cause au début du vingtième siècle avec l'autonomisation de la psychologie puis, plus récemment, avec la psychologie morale ainsi qu'avec la philosophie expérimentale du mouvement XPhi. Il n'en reste pas moins que la philosophie morale, telle qu'elle est pratiquée aujourd'hui, comporte à la fois des préoccupations proches de la mise en œuvre, l'éthique appliquée, des préoccupations proprement morales portant sur le choix et la promotion de la « meilleure » théorie morale et du Bien qu'elle définit, et enfin des préoccupations philosophiques, méta-éthiques, sur ce que peut être une théorie morale et, en particulier, selon quel critère elle pourrait être « meilleure ».

Pour analyser le rapport de la philosophie morale à l'expérimentation en prenant en compte cette contrainte de sa trop grande diversité, j'ai proposé de distinguer quatre perspectives que peut adopter le penseur moral sur le domaine moral. Cette distinction s'est avérée satisfaisante au travers des cinq études des cas et des analyses menées. Ces quatre perspectives sont la perspective descriptive, ce que les humains font, la perspective prescriptive, ce qu'ils devraient faire, en accord avec une théorie morale particulière, la perspective méta-éthique, ce que sont les théories morales, ce qui les fonde, les justifie et pourquoi les humains les suivent, et enfin la perspective appliquée, quand le philosophe cherche à aider les personnes concernées par une situation particulière.

En poursuivant dans ce cadre l'exploitation des cinq études de cas, j'en arrive à plusieurs constats, résultats partiels qui seront à poursuivre et préciser. Tout d'abord, et relativement aux question au départ de mon enquête, ce que l'on peut dire de l'apport de l'expérimentation dans chacune des perspectives évoquées. Puis, observations plus spéculatives, que peut-on extrapoler de l'expérience des XPhi en regard des développements de la philosophie morale et de ses relations à l'ensemble des sciences concernées.

Dans la perspective descriptive

L'approche expérimentale permet d'affiner la description des phénomènes moraux, mais ce que montrent les différentes études de cas, c'est que les ontologies du domaine moral sont trop faiblement structurées pour éviter l'instabilité notionnelle : à chaque nouvelle notion philosophique abordée, à chaque nouvelle expérimentation entreprise, c'est l'ensemble du fragile réseau des entités théoriques postulées qui est remis en cause. Face à cette difficulté, le behaviorisme proposait de faire abstraction des états mentaux et de s'en tenir au schéma Stimuli – Organisme – Réponse, mais cette stratégie n'a pas permis une description crédible du phénomène moral. Comment en finir avec les phlogistiques moraux et trouver la voie moyenne entre la cécité volontaire du behaviorisme, l'opérationnalisme de Bridgman et le dogmatisme moral, pour inventer enfin l'oxygène moral ? Telle pourrait être une reformula-

tion de la question posée par l'utilisation de la démarche scientifique dans la perspective de décrire le phénomène moral.

Adoptons un instant pour répondre à cette question le point de vue du programme négatif des XPhi. Si opérationnaliser les entités, propriétés et relations des théories philosophiques s'avère impossible, c'est peut-être le signe que ces théories ne sont pas valides dans la perspective descriptive. Il convient alors, avec humilité, de reprendre le travail à la base et de tenter de découvrir, empiriquement, quelles entités seraient plus pertinentes pour décrire le comportement humain. Si les théories philosophiques sont descriptivement inaptes, c'est aussi, au moins pour partie, parce que les méthodes utilisées par les philosophes pour les établir ne sont pas fiables, et qu'il faut donc les abandonner au profit de méthodes fiables, c'est-à-dire celles des sciences naturelles. Malheureusement, le bilan en demi-teinte du mouvement XPhi tend à montrer que le défi qui consiste à retrouver des propositions déterminantes sur le plan philosophique à partir de résultats empirique n'est pas à portée. En revanche, des propositions intéressantes ont déjà été apportées, et on peut espérer que d'autres le seront encore, et seront également plus facilement acceptées si, et ce sont là mes propositions, leur appartenance à la seule perspective descriptive est bien clarifiée et si, également, les efforts suffisants sur la qualité de l'opérationnalisation et sur la qualification de l'échelle de temps concernée ont été accomplis.

Dans la perspective prescriptive

Dans le jeu sans fin des arguments, principalement rhétoriques, entre défenseurs des différentes théories morales, les considérations épistémiques ne sont pas au premier plan. Même si l'heure n'est plus, ou du moins plus partout, aux guerres de religions, il n'en reste pas moins que la démarche scientifique expérimentale est peu pertinente pour qui sait déjà ce qu'il doit croire. Monter une expérimentation sur ce qui doit être cru est alors au mieux inutile, puisque la vérité est déjà atteinte. A supposer que les résultats de l'expérience ne soient pas conformes à cette vérité, cela montrera simplement que le Bien n'est pas encore atteint dans notre vallée de larmes. Et, au pire, expérimenter est dangereux, car cela peut conduire à décrédibiliser une morale qui n'est efficace que lorsqu'elle est crue, et donc respectée. On peut conclure qu'il est peu probable que le prosélyte moral s'emparera des résultats expérimentaux avec une ferveur égale quand ils sont favorables à ses thèses et quand ils y sont défavorables. L'apport de la démarche expérimentale scientifique ne peut donc qu'être faible selon cette perspective prescriptive.

Néanmoins, tous les philosophes moraux adeptes d'une théorie morale particulière vivent au quotidien en se conformant à des règles pratiques issues, pour partie, des connaissances

acquises grâces aux sciences expérimentales. Ils ont donc à conjuguer au quotidien les règles morales de leur théorie avec ces règles pratiques qui ont un autre fondement [7]. Et, d'autre part, ils ont à conjuguer, également au quotidien, dans leurs interactions au sein de l'espèce humaine globalisée, les règles morales qu'ils ont choisies avec celles de tous leurs partenaires qui en ont choisi d'autres. Ces deux contraintes obligent à une évolution forte des théories morales, dans la perspective prescriptive, et l'apport de la démarche expérimentale à ces évolutions pourrait être favorisé, c'est là ma proposition, si les trois échelles de temps, de l'espèce, du groupe social et de l'individu, sont mises en avant avec l'ensemble des sciences concernées depuis la biologie, la sociologie, l'anthropologie, etc. jusqu'à la psychologie. L'apport des XPhi sera alors utile en tant qu'incitateur de telles recherches mais restera modeste, dans cette perspective prescriptive, quant aux résultats expérimentaux qui pourront être directement apportés et acceptés car la qualité de l'opérationnalisation ne saurait être satisfaisante pour des prosélytes souhaitant refuser ces résultats.

Dans la perspective méta-éthique

Comparer les théories morales entre-elles est un des axes de travail de la méta-éthique. Elle dispose pour cela des outils de la philosophie mais elle ne peut, sous peine d'être accusée de perdre toute neutralité dans ces comparaisons, adopter les concepts de l'une des théories morales en lice. Les sciences, et en particulier les sciences expérimentales, lui offrent un cadre complémentaire qui partage ce souci de neutralité en regard des jugements moraux. Dans cette perspective méta-éthique, on peut donc attendre un apport important des expérimentations, bien sûr pour la description du phénomène moral, indépendamment de toute théorie morale, mais également, en prolongement des propositions faites plus haut, pour une clarification conceptuelle en analysant les entités (propriétés et relations) postulées par chacune des sciences à chacune des trois échelles temporelles de l'espèce, du groupe et de l'individu, en regard des entités postulées par chaque théorie morale.

Ce travail conceptuel, immense, permettrait d'envisager la description des caractéristiques des liens au sein de l'espèce humaine, en relation avec l'échelle de temps de la spéciation, et, ainsi, les faisceaux de contraintes et les degrés de liberté qui président à la création des groupes sociaux, puis qui donnent un cadre au processus d'individuation des humains. La compréhension de l'inscription des théories morales existantes dans ces contraintes et avec ces degrés de liberté sera le deuxième travail à entreprendre. Cette compréhension ne

7. J'écris ces lignes en mars 2020 alors que la moitié de la population mondiale est confinée pour contrer l'épidémie de coronavirus. Le gouvernement théocratique d'Iran vient de décider de fermer les sites religieux du chiisme. On ne peut mieux montrer que les règles religieuses et pratiques doivent être conjuguées, même par un pouvoir religieux et autoritaire.

résultera pas du seul apport des démarches expérimentales scientifiques car il s'agira ici de décrire des situations particulières, contingentes et historiques, dont les régularités, mais non les spécificités, sont accessibles à cette démarche.

Dans la perspective de l'éthique appliquée

L'éthique appliquée est, par définition, confrontée ici et maintenant à l'obligation d'agir malgré la complexité du comportement humain qui interdit l'analyse complète des conséquences de l'action. Elle a donc renoncé à l'application de théories morales dont les insuffisances descriptives empêchent d'affirmer si et quand elles s'appliquent à des situations réelles. Le sujet central de l'éthique appliquée est de trouver les meilleures heuristiques possibles permettant d'arriver à résoudre les cas concrets, c'est-à-dire à trouver une solution acceptable par les parties prenantes dans un temps compatible avec le déroulement des événements.

L'apport potentiel des démarches expérimentales est, sur le fond des questions morales abordées, très réduit pour les éthiques appliquées dans la mesure où la complexité des situations réelles ne permet pas plus de savoir si les résultats d'une expérimentation sont extrapolables à la situation réelle en cause qu'elle ne permettait de savoir si une règle morale s'appliquait. En revanche, les méthodes expérimentales peuvent être utiles à rechercher quelle est la meilleure méthode, la meilleure heuristique, le qualificatif de meilleur étant accordé à l'heuristique qui conduit assez souvent à un assez bon compromis entre parties prenantes. Tenter de savoir si le compromis est, ou non, moral est, dans cette perspective des éthiques appliquées, une voie sans issue car pour répondre à cette question, il faudrait pouvoir caractériser la situation réelle en regard des règles morales qui s'appliquent, et c'est précisément ce rapprochement entre règles théoriques et situations réelles qui est impossible.

7.5 Perspectives sur l'organisation de la recherche

7.5.1 Le complexe réseau des disciplines académiques

Mon travail n'avait pas pour sujet l'organisation des disciplines académiques, ni la sociologie de leurs acteurs, que ce soit d'un point de vue descriptif ou dans une visée réformatrice. Néanmoins, à titre de perspectives pour des travaux futurs, je pense utile de proposer dans cette section quelques observations sur cette organisation, en lien avec les trois propositions ci-dessus.

Tout d'abord, mon enquête m'a conduit à constater que le réseau des disciplines qui ont,

au moins pour partie, la morale pour sujet est vaste et complexe. De multiples découpages s'entrecroisent. Tout d'abord, ceux de la philosophie morale elle-même avec les trois grandes familles de la philosophie morale substantielle, de la méta-éthique et des éthiques appliquées. Ces familles ne recoupent pas la grande distinction entre philosophes analytiques et continentaux, quelle que soit la signification exacte qu'on puisse donner à ces termes. Dans chaque tradition philosophique, la philosophie morale trouve sa place et elle est différente dans chacune. Mais la philosophie traditionnelle n'est plus, comme mon enquête le montre, la seule approche philosophique de la morale qui existe. Les philosophes expérimentaux l'ont également mise à leur ordre du jour. Et, proches des philosophes moraux expérimentaux, des psychologues se sont dits « psychologues moraux », le premier terme indiquant leur appartenance aux sciences sociales et le second leur centre d'intérêt. Distinguer psychologues moraux et philosophes expérimentaux moraux est souvent difficile tant méthodes et thèmes abordés se recoupent. Les travaux des sociologues sur les normes et les travaux des biologistes de l'évolution sur l'altruisme ont également des dimensions dont il est facile de voir les incidences croisées avec certains travaux de philosophes moraux. Enfin, pour finir ce court inventaire, citons le mouvement récent qui a pris pour dénomination neuroéthique et qui vise à chercher des bases neuronales aux comportements éthiques des humains.

La philosophie a plus pour objet de poser de (philosophiquement) bonnes questions et de les agencer que d'apporter des réponses (et heureusement, car elle en apporte peu). Les sciences expérimentales ont pour objet d'apporter de (scientifiquement) bonnes réponses à des questions limitées héritées ou non des questions (illimitées) des philosophes, et, au passage, donnent aux philosophes des éléments pour déplacer leurs questions. La philosophie morale s'inscrit de façon complexe dans ce schéma car pour partie, elle se donne pour but d'apporter des réponses normatives (exemple : le minimaliste Ruwen Ogien tente de nous convaincre que seules des actions vers des tiers, et non envers soi-même, peuvent être qualifiées de « morales »), pour partie elle tente de construire des critères définissant ce que serait le Bien à rechercher ainsi que des méthodes pour l'atteindre (exemple : l'utilitarisme) et enfin, pour sa partie appliquée, elle se rapproche plus d'une technique pour résoudre ici et maintenant les dilemmes moraux concrètement posés, utilisant pour cela des heuristiques, pour reprendre le mot de Marta Spranzi. Cette présentation du complexe réseau des disciplines académiques est, au moins en première instance, agnostique du point de vue métaphysique, la circularité assumée dans la formulation ci dessus par la définition de « bon » comme (philosophiquement) ou (scientifiquement) bon laisse ouverte toute possibilité de justification de ces définitions en tant que simplement instrumentales ou en tant que comportant un engagement ontologique

plus ou moins fort.

7.5.2 Deux organisations ?

Il s'agit maintenant de comparer, dans la perspective de l'opérationnalisation, deux types d'organisations envisageables pour la recherche expérimentale en philosophie. La première organisation, similaire à celle existant actuellement en philosophie des sciences naturelles, différencie les rôles, la seconde, réformatrice, suivant la suggestion des XPhi, fait descendre les philosophes dans les laboratoires de psychologie.

En prolongement de l'organisation telle qu'on l'observe par exemple dans les sciences physique, la première organisation s'appuie sur une division forte du travail, certes différente pour chaque discipline, mais que l'on peut caricaturer pour les besoins de cette analyse en les répartissant sur plusieurs pôles schématisés en partant du terrain d'application. Tout d'abord les techniciens de laboratoire chargés de mener les expériences et, donc, de maîtriser toutes les techniques nécessaires et d'en développer d'autres. Ensuite, ce qu'on peut appeler les scientifiques plus tournés vers la mise au point des expérimentations (dont l'opérationnalisation, au sens de la section précédente), puis les scientifiques théoriciens, puis, travaillant en collaboration étroite avec les scientifiques, les philosophes spécialistes d'une science particulière, les philosophe des sciences, et enfin les philosophes traitant des thèmes généraux transversaux à tous les domaines de la philosophie, objets ou non d'enquêtes de type scientifique.

En déclinant cette répartition dans le domaine de la philosophie morale, on aurait, en supposant, à titre de simplification de l'exposé, que la psychologie porte l'essentiel de l'approche scientifique des phénomènes moraux : des psychologues de terrain, des psychologues expérimentaux, des psychologues théoriciens, des philosophes de la psychologie, des philosophes moraux, des philosophes généraux. En regard de cette organisation, la suggestion des XPhi est que des philosophes généraux, traitant de tout type de question philosophique, puissent réaliser eux-mêmes le travail des scientifiques expérimentaux, et en particulier dans le domaine moral, aillent jusqu'à jouer le rôle de psychologues expérimentaux.

Les remarques précédentes sur la prise en compte de l'opérationnalisation comme point d'articulation des organisations me conduisent à formuler deux points de vue opposés sur cette question. Premier point de vue, optimiste, l'opérationnalisation telle qu'elle se construit en psychologie est amenée à progresser rapidement avec l'essor des sciences cognitives. En cohérence avec la haute technicité atteinte aujourd'hui par les études psychologiques, en par-

ticulier avec leur lien croissant avec les études neuronales, il est préférable (et sera nécessaire du fait des technicités exigées) de mettre en place des équipes spécialisées, et, de façon analogue à l'ensemble des sciences naturelles, de distinguer fortement les psychologues théoriciens des psychologues praticiens. Ceci suggère, sans bien sûr l'imposer, que l'organisation de la recherche en philosophie morale puisse se calquer sur l'organisation actuelle de la philosophie des sciences. La philosophie morale divergerait alors en, d'une part, une philosophie de la psychologie analogue à, par exemple, la philosophie de la biologie (la psychologie morale et cette nouvelle branche de philosophie de la psychologie recouvrant approximativement à elles-deux ce que j'ai appelé plus haut la perspective descriptive) et d'autre part, une philosophie normative regroupant les perspectives substantielles et méta-éthiques. Dans cet esprit, le mouvement XPhi ne serait que transitoire et aurait principalement pour but de clarifier et accélérer la prise en charge des problèmes philosophiques normatifs par la psychologie morale et, en fin de compte, se dissoudrait dans cette psychologie morale. [8]

Second point de vue, pessimiste, le constat réalisé sur la base de l'effet Knobe n'est ni particulier ni transitoire, mais général et pérenne : de façon durable, il n'y aura pas d'opérationnalisation satisfaisante des concepts issus des problématiques philosophiques et les avancées expérimentales seront très lentes ou inexistantes, sous les effets conjoints de la difficulté du sujet d'étude et des oppositions idéologiques profondes. Dans ce cadre, le mouvement XPhi se pérennise également et répond de façon durable à la nécessité de disposer d'éléments empiriques face à des questions philosophiques, non prises en compte par la psychologie morale qui progressera sur la base de concepts psychologiques différents des concepts philosophiques.

Entre ces deux points de vue extrêmes on peut envisager un ensemble de positions intermédiaires selon le degré et la vitesse d'avancement de l'opérationnalisation en psychologie, et particulièrement en psychologie morale. Il est envisageable que la période de transition qui s'ouvre, si l'on est optimiste, ou la période de confusion, si l'on est pessimiste, soit de longue durée avant que ne soient clarifiés sur quels points (provisoirement et relativement) fixes le théoricien et expérimentateurs psychologues pourront s'appuyer pour structurer l'opérationnalisation de la psychologie morale.

7.6 Que puis-je espérer ?

A l'issue de ce travail, revenons sur la question posée à la fin de l'introduction de la présente thèse, « Que peut la psychologie expérimentale pour la philosophie morale ? ». En m'ap-

8. Ce phénomène qui voit les philosophes engagés dans le mouvement XPhi ne le faire que transitoirement est bien ce qui semble être le cas en France où la psychologie morale a pris le pas sur le mouvement XPhi.

puyant sur l'examen du mouvement XPhi de la philosophie morale expérimentale et sur la mise en abyme qu'ont représenté les cinq études de cas, à l'articulation de la psychologie et de la philosophie morale, j'ai défendu les acquis méthodologiques et les propositions que je viens de rappeler, je voudrai finalement reformuler, encore une fois, comment ces acquis et propositions transforment cette question en la démultipliant.

Les raisons de répondre « rien ou si peu » à la question du « que puis-je savoir de ce que je dois faire » sont nombreuses et profondes, trois questions peuvent marquer ce territoire du doute :

— Les scientifiques peuvent-ils produire des connaissances pertinentes dans un domaine où tout ce qu'ils produisent est immédiatement interprété, et surinterprété, philosophiquement et au prisme des idéologies en place, cette interprétation étant déjà en place avant même qu'ils ne commencent à produire des résultats ?

— Les philosophes moraux, les politiques et les religieux sont-ils prêts à recevoir ces résultats, quelles qu'en soient les conséquences sur les théories morales en place ?

— Comment trouver le juste chemin qui irait de la conception humble de la vérité scientifique, toujours remise en doute, vers le jugement moral, nécessaire justification de l'action, ici et maintenant ?

Mais les raisons d'espérer des réponses plus positive existent, et, à nouveau, formulons-les en trois questions :

— Les sociétés humaines ont fait évoluer leurs règles morales dans les siècles passés sous la pression des changements qu'elles ont subis. L'abolition de l'esclavage et la lente montée des égalités en témoignent. Pourquoi n'en serait-il pas de même aujourd'hui face à la nécessité de refonder les règles morales pour une humanité globalisée ?

— Certes, les morales locales peuvent se durcir en outils d'exclusion mortifères, mais les connaissances transversales à toute frontière s'imposent chaque jour dans tous les domaines de la vie pratique, pourquoi s'arrêteraient-elles là où le siècle en a besoin ?

— La théorie de l'évolution éclaire l'histoire de l'espèce humaine, des sociétés humaines et des individus, offrant un cadre inédit à la réflexion morale, comment ne pas espérer qu'elle contribue, justement, à son évolution ?

Annexe A

Petit lexique des théories morales

Le présent lexique a eu pour principale utilité pendant la réalisation de cette thèse de me permettre de situer les nombreuses variantes de théories morales rencontrées dans les débats entre philosophes moraux, en tant qu'elles peuvent apparaître dans les analyses développées par les philosophes expérimentaux. La description est réduite à une phrase courte ne donnant qu'un sens très général, rappelons que chaque terme a pu faire l'objet d'interprétations multiples et divergentes.

L'ouvrage de Timmons, « Moral Theory : an introduction » [228] a servi de première amorce pour cette liste qui a ensuite été complétée à l'occasion des différentes études réalisées.

Elle est fournie ici non à titre de liste de référence, et encore moins exhaustive, mais comme outil et point de départ pour les travaux futurs.

Absolutisme moral	Certaines interdictions sont absolues, indépendamment du contexte et des conséquences.
Acrasie	Agir contre son jugement moral réfléchi.
Altruisme	Prendre en considération les intérêts d'autrui. Opposé à l'égoïsme.
Anarchisme moral	Les règles morales sont universelles et individuelles, aucune autorité intermédiaire (état, église, société, . . .) ne peut instituer de règle morale.
Animalisme	L'animal n'est pas un acteur moral mais, en tant qu'être vivant sensible, doit être protégé par les règles morales.

Ataraxie	Le Bien est défini comme l'absence de préoccupation, la tranquillité de l'âme.
Atomisme moral	La force d'un argument moral atomique portant sur un des éléments (l'acte, l'acteur, l'intention de l'acteur,...) reste inchangée dans toute situation. C'est le contraire du holisme moral où cette force dépend de l'ensemble du contexte.
Casuistique	L'évaluation morale s'appuie principalement sur la comparaison à des cas semblables qui permettent ainsi de comprendre le sens profond des règles appliquées.
Cognitivisme	Le jugement moral exprime une croyance qui peut être vraie ou fausse. Il résulte d'un processus cognitif.
Commandement divin	Les règles morales existent, sont réelles et sont des commandements divins.
Conséquentialisme de l'acte	Chaque acte est moralement évalué selon ses conséquences.
Conséquentialisme de la règle	Chaque règle morale est évaluée selon les conséquences généralement observées lorsque la règle est suivie.
Constructivisme	Les règles morales se sont construites au cours d'un processus historique culturel. S'oppose aux théories qui cherchent un fondement moral anhistorique.
Contractualisme	Les règles morales sont issues d'un contrat, par exemple entre l'individu et la société.
Déontologisme	La morale est essentiellement affaire de devoirs exprimés dans des règles à suivre.
Discussion	L'éthique de la discussion pose l'accord suite à discussion comme définition du Bien.
Double effet	Il est moralement acceptable de faire un Mal si c'est un effet second non désiré d'un acte visant un plus grand Bien.
Égoïsme	Principe d'action visant au seul intérêt direct de l'acteur.
Émotivisme	Les énoncés évaluatifs sont l'expression d'une émotion. Ils ne réfèrent pas à une propriété externe mais seulement à un état psychologique interne. Nombreux faux amis possibles selon le rôle donné aux émotions dans le jugement moral.

Éthiques appliquées	Branche de l'éthique concernant un domaine particulier, par exemple la santé ou les affaires, et appliquée à des cas particuliers.
Éthique du soin	Le soin (en anglais : care) apporté aux autres est central pour ces théories, en opposition à la règle d'or exprimée négativement qui pose comme central de ne pas nuire.
Eudémonisme	Le bonheur est le but de la vie humaine, et définit le Bien à rechercher.
Euthymie	Le Bien est défini comme le calme intérieur, l'absence de passions et de douleurs.
Euthyphro	Dilemme : Une règle est-elle morale parce que Dieu l'émet ou Dieu est-il bon parce qu'il suit les règles morales ?
Expressivisme	Les énoncés moraux ne sont que l'expression d'une approbation ou d'une réprobation. Ils n'ont pas de valeur de vérité.
Fictionnalisme	Les énoncés moraux sont des fictions, au même titre que les contes ou les romans, constituants d'une culture commune à une société mais sans qu'existent des propriétés morales réelles.
Hédonisme	L'expérience du plaisir est intrinsèquement un Bien et celle de la douleur un Mal et ce sont les seules sources de bien et de mal.
Hétéronomie	Les règles morales ont une source extérieure à l'acteur moral. S'oppose à Autonomie.
Holisme	Le jugement moral porte sur l'ensemble d'une situation et non sur quelques traits moralement plus importants
Impératif catégorique	Un impératif qui s'impose à tout être rationnel en tant que tel dans toute situation. S'oppose à impératif hypothétique qui dépend de la fin recherchée.
Intuitionnisme	L'intuition nous donne accès directement à un jugement moral vrai.
Minimalisme	Ce qu'un individu se fait à lui-même ne relève pas du domaine moral. S'oppose au maximalisme perfectionniste où chacun est moralement engagé à se perfectionner.
Monisme	Il existe un seul Bien (ou une seule valeur) auquel se rapportent toutes les règles morales. S'oppose à pluralisme.

Naturalisme réductionniste	Les propriétés morales sont naturelles et réductibles à la base physique du monde comme toute propriété naturelle.
Naturalisme non réductionniste	Les propriétés morales sont naturelles mais d'une sorte particulière non réductible à la base physique du monde.
Nihilisme moral	Il n'existe pas de propriétés morales.
Objectivisme	Le jugement moral est principalement dépendant des caractéristiques morales du fait jugé.
Particularisme	Tous les cas sont particuliers, il n'existe pas de principes moraux généraux que l'on puisse apprendre.
Perfectionnisme	Est Bien ce qui vise à améliorer notre être.
Pluralisme	Les théories morales, dont les conceptions du Bien, ne sont pas exclusives, il faut les composer en fonction des situations pour comprendre le phénomène moral.
Principisme	Quatre principes sont à la base de l'éthique médicale, ces principes suffisent à déterminer le jugement moral lorsqu'ils sont soigneusement pesés.
Quasi réalisme	Les propriétés morales ne sont pas réelles, elles n'existent que dans nos têtes. Mais à ce titre nous nous comportons comme si elles étaient réelles, et les théories morales doivent prendre en compte ces propriétés quasi réelles.
Quiétisme	Ne rien entreprendre, et s'en remettre à Dieu, est le vrai Bien moral.
Réalisme moral	Les faits moraux et les règles morales existent et font partie de l'ontologie du monde au même titre que les autres propriétés.
Règle d'or	La règle d'or de la réciprocité : ne fais pas à autrui ce que tu ne veux pas qu'autrui te fasse. (en anglais : Golden Rule). Définit le Bien comme ce qui est équilibré par réciprocité.
Relativisme	Les énoncés moraux ne sont vrais que relativement à un contexte social donné.
Rigorisme	Appliquer rigoureusement des règles morales à toute situation.
Sens Moral	Les humains sont équipés d'un sens moral qui leur permet de percevoir la dimension morale d'une situation, au même titre que la vue permet de percevoir les couleurs.

Sentimentalisme	Les sentiments (émotions, sentiments, désirs) jouent un rôle majeur dans nos jugements moraux. S'oppose à rationalisme : la raison joue le premier rôle.
Situationnisme	Chaque situation est unique du point de vue moral. Aucune règle morale ne peut ni rendre compte de ces spécificités ni être utilisée pour guider notre jugement moral.
Subjectivisme	Le jugement moral dépend du sujet qui l'émet et n'est pas objectivement lié au fait jugé.
Universalisme	Les règles morales sont universelles. S'oppose à relativisme.
Utilitarisme	Sorte de conséquentialisme qui pose l'existence d'une fonction d'utilité à maximiser.
Valeur morale	Certaines théories posent l'existence de valeurs morales comme ontologiquement premières, les règles morales étant secondes et ayant pour objet de maximiser les valeurs.
Vertu	Trait de caractère, disposition à Bien agir et à Bien juger.
Vice	Trait de caractère, disposition à Mal agir et à Mal juger.

Annexe B

Liste des articles relatifs à l'effet Knobe

La liste détaille les 33 articles retenus pour réaliser l'étude de cas relative à l'effet Knobe (voir 4.6 , page 219).

La première colonne est le numéro d'ordre utilisé dans la thèse. Il résulte directement de l'ordre produit par l'outil bibliographique de Sorbonne Université, sans intervention ultérieure.

La deuxième colonne précise le type de revue, Phi pour philosophie, Psy pour psychologie, Sci pour sciences généralistes, Man pour management.

La troisième colonne indique le mode de recrutement des participants à l'étude, AMT pour Amazon Mechanical Turk, Child pour des enfants des écoles, Int pour un système internet différent de AMT, Stu pour des étudiants de l'université abritant l'étude, Scan pour l'étude par IRM (qui porte également sur des étudiants) , et enfin No indique qu'il n'y a pas d'expérimentation rapportée dans cet article.

La quatrième colonne indique les auteurs, la cinquième le titre de l'article et la sixième donne les références de la revue.

Ces données proviennent de la base bibliographique de Sorbonne Université et n'ont subi qu'un traitement de forme sans changement de contenu.

Num	Type Revue	Type Expe	Auteurs	Titre	Revue

1	Psy	Stu	Hilton Denis J., McClure John, Moir Briar	Acting knowingly : effects of the agent's awareness of an opportunity on causal attributions.	Thinking & Reasoning. Nov2016, Vol. 22 Issue 4, p461-494. 34p. 5 Charts, 2 Graphs.
2	Phi	Web	Hindriks Frank, Douven Igor, Singmann Henrik	A New Angle on the Knobe Effect : Intentionality Correlates with Blame, not with Praise.	Mind & Language Apr2016, Vol. 31 Issue 2, p204-220, 17p, 1 Black and White Photograph, 1 Graph
3	Psy	AMT	Feldman Gilad, Wong Kin Fai, Ellick Baumeister Roy F.	Bad is freer than good : Positive'negative asymmetry in attributions of free will	Consciousness and Cognition May 2016 42 :26-40
4	Psy	AMT	Chituc Vladimir, Henne Paul et al	Blame, not ability, impacts moral 'ought' judgments for impossible actions : Toward an empirical refutation of 'ought' implies 'can'	Cognition May 2016 150 :20-25
5	Psy	Child	Margoni Francesco, Surian Luca	Children's intention-based moral judgments of helping agents	Cognitive Development January-March 2017 41 :46-64
6	Sci	AMT	Plaks Jason E., Fortune Jennifer L. et al	Effects of Culture and Gender on Judgments of Intent and Responsibility.	PLoS ONE | gPLoS ONE. 4/28/2016, Vol. 11 Issue 4, p1-19. 19p.
7	Psy	AMT	Kim Nancy, Johnson Samuel et al	The effect of abstract versus concrete framing on judgments of biological and psychological bases of behavior.	Cognitive Research : Principles & Implications 3/20/2017, Vol. 2 Issue 1, p1-16, 16p

8	Psy	AMT	Murray Dylan, Lombrozo Tania	Effects of Manipulation on Attributions of Causation, Free Will, and Moral Responsibility.	Cognitive Science; Mar2017, Vol. 41 Issue 2, p447-481, 35p
9	Phi	No	Fischborn Marcelo	Questions for a Science of Moral Responsibility.	Review of Philosophy & Psychology Jun2018, Vol. 9 Issue 2, p381-394, 14p
10	Phi	AMT	Rose David	Folk intuitions of actual causation : a two-pronged debunking explanation.	Philosophical Studies. May2017, Vol. 174 Issue 5, p1323-1361. 39p.
11	Psy	AMT	Martin Justin W., Cushman Fiery	Why we forgive what can't be controlled	Cognition February 2016 147 :133-143
12	Psy	AMT	Clark Cory J., Shniderman Adam et al	Are morally good actions ever free?	Consciousness and Cognition August 2018 63 :161-182
13	Psy	Stu	Siegel Jenifer Z., Crockett Molly J., Dolan Raymond J.	Inferences about moral character moderate the impact of consequences on blame and praise	Cognition Moral Learning, October 2017 167 :201-211
14	Phi	AMT	Cova Florian	Intentional action and the frame-of-mind argument : new experimental challenges to Hindriks.	Philosophical Explorations. Mar2017, Vol. 20 Issue 1, p35-53. 19p.
15	Phi	No	Buckwalter Wesley	Intuition Fail : Philosophical Activity and the Limits of Expertise.	Philosophy & Phenomenological Research Mar2016, Vol. 92 Issue 2, p378-410, 33p

16	Psy	Stu	Parkinson Mary, Byrne Ruth M. J.	Judgments of moral responsibility and wrongness for intentional and accidental harm and purity violations.	Quarterly Journal of Experimental Psychology ; Mar2018, Vol. 71 Issue 3, p779-789, 11p
17	Psy	AMT	Turri John, Friedman Ori, Keefner Ashley	Knowledge central : A central role for knowledge attributions in social evaluations.	Quarterly Journal of Experimental Psychology ; Mar2017, Vol. 70 Issue 3, p504-515, 12p
18	Psy	AMT	Gerstenberg Tobias, Ullman Tomer D. et al	Lucky or clever ? From expectations to responsibility judgments	Cognition August 2018 177 :122-141
19	Phi	No	Friedman Jared P., Jack, Anthony I.	Mapping Cognitive Structure onto the Landscape of Philosophical Debate : an Empirical Framework with Relevance to Problems of Consciousness, Free will and Ethics.	Review of Philosophy & Psychology Mar2018, Vol. 9 Issue 1, p73-113, 41p
20	Psy	Juges	Kneer Markus, Bourgeois-Gironde Sacha	Mens rea ascription, expertise and outcome effects : Professional judges surveyed	Cognition December 2017 169 :139-146
21	Psy	Scan Stu	Chakroff Alek, Dungan James et al	When minds matter for moral judgment : intent information is neurally encoded for harmful but not impure acts.	Social Cognitive & Affective Neuroscience Mar2016, Vol. 11 Issue 3, p476-484, 9p
22	Psy	Stu	Railton Peter	Moral Learning : Conceptual foundations and normative relevance	Cognition Moral Learning, October 2017 167 :172-190

23	Psy	No	Dhar Sharmistha	THE ONTOLOGY OF AGENCY IN THE LIGHT OF DETERMINISTIC CAUSATION : A FOLK-PSYCHOLOGICAL STUDY.	Journal of East-West Thought (JET); Jun2018, Vol. 8 Issue 2, p61-76, 16p
24	Psy	AMT	De Freitas Julian, Johnson, Samuel G.B.	Optimality bias in moral judgment	Journal of Experimental Social Psychology November 2018 79 :149-163
25	Phi	AMT	Björnsson Gunnar	Outsourcing the deep self : Deep self discordance does not explain away intuitions in manipulation arguments./	Philosophical Psychology. Jul2016, Vol. 29 Issue 5, p637-653. 17p.
26	Psy	Web	Redford Liz, Ratliff Kate A.	Perceived moral responsibility for attitude-based discrimination.	British Journal of Social Psychology. Jun2016, Vol. 55 Issue 2, p279-296. 18p.
27	Psy	AMT	Khamitov Mansur, Rotman Jeff D., Piazza Jared	Perceiving the agency of harmful agents : A test of dehumanization versus moral typecasting accounts	Cognition January 2016 146 :33-47
28	Psy	Web	Samland Jana, Waldmann Michael R.	How prescriptive norms influence causal inferences	Cognition November 2016 156 :164-176
29	Psy	Stu	Deutschländer Robert, Pauen Michael, Haynes John-Dylan	Probing folk-psychology : Do Libet-style experiments reflect folk intuitions about free action?	Consciousness and Cognition February 2017 48 :232-245

30	Phi	Stu	Felletti Silvia, Paglieri Fabio	The illusionist and the folk : On the role of conscious planning in intentionality judgments.	Philosophical Psychology. Aug2016, Vol. 29 Issue 6, p871-888. 18p.
31	Psy	Street	Robbins Erin, Shepard Jason, Rochat Philippe	Variations in judgments of intentional action and moral evaluation across eight cultures	Cognition July 2017 164 :22-30
32	Psy	AMT	De Freitas Julian, Alvarez George A.	Your visual system provides all the information you need to make moral judgments about generic visual events	Cognition September 2018 178 :133-146
33	Man	Stu	Kaspar Kai, Newen Albert et al	Whom to blame and whom to praise.	International Journal of Cross Cultural Management ; Dec2016, Vol. 16 Issue 3, p341-365, 25p

Épilogue

Des primates, un jour, on ne sait plus trop pourquoi ni comment, furent atteints d'hypertélie cérébrale [1]. Leur cerveau se mit à croître sans fin et, génération après génération, à l'étroit dans la boîte crânienne, cet organe se replia sur lui-même en de multiples circonvolutions. Au début, rien ne distinguait ces primates de leurs congénères non atteints par la maladie. Ils avaient bien de temps en temps des impressions bizarres, des perceptions qui s'entrechoquaient, provoquant quelques vertiges, des images qui évoquaient des sons et des sons des paysages, des analogies impensées s'imposaient à eux, mais tout cela sans que leur comportement ne change. Peu à peu, ils élaborèrent ces analogies et ces images pour se raconter des mondes qui n'existaient que par leur imagination. Ils devinrent des Homo Fictionalis, comme on allait les baptiser bien plus tard. Certaines de leurs histoires parlaient de ce qu'ils voyaient du monde, d'autres de ce qu'ils auraient aimé y voir, d'autres encore de ce qu'ils attendaient de leurs amis et de ce qu'ils craignaient de leurs ennemis et chaque histoire racontée, chaque fiction partagée et enseignée aux enfants, resserrait un peu plus les liens entre les Homo Fictionalis. Un peu à part, mais admis par leurs congénères, ils vivaient comme leurs ancêtres l'avaient toujours fait.

Et alors advinrent de grandes catastrophes, des sécheresses, des déluges, des canicules, des migrations contraintes, et, avec ces événements terribles, un phénomène curieux se développa. Ces mondes imaginaires qui résultaient des rêveries stimulées par le cerveau prompt aux courts-circuits d'Homo Fictionalis s'avérèrent particulièrement utiles car ils les préparaient à tous ces changements imprévisibles imposés par la dureté des éléments. Et, surtout, les histoires racontées avaient créé des liens d'une solidarité vitale entre ceux qui partageaient ces rêves. Ils étaient préparés, ensemble, à des avenirs incertains là où leurs congénères

1. Merci à Jean-Pierre Gasc pour la découverte de cette belle expression dans (Gasc 2019) [95].

perdaient tout repère. Et c'est ainsi que ceux qui savaient raconter des histoires survécurent.

Les courts-circuits continuèrent, et avec eux la création d'histoires, et même la création d'histoires sur les histoires. Des fictions sur les fictions. Et de certaines fictions, on inventa qu'elles pouvaient dire le monde, d'autres qu'elles étaient la source des courts-circuits, la source de l'imagination, d'autres enfin qu'elles disaient ce que l'on aimait que nos amis fassent, et que nos ennemis ne faisaient pas. Et on dit de ces fictions qu'elles étaient vraies, belles ou morales. Avec ces fictions sur les fictions, ce qu'on appellera bien plus tard la philosophie était née.

Et les courts-circuits continuèrent. Et les fictions sur les fictions se développèrent. Et de ces fictions sur les fictions on se demanda, ce qui est beau est-il vrai ? Ce qui est moral est-il vrai ? Et les vertiges reprirent.

Cette histoire pourrait être la notre. Ce n'est pas certain, c'est seulement possible. Et chercher à le vérifier offre aux expérimentateurs et aux philosophes moraux qui s'intéressent au phénomène moral un cadre général, programmatique, qui rend compte à la fois des découvertes et des analyses prometteuses dues aux approches évolutionnistes et, à la fois, des réserves des philosophes moraux qui hésitent à les accepter. Ce cadre programmatique ne dit pas en lui-même d'où il vient, et cette lacune laisse la place à tous les choix métaphysiques imaginables. Résumons ce cadre à titre d'épilogue de la présente enquête.

— Les humains, Homo Fictionalis, ont la capacité essentielle de produire des fictions qui contribuent à créer et cimenter les groupes sociaux ainsi qu'à anticiper les situations à venir. L'espèce est sociale, rationnelle, projective et dialogique.
— Cette capacité collective repose sur la capacité de chaque individu à intérioriser les fictions partagées, à en faire une composante de son identité.
— Les fictions partagées et intériorisées qui régulent le comportement au sein du groupe reçoivent le nom de « morale ». Les nommer ainsi est auto-réalisateur.
— Ces fictions morales, et ce n'est qu'un constat, cumulent deux effets, cimenter le « nous » et se développer en opposant ce « nous » à des « eux », des « non nous » fictionnels.

Accepter ce cadre programmatique, c'est accepter que la morale ne soit pas « réaliste », au sens académique d'une réalité des propriétés morales en l'absence de « nous ». Sans Homo Fictionalis, plus d'histoire du tout, et donc plus d'histoire morale, et plus non plus d'histoires vraies ou belles. Mais accepter ce cadre programmatique c'est aussi être profondément réaliste dans un second sens : il n'y a pas de « nous » sans propriété morale. Car il n'y a pas de monde où les individus humains vivraient isolés, sans groupes humains, et qu'il n'y a pas

de groupes humains sans fictions qui les cimentent, et certaines sont alors, fiction sur la fiction, vraies, belles ou morales. Cette impossibilité des humains sans fictions est un constat contingent, il est loisible à chacun de bâtir la métaphysique qui lui convient, au-delà de ce constat, et en cela le cadre programmatique ne vise qu'à des connaissances limitées au monde observable.

Accepter ce cadre programmatique, c'est aussi entrevoir qu'il puisse être possible d'expliquer différemment la nécessité des histoires et les contenus de ces histoires. La première, la nécessité des histoires, est une solution résultant des pressions évolutives aux problèmes qui se posent à une espèce à la fois sociale, les individus ne peuvent vivre qu'en groupes, rationnelle, chacun peut calculer son intérêt individuel et collectif, projective et dialogique, les groupes se coordonnent par des discours partagés auxquels s'identifient les individus. Les seconds, les contenus des histoires, sont le résultat historique et contingent de la vie des groupes humains. Aucun contenu d'histoire n'est écrit d'avance, il se détermine avec, et est déterminé par, les circonstances qu'il contribue à faire évoluer. La nécessité de l'existence des histoires, dont la morale, est naturalisable en tant que fonction contribuant à la socialisation au sein d'une espèce dialogique. Le contenu de chaque morale est, pour partie, réalisation de ce rôle social nécessaire et, pour une large partie, contingente. Chaque individu, et chaque groupe, construit son chemin entre cette nécessité et les multiples contingences possibles. La seule morale impossible est de ne pas avoir de morale.

La démarche scientifique expérimentale est une démarche de collaboration, récente mais puissante, qui a pour effet de construire des théories, des histoires d'Homo Fictionalis, qui sont alors dites scientifiques. Jusqu'à un passé récent, elle avait pris pour objet des phénomènes matériels dont elle recherchait les régularités, mettant ainsi en avant, avec une grande fécondité et une impensable efficacité, des lois(fictionnelles) que toute instance de ces phénomènes respecte, aux erreurs, aux approximations ou au hasard près. Le développement de cette démarche l'a menée depuis quelques décennies vers les territoires du comportement humain. Et avec lui, vers la psychologie morale.

Accepter le cadre programmatique, c'est, en regard de la démarche scientifique expérimentale, ouvrir deux perspectives. La première est celle des axes de recherche qui, en prolongement du développement de l'ensemble des disciplines scientifiques concernées, avec la biologie, les sciences cognitives, la sociologie, l'anthropologie, la psychologie, etc. viseront à raconter à nouveaux frais en quoi l'espèce humaine ne peut qu'être morale pour qu'existent les groupes sociaux, et quelles contraintes cette nécessité existentielle induit sur le développement historique des groupes humains et, en particulier, sur les morales qu'ils se donnent.

Mais cette approche ne dira rien, ou si peu de choses, des cas particuliers. De celui de tel individu qui fait telle action ou émet tel jugement dans tel contexte, ou, prenant de la hauteur, de tel groupe qui adopte telle règle morale. C'est la seconde perspective ouverte par ce cadre programmatique que de chercher à combler ce vide entre le général et le particulier, entre les contraintes induites par notre nature, sociale, rationnelle, projective et dialogique, et les histoires morales que « nous » nous raconterons demain, dans ce nouveau contexte évolutif où, d'une part, nous accéderons à la connaissance de ces contraintes de plus en plus finement, et, d'autre part, ce « nous » sera à l'échelle de l'espèce entière et ne pourra plus se construire contre un « eux », contexte rendant les morales existantes pour une large part inopérantes. Dans les deux perspectives, chacune des histoires qui constituent la philosophie morale et la démarche scientifique expérimentale a des atouts, chacune a des limites.

Table des figures

3.1 L'hélice scientifique expérimentale 124

4.1 Tramway, scénarios et théories d'après Bruers et Braeckman 2014 166
4.2 La surestimation du nombre de musulmans (source Ipsos 2016) 175
4.3 La surestimation du nombre d'immigrés (source Ipsos 2014) 176
4.4 Corrélation entre estimations du nombre d'immigrés et de musulmans 181
4.5 Nombre de réponses par tranche d'estimation selon l'intention de vote 182
4.6 Nombre de réponses par tranche de 5 % 183
4.7 IAT Nombre d'articles par année (source Google Scholar) 203

Index des Auteurs

Alexander, 86, 94, 96
Alvarez, 232
Andler, 121, 145, 337, 365
Anquetil, 51
Anscombe, 44, 349
Anstey, 101
Appiah, 101, 357, 391, 396
Appourchaux, 117
Arkes, 205
Arksey, 221
Aubé, 348, 380, 389

Bègue, 172, 331
Baertschi, 167
Banaji, 204
Barberousse, 121, 280, 296
Barwich, 149
Bauman, 246
Baumard, 30, 384
Baumeister, 228
Bensaude, 127
Bernard, 132
Bertin, 141
Bicchieri, 357
Bickle, 92
Bjornsson, 231
Blanton, 205
Boudon, 90
Bourgeois Gironde, 230
Bourget, 376
Bouzeghoub, 89
Boyd, 96
Braeckman, 164
Bricmont, 338
Bridgman, 298
Bruers, 164
Brunel, 216
Buckwalter, 227
Byrne, 228

Caillaud, 315
Canguilhem, 398
Canto-Sperber, 58
Cappelen, 106
Chakravartty, 130
Chakroff, 228
Chalmers, 376
Chang, 149, 301
Chem, 117
Chevassus-au-Louis, 334
Chituc, 229, 351
Chomsky, 337
Churchland, 327
Clark, 229
Clavier, 45
Colaço, 107

Collins, 365
Colombo, 107
Connellan, 97
Coolican, 293, 314
Cova, 86, 91, 106, 155, 229, 284, 326, 340
Crockett, 229
Cullen, 96, 313
Curry, 387
Cushman, 227

Damasio, 88
Danziger, 174
Daston, 139
Dawkins, 130, 251
de Waal, 394
Debove, 385, 395
DeFreitas, 230, 232
Dehaene, 17, 316, 404
Dennett, 384
Desmons, 250, 326
Deutsch, 105, 332
Deutschlander, 231
Dewey, 122, 407
Dhar, 230
Dolan, 229
Doris, 335, 381
Douven, 228
Duffy, 186
Duhem, 144, 149
Duke, 172

El Skaf, 137
Enoch, 74, 325
Evers, 29, 377, 380

Feest, 263, 298, 299
Feldman, 228
Felletti, 231
Feyerabend, 132
Fiedler, 205
Fine, 131
Fischborn, 227, 238
Flick, 315
Foot, 158, 163
Forest, 405
Franceschi, 338
Franklin, 143
Friedman, 230

Galison, 139, 143, 148
Gasc, 441
Gensler, 42
Gerstenberg, 229
Gert, 33
Getstein, 191
Gettier, 95
Ghiglione, 90, 314
Gilligan, 331
Gingras, 148
Giraud, 162
Gleichgerrcht, 171
Gould, 386, 391
Gouverneur, 348
Gracanin, 194
Grawitz, 314
Greene, 46, 62, 65, 88, 158, 161, 381
Greenwald, 197, 201, 202, 204, 205

Hacking, 124, 143, 145, 196
Haidt, 29
Hajek, 215

Haynes, 231
Hilton, 227
Hindriks, 228
Horvath, 106
Houdé, 87, 380
Huemer, 28, 29

Ioannidis, 153

Jack, 230
Jaffro, 384
Jaquet, 326
Jebeile, 138
Jensen, 89
Johnson, 230

Kahane, 249, 337
Kahneman, 177, 204, 338
Kammerer, 140, 319
Kant, 49
Kaspar, 232
Kaufman, 215
Kauppinen, 106, 332, 369
Keefner, 230
Khamitov, 231
Kim, 229
Kistler, 280
Kitcher, 73, 291, 348, 356, 376, 389, 401
Kneer, 230
Knobe, 85, 86, 94, 111, 116, 161, 219
Kripke, 98
Krivine, 126
Kropotkin, 50

Larmore, 333
Lemaire, 250, 326
Libet, 117, 147, 340
Lim, 117
Livengood, 89, 103, 108
Lombrozo, 229
Ludwig, 345
Lutge, 161

Machery, 86, 93, 98, 107, 363, 384
Mallon, 384
Malone, 296
Manzo, 395
Margoni, 229
Markie, 128
Martin, 227
Martinez Navarro, 333
Masala, 29
Mazoyer, 87, 380
Mcarthur, 131
McClure, 227
McConnell, 205
McGhee, 197, 201, 202
Merrill, 395
Michel, 345
Michell, 323
Milgram, 136, 313
Moir, 227
Moore, 35, 348
Mori, 115
Mukerji, 86, 93, 118, 154, 161, 221, 242, 255, 335
Muldoon, 357
Mullins, 387
Munafo, 154
Murray, 229

Nagel, 96
Nahmias, 117
Nichols, 85, 86, 94
Nicolas, 90
Nisbett, 89
Nosek, 200, 204
Nurock, 348

O'Neill, 86
Ockham, 144
Ogien, 29, 35, 47, 58, 348
ONeill, 161
Oswald, 205

Paglieri, 231
Paluck, 214
Pariente-Butterlin, 34
Parkinson, 228
Parris, 338
Pauen, 231
Piazza, 231
Piccinini, 140, 167, 345
Pickering, 124
Plaks, 229
Popper, 144
Prinz, 30, 47, 77, 116, 360
Puech, 385
Putnam, 337

Quine, 144, 149

Railton, 230
Ratliff, 231
Ravat, 26, 62, 69, 359, 396
Rawls, 73, 369, 384
Redford, 231
Renaud, 192
Robbins, 232
Rochat, 232
Rose, 227
Roskies, 376
Rotman, 231
Rousselle, 395
Ruphy, 291, 356, 376
Russell, 97

Samland, 228
Sandberg, 264
Sandel, 353
Sarkissian, 86, 161, 332
Sauer, 349
Savidan, 395
Sayre-McCord, 27
Schwartz, 197, 201, 202
Searle, 349
Serra, 161, 330
Sheeran, 171
Shepard, 232
Sicard, 371
Siegel, 229
Silberstein, 395
Singer, 27, 395
Singmann, 228
Skinner, 297
Skitka, 51
Slovic, 204, 338
Sobel, 191, 194
Sockeel, 314
Sokal, 338
Sontuoso, 357

Sosa, 106, 124, 128
Spranzi, 369, 371, 425
Stevens, 246
Stitch, 95
Stojanovic, 326
Strickland, 159, 175, 253
Stuart, 107
Suikkanen, 369
Surian, 229
Swain, 96
Sytsma, 86, 89, 103, 108, 161

Tallon-Baudry, 365
Tappolet, 27, 49
Tetlock, 205
Thomson, 163
Tiercelin, 346
Timmons, 53, 429
Todd, 326
Tufte, 141
Turmel, 250
Turri, 230
Tversky, 204, 338
Tzourio-Mazoyer, 380

Van Fraassen, 143, 149, 290
Vanzo, 101
Vautier, 323
Vingerhoets, 193

Waal, 148
Waldmann, 228
Watson, 296
Weinberg, 95, 96
Whitehouse, 387
Williamson, 86, 93, 101, 132, 242, 338, 339
Wilson, 384
Wolff, 72
Wong, 228
Wright, 86, 332
Wunsch, 258

Bibliographie

[1] Joshua Alexander. *Experimental Philosophy : An Introduction*. Polity, Cambridge, UK ; Malden, MA, 2012.

[2] Daniel Andler. Dissensus in Science as a Fact and as a Norm. In Hanne Andersen, Dennis Dieks, Wenceslao J. Gonzalez, Thomas Uebel, and Gregory Wheeler, editors, *New Challenges to Philosophy of Science*, pages 493–506. Springer Netherlands, Dordrecht, 2013.

[3] Daniel Andler. *La Silhouette de l'humain : Quelle Place Pour Le Naturalisme Dans Le Monde d'aujourd'hui ?* NRF Essais. Gallimard, Paris, 2016.

[4] Daniel Andler, Anne Fagot-Largeault, and Bertrand Saint-Sernin. *Philosophie Des Sciences*. Number 405 in Collection Folio/Essais. Gallimard, Paris, 2002.

[5] Alain Anquetil. *Qu'est-ce que l'éthique des affaires ?* J. Vrin, Paris, 2008.

[6] Alain Anquetil. *Ethique des affaires : marché, règle, responsabilité*. J. Vrin, Paris, 2011.

[7] Gertrude E.M. Anscombe. Modern Moral Philosophy. *Philosophy*, 33(124) :1, 1958.

[8] Peter R Anstey and Alberto Vanzo. Early Modern Experimental Philosophy. In *A Companion to Experimental Philosophy*, pages 87–102. Blackwell Publishers, 2016.

[9] Kwame Anthony Appiah. *Experiments in Ethics*. The Mary Flexner Lectures. Harvard Univ. Press, Cambridge, Mass., 1. harvard univ. press paperback ed edition, 2009.

[10] Kwame Anthony Appiah. *The Lies That Bind : Rethinking Identity : Creed, Country Colour, Class, Culture*. Profile Books, London, first published edition, 2018.

[11] Krystèle Appourchaux. *Un Nouveau Libre Arbitre : À La Lumière de La Psychologie et Des Neurosciences Contemporaines*. CNRS éditions, Paris, 2014.

[12] Hal R. Arkes and Philip E. Tetlock. TARGET ARTICLE : Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test ?". *Psychological Inquiry*, 15(4) :257–278, October 2004.

[13] Hilary Arksey and Lisa O'Malley. Scoping studies : Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1) :19–32, February 2005.

[14] Beaudoin Aubé. Ethique évolutionniste, version académique. *Kristanek (dir.), l'Encyclopédie philosophique,*, 2017.

[15] William Aubé. *Corrélats neuronaux sous-jacents aux expressions émotionnelles : une comparaison entre musique, voix et visage*. psychologie–recherche & intervention option neuropsychologie clinique, Université de Montréal, Montréal (Québec), 2014.

[16] Anouk Barberousse, Denis Bonnay, and Mikaël Cozic, editors. *Précis de Philosophie Des Sciences*. Collection "Philosophie Des Sciences". Vuibert, Paris, 2011.

[17] Ann-Sophie Barwich and Hasok Chang. Sensory Measurements : Coordination and Standardization. *Biological Theory*, 10(3) :200–211, September 2015.

[18] Christopher W. Bauman, A. Peter McGraw, Daniel M. Bartels, and Caleb Warren. Revisiting External Validity : Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology : External Validity in Moral Psychology. *Social and Personality Psychology Compass*, 8(9) :536–554, September 2014.

[19] Nicolas Baumard. *Une Théorie Naturaliste et Mutualiste de La Morale*. Thèse de Doctorat, EHESS, 2008.

[20] Nicolas Baumard. *Comment Nous Sommes Devenus Moraux : Une Histoire Naturelle Du Bien et Du Mal*. Odile Jacob, Paris, 2010.

[21] Louis Bègue. De la «cognition morale» à l'étude des stratégies du positionnement moral : aperçu théorique et controverses actuelles en psychologie morale. *L'année psychologique*, 98(2) :295–352, 1998.

[22] Bernadette Bensaude-Vincent, Antonio García Belmar, and José Ramón Bertomeu Sánchez. *L'émergence d'une Science Des Manuels : Les Livres de Chimie En France (1789-1852)*. Histoire Des Sciences, Des Techniques et de La Médecine. Archives contemporaines, Paris, 2003.

[23] Jacques Bertin and Marc Barbut. *Sémiologie graphique : les diagrammes, les réseaux, les cartes*. Ed. de l'EHESS, Paris, 2005.

[24] Cristina Bicchieri, Ryan Muldoon, and Alessandro Sontuoso. Social Norms. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition, 2018.

[25] John Bickle. Lessons for experimental philosophy from the rise and “fall” of neurophilosophy. *Philosophical Psychology*, 32(1) :1–22, January 2019.

[26] Gunnar Björnsson. Outsourcing the deep self : Deep self discordance does not explain away intuitions in manipulation arguments. *Philosophical Psychology*, 29(5) :637–653, July 2016.

[27] Hart Blanton, James Jaccard, Patricia M. Gonzales, and Charlene Christie. Decoding the implicit association test : Implications for criterion prediction. *Journal of Experimental Social Psychology*, 42(2) :192–212, March 2006.

[28] David Bourget and David Chalmers. What do philosophers believe ? *Philosophical Studies*, 170(3) :465–500, September 2014.

[29] Mokrane Bouzeghoub and Rémy Mosseri, editors. *Les Big Data à Découvert*. CNRS éditions, Paris, 2017.

[30] Richard Boyd, Philip Gasper, and J. D. Trout, editors. *The Philosophy of Science*. MIT Press, Cambridge, Mass, 1991.

[31] Percy W Bridgman. *The Logic of Modern Physics*. New York : Macmillan, 1927.

[32] Percy W Bridgman. "The Logic of Modern Physics" after Thirty Years. *Daedalus*, 88(3) :518–526, 1959.

[33] Stijn Bruers and Johan Braeckman. A Review and Systematization of the Trolley Problem. *Philosophia*, 42(2) :251–269, June 2014.

[34] Frédéric F. Brunel, Brian C. Tietje, and Anthony G. Greenwald. Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition ? *Journal of Consumer Psychology (Taylor & Francis Ltd)*, 14(4) :385–404, September 2004.

[35] Wesley Buckwalter. Intuition Fail : Philosophical Activity and the Limits of Expertise. *Philosophy and Phenomenological Research*, 92(2) :378–410, March 2016.

[36] Sabine Caillaud and Uwe Flick. Triangulation méthodologique. Ou comment penser son plan de recherche. In *Les représentations sociales.* G. Lo Monaco, S. Delouvée & P. Rateau, Bruxelles, g. lo monaco, s. delouvée & p. rateau edition, 2016.

[37] Georges Canguilhem. *Le normal et le pathologique*. Quadrige, Paris, 2007.

[38] Monique Canto-Sperber and Ruwen Ogien. *La philosophie morale*. PUF, Presses universitaires de France, 2017.

[39] Herman Cappelen. X-Phi without Intuitions ? In Anthony Robert Booth and Darrell P. Rowbottom, editors, *Intuitions*, pages 269–286. Oxford University Press, July 2014.

[40] Anjan Chakravartty. Scientific Realism. In *The Stanford Encyclopedia of Philosophy*. Zalta, April 2011.

[41] Alek Chakroff, James Dungan, Jorie Koster-Hale, Amelia Brown, Rebecca Saxe, and Liane Young. When minds matter for moral judgment : Intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, 11(3) :476–484, March 2016.

[42] Hasok Chang. *Inventing Temperature : Measurement and Scientific Progress*. Oxford Studies in Philosophy of Science. Oxford University Press, Oxford ; New York, 2007.

[43] Hasok Chang. Operationalism. In *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, edward n. zalta edition, 2019.

[44] Nicolas Chevassus-au-Louis. Fraude Scientifique. *E Universalis*, 2017.

[45] Vladimir Chituc, Paul Henne, Walter Sinnott-Armstrong, and Felipe De Brigard. Blame, not ability, impacts moral "ought" judgments for impossible actions : Toward an empirical refutation of "ought" implies "can". *Cognition*, 150 :20–25, May 2016.

[46] Noam Chomsky and Nicolas Calvé. *Quelle sorte de créatures sommes-nous ? : langage, connaissance et liberté*. Lux Editeur, Montréal, 2016.

[47] Patricia Smith Churchland. *Conscience : The Origins of Moral Intuition*. W. W. Norton & Company, New York, first edition, 2019.

[48] Cory J. Clark, Adam Shniderman, Jamie B. Luguri, Roy F. Baumeister, and Peter H. Ditto. Are morally good actions ever free ? *Consciousness and Cognition*, 63 :161–182, August 2018.

[49] Paul Clavier. Kant (A). *Kristanek (dir.), l'Encyclopédie philosophique,*, 2018.

[50] Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*, 349(6251) :4716, August 2015.

[51] Thérèse Collins, Daniel Andler, and Catherine Tallon-Baudry. *La cognition : du neurone à la société*. Folio Essais. Gallimard, Paris, 2018.

[52] Matteo Colombo, Georgi Duev, Michèle B. Nuijten, and Jan Sprenger. Statistical reporting inconsistencies in experimental philosophy. *PLOS ONE*, 13(4) :e0194360, April 2018.

[53] Keith Connellan. Reference and Definite Descriptions. *The Philosophical Review*, pages 281–304, 1966.

[54] Hugh Coolican. *Research Methods and Statistics in Psychology*. Psychology Press, Taylor & Francis Group, London ; New York, sixth edition edition, 2014.

[55] Florian Cova. Philosophie Experimentale. *Klesis*, 27(27) :310, 2013.

[56] Florian Cova. Intentional action and the frame-of-mind argument : New experimental challenges to Hindriks. *Philosophical Explorations*, 20(1) :35–53, January 2017.

[57] Florian Cova, Julien Dutant, Edouard Machery, Joshua Knobe, Shaun Nichols, and Eddy Nahmias. *La philosophie expérimentale*. Vuibert, Paris, 2012.

[58] Florian Cova, Brent Strickland, Angela Abatista, Aurélien Allard, James Andow, Mario Attie, James Beebe, Renatas Berniūnas, Jordane Boudesseul, Matteo Colombo, Fiery Cushman, Rodrigo Diaz, Noah N'Djaye Nikolai van Dongen, Vilius Dranseika, Brian D. Earp, Antonio Gaitán Torres, Ivar Hannikainen, José V. Hernández-Conde, Wenjia Hu, François Jaquet, Kareem Khalifa, Hanna Kim, Markus Kneer, Joshua Knobe, Miklos Kurthy, Anthony Lantian, Shen-yi Liao, Edouard Machery, Tania Moerenhout, Christian Mott, Mark Phelan, Jonathan Phillips, Navin Rambharose, Kevin Reuter, Felipe Romero, Paulo Sousa, Jan Sprenger, Emile Thalabard, Kevin Tobia, Hugo Viciana, Daniel Wilkenfeld, and Xiang Zhou. Estimating the Reproducibility of Experimental Philosophy. *Review of Philosophy and Psychology*, June 2018.

[59] Simon Cullen. Survey-Driven Romanticism. *Review of Philosophy and Psychology*, 1(2) :275–296, June 2010.

[60] Oliver Scott Curry, Daniel Austin Mullins, and Harvey Whitehouse. Is It Good to Cooperate ? : Testing the Theory of Morality-as-Cooperation in 60 Societies. *Current Anthropology*, 60(1) :47–69, February 2019.

[61] Antonio R Damasio and Marcel Blanc. *L'erreur de Descartes : la raison des émotions*. Sciences. Odile Jacob, odile jacob edition, 2010.

[62] Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso. « Qu'a mangé le juge à son petit-déjeuner ? » De l'impact des conditions de travail sur la décision de justice. *Les Cahiers de la Justice*, 4(4) :579–587, 2015.

[63] Lorraine Daston, Peter Galison, and Bruno Latour. *Objectivité*. Fabula. Les Presses du réel, Dijon, 2012.

[64] Richard Dawkins. *Pour en finir avec Dieu*. Perrin, Paris, 2018.

[65] Julian De Freitas and George A. Alvarez. Your visual system provides all the information you need to make moral judgments about generic visual events. *Cognition*, 178 :133–146, September 2018.

[66] Julian De Freitas and Samuel G.B. Johnson. Optimality bias in moral judgment. *Journal of Experimental Social Psychology*, 79 :149–163, November 2018.

[67] Frans de Waal. *Sommes-nous trop bêtes pour comprendre l'intelligence des animaux ?* Les Liens qui libèrent, 2016.

[68] Stéphane Debove. *The Evolutionary Origins of Human Fairness*. PhD thesis, Paris Descartes, 2015.

[69] Stanislas Dehaene. *La bosse des maths : quinze ans après*. O. Jacob, Paris, nouv. édition revue et augmentée edition, 2010.

[70] Daniel C. Dennett. *From Bacteria to Bach and Back : The Evolution of Minds*. Penguin Books, London, 2018.

[71] Ophelie Desmons, Stéphane Lemaire, and Patrick Turmel. *Manuel de metaethique.* Hermann, hermann l'avocat du diable edition, 2019.

[72] Max Deutsch. *The Myth of the Intuitive : Experimental Philosophy and Philosophical Method*. The MIT Press, a Bradford Book, Cambridge, Massachusetts, 2015.

[73] Robert Deutschländer, Michael Pauen, and John-Dylan Haynes. Probing folk-psychology : Do Libet-style experiments reflect folk intuitions about free action? *Consciousness and Cognition*, 48 :232–245, February 2017.

[74] John Dewey and Patrick Savidan. *La quête de certitude : une étude de la relation entre connaissance et action*. Gallimard, Paris, 2014.

[75] Sharmistha Dhar. The Ontology of Intentional Agency in Light of Neurobiological Determinism : Philosophy Meets Folk Psychology. *Journal of Indian Council of Philosophical Research*, 34(1) :129–149, January 2017.

[76] John M. Doris, editor. *The Moral Psychology Handbook*. Oxford Univ. Press, Oxford, 1. pbk. publ edition, 2012.

[77] John M. Doris, Stephen Stich, Jonathan Phillips, and Lachlan Walmsley. Moral Psychology : Empirical Approaches. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition, 2017.

[78] Aaron A. Duke and Laurent Bègue. The drunk utilitarian : Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition*, 134 :121–127, January 2015.

[79] Rawad El Skaf. *La Structure Des Expériences de Pensée Scientifiques*. PhD thesis, Paris 1, 2016.

[80] David Enoch. *Taking Morality Seriously : A Defense of Robust Realism*. Oxford Univ. Press, Oxford, 1. publ. in paperback edition, 2013.

[81] Kathinka Evers. *Neuroéthique : quand la matière s'éveille*. Collège de France. O. Jacob, Paris, 2009.

[82] Uljana Feest. Operationism in psychology : What the debate is about, what the debate should be about. *Journal of the History of the Behavioral Sciences*, 41(2) :131–149, 2005.

[83] Gilad Feldman, Kin Fai Ellick Wong, and Roy F. Baumeister. Bad is freer than good : Positive–negative asymmetry in attributions of free will. *Consciousness and Cognition*, 42 :26–40, May 2016.

[84] Silvia Felletti and Fabio Paglieri. The illusionist and the folk : On the role of conscious planning in intentionality judgments. *Philosophical Psychology*, 29(6) :871–888, August 2016.

[85] Paul Feyerabend, Baudouin Jurdant, and Agnès Schlumberger. *Contre la méthode . esquisse d'une théorie anarchiste de la connaissance*. Seuil, Paris, 1988.

[86] Klaus Fiedler, Claude Messner, and Matthias Bluemke. Unresolved problems with the "I", the "A", and the "T" : A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology*, 17(1) :74–147, January 2006.

[87] Marcelo Fischborn. Questions for a Science of Moral Responsibility. *Review of Philosophy and Psychology*, 9(2) :381–394, June 2018.

[88] Philippa Foot. The Problem of Abortion and the Doctrine of the Double Effect. In *Virtues and Vices : And Other Essays in Moral Philosophy*, pages 19–32. Oxford :Basil-Blackwell, Oxford, 1978.

[89] Denis Forest. *Neuroscepticisme : les sciences du cerveau sous le scalpel de l'épistémologue*. Ithaque, Montreuil-sous-Bois, 2015.

[90] Paul Franceschi. *Introduction à la philosophie analytique : paradoxes, arguments et problèmes contemporains*. Le Manuscrit, Paris, 2005.

[91] Allan D. Franklin. What Makes a 'Good' Experiment? *The British Journal for the Philosophy of Science*, 32(4) :367–374, 1981.

[92] Allan D. Franklin and Slobodan Perovic. Experiment in Physics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

[93] Jared P. Friedman and Anthony I. Jack. Mapping Cognitive Structure onto the Landscape of Philosophical Debate : An Empirical Framework with Relevance to Problems of Consciousness, Free will and Ethics. *Review of Philosophy and Psychology*, 9(1) :73–113, March 2018.

[94] Peter Galison. *How Experiments End*. University of Chicago Press, Chicago, 1987.

[95] Jean-Pierre Gasc. Arts, Sciences, Religions et le surdimensionnement du cerveau humain. *Arts et sciences*, 3(1), 2019.

[96] Harry J. Gensler. *Ethics and the Golden Rule*. Routledge, New York, 2013.

[97] Tobias Gerstenberg, Tomer D. Ullman, Jonas Nagel, Max Kleiman-Weiner, David A. Lagnado, and Joshua B. Tenenbaum. Lucky or clever ? From expectations to responsibility judgments. *Cognition*, 177 :122–141, August 2018.

[98] Bernard Gert and Joshua Gert. The Definition of Morality. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2017 edition, 2017.

[99] Shani Getstein, Yaara Yeshurun, Liron Rozenkrantz, Sagit Shushan, Idan Frumin, Yehudah Roth, and Noam Sobel. Human Tears Contain a Chemosignal (English). *Science (Wash. D.C.)*, 331(6014) :226–230, cover date : 2011.

[100] Edmund Gettier. Is Justified True Belief Knowledge ? *Analysis, 23*, pages 121–123, 1963.

[101] Rodolphe Ghiglione. *Cours de psychologie Bases, Méthodes et épistémologie*. Dunod, Paris, 2007.

[102] Carol Gilligan. *In A Different Voice : Psychological Theory and Women's Development*, volume 326. Harvard University Press, harvard university press edition, January 1982.

[103] Yves Gingras. *L'impossible Dialogue : Sciences et Religions*. Boréal, Montréal (Québec), 2016.

[104] Ezequiel Gleichgerrcht, Teresa Torralva, Alexia Rattazzi, Victoria Marenco, María Roca, and Facundo Manes. Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *Social Cognitive and Affective Neuroscience*, 8(7) :780–788, October 2013.

[105] Stephen Jay Gould and Marcel Blanc. *La structure de la théorie de l'évolution*. Gallimard nrf, 2006.

[106] Asmir Gračanin, Lauren M. Bylsma, and Ad J. J. M. Vingerhoets. Why Only Humans Shed Emotional Tears : Evolutionary and Cultural Perspectives. *Human Nature*, 29(2) :104–133, June 2018.

[107] Madeleine Grawitz. *Méthodes Des Sciences Sociales*. Précis. Droit Public, Science Politique. Dalloz, Paris, 11e éd edition, 2001.

[108] Joshua D. Greene. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293(5537) :2105–2108, September 2001.

[109] Joshua D. Greene. *Moral Tribes : Emotion, Reason, and the Gap between Us and Them*. Atlantic Books, London, paperback edition edition, 2015.

[110] Anthony. G. Greenwald and M. R. Banaji. Implicit social cognition : Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1) :4–27, January 1995.

[111] Anthony G. Greenwald and Mahzarin R. Banaji. The implicit revolution : Reconceiving the relation between conscious and unconscious. *American Psychologist*, 72(9) :861–871, December 2017.

[112] Anthony G. Greenwald, Mahzarin R. Banaji, and Brian A. Nosek. Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, 108(4) :553–561, 2015.

[113] Anthony G. Greenwald, D. E. McGhee, and J. L. Schwartz. Measuring individual differences in implicit cognition : The implicit association test. *Journal of Personality and Social Psychology*, 74(6) :1464–1480, June 1998.

[114] Anthony G. Greenwald, T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji. Understanding and using the Implicit Association Test : III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1) :17–41, 2009.

[115] Ian Hacking. *Representing and Intervening : Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, Cambridge [Cambridgeshire] ; New York, 1983.

[116] Jonathan Haidt and Vintage Books (Nowy Jork). *The Righteous Mind : Why Good People Are Divided by Politics and Religion*. Vintage Books, a division of Random House, New York, 2013.

[117] Denis J. Hilton, John McClure, and Briar Moir. Acting knowingly : Effects of the agent's awareness of an opportunity on causal attributions. *Thinking & Reasoning*, 22(4) :461–494, October 2016.

[118] Frank Hindriks, Igor Douven, and Henrik Singmann. A New Angle on the Knobe Effect : Intentionality Correlates with Blame, not with Praise : A New Angle on the Knobe Effect. *Mind & Language*, 31(2) :204–220, April 2016.

[119] Joachim Horvath. How (not) to react to experimental philosophy. *Philosophical Psychology*, 23(4) :447–480, August 2010.

[120] Olivier Houdé, Bernard Mazoyer, and Nathalie Tzourio-Mazoyer. *Cerveau et psychologie : introduction à l'imagerie cérébrale anatomique et fonctionnelle*. PUF, Presses universitaires de France, Paris, 2010.

[121] Michael Huemer. *Ethical Intuitionism*. Palgrave Macmillan, Houndmills, 2008.

[122] John P. A. Ioannidis. Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8) :e124, August 2005.

[123] Laurent Jaffro, editor. *Le Sens Moral : Une Histoire de La Philosophie Morale de Locke à Kant*. Débats Philosophiques. Presses universitaires de France, Paris, 1re éd edition, 2000.

[124] Julie Jebeile. *Épistémologie des modèles & des simulations numériques : de la repésentation à la compréhension scientifique*. CNRS Editions, Paris, 2019.

[125] Arthur Robert Jensen. *Clocking the Mind : Mental Chronometry and Individual Differences*. Elsevier, Amsterdam ; Boston ; London, 1st ed edition, 2006.

[126] Guy Kahane. Sidetracked by trolleys : Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience*, 10(5) :551–560, September 2015.

[127] Daniel Kahneman, Paul Slovic, and Amos Tversky, editors. *Judgment under Uncertainty : Heuristics and Biases*. Cambridge University Press, Cambridge ; New York, 1982.

[128] François Kammerer. *Conscience et matière. Une solution matérialiste au problème de l'expérience consciente.* Matériologiques Editions, Paris, 2019.

[129] Immanuel Kant and Roger Kempf. *Observations sur le sentiment du beau et du sublime*. Bibliothèque des textes philosophiques. Vrin, vrin edition, 2008.

[130] Kai Kaspar, Albert Newen, Thomas Dratsch, Leon de Bruin, Ahmad Al-Issa, and Gary Bente. Whom to blame and whom to praise : Two cross-cultural studies on the appraisal of positive and negative side effects of company activities. *International Journal of Cross Cultural Management*, 16(3) :341–365, December 2016.

[131] Antti Kauppinen. THE RISE AND FALL OF EXPERIMENTAL PHILOSOPHY. *Philosophical Explorations*, 10(2) :95–118, June 2007.

[132] Mansur Khamitov, Jeff D. Rotman, and Jared Piazza. Perceiving the agency of harmful agents : A test of dehumanization versus moral typecasting accounts. *Cognition*, 146 :33–47, January 2016.

[133] Nancy S. Kim, Samuel G. B. Johnson, Woo-kyoung Ahn, and Joshua Knobe. The effect of abstract versus concrete framing on judgments of biological and psychological bases of behavior. *Cognitive Research : Principles and Implications*, 2(1), December 2017.

[134] Philip Kitcher. *The Ethical Project*. Harvard Univ. Press, Cambridge, Mass., first harvard univ. press paperback ed edition, 2014.

[135] Philip Kitcher and Stephanie Ruphy. *Science, vérité et démocratie*. Presses universitaires de France, Paris, 2010.

[136] KNAW. *Replication Studies - Improving Reproducibility in the Empirical Sciences*. Royal Netherland Academy of Science, Amsterdam, 2018.

[137] Markus Kneer and Sacha Bourgeois-Gironde. Mens rea ascription, expertise and outcome effects : Professional judges surveyed. *Cognition*, 169 :139–146, December 2017.

[138] Joshua Knobe. Intentional action and side effects in ordinary language. *Analysis*, 63(3) :190–194, July 2003.

[139] Joshua Knobe. The Concept of Intentional Action : A Case Study in the Uses of Folk Psychology. *Philosophical Studies*, 130(2) :203–231, August 2006.

[140] Joshua Knobe and Shaun Nichols, editors. *Experimental Philosophy*. Oxford University Press, Oxford ; New York, 2008.

[141] Joshua Knobe and Jesse Prinz. Intuitions about consciousness : Experimental studies. *Phenomenology and the Cognitive Sciences*, 7(1) :67–83, March 2008.

[142] Saul A. Kripke. *Naming and Necessity*. Blackwell Publishers, Oxford, UK ; Cambridge, USA, 1998.

[143] Hubert Krivine and J. C. Ameisen. *Comprendre sans Prévoir, Prévoir sans Comprendre*. Cassini, Paris, 2018.

[144] Petr Alekseevič Kropotkin. *La morale anarchiste*. Mille et un nuits, Paris, 2004.

[145] Charles Larmore. *Modernité et morale*. Philosophie morale. Presses Univ. de France, Paris, 1. éd edition, 1993.

[146] Benjamin Libet. *Mind Time : The Temporal Factor in Consciousness*. Perspectives in Cognitive Neuroscience. Harvard Univ. Press, Cambridge, Mass., harvard edition, 2005.

[147] Daniel Lim and Ju Chen. Is Compatibilism Intuitive ? *Philosophical Psychology*, 31(6) :878–897, July 2018.

[148] Pascal Ludwig and Matthias Michel. Les données en première personne et l'expérimentation en psychologie. *Philosophia Scientae*, 2019(23-2) :111–130, May 2019.

[149] Christoph Lütge. *Experimental Ethics : Toward an Empirical Moral Philosophy*. Palgrave Macmillan, New York, NY, 2014.

[150] Edouard Machery. Semantics, cross-cultural style. *Cognition*, 92(3) :B1–B12, July 2004.

[151] Edouard Machery. *Philosophy within Its Proper Bounds*. Oxford University Press, Oxford, United Kingdom, oup edition, 2017.

[152] John Malone. Did John B. Watson Really "Found" Behaviorism ? *Behavior Analyst*, 37(1) :1, 2014/05/01/Number 1/May 2014.

[153] Gianluca Manzo. *La simulation multi-agents : principes et applications aux phénomènes sociaux*. Presses de Sciences Po, St. Germain, 2014.

[154] Jesse Marczyk and Michael J. Marks. Does it matter who pulls the switch ? Perceptions of intentions in the trolley dilemma. *Evolution and Human Behavior*, 35(4) :272–278, July 2014.

[155] Francesco Margoni and Luca Surian. Children's intention-based moral judgments of helping agents. *Cognitive Development*, 41 :46–64, January 2017.

[156] Peter Markie. Rationalism vs. Empiricism. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2017 edition, 2017.

[157] Justin W. Martin and Fiery Cushman. Why we forgive what can't be controlled. *Cognition*, 147 :133–143, February 2016.

[158] Emilio Martínez Navarro. Filosofía moral experimental : Una revisión del concepto. *Dialogo Filosofico*, 33 :2(98) :229–247, May 2017.

[159] Alberto Masala, Jérôme Ravat, and Luc Faucher. *La morale humaine et les sciences*. Éditions matériologiques, Paris, 2013.

[160] Dan Mcarthur. The Anti-philosophical Stance, the Realism Question and Scientific Practice. *Foundations of Science*, 11(4) :369–397, December 2006.

[161] Allen R. McConnell and Jill M. Leibold. Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology*, 37(5) :435–442, September 2001.

[162] Roberto Merrill and Patrick Savidan. *Du minimalisme moral. Essais pour Ruwen Ogien*, volume 22 of *Raison Publique*. Raison Publique, Paris, 2017.

[163] Joel Michell. Alfred Binet and the concept of heterogeneous orders. *Frontiers in Psychology*, 3, 2012.

[164] Stanley Milgram. Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4) :371–378, 1963.

[165] George E. Moore, Michel Gouverneur, and Ruwen Ogien. *Principia Ethica*. Presses universitaires de France, Paris, 1998.

[166] Masahiro Mori. Bukimi no tani gensho « La vallée de l'étrange ». *Energy*, pages 33–35, 1970.

[167] Nikil Mukerji. *Experimental Philosophy : A Critical Study*. Rowman & Littlefield International, Ltd, London ; New York, 2019.

[168] Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. A manifesto for reproducible science. *Nature Human Behaviour*, 1 :0021, January 2017.

[169] Dylan Murray and Tania Lombrozo. Effects of Manipulation on Attributions of Causation, Free Will, and Moral Responsibility. *Cognitive Science*, 41(2) :447–481, March 2017.

[170] Eddy Nahmias, Stephen G. Morris, Thomas Nadelhoffer, and Jason Turner. Is Incompatibilism Intuitive ? *Philosophy and Phenomenological Research*, 73(1) :28–53, July 2006.

[171] Serge Nicolas. Benjamin Bourdon (1860-1943) : fondateur du laboratoire de psychologie et de linguistique expérimentales à l'Université de Rennes (1896). *L'année psychologique*, 98(2) :271–293, 1998.

[172] Richard E. Nisbett. *The Geography of Thought : How Asians and Westerners Think Differently ... and Why*. Free Press, New York, nachdr. edition, 2004.

[173] Vanessa Nurock. *Sommes-Nous Naturellement Moraux ?* Fondements de La Politique. Série Essais. Presses universitaires de France, Paris, 2011.

[174] Ruwen Ogien. *L'influence de l'odeur Des Croissants Chauds Sur La Bonté Humaine : Et Autres Questions de Philosophie Morale Expérimentale*. Bernard Grasset, Paris, 2011.

[175] Elizabeth O'Neill and Edouard Machery, editors. *Current Controversies in Experimental Philosophy*. Current Controversies in Philosophy. Routledge, New York, 2014.

[176] Frederick L. Oswald, Gregory Mitchell, Hart Blanton, James Jaccard, and Philip E. Tetlock. Predicting ethnic and racial discrimination : A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2) :171–192, 2013.

[177] Isabelle Pariente-Butterlin. L'Émergence de la question éthique et sa présence chez Levinas. In *L'adresse et l'argument, Levinas et La Philosophie Analytique.*, Paris, April 2015.

[178] Mary Parkinson and Ruth M. J. Byrne. Judgments of Moral Responsibility and Wrongness for Intentional and Accidental Harm and Purity Violations. *Quarterly Journal of Experimental Psychology*, page 17470218.2016.1, January 2017.

[179] David L. Parris, Alan Sokal, and Jean Bricmont. Impostures intellectuelles. *The Modern Language Review*, 94(2) :603, April 1999.

[180] Elliot Samuel Paul and Scott Barry Kaufman, editors. *The Philosophy of Creativity : New Essays*. Oxford University Press, New York, 2017.

[181] Gualtiero Piccinini. First Person Data, Publicity and Self-Measurement. *Philosophers'Imprint*, 9(9) :1–16, October 2009.

[182] Andrew Pickering, editor. *Science as Practice and Culture*. University of Chicago Press, Chicago, 1992.

[183] Jason E. Plaks, Jennifer L. Fortune, Lindie H. Liang, and Jeffrey S. Robinson. Effects of Culture and Gender on Judgments of Intent and Responsibility. *PLOS ONE*, 11(4) :e0154467, April 2016.

[184] Karl Raimund Popper, Nicole Thyssen-Rutten, Philippe Devaux, Jacques Monod, and Karl Raimund Popper. *La logique de la découverte scientifique*. Payot, 2014.

[185] Jesse J. Prinz. *The Emotional Construction of Morals*. Oxford Univ. Press, Oxford, 2009.

[186] Michel Puech. *The Ethics of Ordinary Technology*. Number 33 in Routledge Studies in Science, Technology and Society. Routledge, Taylor & Francis Group, New York London, 2016.

[187] Hilary Putnam. *Reason, Truth, and History*. Cambridge University Press, Cambridge ; New York, 1981.

[188] Peter Railton. Maximal Minimalism : Ruwen Ogien and Self-Affecting Actions. *Raison publique*, 22(2) :81, 2017.

[189] Jérôme Ravat. *Éthique et polémiques : les désaccords moraux dans la sphère publique*. CNRS Editions, 2019.

[190] John Rawls. Outline of a Decision Procedure for Ethics. *Philosophical Review*, 60(2) :177–197, 1951.

[191] John Rawls. *Théorie de la justice*. Ed. du Seuil, Paris, 2009.

[192] Liz Redford and Kate A. Ratliff. Perceived moral responsibility for attitude-based discrimination. *British Journal of Social Psychology*, 55(2) :279–296, June 2016.

[193] Marion Renauld. *Philosophie de la fiction : vers une approche pragmatiste du roman*. Collection Æsthetica. Presses Univ. de Rennes, Rennes, 2014.

[194] Erin Robbins, Jason Shepard, and Philippe Rochat. Variations in judgments of intentional action and moral evaluation across eight cultures. *Cognition*, 164 :22–30, July 2017.

[195] David Rose. Folk intuitions of actual causation : A two-pronged debunking explanation. *Philosophical Studies*, 174(5) :1323–1361, May 2017.

[196] Adina Roskies. Neuroethics. *The Stanford Encyclopedia of Philosophy*, Spring 2016 Edition, 2016.

[197] Jana Samland and Michael R. Waldmann. How prescriptive norms influence causal inferences. *Cognition*, 156 :164–176, November 2016.

[198] Anders Sandberg, Heather Bradshaw-Martin, and Mona Gérardin-Laverge. La voiture autonome et ses implications morales. *Multitudes*, 58(1) :62, 2015.

[199] Michael Sandel. What's Wrong with Enhancement. *President's Council on Bioethics, Washington, Dc (Www. Bioethics. Gov)*, 12, 2002.

[200] Hagop (ed) Sarkissian and Jennifer Cole (ed) Wright. *Advances in Experimental Moral Psychology*. Advances in Experimental Philosophy. Bloomsbury Academic, New York, January 2014.

[201] Hanno Sauer. *Debunking Arguments in Ethics*. Cambridge University Press, Cambridge, United Kingdom ; New York, 2018.

[202] Geoff Sayre-McCord. Moral Realism. In *Stanford Encyclopedia of Philosophy*. Edward N Zalta, edward n zalta edition, 2017.

[203] John R. Searle. How to Derive "Ought" From "Is". *The Philosophical Review*, 73(1) :43, January 1964.

[204] Yves Serra. La philosophie morale expérimentale est-elle expérimentale ? *Philosophia Scientiæ. Travaux d'histoire et de philosophie des sciences*, 23(23-2) :149–171, May 2019.

[205] Anne Sheeran and Philip Lucas. Asperger's Syndrome and the Eccentricity and Genius of Jeremy Bentham. *UCL Bentham Project*, Vol 8 2006, 2006.

[206] Didier Sicard. *L'éthique médicale et la bioéthique*. Presses universitaires de France, Paris, 2011.

[207] Jenifer Z. Siegel, Molly J. Crockett, and Raymond J. Dolan. Inferences about moral character moderate the impact of consequences on blame and praise. *Cognition*, 167 :201–211, October 2017.

[208] Marc Silberstein. *Qu'est-ce que la science... pour vous ? Tome 1*. Matériologiques, 2017.

[209] Marc Silberstein. *Qu'est-ce que la science... pour vous ? Tome 2*. Matériologiques, 2018.

[210] Peter Singer. Famine, Affluence, and Morality. *Philosophy and public affairs*, 1(3) :229–243, 1972.

[211] Burrhus F. Skinner. *Science et comportement humain*. In Press, Paris, 2008.

[212] Noam Sobel. Revisiting the revisit : Added evidence for a social chemosignal in human emotional tears. http ://www.tandfonline.com/doi/pdf/10.1080/02699931.2016.1177488, 2017.

[213] Pascal Sockeel and Françoise Anceaux. *La démarche expérimentale en psychologie*. In Press, Paris, 2008.

[214] Ernest Sosa. The Raft and the Pyramid : Coherence versus Foundations in the Theory of Knowledge. *Midwest Studies In Philosophy*, 5(1) :3–26, September 1980.

[215] Ernest Sosa. Experimental philosophy and philosophical intuition. *Philosophical Studies*, 132(1) :99–107, 2007.

[216] Marta Spranzi. *Le Travail de l'éthique : Décision Clinique et Intuitions Morales*. Psy - Théories, Débats, Synthèse. Pierre Mardaga Editeur, Sprimont, 2018.

[217] Stephen Jay Gould and Elisabeth S. Vrba. Exaptation-A Missing Term in the Science of Form. *Paleobiology*, 8(1) :4, 1982.

[218] Stanley Smith Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684) :677–680, 1946.

[219] Brent Strickland and Aysu Suben. Experimenter Philosophy : The Problem of Experimenter Bias in Experimental Philosophy. *Review of Philosophy and Psychology*, 3(3) :457–467, September 2012.

[220] Michael T Stuart, David Colaço, and Edouard Machery. P-curving x-phi : Does experimental philosophy have evidential value ? *Analysis*, 79(4) :669–684, October 2019.

[221] Jussi Suikkanen and Antti Kauppinen. *Methodology and Moral Philosophy*. Routledge, October 2018.

[222] Stacey Swain, Joshua Alexander, and Jonathan M. Weinberg. The Instability of Philosophical Intuitions : Running Hot and Cold on Truetemp : THE INSTABILITY OF PHILOSOPHICAL INTUITIONS. *Philosophy and Phenomenological Research*, 76(1) :138–155, January 2008.

[223] Justin Sytsma and Wesley Buckwalter. *A Companion to Experimental Philosophy*. John Wiley & Sons, March 2016.

[224] Justin Sytsma and Jonathan Livengood. *The Theory and Practice of Experimental Philosophy*. Broadview Press, Peterborough, Ontario, 2016.

[225] Christine Tappolet. *Emotions et Valeurs*. Philosophie Morale. Presses universitaires de France, Paris, 1re éd edition, 2000.

[226] Judith Jarvis Thomson. Killing, Letting Die, and the Trolley Problem. *The Monist*, 59 :204–217, 1976.

[227] Claudine Tiercelin. *Le Doute En Question : Parades Pragmatistes Au Défi Sceptique*. Tiré à Part. Eclat, Paris, 2005.

[228] Mark Timmons. *Moral Theory : An Introduction*. Elements of Philosophy. Rowman & Littlefield Publishers, Lanham, Md, 2nd ed edition, 2013.

[229] Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Conn, 2nd ed edition, 2001.

[230] John Turri, Ori Friedman, and Ashley Keefner. Knowledge Central : A Central Role for Knowledge Attributions in Social Evaluations. *Quarterly Journal of Experimental Psychology*, 70(3) :504–515, March 2017.

[231] Bas C. Van Fraassen. *Scientific Representation : Paradoxes of Perspective*. Oxford University Press, Oxford ; New York, 2010.

[232] Adrianus. J. J. M. Vingerhoets. *Why Only Humans Weep : Unravelling the Mysteries of Tears*. Oxford University Press, Oxford, 2013.

[233] John B. Watson. Psychology as the behaviorist views it. *Psychological Review, 20,*, pages 158–177, 1913.

[234] Jonathan M. Weinberg, Shaun Nichols, Stephen Stich, and University of Arkansas Press. Normativity and Epistemic Intuitions :. *Philosophical Topics*, 29(1) :429–460, 2001.

[235] Timothy Williamson. *The Philosophy of Philosophy*. Number 2 in The Blackwell/Brown Lectures in Philosophy. Blackwell Pub, Malden, MA, 2007.

[236] James Q. Wilson. *The Moral Sense*. The Free Press, New-York Toronto, 1993.

[237] Francis Wolff. *Trois Utopies Contemporaines*. Fayard, Paris, 2017.

[238] Guillaume Wunsch. Confounding and control. *Demographic Research*, 16 :97–120, February 2007.

Philosophie morale et démarche expérimentale, une approche critique

Résumé

Entre éthique et épistémologie, la thèse vise à éclairer les conditions de l'apport des sciences expérimentales à la réflexion morale. Les morales sont soumises aujourd'hui à deux pressions, l'une liée au développement des sciences cognitives et l'autre à la globalisation des interactions humaines. Elles ont à répondre à ces pressions sans nuire à leur rôle de régulation locale des groupes sociaux. Les philosophes expérimentaux ont cherché à mettre les sciences cognitives au service des questions philosophiques, et la thèse s'appuie sur cette expérience dont les résultats, ni insignifiants ni déterminants, ont soulevé de nombreuses réticences de la part des philosophes moraux. La thèse comporte trois chapitres de description du domaine moral et expérimental, puis un chapitre central avec 5 cas d'utilisation de résultats expérimentaux dans des débats philosophiques. Est ensuite analysé le rôle clé de l'opérationnalisation dans ces débats : comment savoir qu'une expérience est pertinente en regard d'une entité psychologique ? Les réticences des philosophes moraux et les réponses des scientifiques sont ensuite décrites. De ces travaux, la thèse tire des leçons méthodologiques et défend trois propositions. 1- L'opérationnalisation des entités est un point nodal de la psychologie expérimentale. 2- La théorie de l'évolution suggère des pistes fécondes, en distinguant trois échelles de temps, la spéciation, les groupes sociaux et l'individuation. 3- Rien ne peut être dit de la philosophie morale en général en regard de l'expérimentation, mais des analyses utiles peuvent être proposées selon les 4 perspectives descriptives, prescriptives, méta-éthique et éthiques appliquées.

Mots-clés : philosophie morale ; philosophie expérimentale ; opérationnalisation ; épistémologie ; expérimentation ; méta-éthique ; XPHI

Moral philosophy and experimental approach, a critical study

Summary

Between ethics and epistemology, the thesis aims to clarify the contribution of experimental sciences to moral philosophy. Moral systems are today subject to two pressures, one due to the development of cognitive sciences and the other to the globalization of human interactions. They have to respond to these pressures without harming their role in regulating local social groups. Experimental philosophers have sought to use cognitive sciences to discuss philosophical questions. The thesis is based on this XPhi experience, the results of which, neither insignificant nor decisive, have raised many reservations from moral philosophers. The thesis has three chapters describing the moral and experimental domains, then a central chapter with five case studies using experimental results in philosophical debates. The key role of operationalization in these debates is then analyzed : how to know that an experience is relevant to study a psychological entity ? The reluctance of moral philosophers and the responses of scientists are finally taken up. From this work, the thesis draws methodological lessons and defends three proposals. 1- The operationalization of entities is a nodal point of experimental psychology. 2- The theory of evolution suggests fruitful paths, distinguishing three time scales, speciation, social groups and individuation. 3- Little can be said of moral philosophy in general with regard to experimentation, but useful analyzes can be offered from four perspectives : descriptive, prescriptive, meta-ethical and applied ethics.

Keywords : Moral philosophy ; experimental philosophy ; XPhi ; operationalism ; epistemology ; experimentation ; metaethics

UNIVERSITÉ SORBONNE UNIVERSITÉ

ÉCOLE DOCTORALE :

ED 5 – Concepts et langages

Maison de la Recherche, 28 rue Serpente, 75006 Paris, France

DISCIPLINE : Philosophie