

UNIVERSITÉ DE MONS & UNIVERSITÉ DE NAMUR

THÈSE DE DOCTORAT

**Vers un modèle unitaire de la
scientificité**

Auteur :

Jeremy ATTARD

Promoteurs :

Anne STAQUET

Olivier SARTENAER

Dominique LAMBERT

soumise au jury

pour l'obtention du grade de Docteur en Philosophie

Date de la défense privée :

10 juin 2024

Membres du jury :

Alban BOUVIER

Damien DARCIS

Bertrand HESPEL

Gianluca MANZO

Claude SEMAY

On a dû te dire qu'il fallait réussir dans la vie; moi je te dis qu'il faut vivre, c'est la plus grande réussite du monde. On t'a dit : "Avec ce que tu sais, tu gagneras de l'argent". Moi je te dis : "Avec ce que tu sais tu gagneras des joies". C'est beaucoup mieux. Tout le monde se rue sur l'argent. Il n'y a plus de place au tas des batailleurs. De temps en temps un d'eux sort de la mêlée, blême, titubant, sentant déjà le cadavre, le regard pareil à la froide clarté de la lune, les mains pleines d'or mais n'ayant plus force et qualité pour vivre; et la vie le rejette. Du côté des joies, nul ne se presse; elles sont libres dans le monde, seules à mener leurs jeux féériques sur l'asphodèle et le serpolet des clairières solitaires. Ne crois pas que l'habitant des hautes terres y soit insensible. Il les connaît, les saisit parfois, danse avec elles. Mais la vérité est que certaines de ces joies plus tendres que les brumes du matin te sont réservées à toi, en plus des autres. Elles veulent un esprit plus averti, des grâces de pensées qui te sont coutumières. Tu es là à te désespérer quand tu es le mieux armé de tous, quand tu as non seulement la science mais encore la jeunesse qui la corrige.

Jean Giono, *Les vrais richesses* (1936)

Table des matières

Remerciements

1	Introduction	1
2	Le problème de la démarcation scientifique	17
2.1	Introduction	17
2.2	Le(s) falsificationnisme(s)	21
2.3	Les programmes de recherche de Lakatos	28
2.4	Quelques constats philosophiques et épistémologiques	30
2.4.1	La charge théorique des observations	30
2.4.2	Le holisme de la confirmation	33
2.4.3	La sous-détermination	34
2.4.4	La non distinction absolue entre énoncés analytiques et synthétiques	38
2.5	Quelques exemples historiques d’ajustements théoriques	39
2.5.1	Un ajustement fécond : la découverte de Neptune	40
2.5.2	Une fausse piste : la planète Vulcain	42
2.5.3	Un cas contemporain non résolu : la matière noire	44
2.6	L’approche socio-historique et ses ramifications contemporaines	47
2.6.1	La scientificité selon Kuhn	47
2.6.2	Les <i>sciences studies</i>	51
2.7	Les champs épistémiques et les approches multi-critères	57
2.8	La théorie meehilienne	59
2.9	La scientificité comme optimisation	61
2.9.1	Le <i>Best System Account</i> des lois	61
2.9.2	Le critère d’information d’Akaike	63
2.9.3	L’utilité épistémique et la <i>concept utility</i>	64
2.10	Conclusion	66
3	Les sciences sociales et leur(s) épistémologie(s)	69
3.1	Introduction	69
3.2	Une pluralité problématique?	75

3.2.1	Approches qualitatives et quantitatives	76
3.2.2	Différents types d'explications	79
	Le constat empirique	80
	L'explication théorique	82
3.2.3	L'utilisation d'outils formels et son interprétation	86
3.2.4	Conclusion	90
3.3	Les arguments pluralistes	90
3.3.1	Quatre familles d'arguments	92
	I - L'historicité	92
	II - La signification	96
	III - L'intentionnalité	99
	IV - La performativité	101
3.3.2	Un <i>système</i> pluraliste	106
3.4	Les réponses classiques	108
3.4.1	Régularités, lois et invariants	109
3.4.2	Conséquences épistémologiques	120
	L'historicité	120
	La signification	121
	L'intentionnalité	122
	La performativité	124
3.5	Conclusion	126
4	Degré de corroboration informationnelle	131
4.1	Introduction	131
4.2	L'unité d'analyse : les modèles empiriques	134
4.2.1	Définition qualitative	134
	Les modèles nomologiques	134
	Les modèles causaux	137
4.2.2	Définition formalisée	139
	Les différents composants d'une unité épistémique élémentaire	139
	Interprétations des termes	140
4.3	Degré de corroboration informationnelle	143
4.3.1	L'héritage poppérien	143
4.3.2	Esquisse de mon approche	145
	Le degré de corroboration comme une quantité d'information	146
	Un principe de maximisation	146

	La charge théorique et la dépendance au langage	147
4.3.3	Les régularités empiriques endogènes	148
	Les régularités <i>E</i> -analytiques	148
	Les artefacts de modélisation	149
4.3.4	Degré d'adéquation et degré de corroboration	152
4.4	Principe de maximisation et conséquences	154
4.4.1	Principe général de scientificité	154
4.4.2	Quelques critères épistémologiques retrouvés	154
	L'adéquation empirique	155
	Opérationnalisation précise des variables de base	159
	La progression des <i>problemshifts</i> lakatosiens	159
	Parcimonie et cohérence	163
4.5	Conclusion	164
5	Épistémologie des modèles théoriques	169
5.1	Introduction	169
5.2	La sociologie analytique	171
5.2.1	Une brève histoire de la sociologie analytique	171
5.2.2	Principes de base	174
	Un premier ensemble de principes	174
	L'individualisme structural et le diagramme de Coleman	175
	Une explicitation programmatique progressive	181
	Quelques critiques basiques	183
5.2.3	La théorie du choix rationnel	185
5.3	Épistémologie des théories de l'action	189
5.3.1	Entre irréalisme et trivialité	189
5.3.2	Une première précision : de quelles hypothèses parle-t-on?	197
5.3.3	Le statut épistémologique des hypothèses : un détour par la physique	202
	Hypothèses fausses, "irréalistes" ou théoriquement inconsistentes	202
	Diverses formulations théoriques ontologiquement incompatibles	205
	Les principes généraux et les hypothèses théoriques sont difficilement testables indépendamment	206
5.4	Critère structural et classification des phénomènes	209
5.4.1	Contraintes épistémologiques structurelles	209

5.4.2	Conséquences épistémologiques	211
5.4.3	Homogénéité des phénomènes et aspects dynamiques	217
5.4.4	Retour sur les théories de l'action en sociologie analy- tique	220
5.5	Conclusion	223
6	Conclusion	227
6.1	Une proposition de modèle unitaire de la scientificité	227
6.2	Limites	233
6.3	Perspectives	233
	Bibliographie	237

Remerciements

Je tiens tout d'abord à remercier Anne Staquet d'avoir accepté, il y a de cela quatre ans maintenant, d'encadrer ce travail de thèse. Elle a fait preuve à mon égard d'une telle confiance sur le plan intellectuel ainsi que d'un tel soutien sur le plan moral comme matériel que sans elle je n'aurais probablement jamais réalisé cette thèse en philosophie des sciences. Pour tout cela, je lui dois toute ma gratitude.

Je voudrais également remercier chaleureusement Dominique Lambert et Olivier Sartenaer d'avoir accepté de co-diriger ce travail. Je leur suis reconnaissant pour leur gentillesse et pour toutes les discussions très riches que j'ai pu partager avec eux pendant ces quatre dernières années. En particulier, je souhaite réitérer mes remerciements envers Olivier pour les critiques détaillées qu'il m'a adressées à ma défense privée ainsi que pour le temps qu'il a pris avec moi ensuite pour que je puisse les intégrer du mieux possible et ainsi améliorer substantiellement mon manuscrit.

Je souhaite également remercier sincèrement Bertrand Hespel, Damien Darcis, Claude Semay, Gianluca Manzo et Alban Bouvier d'avoir accepté de faire partie de mon jury de thèse ainsi que pour leur bienveillance et leurs remarques, conseils et critiques avisés et constructifs.

J'ai contacté Gianluca il y a plus de deux ans maintenant alors que je découvrais – avec quelque émerveillement – la sociologie analytique : son histoire, ses méthodes et les questions et débats épistémologiques qu'elle charriait. Je tiens à le remercier tout particulièrement pour le temps qu'il a pris, alors que l'on ne se connaissait pas, pour répondre à mes questions et m'envoyer moult références dont la lecture a été d'une richesse inestimable pour mes réflexions.

Je le remercie également de m'avoir mis en contact avec l'*Institute for Analytical Sociology* de l'université de Linköping, en Suède, où j'ai eu la chance de pouvoir passer deux mois à l'automne 2023. J'y ai renforcé mes connaissances en sociologie analytique en suivant un certain nombre de cours, de conférences et de séminaires. Je tiens à remercier les personnes qui m'ont gentiment accueilli et avec qui j'ai pu discuter là-bas : Jacob Habinek, Maria Branden, Rodrigo Martinez Pena, Martin Arvidsson, Hendrik Erz, Laura Fürsich, Maël Lecoursonnais, Jesper Lindmarker, Selcan Mutgan ainsi que Petri Ylikoski. J'ai eu la chance d'y rencontrer Gabriel Abend qui était venu au moment où j'y étais pour présenter son dernier ouvrage, et avec qui j'ai pu avoir des échanges très intéressants.

Ma gratitude va également à Dominique Raynaud avec qui j'ai également beaucoup échangé ces dernières années. Je tiens à le remercier pour tout le temps qu'il a pris pour moi, de la discussion passionnante que l'on a eu à Grenoble, en décembre 2021, à propos des types idéaux en sociologie, à tous ses conseils de lecture et ses encouragements permanents.

Si je remonte un peu plus dans le temps, toute cette histoire a probablement germé en grande partie au printemps 2016, lors d'une formation doctorale d'introduction à la pensée critique, donnée par l'éminent Denis Caroti. Les réflexions que cette formation a (ré)suscitées chez moi ont fini par me posséder à un point tel qu'après ma thèse de physique, j'ai voulu y voir plus clair et j'ai plongé tout entier dans la philosophie des sciences et l'épistémologie. Je tiens d'ailleurs à remercier Gabriella Crocco et les autres enseignants du master de philosophie des sciences pour leur accueil chaleureux et leur soutien. Mes pensées vont particulièrement à Gabriel Giovenetti que je remercie profondément de tout le temps qu'il a passé à discuter avec moi et à me guider dans ce nouveau monde qui s'ouvrait à moi. Je tiens également à remercier Thierry Masson, qui a co-dirigé ma thèse de physique, avec qui j'ai partagé beaucoup de discussions épistémologiques qui ont, en bonne partie, participé du renouveau de mon intérêt pour ces questions.

Pour revenir à Denis le vieux, je lui dois non seulement un certain nombre de connaissances, mais également une posture, une attitude face au débat et à la discussion critique, qui me pousse à vouloir lui ressembler autant que faire se peut, en termes d'exigence et d'humilité intellectuelle, et pas uniquement sur le plan capillaire. Et bien entendu, je lui dois mon entrée au Cortecs, le collectif qu'il a fondé avec d'autres amis lors d'une modeste balade dans les calanques il y a quasiment quinze ans aujourd'hui.

Les rencontres que j'ai faites au sein du Cortecs ont également beaucoup joué dans ma construction intellectuelle. La rencontre avec Raul Magni-Berton et Clara Egger, à l'automne 2019, dans leur petite maison forestière sur les hauteurs de Grenoble, a été décisive pour mes réflexions, en particulier concernant la scientificité des sciences sociales. Leur amitié reste encore aujourd'hui, et restera longtemps je l'espère, source de multiples joies et de chambardements intellectuels réguliers.

Mes pensées et remerciements vont également aux autres amitiés que j'ai pu tisser via le Cortecs : Emmanuel, Alex, Eva, Julien, Nico, Sohan, Vivien, David, Delphine, Céline et Richard. Vous avoir avec moi en ces temps troublés est un réconfort à la fois sur le plan des émotions que sur celui de la raison, et les discussions que l'on partage n'en finissent pas de me nourrir.

J'aimerais également remercier les participants aux différentes écoles d'été auto-gérées organisées par la *Basic Research Community for Physics* (BRCP), pendant lesquelles on partait une semaine ou deux se perdre dans la campagne (qu'elle soit autrichienne ou auvergnate) pour parler des fondements de la physique théorique, de science et de philosophie, et où on pouvait prononcer le mot "ontologie" au petit déjeuner sans choquer personne. Merci à Pierre de m'avoir introduit dans ce réseau, merci à Titouan pour toutes ses chansons, et merci à tous les autres pour les discussions que j'ai pu partager avec eux.

En particulier, c'est comme ça qu'un beau jour de juillet 2019 j'ai vu débarquer deux montois¹ qui venaient de traverser la France pour se perdre dans un hameau où le ciel nocturne n'a d'autres spectateurs que ceux qui osent, une bière dans une main et une guitare dans l'autre, chevaucher des montures de paille parsemées au milieu du silence. Je ne me doutais évidemment pas que si je suis en ce moment même en train d'écrire ces remerciements sous une pluie montoise de juin (vous avez bien lu) 2024, c'est principalement dû au croisement, à cet endroit précis, des géodésiques que nous suivons respectivement dans l'espace-temps pseudo-riemannien de nos vies.

Merci également aux divers amis belges et moins belges que j'ai pu me faire ces dernières années. Merci à Orion pour toutes ces discussions passionnées sur la nature de la réalité, sur ce qu'en dit la physique quantique comme sur ce que lui en font dire d'indélicats forbans. Merci à Juliette pour son accueil extraordinaire et son amitié, depuis ma première arrivée au plat pays (qui est le sien), en septembre 2021. Revenir régulièrement dans une ville et savoir que quelqu'un vous y attend, c'est cela se sentir chez soi. Merci également à Ismaël pour la balade sur le Terril de l'Héribus et nos discussions sur les défauts du monde académique, qu'on aime et qu'on déteste tout à la fois, et merci plus généralement aux physiciens de la porte d'à côté pour toutes les choses compliquées qu'ils écrivent sur leur tableau et qui ravivent chez moi des bons souvenirs.

1. L'un s'appelait Ludo... mais le nom du second m'échappe.² En tout cas je n'oublierai jamais le soutien moral dont ils ont honoré mes envolées lyriques à base de pyramides de chaises.

2. Il s'agit pourtant d'un nom tout à fait commun³ pour un individu qui, en ce qui le concerne, a plutôt tendance à s'aventurer (et nous avec, heureusement) hors du commun. Sans lui, pourtant, il faut bien l'avouer, une bonne partie de ma thèse serait restée à l'état un peu vague d'idée pas complètement inintéressante - sans compter un certain nombre d'éclats de rire que je lui dois et m'efforce de lui rembourser année après année.

3. *Des trajectoires d'un genre particulier dans un certain diagramme portent d'ailleurs son nom.*

Une pensée toute particulière va à toutes ces personnes merveilleuses rencontrées sous la grisaille francilienne il y a presque quinze ans aujourd'hui, alors qu'on sortait à peine de l'œuf, au pied du Mont Olympe lors d'une escapade en stop lorsqu'on prenait notre envol, ou un peu plus tard, installés en toute désinvolture dans le salon lumineux d'un grand appartement marseillais, au coin d'une place remplie de joie et de fureur : à Anna, Antoine, Val, Anaïs, Vincent, Alex, Raph, Arthur, Louis, John, Thibaut, Dimitri, Myriam, Louis, Katoche, Adrian, Simon, Antoine, Théo, Julien, Ket, Paco, Nicola, Hugo, Juliette, Mag, Margaux, Anny, Max, JB, Alice, Mathilde, Anabelle, Vivien, Léo et Sarah, ainsi qu'à tous les pitchounets et toutes les pitchounettes qui nous ont rejoint et qui nous rejoindront encore.

Merci à Aude pour ses immenses roues libres, à la grande maison du Pré Chabert et à tous ses habitants et habitantes pour leur accueil et les discussions passionnantes que j'ai pu partager avec elles et eux. Je garderai précieusement le souvenir de ce salon perché au dessus des nuages où philosopher semble être la seule activité raisonnable – ce que je compte revenir faire régulièrement.

Merci à mes vieux amis Arnaud, Jo, Colas, Alex, pour les encouragements qu'ils ont pu me prodiguer ces derniers mois, et plus généralement pour leur présence dans ma vie.

Merci à ma douce Susie pour son amour quotidien et son soutien indéfectible qui sont tout l'oxygène dont j'ai besoin.

Mes derniers remerciements – mais non les moindres – que je formule de tout mon être vont à mes parents, Gerald et Anne-Marie, ainsi qu'à toute ma famille, pour leur amour et leur soutien inconditionnels.

Chapitre 1

Introduction

The demarcation between science and pseudo-science is not merely a problem of armchair philosophy : it is of vital social and political relevance.

(LAKATOS, 1978, p. 1)

Physicien théoricien de formation, j'ai toujours été, depuis le début de mes études supérieures, autant attiré par les questions concernant la nature des théories physiques que par les théories physiques elles-mêmes. J'ai toujours ressenti un besoin profond de comprendre comment des concepts, des théories, des modèles, bref, des objets abstraits créés par des humains intrinsèquement limités semblaient tout de même nous apprendre quelque chose de substantiel à propos du monde, de la nature et de l'univers. De là vient sans aucun doute mon intérêt, qui n'a fait que croître ces quinze dernières années, pour la philosophie des sciences et l'épistémologie, des disciplines qui se posent précisément ce type de questions.

Pendant ma thèse de physique, mon intérêt pour ces questions s'est retrouvé décuplé lorsque j'ai suivi une formation doctorale d'introduction à la pensée critique. Le formateur, Denis Caroti, nous avait notamment fait travailler sur la question de ce qui distingue une croyance ordinaire d'une connaissance et, plus spécifiquement, une science d'une non-science ou d'une pseudo-science.

J'avais trouvé fascinant *la diversité des réponses* qu'un tel exercice pouvait susciter chez des doctorants et des doctorantes provenant de disciplines différentes. Il semblait que le terme "science" pouvait être utilisé pour qualifier des productions cognitives de natures tout à fait distinctes. Surtout, il paraissait difficile de tomber d'accord sur une définition simple, alors même que tout le monde dans la pièce se réclamait, d'une manière ou d'une autre, de la

"science". Lorsque j'entrepris d'étudier plus profondément le sujet, il m'apparut que cette pluralité de sens pouvant être donné au mot "science" était précisément ce qui caractérisait les débats sur le sujet.

C'est à peu près au même moment que je commençais à m'intéresser aux sciences sociales. Cet intérêt était porté par la volonté de mieux comprendre certains phénomènes sociaux et politiques qui m'intéressaient tout particulièrement comme l'origine de la délinquance, l'inégalité des chances devant l'école, la sociologie du milieu carcéral ou encore le processus de ségrégation économique urbaine appelé *gentrification*.

Dans toutes les lectures que j'ai alors pu faire, j'ai été frappé par la différence profonde qu'il semblait y avoir avec la "science" telle que je la connaissais – c'est-à-dire, principalement, avec la physique. Je pense que c'est un sentiment assez communément partagé par les personnes issues d'un parcours de sciences naturelles qui tentent pour la première fois de s'approprier certains contenus de sciences humaines et sociales.

J'ai tout d'abord accepté cet inconfort comme quelque chose dont il me faudrait m'accommoder. Après tout, ce n'était pas si étonnant *a priori* que des disciplines portant sur des phénomènes humains, c'est-à-dire sociaux, culturels ou politiques, produisent des théorisations différentes, dans leur forme, que celles portant sur les constituants les plus élémentaires de la matière ou sur l'évolution de l'univers dans son ensemble.

Malgré tout, des points d'incompréhension grandissaient à mesure que je tentais d'y voir plus clair. La *forme* de ces productions, par exemple, n'en finissait pas de me décontenancer. Proche de l'essai littéraire pour une bonne partie d'entre elles, elles étaient souvent, pour moi en tout cas, assez difficiles à suivre dans leur raisonnement – alors même qu'elles étaient censées être des œuvres fondamentales d'auteurs incontournables. Les concepts de base utilisés me semblaient souvent difficile à cerner, il n'était jamais clair qu'elles étaient les hypothèses testées – ou même s'il y avait des hypothèses qui étaient testées – et encore moins comment en était évaluée l'adéquation avec le réel.

Cependant, ce qui me frustrait le plus, philosophiquement parlant, ce n'était pas autant la forme de ces contenus que l'absence apparente de *critères de choix théoriques clairs*. Si deux explications sociologiques reposent sur des postulats fondamentalement différents mais reproduisent pourtant bien ce qu'il faut expliquer, comment choisir entre les deux ? Quels étaient les critères méta-théoriques guidant un tel choix ? Devait-on simplement adhérer à une "école de pensée" et s'y tenir ?

J'ai réalisé petit à petit, en me documentant sur ces questions, qu'elles avaient en réalité traversé l'histoire des sciences sociales et que les débats à leurs sujets étaient toujours vifs aujourd'hui.

Il n'est donc pas uniquement question de controverses proprement scientifiques à propos de la meilleure théorie à adopter pour aborder, expliquer ou comprendre un phénomène social donné. Il s'agit plus profondément d'une controverse *épistémologique*, c'est-à-dire portant précisément sur ce que signifie une "meilleure théorie". Il existe ainsi des tensions importantes, au sein des sciences sociales, concernant le sens à donner au terme "connaissances scientifiques".

Ces questions épistémologiques spécifiques concernant les sciences sociales s'étaient donc ajoutées aux questionnements généraux qui me préoccupaient à propos de la définition de la science et de la scientificité.

Ce sont ces questionnements, ces inconforts et ces frustrations qui m'ont poussé à entreprendre le travail présenté dans cette thèse, qui s'inscrit donc à l'intersection de deux problèmes épistémologiques majeurs.

D'une part, le problème de *la démarcation scientifique*, qui consiste à identifier ce qui distingue intrinsèquement un système (un énoncé, un modèle, une théorie, ...) scientifique d'un système non scientifique ou pseudo-scientifique. Cela ne consiste pas à élaborer ou analyser une "simple" classification des différents systèmes en science, non-science et pseudo-science, mais bien plutôt à rechercher dans le contenu même de ces systèmes ce qui les distingue intrinsèquement et ce qui les classe dans ces différentes catégories.

D'autre part, le second problème abordé dans ce travail est celui de *l'unité épistémologique des sciences*. Dans quelle mesure toutes les disciplines authentiquement scientifiques, des sciences physiques et biologiques aux sciences humaines et sociales, peuvent-elles être vues comme des instanciations d'une notion unique de la scientificité – et ce, malgré leurs différences assez évidentes au premier abord ?

En d'autres termes, il s'agit de savoir s'il est possible (ou même désirable) d'identifier une caractéristique commune aux systèmes que l'on qualifie de "scientifiques" et qui proviennent de disciplines distinctes, malgré le fait que ces dernières semblent a priori très diverses dans leurs objets, leurs méthodes empiriques, leurs constructions théoriques et leurs pratiques de recherche.

Les domaines scientifiques sont si nombreux et étendus qu'il était dès le départ hors de ma portée, dans le cadre de cette thèse, de prétendre pouvoir fournir une analyse systématique des productions scientifiques provenant de

tous les domaines existants. Je me suis donc focalisé ici sur certains champs disciplinaires, à savoir les sciences physiques et les sciences sociales (et la sociologie en particulier).

La littérature foisonnante sur le problème de la définition de la scientificité, en lien avec celui de l'unité épistémologique des sciences, fait clairement apparaître une grande diversité de réponses et de réactions au manque de consensus qui en résulte, ce que j'interprète comme une situation de *pluralité conceptuelle*. Dans une telle situation, plusieurs versions, ou instanciations, d'un concept donné coexistent sans savoir a priori ce que cette instabilité recouvre.

Concernant les concepts de *science* ou de *scientificité*, cette pluralité se déploie sur trois dimensions.

La première dimension est celle de l'*unité d'analyse*, c'est-à-dire de l'objet de la démarcation. Il s'agit en quelque sorte de définir l'échelle à laquelle on se place dans notre analyse du concept de scientificité.

Cela ne semble pas poser de problèmes conceptuels profonds de définir la scientificité différemment selon que l'on est en train de parler d'une unité purement *cognitive* comme un énoncé, un modèle ou une théorie, ou bien d'un paradigme dans son ensemble, d'une communauté scientifique ou d'un champ épistémique, qui contiennent certains aspects *sociaux* irréductibles. Ces unités étant de natures distinctes, le terme "scientifique" s'instanciera distinctement sans que cela ne pose problème.

La seconde dimension a trait à l'*objectif poursuivi* par l'unité dont on veut juger de la scientificité. Le but est-il uniquement cognitif, c'est-à-dire orienté vers la production de connaissances éventuellement désintéressées, ou bien se place-t-il au delà, dans le fait d'apporter une modification effective à l'objet étudié?

En sociologie, par exemple, il n'est pas rare de distinguer entre plusieurs buts distincts des productions sociologiques (CUIN, 2000; BOUDON, 2002; LAHIRE, 2004). Celles-ci peuvent, par exemple, adopter un objectif *expressif*, décrivant dans une forme éminemment littéraire, proche de l'*essai*, certains aspects de la vie sociale. Elles peuvent également avoir une finalité davantage *descriptive*, afin de renseigner certains acteurs (des responsables politiques, des militants, des citoyens, etc.) sur une réalité sociale particulière qui leur serait inaccessible directement de par leur position sociale, mais sans forcément avoir d'ambition proprement généralisante ni explicative. La sociologie peut aussi se revendiquer de la production de *connaissances* au même

titre que les autres disciplines, c'est-à-dire dont la validité s'autonomiserait petit à petit de ses conditions originales de production. Cette liste n'est pas exhaustive : d'autres buts sont encore possible – j'y reviendrai.

Ces distinctions apparaissent bien entendu dans *absolument toutes les disciplines*, de la physique à la biologie en passant par les neurosciences. La pluralité conceptuelle observée dans les débats sur la scientificité de certaines productions pourrait être dès lors le signe d'une divergence quant aux buts possibles, pertinents ou même désirables de ses productions.

Cette dimension de la pluralité ne pose pas non plus de problème. En effet, il n'est pas pertinent de vouloir comparer des unités d'analyse ne poursuivant pas le même objectif. Le terme "scientifique" qualifiant une mission spatiale destinée à prendre des clichés de la face cachée de la Lune et celui qualifiant la théorie de la relativité générale *ne coïncident pas*, précisément en ce que les objectifs poursuivis sont différents. L'objectif de la mission spatiale est de renseigner de manière descriptive sur un aspect particulier d'un objet singulier, lorsque l'objectif de la théorie de la relativité générale est de fournir un cadre théorique pour mettre en évidence et expliquer des phénomènes généraux. Les concepts de science mobilisés dans ces deux cas ne sont pas les mêmes, bien qu'il puisse y avoir un lien entre eux.

De la même façon, la scientificité de la *théorie des cordes* en tant que production *mathématique* ne sera pas questionnable de la même façon qu'en ce qui concerne sa revendication à constituer une théorie fondamentale de la *physique*. Si son apport conceptuel ne fait pas de doute dans le premier cas, il est en effet beaucoup moins consensuel dans le second (SMOLIN, 2007 ; WOIT, 2007).

En bref, qu'il puisse y avoir plusieurs concepts de science ou de scientificité ne me semble pas problématique si on ne parle pas du même type d'unité d'analyse, ou bien si on parle de démarches ne poursuivant en réalité pas le même objectif.

Enfin, la troisième dimension est disciplinaire, c'est-à-dire relative au type d'objets étudiés par la discipline en question. En d'autres termes, peut-on juger de la scientificité d'une discipline indépendamment du type d'objet qu'elle manipule ? Peut-on étudier les phénomènes sociaux avec les mêmes prémisses épistémologiques que lorsque l'on étudie les phénomènes physiques ? Plus spécifiquement, qualifier une théorie de scientifique en physique veut-il dire la même chose (ou *peut-il* vouloir dire la même chose) qu'en sociologie, en économie ou en biologie, et ce indépendamment de leurs différences évidentes ?

La pluralité se reflète ici dans une diversité des définitions possibles du concept de science *du seul fait que l'objet de l'investigation est d'une nature différente*, lors même que le concept porte a priori sur le même type d'unités d'analyse poursuivant un même objectif de production de connaissances générales.

C'est précisément cette dimension de la pluralité qui traduit l'inconfort qui était au point de départ de mon travail de thèse. Est-il vrai que, selon le type d'objet étudié, les productions cognitives correspondantes ne peuvent pas toutes se rendre intelligibles au sein du même cadre épistémologique? Plus spécifiquement, est-il juste, pertinent ou fécond de considérer que les sciences sociales, en tant qu'elles entendent produire des *connaissances*, doivent jouir d'une épistémologie qui leur est propre?

En d'autres termes, est-il vraiment *nécessaire* de former et de soutenir la pertinence d'un concept distinct de *science*, qui n'ait de sens que pour certaines disciplines? En effet, la création d'un nouveau concept est un processus coûteux en général, en ce qu'elle peut facilement mener à brouiller certains repères – des repères épistémologiques dans le cas qui m'intéresse. En particulier, cela s'incarne concrètement, en sciences sociales, dans l'absence manifeste de critères méta-théoriques clairs et partagés. Il ne semble donc pas évident a priori que ce coût soit compensé par un quelconque bénéfice cognitif.

Pour toutes ces raisons, cette dimension de la pluralité conceptuelle me paraît éminemment plus problématique.

Un cadre d'analyse utile pour appréhender les différentes façons de se positionner face à une telle pluralité est celui développé par Géraldine Carranante dans un travail récent (CARRANANTE, 2023) portant sur la pluralité conceptuelle en science.

Géraldine Carranante identifie quatre réactions typiques face à une situation de pluralité conceptuelle en science, c'est-à-dire quatre types de stratégies mises en œuvre par les scientifiques lorsqu'ils se retrouvent dans une telle situation :

- (i) *La sélection* : une version d'un concept est choisie parmi toutes les propositions disponibles, en fournissant des arguments défendant le fait que cette version est, d'une manière qu'il faut préciser, la meilleure;
- (ii) *L'unification/l'intégration* : une nouvelle version du concept est créée qui unifie toutes les versions déjà proposées, et qui ne correspond

donc pas strictement à une version disponible mais récupère des attributs de chaque version;

- (iii) *L'élimination/le remplacement* : toutes les versions du concept en question sont éliminées, identifiant par exemple une erreur de catégorie, et le cadre conceptuel lui-même est remplacé;
- (iv) *Le pluralisme* : le fait qu'un même concept puisse s'instancier différemment – la *pluralité conceptuelle* – est accepté.

On pourrait transposer ce cadre d'analyse, développé pour s'appliquer aux concepts scientifiques, à notre situation philosophique portant sur le problème de la démarcation. Les différentes stratégies citées correspondent alors à des types de résolution de ce problème.

Sélectionner le meilleur concept de science pourrait correspondre à identifier l'épistémologie d'une discipline particulière comme étant la seule valable, et calquer ensuite les autres disciplines sur cette épistémologie.

L'unification correspondrait au fait de construire un nouveau concept unifié de la scientificité à partir des concepts déjà existants, tel qu'il puisse s'appliquer universellement tout en contenant des éléments importants des différentes épistémologies déjà disponibles.

L'élimination signifierait identifier une erreur de catégorie dans toute tentative de définition du concept de science, c'est-à-dire qu'elle interpréterait la pluralité conceptuelle comme un signe de la non pertinence de cette tentative, et réorienterait les efforts vers un autre cadre conceptuel.

Le pluralisme, enfin, reviendrait à accepter que plusieurs versions distinctes du concept *science* peuvent légitimement co-exister.

On pourrait donc, en toute généralité, qualifier de *pluralistes* les positions soutenant la nécessité d'une épistémologie à part pour les sciences sociales. Dans la littérature, on trouve les qualificatifs plus précis de *dualistes* (il existe une distinction binaire entre les sciences sociales d'une part, les sciences naturelles d'autre part) ou encore *régionalistes* (chaque discipline doit avoir sa propre épistémologie) (RAYNAUD, 2006).

Une position relevant de la *sélection* pourrait revenir à considérer que l'épistémologie des sciences physiques, par exemple, doit être appliquée telle quelle à toutes les autres disciplines, et en particulier aux sciences sociales. Cependant, le fait est qu'il n'existe pas non plus de vision parfaitement unifiée de ce qu'est ou doit être l'épistémologie de la physique. En effet, les questions portant sur la démarcation scientifique *en général* n'ont pour l'instant pas trouvé de réponses définitives, et ce, même en physique.

C'est pourquoi je me place davantage dans une démarche d'*unification*. Dans cette thèse, j'explorerai la possibilité d'une définition générale et unique de la scientificité.

Mon approche comprenant un travail important de clarification conceptuelle, il va être nécessaire de préciser le cadre d'analyse de ma réflexion. Il sera ensuite mobilisé tout au long de cette thèse, et il permettra à la fois de rendre compte d'une manière claire des différentes tentatives de résolution existantes dans la littérature et de préciser ma démarche.

Le *cadre conceptuel* à l'œuvre, c'est-à-dire le ou les concepts que l'on souhaite appliquer à certains objets ou types d'objets bien définis, est le premier élément important à définir. Il peut s'agir du couple {SCIENCE, NON-SCIENCE}, ou encore du triplet {SCIENCE, NON-SCIENCE, PSEUDO-SCIENCE}, mais d'autres cadres sont bien sûr possibles.¹

Les deux caractéristiques centrales d'une analyse conceptuelle sont celles d'*extension* et d'*intension*.

L'*extension* d'un concept, aussi appelée sa référence, consiste en l'ensemble des objets qui répondent positivement à ce concept. Ici, l'*extension* du concept de *science*, par exemple, pourrait être l'ensemble des disciplines effectivement scientifiques.

L'*intension* d'un concept, aussi appelée parfois son rôle inférentiel, est sa définition théorique, c'est-à-dire une caractérisation de ce concept à l'aide d'autres concepts. Le problème de la démarcation scientifique peut ainsi être formulé comme celui de la caractérisation de l'*intension* du concept de *science* (FERNANDEZ-BEANATO, 2020, p. 1).

Une quatrième composante importante de mon outillage analytique est ce que l'on pourrait appeler le *champ d'application*, ou *domaine*, du concept. Le domaine d'un concept ou d'un ensemble de concepts est l'ensemble des objets du monde sur lesquels il est intéressant, ou pertinent, de se demander si ces objets répondent positivement ou non au(x) concept(x) en question. En d'autres termes, dans mon cas, il s'agira ici de définir plus précisément les *unités d'analyse* qui peuvent être qualifiées de scientifiques, non scientifiques ou pseudo-scientifiques (si c'est le cadre que l'on a choisi), c'est-à-dire les "objets de la démarcation" (HANSSON, 2021, p. 13).

Une œuvre littéraire de fiction, par exemple, n'est pas une connaissance

1. Voir par exemple la partie 3 de (HANSSON, 2021) pour une discussion à ce sujet.

scientifique, et il est intéressant de se demander pourquoi, et ce qui la distingue d'objets similaires mais authentiquement scientifiques, comme un ouvrage historique. À l'inverse, un objet de la vie quotidienne comme une table n'est pas une connaissance scientifique non plus, mais il est beaucoup moins intéressant ou pertinent de se demander pourquoi.

Cette constatation montre que, dans toute analyse conceptuelle, il y a toujours déjà un sous-ensemble des objets du monde qui est donné a priori et sur lesquels il est intéressant de se demander à quel(s) concept(s) ils répondent effectivement. Un tel domaine conceptuel est souvent implicite, et je le rendrai explicite afin de résoudre le problème que je me suis posé. Le domaine d'application d'un cadre conceptuel est donc un ensemble plus grand que les extensions des concepts en jeu, mais ne contient pas non plus tous les objets du monde – il y a, comme on l'a remarqué, un certain choix a priori qui est opéré afin de rendre pertinente l'analyse conceptuelle.

Les intensions respectives des concepts, quand à elles, correspondent à une reconstruction théorique de la partition du champ d'application en leurs différentes extensions. Par exemple, comme cela a souvent été fait pour le problème de la démarcation scientifique, il peut s'agir de fournir une propriété que les objets du champ d'application peuvent posséder ou non, et qui les classent en fonction de cela, par définition, dans les différentes catégories visant à correspondre aux extensions respectives des concepts.

Dans l'histoire du problème de la démarcation, un exemple fameux d'un tel critère reposant sur une propriété unique est le critère de *falsifiabilité* introduit par Karl Popper (POPPER, 1934; POPPER, 1959; POPPER, 1962) pour définir la scientificité d'un énoncé ou d'un système d'énoncés. Une critique parfois adressée à ce critère est qu'il ne semble pas permettre de retrouver l'extension du concept de science dans son acception intuitive, en ce qu'il peut désigner la théorie de la Terre plate comme étant scientifique et la théorie de l'évolution darwinienne comme ne l'étant pas. Dans cette thèse, je reviendrai longuement sur ce critère, ses limites et ses possibilités d'amélioration.

Plus généralement, on peut aussi définir un concept à l'aide d'un *ensemble* de propriétés, et il faut alors que celles-ci soit "individuellement nécessaires et collectivement suffisantes" (RAYNAUD, 2021, p. 74), (LAUDAN, 1983, p. 118) pour caractériser le concept en question. En d'autres termes, il faut et il suffit qu'un objet donné possède toutes ces propriétés, sans exception, pour qu'il corresponde au concept.

C'est notamment l'impossibilité de mettre en évidence un tel ensemble

restreint de propriétés qui a poussé Larry Laudan à affirmer que le problème de la démarcation était en réalité un "pseudo-problème" (LAUDAN, 1983, p. 124) – j'aurai également l'occasion d'y revenir.

Un concept peut aussi être défini de manière moins stricte, par exemple comme une *gradation*. Il pourrait s'agir d'une propriété non pas binaire (falsifiable/infalsifiable) mais graduelle (un *degré* de falsifiabilité). Dans ce cas, les unités d'analyse en question ne seraient pas strictement scientifiques ou non scientifiques, mais seraient pourvues d'un certain degré de scientificité. Cela reviendrait à ne plus parler d'une unité scientifique *dans l'absolu*, de la même façon que l'on ne peut pas parler d'un matériau qui serait "chaud" ou "froid" dans l'absolu. Pour autant, définir une scientificité graduelle nous permettrait toutefois de comparer ces différentes unités d'analyse entre elles et éventuellement de les hiérarchiser – ce qui est précisément l'objectif d'un critère de choix théorique.

Un dernier élément qu'il peut sembler pertinent d'introduire pour l'analyse conceptuelle qui nous occupe est le *but épistémique* d'un concept ou d'un cadre conceptuel, une notion introduite par Ingo Brigandt (BRIGANDT, 2009) dans le cadre des concepts scientifiques. Le but épistémique d'un concept scientifique, selon lui, est ce à quoi il sert pour le ou la scientifique qui l'utilise, un même concept pouvant avoir des objectifs différents – par exemple, classer des phénomènes, produire une explication, imaginer de nouvelles hypothèses, et ainsi de suite.

Dans notre cas, qui ne se place pas dans un cadre strictement scientifique mais plutôt philosophique, le but épistémique d'un concept, par exemple celui de SCIENCE, caractérise le type de démarche que l'on adopte, en rejoignant par exemple la distinction classique entre approches *descriptive* et *normative*. Plus précisément, une approche descriptive a pour objectif de clarifier les définitions déjà existantes de la scientificité : partant d'une certaine extension du concept – par exemple, la catégorisation intuitive qu'en font les scientifiques concernés – une telle approche vise à la reconstruire à partir de considérations théoriques. Une approche normative, au contraire, pourrait s'aventurer à remettre en question la classification (l'extension) intuitive des différentes disciplines ou théories en science, non-science et pseudo-science s'il se trouve que celle-ci ne correspond pas à la définition théorique (l'intension) qu'elle propose. Ainsi, bien qu'il soit plus ou moins nécessaire de partir d'une classification donnée, une approche normative a surtout pour objectif de bâtir une caractérisation théorique relativement indépendante de cette classification première, pour pouvoir ensuite lui être appliquée de manière

critique.

Quoi qu'il en soit, puisqu'il apparaît évident qu'une analyse conceptuelle dépend, dans son déploiement d'ensemble, de l'objectif pour laquelle elle est mobilisée, le but épistémique poursuivi dans notre travail (comme dans tout travail de clarification conceptuelle) devra être explicité à un moment donné, pour que l'on puisse ensuite juger clairement de la pertinence effective des outils mis en œuvre pour l'atteindre.

Dans ma thèse, je me focaliserai sur un type d'unités d'analyse particulier, à savoir les unités *cognitives* (telles que les énoncés, les systèmes d'énoncés, les modèles ou les théories). En d'autres termes, je ne m'intéresse pas en premier lieu aux unités plus larges telles que les communautés scientifiques ou les champs épistémiques, c'est-à-dire des structures en partie sociale productrice de connaissance, mais bien *au contenu produit lui-même*. Je me focalise plus précisément sur les *modèles* et les perçoit comme les unités élémentaires de notre connaissance, dans l'esprit de la conception dite "sémantique" des théories.

De plus, ce que je cherche avant tout c'est à définir la scientificité pour des unités ayant une visée proprement *épistémique*, c'est-à-dire avec l'objectif de produire des *connaissances générales* à propos d'un type donné d'objets. Ce qui m'intéresse ce sont donc les conditions de possibilité d'une science *fondamentale* (RAYNAUD, 2021), tournée vers la "recherche désintéressée de nouvelles connaissances scientifiques".²

Plusieurs obstacles se dressent face à une telle approche.

Premièrement, en définissant les modèles comme unités d'analyse, on peut aisément se retrouver piégé dans une *pétition de principe*, où l'on décrète dès le début ce que l'on veut démontrer – circularité fatale qui neutraliserait l'ensemble de mon raisonnement.

Il semble nécessaire de partir d'une unité d'analyse bien identifiée. Afin qu'une comparaison soit intéressante, il faut en effet travailler sur des objets comparables, et ainsi trouver *à la base* une unité commune à toutes les

2. Suivant les conseils avisés d'Oliver Sartenaer, il me semble important de préciser ici une chose à garder en tête à la lecture de cette thèse. Si la question de la *scientificité* (qu'est-ce qui caractérise une théorie scientifique ?) est généralement distincte de celle de la *justification* (quelles sont les bonnes raisons de préférer une théorie plutôt qu'une autre ?), la façon avec laquelle je les aborde dans mon travail les identifie en grande partie. En effet, comme je le développe par la suite, même si je ne parle pas directement de théorie de la justification, c'est bien de cela dont il s'agit, et le degré de scientificité d'une théorie est alors identifié à son degré de justification.

productions à vocation scientifique, avant même d'élaborer une quelconque notion de scientificité, unitaire ou pas.

On pourrait cependant m'objecter que prendre les modèles comme unité de base revient déjà à supposer que toutes les disciplines ont en commun d'utiliser des modèles, et donc, en partie au moins, supposer ce que je souhaite démontrer. Je m'efforcerai, le moment venu, de motiver la focalisation sur les modèles qui est la mienne dans ce travail. En particulier, j'utiliserai une notion relativement *ample* de modèles pour pouvoir englober un ensemble assez étendu de produits cognitifs, afin d'éviter autant que faire se peut cette circularité.

Un deuxième obstacle relève de l'épistémologie générale des sciences. Face à diverses tentatives pour parvenir à une telle définition de la scientificité, par exemple via le concept de falsifiabilité poppérienne, un certain nombre de problèmes ont été mis en évidence dans la littérature. En particulier, il est aujourd'hui assez consensuel en épistémologie de considérer que toute observation est chargée de théorie ou plus généralement que la confirmation empirique est un processus *holiste*, ce qui affaiblit la portée voire la possibilité même d'une "simple" réfutation.

Notre connaissance possède ainsi une structure logico-empirique éminemment complexe qu'il faut donc intégrer dans toute tentative de définition générale de la scientificité. Comme je vais m'attacher à le montrer, il est possible de tenir compte de ces aspects en explicitant le plus possible la structure des unités cognitives, vues comme des "méta-modèles", dont on veut définir le degré de scientificité.

Troisièmement, le fait que l'acceptation d'une théorie ou d'un modèle relève d'une action *collective* pour laquelle tout un ensemble d'aspects "non épistémiques" interviennent est également relativement consensuel aujourd'hui en épistémologie et en philosophie des sciences. En d'autres termes, le choix théorique ne dépend pas uniquement des qualités intrinsèques des théories, mais également de contingences socio-historiques – par exemple, des données empiriques ou des théories concurrentes disponibles à un moment donné, de la structure des rapports de pouvoir au sein de la communauté scientifique, ou plus largement d'une idéologie dominante dans la société.

Au vu de ce constat, il semble difficile de définir la scientificité au niveau des unités cognitives "pures" comme les modèles, sans prendre en compte des aspects plus larges. Afin de surmonter cette difficulté, j'envisage la scientificité non pas comme une qualité intrinsèque que posséderait une telle unité

cognitives, mais comme le résultat d'un processus d'*optimisation*.

Plus précisément, l'optimisation dont il est question est la maximisation d'une certaine quantité définie sur cette unité. Afin de donner à cette démarche un contenu non trivial, j'ai proposé une définition la plus précise possible de cette quantité à maximiser. Il s'agit d'une version sophistiquée du degré de corroboration poppérien, traduisant ce qui selon moi constitue l'apport principal du falsificationnisme : l'idée que la primauté, en termes de scientificité, ne doit pas être donnée à l'adéquation empirique mais à la quantité d'information contenue dans cette adéquation empirique. En d'autres termes, la question fondamentale n'est pas de savoir à quel point un modèle est adéquat aux données, mais quelle information est gagnée à savoir qu'un modèle est à ce point adéquat aux données – autrement dit, *dans quelle mesure il aurait pu ne pas l'être*.

La scientificité est alors définie relativement à la *maximisation* de ce degré de corroboration informationnelle. Cela permet de prendre en compte les aspects socio-historiques dans la définition de la scientificité, en les interprétant comme des contraintes contingentes qui pèsent sur le processus d'optimisation, et en même temps de considérer que la quantité à maximiser, en ce qui la concerne, est propre à l'unité d'analyse considérée.

Enfin, un quatrième type d'obstacles est plus spécifique et concerne le fait que certaines disciplines revendiquent, comme je l'ai déjà mentionné, une épistémologie *à part*. Je me concentre dans cette thèse sur cette revendication au sein des sciences sociales, et notamment en sociologie. En effet, depuis leur émergence au dix-neuvième siècle, un certain nombre d'arguments ont été développés pour défendre la nécessité d'une épistémologie singulière, propres aux sciences sociales.

Ces arguments reposent principalement sur le fait que les objets d'investigation propres à ces champs possèdent certaines propriétés qui leur sont, d'une part, spécifiques, et, d'autre part, qui les empêcheraient d'accéder à la même forme de scientificité que les autres disciplines, comme les sciences dites "naturelles".

Pour aborder ce type de questions, j'emploierai deux types d'arguments mobilisés en synergie : des arguments *théoriques* et des arguments *par l'exemple*.

Un argument théorique pour ou contre l'unité épistémologique des sciences se réfère à des raisons *de principe* défendant le point de vue considéré. On pourrait par exemple soutenir que les productions sociologiques

ne peuvent pas, de par la nature même de leur objet, être rendues intelligibles au sein du même cadre épistémologique que les sciences physiques, en détaillant le lien formel entre la conclusion (l'impossibilité de définir la scientificité de la même façon) et la raison centrale de l'argument, à savoir ici la nature des phénomènes sociaux.

Au contraire, on pourrait soutenir qu'il n'y a aucune raison de penser que les phénomènes physiques aient une spécificité qui puissent nous donner accès plus facilement à certaines régularités empiriques, en répondant aux divers éléments soulevés par la première position.

Les arguments du second type, à savoir des arguments *par l'exemple*, partent au contraire de productions concrètes qui leur permettent de tester de manière critique certains arguments de principes – ou/et d'en construire de nouveaux. Je prendrai l'exemple de *la sociologie analytique*, un courant qui se développe depuis les années 1990 et qui ne semble pas revendiquer la nécessité d'une épistémologie singulière pour rendre intelligible ses productions cognitives.

Un tel contre-exemple suggère qu'il est *possible*, tout du moins, de produire des connaissances sociologiques ayant la même forme que celles produites dans les sciences naturelles – notamment en physique – et fournit une base concrète sur laquelle le débat pourrait se matérialiser de manière moins abstraite. Bien entendu, l'existence de contre-exemples ne signifie pas que ces contre-exemples soient *pertinents*. C'est pour cela que ces deux types d'arguments doivent être mobilisés en interaction constante, et qu'aucun ne se suffit à lui-même.

Les deux premiers chapitres de cette thèse poseront en détail le problème auquel je m'attelle, tout en précisant en profondeur les différents types d'obstacles que ma démarche rencontre. Je m'appuierai pour ce faire sur le cadre d'analyse conceptuelle présenté plus haut, qui me permettra de classer les différentes propositions existantes dans la littérature afin de rendre la situation plus intelligible.

Dans le chapitre 2, je présenterai un historique du débat (moderne) de la démarcation scientifique, c'est-à-dire à la fois des différentes formes sous lesquelles ce problème a été posé, les différentes tentatives de résolution qui ont pu être proposées ainsi que les limites classiques rencontrées. Si le débat s'est initialement structuré autour de la proposition falsificationniste poppérienne, il s'est ensuite diversifié avec l'émergence de plusieurs courants tendant à prendre en considération des unités d'analyse et des critères de

démarcation de plus en plus vastes et hétérogènes.

Le chapitre 3 se concentrera sur le cas plus spécifique des sciences sociales et sur les différents débats existants à propos de leur régime de scientificité. J'exposerai quatre familles principales d'arguments pluralistes, c'est-à-dire soutenant la nécessité d'une épistémologie à part pour les sciences sociales, reposant sur quatre propriétés spécifiques de ces dernières : l'historicité, la signification, l'intentionnalité et la performativité. Je leur opposerai ensuite un certain nombre de réponses qui semblent assez classiques dans la littérature. Cela me permettra de soutenir qu'il ne semble pas y avoir de raisons solides, après analyse, d'adopter une position pluraliste.

Dans les deux chapitres suivants, je m'attellerai à élaborer une notion unitaire de la scientificité en deux temps : d'abord sur la partie empirique des unités cognitives choisies, à savoir les modèles, et ensuite sur leur partie théorique. Comme je l'ai décrit plus haut, je considère la scientificité comme une qualité émergente d'un processus d'optimisation consistant en la maximisation d'une certaine quantité. Pour que cette approche ait un contenu intéressant, il faut préciser la nature et la forme de cette quantité à maximiser.

Le chapitre 4 sera consacré à la partie empirique des modèles. Je montrerai comment on peut, dans ce cadre, forger une notion de *degré de corroboration informationnelle* inspirée de la corroboration poppérienne qui permette de définir proprement un degré de scientificité des modèles empiriques comme la maximisation de cette quantité. La centralité de cette notion dans mon travail se justifie par le fait qu'elle permet de retrouver des critères épistémologiques classiques de manière unifiée, comme des cas particuliers d'un principe plus général de maximisation.

Le chapitre 5 suivra tout d'abord une ligne argumentative *par l'exemple*, en prenant comme cas d'étude la *sociologie analytique*. Constituant un potentiel contre-exemple factuel à la position épistémologique pluraliste, il s'agit donc d'un courant intéressant et pertinent à étudier dans le cadre d'une réflexion sur l'unité épistémologique des sciences.

Ce chapitre me permettra également de compléter l'élaboration d'une notion unitaire de la scientificité, du côté des modèles théoriques à présent. En effet, l'élément clé d'une approche analytique en sociologie est la théorie de l'action utilisée au niveau micro (individuel) afin d'expliquer les phénomènes sociaux (macro). Pour que la sociologie analytique puisse effectivement constituer un contre-exemple valide à la thèse pluraliste, il est nécessaire que les théories de l'action mobilisées au cœur de ses modèles explicatifs jouissent d'une justification épistémologique robuste.

Cette discussion permettra de faire le lien avec des questions épistémologiques plus générales, comme celles du statut épistémologique des principes fondamentaux en science, au cœur des modèles théoriques. Ces principes font face à un dilemme classique qu'il faudra surmonter. D'une part, les modèles théoriques ayant une visée explicative reposent souvent sur des hypothèses idéalisées voire notoirement fausses; d'autre part, des modèles plus complexes reposent sur davantage d'hypothèses et de paramètres libres, ce qui peut donc rendre l'explication triviale – c'est-à-dire sans contenu.

Je proposerai alors un critère épistémologique portant sur la *structure* qui émerge de l'application systématique d'un même type d'hypothèses à un même type de phénomènes. C'est la qualité de la *classification des phénomènes* obtenue qui constituera alors la valeur épistémologique d'un principe fondamental, et non pas son éventuelle correspondance avec la réalité. Ce critère me permettra en retour d'évaluer dans quelle mesure la sociologie analytique fournit effectivement un contre-exemple crédible à la thèse pluraliste.

En conclusion, je reviendrai sur certaines limitations des différentes constructions théoriques proposées dans cette thèse. Cela me permettra d'ouvrir le débat à d'autres disciplines et exemples que je n'aurai pas eu l'occasion de traiter en détail dans le cœur de mon travail.

Chapitre 2

Le problème de la démarcation scientifique

2.1 Introduction

Dans l'histoire de la philosophie occidentale, on peut faire remonter la question de la distinction entre connaissance et croyance non fondée, science et non-science ou encore vérité et opinion au moins aux pré-socratiques, en particulier à Parménide (LAUDAN, 1983, p. 112), il y a donc plus de deux mille cinq cents ans. Comme on va le voir, peu de consensus ressort de ce long débat aux multiples facettes et (re)formulations, mise à part peut-être l'identification de tout un ensemble d'obstacles à l'élaboration d'un (ensemble restreint de) critère(s) simple(s) de scientificité. On ne sait toujours pas ce qui distingue la science de la non-science, mais on sait en tout cas qu'un certain nombre d'approches n'ont pas abouti, et toute nouvelle tentative doit donc intégrer les enseignements tirés de ces échecs.

L'objectif de ce chapitre est davantage de proposer une synthèse des débats et des différentes tentatives de démarcation qui puisse être utile pour mon travail que d'en présenter chronologiquement chaque détail. Dans cette introduction, je vais néanmoins commencer par en évoquer les grandes étapes, en suivant principalement la présentation historique qu'en fait Laudan (LAUDAN, 1983). Je présenterai ensuite le plan du chapitre dans lequel je me concentrerai plus en détail, comme on le verra, sur l'ère moderne et contemporaine – c'est-à-dire *grosso modo* des années 1930 à aujourd'hui.

Malgré leurs divergences, une idée commune à toutes les approches de la démarcation depuis les *Seconds Analytiques* d'Aristote jusqu'aux penseurs du dix-septième siècle est que la connaissance scientifique, entre autres choses, jouit d'une *certitude apodictique*. En d'autres termes, une connaissance scientifique authentique est *infaillible, nécessairement vraie*, au contraire d'une opinion ordinaire, par exemple. Les discussions se concentrent alors davantage

sur la manière d'établir cette infaillibilité, ce sur quoi les vues divergent éventuellement.

Ce n'est qu'au dix-neuvième siècle qu'une vision épistémologique *faillibiliste*, c'est-à-dire se constituant autour de l'idée que la connaissance scientifique peut, tout en étant bien une connaissance authentique, être révisée, corrigée, et même finir par être prouvée fautive, émerge et s'impose finalement. Les tentatives de démarcation entre connaissance scientifique et croyance ordinaire se reportent alors davantage sur la *méthode* utilisée par les scientifiques, qui serait d'une nature particulière et leur permettrait, au contraire de l'activité cognitive ordinaire et intuitive du profane, d'atteindre des connaissances plus assurées.

On voit ici, en utilisant les termes que l'on a présenté en introduction, un changement dans l'*unité d'analyse* sur lequel on prétend appliquer le concept SCIENCE : d'un contenu cognitif (quelles sont les caractéristiques de la *connaissance* scientifique), on passe à la caractérisation d'une pratique (quelle est la bonne *méthode* pour produire de la connaissance scientifique). Cette approche par la méthode a l'avantage d'être un peu plus souple et notamment d'intégrer le fait que la connaissance produite peut tout à fait être faillible, puisque l'efficacité d'une méthode est de fait conditionnée par des éléments contingents, comme par exemple les données empiriques disponibles à une époque particulière.

Pour autant, elle souffre également de multiples défauts. Comme on peut s'y attendre, il y a peu de consensus, déjà à l'époque, sur la caractérisation précise de cette méthode, et les diverses tentatives reposent le plus souvent sur des expressions trop ambiguës pour être réellement utiles concrètement. De plus, comme le remarque plusieurs observateurs, les méthodes prescrites par ces différentes tentatives n'ont de toute façon que très peu à voir avec celles effectivement mises en œuvre par les scientifiques dans leur travail réel.

Notons en passant, et en anticipant un peu, que dans le champ de l'épistémologie des sciences sociales, et en particulier en économie, c'est aussi à ce moment-là (deuxième moitié du dix-neuvième siècle) qu'éclate le *Methodentritt*, le conflit des méthodes. Ce terme caractérise une profonde divergence à propos de savoir si la méthodologie des sciences physiques peut ou doit être appliquée en économie, ou bien si celle-ci (et avec elle les autres sciences sociales naissantes, en particulier la sociologie) doit jouir d'une méthodologie *à part*.

Quoiqu'il en soit et malgré la grande difficulté à se mettre d'accord, la

question de ce qui distingue intrinsèquement une connaissance authentique d'une croyance ordinaire, ou une activité proprement scientifique d'une spéculation métaphysique, continue de refaire surface régulièrement. C'est en particulier l'un des objectifs du Cercle de Vienne à partir des années 1920, comme exprimé dans leur manifeste *La conception scientifique du monde*, que de réaliser une démarcation claire entre les énoncés doués de signification et les autres.

Il ne s'agit pas ici de démarcation scientifique à proprement parler, cela dit, même si le mouvement qui l'accompagne, l'empirisme logique (aussi dénommé plus tard *vision syntaxique* des théories), vise à une reconstruction rationnelle des théories scientifiques – vues comme des structures logiques reposant en dernière instance sur des énoncés empiriques à la validité indiscutable. En particulier, un énoncé est doué de sens s'il est vérifiable expérimentalement, au contraire de la plupart des productions métaphysiques qui sont vues par les membres du cercle et leurs sympathisants comme des spéculations intellectuelles, profondes qu'en apparence, et dont les termes n'ont pas de référents bien identifiés dans le monde.

La version moderne du problème de la démarcation scientifique remonte à Karl Popper (POPPER, 1934) et à son épistémologie fondée sur le critère de *réfutabilité* : un énoncé est scientifique s'il est potentiellement *réfutable* expérimentalement, en opposition à une vision vérificationniste pour laquelle il suffit qu'un énoncé soit vérifiable expérimentalement pour pouvoir être qualifié de scientifique. Ce qu'on appelle également le *falsificationnisme* a offert une nouvelle vision de la scientificité, tout en essuyant de nombreuses critiques sur lesquelles on reviendra en détail par la suite. Popper n'a pas cessé de l'amender suite à ces critiques notamment, en particulier après la sortie en 1959 de la traduction anglaise de son *Logik der Forschung*. Sa proposition a en tout cas structuré le débat moderne sur le problème de la démarcation au moins jusqu'aux années 1960, où il a alors commencé à se diversifier.

Parallèlement, le problème de l'unité épistémologique des sciences a également refait surface à diverses reprises. Même si nous aborderons ce point particulier en détail au chapitre 3, on peut déjà citer la *Positivismusstreit*, controverse épistémologique opposant Karl Popper aux partisans de l'école de Francfort, en particulier à Theodor Adorno (ADORNO, 1976).

Plus proche de nous, on peut aussi penser à l'oeuvre déjà citée de Jean-Claude Passeron *Le raisonnement sociologique* (PASSERON, 1991; PASSERON, 2006) qui défend, comme le sous-titre *L'espace non-poppérien du raisonnement naturel* le laisse deviner, la nécessité d'une épistémologie particulière pour

les "sciences portant sur le monde historique", notamment pour la sociologie. Il conviendra, le moment venu, d'examiner plus précisément les objets de ces différentes controverses, et les différentes positions tenues, pour voir comment elles peuvent nourrir mon approche.

Comme je l'ai déjà précisé, je ne vais pas aborder le problème de la démarcation dans tout son déploiement historique. Je m'en tiendrai à sa version moderne, c'est-à-dire commençant avec le falsificationnisme poppérien, que je présenterai en détail dans une première partie. J'en présenterai notamment une version sophistiquée, avec le concept de programme de recherche, du à Imre Lakatos.

Ensuite, je ferai état d'un certain nombre de constats assez consensuels en philosophie des sciences contemporaine, souvent interprétés comme des critiques classiques du falsificationnisme. Par exemple, les réflexions critiques que Willard Quine a adressées à l'empirisme logique (QUINE, 1951 ; QUINE, 1975) et qui touchent le falsificationnisme par ricochet, la thèse dite de Duhem-Quine du holisme de la confirmation (HARDING, 1975), le fait que toute observation est chargée de théorie, ou encore l'indistinction fondamentale entre les énoncés de types analytique et synthétique.

Tous ces constats s'articulent avec les idées fondamentales du falsificationnisme en ce qu'ils mettent en évidence la non trivialité de la réaction épistémologiquement satisfaisante à adopter lorsqu'une attente théorique rentre en contradiction avec l'expérience. Afin d'illustrer ce point important et d'insister dessus, je présenterai alors quelques exemples d'ajustements théoriques, classiques en histoire des sciences.

Je ferai ensuite un état des lieux de la façon avec lequel le débat sur la question de la démarcation scientifique s'est diversifié depuis les années 1960. En particulier, je présenterai des critiques de type "socio-historique" (KUHN, 1962 ; FEYERABEND, 1975) et ses ramifications contemporaines dans le contexte des *science studies* (BLOOR, 1976 ; DUBOIS, 2001 ; PESTRE, 2006), en passant par la méta-critique de Larry Laudan dans un article déjà cité ici (LAUDAN, 1983) sur la futilité même de chercher un tel critère de démarcation.

Enfin, je présenterai diverses approches qui se sont développées à partir des années 1980 jusqu'à aujourd'hui. En particulier, les approches "multi-critères" et s'appliquant à des unités d'analyse étendues comme les *champs épistémiques* (BUNGE, 1983a ; BUNGE, 1991 ; MAHNER, 2007 ; FERNANDEZ-BEATATO, 2020), mais aussi la théorie meehilienne de la corroboration-similitude, en passant par diverses façons de voir la scientificité comme la manifestation

d'un certain processus d'optimisation.

Je tenterai au maximum de situer ces diverses critiques et développements au sein du cadre d'analyse conceptuelle que j'ai décrit en introduction, c'est-à-dire de les caractériser les unes par rapport aux autres en fonction de : leur cadre conceptuel, leur unité d'analyse, leur type de caractérisation intentionnelle, leur positionnement vis-à-vis du pluralisme, et leur but épistémique.

2.2 Le(s) falsificationnisme(s)

En dépit, comme on va le voir, des différentes formes que son épistémologie a pu prendre, la pensée de Karl Popper reste structurée autour d'une idée fondamentale : ce qui fait le caractère *scientifique* d'une unité cognitive donnée, ce n'est pas son degré d'adéquation empirique mais bien son degré de réfutabilité (*falsifiability* en anglais) – c'est-à-dire non pas à quel point elle est confirmée, d'une manière ou d'une autre, expérimentalement, mais bien à quel point *elle aurait pu ne pas l'être*. Les différentes versions de ce critère que l'on peut retrouver à travers son œuvre se structurent alors selon plusieurs dimensions.

Premièrement, on peut remarquer qu'il n'applique pas exactement le même cadre conceptuel à travers tous ses écrits. Dans (POPPER, 1959), il souhaite démarquer la science empirique de la logique, des mathématiques ou de la métaphysique. Il précise bien que son objectif, au contraire des membres du Cercle de Vienne, n'est pas de distinguer les énoncés doués de sens, ou encore "cognitivement signifiant" (HEMPEL, 1951), des autres, ni même d'ailleurs les énoncés vrais des énoncés erronés :

Ce qui me préoccupait à l'époque [à partir de 1919] n'était pas le problème de savoir "quand une théorie est vraie", ni même "quand celle-ci est recevable". La question que je posais était autre. Je voulais distinguer science et pseudo-science, tout en sachant pertinemment que souvent la science est dans l'erreur, tandis que la pseudo-science peut rencontrer inopinément la vérité [...] J'entendais tracer une frontière – aussi bien que faire se pouvait – entre les énoncés ou systèmes d'énoncés des sciences empiriques et tous les autres énoncés, que ceux-ci fussent de nature religieuse, métaphysique ou, tout simplement, pseudo-scientifique. (POPPER, 1962, p. 59, 68)

Il prend en particulier l'exemple de la prédiction "demain, ici, il pleuvra ou il ne pleuvra pas" (POPPER, 1959, p. 19). Cet énoncé est, d'une part, doué de sens : on peut donner un sens empirique clair à chacun de ses termes; d'autre part, il sera, quoiqu'il arrive, vérifié par l'observation. Pourtant, il a un contenu informationnel *nul*. Ainsi, et c'est le cœur de tout son raisonnement, le fait d'avoir un sens empirique, et qui plus est d'être effectivement vérifié expérimentalement, n'est pas une condition suffisante, pour un énoncé, pour pouvoir être qualifié de scientifique.

Ensuite, son critère ne porte pas toujours sur les mêmes unités d'analyse. En effet, dans (POPPER, 1934; POPPER, 1959), il présente son critère de démarcation comme s'appliquant sur des énoncés singuliers, ou éventuellement sur des systèmes d'énoncés. D'ailleurs, il a conscience que le choix d'une telle unité d'analyse ne va pas de soit, comme en témoigne l'introduction du chapitre "Falsifiability" (POPPER, 1959, p. 57) "The question whether there is such a thing as a falsifiable singular statement (or a 'basic statement') will be examined later", anticipant par là-même une critique très importante de la pertinence du critère de réfutabilité.

Dans (POPPER, 1962) il parle toujours d'énoncés, de systèmes d'énoncés, également de *théories*, mais aussi parfois de *postures* ou d'*attitudes* qui seraient (ou pas) scientifiques. Comme l'écrit Lakatos, citant Popper, *l'honnêteté intellectuelle* consiste alors à préciser, avant tout test d'une théorie, les observations qui feraient abandonner ou du moins remettre en question la théorie par ses tenants : "criteria of refutation have to be laid down beforehand : it must be agreed which observable situations, if actually observed, mean that the theory is refuted." (POPPER, 1962, p. 38).

Dans la controverse qui l'oppose à l'école de Francfort, il défend son critère de réfutabilité comme une condition nécessaire du *débat* rationnel :

Des solutions sont proposées et critiquées. Lorsqu'un essai de solution n'est pas accessible à la critique factuelle [*Sachlich*], il est éliminé du même coup comme non-scientifique, même si ce n'est peut-être que provisoire. (ADORNO, 1979, p. 77)

Un débat qui est nécessaire pour pouvoir avoir une chance de parvenir *collectivement* à une objectivité qui ne peut être définie qu'à cette échelle. Ainsi, on voit que de considérations théoriques s'appliquant au niveau des énoncés ou des systèmes d'énoncés, il finit par tirer des conséquences concrètes s'appliquant sur des pratiques, des attitudes, voir sur l'organisation sociale de la recherche scientifique – l'idée sous-jacente qui guide ses réflexions

reste bien toujours la même mais ne s'instancie pas de la même manière selon l'unité d'analyse dont on parle.

Les buts cognitifs poursuivis par son critère de réfutabilité divergent également selon les cas. S'il l'introduit en premier lieu comme un critère de démarcation scientifique, il soutient également dans ses ouvrages principaux (POPPER, 1959; POPPER, 1962; POPPER, 1972) qu'il lui permet de résoudre dans le même temps le problème de l'induction (posé par exemple par David Hume (HUME, 1748)), et ainsi fournit une condition au progrès de nos connaissances.

Tout d'abord, Popper soutient une position anti-inductiviste, en deux temps. D'une part, partant de la description humienne de l'induction, il soutient qu'il est impossible de justifier celle-ci sur des bases logiques, c'est-à-dire qu'un énoncé universel ne sera *jamais* justifié à partir d'un nombre, même immense, d'instances positives particulières de cet énoncé. D'autre part, il soutient le fait que la démarche scientifique n'est en fait pas inductive, dans le sens où les scientifiques ne partent pas de l'observation de cas particuliers pour en induire une loi générale, mais au contraire ont toujours déjà une hypothèse en tête lorsqu'ils abordent un phénomène particulier, même si elle peut être non explicite.

Popper se refuse donc à élaborer une théorie de la justification de l'induction, et n'en a de toute façon pas besoin. En effet, le deuxième point de son raisonnement est l'identification d'une asymétrie logique entre la vérité et la fausseté d'un énoncé universel : si on ne peut en effet jamais justifier avec certitude un énoncé universel du type "tous les cygnes sont blancs", on peut par contre le réfuter avec certitude, puisqu'il suffit d'observer un seul cygne qui ne soit pas blanc pour savoir que cet énoncé est faux.

Ainsi, on apprend quelque chose à chaque fois que l'on réfute une hypothèse, on augmente effectivement notre connaissance du monde. Selon Popper, la science progresse donc non pas par induction, mais bien par déduction logique, via le *modus tollens*. Comme on l'a déjà évoqué, l'objectif, même s'il est lié, n'est pas strictement le même lorsqu'il s'applique aux scientifiques, individuellement (critère d'honnêteté intellectuelle) ou collectivement (critère de débat rationnel).

Enfin, l'épistémologie falsificationniste de Karl Popper, en tant que système, se présente aussi à divers degrés de sophistication, comme résumé par exemple dans le premier chapitre des *Méthodologies* de son disciple et collègue Imre Lakatos (LAKATOS, 1978). La version présentée ci-dessus correspond à ce qu'il appelle le falsificationnisme "naïf", ou "dogmatique", une

position que Karl Popper n'a en réalité pas vraiment tenue telle quelle.

En effet, ce dernier réalise très vite qu'en fait, s'il n'est pas possible de définitivement valider empiriquement un énoncé, il n'est pas non plus possible de le réfuter définitivement. Comme il le note juste après avoir introduit le critère de démarcation (POPPER, 1959, p. 19-20) :

It might be said that even if the asymmetry is admitted, it is still impossible, for various reasons, that any theoretical system should ever be conclusively falsified. For it is always possible to find some way of evading falsification, for example by introducing ad hoc an auxiliary hypothesis, or by changing ad hoc a definition. It is even possible without logical inconsistency to adopt the position of simply refusing to acknowledge any falsifying experience whatsoever. Admittedly, scientists do not usually proceed in this way, but logically such procedure is possible; and this fact, it might be claimed, makes the logical value of my proposed criterion of demarcation dubious, to say the least.

Il est donc toujours possible de sauver un énoncé de la réfutation en modifiant de manière *ad hoc* un autre énoncé – une condition d'observation, ou bien une définition implicite, par exemple. Selon Lakatos, cette version naïve du falsificationnisme (et donc de ce contre-argument) repose sur deux postulats erronés. D'une part, il y aurait une démarcation naturelle entre les propositions théoriques et les propositions observationnelles et, d'autre part, il existerait des critères absolus permettant de décider qu'une proposition observationnelle est vraie. Les contre-arguments au premier postulat repose sur le fait qu'il n'y a pas d'observations pures, car une observation ou une expérience repose toujours sur l'utilisation d'appareil(s) de mesure, par exemple, dont le fonctionnement n'est intelligible qu'au sein d'un certain cadre théorique. Contre le second postulat, Lakatos ajoute que, même si cette démarcation existait, "the truth-value of the 'observational' propositions cannot be indubitably decided : no factual proposition can ever be proved from an experiment." (LAKATOS, 1978, p. 15). En d'autres termes, on peut juger de la cohérence ou de la consistance d'une proposition avec une observation, mais une observation ne peut jamais, à strictement parler, la prouver ou la réfuter. La raison en est qu'une observation est toujours faillible, et jamais donnée dans un langage pur dénué de toute interprétation.

Dans le but de remédier à ces défauts, le *falsificationnisme méthodologique*

repose sur un sens différent donné à la notion d'*énoncé d'observation*. Il reconnaît qu'il n'y a pas de critère naturel permettant d'accepter dans l'absolu un énoncé comme valide. Cependant, il fait le postulat qu'un critère du genre peut tout de même être partagé, à un moment donné, par des personnes ayant reçu un certain enseignement, "a 'relevant technique' such that 'anyone who has learned it' will be able to decide that the statement is 'acceptable'." (LAKATOS, 1978, p. 19-20).

Ainsi, une distinction entre des énoncés d'observation acceptables et non acceptables est à présent rendue possible, mais au prix d'un aspect qui peut apparaître relativement conventionnel. En réalité, derrière ce postulat, il y a l'idée que toute observation provient d'un appareil de mesure (au sens large du terme) dont le fonctionnement est basé sur certaines théories, théories qui sont alors considérées comme non problématiques, une "unproblematic background knowledge" (LAKATOS, 1978, p. 22) non questionnée au moment où l'on fait l'observation qui va servir à tester une théorie donnée. De ce point de vue, un nouveau sens peut être donné à la réfutation : un énoncé est réfuté par un résultat expérimental si les deux ne correspondent pas, la correspondance étant définie par un ensemble de techniques et de critères acceptés au sein de la communauté scientifique correspondante.

Cette version du falsificationnisme souffre toujours de deux défauts. Le premier, c'est qu'il présente le test d'une théorie comme une comparaison (critique, certes) entre cette théorie seule et l'expérience, alors qu'il semble que cela relève davantage d'une confrontation entre *plusieurs théories rivales* et l'expérience : "tests are – at least – three–cornered fights between rival theories and experiment" (LAKATOS, 1978, p. 31).

Le second, c'est qu'il semble suggérer que seules les réfutations d'hypothèses constituent des découvertes authentiques, alors qu'au contraire, les découvertes résultent généralement, en tout cas dans un premier temps, de la confirmation de prédictions ou d'attentes empiriques. Ces considérations nous mènent alors au *falsificationnisme sophistiqué*, pavant le chemin aux *programmes de recherche* de Lakatos. Pour un falsificationniste sophistiqué, une théorie n'est pas abandonnée à cause d'une contradiction (dans n'importe quel sens du terme) mais parce qu'elle *perd* contre une concurrente. Comme Lakatos le présente (LAKATOS, 1978, p. 32) :

For the sophisticated falsificationist a scientific theory T is falsified if and only if another theory T' has been proposed with the following characteristics : (1) T' has excess empirical content over T : that is, it predicts novel facts, that is, facts improbable in the

light of, or even forbidden, by T ; (2) T' explains the previous success of T , that is, all the unrefuted content of T is included (within the limits of observational error) in the content of T' ; and (3) some of the excess content of T' is corroborated.

Le falsificationisme sophistiqué permet alors d'instancier l'idée maîtresse de Popper d'une façon qui prend en considération l'ajout d'hypothèses suite à une réfutation apparente. La question ne porte alors plus sur la manière de distinguer entre des *théories* scientifiques et pseudo-scientifiques, mais plutôt entre des *ajustements* scientifiques ou pseudoscientifiques de théories. En d'autres termes, il s'agit de distinguer les hypothèses *ad hoc*, les *stratagèmes conventionalistes* comme les appelle Popper, des ajustements épistémologiquement acceptables.

Comme le résume Lakatos (LAKATOS, 1978, p. 33) :

[Popper's] solution - a sophisticated version of methodological falsificationism - is more objective and more rigorous. Popper agrees with the conventionalists that theories and factual propositions can always be harmonized with the help of auxiliary hypotheses : he agrees that the problem is how to demarcate between scientific and pseudoscientific *adjustments*, between rational and irrational changes of theory. According to Popper, saving a theory with the help of auxiliary hypotheses which satisfy certain well-defined conditions represents scientific progress ; but saving a theory with the help of auxiliary hypotheses which do not, represents degeneration. Popper calls such inadmissible auxiliary hypotheses *ad hoc* hypotheses, mere linguistic devices, 'conventionalist stratagem'. But then any scientific theory has to be appraised together with its auxiliary hypotheses, initial conditions, etc., and, especially, together with its predecessors so that we may see by what sort of *change* it was brought about. Then, of course, what we appraise is a *series of theories* rather than isolated *theories*.

Cet objectif est atteint, comme le suggère la dernière phrase de la citation ci-dessus, par un nouveau changement d'unité d'analyse, cette fois opéré par Lakatos lui-même qui remplace alors l'étude d'une théorie par celle d'une *suite de théories*, ou *problemshift*. Une suite de théorie T_1, T_2, \dots est définie telle que pour tout i , T_{i+1} résulte de la modification d'une hypothèse de T_i dans le but d'expliquer une anomalie empirique présente chez T_i .

Un tel *problemshift* est dit théoriquement progressif s'il prédit de nouveaux faits, et empiriquement progressif si certains de ces nouveaux faits

sont effectivement corroborés empiriquement. Ainsi, la condition d'acceptation d'un ajustement est que celui-ci, non seulement explique l'anomalie pour laquelle on l'a introduit mais mène également à la découverte de faits nouveaux.

Au contraire, un *problemshift* qui ne remplit pas ces conditions est dit *dégénèrescent*. Ce changement d'unité d'analyse est vu par Lakatos comme l'identification d'une erreur de catégorie (LAKATOS, 1978, p. 34) : "Not an isolated theory, but only a series of theories can be said to be scientific or unscientific : to apply the term 'scientific' to one single theory is a category mistake."

Le fait que Popper ne fasse pas de distinction entre une théorie et une suite de théories, c'est-à-dire en reste à une approche *synchronique* des unités cognitives d'analyse, explique selon Lakatos le désaccord exprimé par Thomas Kuhn vis-à-vis de la pertinence de la réfutabilité comme critère de démarcation scientifique. En effet, pour Popper, l'astrologie n'est pas scientifique car irréfutable (par exemple, en produisant des descriptions vagues qui peuvent valoir quel que soit le comportement humain ou les événements prédits), alors que pour Kuhn, non seulement l'astrologie est réfutable, mais elle est réfutée depuis longtemps :

[Karl Popper] says, "By making their interpretations and prophecies sufficiently vague they [astrologers] were able to explain away anything that might have been a refutation of the theory had the theory and the prophecies been more precise. In order to escape falsification they destroyed the testability of the theory." Those generalizations catch something of the spirit of the astrological enterprise. But taken at all literally, as they must be if they are to provide a demarcation criterion, they are impossible to support. The history of astrology during the centuries when it was intellectually reputable records many predictions that categorically failed. Not even astrology's most convinced and vehement exponents doubted the recurrence of such failures. Astrology cannot be barred from the sciences because of the form in which its predictions were cast. (KUHN, 1977, p. 274)

Ce désaccord est neutralisé dans la vision *diachronique* introduite par Lakatos. En effet, de ce point de vue, lorsque Popper dit que l'astrologie n'est pas scientifique car infalsifiable, cela veut dire que l'astrologie n'est pas scientifique car elle constitue une suite de théories *dégénèrescente*, ce qui du même coup est bien en accord avec les observations de Kuhn.

En résumé, le falsificationnisme a pris différentes formes selon :

- les concepts qu'il s'attachait à distinguer : science, métaphysique, pseudo-science, non-science, ...
- les unités d'analyse auxquelles il s'appliquait : un énoncé singulier, un système d'énoncés, une théorie, une série de théorie, une posture intellectuelle, une organisation sociale, ...
- son but épistémique : démarquer la science de ce qui n'est pas de la science, donner une base pour le progrès des connaissances, définir les règles d'un débat fécond, ...
- et finalement, le degré de sophistication sous lequel il se présente : approche diachronique ou synchronique, prise en compte (ou pas) des ajouts éventuels d'hypothèses suite à une réfutation, ...

Ainsi, il est possible d'être en accord, encore aujourd'hui, avec certains points particuliers sans adhérer à d'autres – par exemple, on peut toujours considérer que la réfutabilité reste une *vertu épistémologique*¹ fondamentale sans toutefois adhérer au fait que cela résout le problème de l'induction et par rebond celui du progrès de nos connaissances.

2.3 Les programmes de recherche de Lakatos

Dans la lignée du falsificationisme sophistiqué, et dans l'idée qu'il est plus pertinent que les critères de scientificité portent sur des suites de théories, des *problemshifts*, que sur des théories isolées, Lakatos définit une nouvelle unité d'analyse, les *programmes de recherche* (LAKATOS, 1978). Un programme de recherche est composé de deux éléments principaux :

- un noyau dur (*hard core*) qui contient tout un ensemble de principes fondamentaux, de lois théoriques, de définitions de base, de critères de comparaison empirique, etc. En bref, un cadre fixé, en quelque sorte la partie commune à toutes les théories d'un *problemshift* donné ;
- un ensemble d'hypothèses auxiliaires, composées d'hypothèses théoriques (par exemple, des hypothèses théoriques concernant certaines formes d'énergie en physique), ou encore qui précisent les conditions initiales correspondant à une situation particulière.

Les principes généraux du noyau dur, à l'aide de certaines hypothèses auxiliaires, permettent alors de faire des prédictions qui peuvent être comparées à des données empiriques, issues d'expériences ou d'observations,

1. Je remercie Aurélien Allard de cette suggestion suite à ma communication au congrès de la SPS à Nanterre en Mai 2023.

et être confirmées ou réfutées. Si la prédiction est considérée comme réfutée, alors le *modus tollens* implique que quelque chose doit être changé, mais comme on l'a noté plus haut, rien n'indique a priori ce qui doit être modifié.

L'idée fondamentale des programmes de recherche de Lakatos est d'explicitier le fait que certaines hypothèses sont, par *décision méthodologique*, protégées de la réfutation : ce sont celles contenues dans le noyau dur. Les hypothèses auxiliaires forment alors une *ceinture de protection* autour du noyau dur : lors d'une réfutation apparente, le *modus tollens* est dirigé vers ces hypothèses auxiliaires. Lakatos nomme cela l'*heuristique négative* d'un programme de recherche : le noyau dur est considéré comme irréfutable, les réfutations empiriques apparentes sont absorbées par la modification de certaines hypothèses auxiliaires, et cette position est considérée comme justifiée tant que cela mène à découvrir de nouveaux faits, augmentant le contenu empirique de la théorie.

Il s'agit ici de la même idée que le falsificationnisme sophistiqué : ce n'est pas un problème en soi de protéger une théorie de la falsification, pour autant que cela génère de nouvelles découvertes. Ainsi, le noyau dur d'un programme de recherche peut aussi finir par s'effondrer sous le poids de trop d'ajustements stériles dans la ceinture de protection.

Un programme de recherche est également constitué d'une *heuristique positive*. Il s'agit de la possibilité d'explorer, une fois un programme de recherche constitué, les possibilités théoriques qu'il offre. C'est-à-dire, considérer le noyau dur comme irréfutable, modifier certaines hypothèses pour mener à de nouvelles prédictions, mais sans forcément être guidé directement par l'expérience ou les données empiriques. Cette heuristique mène ainsi à modifier et à sophistiquer la ceinture de protection, "refus[ing] to be drawn into observation. (...) Occasionally, of course, [we] will ask Nature a shrewd question : [we] will then be encouraged by Nature's YES, but not discouraged by its NO." (LAKATOS, 1978, p. 50).

Cette stratégie permet de ne pas se noyer dans l'"océan d'anomalies" qui caractérise la plupart des programmes de recherche tout au long de leur développement – coupant l'herbe sous le pied à un certain falsificationnisme naïf qui voudrait que l'on abandonne une théorie dès l'instant où on en trouve une réfutation empirique.

Ainsi, cette propriété rend compte de l'autonomie relative des sciences théoriques, et en particulier celle de la physique théorique vis-à-vis de la physique expérimentale. Lakatos décrit notamment comment la publication

des *Principia* de Newton a été repoussée non pas à cause de simples inconsistances empiriques mais surtout pour des inconsistances internes, voire métaphysiques. Ceci s'avère caractéristique de comment la science fonctionne, dans le développement des programmes de recherche comme dans leur formation. Un autre exemple important est la constitution de la théorie de la relativité générale, qui a été guidée initialement par la conviction qu'Albert Einstein avait d'une certaine "harmonie de l'univers" (REICHENBACH, 1949, p. 292), et non pas dans le but d'expliquer une quelconque anomalie observationnelle – comme l'avancée du périhélie de Mercure, par exemple.

2.4 Quelques constats philosophiques et épistémologiques

Les différents obstacles argumentatifs dont l'épistémologie falsificationniste a du s'accommoder, cités en filigrane ci-dessus, peuvent être mis en connexion avec divers constats philosophiques et épistémologiques qui ont émergés simultanément, en partie en réponse à la proposition poppérienne mais surtout au développement de l'empirisme logique dans la première moitié du vingtième siècle. D'autres, encore, avaient déjà été formulés à la fin du dix-neuvième siècle par différents tenants du conventionnalisme.

2.4.1 La charge théorique des observations

Toute observation, toute expérience, toute mesure quelle qu'elle soit est toujours, d'une manière ou d'une autre, *chargée de théorie* (HANSON, 1958; KUHN, 1962; FEYERABEND, 1975; LONGINO, 1979; BOYD et BOGEN, 2021). Cela rejoint par exemple l'impossibilité de distinguer dans l'absolu des propositions "purement empiriques" d'un côté et des propositions théoriques de l'autre. En d'autres termes, la production de données empiriques comme leur utilisation en tant que test d'une théorie nécessitent toujours que certaines hypothèses soient tenues pour vraies. Notons que ces hypothèses peuvent être indépendantes *ou pas* de la théorie en question.

Prenons quelques exemples. Si des données sont rapportées par un observateur, par exemple une scientifique qui fait les mesures, alors leur utilisation dépend du fait que cette scientifique était en pleine possession de ses facultés intellectuelles à ce moment-là, qu'elle ne mentait pas, etc. Ensuite, l'appareil de mesure fonctionne d'une certaine façon, plus ou moins complexe

d'ailleurs, et ce fonctionnement repose sur des présupposés théoriques plus ou moins élaborés.

L'exemple typique, cité par Lakatos dans le cas de Galilée, est celui d'un astronome pointant sa lunette ou son télescope dans le ciel. Le test de la théorie qu'il met à l'épreuve (par exemple, l'existence d'une planète) dépend alors du fonctionnement du télescope, c'est-à-dire de présupposés théoriques concernant l'optique géométrique. Sans ces présupposés, notre astronome peut difficilement donner du sens à ce qu'il voit. Comme l'écrivait même le chevalier florentin Francesco Sizzi dans un ouvrage (SIZZI, 1611) ayant pour objectif de réfuter les affirmations de Galilée concernant l'existence de corps inconnus dans le système solaire : "La lunette déforme et trompe le jugement, aucune vérité ne peut être établie par son moyen" (TESTART, 1991, p. 49). C'est bien l'exploitation de la charge théorique de l'expérience qui lui permet d'élaborer un tel argument.

Parfois, le lien entre les données et les hypothèses sous-jacentes peut être encore plus complexe. Par exemple, en neurosciences, on peut "observer" l'activité dans une zone du cerveau grâce à l'imagerie. Or, ce qui est réellement observé, ce sont des flux d'oxygène dans le cerveau, qui sont *associés par hypothèse* à une activité neuronale – bien sûr, les neuroscientifiques ont des raisons de penser que ces hypothèses sont bonnes, mais elles n'en restent pas moins des hypothèses (BUXTON, 2010).

On peut même aller plus loin : observer ces flux d'oxygène dans le cerveau – et les interpréter de la sorte – nécessite de connaître le fonctionnement d'une machine d'IRM, qui est basé sur le phénomène de *résonance magnétique nucléaire*, décrit très précisément par la *physique quantique*. Ainsi, le test d'une hypothèse en neurosciences n'en est un que si l'on tient pour vrai tout un ensemble d'autres hypothèses appartenant possiblement à d'autres théories, voire à d'autres disciplines.

Là où la charge théorique des observations donne lieu à des inquiétudes, c'est lorsqu'elle laisse entendre qu'une certaine circularité pourrait s'immiscer dans le dispositif logico-argumentatif à l'œuvre. Prenons un exemple simple en mécanique classique. On veut tester un modèle de chute d'un objet avec frottement. Pour cela, on applique la loi fondamentale de la dynamique de Newton sur l'objet de masse m , et on en déduit, en rajoutant les conditions initiales adéquates, sa trajectoire. Pour pouvoir tester cette trajectoire théorique vis-à-vis de données empiriques concrètes, il nous faut connaître la masse m de l'objet en chute.

On peut, pour cela, utiliser par exemple une balance fonctionnant grâce à

un ressort. Or, une telle balance ne nous donne une masse exacte que parce qu'elle convertit la compression du ressort en un poids qu'elle convertit ensuite en masse, opérations uniquement rendues possibles au sein d'un cadre théorique particulier, celui de la mécanique Newtonienne elle-même. Ainsi, on ne peut pas dire que l'on "teste" la loi fondamentale de la dynamique via ce procédé sans tomber dans une circularité logique, puisque pour que ce test ait un sens, il faut avoir déjà au préalable tenu cette loi pour vraie.

(CHANG, 2004) aborde un problème similaire de circularité dans la définition historiquement laborieuse de la température, et notamment dans les premières tentatives de mesurer une telle grandeur. Il prend l'exemple du thermoscope, qui est basé sur le fait que le volume d'un objet physique, comme une barre métallique, varie en fonction de la chaleur qu'il reçoit. La température peut alors être mesurée relativement à la variation de volume d'un tel objet standard. Le problème est que dans le but de définir une échelle de température, on doit savoir quelle variation de volume correspond à quelle variation de température. Dans ce but, une mesure de la température qui soit indépendante de celle fournie par le thermoscope est nécessaire, ce qui nous ramène au point de départ.

Plus généralement, la création d'échelles de mesure dans une période pré-théorique rencontre le problème suivant (CHANG, 2004, p. 59) :

1. On souhaite mesurer la quantité X .
2. La quantité X n'est pas directement observable, donc on l'infère à partir d'une autre quantité Y , qui est directement observable.
3. Pour cette inférence, on a besoin d'une loi exprimant X comme une fonction de Y : $X : f(Y)$
4. La forme de la fonction f ne peut pas être découverte ou testée empiriquement, puisque cela impliquerait de connaître les valeurs à la fois de X et de Y , et X est précisément la variable inconnue que l'on cherche à mesurer.

Chang propose dans son ouvrage de résoudre ce problème en faisant appel à des justifications théoriques plutôt que directement empiriques. La théorie sous-jacente doit faire des hypothèses à propos de la forme de la fonction f , voire déduire cette forme à partir d'autres lois de la théorie elle-même.

Même si j'ai pris là des exemples en physique, la charge théorique des observations est bien entendu une propriété épistémologique commune à toutes les disciplines scientifiques, et donc pas spécifiquement propre à la physique.

En effet, dans les sciences humaines et sociales, par exemple, une question

fondamentale est celle de la définition de certaines variables de base comme l'intelligence, la pauvreté, le niveau de développement d'un pays, la délinquance, les traits de personnalité, et ainsi de suite. Comme dans les autres sciences, un changement dans les manières d'observer le monde résulte en une observation distincte.

Par exemple, il a été récemment montré (STOET et GEARY, 2018) qu'il existait une corrélation négative entre le degré d'égalité entre les genres dans un pays (mesuré à l'aide de l'indicateur *Global Gender Gap Index* comme défini dans (FORUM, 2022)) et la proportion, dans le pays, de femmes diplômées en science, technologie, ingénierie et mathématiques (*STEM*). Cette association statistique est mesurée alors qu'il n'y a généralement pas de différence entre les hommes et les femmes en termes de niveau de compétence scolaire. En d'autres termes, plus les femmes et les hommes sont égaux dans un pays, et donc plus les femmes semblent "libres" de choisir leur carrière académique, moins elles choisissent les *STEM*. Cette observation est appelée le *gender-equality paradox in STEM*, puisqu'on se serait a priori attendu à la corrélation inverse.

Ce résultat empirique, comme ses interprétations, ont donné lieu à plusieurs critiques. La controverse s'est en partie focalisé, et c'est cela qui m'intéresse particulièrement ici, sur la façon de mesurer le *gender gap* dans un pays. En particulier, (RICHARDSON et al., 2020) a montré que si un autre indicateur est utilisé (à savoir : le *Basic Index of Gender Inequality*), alors la corrélation disparaissait.

Ce n'est pas mon propos d'entrer dans la controverse à proprement parler, mais simplement d'illustrer de nouveau la dépendance des données empiriques aux variables de bases utilisées pour les produire – en particulier lorsque ces variables sont des *indicateurs complexes*, c'est-à-dire composés eux-mêmes d'autres variables.

2.4.2 Le holisme de la confirmation

Un second constat qui fait le lien avec le précédent est ce qu'on appelle le holisme de la confirmation, ou parfois thèse de Duhem–Quine (DUHEM, 1906; QUINE, 1951; QUINE, 1975; HARDING, 1975). Il s'agit de la constatation, déjà présente en filigrane dans les discussions ci-dessus, qu'en réalité on ne teste jamais un seul énoncé isolé, mais toujours un ensemble d'énoncés, d'hypothèses plus ou moins explicites : "If this view is right, it is misleading

to speak of the empirical content of an individual statement" (QUINE, 1951, p. 40).

Le fait que les observations soient chargées de théorie est une instance particulière du holisme de la confirmation : ce qui est testé lorsqu'on observe le ciel à travers un télescope, c'est tout à la fois l'existence de la planète, les lois de l'optique et la stabilité psychologique de l'astronome. De même, ce que l'on teste par une imagerie cérébrale, c'est tout à la fois une hypothèse portant sur le fonctionnement neuronale, l'hypothèse que les flux d'oxygène correspondent effectivement à une activité neuronale, et les lois de la physique quantique qui sous-tendent le fonctionnement de l'IRM. Ce point de vue met donc en perspective la vision linéaire de la science comme un test d'hypothèse isolé.

Cela ne veut pas dire, bien sûr, que cette vision linéaire n'est pas pertinente dans certains cas précis. En effet, un grand nombre de modèles de phénomènes physiques se produisant non loin de la surface terrestre suppose (raisonnablement) que la Terre est plate (en tout cas localement). Par analogie, on peut dans certains cas *approximer* la structure des théories scientifiques à une structure linéaire : *hypothèse* → prédiction → comparaison avec des données. C'est particulièrement le cas lorsque le réseau d'hypothèses dans lequel cette progression prend place est par ailleurs bien éprouvé – l'équivalent de l'*unproblematic background knowledge*.

Pour autant, en général, nos connaissances s'organisent dans un vaste réseau *complexe* d'hypothèses, "a man-made fabric which impinges on experience only along the edges" comme l'exprime Quine (QUINE, 1951, p. 39), c'est-à-dire une structure logico-empirique dont la topologie globale est loin d'être triviale. Ce n'est pas forcément un mal en soi, c'est par contre un fait qu'il faut prendre en considération dans toute tentative épistémologique de reconstruction rationnelle des théories.

2.4.3 La sous-détermination

Une sorte de corollaire du constat précédent est le fait que les théories sont toujours sous-déterminées par l'expérience. Si l'on associe souvent la thèse de la sous-détermination à Quine (QUINE, 1951 ; QUINE, 1975), c'est un constat qui avait déjà été fait par d'autres auteurs, en particulier des "conventionnalistes" comme par exemple Henri Poincaré. Dans *La science et l'hypothèse*, Henri Poincaré prend l'exemple de l'interpolation d'un certain nombre

de points par une courbe, et remarque que les données elles-mêmes ne déterminent pas quelle est la "bonne" courbe, et que cette détermination passe par d'autres voies, comme des considérations de simplicité :

Il est clair qu'un fait quelconque peut se généraliser d'une infinité de manières, et il s'agit de choisir; le choix ne peut être guidé que par des considérations de simplicité. Prenons le cas le plus banal, celui de l'interpolation. Nous faisons passer un trait continu, aussi régulier que possible, entre les points donnés par l'observation. Pourquoi évitons-nous les points anguleux, les inflexions trop brusques? Pourquoi ne faisons-nous pas décrire à notre courbe les zigzags les plus capricieux? C'est parce que nous savons d'avance, ou que nous croyons savoir que la loi à exprimer ne peut pas être si compliquée que ça. (POINCARÉ, 1902, p. 161-162)

Il n'est pas utile ici d'adhérer au conventionnalisme de Poincaré, et en particulier à son critère de simplicité commode, pour reconnaître avec lui que les données, étant en nombre finies, sous-déterminent bien la théorie, dans le sens où elles ne peuvent pas trancher toutes seules entre différentes courbes qui s'y ajusteraient aussi bien. C'est ce que Quine reconnaît dans (QUINE, 1975) : il existe toujours plusieurs systèmes (modèles, théories, ...) en contradiction les uns avec autres mais avec le même degré d'adéquation empirique, vis-à-vis de données particulières :

If all observable events can be accounted for in one comprehensive scientific theory - one system of the world, to echo Duhem's echo of Newton - then we may expect that they can all be accounted for equally in another, conflicting system of the world. (...) Such is the doctrine that natural science is empirically under-determined; under-determined not just by past observation but by all observable events. (QUINE, 1975, p. 1)

Une conséquence très importante de ce constat, qui va faire immédiatement le lien avec le falsificationnisme (en particulier dans sa forme diachronique), est que lors de la réfutation empirique apparente d'une théorie, le *modus tollens* implique effectivement que *quelque chose ne va pas*, mais les données empiriques en elles-mêmes sont incapables de déterminer ce qu'il faut modifier. En d'autres termes, lorsqu'une théorie se retrouve en conflit avec des observations ou des résultats d'expérience, cette "théorie" étant en

réalité constituée d'un réseau logico-empirique complexe et hautement non-linéaire, on sait qu'un certain ajustement est à réaliser, mais les données seules ne suffisent pas à déterminer quel est le bon ajustement à réaliser. Quine exprime cela à l'aide d'une nouvelle métaphore :

total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Re-evaluation of some statements entails re-evaluation of others, because of their logical interconnections - the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having re-evaluated one statement we must re-evaluate some others, whether they be statements logically connected with the first or whether they be the statements of logical connections themselves. But the total field is so undetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to re-evaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole. (QUINE, 1951, p. 39-40)

Cette constatation de Quine est très profonde, et un coup de massue pour l'espoir empiriste logique, puisqu'il met en évidence le fait que les règles de la logique ne sont pas suffisantes en elles-mêmes pour trancher entre différentes explications qui redonneraient toutes, *déductivement*, le phénomène à expliquer. Ce point de vue rejoint en quelque sorte ce qu'ont également en tête Popper et Lakatos lorsqu'ils réalisent que ce n'est pas autant sur les théories que doit porter un critère de démarcation scientifique, que sur les ajustements de théories réalisés suite à des réfutations apparentes : cela sous-entend nécessairement qu'il faut pouvoir trancher entre différents ajustements possibles, et donc suppose que cela ne va pas de soi. Si cela ne va pas de soi, c'est précisément parce que suite à une réfutation apparente, un grand nombre – sinon une infinité – d'ajustements sont *logiquement admissibles* : ils redonnent tous, en accord avec les lois de la logique déductive, les données empiriques en question.

Cependant, même s'il affirme que ni l'expérience ni les lois de la logique seules peuvent nous faire trancher entre différents ajustements possibles, il

remarque également que l'on a une certaine tendance à préférer la modification *qui dérange le moins possible* l'ensemble du réseau :

A recalcitrant experience can, I have already urged, be accommodated by any of various alternative re-evaluations in various alternative quarters of the total system; but, in the cases which we are now imagining, our natural tendency to disturb the total system as little as possible would lead us to focus our revisions upon these specific statements concerning brick houses or centaurs. These statements are felt, therefore, to have a sharper empirical reference than highly theoretical statements of physics or logic or ontology. The latter statements may be thought of as relatively centrally located within the total network, meaning merely that little preferential connection with any particular sense data obtrudes itself. (QUINE, 1951, p. 40-41)

C'est là où ces considérations rappellent l'idée derrière le noyau dur des programmes de recherche de Lakatos : il y a certaines hypothèses que l'on considère comme intouchables, et on continue à les considérer intouchables tant que cela nous fait découvrir de nouvelles choses. Ce qu'il faut remarquer, bien sûr, c'est la nécessité d'introduire *par principe* une manière de trier entre différents ajustements, et que ce principe, comme n'importe quelle hypothèse fondamentale en science, en fait, ne peut pas être entièrement justifié "de l'extérieur", par exemple en le réduisant aux lois de la logique.

De même, ce que Quine suggère ici, c'est que pour rendre compte de la manière avec laquelle on choisit entre différents ajustements, on peut reconnaître qu'une façon de faire serait de décider de choisir l'ajustement qui modifie le moins l'ensemble de nos connaissances. En d'autres termes, le meilleur ajustement serait celui qui a le moins possible de répercussions sur le reste du réseau.

Dans le livre III de son *Système de Logique*, John Stuart Mill file peu ou prou la même métaphore. Il présente l'ensemble de notre connaissance, la "régularité générale de la nature", comme un "tissu de régularités partielles, appelées lois". Ou encore :

la régularité existant dans la nature est un tissu composé de fils distincts, qui ne peut être étudié qu'en suivant chaque fil séparément, travail pour lequel il est nécessaire souvent de défaire quelque morceau de la toile et d'examiner les fils un à un. Les règles de l'expérimentation sont les moyens inventés pour effiler le tissu. page 35

Et de la même façon, il reconnaît que certaines de ces régularités sont mieux établies que d'autres, et que même si rien n'est jamais acquis, il semble préférable d'utiliser les régularités les plus établies pour juger de la robustesse des nouvelles :

Lorsque vint aux hommes l'idée d'étudier les phénomènes par une méthode plus rigoureuse et plus sûre que celle qu'ils avaient d'abord adoptée spontanément, ils ne pouvaient pas, pour se conformer à l'excellent, mais impraticable, précepte de Descartes, partir de la supposition que rien n'était encore assuré. (...) Les premiers investigateurs scientifiques admirent [certains] faits et autres semblables comme des vérités connues, et partirent de celles-ci pour en découvrir de nouvelles; et ils n'avaient pas tort en cela, bien qu'obligés, comme il s'en aperçurent dans la suite, de soumettre à une révision ultérieure même ces généralisations spontanées, lorsqu'une connaissance plus avancée leur fit assigner des limites et fit voir que leur vérité dépendait de quelque circonstance non remarquée d'abord. La suite de notre recherche montrera, je pense, qu'il n'y a aucun vice logique dans cette manière de procéder; mais on peut voir déjà que tout autre mode est absolument impraticable, puisqu'il est impossible d'instituer une méthode scientifique inductive ou un moyen de garantir la validité des inductions, sinon dans l'hypothèse que quelques inductions méritant crédit ont déjà été faites. (MILL, 1866, ???)

Le contenu principal de ce constat épistémologique, quelle que soit sa formulation, est en tout cas le même : bien qu'un tel principe de parcimonie apparaisse à première vue tout à fait intuitif, raisonnable et qu'il semble fécond de le suivre, il n'y a aucune raison strictement logique qui pourrait le justifier a priori.

2.4.4 La non distinction absolue entre énoncés analytiques et synthétiques

Un dernier apport important de Quine qui est utile à prendre en compte dans mon travail et qui est dans la continuité directe de ce que j'ai présenté ci-dessus a été l'abolition de toute démarcation *absolue* entre énoncés analytiques et énoncés synthétiques. Les énoncés analytiques sont des énoncés qui sont vraies de par la définition même de leurs termes, adjoints aux règles de la logique. Les énoncés synthétiques, quant à eux, sont des énoncés qui

peuvent être éventuellement vrais mais dont la vérité ne peut pas se déduire uniquement de la définition de leur termes et des règles de la logique.

Par exemple, "tous les cygnes sont des oiseaux" est un énoncé analytique. En effet, être un oiseau est une condition nécessaire, par définition, pour être un cygne. Par contre, l'énoncé "tous les cygnes sont blancs" est un énoncé synthétique, puisque le fait d'être blanc n'est pas inclus a priori dans la définition de "cygne". La remarque de Quine est que la distinction entre ces deux types d'énoncés dépend en réalité des définitions sous-jacentes, justement, qui sont données aux différents termes composants ces énoncés, et qu'une grande part de ces définitions sont arbitraires – ou en tout cas pourrait en toute logique être modifiée.

Par exemple, si l'on observe un cygne noir, qui semble réfuter donc la conjecture que tous les cygnes sont blancs, on peut toujours changer la définition de "cygne" et y adjoindre, dans la liste de ses propriétés de définition, le fait d'être blanc. Avec ces nouvelles définitions, l'énoncé "tous les cygnes sont blancs" est maintenant analytique, c'est-à-dire vraie par définition, puisque observer un cygne noir est devenu impossible.

Cet exemple peut sembler caricatural, mais ce sont des choses qui arrivent parfois dans le cours de la recherche scientifique. L'introduction des nombres complexes par Cardan et Bombelli au seizième siècle n'est rien de plus qu'un changement profond dans la définition de "nombre" permettant de donner un sens à l'énoncé " $z^2 < 0$ ", où z est un nombre.

Encore une fois, ce n'est pas nécessairement un mal en soi, il s'agit juste d'un fait qu'il faut nécessairement prendre en considération dans toute entreprise épistémologique : la distinction entre énoncés analytiques et synthétiques, et donc entre énoncés réfutables empiriquement et énoncés irréfutables, dépend intrinsèquement du langage d'énonciation choisi. En particulier, cette distinction dépend des définitions des différentes grandeurs et variables qui sont choisies pour observer un pan du réel.

2.5 Quelques exemples historiques d'ajustements théoriques

Avant de passer à la présentation d'approches plus contemporaines du problème de la démarcation, il m'a semblé intéressant de donner quelques exemples d'ajustements théoriques effectués dans l'histoire des sciences – particulièrement en physique. Mon objectif est principalement d'offrir des

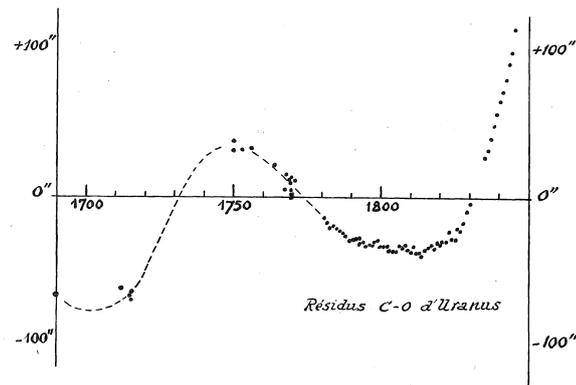


FIG. 1. — Différences entre les longitudes calculées et les longitudes observées d'Uranus (1690-1845). Valeurs rapportées par Le Verrier aux tables de Bouvard, après correction de celles-ci.

FIGURE 2.1 – Écart entre les longitudes observées et les longitudes théoriques d'Uranus (DANJON, 1946, p. 372)

illustrations concrètes des divers aspects épistémologiques théoriques présentés dans les parties précédentes.

2.5.1 Un ajustement fécond : la découverte de Neptune

Au début du dix-neuvième siècle, seules sept planètes étaient connues dans le système solaire. Il se trouve que la trajectoire d'Uranus, la planète la plus éloignée du soleil, découverte à la fin du dix-huitième siècle, ne se conformait pas aux prédictions théoriques que l'on pouvait déduire des lois de la gravitation newtonienne.

Plus précisément, l'astronome Alexis Bouvard, en 1821, avait tenté d'appliquer à Uranus la "théorie des perturbations", développée par Laplace dans les décennies précédentes. Il s'agit d'un ensemble de techniques mathématiques et calculatoires permettant d'appliquer la théorie newtonienne à des systèmes gravitationnels constitués de plusieurs corps. Elle permet de simplifier les calculs en tenant compte du fait que l'effet gravitationnel de certaines planètes (comme Jupiter et Saturne) est beaucoup plus intense que pour d'autres.

Or, en appliquant cette méthode pour prédire la trajectoire d'Uranus, Bouvard se rend compte de l'existence d'un fort écart entre ses prédictions et la trajectoire observée, comme on peut le voir figure 2.1.

Les observations ultérieures à celles de Bouvard n'avaient fait que confirmer cette anomalie, qui resta sans réponse satisfaisante pendant plus de vingt ans. Il fallut pour cela attendre les années 1840 et les travaux de l'astronome Urbain Le Verrier. Ce dernier commença à s'intéresser à ce problème spécifique en 1845. Lire ce qu'écrivait Le Verrier à cette époque est très instructif

pour avoir une idée de ce qu'il se passe en général lorsqu'une théorie rentre en contradiction avec une observation :

A peine avait-on commencé, il y a quelques années, à soupçonner que le mouvement d'Uranus était modifié par quelque cause inconnue, que déjà toutes les hypothèses possibles étaient hasardées sur la nature de cette cause. Chacun, il est vrai, suivit simplement le penchant de son imagination, sans apporter aucune considération à l'appui de son assertion. On songea à la résistance de l'éther; on parla d'un gros satellite qui accompagnerait Uranus, ou bien d'une planète encore inconnue, dont la force perturbatrice devrait être prise en considération; on alla même jusqu'à supposer qu'à cette énorme distance du Soleil, la loi de la gravitation pourrait perdre quelque chose de sa rigueur. Enfin, une comète n'aurait-elle pas pu troubler brusquement Uranus dans sa marche? (LE VERRIER, 1846, p. 69)

Plusieurs modifications de la théorie sont donc imaginées pour faire face à cette contradiction empirique, et rien a priori n'empêche de remettre en question les fondements de celle-ci, à savoir la loi de gravitation newtonienne. Or, comme le note Lakatos, ce noyau dur théorique n'est remis en question qu'en dernier recours, et les scientifiques vont d'abord faire tout ce qu'ils peuvent pour rendre compte de l'anomalie au sein du cadre théorique qui est déjà en vigueur.

Cette pondération différenciée des hypothèses est bien illustrée par une autre citation de Le Verrier, suivant la précédente :

Je ne m'arrêterai pas à cette idée que les lois de la gravitation pourraient cesser d'être rigoureuses, à la grande distance à laquelle Uranus est situé du Soleil. Ce n'est pas la première fois que, pour expliquer des inégalités dont on n'avait pu se rendre compte, on s'en est pris au principe de la gravitation universelle. Mais on sait aussi que ces hypothèses ont toujours été anéanties par un examen plus approfondi des faits. L'altération des lois de la gravitation serait une dernière ressource à laquelle il ne pourrait être permis d'avoir recours qu'après avoir épuisé l'examen des autres causes, qu'après les avoir reconnues impuissantes à produire les effets observés. (LE VERRIER, 1846, p. 70)

Ainsi, contrairement aux autres possibilités qu'il a écartées, Le Verrier ne donne pas pour celle-ci de raisons *physiques*. Il se réfère plutôt à ce que l'on

pourrait appeler un argument méta-inductif, du type : à chaque fois qu'un tel problème est survenu, il a pu être résolu sans toucher aux lois fondamentales de la gravitation, donc en toute probabilité il va se produire la même chose pour le problème d'Uranus.

L'hypothèse principale de Le Verrier est donc que la loi de gravitation est toujours valable à cette échelle, et qu'une planète inconnue jusqu'alors, se trouvant au-delà de l'orbite d'Uranus, doit perturber la trajectoire de cette dernière, *précisément en accord avec la théorie newtonienne*. Le 31 août 1846, Le Verrier présente des résultats concrets prédisant la position de cette nouvelle planète, qui finira par être observée moins d'un mois plus tard à l'observatoire de Berlin par les deux astronomes Johann Gottfried Galle et Heinrich Louis d'Arrest.

La découverte de cette planète, Neptune, donna alors lieu à un conflit de paternité entre Le Verrier et John Couch Adams, un astronome britannique qui avait fait les mêmes calculs que Le Verrier au début des années 1840.

Ainsi, face à une contradiction empirique, l'ajustement consistant à faire confiance aux principes fondamentaux de la théorie a porté ces fruits. Des cas similaires ont traversé l'histoire de la physique. On peut penser par exemple aux neutrinos, des particules qui ont été prédites par Wolfgang Pauli dans les années 1930 pour expliquer une anomalie expérimentale connue depuis les années 1910 concernant la désintégration bêta (AMALDI, 1984). Pauli, à l'inverse de Niels Bohr, n'interprétait pas cette anomalie comme le signe d'une violation de la loi de conservation de l'énergie, mais comme le fait qu'une particule inconnue jusqu'alors était émise elle aussi dans la désintégration. La découverte des neutrinos a finalement été confirmées dans une expérience en 1954 (COWAN et al., 1956), donnant encore une fois raison au fait de sauver une théorie contre la réfutation.

L'exemple suivant se place dans la continuité de la découverte de Neptune. Il illustre comment une tentative qui semblait pourtant similaire en tout point n'a finalement pas abouti.

2.5.2 Une fausse piste : la planète Vulcain

Après être devenu directeur de l'Observatoire de Paris en 1854, Le Verrier reprit ses travaux sur Mercure, la planète la plus proche du soleil. Il trouva, de la même façon, une anomalie observationnelle : le périhélie de Mercure avançait trop vite (d'environ 38 secondes d'arc par siècle) par rapport à ce qui était prédit par un modèle issu de la théorie newtonienne.

En d'autres termes, Le Verrier tente de prédire la trajectoire de Mercure en faisant un modèle du système solaire composé des huit planètes connues alors, et soumis aux lois de la gravitation newtonienne, et il trouve qu'il existe un écart, faible mais significatif, entre ses prédictions et les (nombreuses) observations disponibles.

En toute rationalité, Le Verrier explore alors ce qui pour lui est l'hypothèse la plus plausible : l'existence d'une planète inconnue, entre le Soleil et Mercure, qui perturberait la trajectoire de cette dernière. Tellement plausible, que les astronomes la nomme déjà : Vulcain.

Le même schéma épistémologique est à l'œuvre ici : on tente de résoudre le problème en gardant le noyau dur de la théorie, au sein du même "paradigme". Comme on l'a vu avec Neptune, il s'agit bien d'une illustration on ne peut plus manifeste du fait que si le *modus tollens* enjoint, sur des bases logiques, à la modification de la théorie, celle-ci fait montre d'une structure si ramifiée que rien dans les seules lois de la logique ne détermine quelle est la bonne ou la meilleure modification à apporter.

Il est important de noter que mathématiquement, la solution du problème était connue. En d'autres termes, les astronomes, par la même méthode utilisée précédemment pour prédire la position de Neptune, savaient quel était le terme mathématique à rajouter au modèle pour qu'il prédise la bonne trajectoire pour Mercure. C'est l'interprétation de ce terme, à savoir l'existence d'une nouvelle planète entre Mercure et le Soleil, n'avait pas été confirmé empiriquement. Et tant qu'une telle confirmation n'était pas arrivée, le fait de connaître la solution mathématique n'était absolument pas satisfaisant en soi.

En outre, même si plusieurs astronomes, dont des amateurs comme le médecin Lescaubault, ont annoncé avoir observé Vulcain à plusieurs reprises, et même si Le Verrier lui-même l'a annoncé à l'Académie des Sciences (LEVENSON, 2015), des observations contradictoires s'accumulent. Il a fallu attendre l'émergence d'une nouvelle théorie de la gravitation, à savoir la relativité générale d'Albert Einstein en 1915, pour que la trajectoire apparemment anormale de Mercure trouve enfin une explication.

Dans ce dernier cas, il s'agissait bien des limitations intrinsèques du noyau dur de la théorie newtonienne. En effet, celle-ci n'est plus valide lorsque le champ de gravitation est trop élevé, comme c'est le cas pour Mercure qui est relativement proche du Soleil. Mais pour connaître les limitations du paradigme newtonien, il a fallu qu'une nouvelle théorie, c'est-à-dire un nouveau noyau dur, émerge et englobe la théorie précédente.

Ces deux cas, très classiques en histoire et en philosophie des sciences, illustrent parfaitement les diverses réactions possibles lorsqu'une théorie fait face à des contradictions empiriques. Comme on l'a vu avec Popper et Lakatos, conserver le noyau dur de la théorie en le protégeant par l'ajout ou la modification d'hypothèses auxiliaires s'avère souvent être l'ajustement le plus fécond. Et même s'il ne l'est pas, comme dans le cas de Vulcain, il n'en reste pas moins qu'au vu des données disponibles à un moment précis et des autres théories concurrentes, il peut rester le choix théorique le plus rationnel.

Dans la section suivante, je voudrais donner un exemple tout à fait contemporain d'une telle stratégie, pour lequel aucune solution satisfaisante ne fait encore consensus.

2.5.3 Un cas contemporain non résolu : la matière noire

En cosmologie, l'hypothèse de l'existence d'une matière "noire" ou "sombre" (*dark matter*) constitue une tentative théorique de résoudre un certain nombre d'anomalies observationnelles, c'est-à-dire des écarts entre les prédictions théoriques que l'on peut faire à partir de la théorie de la gravitation (newtonienne ou einsteinienne selon les cas) et des observations effectives.

La première observation de ce genre remonte à Fritz Zwicky (ZWICKY, 1933). Il se rendit compte que la vitesse de rotation des galaxies dans l'amas de Coma (*Coma cluster*) était trop élevée vis-à-vis de la masse totale de l'amas, insuffisante pour assurer sa cohésion gravitationnelle.

De nos jours, un grand nombre d'observations ne peut être expliqué que si l'on suppose l'existence d'une matière invisible, de nature inconnue et uniquement soumise à l'interaction gravitationnelle.

Je vais me limiter ici à un cas précis où l'introduction de la matière noire est nécessaire : le *lensing* gravitationnel. La relativité générale prédit en effet que la trajectoire des rayons lumineux est déviée significativement lorsqu'ils passent à proximité d'un objet possédant une masse assez importante. Cet effet est par exemple magistralement illustré par le phénomène d'anneaux d'Einstein, comme présenté sur la figure 2.2.

Ce que l'on voit sur la figure 2.2, c'est l'image de galaxies *en arrière plan* qui se trouve déformée par le champ de gravitation de l'amas, et qui s'étire en arcs autour de lui. C'est un phénomène optique qui est similaire à l'effet d'une lentille optique, d'où son nom de *lensing* gravitationnel.

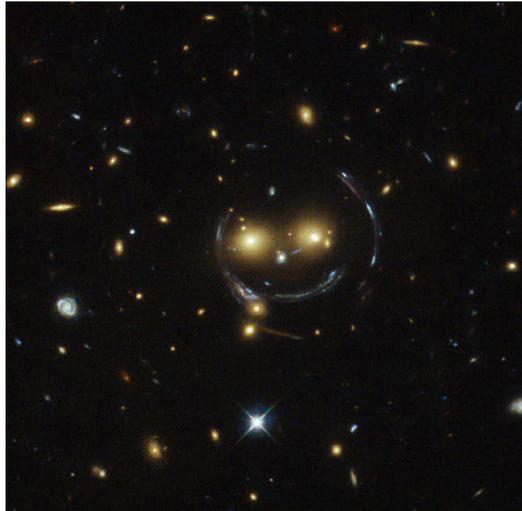


FIGURE 2.2 – Anneaux d'Einstein autour de l'amas de galaxies SDSS J1038, NASA/ESA Hubble Space Telescope, 2015.

À partir de la théorie de la relativité générale, il est alors possible de calculer la masse que l'amas doit avoir pour produire cet effet de *lensing*, c'est-à-dire pour retrouver les données d'observation. Ce phénomène donne donc une manière de mesurer la masse d'un amas de galaxies.

Or, il existe un autre moyen, plus "direct", pour calculer cette masse : à partir de sa *luminosité*. Pour cela, il faut d'abord calculer la luminosité intrinsèque de l'amas à partir de sa luminosité apparente, telle que mesurée sur Terre, et de la mesure de la distance à laquelle il se trouve de nous. Cette distance se calcule elle-même à partir d'autres types d'objets et d'évènements, à savoir les étoiles dites céphéïdes et les supernovae de type Ia.

En outre, une fois que la luminosité intrinsèque d'un amas est connue, il est possible de la relier à sa masse, puisque l'émission de lumière est directement liée à la quantité de "carburant" nucléaire disponible dans l'étoile. Pour autant, cette connexion est elle-même loin d'être évidente (MITCHELL et al., 2013) : "stellar mass is a derived quantity that can only be estimated from observational data through the application of a series of models and assumptions."

Ainsi, on dispose de deux manières de mesurer, de manière hautement *chargée de théorie*, la masse d'un amas de galaxies. Il se trouve qu'en faisant cela, un écart très élevé et systématique est trouvé entre les résultats. Plus précisément, pour que le tout soit cohérent, il faut supposer que la masse "visible" de l'amas, mesurée via sa luminosité, est au plus de l'ordre de *quinze pour cent* de la masse totale, mesurée à l'aide du *lensing*.

C'est ici qu'intervient l'hypothèse de la matière noire. Son existence est

supposée précisément pour pouvoir expliquer ces écarts sans remettre profondément en question le noyau dur de la théorie en vigueur, à savoir la relativité générale d'Einstein et les hypothèses particulières du modèle standard de la cosmologie.

Il semble très difficile, à l'heure actuelle, de savoir si cet ajustement est épistémologiquement acceptable, c'est-à-dire s'il va mener à découvrir effectivement un nouveau type de matière, ou bien si la cosmologie, en tant que programme de recherche, est dans une phase de dégénérescence.

Cette seconde possibilité est soutenue notamment par David Merritt (MERRITT, 2017), un tenant d'une théorie alternative, à savoir la théorie MOND (*MOdified Newtonian Dynamics*). Selon lui, les modèles de la matière noire effectivement utilisés pour rendre compte de ces divers écarts observationnels ont dû s'adapter petit à petit à tout un ensemble de réfutations via ce qu'il nomme, d'après Popper, des "stratagèmes conventionnalistes".

En particulier, ils semblent pouvoir s'adapter a posteriori à n'importe quelle courbe de vitesse galactique, ce qui les rend très difficilement réfutables. Pour rendre compte de ce fait, on peut envisager que la distribution de matière dans une galaxie ou dans un amas de galaxie dépend fortement de leur histoire. Dans ce cas, il faut supposer que la matière noire est "froide" (*cold dark matter*), c'est-à-dire non relativiste, pour pouvoir expliquer la formation de certaines structures en un temps donné. Cette hypothèse, en retour, contraint la masse possible des particules pouvant constituer cette matière noire, particules qui n'ont toujours pas été détectées dans les accélérateurs à notre disposition.

Ce cas est similaire, dans sa structure logico-empirique, à d'autres exemples d'anomalies observationnelles en cosmologie, comme l'expansion accélérée de l'univers ou la tension de Hubble. Ils constituent des illustrations parfaites des divers constats épistémologiques que l'on a recensés dans la section précédente : le fait que toute observation est toujours, dans une certaine mesure, chargée de théorie, et que face à une contradiction empirique, on ne peut pas savoir a priori quel élément modifier dans le vaste "tissu" de notre connaissance.

2.6 L'approche socio-historique et ses ramifications contemporaines

À partir des années soixante, les approches du problème de la démarcation scientifique se diversifient dans plusieurs directions. En particulier, un intérêt particulier est porté aux dimensions *socio-historiques* de la production de connaissances, ce à quoi cette section est dédiée.

2.6.1 La scientificité selon Kuhn

La vision de la production et du progrès scientifique de l'historien des sciences Thomas Kuhn détone radicalement avec l'approche poppérienne. Dans le tout dernier chapitre de *La structure des révolutions scientifiques*, il adopte en effet les communautés scientifiques, "ces communautés très spéciales dont provient la productivité scientifique" (KUHN, 1962, p. 273), comme unités privilégiées d'analyse. Après avoir critiqué l'énergie dépensée dans la recherche d'une définition du mot "science" et les débats que ce type d'approches suscitent, il propose un ensemble de "caractéristiques essentielles de ces groupes" :

L'homme de science doit ainsi s'occuper de résoudre des problèmes concernant le comportement de la nature. De plus, bien que son intérêt à l'égard de la nature puisse être global par sa portée, les problèmes sur lesquels il travaille doivent être des problèmes de détail. Fait plus important, les solutions qui le satisfont ne doivent pas être uniquement personnelles : il faut qu'elles soient acceptées par un groupe nombreux. Ce groupe ne peut cependant pas être tiré au hasard de la société dans son ensemble, c'est plutôt le cercle bien défini des spécialistes ayant la même activité professionnelle. L'une des règles les plus strictes, quoique non écrite, de la vie scientifique est l'interdiction de faire appel, en matière de science, aux chefs d'État ou à la masse du public. La reconnaissance de l'existence d'un groupe professionnel seul compétent et accepté comme arbitre exclusif des réalisations professionnelles a d'autres conséquences. Les membres du groupe, en tant que individus et en vertu de leur formation et de leur expérience commune, doivent être considérés comme les seuls

connaisseurs des règles du jeu ou d'un critère équivalent de jugement sans équivoque. Douter de ces critères communs d'évaluation reviendrait à admettre l'existence de critères incompatibles de la réussite scientifique. (KUHN, 1962, p. 273)

Cette caractérisation porte à la fois sur le contenu du travail du groupe des scientifiques (étudier dans ses détails le comportement de la nature) et sur le groupe lui-même, notamment en termes d'acceptation des solutions proposées aux problèmes. Cette évaluation doit alors être collective et issue du travail de personnes ayant reçu une certaine formation, partageant un certain *paradigme*, c'est-à-dire un ensemble bien déterminé de croyances considérées comme des connaissances établies, des critères communs pour reconnaître une situation comme problématique (c'est-à-dire comme une *énigme*, un problème à résoudre) et un ensemble de méthodes et de techniques communes pour résoudre ces énigmes.

Bien que cette vision ne se réduise pas à une définition sociologique tautologique de la science, comme par exemple "la science est ce que font des personnes s'identifiant et identifiées comme scientifiques", elle reçut un certain nombre de critiques portant notamment sur l'absence d'énonciation de critères rationnels de choix (KUHN, 1977, p. 320) : "Kuhn believes, I am told, that 'the decision of a scientific group to adopt a new paradigm cannot be based on good reasons of any kind, factual or otherwise' ". C'est pourquoi Kuhn revient sur ce point dans un ouvrage plus tardif, où il énonce un ensemble de critères auxquels doit satisfaire une bonne théorie scientifique :

What, I ask to begin with, are the characteristics of a good scientific theory? (...) First, a theory should be accurate : within its domain, that is, consequences deducible from a theory should be in demonstrated agreement with the results of existing experiments and observations. Second, a theory should be consistent, not only internally or with itself, but also with other currently accepted theories applicable to related aspects of nature. Third, it should have broad scope : in particular, a theory's consequences should extend far beyond the particular observations, laws, or subtheories it was initially designed to explain. Fourth, and closely related, it should be simple, bringing order to phenomena that in its absence would be individually isolated and, as a set, confused. Fifth – a somewhat less standard item, but one of special importance to actual scientific decisions – a theory should be fruitful of new research findings : if should, that is, disclose new phenomena

or previously unnoted relationships among those already known.
(KUHN, 1977, p. 321-322)

Il admet que ces cinq critères, donc : précision, consistance, portée, simplicité et fécondité, sont parfaitement standards dans la littérature, et sont effectivement probablement mis en oeuvre par les scientifiques, dans une certaine mesure, dans leur travaux quotidiens. Ils ne sont pas d'ailleurs sans rappeler les critères énoncés par Carl Hempel dans un article où il se demande ce qui fonde la signification d'un énoncé :

Instead of dichotomizing this array into significant and no significant systems, it would seem less arbitrary and more promising to appraise or compare different theoretical systems in regard to such characteristics as these :

- a. the clarity and precision with which the theories are formulated, and with which the logical relationships of their elements to each other and to expressions couched in observational terms have been made explicit;
- b. the systematic, i.e., explanatory and predictive, power of the systems in regard to observable phenomena;
- c. the formal simplicity of the theoretical system with which a certain systematic power is attained;
- d. the extent to which the theories have been confirmed by experiential evidence. (HEMPEL, 1951, p. 74)

Pour autant, Kuhn met en garde tout de suite contre deux écueils que rencontrent inévitablement de telles listes de critères : d'une part, les scientifiques qui les utilisent peuvent les interpréter différemment et donc, avec les mêmes critères, faire des choix différents ; d'autre part, ces critères sont souvent contradictoires les uns avec les autres une fois appliqués à une situation concrète où deux théories, par exemple, doivent être départagées, ce qui rend leur utilisation mal aisée, et donc leur existence même caduque.

C'est suite à l'identification de ces écueils, inévitables selon Kuhn, que ce dernier fait reposer une quelconque définition de la scientificité à un niveau sociologique, comme un ensemble de normes partagées par un groupe social donné, c'est-à-dire des règles communes mais dont une grande partie sont *implicites*. C'est pour autant le fait que ces règles, ces normes soient communément partagées et servent une évaluation critique collective des propositions théoriques qui fonde la confiance que l'on peut déléguer dans la science : "In the absence of criteria able to dictate the choice of each individual,

I argued, we do well to trust the collective judgment of scientists trained in this way."

Le point de vue kuhnien rapidement résumé ici a donné lieu à de multiples développements dont certaines ramifications sont encore vivaces aujourd'hui. Il laisse en particulier déjà entrevoir deux chemins menant, suivant (FERNANDEZ-BEANATO, 2020), à *l'anarchie* et la *capitulation*.

Le premier terme fait bien sûr référence à la position "anarchiste" de Paul Feyerabend développée dans (FEYERABEND, 1975), ouvrage qui devait à l'origine être un débat avec Lakatos, mais auquel ce dernier n'a pas pu participer dû à son décès prématuré en 1974. Feyerabend y défend une position épistémologique anarchiste, c'est-à-dire où non seulement la mise en évidence et (d'autant plus) l'imposition de méthodes et de critères de choix rationnel de la part des scientifiques sont probablement vaines, mais où elles seraient même, en général, néfastes à la science.

Le second terme fait référence à la position défendue en 1983 par Larry Laudan dans un article que j'ai déjà cité plus haut (LAUDAN, 1983). Il y développe également l'idée que la recherche d'une définition de la science ou de la scientificité est vouée à l'échec. D'une part, aucun ensemble de critères individuellement nécessaires et collectivement suffisants pour caractériser théoriquement ce qui est scientifique et ce qui ne l'est pas n'a jusqu'ici été mis en évidence. D'autre part, le terme de "science" ou "scientifique" recouvre des éléments si hétérogènes qu'une telle démarche est intrinsèquement, sinon impossible, du moins inintéressante :

Some scientific theories are well tested ; some are not. Some branches of science are presently showing high rates of growth ; others are not. Some scientific theories have made a host of successful predictions of surprising phenomena ; some have made few if any such predictions. Some scientific hypotheses are ad hoc ; others are not. Some have achieved a 'consilience of inductions' ; others have not. (Similar remarks could be made about several non-scientific theories and disciplines.) The evident epistemic heterogeneity of the activities and beliefs customarily regarded as scientific should alert us to the probable futility of seeking an epistemic version of a demarcation criterion. Where, even after detailed analysis, there appear to be no epistemic invariants, one is well advised not to take their existence for granted. But to say as much is in effect to say that the problem of demarcation - the very problem which Popper labelled 'the central problem of epistemology' - is

spurious, for that problem presupposes the existence of just such invariants. (LAUDAN, 1983, p. 124)

Il semble alors bien plus intéressant, selon Laudan, de concentrer les efforts sur la question de savoir ce qui fait qu'une croyance est bien fondée, fiable ou féconde, et non sur ce qui fait qu'elle est (ou non) scientifique.

2.6.2 *Les sciences studies*

L'idée que la recherche d'invariants qui pourraient définir la scientificité est vaine, ou en d'autres termes que la science, comme tout autre production socio-culturelle, "n'a pas d'essence" (PESTRE, 2006) qui soit indépendante des conditions socio-historique de sa production, rappelle les positions contemporaines fondamentales des *science studies*. Ce dernier terme recouvre des champs de recherche héritiers de plusieurs courants se réclamant de l'étude sociale des sciences, nés dans les années 1970. Ils se placent à la fois en opposition à la philosophie des sciences classique héritière par exemple de Popper ou du cercle de Vienne, et à la sociologie des sciences mertonienne alors dominante à l'époque.

On peut, globalement, distinguer quatre aspects ou tendances caractéristiques de ces courants (DUBOIS, 2001, p. 28).

Premièrement, la focalisation sur les pratiques concrètes des scientifiques vus comme des acteurs sociaux plongés dans un certain contexte historique et soumis à certaines normes sociales.

Deuxièmement, l'analyse de l'activité scientifique comme intrinsèquement assortie à un certain cadre conceptuel ou "sociocognitif" permettant aux scientifiques de donner du sens à leurs objets d'étude. L'impossibilité de penser la production de connaissances indépendamment de son contexte historique est justifiée notamment par la pluralité d'interprétations qui peuvent parfois exister d'un même concept ou d'une même controverse, et mener même à un relativisme cognitif – la justification d'une théorie, par exemple, étant dépendante du cadre sociocognitif au sein duquel se trouvent les scientifiques qui y adhèrent ou l'utilisent.

Troisièmement, la prise en compte de l'interdépendance d'une grande quantité de facteurs (cognitifs, sociaux, politiques, ...) dans la production et la justification des productions scientifiques, et le refus de traiter ces facteurs de manière indépendante les uns des autres. Ce positionnement se traduit par l'abolition de certaines distinctions classiques, comme par exemple entre les différentes disciplines, entre science et technique, ou encore entre la science

fondamentale et la science appliquée. Du point de vue épistémologique, on pourrait également rajouter l'abolition, pour les chercheurs se réclamant de ces courants, de la distinction entre le contexte de découverte et le contexte de justification des théories.

Enfin, quatrièmement, ces courants se caractérisent également par l'accent mis sur les conséquences (notamment sociales) de la science, ou encore la responsabilité politique des scientifiques (en particulier, des applications éventuellement néfastes de leurs découvertes).

Ainsi, les études sociales de la science sont héritières, d'une part, de certaines considérations théoriques, en particulier kuhniennes, sur la dépendance intrinsèque de tout discours scientifique à un paradigme donné. Elles sont, d'autre part, héritières également du contexte historique de leur naissance, à savoir les années 1970 et la remise en question politique de la toute puissance de "la science" et de ses promesses en tant que nécessairement chargée de progrès humain et social.

Pour autant, différents courants co-existent et, tout en se réclamant de cet héritage, ne mettent pas l'accent sur les mêmes aspects parmi les quatre cités plus haut, et sont parfois en désaccord sur certains points.

L'école d'Edimbourg et son *programme fort* (BLOOR, 1976), par exemple, proposent d'interpréter la résolution d'une controverse scientifique non pas comme la victoire d'arguments rationnels convaincants échangés au sein d'une communauté partageant certaines règles de débat, mais bien plutôt comme le reflet des rapports de force sociaux au sein de ladite communauté, conduits par les "intérêts sociaux" de ses membres.

L'approche socioconstructiviste, représentée en France par Bruno Latour, met, de son côté, davantage l'accent sur l'étude *in situ* des pratiques scientifiques. On peut par exemple penser à l'ouvrage *La vie de laboratoire* (LATOUR et WOOLGAR, 2006), parangon de cette approche, dans lequel Bruno Latour et Steve Woolgar retranscrivent leur analyse de terrain dans un laboratoire de biologie de l'Institut Salk en Californie, où ils ont enquêté tels que le feraient deux anthropologues à propos des pratiques et des coutumes d'un peuple inconnu.

On peut qualifier, pour des raisons de facilité, l'ensemble de ces approches de *constructivistes*. Il faut garder en tête, toutefois, qu'en faisant cela on regroupe sous le même terme, encore une fois, des courants différents qui se sont aussi construits en opposition les uns aux autres, comme le montre brillamment par exemple Michel Dubois (DUBOIS, 2001, p. 40).

Ces approches, néanmoins, ont en commun de s'opposer à une vision

strictement *rationaliste* du progrès de la connaissance. Pour les rationalistes, la rationalité est une faculté universelle, également partagée par tous les êtres humains, une capacité à être convaincu par des arguments ou à accepter un argument comme étant valide qui est indépendante du contexte historique, social, institutionnel, ou culturel. Ce qui est dépendant de ces contextes, par contre, c'est la façon avec laquelle cette rationalité s'incarne effectivement. C'est cela qui explique que certaines théories pourtant "bonnes" mettent parfois beaucoup de temps à s'imposer. D'un point de vue rationaliste, elles étaient déjà des "bonnes" théories même lorsque seulement une minorité de personnes en étaient convaincues. D'un point de vue constructiviste, le fait qu'une théorie soit meilleure qu'une autre ne peut pas avoir un sens qui soit indépendant du contexte socio-historique de son évaluation.

Une sociologie des sciences rationaliste à la Merton, par exemple, ne nie absolument pas que des facteurs externes à la science interviennent abondamment dans le processus de production de connaissance et dans leur acceptation à un moment donné. Tout simplement, une telle vision considère qu'on ne peut pas expliquer le succès scientifique *uniquement* par ce type de raisons sociologiques.

Une sociologie des sciences constructiviste, au contraire, soutient qu'on ne peut pas distinguer fondamentalement la rationalité (entendue comme une démarche de tri entre différentes options) de la façon avec laquelle elle s'incarne dans un contexte historique, culturel ou social donné. Pour autant, cela ne veut pas nécessairement dire que tout se vaut et qu'il faille forcément embrasser une certaine forme de relativisme. En effet, à un instant donné, les scientifiques sont munis de certains critères, pour certains implicites, qui leur permettent de faire leur choix. C'est simplement que ces critères sont vus comme n'ayant rien d'universel, et sont également situés socialement, dans le sens où ils sont également emprunts des normes sociales et des valeurs ayant cours dans la société au moment où ils sont utilisés.

En d'autres termes, les constructivistes, pour la plupart, reconnaissent que le réel s'exprime dans la science, mais c'est simplement qu'il ne s'exprime pas de manière univoque. Le choix entre différentes interprétations possibles n'est alors pas uniquement guidé par des valeurs épistémiques mais également par des valeurs non épistémiques, par exemple (et consciemment ou non) par des préférences idéologiques.

Du point de vue de notre cadre d'analyse conceptuelle, les science studies se caractérisent principalement par un changement d'unité d'analyse, en affirmant qu'on ne peut pas parler de la scientificité d'une hypothèse ou d'une

théorie indépendamment de son contexte concret de production, et de la façon avec laquelle elle est manipulée par des scientifiques. En d'autres termes, que les unités d'analyse habituellement considérées comme pertinentes en philosophie des sciences (comme les théories scientifiques) sont, au mieux, très limitées, et, au pire, tout à fait caduques.

Bien qu'elles soient parvenues à acquérir une certaine légitimité dans le champ de la philosophie des sciences et de l'épistémologie, les *science studies* reposent sur des prémisses et soutiennent des thèses qui ne font pas nécessairement l'unanimité pour autant. L'ouvrage déjà cité de Michel Dubois, *La nouvelle sociologie des sciences* (DUBOIS, 2001), propose un exposé historique du développement de ces différents courants depuis les années 1970, de leurs différentes positions philosophiques ainsi que des thèses qu'ils ont été amenés à défendre.

Un reproche important que l'on peut retenir à l'encontre de ces courants, sans que cela n'enlève quoi que ce soit, d'ailleurs, à l'intérêt intrinsèque que peuvent contenir leurs diverses productions, est la tendance qu'ils ont à se définir par des oppositions artificielles. En d'autres termes, ils se présentent parfois comme soulevant une problématique n'ayant pas été posée avant alors que le plus souvent, ils proposent simplement *une résolution différente* de ces problèmes.

On peut penser, par exemple, à la sociologie mertonienne. D'une part, Merton lui-même ainsi que ses supporters n'ont pas du tout interprété le travail de Kuhn comme une critique des fondements de sa propre approche sociologique mais bien comme un apport historique très riche ne rentrant pas en contradiction avec la sienne mais qui doit y être au contraire intégré. Selon Dubois, les multiples interprétations différentes (et contradictoires) des thèses kuhniennes sont probablement dues à l'absence de définition claire et univoque de ses différents concepts de base, à commencer par celui de paradigme.

De plus, Merton, comme n'importe quel sociologue des sciences, est bien conscient que des facteurs externes interviennent dans la production de connaissances scientifiques par une communauté de chercheurs. En témoignent par exemple ses propres travaux à propos de l'effet Matthieu, qui caractérise le fait qu'un chercheur va recevoir du prestige en fonction du prestige qu'il a déjà accumulé, menant parfois à l'amplification artificielle et auto-réalisatrice de simples différences aléatoires, et donc à la sur-reconnaissance, dé-corrélée de leur qualité réelle de chercheur, de certaines personnes au détriment d'autres.

Comme le montre également Dubois, il y a une certaine cyclicité des idées en philosophie et les différents constats épistémologiques censés être le fondement du renouveau de la philosophie et de la sociologie des sciences dans les années 1970 peuvent déjà se retrouver à la fin du dix-neuvième siècle, en particulier dans les différentes écoles dites "conventionnalistes", d'Edouard Le Roy à Pierre Duhem en passant par Henri Poincaré. Bien entendu, s'ils font le même constat concernant, entre autres, la place des conventions ou la marge de manœuvre tolérée par les données dans les constructions théoriques, ils n'en tirent pas du tout les mêmes conclusions.

De même, Popper est tout à fait conscient, dès le début, des différents constats épistémologiques que l'on a énumérés plus haut. En particulier, il construit son critère de falsifiabilité de telle façon à ce que ça réponde, justement, aux critiques éventuelles qui pourraient venir des conventionnalistes.

De plus, il reconnaît même que tout critère de sélection des théories est forcément basé sur des valeurs, la rationalité ou la recherche de la vérité elles-mêmes étant des valeurs à part entière :

La question n'est donc pas simplement que l'objectivité et l'absence de jugement de valeur sont pratiquement hors de portée de l'homme de science isolé, mais que l'objectivité et l'absence de jugement de valeur sont en elles-mêmes des *valeurs*. Et comme l'absence de jugement de valeur est en elle-même une valeur, l'exigence d'une absence absolue de jugement de valeur est un paradoxe. (...) il faut tout de même remarquer que le paradoxe disparaît entièrement de lui-même si nous remplaçons l'exigence d'absence de jugement de valeur par cette exigence selon laquelle l'une des tâches de la critique scientifique doit être de mettre à jour les confusions de valeur et de séparer les questions de valeur purement scientifiques (...) des questions de valeur extra-scientifiques. (ADORNO, 1979, p. 84)

Non seulement, pour Popper, l'objectivité est elle-même une valeur, mais l'existence de telles valeurs est nécessaire dans la production de connaissance, puisque ce sont elles qui guident les chercheurs dans leur travail. Toutefois, l'effet de ces valeurs doit être neutralisé par le débat collectif, la "critique scientifique" comme citée plus haut. C'est précisément pour cette raison que Popper reconnaît que le processus de sélection théorique est un processus éminemment social contenant également des aspects politiques, comme en témoigne cette citation fameuse que j'ai déjà énoncée plus haut :

[...] l'objectivité de la science n'est pas une question d'individu [...] mais une question sociale qui résulte de leur critique mutuelle, de la division du travail amicale-hostile entre scientifiques, de leur collaboration autant que de leur rivalité. Elle dépend donc partiellement d'une série de conditions sociales et politiques qui rendent cette critique possible. (ADORNO, 1979, p. 82)

Enfin, sur le sujet des controverses, on peut par exemple se référer à l'ouvrage *Sociologie des controverses scientifiques* de Dominique Raynaud (RAYNAUD, 2018) qui aborde un certain nombre de controverses scientifiques d'un point de vue rationaliste. Plus précisément, il teste le modèle normatif mertonien sur une petite dizaine d'exemples historiques, dont la controverse opposant Pasteur et Pouchet à propos de la génération spontanée, exemple abondamment traité par les tenants du programme fort.

Loin de négliger les éléments externes intervenant dans le travail collectif des scientifiques, Raynaud montre que si ces éléments interviennent, ils ne déterminent généralement pas l'issue de la controverse, au contraire de ce que prétend (par principe) l'approche constructiviste.

Le but de cette courte présentation simplifiée des *science studies* est d'affirmer mon opposition, dans ce travail de thèse, à leurs principales prémisses méthodologiques. Ce qui m'intéresse, c'est chercher à reconstruire rationnellement la structure et le contenu des productions scientifiques, en focalisant sur les produits et pas nécessairement sur le processus de production lui-même ou sur les producteurs.

Une telle démarche exige toutefois, comme on va le voir, d'intégrer dès le début l'ensemble des constats épistémologiques critiques que l'on a mis en exergue plus haut. Encore une fois, cela n'implique pas de nier qu'un grand nombre de facteurs externes, comme des valeurs non épistémiques ou des biais implicites, rentrent activement en ligne de compte quand il s'agit de juger concrètement de la scientificité d'une théorie, ou de sa valeur épistémologique. Ils jouent même un rôle causal nécessaire à prendre en considération dans le processus historique d'acceptation d'une théorie. Simplement, je considère que cela n'implique pas d'abolir la distinction *méthodologique* classique entre l'étude des aspects propres à la *découverte* scientifique et celle des modes de *justification* intrinsèques des productions scientifiques.

2.7 Les champs épistémiques et les approches multi-critères

Malgré ces divers courants pessimistes quant à la pertinence de travailler sur un modèle général de scientificité d'unités cognitives, la recherche de tels critères n'a jamais vraiment cessé. Dans les années 1980, on peut noter par exemple les travaux de Mario Bunge qui se place à l'échelle de ce qu'il nomme les *champs épistémiques* (*epistemic fields*), ou encore *family of factual scientific research fields* (BUNGE, 1983a; BUNGE, 1991). Comme le synthétise par exemple Martin Mahner (MAHNER, 2007, p. 523) : "Roughly speaking, an epistemic field is a group of people and their practices, aiming at gaining knowledge of some sort." Dans le même article, suivant Bunge, Mahner définit plus précisément un champ épistémique E comme un 10-uplet :

$$E = \langle C, S, D, G, F, B, P, K, A, M \rangle,$$

où :

1. C est la communauté de chercheurs ;
2. S est la société dans laquelle les activités de C prennent place ;
3. D est le domaine ou l'univers de discours des membres de C ;
4. G est le background philosophique partagé par les membres de C , qui consiste en (a) une ontologie (une vision générale de la nature des choses), (b) une épistémologie (une vision générale de la nature de la connaissance), et (c) une méthodologie, une axiologie et une moralité (une vision générale à propos des façons correctes d'acquérir de la connaissance) ;
5. F est le background formel (règles mathématiques par exemple) partagé par les membres de C ;
6. B est un background spécifique, un ensemble de connaissance emprunté à d'autres champs épistémiques ;
7. P est une collection de problèmes concernant D , G ou F ;
8. K est un fond de connaissance commun ;
9. A est un ensemble de buts cognitifs, pratiques ou moraux poursuivis par les membres de C ;
10. M est un ensemble de méthodes générales et spécifiques utilisées par les membres de C pour étudier les éléments de D .

On voit donc que l'unité d'analyse choisie par Mario Bunge est une structure complexe et surtout hétérogène, puisqu'elle implique à la fois des groupes

sociaux (C, S), des objets du monde (D), des éléments cognitifs (G, F, B, P, K, A) eux-mêmes de types divers, et des méthodes concrètes (M). À partir de cette unité d'analyse sous forme de 10-uplet, Bunge peut ensuite définir les champs épistémiques *scientifiques*, ou *factuels*, comme des champs épistémiques satisfaisant approximativement un ensemble de conditions portant sur chaque élément du 10-uple E (MAHNER, 2007, p. 525-536).

C'est ce qui poussa Mahner, par analogie avec la difficile question de la définition d'espèce en biologie, à développer une approche de la scientificité vue comme un *cluster* (amas, regroupement) de propriétés (MAHNER, 2007; MAHNER, 2013). Une telle approche par *clustering*, en biologie, permet d'avoir une définition utile de ce qu'est une espèce sans pour autant avoir à s'engager quant à son statut ontologique. De la même façon, cette approche utilisée en épistémologie permet de caractériser utilement des champs scientifiques vis-à-vis de champs non scientifiques ou pseudo scientifiques, sans forcément avoir à supposer une quelconque *essence* à la science.

La définition de la scientificité dans le programme de recherche esquissé par Mahner repose sur la satisfaction d'un certain nombre de critères, par exemple sept critères sur dix, laissant ainsi 176^2 façons possibles, pour un champ épistémique, d'être scientifique.

Ce programme de recherche a récemment été concrétisé par Damian Fernandez-Beanato (FERNANDEZ-BEANATO, 2020; FERNANDEZ-BEANATO, 2021; FERNANDEZ BEANATO, 2022) dans son travail de thèse de doctorat. Dans (FERNANDEZ-BEANATO, 2021), il juge par exemple de la (non) scientificité du Feng Shui sur la base d'une liste de 36 critères : "Predictive power, General legality, Explanatory legality, Empirical adequacy, Has been replicated, High degree of theoretical confirmation, Does not use unfounded immunizations, Connectivity, Contiguity, Bulk external consistency, Uses, and abides by, evidence, Does not use cherry-picking, Factual screening and sorting, It is evidenced, Methodical scientificity, Independent recognizability of correctness, Confirmations and disconfirmations, Checks for variation of variables, Systemicity, Criticalness, Revises or checks, Considers its disconfirmations, Reactive dynamicity, It is peer reviewed, Avoidance of cognitive errors and biases, Concern for evidentiary criteria, Lack of appeal to mysteries, Does not create mystery, Rationality, Ethos, Willingness to test, Disposition to comply, Does not evade scientific peer review, Promise, Progressiveness, Self-correctiveness." Ces 36 critères sont décrits précisément les uns après les

2. Il y a $176=10!/(7!3!)$ façons de choisir 7 éléments (discernables) dans un ensemble de 10 éléments.

autres (pp. 6–10), et sont reliés à leurs diverses occurrences dans la littérature en philosophie des sciences.

2.8 La théorie meehilienne

Une tentative de prolonger les travaux de Popper et Lakatos qui s'est développée dans les années 1980 et qui vaut le coup d'être notifiée est la théorie de Paul Meehl de la corroboration–vérisimilitude (MEEHL, 1990; CAMPBELL, 1990). Originellement appliquée aux tests d'hypothèses en psychologie, la portée théorique est cependant plus large que ce champ d'applications particulier.

Meehl voit les théories comme des réseaux inter-connectés d'énoncés, dans le même style que la description que l'on en a donné plus haut. Il décrit le test empirique d'une "théorie substantielle" T comme ayant la structure logique suivante (MEEHL, 1990, p. 109) :

$$(T.A_t.C_p.A_i.C_n) \rightarrow (O_1 \supset O_2) \quad (2.1)$$

où :

- T est la théorie à tester,
- A_t est une conjonction de théories auxiliaires nécessaires pour en dériver des observations,
- C_p est une clause *ceteris paribus* (toute chose égale par ailleurs),
- A_i est une théorie auxiliaire ayant trait aux instruments de mesure,
- C_n énonce les conditions détaillées dans lesquelles l'expérience est réalisée,
- \rightarrow dénote la déduction logique,
- $O_1 \supset O_2$ est une prédiction conditionnelle, symbolisée par l'implication matérielle \supset , et signifie que si l'on observe O_1 , alors on observera O_2 .

Ce modèle simple a le bénéfice d'illustrer concrètement un point que l'on a déjà soulevé plus haut et qui était, comme on l'a vu, connu de Popper lui-même comme une faiblesse de sa théorie : si on observe une réfutation de la prédiction (à droite de la flèche), par exemple $O_1 \cdot \neg O_2$, alors le *modus tollens* implique la réfutation du membre de gauche, mais rien ne peut nous dire quel élément de la conjonction $T.A_t.C_p.A_i.C_n$ il faut modifier ou abandonner. Meehl rappelle alors la stratégie lakatosienne de défense d'une théorie : on sauve le noyau dur, représenté ici par T , de la réfutation en sacrifiant des

hypothèses auxiliaires. Son apport consiste alors à proposer deux principes garantissant que cette stratégie est payante.

Le premier, qu'il nomme le principe de "l'argent à la banque", ou tout simplement principe de Lakatos, stipule que l'on fait bien de ne pas rejeter une théorie si celle-ci a accumulé une certaine quantité de succès passés (MEEHL, 1990, p. 115) : "we are warranted in continuing to conjecture that a theory has high verisimilitude when it has accumulated "money in the bank" by passing several stiff tests."

Le second principe, qu'il nomme principe des "coïncidences sacrément étranges" ("damn strange coincidences"), ou encore principe de Salmon en référence aux travaux de Wesley Salmon (SALMON, 1984), énonce qu'une théorie accumule de "l'argent à la banque" en faisant des prédictions risquées, c'est-à-dire en prédisant des faits qui sont effectivement observés mais qui auraient été hautement improbables hors de la théorie T .

Il formalise alors cette idée à l'aide de plusieurs concepts, en prenant pour contexte le cas où la prédiction se réduit à celle de la valeur d'une variable réelle x . Premièrement, il définit le *Spielraum* S comme étant le domaine a priori de valeurs que peut prendre x , indépendamment de la théorie que l'on veut tester. Ensuite, il définit plusieurs autres intervalles (MEEHL, 1990, p. 128-129) :

- I : l'intervalle toléré par la théorie T – c'est-à-dire la gamme de valeurs, dans S , prédite par la théorie T ,
- I/S : la tolérance relative de la théorie T ,
- $In = 1 - (I/S)$: l'intolérance de T ,
- D : déviation entre la valeur observée x_0 et le bord le plus proche de l'intervalle de tolérance – en d'autres termes, l'erreur commise dans la prédiction,
- D/S : l'erreur relative,
- $Cl = 1 - (D/S)$: la proximité ("closeness").

Bien entendu, on désigne ici par S, I, D, \dots à la fois les intervalles et leur taille, ce qui donne du sens à des expressions comme I/S ou D/S . Ainsi, on peut opérationnaliser une notion plus fluide d'adéquation empirique, et de ne pas se restreindre aux cas binaires (vérifié/réfuté) qui font passer à côté de certains aspects intéressants et surtout plus proche de la réalité. Meehl définit alors l'indice de corroboration C_i pour une telle expérience particulière comme étant égale au produit de l'intolérance et de la proximité :

$$C_i = (In) \times (Cl) \tag{2.2}$$

L'idée de prendre le produit est sous-tendue par le fait que les deux quantités In et Cl sont en compétition l'une contre l'autre, un peu comme la simplicité et la force dans le *best system account* des lois, comme on va le voir plus bas. Et en effet, plus l'intolérance est forte (donc plus la gamme de valeurs permises par la théorie est réduite relativement à la taille totale de S), plus l'erreur sera probablement élevée, et inversement : si l'intolérance est faible, alors la proximité sera élevée, mais un peu trivialement, en quelque sorte.

Le degré de corroboration forgé par Meehl transmet donc ce qui pour moi constitue l'enseignement fondamental du falsificationnisme, à savoir qu'une bonne théorie scientifique est une théorie qui a raison alors qu'elle aurait très bien pu avoir tort. Un bon degré de corroboration doit pouvoir capturer cette intuition en la formalisant.

2.9 La scientificité comme optimisation

Dans les sections précédentes, on a commencé à pressentir que non seulement la scientificité était davantage une question de degré que le résultat d'une catégorisation stricte, mais surtout qu'il semble fécond de la percevoir comme l'optimisation, sous certaines contraintes, d'une certaine quantité. Tout le défi repose alors sur la bonne définition de cette quantité.

D'autres approches qui proposent de définir une certaine propriété via un processus ou comme un résultat d'optimisation, même si elles n'ont pas été explicitement conçues pour répondre au problème de la démarcation scientifique, me semblent intéressantes à citer ici, comme des sources fécondes d'inspiration.

2.9.1 Le *Best System Account* des lois

Le *best system account* (BSA) entend fournir une définition formelle d'une *loi scientifique*, pour distinguer cette dernière d'une simple généralisation contingente. On trouve des prémisses de ce point de vue chez Mill :

Dans une autre manière de parler, la question : Que sont les lois de la nature ? peut se traduire ainsi - Quelles sont les suppositions les plus simples et les moins nombreuses qui, étant accordées, tout l'ordre existant de la nature en résulterait ? On pourrait encore l'exprimer autrement en disant : Quelles sont les propositions générales les moins nombreuses desquelles pourraient

être inférées déductivement toutes les uniformités de la nature ?
(MILL, 1866,???)

Cependant, c'est la version contemporaine de David Lewis (LEWIS, 1973 ; LEWIS, 1983 ; LEWIS, 1994) que je me propose de présenter ici. Un système théorique est vu comme une structure axiomatique au sein de laquelle se confrontent deux tendances : un système peut être plus ou moins *simple* (par exemple, selon le nombre d'axiomes sur lequel il est construit), et il peut être plus ou moins *fort* (selon le nombre de faits expérimentaux qu'il parvient à expliquer). Ces deux aspects sont a priori contradictoires : d'une part, on peut augmenter artificiellement la simplicité en diminuant le nombre d'axiomes, mais on va probablement perdre en force, puisque l'on va pouvoir en déduire moins de faits empiriques ; d'autre part, on peut aussi augmenter le nombre d'axiomes pour pouvoir expliquer davantage de faits, mais le gain de force se traduira alors par une perte de simplicité.

Pour le BSA, une loi scientifique est un axiome ou un théorème du *meilleur* système, c'est-à-dire de celui qui possède le meilleur équilibre entre simplicité et force, en d'autres termes l'équilibre *optimal*.

D'une certaine façon, cette vision peut être reliée à la sous-détermination théorique, décrite plus haut : une fois donnés un ensemble de faits empiriques, on reconnaît que plusieurs configurations théoriques sont toujours possibles, selon les énoncés (par exemple, les définitions de base) que l'on considère plus fondamentaux que les autres. Les lois de la logique seules ne pouvant pas départager en elles-mêmes différentes configurations théoriques, on peut alors choisir la configuration qui optimise le produit entre simplicité et force. Il s'agit donc d'un principe (ici, de légalité scientifique) qu'il faut rajouter "à la main" et qui ne peut pas en lui-même se déduire des lois de la logique seules.

Le BSA, qui a été beaucoup discuté dans la littérature, rencontre une difficulté majeure : celle consistant à définir ses concepts de base (simplicité, force et équilibre entre les deux) d'une manière qui soit à la fois claire, opérationnelle et indépendante des systèmes théoriques en question – pour pouvoir effectivement comparer plusieurs systèmes entre eux.

Pour autant, l'intuition sous-jacente reste très stimulante, et a donné lieu à des développements intéressants allant dans le sens d'une telle opérationnalisation. Par exemple, on peut citer les travaux récents de Billy Wheeler (WHEELER, 2018 ; WHEELER, 2019) où il propose une vision des lois comme

des algorithmes de compression de données. Un tel point de vue lui permet d'utiliser la théorie algorithmique de l'information ("algorithmic information theory"), et en particulier une notion de simplicité définie à partir de la *complexité de Kolmogorov*, une notion centrale de cette théorie. Soit un énoncé x que l'on retrouve comme output d'un programme p et d'input y (ici, on pourrait penser à un ensemble de données que l'on souhaite retrouver à partir d'une loi empirique et de conditions initiales). La complexité de Kolmogorov de x est la taille l de la plus petite paire (p, y) (programme plus input) redonnant x comme output : $K(x) = l(p) + l(y)$. En d'autres termes, la loi la plus simple pour décrire des données est celle qui en fournit la plus courte description.

2.9.2 Le critère d'information d'Akaike

On peut également faire le parallèle avec des critères de sélection de modèles statistiques comme le *critère d'information* développé par Hirotogu Akaike (AKAIKE, 1998) dans les années 1970. Il s'agit ici encore une fois d'opérationnaliser la notion de simplicité ou de parcimonie en voyant le meilleur modèle comme celui qui a le meilleur rapport coût/bénéfice. En d'autres termes, on peut améliorer un modèle en augmentant sa fonction de vraisemblance ou *likelihood* L – qui correspond à la probabilité du modèle vis-à-vis d'un certain nombre de données – par l'ajout de paramètres supplémentaires. Cette modification a toutefois un coût, mesuré par le nombre de paramètres k que l'on rajoute. Le critère d'Aikake consiste alors à choisir le modèle qui minimise la quantité :

$$AIC = 2k - 2\log(L). \quad (2.3)$$

Le contraste est saillant entre le BSA de Lewis et le critère d'information d'Akaike. Le premier cas est très général (il s'applique à tous les systèmes théoriques vus comme structure axiomatique) mais souffre rapidement, du même coup, d'un manque d'opérationnalisation. Le second cas est, quant à lui, très spécifique (il ne s'applique qu'à un type bien identifié de modèles) mais fournit dans le même temps un outil explicite pour effectivement sélectionner entre différents modèles.

2.9.3 L'utilité épistémique et la *concept utility*

Enfin, une dernière approche qui me semble pertinente à évoquer ici est celle consistant à évaluer les théories en fonction de leur *utilité épistémique* (HUBER, 2008). L'idée sous-jacente à cette approche est qu'une théorie doit à la fois être *plausible* (vis-à-vis des faits connus et des autres théories acceptées) et *informative* (nous apprendre quelque chose de substantielle sur le monde, que l'on ne savait pas déjà).

Encore une fois, ces deux aspects sont contradictoires. Par exemple, on peut augmenter la plausibilité d'une théorie en y ajoutant des énoncés trivialement vrais, mais on fera ainsi baisser du même coup son *informativité* ("informativeness"), et inversement. L'utilité épistémique d'une théorie (mais plus généralement d'une croyance) mesure l'équilibre entre ces deux aspects, et la meilleure théorie ou croyance est celle qui maximise cet équilibre.

En particulier, des auteurs ont proposé récemment de quantifier, dans la même lignée que l'utilité épistémique, l'utilité associée à une conceptualisation, c'est-à-dire à la classification d'un ensemble d'objets formant un domaine D en un nombre fini de concepts (C_1, C_2, \dots, C_N) (EGRÉ et O'MADAGAIN, 2019). On peut considérer un concept comme un sous-ensemble de D , et une conceptualisation (C_1, C_2, \dots, C_N) comme une *partition* de D , vérifiant donc $D = \cup_i C_i$ avec $C_i \cap C_j = \emptyset$.

Les auteurs s'attachent alors à formaliser l'intuition suivante : d'une part, pour être *informative*, une conceptualisation doit pouvoir s'appliquer au plus de cas concrets possibles, c'est-à-dire doit posséder une certaine *inclusivité*; d'autre part, pour que les inférences que l'on puisse tirer à propos d'objets soient *plausibles*, il faut que les objets regroupés sous un certain concept jouissent d'une certaine *homogénéité*.

De la même façon ici, l'idée sous-jacente est que pour juger de la qualité d'une conceptualisation, deux aspects contradictoires sont à prendre en compte, et que la meilleure conceptualisation est celle qui donne l'équilibre optimal entre les deux. Ici, si on augmente l'inclusivité d'un concept en incluant par exemple davantage d'objets, alors probablement qu'ils vont faire montre d'une certaine hétérogénéité qui va diminuer la qualité des inférences qu'on pourra y associer. D'un autre côté, si on regroupe les objets dans des catégories très homogènes alors elles seront probablement de taille plus réduite. La meilleure conceptualisation est celle qui fournit les quantités optimales d'inclusivité en dégradant le moins possible l'homogénéité, et vice versa.

L'inclusivité d'un concept C est définie comme la proportion des objets de D qui sont regroupés sous le concept C , que je note donc :

$$Inc(C) = \frac{|C|}{|D|}, \quad (2.4)$$

où $|X|$ dénote le cardinal de l'ensemble X .

Quant à l'homogénéité, les auteurs considèrent que les différents objets de D peuvent posséder un certain nombre (fini) de propriétés, des *features* F_1, F_2, \dots, F_n vus comme des variables binaires ($F_j \in \{0, 1\}$). L'homogénéité d'un concept C vis-à-vis d'une propriété F , notée $Hom(C, F)$, est définie comme le nombre maximum de points communs, concernant F , entre les différents objets du concept. En d'autres termes, on prend la plus grande valeur entre la proportion des objets dans C qui possèdent la propriété F et la proportion des objets dans C qui ne la possèdent pas.

En considérant une propriété comme un sous-ensemble de D , c'est-à-dire $F = \{x \in D | F(x) = 1\}$, l'homogénéité d'un concept C vis-à-vis d'une propriété F est alors la plus grande valeur entre $\frac{|F \cap C|}{|C|}$ et $1 - \frac{|F \cap C|}{|C|}$:

$$Hom(C, F) = \max\left(\frac{|F \cap C|}{|C|}, 1 - \frac{|F \cap C|}{|C|}\right). \quad (2.5)$$

L'homogénéité d'un concept C vis-à-vis d'un ensemble de propriétés $\mathcal{F} = (F_1, F_2, \dots, F_n)$ est alors la *moyenne* de l'homogénéité de C pour chaque propriété. La *concept utility* d'un concept vis-à-vis d'un ensemble de propriétés \mathcal{F} est alors définie comme le produit de son inclusivité et de son homogénéité (EGRÉ et O'MADAGAIN, 2019, p. 24) :

$$U(C, \mathcal{F}) = Inc(C) \times Hom(C, \mathcal{F}) \quad (2.6)$$

La *concept utility* d'une conceptualisation, c'est-à-dire d'une partition d'un domaine D en un ensemble fini de concepts est alors la moyenne de la *concept utility* de tous les concepts. Parmi toutes les partitions possibles de l'ensemble D en différents concepts, la meilleure est définie comme celle qui maximise la quantité 2.6. Dans leur papier, les auteurs utilisent ensuite cette construction théorique sur le cas concret de la recatégorisation de la planète Pluton en planète naine par l'*International Astronomical Union* en 2006.

Avant de conclure, notons bien que ces différents développements théoriques (le BSA, le critère d'information d'Akaike, la *concept utility*, ...) ne visaient pas à résoudre spécifiquement le problème de la démarcation scientifique. Pour autant, ils fournissent une source importante d'inspiration pour

mon travail, étant donné que je développe dans cette thèse une vision de la scientificité comme l'optimisation d'une certaine quantité, sous contraintes.

2.10 Conclusion

Au cours des cinquante dernières années, la question de la démarcation scientifique a donc énormément muté suite à un ensemble de critiques à la fois de types logique et socio-historique des différentes tentatives de démarcation.

En particulier, un certain nombre de constats ont été établis concernant les difficultés qu'il faut prendre en considération lors de toute élaboration d'un concept de scientificité :

LA CHARGE THÉORIQUE DE L'OBSERVATION

Le cadre empirique au sein duquel on produit des données auxquelles on souhaite comparer de manière critique nos suppositions théoriques repose lui aussi sur d'autres présupposés théoriques qui doivent être tenus pour vrais pour que les données aient un sens exploitable.

LE HOLISME DE LA CONFIRMATION

Un énoncé empirique n'est jamais testé seul face à des données empiriques "absolues", l'unité qui est mise à l'épreuve est, a minima, un système d'énoncés possédant une certaine structure logique, et qui inclue le cadre empirique.

LA COMPLEXITÉ STRUCTURELLE LOGICO-EMPIRIQUE

L'ensemble de notre connaissance est un système dont la structure logique est hautement complexe, ce qui implique que la justification d'un énoncé particulier fait appel à tout un ensemble d'autres énoncés appartenant à d'autres modèles, éventuellement à d'autres théories voire à d'autres disciplines.

LA SOUS-DÉTERMINATION DE LA THÉORIE

Face à une contradiction empirique, et en application du *modus tollens*, rien dans la théorie ni dans les données ne peut nous indiquer ce qu'il faut modifier dans l'ensemble complexe de notre connaissance pour résoudre la contradiction.

LES LIMITES ÉPISTÉMOLOGIQUES DE LA LOGIQUE

La logique seule ne suffit pas à justifier notre connaissance scientifique.

LA SOCIO–HISTORICITÉ

Le fait d'attribuer à une unité épistémique le qualificatif de "scientifique" est hautement tributaire de contingences socio–historiques, comme les données empiriques disponibles à un moment donné, les autres hypothèses ou théories sur le marché à ce moment–là, ou bien les rapports de force sociaux et les intérêts non épistémiques pesant dans le choix théorique.

LA DIACHRONICITÉ

La scientificité d'une unité cognitive a aussi un aspect *diachronique* inhérent, dans le sens où c'est souvent après une certaine durée que l'on peut connaître sa valeur scientifique : la frontière est parfois floue entre une théorie pseudo–scientifique et une proto–théorie scientifique, c'est-à-dire une théorie qui doit (et peut) arriver à un certain stade de maturation pour satisfaire les critères de scientificité.

Pour répondre à ces difficultés, un certain nombre de stratégies ont été mises en œuvre. Au sein du cadre d'analyse conceptuel que j'ai décrit en introduction, ces diverses réponses se caractérisent par plusieurs aspects.

On a pu modifier *le cadre conceptuel*, passant de {SCIENCE, PSEUDO–SCIENCE} à {SCIENCE, NON SCIENCE, PSEUDO–SCIENCE, PROTO–SCIENCE}, par exemple.

Le domaine conceptuel, c'est-à-dire l'unité d'analyse sur laquelle on s'est attelé à définir la scientificité, a beaucoup varié également, notamment en se complexifiant : les énoncés singuliers, les systèmes d'énoncés, les théories, les communautés, les pratiques, voire les champs épistémiques entiers – qui forment, à ma connaissance, l'unité d'analyse la plus vaste et bien définie que l'on trouve dans la littérature.

Les diverses tendances des *science studies* partent même du principe qu'une définition de la scientificité est non pertinente à rechercher à l'échelle des unités cognitives uniquement, c'est-à-dire sans prendre en compte le contexte socio–historique de sa production.

On a également complexifié *le type de définition intensionnelle*, en passant d'un petit nombre de critères, voire d'un mono–critère (la falsifiabilité), à des approches multi–critères, ou en reconnaissant à la scientificité une certaine forme de *gradation* difficile à capturer aussi facilement.

La théorie meehilienne de la corroboration–vérisimilitude est un exemple de sophistication de l'unité d'analyse sur laquelle un certain degré de corroboration est défini de telle sorte à voir la scientificité comme une *maximisation* de ce degré de corroboration. On trouve des points de vue similaires dans le *best system account* des lois scientifiques, par exemple, bien qu'ils n'aient pas

été développés au départ pour répondre à la question de la scientificité.

Face à la pluralité conceptuelle autour du concept de SCIENTIFICITÉ, on a également adopté des stratégies diverses, comme l'élimination (la recherche d'une définition de la scientificité comme pseudo-problème) ou le pluralisme (l'acceptation que plusieurs conceptions de ce qu'est une science peuvent co-exister).

On a reconnu enfin que le qualificatif "scientifique" ne semble pas n'avoir qu'une dimension descriptive mais a également une dimension normative, de la même façon que le fait de qualifier des travaux ou des personnes de "pseudo-scientifiques" n'est pas uniquement une démarche descriptive mais poursuit certains objectifs, comme disqualifier ou dé-légitimer un discours.

Face à ce constat, on peut, comme le font par exemple les *science studies*, renoncer à la possibilité de pouvoir distinguer entre ces deux dimensions, voire même revendiquer qu'il n'est pas pertinent de les distinguer.

Le problème de la démarcation scientifique, sous quelque forme que ce soit, reste un projet de clarification conceptuelle intéressant à mener à bien. Dans l'introduction du *Handbook Philosophy of pseudoscience : Reconsidering the demarcation problem* (PIGLIUCCI et BOUDRY, 2013), les éditeurs appuient sur un point qui me semble aller dans ce sens, et à côté duquel semble être justement passé Laudan. Cela concerne la pertinence de la démarche de recherche elle-même : même si on n'est pas parvenu à une réponse définitive concernant la distinction entre science et pseudo-science, *on a beaucoup appris en chemin* concernant la science, à la fois dans ses aspects structurels, c'est-à-dire logico-empiriques, que socio-historiques. Ainsi, *la démarche est féconde* quand bien même aucune solution définitive n'a été atteinte pour le moment.

Les réflexions sur le problème de la démarcation scientifique que j'ai exposées dans ce chapitre n'ont cependant pas abordé en profondeur une importante question *transversale* qui est de savoir dans quelle mesure les critères épistémologiques invoqués dépendent de la discipline étudiée. Par exemple, est-ce que la scientificité d'une discipline peut être définie indépendamment du type d'objet qu'elle investigate? Plus spécifiquement, est-il possible de trouver des critères épistémologiques auxquels peuvent se soumettre de la même manière les sciences sociales que les sciences physiques? Quels sont les obstacles a priori d'une telle démarche et existe-t-il des pistes pour les surmonter? Le chapitre suivant va explorer ces questions en profondeur et tenter de leur apporter des éléments de réponse.

Chapitre 3

Les sciences sociales et leur(s) épistémologie(s)

3.1 Introduction

Les *sciences sociales* recouvrent des disciplines aussi diverses que la sociologie, l'anthropologie, l'histoire, la linguistique, les sciences politiques ou encore l'économie.¹ La liste n'est pas exhaustive, et c'est sans compter les multiples subdivisions pouvant exister au sein de chaque catégorie. Dans toute la suite de cette thèse, je me concentrerai plus spécifiquement sur la *sociologie* et je traiterai peu voire ne traiterai pas du tout des autres disciplines. Je suis toutefois convaincu que les questionnements épistémologiques développés ici pourront s'y appliquer également.

Si le type de réflexions que les sciences sociales recouvrent ont probablement toujours existé dans l'histoire de la pensée humaine, leur constitution en tant que disciplines scientifiques est relativement récente. Les sciences économiques, par exemple, et avec elles la "théorie de la décision", remontent à la fin du dix-huitième siècle. La "physique sociale", quant à elle, émerge au début du dix-neuvième siècle pour devenir "sociologie" sous la plume d'Auguste Comte en 1830 (COMTE, 1830).

Après deux siècles d'existence, ces disciplines se présentent sous des formes extrêmement diverses. Se reconnaissent comme *sciences sociales*, en effet : des approches sociographiques très descriptives n'ayant aucune prétention à la généralisation ; des travaux se basant sur l'utilisation extensive de statistiques à plus ou moins grande échelle ; des tentatives d'expérimentations directes en laboratoire ou sur le terrain ; des productions principalement conceptuelles, d'aspect très littéraire, dont le but est l'explicitation du

1. Je ne parlerai pas de la psychologie et de ses différentes sous-disciplines comme de sciences sociales, même si je pourrai parfois les évoquer à l'appui de certains arguments, et même si la frontière n'est pas toujours nette.

sens ou de la signification du vécu des acteurs en présence via l'interprétation et la compréhension; l'application de modèles explicatifs mathématisés, voire de simulations numériques; etc.

Si ce constat d'hétérogénéité des productions est unanime dans le champ, les positions divergent par contre fortement sur la *signification* à lui donner. En particulier, un point de dissension important porte sur la question de savoir dans quelle mesure cet état est passager dans l'histoire des sciences sociales. Celles-ci vont-elles, comme la plupart des autres champs disciplinaires, finir par voir le nombre de leurs cadres théoriques se réduire, tandis que leurs méthodes empiriques et leurs épistémologies s'harmoniseront? Ou bien, au contraire, cet état est-il permanent, c'est-à-dire propre à ces disciplines en tant qu'elles ont pour objet des phénomènes humains, sociaux et culturels, donc intrinsèquement hétérogènes car historiques, et dont le sens est toujours à reconstruire au sein de cadres conceptuels eux-mêmes indexés temporellement?

Ces discussions ont accompagné toute l'histoire des sciences sociales, sous différentes formes, et les protagonistes sont loin de se diviser en deux camps bien distincts – il faut donc faire très attention ici aux éventuels raccourcis binaires qui pourraient nous faire passer à côté d'une complexité significative.

Plus précisément, cette pluralité est visible sur (au moins) trois aspects importants :

MÉTHODOLOGIQUE : Comment rentrer en contact avec le réel social? Ici, de nombreuses méthodes de natures complètement différentes coexistent : témoignages d'archive, différents genres d'entretiens plus ou moins directs, des observations participantes, du traitement de données statistiques, des expériences de laboratoire ou de terrain, ...

THÉORIQUE : Comment organiser les faits et phénomènes humains et sociaux dans un ensemble logiquement structuré? Ici, un grand nombre de "courants" théoriques, voire d'écoles de pensée coexistent également : structuralisme, fonctionnalisme, holisme, individualisme, interactionnisme, ...

ÉPISTÉMOLOGIQUE : À quoi doit ou peut ressembler une connaissance scientifique du monde social, et dans quel but? Les objectifs épistémiques des sciences sociales semblent aussi nombreux que les cadres théoriques : la description la plus précise et loyale possible d'une situation sociale ou au contraire la mise en évidence de lois générales, la compréhension des motivations des acteurs sociaux dans certaines

situations particulières ou au contraire l'élaboration de cadres explicatifs systématiques des phénomènes collectifs, ...

Remarquons tout de suite qu'une telle diversité n'est pas un problème en soi. Par exemple, que des méthodes empiriques ou des constructions théoriques différentes soient appliquées à différents types de phénomènes semble tout à fait naturel – c'est le cas dans toutes les autres disciplines scientifiques.

En effet, les productions scientifiques en elles-mêmes font toujours montre d'une grande hétérogénéité, quelle que soit la discipline. En physique, par exemple, on peut penser à la différence entre un article en cosmologie observationnelle, comme par exemple (AMENDOLA et al., 2018), issu d'une grande collaboration comme *Euclid*, regroupant des dizaines d'auteurs, et dont le but est de produire un certain nombre d'observations bien définies pour répondre à une question bien identifiée (l'origine de l'accélération de l'expansion de l'univers), et un article de physique théorique comme par exemple (RUSSO, SUSSKIND et THORLACIUS, 1992), signé par trois auteurs, et qui étudie le phénomène d'évaporation des trous noirs dans un espace-temps de dimension $1+1$.² C'est un exemple extrême mais assez révélateur de l'hétérogénéité possible de productions scientifiques qui pourtant appartiennent toutes à la même discipline.

Que plusieurs cadres théoriques co-existent n'a rien non plus de bien problématique a priori. Pour reprendre l'exemple de la physique, depuis à peu près un siècle deux grands cadres théoriques principaux – la physique quantique d'une part, la relativité générale d'autre part – co-existent, sont féconds dans leur domaine d'application respectifs tout en décrivant des phénomènes bien différents dans leur nature et leur fonctionnement. Certes, cette situation est perçue comme insatisfaisante par nombre de physiciens qui tentent, depuis Albert Einstein, de concilier les deux dans des théories de *gravité quantique* (toutes restées à un stade hautement spéculatif jusqu'à présent). Cependant, cela n'empêche pas de considérer unanimement chaque théorie, dans son domaine propre d'application, comme étant la *meilleure* théorie disponible jusqu'à preuve du contraire. Cette situation pluri-paradigmatique ne pose donc pas fondamentalement de problème quand il s'agit d'utiliser une théorie particulière pour une classe de phénomènes particuliers.

Cependant, dans les sciences sociales, cette situation pluri-paradigmatique pose des questions d'une nature distincte. En particulier, il y a selon moi au

2. Sachant que ce phénomène en lui-même est déjà spéculatif, il faut se souvenir de plus que l'espace-temps usuellement utilisé en physique est de dimension $3+1$.

moins trois raisons de considérer cet état de chose comme étant problématique : l'absence de critères méta-théoriques (ou épistémologiques) clairs, la défense de la nécessité d'un concept distinct et propre de *scientificité*, et la difficulté à accumuler de la connaissance.

Il n'est pas rare, comme on va le voir dans la partie 3.2, de trouver en sciences sociales des approches issues de cadres théoriques différents qui tentent de s'appliquer à des situations identiques et, tout en étant (à peu près) d'accord sur les constats empiriques, se retrouvent en désaccord sur les causes (par exemple) de ce constat, ou plus généralement sur la façon la plus pertinente (dans un sens qui est propre à chaque approche) de rendre compte du phénomène en question. Appartenant à des cadres théoriques différents, et donc reposant sur des prémisses distinctes, il peut sembler assez difficile de les comparer et donc de les départager rationnellement, par exemple en se fondant sur des éléments de preuve empirique.

Il ne s'agit donc pas uniquement d'un débat portant sur la meilleure théorie ou le meilleur modèle à appliquer à telle ou telle situation, comme il peut en exister ailleurs et qui traduit d'ailleurs un certain degré de santé au sein d'un champ scientifique donné. Il s'agit plus profondément d'un débat *épistémologique*, c'est-à-dire une dissension sur les critères mêmes à appliquer lorsqu'il s'agit de choisir la meilleure théorie ou le meilleur modèle. Or, un désaccord sur ce point a des conséquences bien plus problématiques. En effet, sans critères clairs de choix théoriques, comment faire pour trancher, collectivement, entre différentes approches d'un même phénomène proposant des explications contradictoires ?

Cette situation étant parfois perçue comme normale ou en tout cas non problématique, certains auteurs ont régulièrement tenté de la justifier en développant des argumentaires défendant pour ces disciplines la nécessité d'une épistémologie à part, c'est-à-dire d'une définition propre de la scientificité. Suivant (RAYNAUD, 2006), on peut qualifier cette position de *régionaliste* (chaque discipline scientifique a sa propre épistémologie), ou encore de *dualiste* (il existe une différence *essentielle* entre les sciences de la nature comme la physique et les sciences sociales comme la sociologie).

Conséquemment, une telle dissension épistémologique favorise la coexistence passive de cadres théoriques parmi lesquels le chercheur en sciences sociales peut être appelé à "faire son choix" (FABIANI, 2013). Outre le fait qu'en l'absence de critères épistémologiques consensuels, un tel choix semble gagner en arbitraire, une autre conséquence claire d'une telle situation d'instabilité conceptuelle est la limitation que cela engendre quant à la cumulativité

des connaissances. Or, la cumulativité des connaissances est un élément fondamental, à la fois cause et conséquence d'une division sociale efficace du travail scientifique.

Ce débat peut être perçu comme une situation de pluralité conceptuelle, dans laquelle plusieurs versions, parfois contradictoires, d'un même concept (ici "la science" ou "la scientificité") coexistent néanmoins. Cette pluralité se perçoit entre différentes disciplines, bien sûr, mais également au sein même des sciences sociales. En témoignent par exemple les divers *schèmes d'intelligibilité* recensés par Jean-Michel Berthelot dans son ouvrage *L'intelligence du social* (BERTHELOT, 1990), et dont le sous-titre *Le pluralisme explicatif en sociologie* est parlant pour notre propos. Il y distingue sept schèmes différents : le schème *causal*, le schème *fonctionnel*, le schème *structural*, le schème *herméneutique*, le schème *actanciel* et enfin le schème *dialectique*, un schème étant défini comme "une matrice d'opérations permettant d'inscrire un ensemble de faits dans un système d'intelligibilité, c'est-à-dire, d'en rendre raison ou d'en fournir une explication (au sens non restrictif)" (BERTHELOT, 1990, p. 23).

On l'a vu dans l'introduction de cette thèse, suivant (CARRANANTE, 2023), au moins quatre réactions différentes peuvent être identifiées face à une telle pluralité : la sélection (un concept est choisi et les autres sont supprimés), l'unification (un nouveau concept est élaboré qui unifie les concepts existants), l'élimination (tous les concepts existants sont éliminés et un nouveau cadre conceptuel est éventuellement envisagé), et le pluralisme (la coexistence de différents concepts est acceptée, éventuellement sous certaines conditions).

Les positions défendant un monisme épistémologique, c'est-à-dire la non nécessité de fonder les sciences sociales sur une épistémologie "à part" voire même "dérogatoire" (RAYNAUD, 2006), peuvent correspondre aux réactions de type "sélection" (on impose l'épistémologie d'une discipline particulière, par exemple celle de la physique, à toutes les autres) ou "unification" (on construit un concept de scientificité qui unifie les concepts existants).

Les positions "régionalistes" ou "dualistes", quant à elles, correspondraient davantage à une certaine acceptation de la pluralité conceptuelle. C'est pourquoi on qualifiera également ces positions de "pluralistes" dans la suite de ce manuscrit.

Dans ce chapitre, mon objectif est de soutenir, contre ces positions pluralistes, qu'il n'y a pas de raison de principe pour que les sciences sociales (et en particulier la sociologie) jouissent d'une épistémologie singulière. Pour

cela, je vais procéder en trois temps.

Premièrement, dans la section 3.2, je vais illustrer les diverses tensions citées plus haut (méthodologiques, théoriques et épistémologiques) sur quelques exemples concrets de travaux en sciences sociales. En particulier, j'insisterai sur des cas pour lesquels des cadres essentiellement différents tentent de s'appliquer à la description d'un même phénomène, posant la question épistémologique fondamentale des critères de choix théoriques dans une telle situation. Cela me permettra de mieux préciser pourquoi ce type de situations peut être considéré comme problématique.

Dans la section 3.3, ensuite, je tenterai de présenter le plus clairement possible la position *pluraliste*. Les partisans de ce point de vue admettent éventuellement que cette situation puisse être considérées comme problématique, mais argumentent suivant l'idée qu'il ne semble pas possible de faire autrement, étant donné les spécificités des sciences sociales vis-à-vis des sciences naturelles. Ils tentent alors de développer une authentique épistémologie alternative à celle ayant cours ailleurs, afin de *justifier* cet état de fait.

Je présenterai alors ces arguments catégorisés en quatre grandes familles, selon quatre propriétés principales qui sont souvent considérées comme spécifiques aux sciences sociales : l'historicité (des phénomènes), la signification (que les acteurs donnent aux situations et phénomènes qu'ils vivent), l'intentionnalité (des acteurs), et la performativité (des connaissances du social sur le social). J'illustrerai le tout par un grand nombre d'exemples.

Ces épistémologies "pluralistes" forment elles-mêmes un ensemble relativement hétérogène. Je me suis alors efforcé d'en présenter une version *idéaltypique* sous la forme d'un système philosophique adoptant de manière cohérente des positions ontologique, épistémologique et méthodologique. Un tel *idealtyp*, même s'il ne correspond pas nécessairement à une position précise ayant été adoptée tel quel dans la littérature, permet au moins de fixer les idées sur un modèle simple et de pouvoir raisonner à partir de celui-ci. Je me suis efforcé le plus possible à ne pas caricaturer les positions vis-à-vis desquelles je me place en désaccord, même si leur hétérogénéité rend la tâche quelque peu ardue.

Enfin, dans la section 3.4, j'argumenterai, à l'inverse, en faveur d'une vision unitaire de la scientificité ou, à la suite de (RAYNAUD, 2006), d'un "monisme épistémologique". Je présenterai les diverses réponses classiques que l'on peut opposer aux arguments pluralistes. Aucune ne nie que les sciences humaines et sociales, et en particulier leurs objets, aient des particularités et des spécificités vis-à-vis des objets des autres disciplines. La ligne

argumentative principale consiste cependant à considérer que ces spécificités n'impliquent pas nécessairement une spécificité épistémologique, c'est-à-dire une manière singulière, particulière à ces disciplines, de considérer la connaissance que l'on peut avoir de ces objets.

3.2 Une pluralité problématique ?

L'objectif de cette première partie est d'illustrer la dissension épistémologique, évoquée plus haut, sur un certain nombre d'exemples. En particulier, j'aimerais ici illustrer trois cas qui me semblent représentatifs des différentes dimensions dans lesquelles cette dissension se déploie habituellement.

Le premier cas, que je vais présenter en section 3.2.1, est celui où des auteurs tombent d'accord sur l'existence d'un phénomène mais l'abordent de manières profondément distinctes. Je vais prendre l'exemple de deux travaux récents en sciences politiques portant sur le même sujet, à savoir le lien de causalité éventuel entre interventionnisme militaire et terrorisme. Le premier se réfère à des traditions sociologiques relevant de l'*interprétation par compréhension*, tandis que le second s'applique à mettre en évidence le phénomène à l'aide de méthodes *quantitatives*.

Dans la section 3.2.2 j'aborderai un exemple où ce n'est pas autant sur les méthodes empiriques elles-mêmes que se situe la dissension, que sur le but (et donc la forme) de l'investigation ainsi que sur le mode d'explication à mobiliser. J'illustrerai ce point avec le cas de la réussite scolaire socialement différenciée, et plus particulièrement deux familles d'approches classiques en sociologie. La première est celle représentée initialement par le travail de Bourdieu et Passeron, et si elles reposent sur des analyses en partie quantitatives, elles relèvent également d'une certaine forme d'interprétation par compréhension, c'est-à-dire d'une forme plus qualitative et qui n'a pas le même objectif. La seconde est basée sur le paradigme explicatif de l'individualisme méthodologique, longtemps représenté en France par Raymond Boudon. Dans son cas, l'explication prend la forme d'un modèle qui tente de dériver une observation "macro", au niveau statistique, à partir de l'agrégation de comportements "micro", c'est-à-dire individuels.

Enfin, le dernier cas, présenté dans la section 3.2.3, me permettra d'exemplifier le débat qui peut exister à propos de l'utilisation d'outils formels en sociologie. Je présenterai une étude de Dominique Raynaud mobilisant des éléments de la théorie des réseaux complexes (ou *théorie des graphes*), et surtout la réaction d'un autre chercheur, appartenant à une tradition sociologique

différente, à propos de ce travail. Selon moi, cette opposition particulière de points de vue illustrent à la perfection une incompréhension fondamentale à la base de la dissension épistémologique plus générale que je tente de circonscrire.

3.2.1 Approches qualitatives et quantitatives

Dans un numéro récent de la revue de sciences politiques *Cultures & Conflits*, un certain nombre d'auteurs et d'autrices ont abordé, à travers une dizaine d'articles, la question du lien entre les interventions militaires de pays occidentaux et leur propension à être frappés d'attaques terroristes. En particulier, ils défendent la thèse dite du "pompier pyromane", à savoir que les interventions militaires occidentales dans certains pays du moyen orient, fussent-elles justifiées par leur défenseurs par la nécessité de combattre le terrorisme, sont en réalité des causes de celui-ci – en particulier, des causes des attaques terroristes sur le sol occidental.

En plus de l'intérêt intrinsèque d'une telle question, ce qui m'intéresse particulièrement ici est que les contributeurs de ce numéro, bien qu'ils soient à peu près d'accord sur la thèse globale défendue, divergent radicalement dans leur approche du problème, et notamment sur les postulats épistémologiques et les outils méthodologiques qu'ils mettent en œuvre pour poser une telle question et y répondre. Cela va me permettre toutefois de mettre en exergue le fait que des approches appartenant a priori à la même discipline (les sciences politiques), et abordant le même phénomène (le lien entre les interventions militaires et le terrorisme) peuvent pourtant avoir des formes, des buts, des méthodes et des modes de justification profondément différents.

Cette divergence est particulièrement illustrée par les deux premières contributions, sur lesquelles je vais me focaliser ici. La première : *La rivalité mimétique, une matrice de la guerre contre le terrorisme et de ses stratégies discursives?* (BIGO et al., 2021) signée par Didier Bigo, Laurent Bonnefoy, Mathias Delori, Anastassia Tsoukala et Christophe Wasinski, met en avant une approche basée sur "l'interprétation par la compréhension". La seconde : *Les interventions militaires sont-elles une cause du "terrorisme"?* (DELORI et al., 2021) signée par Mathias Delori, Clara Egger, Raül Magni-Berton et Simon Varaine, met quant à elle en avant "l'utilité des analyses quantitatives pour les études critiques de la sécurité".

La première contribution citée offre une approche que l'on pourrait qualifier de "qualitative", basée sur l'analyse des discours et des pratiques des acteurs en présence et de leurs interactions. Le cadre conceptuel choisi ici est celui de la "rivalité mimétique" développé par l'anthropologue René Girard, et appliqué au cas précis du terrorisme et des interventions militaires censées le combattre. Le concept de rivalité mimétique n'est pas précisément défini tel quel dans l'article, mais suggère une construction symétrique dans les discours de justification ainsi que dans les pratiques concrètes des deux parties en présence, à savoir les groupes terroristes d'une part, et les États occidentaux interventionnistes d'autre part. De ce point de vue, les discours adoptés vont empêcher de percevoir la violence de l'autre comme légitime et dans le même temps légitimer la sienne propre en tant que réponse justifiée à la première :

Les discours (...) jouent un rôle central dans la construction de la rivalité mimétique car ils empêchent les acteurs de comprendre le caractère relationnel de la violence qu'ils perpétuent et de celle qu'ils reçoivent en retour. La violence, à leurs yeux, apparaît comme le seul résultat de la violence primaire de l'adversaire, que celle-ci soit attribuée à l'hostilité présumés des "croisés" à l'égard des musulmans ou, selon l'autre discours, au fanatisme des "djihadistes". (BIGO et al., 2021, p. 28)

En d'autres termes, chacun des deux partis en présence va développer un certain discours pour justifier ses propres actions, sans voir que ces dernières ne peuvent en réalité se comprendre qu'en termes *relationnel*, c'est-à-dire comme définies uniquement vis-à-vis des actions de l'autre parti. Cela mène alors à une escalade de la violence des deux cotés, et ainsi, des interventions militaires qui étaient censée combattre le terrorisme peuvent très bien se retrouver le causer au moins en partie.

La seconde contribution, que je souhaiterais mettre en parallèle à la première ici, utilise des méthodes statistiques pour quantifier précisément à quel point les interventions militaires sont une cause du terrorisme. Ici, la méthode ne consiste pas à se focaliser sur les discours de justification produits par les deux partis en présence, mais bien plutôt à prendre un certain recul – permis par les outils statistiques – pour voir s'il on peut détecter, *toute chose étant égale par ailleurs*, une corrélation positive entre l'interventionnisme d'un pays occidental et les attaques terroristes sur son sol.

Une première étape fondamentale est le choix d'une définition opérationnelle pour les concepts d'interventionnisme et surtout de terrorisme : quel

type d'intervention militaire (ou d'attaque terroriste) est compté? à partir de combien de soldats sur place parle-t-on d'intervention militaire? une fois ces questions-là fixées, quelle base de données est utilisée, et pourquoi?

On voit ici à l'œuvre la notion de charge théorique de l'observation, déjà évoquée : toute observation, toute production de données repose forcément sur des hypothèses, des définitions, certaines conventions que l'on adopte plutôt que d'autres, et non pas sur des variables qui reflèteraient d'elles-mêmes parfaitement la réalité.

Les auteurs ne perçoivent d'ailleurs pas cela comme un problème épistémologique fondamental :

Aucune méthode d'enquête n'offrant une représentation parfaite de la réalité, la question pertinente n'est pas de savoir si ces indicateurs sont neutres et objectifs – ils ne le sont pas – mais si les biais qu'ils contiennent sont explicites, font l'objet d'une analyse critique par les chercheurs qui mobilisent ces données chiffrées et dans quelle mesure ils influencent les résultats obtenus. (DELORI et al., 2021, p. 39)

Ils choisissent d'ailleurs de considérer le "terrorisme" comme étant un concept "épais", c'est-à-dire relevant tout à la fois d'une dimension descriptive et d'une dimension normative. Ils utilisent pour cela des bases de données spécifiques, qui définissent le terrorisme du point de vue des gouvernements en place qui souhaitent le combattre, donc comme un objet "politiquement construit".

Cela ne leur pose aucun problème non plus puisque c'est précisément ce que ces auteurs cherchent à mesurer :

En d'autres termes, le fait que le thermomètre ne mesure pas la chaleur objective n'est pas un problème. Il mesure très bien le sentiment de chaleur produit par un groupe d'acteurs particulier, or c'est précisément ce qui nous intéresse. Un usage réflexif de cette littérature est donc non seulement possible mais également utile et urgent pour comprendre les origines du phénomène que le discours dominant appelle, en Europe et en Amérique du Nord, "terrorisme". (DELORI et al., 2021, p. 48)

Une fois le choix de certains indicateurs justifié, les auteurs montrent alors que la littérature quantitative sur cette question, foisonnante à l'international depuis plus d'une vingtaine d'années mais largement méconnue dans la littérature francophone, est relativement *consensuelle* : une fois contrôlées

les variables de confusion potentielles, persiste une forte corrélation entre le degré d'interventionnisme militaire d'un État et sa probabilité d'être touché par une attaque terroriste, suggérant par là même une relation *causale* entre les deux.

Par exemple, citant les résultats de (COLLARD-WEXLER, PISCHEDDA et SMITH, 2014) :

Toutes choses égales par ailleurs (PIB, population, régime politique, religion des auteurs de l'attaque, religion majoritaire du pays, etc.), la probabilité pour un État de voir ses troupes ou ses ressortissants visés par un attentat-suicide est 17 fois plus grande pour les pays "occupants" que pour les pays "non-occupants".

(DELORI et al., 2021, p. 51)

Cet exemple permet d'illustrer la différence profonde qu'il peut exister entre différentes approches d'un même phénomène se réclamant pourtant de la même discipline. Les deux contributions reposent malgré tout sur un constat à peu près identique, à savoir une certaine relation de causalité entre l'interventionnisme militaire et le terrorisme. Mais qu'en serait-il si les deux approches avaient mis en exergue des causes différentes ?

C'est ce cas-là que je voudrais aborder à présent, illustré par la célèbre controverse entre Pierre Bourdieu et Raymond Boudon à propos des causes de la réussite scolaire.

3.2.2 Différents types d'explications

La question de la réussite scolaire socialement différenciée et celle, fortement corrélée, du rôle de l'institution scolaire dans la reproduction des inégalités socio-économiques, ont été *abondamment* traitées dans la littérature sociologique des cinquante dernières années.

En France, le débat a été longtemps structuré par l'opposition entre deux grandes approches. L'une, dont les représentants principaux sont Pierre Bourdieu et Jean-Claude Passeron (BOURDIEU et PASSERON, 1984; BOURDIEU et PASSERON, 1970), est basée sur la prévalence explicative du concept d'*habitus*, de dispositions sociales incorporées et de distribution différenciée du capital culturel (adjointe à celle du capital économique). Quant à l'autre, représentée par Raymond Boudon (BOUDON, 1973; BOUDON, 1990), son cœur explicatif prend la forme de modèles basés sur l'hypothèse de rationalité des acteurs sociaux, qui font les meilleurs choix possibles en fonction des ressources (cognitives, économiques, etc.) à leur disposition.

Bien que dans les décennies suivantes, les travaux sur la question se soient quelque peu diversifiés (en intégrant notamment d'autres types de variables, comme le sexe ou la trajectoire migratoire (BLANCHARD et CAYOUILLE-REMBLIÈRE, 2017)), je m'en tiendrai ici à une comparaison entre les deux approches originelles, citées plus haut. D'une part, la littérature consacrée à cette question est très vaste et en faire une présentation exhaustive serait hors de portée du présent travail. D'autre part, une comparaison entre ces deux approches contient certaines tensions épistémologiques fondamentales qui se retrouvent de toutes façons dans les travaux plus tardifs qui s'en réclament.

Tout d'abord, il faut mettre en exergue un point important : en dépit de leurs différences fondamentales, ces deux approches tombent généralement d'accord sur le constat empirique de base qui est qu'il existe un lien très fort entre la réussite scolaire d'un élève et son origine sociale. Outre les débats méthodologiques portant sur la meilleure façon d'opérationnaliser ces deux variables, les distinctions se portent davantage sur la signification, en termes notamment de causalité, du lien entre elles, ainsi que sur le traitement théorique qu'il s'agit d'opérer vis-à-vis d'un tel constat. Ce sont ces deux points précis sur lesquels je vais me concentrer ici.

Le constat empirique

Le travail de Bourdieu et Passeron met en lumière, à grand renfort de statistiques, un constat assez clair : la réussite scolaire en général, et l'accès à l'université en particulier, est fortement dépendante de l'origine sociale. En effet :

Un fils de cadre supérieur a quatre-vingts fois plus de chances d'entrer à l'université qu'un fils de salarié agricole et quarante fois plus qu'un fils d'ouvrier ; ses chances sont encore le double de celles d'un fils de cadre moyen. (BOURDIEU et PASSERON, 1984, p. 12)

Ayant contrôlé l'effet d'autres variables, ils constatent que :

De tous les facteurs de différenciation, l'origine sociale est sans doute celui dont l'influence s'exerce le plus fortement sur le milieu étudiant, plus fortement en tout cas que le sexe et l'âge et surtout plus que tel ou tel facteur clairement aperçu, l'affiliation religieuse par exemple. (BOURDIEU et PASSERON, 1984, p. 22)

Pour autant, selon eux, ces inégalités devant l'école ne sont pas uniquement dues à une différence d'accès aux ressources matérielles, autrement dit à une différence de capital économique, mais *avant tout* à une différence d'accès à des ressources symboliques comme le capital culturel :

Les obstacles économiques ne suffisent pas à expliquer que les taux de "mortalité scolaire" puissent différer autant selon les classes sociales. N'en aurait-on aucun autre indice et ignorerait-on les voies multiples et souvent détournées par lesquelles l'École élimine continûment les enfants originaires des milieux les plus défavorisés, on trouverait une preuve de l'importance des obstacles culturels que doivent surmonter ces sujets dans le fait que l'on constate encore au niveau de l'enseignement supérieur des différences d'attitudes et d'aptitudes significativement liées à l'origine sociale, bien que les étudiants qu'elles séparent aient tous subi pendant quinze à vingt années l'action homogénéisante de l'École et que les plus défavorisés d'entre eux n'aient dû qu'à une plus grande adaptabilité ou à un milieu familial plus favorable d'échapper à l'élimination. (BOURDIEU et PASSERON, 1984, p. 19, 22)

Ainsi, selon ces auteurs, la preuve que la différence de capital *culturel* est prépondérante dans la production des inégalités devant l'école se lit dans le fait que même si on ne regarde que les élèves ayant réussi à atteindre l'université, on observe encore une différence d'aptitude liée à l'origine sociale.

L'approche à laquelle je souhaiterais comparer celle-ci, à savoir celle développée par Boudon, mobilise également un grand nombre de données statistiques. En particulier, son travail sur le sujet des inégalités devant l'école (BOUDON, 1973 ; BOUDON, 1990) se base sur une enquête de l'INED (GIRARD, 1970) datant des années 1970 portant sur l'orientation, au début des années 1960, d'un grand échantillon d'élèves sortant de l'école primaire et auxquels trois voies sont proposées (BOUDON, 1990, p. 535) : "le lycée (débouchant sur le secondaire long), le collège d'enseignement général (débouchant sur le secondaire court), et la classe de fin d'études (correspondant à un supplément d'enseignement post–primaire d'un an)."

À partir de ces données, Boudon observe alors tout d'abord la chose suivante (BOUDON, 1990, p. 536) : "au début de la scolarité, la réussite scolaire tend à être en moyenne d'autant plus faible que le niveau social de la famille est plus modeste."

Dans un second temps, il s'intéresse alors au lien entre l'origine sociale et une autre variable : le choix d'orientation, à la fin du primaire, entre les trois possibilités (lycée, collège d'enseignement général ou classe de fin d'études). Plus précisément, il regarde tout d'abord le lien entre l'origine sociale et le fait de s'orienter vers le secondaire (long ou court) ou bien de choisir la classe de fin d'étude, et ce à *résultat scolaire égal*.

Il observe alors que la majorité des bons élèves vont dans le secondaire, quelle que soit leur origine sociale (BOUDON, 1990, p. 536) : "90% des enfants d'ouvriers, 96% des enfants d'employés, 99% des enfants de cadres moyens ou supérieurs (...)"

Par contre, et c'est là où ça devient intéressant, l'écart se creuse quand le niveau de réussite diminue : pour les élèves moyens, "57% des enfants d'ouvriers vont dans le secondaire, 78% des enfants d'employés, 81% des enfants de cadres moyens, 92% des enfants de professions libérales et industriels, et 99% des enfants de cadres supérieurs. Les contrastes deviennent encore plus marqués dans le cas des élèves faibles (désignés comme "médiocres" ou "mauvais" dans l'enquête)."

Il présente ensuite des résultats un peu plus affinés, mais le constat est le même : sur l'échantillon de l'enquête, plus le niveau scolaire est faible en sortie de primaire, plus l'origine sociale est statistiquement associée au choix d'orientation. Ces résultats empiriques, notamment l'existence d'une telle association statistique restant élevée même lorsque la variable "niveau scolaire" est contrôlée, lui permettent de soutenir l'existence d'une relation de *causalité* entre l'origine sociale et le choix d'orientation.

En d'autres termes, le capital culturel hérité de la famille joue un rôle évident dans le fait qu'on observe en général que le niveau scolaire est meilleur lorsque l'origine sociale est plus élevée; par contre, puisque ces inégalités restent élevées même lorsqu'elles sont observées à niveau scolaire égal, c'est la preuve pour Boudon que c'est bien l'origine sociale dans sa dimension *économique* qui est la cause prépondérante des inégalités devant l'école.

L'explication théorique

Comme je l'ai suggéré plus haut, en plus de ces dissensions quant au constat empirique à dresser au sujet des inégalités sociales devant l'école, ces deux types d'approches se distinguent également dans leur contenu théorique.

La première approche semble avoir à la fois une visée explicative et une visée interprétative. L'explication du fait que le capital culturel est la variable prépondérante dans l'origine et le maintien de l'inégalité sociale devant l'école repose sur la notion d'*habitus*, un concept qui remonte à Thomas d'Aquin, mobilisé par d'importants sociologues comme Marcel Mauss ou Norbert Elias, et dont Bourdieu donne la définition suivante :

[...] systèmes de dispositions durables et transposables, structures structurées prédisposées à fonctionner comme structures structurantes, c'est-à-dire en tant que principes générateurs et organisateurs de pratiques et de représentations qui peuvent être objectivement adaptées à leur but sans supposer la visée consciente des fins et la maîtrise expresse des opérations nécessaires pour les atteindre, objectivement "réglées" et "régulières" sans être en rien le produit de l'obéissance à des règles, et, étant tout cela, collectivement orchestrées sans être le produit de l'action organisatrice d'un chef d'orchestre. (BOURDIEU, 1980, p. 88)

L'*habitus* désigne donc la façon avec laquelle le social s'incorpore dans l'individu. En d'autres termes, les gestes, la manière de parler ou de se comporter, les goûts et les préférences des individus ainsi que leur intelligence sociale (la connaissance de règles et normes sociales implicites qui leur permet de donner un sens à certaines situations) tout ceci est en grande partie déterminé par l'origine sociale des individus en question.

Or, les milieux sociaux, comme l'école, ont tendance à favoriser les individus porteurs d'un certain *habitus*. Par exemple, certaines façons de se comporter en classe, ou certains rapports au savoir et à l'apprentissage vont être davantage valorisés que d'autres dans le champ scolaire. Ces *habitus* sont de fait toujours sous-tendus par un capital culturel lui-même hautement dépendant de l'origine sociale : les livres disponibles dans l'environnement familial de l'élève, les sorties culturelles et la façon avec laquelle elles sont perçues et valorisées selon le milieu social, et ainsi de suite.

La reproduction des inégalités sociales passe donc par l'héritage socialement différencié d'un certain capital culturel et d'un certain type d'*habitus* : cette différence va se traduire par une réussite scolaire elle-même différenciée, et donc par des probabilités différentes d'atteindre un statut social élevé.

En plus de cette visée explicative, on trouve également un autre type de contenu théorique dans ce premier type d'approche, qui a trait davantage à l'interprétation du comportement des acteurs dans une optique critique de dévoilement.

L'ouvrage de Passeron et Bourdieu déjà cité plus haut, *Les héritiers*, est exemplaire en la matière. En plus des données statistiques à disposition des auteurs, cet ouvrage rend également compte de résultats d'enquêtes directes auprès d'étudiants de facultés de lettres, à Paris comme dans le reste de la France. Toute une partie de ce livre est alors consacré à l'interprétation du sens que les acteurs donnent à leur position d'étudiant. En particulier, est décrit en détails le jeu social auquel ils s'adonnent afin de se constituer effectivement comme étudiants, c'est-à-dire afin de correspondre à l'image mythique qui s'y rattache.

Le café où ils se réunissent d'ordinaire, par exemple, devient un endroit où l'étudiant consomme "avant tout les significations symboliques dont sont investis le café et le travail solitaire au café", et où il vient rejoindre "l'étudiant archétypal" plus que d'autres étudiants réels. Ainsi, ce travail d'interprétation a pour but de mettre en lumière l'effort (social) constant qu'un étudiant doit fournir afin d'être perçu et de se percevoir comme tel. Cela permet de *dévoiler* ainsi le caractère perpétuellement construit du statut social de l'étudiant, et de l'effort permanent pour (se) dissimuler cette construction.

L'autre dissimulation dévoilée a une dimension autrement plus politique. Elle porte sur l'hypocrisie du système scolaire à se déclarer et à se présenter égalitaire lorsqu'il consacre les inégalités sociales en les faisant passer pour des inégalités d'aptitudes et par là, naturellement méritées. L'objectif d'une telle œuvre sociologique n'est donc pas uniquement descriptive ou explicative, mais également *critique*.

En ce qui concerne Boudon, le contenu théorique de ses travaux s'en tient principalement aux aspects explicatifs, sur lesquels je vais me concentrer ici – même s'il place de temps à autre certaines formules rhétoriques à vocations normatives à peine voilées.

Sa proposition principale est celle d'un *modèle* théorique ayant pour but d'expliquer, par l'agrégation des choix individuels, la relation de causalité observée au niveau statistique. Comme j'aurai l'occasion de revenir plus longuement sur ce type de modèle, je vais ici simplement le présenter schématiquement.

L'idée fondamentale est la suivante. Pour expliquer une telle relation observée au niveau collectif (statistique, macro), il faut modéliser le choix d'entités individuelles : les contraintes qui pèsent sur ce choix ainsi que la façon dont ces contraintes vont influencer effectivement ce choix. Boudon considère une famille typique qui doit faire un choix entre différents parcours scolaires possibles. Plus un parcours peut rapporter de bénéfice à long terme,

plus il nécessite de ressources économiques pour pouvoir supporter financièrement des études longues. À l'inverse, les parcours scolaires les plus courts sont donc moins coûteux et rapportent plus rapidement, même s'ils rapportent moins, en moyenne, que de longues études.

Le choix d'un acteur individuel (ici, la famille, les parents ou l'élève) est donc guidé par la maximisation d'un bénéfice espéré associé à chacun des choix possibles. Ce bénéfice espéré dépend donc du coût de chaque choix, du bénéfice à plus ou moins long terme, ainsi que de la probabilité perçue de réussir effectivement à atteindre le but fixé lors du choix d'un certain parcours.

Or, le coût des différents choix dépend directement de la position sociale de l'acteur individuel en question. Pour les familles aisées, par exemple, le coût (relativement à leur niveau de vie global) est plus faible que pour une famille pauvre. La probabilité de réussite, quant à elle, dépend du niveau scolaire atteint à la fin de la primaire. On peut imaginer que plus celui-ci est élevé, plus la probabilité (perçue) de réussite l'est aussi.

On suppose ensuite qu'un acteur individuel va alors faire le choix qui maximise le bénéfice espéré. En particulier, à réussite scolaire égale (donc à probabilité perçue de réussir égale), c'est principalement l'origine sociale qui va contraindre le choix : à réussite scolaire égale, une famille aisée va avoir davantage tendance à laisser son enfant dans le système scolaire qu'une famille pauvre, car le coût relatif de ce choix pour la première sera plus faible que pour la seconde.

On voit dans cet exemple comparatif se refléter des dissensions méthodologiques et empiriques, c'est-à-dire portant sur le constat effectif d'une situation sociale donnée : est-ce la différence de capital économique ou culturel qui est la principale source du maintien des inégalités sociales par l'école ? Cependant – et c'est cette question qui m'intéresse ici fondamentalement – se reflète également une dissension épistémologique importante, portant sur les objectifs et la forme du contenu théorique des productions sociologiques. Or, cette question épistémologique ne m'apparaît pas accessoire ou secondaire, mais bien plutôt *fondamentale*, puisqu'elle porte sur les conditions mêmes qui peuvent permettre une discussion quant à la meilleure façon de mettre en évidence un phénomène et de l'expliquer.

On observe aussi une dissension dans les outils conceptuels mobilisés.

Le premier type d'approches repose principalement sur des concepts qualitatifs et peu – voire pas du tout – modélisés formellement. L'approche bou-donienne, ainsi que les travaux qui s'en réclament,³ adoptent au contraire massivement l'utilisation d'outils formels et mathématiques.

Cela pose une question qui va prendre davantage de place dans la suite de cette thèse, à savoir : l'utilité, la possibilité ou encore la pertinence d'utiliser ce type d'outils dans l'explication de phénomènes sociaux. Je vais illustrer cette problématique dans la section suivante à l'aide d'un dernier exemple, tiré de la sociologie des réseaux.

3.2.3 L'utilisation d'outils formels et son interprétation

Dans l'objectif d'illustrer la possibilité d'un raisonnement expérimental en sociologie, Dominique Raynaud présente à plusieurs endroits de ses travaux (RAYNAUD, 2005 ; RAYNAUD, 2006 ; RAYNAUD, 2019) une étude de sociologie historique dans laquelle il lui est possible de faire une *prédiction* couronnée de succès.

Ce qui me semble intéressant ici, outre les outils de sociologie des réseaux mobilisés, est une réaction particulière qu'a suscité la présentation de ces résultats au séminaire *Sociologie : Théorisation et expérimentation* à l'Université de Paris Sorbonne le 10 mars 2004, et dont (RAYNAUD, 2005) est une retranscription. C'est cette réaction qui traduit selon moi une dissension épistémologique profonde quant à l'utilisation de tels outils formels, comme on va le voir.

Le propos de Raynaud est l'étude du taux de cosmopolitisme des universités médiévales (les *studia generalia*) d'Oxford, Paris et Bologne aux treizième et quatorzième siècles. Il trouve, notamment, qu'il existe une relation empirique entre cette grandeur et leur centralité dans le réseau global des universités médiévales de l'époque.

Le taux de cosmopolitisme est défini simplement comme la fraction des étrangers sur le nombre total de *lecteurs*. Quant à la centralité, il s'agit d'un concept mathématique tiré de la théorie des graphes (ou des "réseaux complexes"). Un réseau, ou graphe, est un ensemble fini de $g \in \mathbb{N}^*$ sommets connectés les uns aux autres d'une certaine façon, traduisant la topologie du réseau. Ici, chaque sommet représente un *studium generale* et une connexion entre deux sommets représente le fait que deux *studia* sont en contact. En notant \mathcal{J}_i le nombre de sommets du graphe desquels le sommet i peut recevoir

3. Par exemple, le modèle de Breen et Golthorpe (GOLDTHORPE, 1996) ou bien l'approche *agent based modeling* de Manzo (MANZO, 2013a), sur lesquelles je reviendrai au chapitre 4.

de l'information en un nombre fini de pas, la centralité du sommet i est alors définie comme :⁴

$$C_i^* = \frac{\mathcal{J}_i}{g-1}, \quad (3.1)$$

$g-1$ représentant alors le nombre de sommets maximal auquel pourrait être lié le sommet i . Selon la définition donnée ci-dessus, plus un nœud est central, plus la proportion de noeuds auxquels il est lié, directement ou indirectement, est élevée.

Raynaud mesure alors le taux de cosmopolitisme C_i des trois *studia generalia* cités, ainsi que leur centralité respective C_i^* dans le réseau des universités dans lequel elles se trouvent. Il observe alors que ces deux grandeurs sont empiriquement liées. Il postule même la relation suivante entre les deux :

$$C_i = C_{max}(1 - e^{-\lambda C_i^*}). \quad (3.2)$$

En d'autres termes, le cosmopolitisme d'une université médiévale semble directement liée à sa centralité dans le réseau des universités dans lequel elle se trouve. Il s'agit d'une relation observée entre des variables macro, "qui ne dévoile rien de sa raison d'être".

Pour autant, on peut en donner une interprétation assez direct via un modèle explicatif simple intervenant au niveau micro, c'est-à-dire à l'échelle des *studia* :

plus un studium est central dans le réseau, plus il est aisé à ses membres de nouer des relations avec ceux des autres studia ; ils sont donc mieux à même de comparer les talents et d'attirer à eux les lecteurs étrangers jugés les plus compétents. Un studium central devient ainsi une organisation plus cosmopolite que les studia excentrés. (En italique dans le texte). (RAYNAUD, 2005, p. 14)

Il reconstruit donc la relation observée au niveau macro à l'aide d'un raisonnement s'appliquant au niveau micro, basé sur les actions des individus, s'inscrivant ainsi dans un paradigme *actionniste* tel que préconisé par Weber ou Simmel. Nous aurons l'occasion de revenir plus en détail sur ce type d'explications lorsque je parlerai de l'individualisme méthodologique.

De plus, Raynaud met ensuite à l'épreuve son modèle en prédisant le taux de cosmopolitisme de l'université de Cambridge à la même période. Prédissant en 2002, à une époque où il n'avait pas encore eu accès aux données

4. Raynaud donne également une deuxième façon de mesurer la centralité, notée C^0 , mais j'ai choisi ici de n'en présenter qu'une seule pour simplifier la présentation.

lui permettant de calculer le taux de cosmopolitisme, une valeur de 8% pour celui-ci uniquement à partir de la relation (3.2) et de la mesure de la centralité de l'université de Cambridge, il mesure quelques années plus tard un taux de 9,5% (intervalle de confiance à 95% : [8%–11%]) à peine supérieur à sa prédiction. Les détails des hypothèses auxiliaires sous-tendant cette prédiction peuvent être trouvés dans les travaux déjà cités, par exemple (RAYNAUD, 2005, p. 15).

Lors de la présentation de ces résultats au séminaire cité plus haut, le commentaire du sociologue Arnaud Saint-Martin me semble représentatif d'une manière assez commune de comprendre l'utilisation de ces outils de modélisation (comme la théorie des graphes) en sciences sociales. Il déplore notamment que dans cet exemple, la relation quantitative mesurée au niveau macro, reposant sur des "designateurs ultra-rigides", semble tout expliquer, sans laisser de place aux acteurs et institutions concrètement impliqués à ce moment-là. Plus précisément :

Les interprétations sociologiques classiques qui utilisent des notions plus qualitatives sont hélas reléguées au second plan; pire, elles sont frappées de désuétude. Cet effort de *désubstantialisation* du matériau sociologique, qui n'est figuré que par des chiffres, aboutit à un tableau désincarné de cette réalité, à des lieues de la collection patiente des traces, de l'enquête qui s'attache à décrire le moindre détail, l'identité et la biographie des membres de ces mondes-là, l'histoire des *studia*, des luttes qui y eurent lieu, etc. Si les *studia* d'Italie centrale attirent autant d'étrangers, peut-être est-ce aussi du aux modes d'organisation de ces centres, à l'idéologie qu'ils légitiment, plus ouvertes sur l'extérieur, moins « xénophobe » qu'en Angleterre, etc. Pour expliquer les dispersions et les choix des acteurs, ces descriptions pourtant instructives apparaissent bien superflues au regard de la puissance explicative de la sociologie des réseaux. (RAYNAUD, 2005, p. 31)

Ainsi, ce qui est déploré ici est une forme de "réductionnisme" consistant à passer outre tout un ensemble de détails qui pourraient sembler au premier abord instructifs pour réduire l'analyse globale à une modélisation ultra-simplifiée se basant sur la mesure unique de deux variables.

À ceci, il est intéressant d'opposer la réponse qu'en fait Dominique Raynaud, soulignant selon moi une réelle fracture épistémologique existante entre ces deux manières de percevoir les sciences sociales, la sociologie en particulier, et également l'utilisation d'outils formels en son sein.

D'abord, Raynaud est tout à fait d'accord sur le fait que les acteurs et les institutions jouent un rôle particulier, puisqu'ils sont à la base de l'explication de la relation macro observée, comme décrit plus haut. De plus, l'analyse en termes de réseaux complexes et d'indices de centralité ne signifie absolument pas que ces rôles individuels sont niés, au contraire : ceux-ci sont tout simplement mis en modèle, traduits en hypothèses claires sous une forme mathématique.

En effet :

Tout d'abord, je dois vous dire que je n'endosse pas la dichotomie que vous introduisez entre indices mathématiques et acteurs historiques. Car les propriétés réticulaires du réseau social ne sont qu'une formalisation des relations entre les acteurs. La possibilité de retranscrire ces relations dans le cadre de la théorie des graphes ne se justifie que parce que les acteurs créent ces relations à partir des motivations qu'ils ont d'entrer en relation avec autrui, ce qui rejoint directement la perspective actionniste. [...] Vous m'opposez que "si les *studia* d'Italie centrale attirent autant d'étrangers, c'est aussi parce que ces centres sont plus ouverts et moins xénophobes qu'en Angleterre". Vous ne faites là que retraduire en langage naturel les concepts de la sociologie des réseaux. Car si un *studium* xénophobe se ferme sur lui-même, cela équivaut à dire qu'il perd ses relations avec les *studia* adjacents. Cela ne constitue en rien une différence quant au sens, mais seulement quant à la forme (qualitative vs. quantitative). Je vous renvoie à nouveau à ce que j'ai dit plus haut : le choix de l'analyse réticulaire se justifie par des raisons méthodologiques, à savoir la possibilité de faire des prédictions quantitatives. (RAYNAUD, 2005, p. 31)

Ainsi, il n'y a pas de différence profonde, selon Raynaud, entre les hypothèses micro avancées par Arnaud Saint-Martin, reposant par exemple sur la xénophobie des *studia*, et les hypothèses micro avancées par lui-même, reposant sur des mesures de centralité. La seule différence est une différence de forme : qualitative (exprimée dans un langage naturel) pour l'un, quantitative (exprimée dans le langage de la théorie des graphes) pour l'autre.

Raynaud défend toutefois que cette manière de faire a un avantage qui, selon lui, compense la perte de détails et peut-être l'éloignement du sens commun produite par l'utilisation de ces outils formels : ces derniers permettent en effet de formuler des prédictions quantitatives et donc plus précises, pouvant se soumettre plus aisément à des tests sévères.

3.2.4 Conclusion

Cette première partie a servi à illustrer la pluralité des points de vue existants concernant les buts des sciences sociales et les moyens à investir pour atteindre ces objectifs. Cette situation peut soit être interprétée comme étant normale, ou bien comme problématique, c'est-à-dire non satisfaisante, à améliorer d'une manière ou d'une autre.

Dans la partie suivante, je vais présenter les principaux arguments "pluralistes", c'est-à-dire des arguments défendant l'idée que cette situation est tout à fait *normale* étant donné les spécificités que les sciences sociales possèdent vis-à-vis des sciences naturelles. Pour les raisons données en introduction de ce chapitre, je considère, à l'encontre de ces positions pluralistes, que la situation de pluralité conceptuelle dans laquelle se retrouve les sciences sociales est au contraire *problématique*.

Je soutiendrai notamment, en mobilisant un certain nombre d'arguments classiques, qu'il n'y a aucune bonne raison en principe que les sciences sociales jouissent d'une épistémologie singulière, c'est-à-dire d'une définition propre de la scientificité. En d'autres termes, l'état d'éclatement épistémologique actuel des sciences sociales est un état que l'on peut et qu'il faut améliorer, en un sens qu'il me faudra alors préciser.

3.3 Les arguments pluralistes

La pluralité méthodologique, théorique mais surtout épistémologique exposée dans la section précédente est sous-tendue par une certaine tension quant à la question de savoir si les sciences sociales peuvent ou doivent être considérées et construites comme des sciences comme les autres. Ces débats sur la nécessité ou non d'adopter une épistémologie à part pour les sciences sociales ont accompagné celles-ci depuis leur naissance. À la fin du dix-neuvième siècle, le fameux *Methodenstreit* (le conflit des méthodes) oppose déjà plusieurs écoles de pensée économique, en se nourrissant par ailleurs de la distinction très influente entre sciences de la nature (*Naturwissenschaften*) et sciences de l'esprit (*Geisteswissenschaften*) due à Wilhelm Dilthey (DILTHEY, 1989). Ce type de distinction se retrouve à plusieurs moments du développement de ces disciplines, comme par exemple lors de la controverse entre Karl Popper et l'école de Francfort dans les années 1960, ou encore dans l'ouvrage de Jean-Claude Passeron, *Le raisonnement sociologique* (PASSERON, 1991), paru

pour la première fois au début des années 1990, qui a été très influent dans la sociologie francophone.

Le but de cette partie est de présenter la position "pluraliste" le plus exhaustivement possible. Celle-ci, comme toute position philosophique complexe, n'est pas strictement unifiée ni homogène, et a pu prendre des formes historiques différentes qui ne se raccordent pas nécessairement parfaitement les unes avec les autres.

Ici, je me concentre sur les différents *arguments* avancés pour soutenir la nécessité d'une épistémologie singulière pour les sciences sociales. Mon objectif est de synthétiser ces différents arguments sous la forme de quatre grandes familles de propriétés qui semblent être spécifiques à ces disciplines :

I – L'HISTORICITÉ : Les phénomènes sociaux sont des phénomènes historiques et donc singuliers, c'est-à-dire non reproductibles, et multifactoriels, donc irréductibles à un nombre restreint de variables que l'on pourrait contrôler (les deux points étant fortement liés).

II – LA SIGNIFICATION : Les phénomènes sociaux et culturels ont une signification, un sens particulier pour les individus qui les vivent, sens qui est à la portée du chercheur en tant qu'il est et parce qu'il est, lui aussi, un être de même nature que l'objet qu'il étudie.

III – L'INTENTIONNALITÉ : Les acteurs sociaux, qu'ils soient individus ou institutions, ont des intentions, des motivations, ce qui fait que leurs actions sont très souvent finalisées, orientées vers une fin qui pré-existe à l'action sous forme de représentation.

IV – LA PERFORMATIVITÉ : Les chercheurs tout comme les connaissances elles-mêmes peuvent avoir un effet potentiellement imprévu sur les objets étudiés, par exemple sur leur comportement.

Je vais étudier plus en détail ces quatre familles d'arguments, en rattachant certains aspects à des auteurs bien identifiés, mais surtout en tentant de présenter la façon avec laquelle ils sont utilisés pour justifier la nécessité d'une épistémologie à part.

En effet, les positions pluralistes vont habituellement plus loin que simplement la reconnaissance que ces aspects-là sont spécifiques aux sciences sociales : elles soutiennent que ces spécificités ont des conséquences épistémologiques importantes. En particulier, elles justifieraient de rendre compte de la scientificité de ces disciplines au sein de cadres épistémologiques alternatifs, c'est-à-dire répondant à des critères différents des autres productions scientifiques.

3.3.1 Quatre familles d'arguments

I - L'historicité

Les phénomènes qu'étudient les sciences sociales sont des phénomènes humains et donc, en cela, intrinsèquement historiques, singuliers, non reproductibles artificiellement. Par exemple, on ne pourra jamais étudier l'émergence de l'agriculture ou du capitalisme, ou encore les causes de la première guerre mondiale, de la même façon que l'on étudie la croissance d'un arbre dans un sol plus ou moins acide ou la trajectoire d'un électron dans un champ magnétique plus ou moins fort.

Cette particularité est reconnue par tous les tenants d'une épistémologie alternative pour les sciences sociales. On retrouve par exemple chez Theodor Adorno l'idée que la nature même de l'objet sociologique l'empêche d'être approché par les mêmes procédés que ceux utilisés en sciences naturelles :

Cette discordance [la tension qui oppose le général au particulier] fonde le fait que l'objet de la sociologie – la société et ses phénomènes – n'a pas la sorte d'homogénéité sur laquelle peut compter ce qu'on appelle science classique de la nature. Alors qu'on était habitué à conclure de l'observation des propriétés d'un morceau de plomb à celles de tout plomb, on ne peut pas, en sociologie, procéder semblablement à partir d'affirmations partielles sur des situations sociales à leur généralisation, même restrictive. La généralité des lois en sciences sociales n'est pas du tout celle d'une extension conceptuelle, dans laquelle les exemplaires isolés viendraient s'encaster sans rupture, elle se rapporte toujours et essentiellement à la relation du général et du particulier dans sa concrétion historique. De cela témoigne, en négatif, l'absence d'homogénéité de la condition sociale, l'"anarchie" de toute histoire jusqu'à aujourd'hui, comme, en positif, le moment de la spontanéité qui ne se laisse pas capter par la loi des grands nombres. (ADORNO, 1979, p. 66-67)

On retrouve la même sorte d'idée chez Jean-Claude Passeron qui insiste beaucoup sur la nécessaire indexation de tout concept sociologique à un contexte historique particulier hors duquel il perd irrémédiablement son sens et donc son intérêt théorique. Ce constat a pour conséquence notamment l'impossibilité de penser le phénomène indépendamment de son contexte historique, c'est-à-dire "[désassorti] de [ses] coordonnées spatio-temporelles" (PASSERON, 2006, p. 615).

Plus important encore, cela condamne les sciences portant sur le monde historique à une certaine forme d'instabilité conceptuelle, puisque :

[ne sont pas réunies les] conditions qui permettent ailleurs la stabilisation du sens des concepts que ce soit dans un paradigme ou dans un système d'idées reçues, à savoir, soit la constance des fonctions pratiques, corrélative de la sociologie spontanée ou de l'idéologie, soit l'universalité formelle des opérations non empiriques dans un système logique, soit la généralité expérimentale capable de stabiliser l'espace logique d'une théorie physique.

(PASSERON, 2006, p. 268)

Comme on le voit ici, l'impossibilité d'une expérimentation reproductible, du fait de la nature même des phénomènes sociaux, ne permet donc pas de stabiliser l'espace conceptuel au sein duquel s'énoncent les assertions sur le monde historique. Ainsi, cela permet à Passeron d'éviter le contre-argument consistant à souligner qu'il existe tout un ensemble de disciplines relevant des sciences naturelles et qui portent pourtant sur des phénomènes tout aussi historiques, comme la cosmologie, la géologie ou la biologie évolutive par exemple. En effet, il reconnaît que :

Il existe bien [...] des questions de forme historique dans toutes les sciences de la matière et de la vie (histoire de la terre, de l'univers, des espèces vivantes ou des écosystèmes, ou tout simplement quand il s'agit d'expliquer un accident de chemin de fer).

(PASSERON, 2006, p. 596)

Cependant, la différence fondamentale est que ces disciplines peuvent organiser ces observations au sein d'un cadre nomologique, c'est-à-dire structuré autour d'une ou plusieurs *lois* – ce dernier terme étant compris ici comme une relation constante entre des variables ou des grandeurs bien identifiées et dont la pertinence ne dépend pas du contexte historique.⁵

L'impossibilité d'établir des lois, en sociologie, qui seraient valables indépendamment du contexte historique de leur application (pour les raisons soulevées plus haut) empêche donc d'organiser les observations historiques de la même manière qu'en cosmologie ou en astrophysique, par exemple, où

5. Une chose importante à noter ici et à garder en tête à la lecture de tout le manuscrit est que le terme de "loi" n'est pas employé dans le même sens qu'en métaphysique des sciences, mais bien dans un sens plus faible de régularités ayant possiblement un certain domaine de validité – ce qui n'est pas, pour l'immense majorité des auteurs, le cas d'une loi au sens métaphysique.

les théories sous-jacentes peuvent être testées indépendamment des observations de type historique (comme la formation des planètes, l'évolution des galaxies où l'émission du fond diffus cosmologique), même si ces dernières se dérobent à la manipulation voire à l'observation directe.

La mise en évidence de cette caractéristique est aussi vieille dans l'histoire de ces disciplines que le *Methodenstreit* qui éclate, comme on l'a déjà souligné, à la fin du dix-neuvième siècle, alors que domine le paradigme comtien (COMTE, 1830) déployant les différentes disciplines scientifiques, par ordre de complexité, le long d'un "*continuum* épistémologique et méthodologique" (CUIN, 2000, p. 115). Cette controverse est justement caractérisée par une distinction entre les sciences comme la physique qui décrivent des phénomènes reproductibles, généraux, pouvant se subsumer sous des lois, et les autres sciences comme la sociologie ou l'histoire qui ne peuvent qu'étudier des phénomènes dans leur singularité historique.

Deux positions sont en réalité possibles, à ce point. D'une part, une position descriptive, faisons montre d'un pessimisme fondé historiquement : l'histoire de la sociologie atteste que malgré tout leurs efforts, les sociologues ne sont pas parvenus à mettre en exergue des lois, et c'est probablement parce que c'est impossible, dû à l'aspect essentiellement chaotique de la vie sociale. D'autre part, une position plus normative : même si c'était possible de mettre à jour des régularités stables indépendantes du contexte historique, cela ne serait pas pertinent puisque justement, si ces régularités ne dépendent pas du contexte historique, elles ne disent rien de bien intéressant sur les phénomènes qui nous occupent.

Il semble que l'on peut trouver ces deux positions chez Passeron. Lorsqu'il parle de la différence entre les sciences naturelles qui s'intéressent à des phénomènes historiques, comme on l'a cité plus haut, il contraste ensuite cette situation avec celles des sciences sociales :

... il n'existe aucun savoir nomologique de ce type qui guiderait le questionnement historique des sciences sociales. L'histoire des sciences sociales suffit ici pour trancher, en nous montrant comment ont été déçus tous les espoirs naturalistes qu'on a pu placer dans le rôle fondatif d'une science nomologique-mère (successivement ou contradictoirement, économie, démographie, psychologie ou linguistique). Sauf espoir de l'apparition inattendue d'un génie galiléen, newtonien ou einsteinien encore à venir, l'illusion de Merton (...) qui renvoie l'avènement du royaume nomologique des sciences sociales à des surlendemains radieux, est de type messianique. Il

existe, il est vrai, nombre de prétendants au rôle de messie théorique; ils ne font des miracles qu'aux yeux de la secte. (PASSERON, 2006, p. 596)

À ce type de constat, on pourrait par exemple avoir le réflexe d'opposer l'*existence* d'un certain nombre de régularités empiriques qui ne semblent pas dépendre fortement du contexte historique, ou en tout cas dont le domaine de validité est assez étendu pour pouvoir prétendre au rôle de loi. C'est par exemple ce à quoi Lahire consacre tout un chapitre dans (LAHIRE, 2023).

Le problème de ce contre-argument est qu'il manque sa cible, puisque le sociologue pluraliste pourra alors se rabattre sur la deuxième version de son argument : si on met en évidence des lois, c'est-à-dire des relations, des structures, des régularités qui ne dépendent pas du contexte historique, alors elle n'auront, du même coup, aucun contenu sociologique intéressant à exploiter.

En effet, il suffit de lire la définition que Jean-Claude Passeron donne au concept de "monde historique", terme qu'il utilise tout le long de son ouvrage (PASSERON, 2006, p. 615) : "Ensemble des occurrences observables lorsqu'elles ne peuvent être désassorties de leurs coordonnées spatio-temporelles sauf à perdre le sens que l'on vise en assertant sur elles." L'argument est donc essentiellement le suivant : ce n'est pas qu'il soit impossible de mettre au jour des régularités indépendantes du contexte historique, c'est simplement qu'elles n'ont aucun intérêt sociologique.

Bien sûr, cette seconde version trivialise du même coup l'argument, puisque dans ce cas on présuppose dès le début ce dont on souhaite nous convaincre. Il semble que cela soit également la position de Wilhelm Dilthey :

Each particular human science is formed through the technique of isolating a partial content of socio-historical reality. Even history disregards those factors in the life of individuals and of societies which are the same in all epochs; its view is directed toward what is distinctive and singular. (DILTHEY, 1989, p. 79)

L'intérêt du monde social, historique, l'information pertinente que la sociologie doit mettre en lumière et organiser, est contenu dans le sens particulier que les acteurs donnent aux phénomènes et aux événements dans lesquels ils se trouvent plongés – et ce sens, par définition, ne peut être pensé de manière universelle.

On retrouve le même type d'argument chez les défenseurs d'une individualisation nécessaire en psychanalyse, pour justifier en particulier la non pertinence de se soumettre aux mêmes types de tests expérimentaux qu'en psychologie cognitive, par exemple :

La psychanalyse se dégage [du modèle causaliste de la psychologie cognitive] pour envisager le sujet d'abord dans son dire, plutôt que dans les faits de son histoire. C'est ce que le sujet a élaboré fantasmatiquement, ce qu'il s'est représenté de son histoire, qui compte plus qu'une réalité des faits, ce qui fait événement pour lui. (...) [L]e cadre épistémologique de la psychanalyse est créé par son objet même, qui n'est pas le comportement humain, mais un vécu humain en tant qu'il est spécifié, référé à une conduite. Or, le vécu humain est subjectif. C'est parce que la psychanalyse a affaire à du subjectif qu'elle ne peut pas faire autrement que d'interpréter à partir de chaque sujet. (SIGNOL, 2011, p. 55)

L'analyste ressemble alors bien plus à un *archéologue* s'attelant à découvrir une histoire vécue – telle que *perçue* par le sujet – singulière, et ce, couche après couche, lentement, avec l'impossibilité de ranger ce travail dans une catégorie de modèles donnée a priori au sein d'un cadre théorique précédant l'analyse en question.

Cela nous mène alors à la deuxième famille d'arguments pluralistes.

II - La signification

La seconde propriété distinctive des sciences portant sur le monde historique et social est que les individus donnent du sens aux phénomènes et aux événements qu'ils vivent. Et non seulement ils leur donnent du sens, mais ce sens, cette signification, fait partie intégrante du phénomène ou de la situation étudié, et dépend donc intrinsèquement du contexte historique de cette situation.

De ce point de vue, décrire un tel phénomène sans intégrer cette signification – par exemple, en mettant en évidence des régularités qui ne dépendent pas du contexte historique, et donc pas non plus du sens éventuel que les acteurs leurs donnent – c'est passer à côté de ce qui fait que c'est un phénomène sociologique, et donc c'est passer à côté de ce que l'on tente précisément de faire.

Cette propriété provient à la fois de la nature de l'objet sociologique, c'est-à-dire l'humain, qui est un être de significations et de représentations, mais aussi du fait que le chercheur lui-même est de même nature, et donc peut avoir accès – par un travail méthodique toutefois – à ces significations, en en reconstruisant le sens par la *compréhension*.

Comme l'exprime clairement Dilthey :

The practice of regarding these disciplines as a unity distinct from the natural sciences is rooted in the depth and totality of human self-consciousness. Even before he is concerned to investigate the origin of the human spirit, man find within his self-consciousness a sovereignty of the will, a responsibility for actions, a capacity for subjecting everything to thought and for resisting, from within the stronghold of his personal freedom, any and every encroachment. This differentiates him from the rest of nature. He exists in nature as a realm within a realm – *imperium in imperio*, to use an expression of Spinoza. (DILTHEY, 1989, p. 58)

Social states are intelligible to us from within; we can, up to a certain point, reproduce them in ourselves on the basis of the perception of our own states; our representations of the historical world are enlivened by love and hatred, by passionate joy, by the entire gamut of our emotions. Nature, however, is dead for us. (...) Nature is alien to us. It is a mere exterior for us without any inner life. Society is our world. We sympathetically experience the interplay of social conditions with the power of our total being. (DILTHEY, 1989, p. 88)

À la distinction entre sciences de la nature et sciences de l'esprit, correspond celle entre recherche de régularités et focalisation sur les singularités, et s'y ajoute donc la dernière entre la visée explicative des premières, et la visée compréhensive des secondes. Le débat explication/compréhension ("on explique la Nature, on comprend le social"), typique du *Methodenstreit*, a toujours des répercussions aujourd'hui sur le point de vue épistémologique à tenir en sciences sociales, en sociologie tout particulièrement.

Passeron insiste également sur le fait que même si l'utilisation d'outils statistiques n'est pas à proscrire en sciences sociales, et en particulier en sociologie, leur utilisation doit être conditionnée à une interprétation minutieuse de ce que l'on énonce dans leur langage. En d'autres termes, les variables et les relations statistiques que l'on peut établir entre elles ne sont pas données telles quelles par la réalité sociale mais sont toujours construites à travers un certain processus. L'oubli de ce processus, de ce contexte de production ou captation de données, peut alors mener à des interprétations non sociologiques, c'est-à-dire biaisées, erronées.

Il prend en particulier l'exemple d'un tableau statistique représentant, à trois stades du parcours scolaire (1^{er}, 2^{me} ou 3^{me} cycle), la corrélation éventuelle entre réussite scolaire ('faible' ou 'forte') et le niveau socio-économique

('classe supérieure', 'classe populaire'). Je l'ai reproduit en figure 3.1.

	forte	faible		forte	faible		forte	faible
Classe sup.	70	30	Classe sup.	55	45	Classe sup.	30	70
Classe pop.	30	70	Classe pop.	45	55	Classe pop.	70	30

FIGURE 3.1 – Reproduction du Tableau II (PASSERON, 2006, p. 215)

En s'en tenant à une lecture "directe" du tableau, on peut énoncer le fait statistique qui est que "la relation entre origine sociale et réussite scolaire s'inverse du stade 1 au stade 3", laissant penser que par un mécanisme quelconque, les élèves des classes populaires se mettent à mieux réussir que les élèves des classes supérieures une fois arrivés à un certain stade.

Mais cette interprétation, nous dit Passeron, n'a bien sûr pas de sens sociologiquement, puisque :

Pour énoncer ce que dit le Tableau II, sans déformer par cette énonciation même la relation qui unit l'appartenance de classe (ou de sexe) à des chances relativement plus ou moins fortes de succès scolaire, le raisonnement sociologique doit prendre en compte le fait (absent des relations explicitement formulées par le tableau croisé lui-même) que les populations sur lesquelles on établit la variation de la relation entre appartenance à une catégorie (de sexe ou de classe) et réussite scolaire sont progressivement sélectionnées tout au long du cursus avec une inégale sévérité sous le rapport même des variables dont on entend énoncer les effets sur cette réussite. Le contextualisateur de l'interprétation est évidemment ici le concept de "sur- et de sous-sélection scolaire relatives" qui, en résumant d'autres descriptions (statistiques, biographiques ou institutionnelles), permet d'incorporer à la formulation des relations entre origine sociale (ou sexe) et réussite (ou choix) scolaire le fait que les mesures successives s'opèrent ici sur des échantillons qui, par rapport aux populations de départ, ont vu les valeurs initiales des principales variables explicatives être progressivement et inégalement biaisées par la sélection et l'orientation scolaires. (PASSERON, 2006, p. 215)

En d'autres termes, les données collectées sont biaisées précisément par le phénomène que l'on souhaite mettre en évidence en collectant ces données. Tenir compte de ce fait, par un raisonnement "sociologique" et non pas uniquement "expérimental", est donc indispensable pour étudier avec pertinence ce type de phénomènes.

Adorno, de son côté, remarque également que "les données dont cette discipline dispose ne sont pas des données brutes sans autre qualification mais sont structurées par le contexte de la totalité sociale" (ADORNO, 1979, p. 92). Ainsi, les variables mêmes avec lesquelles on va mettre en évidence un phénomène, ayant une certaine signification pour les différents acteurs, sont très souvent intriquées avec certains aspects du phénomène lui-même.

On peut penser à l'analyse quantitative vue plus haut sur le lien entre interventionnisme militaire et terrorisme. Pour ce phénomène, il n'y a déjà pas de "bonne" définition a priori. De plus, les données disponibles (par exemple, via les différentes bases de données dont les auteurs parlent pour "mesurer" le terrorisme, comme le GTD, la RAND-DWTI ou la base ITERATE) ne sont pas des données "brutes" mais dépendent elles-mêmes de la signification que les acteurs (par exemple, les gouvernements occidentaux qui récoltent ces données) en donnent. En particulier, le concept de terrorisme étant un concept "épais", c'est-à-dire chargé également d'un point de vue normatif, il va être intrinsèquement difficile de se départir de cet aspect s'il on veut étudier le phénomène, car on perdrait une bonne partie de sa pertinence, justement, en le faisant.

III - L'intentionnalité

La troisième famille d'arguments pluralistes repose sur le constat que les acteurs sociaux, que l'on parle d'individus ou bien de structures sociales plus importantes (des États, par exemple), ont des intentions, des motivations qui guident leurs actions. Ainsi, alors que les sciences naturelles ont mis littéralement des siècles à bannir les "causes finales" de leurs schèmes explicatifs, celles-ci paraissent être d'une pertinence toute particulière concernant les sciences sociales.

En effet, un objet ne tombe pas à Terre parce qu'il poursuit un but précis mais simplement parce que des forces sans intention ni finalité se trouvent s'appliquer sur lui à cet instant, dans des conditions particulières. De même, le long cou de la girafe n'est pas apparu tel quel pour permettre aux girafes d'atteindre des branches d'arbre plus élevées et dans le but de leur permettre de mieux survivre. Selon la théorie de l'évolution, ce caractère particulier est apparu graduellement, via un ensemble de mécanismes complexes de mutation, sélection et transmission, le tout en interaction avec l'environnement même de ces animaux (les autres individus, les autres espèces animales, mais aussi les arbres en question). Le point clé est qu'un tel processus, en plus

d'être lent et progressif, est surtout aveugle, c'est-à-dire non orienté vers l'accomplissement d'une fonction précise ou la résolution d'un problème particulier.

Au contraire, il ne semble pas absurde de dire qu'un étudiant travaille beaucoup parce qu'il veut réussir ses examens, qu'une électrice met tel bulletin dans l'urne parce qu'elle espère que son candidat va gagner les élections, ou bien qu'un pays envahit un autre parce qu'il compte lui ponctionner certaines de ces ressources. Ainsi, en sciences sociales, le fait que les unités élémentaires étudiées aient des intentions, des motivations, qu'elles poursuivent des buts précis, doit être pris en considération d'une façon ou d'une autre.

En d'autres termes, alors que la scientificité des sciences de la nature comme la physique, la chimie ou la biologie, est passée par leur abandon des causes finales dans leurs explications, les sciences sociales n'ont aucune raison de les abandonner. Il s'agit donc bien d'une propriété épistémologique que les sciences sociales possèdent spécifiquement.

De plus, il semble que le lien épistémique particulier qui relie le chercheur à son objet, dont on a déjà parlé plus haut, permet à celui-ci d'avoir accès *plus directement* à ces causes finales, en ce qu'il peut enquêter en interaction "directe" avec les différents acteurs dont il étudie le comportement. Le fait de pouvoir interroger directement les unités élémentaires d'un phénomène social (par exemple, les individus directement impliqués) semble effectivement être spécifique aux sciences humaines et sociales, là où le chercheur et les acteurs qu'il étudie partagent un certain langage ("naturel", comme dirait Passeron).

C'est pour cela que s'extraire du monde social à l'aide d'un langage artificiel qui permettrait d'en observer les structures causales à l'œuvre indépendamment de l'interprétation qu'en font les acteurs est non seulement difficile, mais peut sembler également peu pertinent puisqu'on peut avoir accès à des sortes de causes qui sont justement spécifiques à ces disciplines, du fait même du lien épistémique particulier qu'entretiennent chercheurs et acteurs sociaux.

L'intentionnalité des acteurs sociaux a d'autres conséquences. En particulier, l'hétérogénéité et la diversité des comportements humains laissent penser qu'y trouver des régularités est peine perdue, ce qui nous ramène à la considération (I) ci-dessus : le réel social est trop chaotique (du fait de l'intentionnalité elle-même imprévisible de ses unités élémentaires) pour établir des lois qui soient valables en tout temps et en tout lieu.

L'absence de régularités macroscopiques reflété dans l'historicité intrinsèque des phénomènes trouve donc ici son pendant microscopique dans l'intentionnalité individuelle difficile voire impossible à subsumer sous des lois simples et stables. En physique, au contraire, les lois régissant les particules élémentaires d'un gaz sont relativement simples et permettent également de comprendre les phénomènes émergent, donc macroscopiques, mesurables à l'échelle du gaz, à partir de considérations microscopiques.

IV - La performativité

Enfin, il y a une dernière famille d'arguments rencontrés pour soutenir la spécificité épistémologique des sciences sociales, c'est celle de la performativité de l'investigation scientifique sur l'objet lui-même. Elle peut se décliner de trois manières : via l'interaction entre le chercheur et ses objets de recherche, via l'interaction des connaissances du social avec le social lui-même, et enfin lorsque des injonctions politiques s'entremêlent fortement à la mesure d'une ou de plusieurs variables caractérisant un phénomène.

La première version est bien connue de tout chercheur visant à expérimenter sur ou observer directement le monde social. Le fait même de participer à une expérience va modifier en partie le comportement des sujets. En particulier, ces derniers vont chercher à savoir ce que l'on attend d'eux, et ainsi en partie d'y conformer, par exemple – ce qui, il va sans dire, peut fortement biaiser les observations elles-mêmes. Si on fait une expérimentation en aveugle sur des plantes en les assignant à deux groupes aléatoirement, les plantes n'essaieront pas de savoir dans quel groupe elles sont. Lors d'essais cliniques en double aveugle sur des humains, par contre, ces derniers chercheront à savoir s'ils ont reçu le placebo ou bien le traitement testé, et devineront, en général, mieux que ce dont on pourrait s'attendre du hasard total (STEGENGA, 2018, p. 155).

Il peut aussi s'agir de l'interaction directe entre le chercheur et les sujets d'observation : la simple présence d'un chercheur lors d'une observation participante, par exemple, peut en elle-même modifier le comportement des personnes observées. La façon dont on pose les questions dans une enquête par questionnaire, de la même façon, va également pouvoir influencer les réponses des sujets dans un sens particulier, ce qui fait qu'il est en réalité difficile (mais loin d'être impossible) de réaliser des bons questionnaires. Ce sont des choses bien connues des chercheurs en sciences humaines et sociales, et

les manuels de méthodologie de terrain, par exemple, sont remplis de recommandations visant à limiter ou en tout cas à contrôler les biais introduits par l'observation elle-même.

C'est en tout cas une particularité de ces disciplines-là, pour lesquelles les sujets sont humains et donc savent qu'ils sont observés et peuvent modifier, même sans le vouloir, leur comportement en fonction. Concernant ce point précis, il faudra éviter la mauvaise analogie avec la physique quantique, dans laquelle on entend parfois que *l'observation modifie le résultat de la mesure*. Ce n'est pas vraiment très précis. En fait, selon l'interprétation dominante (celle de Copenhague), c'est encore pire que ça : il devient impossible de parler de l'état d'un système quantique indépendamment de sa mesure, c'est-à-dire avant qu'il interagisse avec un appareil de mesure. Ainsi, on ne peut pas dire que la mesure modifie l'état du système, mais le crée : avant la mesure, le système n'est pas dans un état particulier mais dans une superposition d'états possibles, superposition pondérée par des facteurs qui donnent la probabilité de trouver le système, lors d'une mesure, dans l'état correspondant. Donc, même si la situation est "pire" que ce qu'on peut entendre, elle est aussi moins grave puisque même si un résultat unique de mesure est aléatoire, la théorie nous donne ce à quoi on doit s'attendre statistiquement – et cette distribution est parfaitement déterminée par la situation expérimentale donnée. C'est pour cela que le fait qu'un observateur modifie le comportement des personnes qu'il observe est incomparable avec la situation décrite ici en physique quantique.

De même, on pourrait penser qu'il s'agit simplement d'une instance de la charge théorique de l'observation : aucune observation n'est pure ou directe, toutes sont chargées de théories, c'est-à-dire nécessitent pour avoir du sens que l'on tienne au préalable certaines hypothèses pour vraies (par exemple sur le fonctionnement des appareils de mesure, ou sur l'opérationnalisation de certaines variables, voir section 2.4.1). Par exemple, mesurer une température avec un thermomètre modifie aussi l'état du fluide dont on mesure la température, puisqu'il interagit localement avec le thermomètre, et donc la mesure de la température, en plus de reposer sur des hypothèses de fonctionnement du thermomètre, repose aussi sur des hypothèses concernant les conditions d'utilisation du thermomètre – par exemple, le fait que l'interaction est assez faible pour que les effets soient négligeables par rapport aux effets que l'on veut effectivement observer.

La différence, c'est que la charge théorique de l'observation ne pose pas

vraiment de problème fondamental dès lors que l'on a une idée claire des hypothèses sous-jacentes, justement, à l'observation – par exemple, une théorie robuste de l'optique pour le télescope, ou une théorie robuste de la chaleur pour le thermomètre. Et c'est précisément cela qui n'est pas le cas en sciences humaines et sociales : le sociologue n'a pas à sa disposition une théorie robuste du comportement humain qui lui permettrait de savoir dans quelle mesure corriger ses observations. La question reste entière de savoir si cette situation est intrinsèque aux sciences humaines et sociales ou bien si, dans le futur, on pourrait imaginer avoir en notre possession une telle théorie qui nous permettrait de mieux interpréter les données d'observations en détectant et en supprimant leurs éventuels biais systématiques.

La deuxième instance de la performativité concerne l'effet que les connaissances elles-mêmes peuvent avoir sur le monde social, un effet qui n'existe pas en sciences naturelles.

Il peut s'agir d'une connaissance à court terme, par exemple le fait de connaître les résultats d'un sondage d'intention de votes juste avant des élections peut avoir un effet sur les votes effectifs des électeurs. En effet, si je me rends compte que la candidate X, qui est ma préférée, a plus de chances que je l'imaginai de gagner, je peux peut-être changer mon vote pour elle alors que je pensais au départ voter stratégiquement, en donnant ma voix à un autre candidat.

Il peut s'agir aussi d'un effet plus global. La psychologie sociale a par exemple mis en évidence des phénomènes tel que *l'ignorance pluraliste* (MILLER, 2023). Ce terme désigne le fait qu'une majorité de personnes peut avoir de fausses croyances à propos de ce que les autres croient, et penser par exemple appartenir à la minorité alors que ce n'est pas le cas. On peut expliquer ainsi des cas où une majorité de personnes se soumet à des normes sociales auxquelles elle n'adhère en fait pas du simple fait qu'elles pensent être une minorité à ne pas y adhérer.

On peut penser également à l'effet de prophétie auto-réalisatrice (MERTON, 1948) où le fait de croire collectivement à un état du monde (par exemple, une pénurie d'essence) peut le faire advenir. La croyance (d'une pénurie) va modifier les comportements des personnes (qui vont aller faire des stocks, donc acheter plus d'essence que d'habitude), comportements qui, en s'agrégeant, peut faire advenir effectivement le contenu de la croyance en question (la demande augmente trop vite par rapport à l'offre, il y a pénurie).

Il semble pertinent de citer ici également ce que l'on pourrait appeler l'effet "Bison Futé",⁶ qui donne les prévisions de circulation sur les routes. Si un certain nombre de personnes veulent faire un trajet un jour donné, Bison Futé prévoira une journée noire sur les routes. Cette prévision pourra alors modifier le comportement des personnes, par exemple en les faisant partir plus tôt, un autre jour ou bien en les faisant renoncer à leur voyage. Finalement, le jour en question, la prévision se révèlera fautive. Ce qui est paradoxal ici, c'est que si elle est fautive, c'est précisément parce qu'elle était vraie au départ. Ainsi, une réfutation apparente, dans ce cas, n'est pas forcément une réelle réfutation de la prédiction au départ, mais peut même en être une confirmation. Il faut avouer que c'est un schéma logique assez particulier, mais qui est pourtant assez plausible appliqué au monde social.

De manière plus générale, on pourrait aussi penser à des humains facétieux qui voudraient systématiquement mettre en défaut les théories sociologiques que l'on élabore à partir de l'observation de leur comportement : le fait même que cela soit en théorie possible est une particularité des sciences humaines et sociales. C'est parce que la connaissance que l'on a du monde social est accessible aux unités élémentaires même de ce monde social que ce type de (méta-)phénomènes peuvent se produire, ce qui n'est effectivement pas le cas dans les sciences naturelles.

Enfin, il nous faut aborder le pouvoir performatif que peuvent avoir les variables statistiques elles-mêmes. Le développement des statistiques est intrinsèquement lié aux enjeux sociaux et politiques de mesure visant à un certain *contrôle* (DESROSIÈRES, 2014). C'est en cela que les variables utilisées, même les plus "basiques", comme on l'a déjà remarqué plus haut, ne sont jamais données de l'extérieur du monde social, en toute neutralité, mais émergent toujours d'un certain cadre, d'une certaine manière de poser certaines questions *dans certains buts précis*. Ainsi, les variables font elles-mêmes montre d'une certaine performativité, notamment lorsque leur utilisation s'insère dans une injonction politique de résultats quantifiés.

Le cas emblématique de ce phénomène est bien sûr la mesure de la délinquance. Ce concept, tout aussi *épais* que celui de "terrorisme", est, déjà, loin d'être évident à opérationnaliser. De plus, cette amplitude conceptuelle se traduit dans les faits par une certaine marge de manœuvre laissée aux acteurs chargés de la mesurer, ici les forces de l'ordre. Si, enfin, ceux-ci se retrouvent

6. L'expression n'est pas de moi, mais je ne me souviens plus de son origine. J'ai trouvé un article de blog qui utilise le même terme pour décrire le même phénomène : <https://lehollandaisvolant.net/?d=2022/12/12/18/19/41-edf-leffet-bison-fute-ou-comment-esperer-a-tout-prix-se-tromper>.

pris, comme c'est le cas depuis le début des années 2000 en France, dans une politique du chiffre, c'est-à-dire dans un certain nombre d'injonctions à produire des résultats quantifiés reflétant l'efficacité des mesures mises en œuvre pour lutter contre la délinquance (augmentation du nombre de cas résolus, baisse de la délinquance, etc.), alors on peut s'attendre à une certaine forme de prophétie auto-réalisatrice.

Concrètement, l'incitation à des résultats particuliers va modifier le comportement des agents qui vont alors adopter certaines stratégies visant à maximiser leur gain (monétaire ou bien en termes de promotion professionnelle). La variable "délinquance" sera alors manipulée afin de servir cette fin particulière, et ne servira alors plus à mesurer le plus objectivement possible un phénomène socio-politique.

L'analyse de Laurent Mucchielli (MUCCHIELLI, 2008) à propos de la politique de lutte contre la délinquance déployée par Nicolas Sarkozy à partir de 2002 détaille les différents phénomènes sus-cités. Il illustre en particulier la performativité de la variable "délinquance" une fois adjointe aux injonctions de résultats formant le corps des discours que le tout nouveau ministre de l'intérieur a prononcé fin juin 2002 devant des milliers de cadres de la police, puis début juillet devant les gendarmes :

A peine le ministre était-il installé dans son bureau et avant même qu'il ait pu prendre la moindre mesure, rédiger la moindre circulaire ou signer le moindre décret, que la statistique, après des années de hausse continue, indiquait pour le mois de mai 2002 une baisse miraculeuse de 0,69 % en zone police (et même de 1,6 % pour la Préfecture de police de Paris). Le mois suivant, alors que la politique du ministre en était toujours au stade des discours, la baisse devenait spectaculaire et atteignait 7,38 % chez les policiers (8,6 % à Paris). (MUCCHIELLI, 2008, §17)

Ce constat semble lui aussi *propre* aux sciences humaines et sociales. Bien sûr, les enjeux économiques et politiques interviennent profondément dans le développement et l'utilisation des sciences naturelles, en particulier en sciences physiques : on peut penser à la thermodynamique durant les révolutions industrielles ou encore à la physique nucléaire durant la seconde guerre mondiale. Cependant, il semble beaucoup plus difficilement concevable pour les théories physiques et les connaissances produites d'avoir directement un tel effet sur le niveau de réalité (physique, donc) sur lequel elles portent.

3.3.2 Un système pluraliste

À la présentation de ces différentes familles d'arguments, il faut remarquer une chose importante : les lignes argumentatives défendant une épistémologie dualiste se positionnent à différents degrés : la difficulté, l'impossibilité et l'absence de pertinence d'une autre épistémologie.

La première consiste à soutenir que certaines propriétés spécifiques des sciences humaines et sociales rendent plus difficiles qu'ailleurs l'application de la même épistémologie, ou plus précisément le développement de cadres théoriques similaires (dans leur forme globale et leur mode de justification) que, par exemple, dans les sciences physiques.

La seconde consiste à énoncer et à argumenter l'*impossibilité* d'une épistémologie unitaire, c'est-à-dire l'impossibilité de mettre en œuvre les mêmes approches et méthodes empiriques en sciences sociales qu'en sciences physiques (par exemple), et l'impossibilité que les théories du monde social ressemblent, dans leur forme comme dans leur mode de justification, aux théories physiques.

On peut opposer à ces deux stratégies argumentatives l'*existence* de tout un ensemble de travaux, empiriques et théoriques, montrant qu'il est tout à fait possible de faire peu ou prou la même chose qu'en sciences physiques, comme par exemple mettre en exergue de fortes régularités empiriques, expliquer les phénomènes sociaux à l'aide de modèles idéalisés, et ainsi de suite. C'est en partie ce à quoi le chapitre 5 est consacré.

La troisième ligne argumentative, quant à elle, est plus difficile à réfuter – par construction. Elle prétend que ce n'est pas autant la difficulté ou l'impossibilité d'une telle épistémologie qui est problématique, mais bien son *absence de pertinence*. Au vu de la nature *chargée de signification* de ses objets (par exemple), il n'est pas pertinent, pour la sociologie ou les sciences sociales en général, de chercher à isoler des régularités qui seraient indépendantes du contexte historique de leur instanciation concrète, car en faisant cela on passe à côté, justement, du caractère social et culturel des objets étudiés. En faisant cela, en d'autres termes, on vide la démarche de tout contenu *précisément sociologique*.

Cette position, sur laquelle je vais me concentrer ici, n'est pas aussi simple à opposer "de l'extérieur" puisqu'elle repose, comme on va le voir, sur une argumentation circulaire. Ce n'est pas forcément péjoratif, c'est simplement qu'elle se présente sous la forme d'un *système philosophique* prenant position, classiquement (MAHNER, 2007, p. 524), sur trois aspects : *ontologique* (quelle est la nature des objets que l'on étudie), *épistémologique* (à quoi ressemble

la connaissance que l'on peut avoir de ces phénomènes) et *méthodologique* (comment faire concrètement pour produire une telle connaissance).

Bien qu'il soit possible, dans une certaine mesure, de critiquer les arguments cités ci-dessus individuellement, il faut donc bien retenir que la plupart du temps, ces trois aspects se renforcent les uns les autres, formant un *système* possédant une certaine cohérence interne. Je vais proposer une synthèse des positions pluralistes de ce type en mettant en exergue sa structure systémique.

Précisons toutefois une chose très importante : comme je l'ai annoncé, il ne s'agit pas là d'une position philosophique qui a été tenue de manière stricte, telle quelle, par des auteurs bien identifiés. On l'a vu, il y a toute une diversité de positions et traditions pluralistes, et les auteurs rencontrés dans la littérature tiennent très souvent des positions mixtes. En d'autres termes, il s'agit ici de la reconstruction artificielle d'une position pluraliste qui n'a pas forcément existé sous cette forme précise, mais qui est proposée ici comme une bonne représentation clarifiée de cette position, et qui suffira donc pour lui opposer des contre-arguments.

Cette position, la voici :

Les objets des sciences sociales (les individus, les acteurs sociaux, les institutions, les phénomènes sociaux, la société, ...) sont par nature des êtres de signification, signification intrinsèquement indexée à un contexte historique particulier, voire même à la "totalité sociale", hors desquels elle se vide de sa substance (positionnement ontologique). De ce fait, la connaissance que l'on peut en acquérir se focalise nécessairement sur les singularités puisque ce sont ces singularités, et non les régularités transculturelles ou transhistoriques, par exemple, qui contiennent l'information pertinente, et qui constituent donc le but de l'investigation scientifique. En d'autres termes, ces singularités ne sont pas des contingences historiques ou des fluctuations statistiques qu'il s'agirait d'effacer au profit d'une régularité sous-jacente, comme en sciences physiques, mais constituent, de par la nature même des phénomènes étudiés, la connaissance effective à atteindre (positionnement épistémologique). Ainsi, les méthodes expérimentales, où l'on recrée artificiellement un contexte social miniature, épuré et simplifié, réduit à la manipulation de quelques variables, pour voir comment un échantillon d'individus réagit à des modifications contrôlées de ces variables, ne peut pas produire une connaissance véritable du monde social, mais simplement une forme de pseudo-connaissance, une sorte d'avatar grotesque désespérément envieux

des connaissances nomologiques produites avec succès dans d'autres domaines comme la physique qui s'y prêtent mieux de par la nature de leur objet et les contraintes épistémologiques qui leur sont associées. Il faut alors leur préférer des méthodes compréhensives, interprétatives, qui reconstruisent le sens que les acteurs donnent à leurs comportements, sans jamais figer ce sens dans un cadre conceptuel dont la validité pourrait de trop dépasser un contexte historique, social ou politique particulier (positionnement méthodologique).

Présenté ainsi, le positionnement ontologique légitime, ou du moins forme une base de justification aux deux autres positionnements : si la connaissance du monde social ne peut que ressembler à cela, et que donc les méthodes à utiliser pour la produire sont celles-ci, c'est parce que le monde social *est* comme ceci. Cependant, les positions épistémologiques et méthodologiques renforcent également le postulat ontologique. En effet, la multitude des objets et phénomènes humains et sociaux – comme, d'ailleurs, tous les autres – nous apparaissent a priori dans une grande diversité. Ainsi, si l'on souhaite se concentrer sur ce qui les distingue les uns des autres, on pourra toujours trouver de quoi satisfaire cet objectif. La décision épistémologique citée au-dessus, mise en œuvre par des méthodes interprétatives et compréhensives, va donc nécessairement confirmer, mécaniquement, le postulat ontologique, c'est-à-dire l'idée que les objets de telles disciplines (l'humain, la société, etc.) sont trop diversifiés pour pouvoir se subsumer sous des lois.

Ainsi, c'est pour cette raison que ces trois positionnements ne peuvent se comprendre qu'en *système* auto-cohérent, ce qui en rend difficile une évaluation extérieure.

3.4 Les réponses classiques

Le principal défaut de ces lignes argumentatives dualistes est qu'elles semblent reposer sur une accentuation artificielle des différences entre les sciences sociales et les sciences naturelles. Ces dernières, en particulier la physique, sont souvent perçues comme ayant plus facilement et naturellement accès à des régularités empiriques, du fait principalement de la nature soit disant "homogène" de ses objets. On prête souvent aux sciences sociales, pour leur part, une capacité particulière à avoir accès plus directement à leurs objets, de par l'homogénéité épistémique supposée entre ces derniers et les chercheurs qui les étudient.

Comme on va le voir, cette double idéalisation (au sens péjoratif du terme) et l'amplification des différences qui en résulte peuvent facilement mener à tirer certaines conséquences épistémologiques que l'on peut juger inadéquates.

3.4.1 Régularités, lois et invariants

Parmi les différentes propriétés distinctives que l'on a citées plus haut, la question des *régularités empiriques* forme la pierre de voute du raisonnement. Celles-ci seraient, pour les sciences portant sur des phénomènes historiques, inexistantes, rares, ou bien juste de bien moindre portée qu'en physique. Les positions pluralistes, comme on l'a déjà énoncé, partent souvent du principe que la réalité sociale et historique est trop chaotique, toujours changeante pour que l'on puisse mettre en lumière de manière pertinente des régularités.

Mais on peut également retourner ce raisonnement, et dire que si l'on ne perçoit pas autant de régularités qu'en physique c'est parce que, *partant du principe* qu'elles n'existent pas ou qu'elles n'ont pas de pertinence, on ne travaille pas assez à la recherche des bonnes variables, du bon cadre empirique qui nous ferait apparaître le monde social comme (un peu) plus régulier.

Comme on l'a vu à travers plusieurs auteurs, la nature même du monde social, toujours changeant, empêcherait de stabiliser l'espace conceptuel que l'on peut lui appliquer. Mais penser cela suppose de considérer que l'on connaît la nature profonde du monde social avant toute investigation rigoureuse et méthodique, en d'autres termes avant toute investigation de type scientifique. Ce postulat est effectivement cohérent dans le système philosophique pluraliste considéré plus haut : on a un accès "direct" ou en tout cas privilégié au monde social à travers la signification que les phénomènes sociaux ont pour nous et la compréhension que l'on peut en avoir.

C'est justement ce système de postulats qui est fortement remis en question dans les réponses monistes ou unitaristes. Comme l'exprime clairement Charles-Henri Cuin :

Si deux fleurs de la même tige de pissenlit paraissent, de prime abord, se ressembler davantage que deux religions occidentales, c'est par pur effet d'une construction de l'esprit rendue plus facile dans le premier cas que dans le second par une similitude des sensations physiques éprouvées par l'observateur. Or, si nous nous représentons aujourd'hui la nature comme un espace de régularités et d'identités, c'est évidemment à la suite d'une longue

série d'expériences pratiques et cognitives dont la plupart nous ont d'ailleurs été enseignées. (CUIN, 2000, p. 105)

Alain Testart énonce une chose similaire (TESTART, 1991, p. 11) "ce n'est pas l'identité des choses qui fonde la théorie, c'est la théorie une fois constituée dans la systématisme de ses concepts qui justifie le jugement d'identité." C'est simplement, en réalité, le rappel que toute observation est chargée de concepts et/ou de théorie, comme on en a déjà parlé à la section 2.4.1 : nous n'avons jamais accès à la réalité, qu'elle soit physique, biologique, psychologique ou sociale, d'une manière *immédiate*, mais toujours à travers un certain cadre conceptuel, théorique ou empirique.

La différence entre la physique et la sociologie ne résiderait donc pas, de ce point de vue, dans la nature de leurs objets mais bien dans le niveau de développement des cadres empiriques à l'intérieur desquels on observe ceux-ci.

On peut avoir l'impression que le monde physique *est* bien plus régulier que le monde social, mais on pourrait également penser qu'il nous *apparaît* simplement plus régulier car on l'observe à travers un cadre particulier qui semble évident peut-être aujourd'hui mais qui était tout sauf donné au départ. On peut par exemple repenser au concept de température, dont l'opérationnalisation nous apparaît évidente et facile aujourd'hui, mais qui est en réalité passé par tout un chemin historique tortueux et a dû surmonter des obstacles techniques et conceptuels importants (CHANG, 2004).

Un autre exemple peut-être encore plus parlant est celui de la trajectoire des planètes dans le système solaire. Celles—ci nous apparaissent régulières (par exemple, approximativement en forme d'ellipses) uniquement si on les mesure avec une certaine marge d'erreur, ou sur une durée limitée. Pourtant, si on relève leur mouvement de manière plus précise (en parvenant à observer la position des planètes avec de meilleurs télescopes) et pendant plus longtemps (disons, des dizaines voire des centaines d'années), leurs trajectoires nous apparaissent alors beaucoup moins elliptiques – et c'est là que la théorie explicative doit alors prendre le relai.

La régularité perçue d'un phénomène dépend donc *aussi* des lunettes que l'on porte pour l'observer.

Un dernier exemple qui est beaucoup moins trivial qu'il n'y paraît : le concept de *distance* en physique. Comme l'écrivent Delori et al., en citant Alain Desrosières, dans la discussion déjà citée en section 3.2.1 à propos de la difficulté d'opérationnaliser de manière quantitative certains concepts comme le terrorisme :

Cet auteur donne l'exemple de la différence entre la mesure de la distance Terre-Lune et celle, par les sondages, de l'opinion publique. On peut facilement convenir du fait que l'opération de quantification n'est pas de la même nature. La distance Terre-Lune existe indépendamment de sa quantification. (DELORI et al., 2021, p. 39)

Même si cette remarque est tout à fait juste en un sens – mesurer une distance ne modifie pas la distance en elle-même, alors que le fait même de sonder la population sur une question peut avoir un impact sur les opinions rapportées – il est toutefois assez ironique de noter que depuis le développement des théories de la relativité, et en particulier la relativité générale, la notion de distance est beaucoup moins "évidente" que dans son sens commun. Comme je vais le montrer rapidement, il semble que la citation ci-dessus tombe encore dans une simplification qui voudrait que mesurer le monde physique est toujours plus simple en soi que mesurer le monde sociale et politique.

En effet, il se trouve que depuis 1983, à la suite de la 17^{me} Conférence Générale des Poids et Mesure, il a été décidé de redéfinir le mètre à partir de la mesure de la seconde (devenue extrêmement précise avec le développement des horloges atomiques) et de la vitesse de la lumière *définie* à présent comme étant strictement *égale* à $c=299\,792\,458$ m/s.

Or, il se trouve que l'on sait, au sein de la relativité générale, que le temps propre d'un observateur dépend du champ local de gravitation dans lequel il est plongé. De plus, pour mesurer la distance Terre-Lune, une manière classique de faire est d'envoyer un laser, depuis la Terre, sur un réflecteur placé sur le sol lunaire, et de mesurer (très précisément) le temps Δt que ce laser met pour faire l'aller-retour. La distance Terre-Lune sera alors égale à $2c/\Delta t$.

Pendant, comme on l'a dit, la durée Δt dépend du champ de gravité dans lequel est placée l'horloge atomique qui mesure cette durée. Il s'en suit alors, de manière tout à fait étrange à notre intuition première, que "la" distance Terre-Lune *dépend en réalité de l'endroit où on la mesure*. Donc, dire que "La distance Terre-Lune existe indépendamment de sa quantification" manque de justesse. Bien sûr, cela n'enlève absolument rien à ce qui est défendu par ailleurs en termes de méthodologie, mais cela sous-tend encore une fois une conception de la mesure qui serait davantage "donnée" en physique, en contraste avec le monde social et politique.

Il n'en est rien. Pour insister davantage, concernant la distance, il suffit de

s'aventurer dans n'importe quel ouvrage d'introduction à la cosmologie moderne (LIDDLE, 2000) pour réaliser que ce concept est très loin d'être donné tel quel, par un geste amical du réel physique, mais qu'il est au contraire sujet à d'âpres débats et à un certain nombre d'obstacles méthodologiques tout à fait fondamentaux (CZERNY et al., 2018; RIESS, 2020).

Ainsi, le niveau de régularité observé dans le monde dépend du cadre conceptuel ou empirique qu'on lui applique, c'est-à-dire les concepts, grandeurs et variables de base que l'on se donne pour l'observer, ainsi que leur définition opérationnelle. Ce constat vaut même dans un cas paradigmatique comme les mouvements célestes qui sont censés être parmi les plus réguliers *et par cette régularité permettre l'émergence de la science astronomique telle qu'on la connaît*, ou bien le concept si "évident" à première vue de distance.

On note ici une chose importante et qui va rester au cœur de mon travail : le langage avec lequel on parle du monde *et* ce que l'on dit du monde *co-évoluent* en réalité au cours de la progression scientifique.

Le même phénomène, comme on peut s'y attendre, survient donc en sciences sociales.

Des travaux relativement récents (FORTUNATO et CASTELLANO, 2007; CHATTERJEE, MITROVIĆ et FORTUNATO, 2013) défendant la pertinence d'acertains phénomènes collectifs en utilisant des outils ou approches issus de la physique statistique en offrent une illustration d'une clarté cristalline.

Les auteurs abordent la détection de *patterns* universels dans la distribution des votes lors d'un certain type d'élections : des élections proportionnelles avec circonscriptions multi-sièges et listes ouvertes. En d'autres termes, les électeurs votent pour un candidat qui reçoit un certain nombre v de votes, qui appartient à une certaine liste composée de Q candidats au total, et les listes sont ensuite représentées proportionnellement au nombre total de votes N qu'elles ont reçues – les candidats sont ensuite élus, à l'intérieur d'une liste, en fonction du nombre de votes qu'ils ont obtenu.

En s'intéressant à de nombreux cas pour lesquels ce type d'élections a eu lieu, à savoir des pays culturellement et historiquement très différents, à des moments également différents de leurs histoires,⁷ ces auteurs montrent que la distribution des votes par candidat apparaît malgré tout suivre *la même loi* avec une particulière précision (Cf figure 3.2).

7. L'Italie en 1958, 1972, 1976, 1979 et 1987, la Pologne en 2001, 2005, 2007 et 2011, la Finlande en 1995, 1999, 2003 et 2007, le Danemark en 1990, 1994, 1998, 2001, 2005, 2007 et 2011 et l'Estonie en 2003, 2007 et 2011.

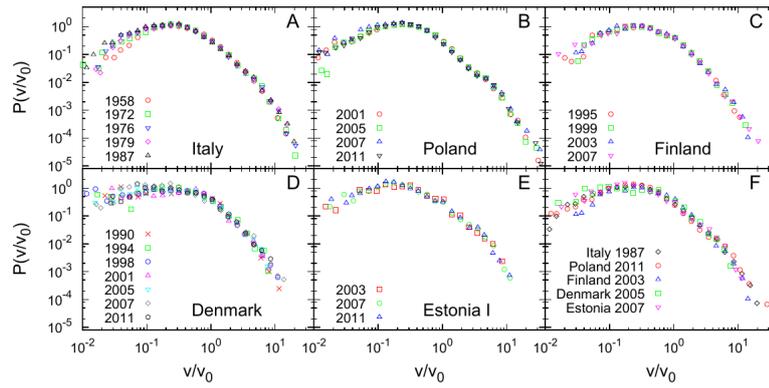


FIGURE 3.2 – Figure 1 de (CHATTERJEE, MITROVIĆ et FORTUNATO, 2013, p. 3) mettant en exergue les pattern réguliers dans la distribution des votes pour différents pays sur différentes années.

Plus précisément, la distribution des votes par candidat $P(v, Q, N)$ ne semble dépendre que d'une seule variable vQ/N où $v_0 = N/Q$ représente le nombre moyen de vote obtenus dans la liste du candidat en question :

$$P(v, Q, N) = F(vQ/N) = \frac{N}{\sqrt{2\pi\sigma vQ}} e^{-(\log(vQ/N) - \mu)^2 / 2\sigma^2}, \quad (3.3)$$

(avec $\mu = -0.54$ et $\sigma^2 = 1.08$). Il semble se dessiner un pattern, à l'échelle macroscopique, indépendant du pays ou de l'année en question. Ce constat est à première vue très étonnant, puisque le phénomène étudié est un phénomène social et qu'il existe une forte diversité culturelle et historique entre ces différents pays, à ces différentes époques.

Or, une telle régularité *n'apparaît pas* a priori lorsque l'on regarde la distribution "brute" $P(v, Q, N)$ du nombre de votes v reçus par candidat (voir figure 3.3 A), distribution qui dépend aussi des paramètres Q (le nombre de candidats dans la liste du candidat en question) et N (le nombre de votes total reçu par tous les candidats de la liste du candidat en question).

Cependant, lorsque l'on étudie la distribution *rescaled* $F(vQ/N)$ ne dépendant plus que de l'unique variable vQ/N (voir figure 3.3 B), émerge alors une régularité saillante qui ne semble dépendre que de la structure du système électoral et plus du tout, ou très peu, du contexte historique ou politique particulier.

Il s'agit donc bien d'une illustration du phénomène déjà décrit plus haut par plusieurs auteurs : le niveau de régularité perçu dans le monde dépend du cadre empirique à travers lequel on l'observe.

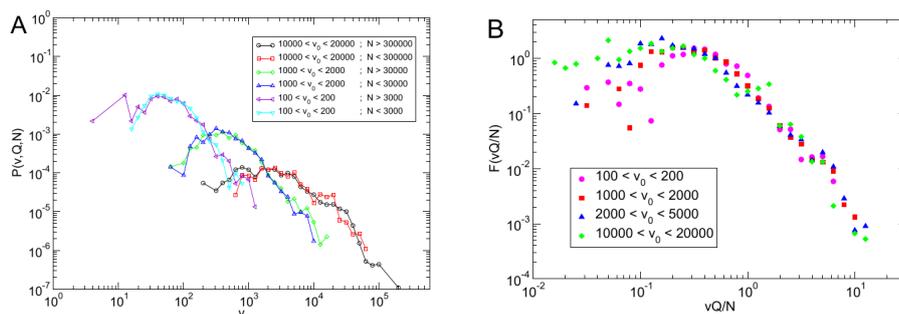


FIGURE 3.3 – Figure 1 de (FORTUNATO et CASTELLANO, 2007, p. 2). Gauche : tracé de la distribution "brute" $P(v, Q, N)$, résultant en un pattern irrégulier. Droite : tracé de la distribution "rescaled" $F(vQ/N)$ en fonction de la seule variable vQ/N , résultant en un pattern régulier. "Data refer to the Italian parliamentary elections in 1972, but we obtained very similar results from the analysis of each dataset."

De plus, l'identification d'une telle régularité n'empêche cependant absolument pas les auteurs de tenir un raisonnement sociologique :

This lack of universality [of earlier studies] is a consequence of the fact that the number of votes a candidate receives is the combination of two distinct factors : how many of the total number of electors vote for the candidate's party and the personal appeal of the candidate within the restricted pool of voters for his/her party. The first factor strongly depends on policy-related issues : typically voters know the position of all parties with respect to the political issues they deem more relevant and they select the party that best matches their personal views. The second factor is instead practically independent from political issues. Since candidates of the same party mostly share a common set of opinions on ethical, social and economical issues, the selection of a specific candidate has not to do with such issues, rather it depends on a "personal" interaction between the candidate and the voters. Typically voters know at most a few of the candidates in their party list, and in this small subset they select the one they will support. Successful candidates are those able to establish some form of direct or indirect contact with many potential voters during the electoral campaign. This type of opinion dynamics is likely to give rise to universal phenomena. (FORTUNATO et CASTELLANO, 2007, p. 1-2)

On pourrait même aller jusqu'à dire que c'est le contraire : la détection de

telles régularités est *consubstantielle* du raisonnement sociologique qu'il est pertinent de tenir sur un phénomène donné, en le guidant et en le structurant fondamentalement.

Cette réflexion peut être mise en parallèle avec un autre argument soutenant l'importance de mettre au jour des régularités ou des lois : le fait que l'étude de variations supposent nécessairement l'existence d'invariants sous-jacents.

Cet argument a été développé notamment par Bernard Lahire récemment (LAHIRE, 2023). Selon Lahire, la frilosité que les sciences sociales expriment vis-à-vis de la recherche d'invariants peut en partie s'expliquer par l'équivalence supposée entre l'existence de tels invariants et l'invariabilité du monde social qui en découlerait.

En réalité, comme il l'exprime très clairement au chapitre 4 (intitulé "Lois, principes, invariants"), les lois et mécanismes invariants mis en exergue en physique ou en biologie n'empêchent absolument pas de rendre compte des variations temporelles de l'univers dans son ensemble ou de la vie terrestre. *Au contraire* : ces principes invariants sont justement ce qui permet de rendre compte du changement, d'inscrire même la diversité apparente du monde au sein d'un cadre qui lui donne du sens :

L'un des grands points de résistance des chercheurs en sciences sociales est lié au fait que, pour eux, qui dit "lois générales" ou "invariants" dit "stabilité" ou "état permanent", alors que ce qui constitue à leurs yeux l'objet central de leurs sciences, c'est la variation culturelle, l'histoire, le changement permanent. En pensant cela, ils ne se rendent pas compte que l'univers et le vivant sont eux aussi en transformation permanente. Ce que l'univers nous donne à voir aujourd'hui ne ressemble en rien à ce qu'il était au moment du Big Bang et à différents moments de son histoire [...] La variation, la transformation, le changement ou l'histoire ne sont pas incompatibles avec l'existence de lois générales qui agissent de façon continue depuis le début de la Terre, de la vie ou de l'humanité [...] L'existence des mêmes lois qui agissent tout le temps et partout n'implique pas la répétition à l'identique des mêmes états (de la matière, de la vie ou des sociétés). Ces lois sont précisément ce qui explique les changements permanents. (LAHIRE, 2023, p. 122-123).

Il s'agit donc d'un contre-sens complet de croire que la recherche et la mise en exergue d'invariants (physiques, biologiques, sociaux) impliqueraient

une vue figée et invariable des phénomènes correspondants. Il décèle d'ailleurs une sorte de contradiction au fait de se focaliser systématiquement sur les variations sans jamais chercher d'invariants :

Mais les variations ont-elles un sens indépendamment des invariants à partir desquels elles se déploient? Et si toute variation suppose en toute logique un invariant, pourquoi les chercheurs ne s'interrogent-ils pas sur l'origine et la nécessité de ces invariants, ou considèrent-ils leur étude comme très secondaire? Prendre conscience du fait que ce qui varie repose sur de grands invariants propres (ou non) à l'espèce ouvre la possibilité de ne pas donner le même sens à l'étude des différentes variantes. (LAHIRE, 2023, p. 26)

On pourrait reformuler l'idée sous-jacente ainsi : il n'y a pas d'hétérogénéité intéressante sans un minimum d'homogénéité. Cela rejoint les remarques plus générales à propos des analyses conceptuelles que j'ai données dans l'introduction de cette thèse : toute classification conceptuelle présuppose, même implicitement, la donnée d'un ensemble d'objets sur lesquels il est intéressant (c'est-à-dire non trivial) de se demander s'ils appartiennent au concept ou pas.

Il se passe la même chose ici : sous-entendre qu'il est pertinent de dire qu'il y a des variations entre les cultures humaines, c'est supposer déjà que ces cultures humaines ont en commun, malgré tout, certaines caractéristiques – sans quoi, cette remarque devient parfaitement triviale. Ainsi, même les positions philosophiques soutenant que le monde social et historique est trop chaotique et diversifié pour pouvoir mettre au jour des régularités doivent tenir pour vrai, sous peine de voir leur remarque se trivialisier, le fait que malgré tout derrière cette diversité il y a des choses communes – ne serait-ce que pour parler du "monde historique".

Dans ce cas-là, cependant, ce discours semble rentrer en contradiction avec son postulat de départ. Il semble pris dans un dilemme fatidique entre une position triviale et une position logiquement contradictoire.

Afin de clarifier cette discussion entre variations observées et invariants sous-jacents, on peut faire une remarque importante à propos des "régularités" ou des "lois" dont seraient privées les sciences sociales : on distingue typiquement *deux* types de lois : les lois empiriques et les lois théoriques (FEIGL, 1970; CARTWRIGHT, 1983; MAYO, 1996).

En physique, par exemple, la "loi des gaz parfaits" $PV = nRT$ liant diverses caractéristiques macroscopiques mesurables d'un gaz (sa pression P ,

son volume V , sa quantité de matière n et sa température T) est une *loi empirique* – qui n'est valable, de plus, que dans *certaines* conditions expérimentales.

Pour expliquer cette loi à partir de considérations microscopiques, on suppose que le gaz est composé de particules dont les vitesses sont distribuées selon la loi de Maxwell-Boltzman :

$$d\mu(v_x) = \left(\frac{m}{2\pi k_B T}\right)^{1/2} \exp\left(-\frac{mv^2}{2k_B T}\right) dv_x, \quad (3.4)$$

où k_B est la constante de Boltzman. Plus précisément, (3.4) donne la probabilité $d\mu(v_x)$ qu'une particule ait sa composante de vitesse selon x comprise entre les valeurs v_x et $v_x + dv_x$. Il s'agit d'une *loi théorique* à partir de laquelle – adjointe à quelques hypothèses auxiliaires – on peut retrouver $PV = nRT$.

En d'autres termes, les lois empiriques constituent les *explicanda* en physique (ce qu'il faut expliquer), tandis que les lois théoriques font partie des *explicans* (ce qui explique). Elles n'ont donc pas le même statut épistémologique.

C'est une remarque importante lorsque l'on transpose la discussion aux lois en sciences humaines et sociales. On a parlé plus haut de régularités empiriques, donc du premier type de lois. Mais le même type d'arguments (pluraliste ou non) s'appliquent aux lois théoriques.

En sociologie, de telles lois théoriques se retrouvent par exemple au coeur de ce que l'on appelle les théories de l'action, comme par exemple la théorie du choix rationnel. De telles théories donnent, dans des conditions données, la probabilité que les agents agissent de telle ou telle façon. Par contraste, les lois empiriques correspondraient dans ce cas-là aux régularités macroscopiques que l'on pourrait mettre en évidence.

Les mêmes arguments s'appliquent, donc : on pourrait soutenir de la même façon que les comportements humains sont si diversifiés qu'il est impossible (ou, encore, non pertinent) de chercher à les interpréter au sein de cadres nomologiques (cherchant et s'appuyant *par décision* épistémologique et méthodologique sur les régularités), et non herméneutiques (s'appuyant sur le sens que les acteurs donnent à leurs actions, sens toujours renouvelé, jamais figé, toujours indexé sur un contexte social ou historique, etc.)

On peut alors répondre de la même façon que pour les régularités empiriques : la question n'est en fait pas de savoir si de telles lois "existent" dans le monde social, mais si on *peut* trouver *le* ou *un* bon langage théorique pour

mettre de l'ordre dans nos observations. La question du statut épistémologique de telles lois ou principes fondamentaux est au cœur de mon travail, on y reviendra donc en détail plus loin, en particulier au chapitre 5.

Il est un élément – décisif – que j'ai glissé à plusieurs reprises dans le raisonnement ci-dessus, et qu'il s'agit maintenant d'explicitier. Toutes les lois, toutes les régularités empiriques que l'on pourra jamais mettre en exergue, quelle que soit la discipline, ont un certain *domaine de validité*.

Par exemple, il n'existe pas (et ne peut probablement pas exister) de lois universelles du comportement humain comme de la gravitation. L'erreur de raisonnement provient probablement du fait que cette dernière ainsi que la plupart des lois théoriques en physique sont, par principe, valables en tout point de l'espace et du temps. En d'autres termes, leur domaine de validité n'est pas limité spatio-temporellement.

Pour autant, elles ont quand même un domaine de validité restreint. On a pensé très longtemps que la théorie de la gravitation newtonienne, par exemple, était indépassable. En fait, comme toute la mécanique dite classique qui l'englobe, elle ne fonctionne bien que sous certaines conditions : il faut que $v/c \ll 1$, c'est-à-dire que la vitesse v des systèmes physiques en jeu soit très faible devant la célérité c (qui correspond à la vitesse de la lumière dans le vide), et que les champs de gravitation en jeu soient assez faibles, ce que l'on note habituellement $\Phi/c^2 \ll 1$ où Φ est le potentiel gravitationnel.

Tout cela dépend bien sûr du niveau de précision que l'on souhaite atteindre. Si l'on est dans ces conditions, alors, certes, ces lois sont "universelles" dans le sens où elles ne sont plus dépendantes des coordonnées d'espace-temps dans lesquelles on les exprime. Pourtant, ce n'est déjà *plus* le cas si on veut faire des GPS précis ou décrire correctement la trajectoire de la planète Mercure, qui ne sont pourtant pas les phénomènes les plus exotiques que l'on trouve dans l'univers.

Les lois empiriques répondent à la même contraintes : la loi des gaz parfaits $PV = nRT$ est valable dans un certain domaine, plus précisément dans le cas où la pression des gaz en question est assez faible. Plus la pression est élevée, moins la loi est valable.

Ainsi, toutes les lois sont conditionnées à un certain domaine de validité. La seule différence avec les "lois" que l'on pourrait mettre en évidence en sciences sociales, pourrait éventuellement être que les conditions de validité soient limitées dans le temps (historique) et l'espace (géographique).

Mais ça n'est même pas forcément le cas. On a vu qu'on pouvait détecter des régularités empiriques dans le monde social qui ne soient que très

peu dépendant du contexte politique ou historique. On pourrait aussi imaginer (cela ne veut pas dire que c'est forcément pertinent, possible ou exempt de critiques) que les lois fondamentales du comportement humain ne soient elles-mêmes pas dépendantes du contexte historique mais, par exemple, directement héritées de l'évolution biologique de l'espèce humaine.

Cependant, ces lois, aussi générales soient-elles, vont de toute façon probablement s'instancier différemment selon le contexte historique, social et politique. C'est ce qu'il se passe en physique : même si la loi fondamentale de la mécanique reliant l'accélération d'un corps aux forces qui s'y exercent est invariante (par principe) par translation dans l'espace, un même objet lâché sur Terre d'une même hauteur par rapport au sol ne tombera pas aussi vite au pôle nord qu'à l'équateur, car les conditions expérimentales (traduites par les hypothèses auxiliaires du modèle correspondant) ne sont pas les mêmes.

Pour résumer, tout cela revient à ce que l'on a souligné plus haut : l'existence ou non de lois, empiriques comme théoriques, n'est pas (uniquement) donné *par le réel* que l'on investigate (puisque l'on l'investigue justement pour le connaître, on doit donc se laisser éventuellement surprendre et non pas lui accoler toutes nos idées préconçues) mais (aussi) *via une décision épistémologique* et méthodologique.

Le débat sur l'unité épistémologique des sciences se déporte alors de discussions métaphysiques interminables à la question, bien plus abordable, de la pertinence d'une telle décision. Il ne s'agit pas de chercher des lois qui vaudraient en tout temps, en tout lieu et en toute condition, car cela n'existe dans aucune science, *surtout pas en physique*. Insistons et rappelons que dans cette discipline, les plus grandes avancées correspondent justement à des cas où l'on a compris qu'on travaillait en fait au sein d'un certain domaine de validité, et que l'on commençait, avec telle ou telle observation récalcitrante, à en toucher le bord.

Il s'agit davantage de poser cette recherche comme un principe – et non pas de la rejeter a priori – et de voir jusqu'où cela nous mène. Bien sûr, de la même façon que la position pluraliste décrite plus haut est un système possédant une certaine cohérence interne qu'il est difficile de directement attaquer telle qu'elle de l'extérieur, la position développée dans ce manuscrit a la même propriété.

Nous faisons donc face ici à un problème de méthode philosophique : il semble impossible de trancher directement entre ces deux systèmes philosophiques en les confrontant directement l'un à l'autre, puisque chacun possède déjà, structurellement, les réponses de principe à apporter à chacune

des oppositions. Tout ce que nous pouvons faire, c'est donc, comme on l'a déjà vu au dessus, poser cela comme un principe et en explorer les diverses conséquences.

L'impossibilité de confronter directement, dans un cadre plus large qui les engloberait, ces deux systèmes philosophiques fondés sur des postulats contraires explique probablement la stérilité apparente des débats épistémologiques comparatifs des différentes disciplines – comme par exemple celui ayant porté le nom de *Science Wars* dans les années 1990.

3.4.2 Conséquences épistémologiques

Pour terminer ce chapitre, nous allons donc appliquer notre méthode, qui consiste à adopter le postulat moniste et observer les conséquences que l'on peut en tirer pour les diverses familles d'arguments pluralistes présentées au début du chapitre. Mon objectif, pour évaluer sa pertinence, est notamment de montrer en quoi ce point de vue semble résoudre, en principe, un certain nombre de problèmes.

J'ai présenté au début de ce chapitre les différents arguments pluralistes comme étant basés sur un ensemble de propriétés censées être spécifiques aux sciences humaines et sociales, et divisées en quatre familles : l'historicité, la signification, l'intentionnalité et la performativité. Si on adopte le point de vue moniste, ou unitaire, c'est-à-dire celui consistant à décréter par principe que le but premier de la connaissance scientifique est d'être organisée sous forme de lois, de régularités de portée plus ou moins étendue, alors on parvient à neutraliser les conséquences épistémologiques prétendues être déduites de ces propriétés spécifiques.

L'historicité

Même si l'historicité restera toujours une source de problèmes méthodologiques qui sont bien sûr loin d'être triviaux, elle n'apparaît plus comme une limitation épistémologique fondamentale et spécifique aux sciences sociales, pour deux raisons.

D'une part, il existe un certain nombre d'exemples de régularités empiriques que l'on a réussi à mettre au jour, une fois adopté la pertinence de cette recherche comme principe. D'autre part, même si le phénomène que l'on étudie ne présente pas de fortes régularités apparente, voire même que l'on souhaite étudier son évolution temporelle, on peut alors défendre que la recherche de lois théoriques va permettre d'éclairer des situations historiques

a priori différentes comme étant des instanciations, sous certains aspects, de cette même loi.

On a déjà rencontré cet argument plus haut concernant la physique ou la biologie. Les données d'observation cosmologique ne prennent sens comme traces d'une longue évolution qu'à travers un cadre théorique qui permet, justement, de reconstruire cette évolution. La diversité des formes de vie sur Terre, pour leur part, et leur adaptation à leur milieu, extraordinaire à première vue, ne prennent sens qu'à travers un cadre théorique qui permet, de même, de rendre compte d'une évolution sur laquelle on ne peut pas expérimenter directement.

Ainsi, avoir des lois sociologiques théoriques, c'est-à-dire des mécanismes théoriques dont on connaît bien le domaine de validité, permettrait non pas de s'extraire de l'historicité intrinsèque des sciences sociales, historiques et politiques, mais bien au contraire de lui rendre grâce de la meilleure des manières : en lui donnant sens à travers un cadre théorique robuste.

La signification

Pour les mêmes raisons, le fait que les phénomènes ou les faits sociaux, politiques et historiques soient signifiants pour les agents n'implique pas de problèmes autre que méthodologiques.

Comme on le voit avec l'exemple du terrorisme, le fait d'utiliser un concept défini à partir du sens précis qu'en donnent certains acteurs sociaux n'est pas un problème, au contraire : dans le contexte précis de la recherche en question, cela rend son utilisation *encore plus pertinente*.

Pour autant, de manière plus générale, on peut remarquer que rien ne nous oblige a priori à nous tenir strictement à ces significations pour faire de la science. Bien sûr, il faut bien partir de quelque part, et les significations brutes, communes, sont parfois de bons points de départ. Mais si notre but est de mettre au jour des mécanismes causaux, c'est-à-dire des modèles de comportements humains et de phénomènes sociaux, alors il est probable que l'on mette au jour des schémas causaux qui ne correspondent pas aux significations ou interprétations spontanées que les agents bâtissent des phénomènes dans lesquels ils sont plongés.

L'intentionnalité

Le même argument s'applique à l'intentionnalité. D'une part, rien n'empêche d'intégrer les finalités poursuivies par les acteurs sociaux dans la modélisation de leur comportement. D'autre part, les intentions ou motivations déclarées par les agents ne sont pas nécessairement celles qui ont un réel effet causal, ou qui sont pertinentes pour expliquer le phénomène en question.

Comme le montre très bien Dominique Raynaud (RAYNAUD, 2021), il existe tout un ensemble de régularités empiriques détectées dans le monde social que l'on peut expliquer par des mécanismes simples d'agrégation sans jamais avoir recours aux intentions précises des acteurs en jeu, ni mêmes aux significations ou interprétations qu'ils pourraient en donner – quand bien même leurs actions constituent pourtant la substance des phénomènes en question.

On peut même aller plus loin : c'est bien une force de l'approche scientifique de pouvoir éventuellement se confronter aux interprétations spontanées, aux significations que les agents donnent des phénomènes qu'ils vivent, et donc de mettre en évidence des phénomènes contre-intuitifs. Sans cela, faire de la science ne servirait pas à grand chose.

Le modèle explicatif sous-tendant l'universalité apparente dans la distribution des votes lors de certains types d'élections, que j'ai déjà évoqué plus haut dans ce chapitre, offre un bon exemple d'un tel cas de figure.

Dans (FORTUNATO et CASTELLANO, 2007), un mécanisme explicatif relativement simple est effectivement proposé pour retrouver la loi macroscopique (4.5). Il est supposé que chaque électeur, à commencer par le candidat lui-même, essaye de convaincre des personnes de son entourage, qu'il y parvient avec une probabilité r , et que les personnes convaincues tentent ensuite de convaincre les personnes de leur entourage, et ainsi de suite. On suppose qu'une personne prise au hasard a une probabilité $p(k) \sim k^{-\alpha}$ d'avoir un nombre k de personnes à convaincre potentiellement dans son entourage, avec $\alpha > 0$, et un nombre minimum k_{min} . Ce modèle, à trois paramètres (r, k_{min}, α) est un modèle classique de *processus de branchement* et reproduit les données avec une remarquable précision, comme le montre la figure 4.2.

Pour rappel, les auteurs (FORTUNATO et CASTELLANO, 2007) supposent un modèle simple où chaque électeur, à commencer par le candidat, tente de convaincre les personnes avec qui il est en contact de voter pour son candidat préféré. La distribution des personnes avec qui une personne donnée est en contact est supposée être en loi de puissance, et la probabilité qu'une personne en convainque une autre est la même pour tout le monde.

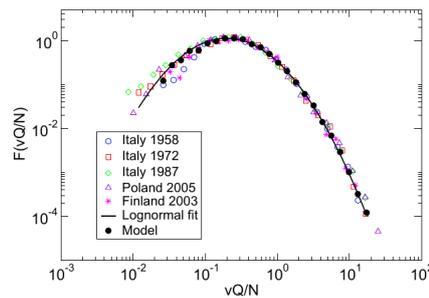


FIGURE 3.4 – Figure 2 de (FORTUNATO et CASTELLANO, 2007, p. 3) comparant des données, l’ajustement par la loi log-normale ainsi que les résultats issus du modèle ($r = 0.25, k_{min} = 10, \alpha = 2.45$).

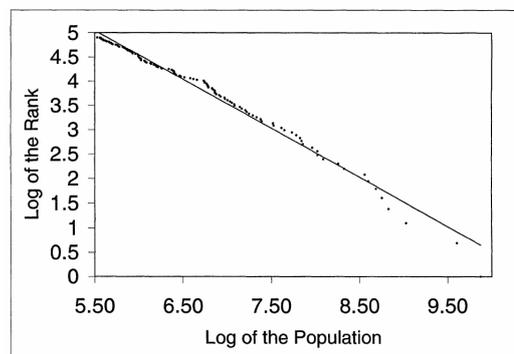


FIGURE 3.5 – Figure 1 de (GABAIX, 1999, p. 740) représentant le logarithme de la population en fonction du logarithme du rang de la ville.

Avec ces simples hypothèses, une fois la dynamique simulée, on retrouve la loi empirique observée avec une grande précision. Pourtant, les hypothèses sur les intentions des individus sont extrêmement minimalistes. De plus, les auteurs peuvent expliquer le phénomène observé sans jamais aller effectivement interroger les individus sur leur façon de procéder pour convaincre d’autres personnes de voter pour leur candidat préféré. Non pas que cela soit inintéressant de poursuivre une telle démarche, mais c’est en tout parfaitement inutile pour expliquer ce phénomène particulier.

Un modèle théorique encore plus minimaliste que le précédent est celui sous-tendant l’émergence systématique de lois de puissance dans la distribution de certains phénomènes, aussi appelé *lois de Zipf* du nom du chercheur qui les a mis en évidence le premier. Empiriquement, on observe que la distribution des villes, dans plusieurs pays très différents et à des époques différentes, suit pourtant une même loi, représenté sur la figure 3.5, tiré de (GABAIX, 1999).

Le modèle explicatif donné par Xavier Gabaix en 1999 (GABAIX, 1999)

offre une bonne illustration de la non pertinence de prendre en compte les intentions ou motivations individuelles précises dans l'explication de *certain*s phénomènes sociaux ou économiques.

Dans son papier, Gabaix prend les villes comme unités élémentaires d'analyse, et suppose simplement qu'elles suivent toutes le même processus de croissance "proportionnel". Chaque ville est supposée croître, d'un pas de temps à l'autre, d'une certaine proportion distribuée de la même façon pour toutes les villes. En d'autres termes, à chaque pas de temps, chaque ville a une certaine probabilité de voir sa population croître ou décroître de $x\%$, x pouvant être positif ou négatif. Cette distribution de probabilité est la même pour toutes les villes. Avec ces simples hypothèses, même en commençant avec des villes ayant toutes des populations similaires, un tel processus va mener à amplifier les petites variations, et la distribution va inexorablement tendre vers une loi de Zipf comme sur la figure 3.5.

Du même coup, un tel exemple illustre également de manière concrète la remarque que l'on a faite plus haut concernant la relation entre invariants, variations et variabilité. En effet, ici on suppose que la même loi (la probabilité de croître d'un certain pourcentage) agit pour toutes les villes, et on observe à la fin du processus une distribution (certes, bien identifiée) hautement hétérogène.

Une telle hétérogénéité se retrouve pour autant expliquée précisément parce que l'on a supposé une certaine homogénéité dans le mécanisme micro.

La performativité

Enfin, en ce qui concerne la performativité, l'adoption du postulat moniste et nomologique fournit également des contre-arguments aux inconvénients qu'elle pourrait engendrer.

Si on se munie de cadres théoriques plus robustes, alors on pourra *d'autant mieux* corriger les biais impliqués dans toute démarche d'enquête empirique directe : expérimentations, observations participantes, entretiens, questionnaires, etc.

De manière plus générale, le fait qu'une information sur le monde social (par exemple, les résultats d'un sondage) modifie le comportement des agents et in fine un état macrosocial (l'issue d'une élection) n'implique pas que toute loi du comportement est impossible, pour la raison suivante : ce que l'information sur le monde social modifie, ce n'est pas la loi en elle-même mais ses conditions particulières d'instanciation (ce qu'on appelle parfois les "conditions initiales").

Bien sûr, cela implique certains phénomènes de non-linéarité qui mènent à des modèles plus compliqués, et à des pouvoirs prédictifs plus faibles – mais rien qui ne soit particulier ou spécifique aux phénomènes sociaux. Au contraire, plus on s'efforcera de développer de telles lois fondamentales, plus on peut s'attendre à ce que ces phénomènes soient mieux compris, et donc moins ils poseront de problèmes sur le plan méthodologique.

Imaginons, maintenant, des personnes qui voudraient donner tort aux sociologues ou aux psychologues en n'agissant pas, volontairement, suivant leurs théories, pourrait-on dire que leurs théories sont réfutées, et par là même que la recherche de lois est chimérique ? Il ne me semble pas. S'il s'agit de bonnes lois, alors il est peu probable que des personnes puissent fondamentalement les violer. En effet, il ne suffit pas de vouloir donner tort aux sociologues pour obtenir un haut diplôme universitaire en venant d'un milieu populaire. Il ne suffit pas non plus de vouloir donner tort aux psychologues cognitivistes et à leurs théories de la cécité cognitive pour remarquer qu'il y a un "pas" de trop au début de cette phrase. Les phénomènes sociaux et psychologiques, s'il s'agit bien de phénomènes, agissent qu'on le veuille ou non, et les connaître ne les fait pas disparaître mais tient leur influence sur nous seulement un peu plus éloignée.

De plus, le fait même de vouloir donner tort aux sociologues peut très bien être interprété à l'aide de la loi même que la personne tente de violer. Par exemple, le comportement d'un individu qui ferait exprès de ne pas faire les choix les plus "rationnels" pour donner tort aux théoriciens du choix rationnel peut très bien être décrit par cette même théorie du choix rationnel, en intégrant dans l'utilité espérée à maximiser un terme correspondant au fait de donner tort à la théorie.

Encore une fois, s'il s'agit d'une vraie loi, alors tout ce que l'on peut faire c'est modifier des conditions initiales, éventuellement, mais probablement pas la loi en elle-même. Bien entendu, il s'agit alors de savoir dans quelle mesure il s'agit d'une loi féconde et non tautologique : il s'agit du cœur de ma discussion à propos du statut épistémologique des principes fondamentaux en science, que je laisse donc pour un chapitre ultérieur.

On peut également, à la suite de Testart ou Cuin, dénoncer le "mythe du rapport épistémique" (CUIN, 2000), c'est-à-dire l'idée que l'humain aurait un accès privilégié à la réalité sociale, là où le physicien a besoin de construire des modèles pour obtenir de l'information sur la réalité physique.

Il y a plusieurs aspects ici. D'une part, on l'a vu, il n'y a aucune raison de penser que le sociologue ou l'anthropologue ait un accès direct aux causes

des phénomènes qu'il étudie – si c'est cela qui l'intéresse. Au contraire, on aura le plus souvent accès aux rationalisations qu'en font les acteurs et pas forcément aux causes directes. Pour trouver celles-ci, précisément parce que le monde social et politique est complexe et diversifié, il nous faut appliquer des outils statistiques, comme le montre l'exemple du lien entre interventions militaires et terrorisme.

D'autre part, cette illusion, comme le note Alain Testart, est probablement dû au fait que les cadres conceptuels de la physique, ses variables de base, sont éprouvés et ont été rendus robustes depuis plus longtemps, par un processus qui est tout sauf donné. Il montre notamment que le sujet n'occupe une place à part en sciences sociales que suite à ce même type de contingence historique, et non pas du fait d'une nature particulière des phénomènes sociaux.

On ne peut pas observer le monde de l'extérieur du monde, mais c'est justement pour cette raison qu'on peut faire de la science :

J'entends bien que l'homme est social et n'y vois toutefois pas une source de difficultés mais bien la première condition épistémologique qui doit rendre un jour possibles les sciences sociales. Car l'homme est physique aussi, ce qui n'a pas empêché l'essor des sciences physiques. C'est bien au contraire cette commune mesure de l'homme et du monde dans lequel il se situe, cette commune nature entre un sujet et un objet qui doivent être dits également « physiques », qui permet une science comme celle de la physique. Ce que nous avons déjà dit et qui nous permet de conclure : *Si l'homme peut faire un jour la théorie du social c'est parce qu'il est social, exactement pour la même raison qu'il a pu faire la théorie de la physique parce qu'il est physique.* (TESTART, 1991, p. 58)

3.5 Conclusion

Les positions pluralistes reposent ainsi le plus souvent sur une accentuation des différences existantes entre les sciences sociales et les autres, accentuation souvent due à une oblitération de l'histoire de la physique qui pousse à lui donner des propriétés qu'elle n'a pas et n'a jamais eues, doublé de l'illusion d'une connexion épistémologique particulière entre le sujet et l'objet sociologiques.

Les productions épistémologiques défendant l'impossibilité, pour les sciences sociales, d'être des sciences comme les autres apparaissent alors comme

des tentatives de justification d'un certain état de fait, à savoir une certaine pluralité conceptuelle. En d'autres termes, au lieu de percevoir, comme c'est mon cas, cette situation comme problématique, on s'attache au contraire à la justifier.

Ce type de situation, où une définition différente de la science est développée et défendue pour répondre à certaines attaques, est-il propre aux sciences sociales? Pour répondre à cette question, il est intéressant de noter que certains physiciens critiques de la *théorie des cordes* ont fait le même type de constat concernant la réaction des tenants de la théorie face à ses détracteurs. En particulier, dans (SMOLIN, 2007), Lee Smolin, ancien théoricien des cordes, aborde son histoire, son épistémologie et sa sociologie avec force détails, et nous livre une critique fouillée de ce qu'il peine lui-même à nommer une "théorie".

En effet, la théorie des cordes, après plus de vingt-cinq ans d'existence (dans les années 2000) ressemble davantage à un ensemble disparate de *familles* de "résultats approximatifs" et de "conjectures" (SMOLIN, 2007, p. 244), qui forment un *paysage* à la structure hautement complexe et opaque. En dépit de la différence évidente avec la forme que prennent les théories physiques habituelles après un certain temps de maturation, les tenants de la théorie des cordes espèrent depuis les années 1980 qu'elles puissent un jour révéler une théorie plus profonde qui unifierait les quatre interactions physiques fondamentales actuellement connues, ainsi que la gravitation avec la physique quantique.

Il est ainsi intéressant de noter comment les tenants de la théorie des cordes réagissent habituellement, selon Lee Smolin qui les a beaucoup côtoyés, quand certaines de ses apories sont pointées du doigt :

Si la discussion [avec des théoriciens des cordes] en vient au fait que la théorie des cordes prédit le paysage et, par conséquent, ne fait pas de prédictions concrètes, alors quelques théoriciens des cordes nous chanteront qu'*il faut modifier le sens de ce qu'est la science*. (SMOLIN, 2007, p. 351) (C'est moi qui souligne.)

Le but de cette comparaison est de suggérer que l'idée de la nécessité d'une épistémologie particulière, c'est-à-dire d'un autre "sens de ce qu'est la science", n'a aucune raison d'être propre aux sciences *sociales*. Cette idée émerge au contraire à chaque fois que les tenants d'une quelconque production cognitive se trouvent attaqués dans leur légitimité scientifique.

On peut toutefois comprendre cette pluralité autrement : comme des *but*s différents donnés à la production sociologique. L'existence de ces buts éventuellement distincts a l'air d'être relativement consensuel.

Boudon, en particulier (BOUDON, 2002), distingue "quatre types idéaux majeurs et permanents de productions sociologiques", à savoir :

1. Le genre *expressif* ou *esthétique* : l'objectif ici est de rendre compte dans une forme hautement *littéraire* de certaines vicissitudes communes de la vie sociale. Un tel genre est très proche de la littérature, à travers une forme que l'on pourrait qualifier d'"essayisme social".
2. Le genre *caméraliste* ou *descriptif* : il s'agit ici de *renseigner* un public donné (par exemple, des décideurs politiques) à propos de la réalité sociale vécue par certains acteurs particuliers. Il ne s'agit alors pas d'expliquer les phénomènes sociaux, mais davantage d'offrir à ce public une description d'un pan de la vie sociale qui ne lui est pas directement accessible.
3. Le genre *critique* ou *engagé* : de telles productions ne font pas que décrire une situation sociale, mais se donnent une mission *critique* de dévoiler certains aspects du fonctionnement de la société considérés comme intolérables, ou à minima problématiques.
- 4 Le genre *cognitif*, ou *scientifique* : le quatrième type de production sociologique correspond à la production de *connaissances scientifiques* sur le monde social, au même titre que ce qui est fait dans les autres disciplines. En d'autres termes, il s'agit de mettre en évidence des phénomènes sociaux et d'en proposer des explications au sein de cadres théoriques bien identifiés.

Charle-Henri Cuin fait également part d'une pluralité d'objectifs similaire :

[...] certains conçoivent la sociologie comme un des métiers de l'information, d'autres comme une science quasi naturelle, d'autres comme un art de décryptage d'un sens caché, d'autres comme une vocation d'élaboration de cet indispensable sens, d'autres comme un devoir de "dévoilement" et de dénonciation, d'autres encore comme une lutte pour la défense de valeurs plus ou moins transcendantes. (CUIN, 2000, p. 13)

On peut trouver le même genre de distinctions à la base de l'ouvrage collectif dirigé par Bernard Lahire (LAHIRE, 2004) :

si la sociologie doit avoir une quelconque utilité, quelle doit en être la nature : politique (sociologue-expert, sociologue-conseiller

du prince, sociologue au service des luttes des dominés), thérapeutique (la sociologie comme socio-analyse et moyen de diminuer ses souffrances grâce à la compréhension du monde social), cognitive (la sociologie comme savoir n'ayant d'autre objectif que d'être le plus scientifique possible)?

Cette mise en évidence d'objectifs distincts pour les productions sociologiques peut éclairer la situation de pluralité sur laquelle j'ai insisté dans ce chapitre. En effet, il ne s'agit plus ici de savoir si les sciences sociales doivent jouir, en général, d'une épistémologie à part, mais bien si elles peuvent se développer davantage dans leur dimension cognitive.

Les autres formes de sciences sociales devront alors éventuellement se justifier autrement, mais étant donné qu'elles n'ont pas les mêmes objectifs, une telle distinction ne posera plus les mêmes problèmes. Cette pluralité sera alors tout à fait similaire à celle qui existe dans les autres disciplines, par exemple entre les modes de justification d'une théorie physique et ses applications concrètes possibles. Cette différence de sens rejoindrait alors celle qui existe, et qui est *non problématique*, entre la scientificité d'une mission spatiale et la scientificité d'une théorie physique fondamentale, pour reprendre l'exemple donné dans l'introduction de ma thèse.

Dans la suite de mon travail, fort de ces diverses réflexions, j'adopte une position moniste, ou *unitariste*. Plus précisément, je considère que, si tant est que l'objectif poursuivi soit la production de connaissances générales sur le monde social, c'est-à-dire la constitution d'une *science fondamentale*, il n'y a aucune raison de principe pour que les sciences sociales nécessitent de s'interpréter au sein d'un cadre épistémologique différent des autres sciences.

Cependant, il reste à *identifier* plus précisément ce cadre épistémologique unitaire. En effet, comme on l'a vu au chapitre 2, des questions épistémologiques importantes se posent aux disciplines scientifiques en général, pas uniquement aux sciences sociales. C'est à une clarification de ces questions, en lien serré avec les problèmes soulevés dans ce chapitre, qu'est consacrée la suite de cette thèse.

Chapitre 4

Degré de corroboration informationnelle

4.1 Introduction

Le présent chapitre constitue une première étape de mon élaboration d'un modèle unitaire de la scientificité. Comme on l'a vu dans les chapitres précédents, et comme on l'a résumé dans la section 2.10, toute démarche cherchant à élaborer une conception un tant soit peu pertinente de la scientificité doit intégrer un ensemble de critiques traduisant un certain nombre de difficultés. Dans cette introduction je vais présenter, dans les grandes lignes, l'approche que j'ai choisie dans mon travail, et en particulier la façon avec laquelle je compte surmonter ces difficultés.

La démarche générale que j'embrasse considère la scientificité comme une propriété émergente d'un processus d'optimisation caractérisé par la maximisation, sous diverses contraintes, d'une certaine quantité.

J'ai déjà évoqué cette façon de percevoir la scientificité dans la dernière partie du chapitre 2 et comme on l'avait noté alors, toute la pertinence d'une telle démarche repose sur la définition que l'on donne à la quantité à maximiser. En effet, si l'on ne précise rien de plus, alors cette caractérisation est triviale, puisque à peu près n'importe quoi peut être vu comme la maximisation de quelque chose, si ce "quelque chose" est laissé sans plus de précision. C'est à cette définition qu'est consacré ce chapitre.

Un premier élément important de ma démarche est de me focaliser sur des unités d'analyse bien définies, à savoir les *modèles*.

Prenant à contre-pied la tendance à définir la scientificité à des échelles de plus en plus vastes, comme par exemple les champs épistémiques bungeins, je prends le parti de rester à l'échelle des objets cognitifs eux-mêmes. Dans la suite de mon travail, j'utiliserai le terme générique d'*unités épistémiques* pour qualifier ces unités d'analyse.

La principale difficulté d'un tel choix d'unités d'analyse est que cette démarche sous-entend qu'il est pertinent de les étudier comme si elles pouvaient être désassorties du contexte socio-historique de leur production. Or, comme on l'a vu, le qualificatif de "scientifique" pour désigner une hypothèse, un modèle ou une théorie dépend intrinsèquement d'éléments qui ne sont pas directement dépendants de ces unités cognitives : les données empiriques disponibles, les autres hypothèses, modèles ou théories en concurrence, ou encore les rapports de pouvoirs au sein des communautés scientifiques considérées, entre autres choses.

Cette difficulté peut toutefois être écartée, je pense, précisément en considérant la scientificité non pas comme une qualité intrinsèque d'une unité cognitive donnée mais comme la manifestation d'un processus d'optimisation. Or, cette optimisation dépend à la fois de l'unité cognitive en elle-même, mais également d'autres éléments, comme par exemple les données qui sont disponibles à un moment particulier, et ainsi de suite.

Plus précisément, je souhaite définir sur ces unités cognitives un certain *degré de corroboration informationnelle*. C'est la maximisation de cette quantité qui va alors caractériser le degré de scientificité d'une telle unité. La maximisation *effective*, c'est-à-dire le processus d'optimisation sous-jacent, est contrainte par des éléments contingents, comme les facteurs socio-historiques que j'ai déjà évoqués ; pour autant, la définition de cette quantité à maximiser, elle, n'en dépend pas.

On peut également voir cette démarche comme la définition d'un algorithme qui prend plusieurs unités épistémiques et des données empiriques en *input* et ressort l'unité épistémique la plus corroborée (dans un sens qu'il reste à définir) en *output* : le degré de scientificité d'une unité cognitive (l'*output*) dépend de contingences historiques et sociales : quelles données sont disponibles, ou bien quels modèles sont en compétition (l'*input*). Cependant, *les règles intrinsèques de cet algorithme*, elles, restent par définition anhistoriques.

Ma vision de l'épistémologie théorique est que sa tâche est précisément d'isoler et d'étudier ces règles, indépendamment ensuite de leurs applications concrètes – qui, elles, seront bien entendu hautement dépendantes du contexte socio-historique.

Un deuxième élément important de ma démarche est que je tente le plus possible d'explicitier les *méta-modèles* qui forment le substrat cognitif de mes réflexions.

Ici, j'entends par méta-modèle un modèle d'unité épistémique, c'est-à-dire une description de la structure de notre connaissance, à l'échelle à laquelle on souhaite l'étudier. Le postulat sous-jacent ici est que toute réflexion philosophique ou épistémologique, qu'on le veuille ou non, porte sur des *modèles de notre connaissance*, aussi complexes et détaillés soient-ils. Comme tout modèle utilisé dans un contexte scientifique, ils ne sont pas strictement vrais ou faux, mais ont plutôt une certaine gamme de validité.

Ainsi, le modèle "hypothèse \rightarrow prédiction \rightarrow comparaison avec des données" n'est pas *mauvais* ou *faux* en lui-même, il s'applique simplement à une gamme assez restreinte de situations : celles pour lesquelles on s'est mis d'accord sur une certaine *unproblematic background knowledge*. De la même façon, des unités plus complexes comme les programmes de recherche lakatosiens, les paradigmes kuhnien, les champs épistémiques bungeins ou la théorie meehilienne de la corroboration-similitude, bien qu'elles prennent en considération davantage d'éléments et reposent sur des structures un peu plus ramifiées, n'en restent pas moins des (méta-)modèles.

Mon postulat méthodologique est qu'il ne faut pas les voir comme une *affirmation* d'à quoi notre connaissance scientifique ressemble effectivement, mais plutôt comme une *proposition* de *schéma* de notre connaissance dont le niveau de détail à implémenter, comme pour tout modèle, dépend du but recherché dans son utilisation. En d'autres termes, il faut les voir ici comme les substrats cognitifs de nos réflexions épistémologiques.

Utiliser un tel méta-modèle n'implique donc pas de nier la complexité intrinsèque de la connaissance scientifique en tant que système logico-empirique. Il s'agit juste de faire la même chose que lorsqu'on utilise un modèle (scientifique) pour étudier une réalité intrinsèquement complexe : reposer sur une version simplifiée de cette dernière via des idéalizations et des abstractions pour rendre son étude plus abordable. En d'autres termes : se focaliser sur certains aspects de ce réseau complexe, en y définissant une zone d'analyse bien circonscrite, et en laissant de côté d'autres aspects considérés comme non pertinents pour le propos en question.

Dans ce chapitre, je vais restreindre mes réflexions à la partie de l'unité épistémique "la plus proche" des données empiriques, à savoir les *modèles empiriques* (j'aborderai les modèles théoriques dans le chapitre 5). La section 4.2 est dédiée à leur définition formelle.

Je vais ensuite m'attacher, dans la section 4.3, à définir un degré de corroboration informationnelle sur ces unités d'analyse particulières. L'idée sous-jacente principale développée ici est hautement inspirée du falsificationnisme :

ce qui compte, relativement à la scientificité, ce n'est pas autant que nos constructions théoriques soient validées par l'expérience, mais à *quel point elles auraient pu ne pas l'être*. En d'autres termes, la bonne question à se poser est : quelle *information* gagne-t-on à savoir que tel modèle est à ce point empiriquement adéquat ? C'est cette quantité d'information que j'appelle le degré de corroboration informationnelle.

Enfin, dans la section 4.4, je discuterai de certaines propriétés de ce degré de corroboration informationnelle. Je montrerai notamment que de l'exigence de maximisation de cette quantité découlent naturellement plusieurs critères épistémologiques bien connus, qui apparaissent alors comme des cas particuliers.

4.2 L'unité d'analyse : les modèles empiriques

Une première étape, comme je l'ai annoncé, consiste donc à décrire précisément l'unité d'analyse que je prends pour base de ma construction dans ce chapitre : les modèles empiriques. Si l'on voit, à la suite de Mill ou de Quine, l'ensemble de notre connaissance comme un grand tissu reposant en dernière instance ("sur les bords") sur les données empiriques, les modèles empiriques que l'on va définir ici peuvent être vus comme les parties de ce tissu les plus proches des données.

Dans une première partie, je vais définir une telle unité d'analyse qualitativement, sur un certain nombre d'exemples. Ensuite, j'en proposerai une définition plus formelle dans un langage mathématique inspiré de la théorie des probabilités.

4.2.1 Définition qualitative

Les modèles empiriques auxquels on s'intéresse ici peuvent être possiblement de deux types : les modèles nomologiques et les modèles causaux. Je vais introduire ces deux types de modèles à l'aide d'exemples concrets.

Les modèles nomologiques

En physique classique, un exemple de modèle nomologique est la loi des gaz parfaits :

$$PV = nRT, \tag{4.1}$$

qui s'applique à un gaz, où P désigne sa pression, V son volume, n sa quantité de matière, T sa température, et R est la constante des gaz parfaits, qui vaut $8,314 \text{ J mole}^{-1}\text{K}^{-1}$.

Les variables (P, V, n, T) données avec leurs définitions *opérationnelles* (les opérations concrètes qu'il faut faire pour les mesurer sur un gaz réel particulier) forment le *cadre empirique*.

Ce cadre empirique peut être vu comme définissant un certain *espace* mathématique E . Ce que l'on sous-entend en disant cela est que le cadre empirique a une certaine structure mathématique plus riche qu'un simple ensemble.

Le modèle empirique (4.1), dans ce cas, se traduit comme une *relation* entre ces variables, relation qui définit un sous-espace de l'espace E .

Pour le même type de systèmes physiques, à savoir des gaz, observés au sein du même cadre empirique, un autre modèle empirique possible est l'équation dite de van der Waals :

$$\left(P + \frac{an^2}{V^2}\right)(V - nb) = nRT, \quad (4.2)$$

où a et b sont des paramètres supplémentaires de l'équation et dépendent du gaz en question. On peut remarquer que (4.1) est un cas particulier de (4.2) pour $b = 0$ et $n/V \rightarrow 0$.

Un autre exemple pourrait être, toujours en physique, celui de la chute avec frottement. L'expérience typique associée est celle de la chute d'une bille sphérique dans un fluide (par exemple, dans l'air) à partir d'une certaine hauteur h_0 . Les variables de base que l'on peut mesurer, et qui forment ainsi le cadre empirique, sont la masse m , la vitesse v et la durée t (avec $t = 0$ au moment où on lâche la bille).

Un modèle empirique pour cette situation est le suivant :

$$|v(t)| = v_\infty(1 - e^{-t/\tau}), \quad (4.3)$$

où v_∞ et τ sont a priori des paramètres libres, et la masse m n'apparaît pas explicitement. Un autre modèle possible est le suivant :

$$|v(t)| = v_\infty(1 - e^{-t/\tau}), \quad v_\infty = g\tau, \quad \tau = \frac{m}{\alpha}, \quad g = 10 \text{ m.s}^{-2}. \quad (4.4)$$

Dans celui-ci, la masse apparaît explicitement et le seul paramètre libre restant est α . (4.4) est un cas particulier de (4.3) pour lequel on suppose que v_∞ et τ sont linéairement reliés.

Dans l'exemple déjà cité au chapitre 3 à propos des patterns universels pouvant être détectés lors d'élections proportionnelles, la relation :

$$P(v, Q, N) = F(vQ/N) = \frac{N}{\sqrt{2\pi\sigma vQ}} e^{-(\log(vQ/N) - \mu)^2 / 2\sigma^2} \quad (4.5)$$

constitue un modèle empirique nomologique. Elle exprime une relation entre les variables v , le nombre de vote reçus par un candidat, Q le nombre de candidats dans la liste du candidat en question, N le nombre total de votes reçu par la liste du candidat, et donne la proportion de candidats pour chaque valeur de vQ/N .

De tels modèles, par définition, peuvent être comparés à des *données*. Ils expriment en effet des relations entre des variables qui ont un sens empirique précis, puisqu'elles sont supposées être fournies avec leur définition opérationnelle.

Dans l'exemple du gaz, les données empiriques prennent la forme d'un ensemble de points dans l'espace E défini par ces variables. Par exemple, les données typiques pour une expérience de chute avec frottement prennent la forme donnée en figure 4.1.

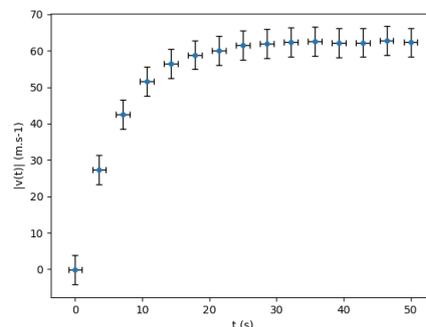


FIGURE 4.1 – Des points de données typiques d'une chute avec frottement.

Dans le cas du pattern d'universalité dans les élections proportionnelles, comme on l'a déjà vu, les données prennent la forme donnée en figure 4.2.

Cet exemple va nous permettre de préciser ce qui nous intéresse ici : les points de données en couleur, ainsi que le modèle empirique représenté par la courbe lognormal de fit, mais *pas* les points noirs issus des prédictions du modèle théorique, dont on ne s'occupe pas ici. En effet, on se focalise dans ce chapitre sur les modèles empiriques sans prendre en compte d'éventuel modèles théoriques explicatifs.

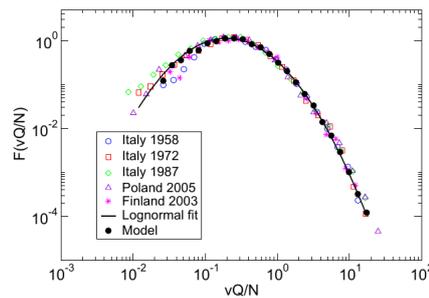


FIGURE 4.2 – Figure 2 de (FORTUNATO et CASTELLANO, 2007, p. 3) comparant des données, l'ajustement par la loi log-normale ainsi que les résultats issus du modèle ($r = 0.25, k_{min} = 10, \alpha = 2.45$).

Les modèles causaux

Un modèle empirique de type causal traduit simplement l'hypothèse d'une relation de causalité entre différentes variables sur une certaine population statistique.

Dans ce cas, classiquement, les variables peuvent être de quatre types différents : quantitatives et continues (leur valeur est un nombre réel, comme la taille ou le QI), quantitatives et discrètes (leur valeur est un nombre entier naturel, comme le nombre d'habitants dans une ville), catégorielles et ordinales (leur valeur est une catégorie ordonnée, comme par exemple le niveau de douleur sur une échelle verbale simple à quatre niveaux), ou bien catégorielles et nominales (leur valeur est une catégorie non ordonnée, comme le genre, ou un ensemble de villes).

Rappelons que je ne parle pas ici des mécanismes théoriques qui pourraient éventuellement expliquer la relation de causalité : je me focalise sur le niveau statistique (empirique et non théorique) uniquement. Ce niveau correspond par exemple à la relation macro-macro du diagramme de Coleman (COLEMAN, 1990), ou encore aux "associations statistiques observés" en épidémiologie, en termes de critères de Bradford Hill (BRADFORD HILL, 1965).

La causalité est un concept abondamment discuté en philosophie des sciences, et je prends ici la définition opératoire suivante, relativement consensuelle : deux variables A et B sont reliées par une relation de causalité si on peut mesurer une association statistique élevée entre elles, et si cette association statistique reste relativement élevée *ceteris paribus*, c'est-à-dire toute chose étant égale par ailleurs.

La forme de l'association statistique dépend de la nature des variables – par exemple, on peut prendre le coefficient de corrélation de Pearson ou son

carré, le coefficient de détermination, comme une bonne mesure de l'association statistique entre des variables quantitatives et continues.¹

Le cadre empirique, dans ce cas, est constitué de toutes les variables pertinentes à mesurer sur le phénomène en question accompagnées de leurs diverses définitions opérationnelles. Des données empiriques vont alors pouvoir être produites en appliquant ce cadre sur *un échantillon* d'une population.

Le modèle empirique, c'est-à-dire l'énonciation d'une relation causale, porte, quant à lui, sur la population dans son ensemble, et repose sur des données produites sur un ou plusieurs échantillons. Un modèle empirique peut donc être également représenté comme un sous-espace (à la topologie éventuellement un peu plus compliquée) d'un espace E défini par les différentes variables choisies comme pertinentes pour la situation en question.

Un exemple parmi ceux déjà cités dans les chapitres précédents pourrait être le lien causal existant entre l'interventionnisme militaire d'un État et les actes terroristes commis sur son territoire. Comme le souligne les auteurs :

Toutes choses égales par ailleurs (PIB, population, régime politique, religion des auteurs de l'attaque, religion majoritaire du pays, etc.), la probabilité pour un État de voir ses troupes ou ses ressortissants visés par un attentat-suicide est 17 fois plus grande pour les pays "occupants" que pour les pays "non-occupants".

(DELORI et al., 2021, p. 51)

Ici, le cadre empirique est donc constitué, entre autres, des variables suivantes : degrés d'interventionnisme, nombre d'attentats recensés sur le territoire, PIB, population, régime politique, religion des auteurs de l'attaque, religion majoritaire du pays, et ainsi de suite – ces variables étant éventuellement de natures différentes.

L'énoncé d'une relation de causalité repose sur l'observation d'une forte association statistique entre les deux premières variables, une fois contrôlées toutes les autres – c'est-à-dire sur chaque sous-ensemble des échantillons pour lesquels toutes ces variables sont les mêmes.

Encore une fois, on ne parle pas des explications éventuelles, c'est-à-dire des modèles théoriques qui pourraient être mobilisés pour rendre compte de ces associations statistiques, en restant pour le moment au niveau empirique, en l'occurrence statistique.

1. Un panorama général de ces différentes façons d'évaluer la prééminence d'une association statistique, aussi appelée "taille d'effet" dans la littérature consacrée, est donnée par exemple dans cet article : (SULLIVAN et FEINN, 2012, p. 280), comme dans n'importe quel manuel d'introduction aux statistiques.

Un autre exemple, dont on a déjà parlé également, est la relation existante entre le milieu social d'origine d'une personne et sa réussite scolaire ou académique. Comme on l'a décrit plus en détail en section 3.2.2, plusieurs auteurs dont Raymond Boudon ont mis en évidence une association statistique forte entre le milieu social d'origine d'un élève (ouvriers, cadres moyens, cadres supérieurs, industriels et professions libérales, ...) et le choix d'orientation à la fin de la primaire entre trois possibilités (lycée, collège d'enseignement général ou classe de fin d'études).

Il observe alors que plus l'origine sociale est élevée, plus l'élève a tendance à rester dans le système scolaire en choisissant une voie longue, et ce à *résultat scolaire égal* à la fin du primaire. Tout ceci le mène à énoncer une relation de causalité entre l'origine sociale des élèves et leurs choix d'orientation (et donc in fine sur la réussite scolaire), relation portant sur la population en général et donc constituant une hypothèse inductive authentique, c'est-à-dire un modèle empirique (de type causal) tel qu'on les a définis.

Cette relation de causalité, comme dans les cas précédemment évoqués, se manifeste donc à travers des associations statistiques qui persistent même une fois que les principaux facteurs de confusion sont contrôlés.

4.2.2 Définition formalisée

De manière générale, ce que j'appelle ici modèle empirique est simplement *l'énoncé d'une certaine relation entre un ensemble fini de variables, associées à une certaine situation empirique, de manière à ce qu'elle puisse être comparée à des données empiriques, données elles-mêmes produites en appliquant ces différentes variables à la situation empirique donnée.*

Les différents composants d'une unité épistémique élémentaire

L'unité épistémique élémentaire sur laquelle porte ma réflexion dans ce chapitre consiste donc en la donnée de plusieurs composants.

UNE SITUATION EMPIRIQUE ou un ensemble de situations empiriques bien définies, notée σ .

UN CADRE EMPIRIQUE noté E et constitué d'un ensemble fini de variables, données avec leur définition opérationnelle respective. J'entends par définition opérationnelle une façon concrète de les mesurer, tout en sachant que cette dernière n'est pas forcément unique. E peut être vu

comme un certain *espace*, c'est-à-dire un ensemble possédant une certaine structure donnée par la nature des différentes variables qui le définissent.

UN FOND DE CONNAISSANCE à propos des variables définissant le cadre empirique : leurs définitions opérationnelles leur imposent des contraintes *a priori* (sur les valeurs qu'elles peuvent prendre, par exemple), ce que l'on traduit par la donnée d'une mesure de probabilité \mathbb{P} sur l'espace E .²

DES DONNÉES EMPIRIQUES produites en appliquant le cadre empirique E aux situations σ , notées D , pouvant être vues comme un ensemble de points de E .

UN MODÈLE EMPIRIQUE énonçant une certaine relation ou une famille de relations entre les variables définissant le cadre empirique, noté M . Un modèle empirique peut être vu comme un sous-espace (ou une famille de sous-espace) de E . On pourra le noter plus généralement $(M_\theta)_{\theta \in \Theta}$ où Θ désigne un espace de paramètres.

UN COMPARATEUR EMPIRIQUE permettant de comparer les données D au modèle empirique M , par exemple via *une distance* d définie sur E . Cette distance est ensuite élargie pour comparer non plus deux points, mais un point ou un ensemble de points (D) et un sous-espace (M). Cette distance va ensuite être utilisée pour définir un certain degré d'adéquation empirique entre le modèle M et les données D .

L'unité épistémique élémentaire, que l'on va noter ϵ , peut donc être décrite comme la donnée de ces différents éléments :

$$\epsilon = (\sigma, E, \mathbb{P}, D, (M_\theta)_{\theta \in \Theta}, d). \quad (4.6)$$

Interprétations des termes

La ou les situation(s) empirique(s) σ .

Il semble nécessaire de prendre en compte explicitement la situation empirique concrète sur laquelle repose l'investigation envisagée. Une telle situation empirique est définie en grande partie implicitement, par une sorte de

2. Ainsi, le cadre empirique est représenté par un certain espace probabilisé $(E, \mathcal{B}, \mathbb{P})$ où \mathcal{B} est une tribu sur E , lui-même vu comme l'espace cible d'une certaine variable aléatoire $X : \Omega \rightarrow E$ et la mesure de probabilité \mathbb{P} sur E comme la mesure associée à X à partir de la probabilité définie de base sur l'univers Ω . Par abus de langage, et puisque les autres éléments ne nous sont d'aucune utilité dans notre réflexion, on identifiera simplement (E, \mathbb{P}) à l'espace probabilisé représentant le cadre empirique.

conceptualisation a priori dont sont munis les individus engagés dans l'investigation en question. Comme je souhaite me concentrer sur des éléments purement cognitifs, je me garde bien ici de décrire davantage cette conceptualisation, et j'en donne donc une description minimale, quasiment uniquement symbolique, sous la forme d'un élément, ou d'un ensemble d'éléments, sans plus de précision.

Le cadre empirique E et la mesure de probabilité associée \mathbb{P} .

Comme on l'a déjà énoncé, l'espace probabilisé (E, \mathbb{P}) représente le cadre empirique. La mesure \mathbb{P} encode en quelque sorte ce que l'on sait par ailleurs sur les variables définissant l'espace E , à la fois sur les façons concrètes de les mesurer et sur les contraintes qui en découle (les valeurs qu'elles ne peuvent probablement pas prendre, par exemple).

Bien entendu, on rend explicite cette mesure pour représenter le fait que cet élément, pour un même espace E , peut très bien être modifié lors d'une évolution théorique.

Rendre explicite le cadre empirique permet de représenter concrètement le concept de charge théorique de l'observation et d'intégrer dès le début de la réflexion le fait que les hypothèses sous-jacentes à la production de données sont elles aussi susceptibles de changer, et qu'aucune n'est a priori immunisée définitivement contre la réfutation.

Comme on l'a déjà mentionné plusieurs fois et notamment au chapitre 3, la recherche de bonnes variables de base est une tâche aussi importante, dans l'histoire d'une discipline scientifique, que la recherche de régularités empiriques pouvant être détectées à l'aide de ces variables. En d'autres termes, et comme je l'ai déjà exprimé dans le chapitre précédent : le langage avec lequel on parle du monde *et* ce que l'on dit du monde *co-évoluent* en réalité au cours de la progression scientifique.

Les modèles empiriques.

Voir les modèles empiriques comme des (familles de) sous-espaces du cadre empirique E rappelle bien entendu l'approche dite "espace de phase" développée notamment par Frederick Suppe (SUPPE, 1974) et Bas van Fraassen (FRAASSEN, 1980) dans le contexte de la conception sémantique des théories. J'appellerai parfois l'espace E "espace des phases"³ par abus de langage.

3. En physique, l'espace des phases d'un système consiste en l'ensemble des états dans lequel un système peut se trouver, et possède une certaine structure.

Dans notre cas, cependant, on ne parle pas de l'espace des phases d'une théorie dans son ensemble, mais bien de celui associé à un modèle empirique particulier.

Par exemple, la loi des gaz parfaits déjà citée plus haut est un exemple de modèle empirique. Le seul paramètre a priori libre est $\theta = R \in \Theta = \mathbb{R}$, si on se place dans le cas général pour lequel cette valeur serait inconnue. Pour chaque $R \in \mathbb{R}$, l'équation (4.1) définit un sous-espace topologique de dimension 3 de l'espace de dimension 4 défini par les variables (P, V, n, T) .

Le cas de la chute avec frottement exhibe un autre exemple de famille paramétrée de modèles, où $\theta = (v_\infty, \tau) \in \Theta = \mathbb{R}^{+2}$. De la même façon, l'équation (4.3) définit une courbe (un sous-espace de dimension 1) dans $E = \mathbb{R}^2$, ce dernier étant défini à partir des variables t, v .

Le cas des modèles empiriques causaux est similaire. Considérons trois variables A, B et C telles que A et B sont quantitatives et continues, et C catégorielle et nominale, à trois modalités C_1, C_2 et C_3 . L'hypothèse d'une relation de causalité entre A et B , par exemple $A \longrightarrow B$, implique l'observation d'une corrélation (par exemple, affine) entre A et B de la forme de la forme $B = aA + b$, sur des échantillons pour lesquels la valeur de C est *fixée* (à C_1, C_2 ou C_3). Dans ce cas, le cadre empirique prend la forme $E = E_1 \times E_2 \times E_3$, où $E_i = \mathbb{R}^2$ correspond à $C = C_i$ pour $i = 1, 2, 3$. Une relation causale entre A et B est donc caractérisée par un ensemble $\theta = \{(a_i, b_i)\}_{i=1,2,3}$. Pour chaque θ , le modèle causal est représenté par un sous-espace de E , pouvant être vu comme l'union pour $i=1, 2$ et 3 de droites dans E_i définies par $(a_i, b_i) \in \mathbb{R}^2$. Bien entendu, les cas concrets de corrélations sont plus complexes en général, mais l'idée principale reste la même.

On fait ainsi l'hypothèse de (méta-)modélisation que tout modèle empirique, aussi compliqué soit-il, peut toujours être vu comme une famille de sous-espaces de E .

Les données empiriques et le degré d'adéquation.

On suppose que les données empiriques, aussi diverses puissent-elles être, peuvent toujours être vues comme un ensemble fini d'éléments de E . C'est-à-dire, qu'il existe $n \in \mathbb{N}^*$ tel que $D \in E^n$.

Dans le cas d'un modèle nomologique exprimant une relation entre deux variables quantitatives et continues x et y , E peut être identifié à \mathbb{R}^2 et les données empiriques peuvent alors être représentées comme un élément de $\{(x_i, y_i)\}_{i=1..n} \in E^n$. De même pour un modèle causal, qui suppose une relation causale entre deux variables A et B définissant le cadre empirique E .

Dans ce cas, les données prennent la forme d'un ensemble de valeurs de A et B sur un certain échantillon de taille $n \in \mathbb{N}$, c'est-à-dire qu'elles peuvent encore être vues comme un élément de E^n .

On suppose également que l'espace E est métrique, c'est-à-dire muni d'une distance d . On élargit naturellement cette notion de distance afin de pouvoir comparer un ensemble de points de E et un sous-espace de E .

En d'autres termes, on se donne une façon d'évaluer à quel point le modèle et les données empiriques sont "éloignés", et sa forme concrète dépend de beaucoup d'aspects, en particulier de la nature exacte des données. Cette distance est ensuite utilisée pour définir formellement *un degré d'adéquation empirique*, comme on le verra plus loin.

4.3 Degré de corroboration informationnelle

4.3.1 L'héritage poppérien

L'idée fondamentale du falsificationnisme poppérien et qui constitue le squelette logique et principal de notre argumentation est que l'adéquation empirique d'une unité épistémique n'est pas une condition suffisante pour lui garantir un degré de scientificité élevé.

De ce point de vue, la scientificité n'est pas mesurée à la lumière de son degré d'adéquation empirique mais aussi et surtout à la lumière de son *degré de falsifiabilité*. Comme on l'a déjà énoncé, la question pertinente n'est pas seulement "à quel point ce modèle est adéquat empiriquement?" mais "à quel point aurait-il pu ne pas l'être?". Selon cet aspect précis du falsificationnisme poppérien, "Theories are not verifiable, but they can be 'corroborated'" (POPPER, 1959, p. 248). Les données empiriques ne peuvent jamais justifier définitivement un modèle, mais celui-ci peut être tout du moins *corroboré* : pour cela, il doit *avoir raison* alors qu'il aurait pu *avoir tort*.

Pour percevoir la pertinence toujours actuelle de ce point de vue, je pars du constat que cadre empirique donné va exhiber des *régularités endogènes*. Il s'agit de relations exprimées entre des variables qui sont vraies de par la définition même de ces variables, et donc seront (par définition) effectivement vérifiées empiriquement.

On peut, dans le langage de Popper, appeler ces énoncés des vérités tautologiques ou encore irréfutables. Ce n'est pas un mal en soi, c'est simplement *un fait*. La leçon principale du falsificationnisme est simplement que ce genre de régularités n'est d'aucun support épistémologique.

Une autre façon de comprendre cela est de dire qu'un modèle scientifique prétend dire quelque chose à propos du monde. Ainsi, la vérité de ses énoncés doit dépendre, au moins en partie, du monde – c'est-à-dire des données empiriques. Si un énoncé particulier est empiriquement adéquat à un certain degré, plus cet énoncé aurait pu être réfuté, *plus on gagne de l'information* sur la pertinence du modèle sous-jacent à dire quelque chose de substantiel sur le monde, c'est-à-dire quelque chose qui ne dépende pas uniquement du langage dans lequel on l'exprime.

C'est ce que Popper exprime à plusieurs endroits de son œuvre. Par exemple, dans (POPPER, 1959), il définit les théories comme des ensemble d'énoncés, et définit leur falsifiabilité à partir des "classes de leur falsificateurs potentiels". Plus précisément :

[If] we represent the class of all possible basic statements by a circular area, and the possible events by the radii of the circle, then we can say : At least one radius—or perhaps better, one narrow sector whose width may represent the fact that the event is to be 'observable'—must be incompatible with the theory and ruled out by it. One might then represent the potential falsifiers of various theories by sectors of various widths. And according to the greater and lesser width of the sectors ruled out by them, theories might then be said to have more, or fewer, potential falsifiers. (The question whether this 'more' or 'fewer' could be made at all precise will be left open for the moment.) (POPPER, 1959, p. 95-96)

Un peu plus loin, dans la section 32 "How are classes of potential falsifiers to be compared", il revient sur la dernière remarque de la citation ci-dessus, à savoir comment définir que certaines théories sont plus falsifiables que d'autres? Ou encore, de son point de vue : comment exprimer le fait que certaines théories ont plus de *falsificateurs* que d'autres, que leurs classes de falsificateurs est plus grande? Le problème principal étant, bien entendu, que ces classes sont la plupart du temps des ensembles infinis.

Son choix se porte finalement sur une relation d'inclusion au sens de la théorie des ensembles, entre différentes classes de falsificateurs. Deux théories T et T' peuvent alors être comparées à l'aide de leur classe respective de falsificateurs α et β . T est plus falsifiable que T' si et seulement si $\alpha \subset \beta$. Comme Popper le remarque, cette relation permet de formaliser "the intuitive 'more' and 'fewer', but it suffers from the disadvantage that this relation can only be used to compare the two classes if one includes the other." (POPPER, 1959, p. 98)

Il précise que l'on peut considérer les preuves *confirmant* un énoncé particulier "que dans les cas où elles procèdent de tests authentiques subis par la théorie en question; on peut donc définir celles-ci comme des tentatives sérieuses, quoique infructueuses, pour invalider (falsify) telle théorie (j'emploierai désormais pour les désigner le terme de "preuves corroborantes")." (POPPER, 1962, p. 64-65). Il s'agit donc bien de cette idée : la validation d'un énoncé empirique par des données ne vaut quelque chose que dans le cas où cette validation est en quelque sorte inattendue, voire même extraordinaire, dans le cas où la conjecture est "audacieuse".

Il formalise à nouveau l'idée d'un certain degré de corroboration à travers la notion de *vérisimilitude*. L'idée est la suivante : il définit le "contenu de vérité" d'un énoncé comme "la classe de tous les énoncés vrais qui [en] découlent (...) et qui ne sont pas tautologiques" (POPPER, 1972, p. 104), ainsi que son "contenu de fausseté" comme (moyennant quelques subtilités) la classe des énoncés faux qui en découlent, de telle façon à ce que ça soit bien un contenu et qu'elle ne contienne pas d'énoncés vrais.

Les énoncés tautologiques exclus par définition du contenu de vérité d'un énoncé correspondent aux régularités endogènes que nous avons décrit plus haut – leur nature est différente, mais l'idée sous-jacente est strictement la même. La *vérisimilitude* d'un énoncé, ou d'une théorie, est alors définie comme *le rapport de son contenu de vérité sur son contenu de fausseté*.

En d'autres termes, et c'est aussi la base de ma réflexion, il est évident que l'adéquation empirique est une vertu pour une théorie : Popper remet simplement en question l'idée qu'une telle adéquation suffirait en elle-même, en rajoutant des contraintes à respecter pour que cette adéquation soit informative.

La construction que je vais présenter dans ce chapitre n'est rien de plus que la mise à jour de ces idées en prenant comme substrat de base non plus les énoncés ou systèmes d'énoncés, mais les unités d'analyse basées sur les modèles empiriques que j'ai décrit plus haut, en section 4.2.

4.3.2 Esquisse de mon approche

Avant de rentrer plus en avant dans le cœur de la construction, je vais donner un aperçu des différentes étapes pour pouvoir garder une vue d'ensemble à l'esprit.

Le degré de corroboration comme une quantité d'information

Le point de départ de ma construction est l'idée, d'inspiration hautement poppérienne, que ce qui compte en ce qui concerne la scientificité, ce n'est non pas à quel point un modèle est confirmé empiriquement, mais *quelle information obtient-on à savoir qu'un modèle est à ce point confirmé empiriquement ?* C'est très précisément cette quantité que j'appelle *degré de corroboration informationnelle*. J'en donnerai une présentation formalisée dans une section ultérieure.

Ce degré de corroboration informationnelle peut être vu, qualitativement, de la façon suivante. Soit une unité épistémique donnée ϵ comme décrit plus haut. On a, en particulier, un espace E , des données empiriques D , un modèle empirique M et une manière de comparer D et M , donnant un degré d'adéquation que l'on note $\alpha(M, D) > 0$ (de telle façon, par convention, que $\alpha(M, D) = 0$ correspond à une adéquation parfaite et que plus $\alpha(M, D)$ est élevé, moins M et D sont adéquats).

On définit alors l'ensemble suivant : *l'ensemble de toutes les données qui sont au moins aussi adéquates que D avec M* . Cet ensemble, comparé à l'ensemble des données qui peuvent être a priori générées par le cadre empirique E , mesure à quel point il est surprenant que M et D soit à ce point adéquats. En d'autres termes, il quantifie l'information contenue dans une adéquation empirique de taille $\alpha(M, D)$.

Les régularités endogènes, c'est-à-dire les "vérités tautologiques", sont des modèles qui sont, par définition même des variables définissant E , parfaitement adéquats avec n'importe quel ensemble de données que l'on pourrait imaginer dans E . On n'apprend rien à savoir qu'ils sont à ce point adéquats, puisqu'on *le savait déjà*, en quelque sorte.

Au contraire, plus cet ensemble est "petit" à l'intérieur de l'ensemble des données possibles, plus un tel modèle est a priori falsifiable, c'est-à-dire plus on gagne de l'information à partir de son adéquation empirique.

La notion de taille relative de ces deux ensemble est opérationnalisée à l'aide des différents éléments dont on a doté notre unité épistémique, en particulier grâce à la mesure de probabilité \mathbb{P} définie sur E , comme on le verra plus bas.

Un principe de maximisation

Une fois défini un tel degré de corroboration, vu comme une certaine quantité d'information contenue dans un degré d'adéquation empirique entre

un modèle et des données, on peut alors comparer plusieurs unités épistémiques entre elles. Le degré de scientificité (empirique) d'un tel modèle correspond alors à sa place dans la concurrence qui l'oppose à d'autres modèles qui tous entendent reproduire certaines données.

Ce principe, bien qu'il soit ici un peu abstrait, permet de retrouver naturellement tout un ensemble de critères épistémologiques bien connus, comme on le verra aussi plus loin.

La charge théorique et la dépendance au langage

La charge théorique de l'expérience, ou encore la dépendance au langage de la distinction entre énoncés analytiques et synthétiques, ou entre énoncés réfutables et irréfutables, peuvent apparaître parfois comme des limitations intrinsèques d'une approche falsificationniste. L'approche que je développe ici entend dépasser cette limitation tout en conservant le bénéfice du falsificationnisme.

Premièrement, le langage empirique, représenté par le cadre empirique (E, \mathbb{P}) , est lui-même un élément pouvant varier lors du processus de maximisation du degré de corroboration. Le principe de maximisation sous-jacent prend donc dès le début cette caractéristique en compte, par construction.

Ensuite, les régularités empiriques authentiques (celles qui contiennent le plus d'information) et le langage dans lequel elles sont exprimées sont co-sélectionnées dans le processus de maximisation de la corroboration. Il est donc possible d'imaginer pouvoir sélectionner des langages qui sont, *ceteris paribus*, meilleurs que d'autres (pour certains buts précis). Ainsi, le fait de faire du cadre empirique une variable comme une autre n'implique aucun relativisme cognitif.

Enfin, on pourrait même aller jusqu'à affirmer que la charge théorique des observations est en fait *une raison de plus* de considérer la falsifiabilité comme un critère pertinent de scientificité. En effet, cette notion précise de falsifiabilité permet de distinguer les régularités empiriques authentiques (celles qui ont une chance de contenir une information pertinente sur le monde) des régularités qui sont endogènes au langage que l'on utilise.

En d'autres termes, *on force le monde à parler notre langage, mais on veut faire en sorte de ne pas parler à sa place*. La falsifiabilité nous en offre une certaine garantie.

4.3.3 Les régularités empiriques endogènes

Détecter des régularités au sein de données empiriques est une tâche fondamentale de toute investigation scientifique. Ces régularités dépendent bien sûr des variables mesurées et de la situation à l'œuvre. L'intuition fondamentale – celle que je cherche à formaliser – est cependant la même : une régularité est quelque chose d'intrigant, d'intéressant, qui appelle, en quelque sorte, une explication.

Comme mentionné précédemment, on est forcé d'exprimer ces régularités au sein d'un certain langage représenté ici par un cadre empirique E . Il s'avère que le fait même d'utiliser un tel cadre peut générer des régularités qui ne sont que de simples artefacts de ce cadre. Si le cadre empirique est vu comme une paire de lunettes que l'on porte pour observer le monde, les régularités endogènes sont des sortes de tâches, ou de reflets sur ces lunettes : on peut les confondre avec quelque chose d'intéressant présent dans le monde extérieur, mais elles ne sont que des artefacts produit par le fait même que l'on porte des lunettes.

Je vais présenter ici, à l'aide d'exemples concrets, deux types de telles régularités endogènes : celles qui sont dues aux définitions des variables composant E , que je vais appeler E -analytiques,⁴ et celles qui sont des artefacts de notre modélisation, c'est-à-dire qui ne sont pas dues aux variables de base ou à leur définition, mais aux hypothèses de modélisation à l'œuvre.

Les régularités E -analytiques

Voici deux exemples de relations empiriques observées qui sont directement dues aux définitions opérationnelles des termes qu'elles connectent.

Des énoncés analytiques.

Les cas les plus simples de telles régularités sont les énoncés analytiques. Un exemple un peu artificiel déjà cité est : "tous les cygnes sont des oiseaux". Cet énoncé a un sens empirique précis (dès l'instant où l'on s'est donné une manière précise de détecter lorsqu'un animal est un cygne et lorsqu'il est un oiseau), et il est vrai, justement par définition : le fait d'être un oiseau est une condition préalable au fait d'être un cygne. Pourtant, la confirmation empirique d'un tel énoncé ne nous apporte aucune information que nous n'avions déjà.

4. J'utilise cette notation pour rappeler que la distinction entre analytique et synthétique dépend, justement, du langage utilisé.

Un autre exemple pourrait porter sur le taux de pauvreté d'un pays. Si l'on définit le seuil de pauvreté comme un certain pourcentage du revenu médian, alors le taux de pauvreté d'un pays pour lequel il est possible de le mesurer ne pourra jamais être supérieur à 50%. C'est une régularité empirique, qui est parfaitement vraie, mais de même, cette adéquation empirique parfaite ne nous apprend rien sur le phénomène socio-politique sous-jacent.

La fonction de production de Cobb-Douglas.

Voici un autre exemple, correspondant à un cas réel. Un concept important des théories macro-économiques est celui de fonction de production agrégée, qui relie la quantité d'output économique (comme la production globale d'un pays) à certains inputs économiques (comme le travail, le capital, etc.)

En particulier, Charles Cobb et Paul Douglas ont proposé en 1928 une fonction (COBB et DOUGLAS, 1928) qui relie la production totale Q en un an à la quantité totale de travail L et de capital K (mesurés en unités homogènes sur lesquelles on ne s'attarde pas ici) :

$$Q = AL^\alpha K^\beta, \quad (4.7)$$

où A , α et β sont des paramètres. Cette fonction a été testée à l'aide de méthodes économétriques et une adéquation empirique très forte a été trouvée. En particulier, les paramètres de ce modèle empirique se trouvent satisfaire la relation : $\alpha + \beta = 1$, ce qui constitue apparemment un argument en faveur de certaines hypothèses de l'économie néo-classique.

Cependant, des critiques se sont élevés avec le temps, et il a été récemment prétendu (FELIPE et MCCOMBIE, 2013) qu'en réalité, cette relation se réduit à une *identité comptable*. En d'autres termes, la relation (4.7), avec $\alpha + \beta = 1$, peut se *déduire* mathématiquement d'une identité qui est vraie par définition – plus précisément, de la définition sous-jacente à la mesure des différentes variables en jeu.

C'est un exemple assez frappant d'une régularité qui semble à première vue nous apprendre quelque chose sur le monde extérieur, mais qui s'avère n'être rien d'autre qu'une tâche sur nos lunettes.

Les artefacts de modélisation

Je vais maintenant donner deux nouveaux cas pour lesquels la tautologie ne provient pas directement de la définition des variables de base mais du cadre utilisé pour modéliser les données.

La méthode de Lagrange.

Un exemple bien connu de ce phénomène, bien qu'il ne soit pas forcément perçu a priori dans sa dimension épistémologique, est le suivant : pour tout ensemble de n points de données $\{(x_i, y_i)\}_{i=1..n}$, il existe toujours une fonction polynomiale P_n de degré au plus $n - 1$ (donc, à n paramètres) qui s'ajuste parfaitement aux données. En d'autres termes, qui est telle que

$$\text{pour tout } i \in [1, n], y_i = P_n(x_i). \quad (4.8)$$

On peut difficilement imaginer un meilleur ajustement empirique.

Cependant, pour un ensemble donné de n points de données D , l'énoncé : "il existe n paramètres $\{a_i\}_{i=1..n}$ tel que la fonction polynomiale :

$$P_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} \quad (4.9)$$

s'ajuste parfaitement à D " n'est pas une prédiction empirique authentique, mettant au jour une régularité significative et contenant une information pertinente sur le phénomène en question. En effet, il est *toujours* possible de trouver ces paramètres. En d'autres termes, et même si l'énoncé ne semble pas simple dans sa forme, il est logiquement trivial, et donc empiriquement vide.

La procédure pour trouver le polynôme qui s'ajuste parfaitement à certaines données est appelée la méthode de Lagrange, et un exemple est donné figure 4.3.⁵

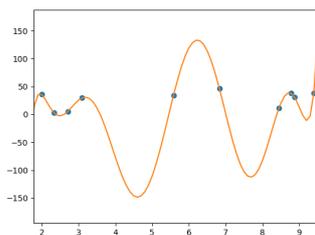


FIGURE 4.3 – La méthode de Lagrange appliquée à un ensemble de 10 points générés aléatoirement.

Plus précisément, dans le langage formalisé que nous avons donné plus haut, on a une famille de modèles $(M_\theta)_{\theta \in \Theta}$ où $\Theta = \mathbb{R}^n$, où $\theta \in \Theta$ est sous la forme : $\theta = (a_0, a_1, \dots, a_{n-1})$, et donc pour $\theta \in \Theta$, $M_\theta : (y(x) = P_n(x))$ avec P_n comme défini par (4.9).

5. Dans cet exemple, j'ai généré aléatoirement 10 points dans un graph à deux dimensions et calculé l'unique polynôme à 10 paramètres qui s'ajuste parfaitement à ceux-ci.

Pour tout $D = \{(x_i, y_i)\}_{i=1..n}$, l'énoncé " $\exists \theta \in \Theta$ tels que M_θ et D sont parfaitement adéquats" est vrai quel que soit l'ensemble de données $D \in \mathbb{R}^{2n}$ de départ. Aucune information n'est contenue dans une telle adéquation empirique, peu importe que cette adéquation soit parfaite ou non.

Les épicycloïdes de Ptolémée.

Un autre cas important, peut-être plus subtil, est celui du modèle géocentrique de Ptolémée, qui a dominé l'astronomie jusqu'à la fin du seizième siècle. Dans ce modèle, la Terre est au centre de l'Univers et les autres planètes ainsi que le soleil gravitent autour d'elle. Il se trouve que certaines de ces planètes ne suivent pas des orbites strictement circulaires, mais des trajectoires plus compliquées. Le modèle de Ptolémée rend compte de ces trajectoires sous la forme d'*épicycloïdes*.

Une épicycloïde est générée par la rotation d'un point le long d'un cercle dont le centre est lui-même en rotation le long d'un autre cercle (d'un rayon éventuellement différent), et ainsi de suite. Il se trouve que ce modèle possède un certain succès empirique, une certaine adéquation avec des données d'observations – dont une grande partie sont connues depuis longtemps. De plus, à chaque fois qu'une anomalie empirique apparaît, elle était accommodée en rajoutant un nouvel épicycle, c'est-à-dire un nouveau cercle dont le centre est en rotation autour du cercle précédent. A priori, ce modèle semble donc plutôt robuste, puisqu'il peut accommoder tout un ensemble d'anomalies sans modifier ses hypothèses de base.

Cependant, il se trouve que l'adéquation empirique de ce modèle est en réalité un artefact mathématique, un peu comme le précédent. Il se trouve que n'importe quelle trajectoire fermée, dans le repère terrestre, peut en réalité être ajustée par une épicycloïde possédant un nombre fini de cercles, à la condition que la précision observationnelle soit elle-même finie – ce qui est toujours le cas.

Il s'agit de la traduction d'un résultat mathématique bien connu, à savoir la décomposition de n'importe quelle fonction périodique (respectant certaines conditions) en série de Fourier. C'est son corolaire qui nous intéresse plus particulièrement ici : il est toujours possible d'approximer, à une précision donnée, une telle fonction par un nombre fini de termes de sa série

de Fourier.⁶ Or, une telle décomposition revient à une somme (pondérée) finie de fonctions cosinus et sinus. En deux dimensions, cette décomposition donne exactement des épicycloïdes finies.

Cela ne signifie pas que ce modèle n'a aucun intérêt. Il peut être et a été utilisé pour faire des prédictions précises. C'est simplement que la forme épicycloïdale en elle-même ne constitue pas quelque chose de profond à expliquer. En effet, elle se réduit à un simple artefact mathématique dû à la façon avec laquelle on modélise les trajectoires des planètes (comme des courbes fermées).

Ce n'est pas le cas de la forme elliptique des planètes dans le modèle héliocentrique de Kepler : il n'est pas vrai, dans ce cas, que n'importe quelle trajectoire, une fois fixée une certaine précision observationnelle, peut être ajustée par une ellipse. Le fait que ce modèle soit (dans une certaine mesure) adéquat empirique contient alors une information substantielle.

On retrouve l'idée, toujours poppérienne, qu'il vaut mieux un modèle non parfaitement adéquat ("faux") mais non trivial, plutôt qu'un modèle trivialement vrai.

4.3.4 Degré d'adéquation et degré de corroboration

Je vais maintenant m'atteler à une définition un peu plus formelle du degré de corroboration informationnelle que les modèles scientifiques maximisent par définition.

Cette formalisation n'est peut-être pas strictement nécessaire pour comprendre le fond de mon propos. Cependant, comme je l'ai déjà remarqué, pour que mon approche de la scientificité comme maximisation d'une certaine quantité ait un contenu non trivial et puisse former une base pour une discussion critique, je me dois d'en proposer une définition aussi précise que possible – quitte à ce que ses limites soient rapidement atteintes.

Il faut garder à l'esprit toutefois que cette construction formelle précise est en partie indépendante du reste de mon travail. En effet, même si la proposition que je m'appête à faire ici est, sans aucun doute, hautement perfectible, cela ne remettra pas nécessairement en question la pertinence de mon approche globale consistant à percevoir la scientificité comme la manifestation d'un certain processus d'optimisation.

6. Plus précisément, pour n'importe quelle fonction f périodique, continue et dérivable, sa série de Fourier converge uniformément vers f . Cela signifie qu'une fois donné un intervalle de précision δ , la série de Fourier de f va se retrouver à une distance inférieure à δ en un nombre fini d'itérations.

À présent, soit $\epsilon = (\sigma, E, \mathbb{P}, D, (M_\theta)_{\theta \in \Theta}, d)$ une unité épistémique. On a supposé que la distance d sur E permettait de comparer un modèle donné, M_θ (pour $\theta \in \Theta$), avec des données D , et donc de donner un sens à l'expression : $d(m_\theta, D) \geq 0$.

Le degré d'adéquation de ϵ peut alors être défini comme :

$$\alpha(\epsilon) = \inf_{\theta \in \Theta} d(m_\theta, D). \quad (4.10)$$

Le degré d'adéquation est un nombre réel positif, et plus $\alpha(\epsilon)$ est élevé, moins ϵ est empiriquement adéquat. Au contraire, $\alpha(\epsilon) = 0$ signifie une adéquation parfaite, comme dans le cas des exemples de régularités endogènes à E cités plus haut.

Maintenant que l'on s'est muni d'une façon de définir le degré d'adéquation empirique d'une unité épistémique $\epsilon = (E, \mathbb{P}, D, (M_\theta)_{\theta \in \Theta}, d)$, on peut définir l'ensemble de toutes les données empiriques de même taille que D qui sont au moins aussi adéquates avec $(M_\theta)_{\theta \in \Theta}$ que D .

En d'autres termes :

$$\mathcal{D}[\epsilon] = \{D' \in E^{|D|} \mid \exists \theta \in \Theta, d(m_\theta, D') \leq \alpha(M)\}. \quad (4.11)$$

$E^{|D|} = E^n$ est équipé de la mesure produit dérivée de \mathbb{P} sur E , que l'on note aussi \mathbb{P} par abus de langage.

$\mathcal{D}[\epsilon]$ est donc un espace mesurable, sous-espace de $E^{|D|}$ et son "volume" $\mathbb{P}(\mathcal{D}[\epsilon]) \in [0, 1]$ représente à quel point ϵ aurait pu ne pas être aussi adéquat qu'il ne l'est, ce qui correspond précisément à ce que la notion de degré de falsifiabilité vise à recouvrir. Plus précisément, plus $\mathbb{P}(\mathcal{D}[\epsilon])$ est élevé, moins il est étonnant que l'on est pu trouver un θ tel que M_θ s'ajuste à ce point aux données D .

On définit le degré de corroboration informationnelle, c'est-à-dire la quantité d'information que l'on obtient à savoir que ϵ est à ce point empiriquement adéquat, comme :

$$\mathcal{C}[\epsilon] = -\log(\mathbb{P}(\mathcal{D}[\epsilon])). \quad (4.12)$$

Par construction, $\mathcal{C}[\epsilon]$ est positive pour tout ϵ . Toute unité épistémique ϵ contenant un modèle empirique correspondant à une régularité endogène est alors par construction telle que $\mathcal{C}[\epsilon] = 0$, et plus $\mathcal{C}[\epsilon]$ est élevé, plus l'information gagnée à savoir que le modèle empirique est adéquat à ce point avec les données est importante.

À partir de cette définition formelle de degré de corroboration informationnelle, on peut maintenant voir la scientificité (au niveau des modèles empiriques) comme la maximisation de cette quantité. C'est ce que je vais définir dans la section suivante.

4.4 Principe de maximisation et conséquences

4.4.1 Principe général de scientificité

Une tâche centrale de l'épistémologie est de pouvoir comparer différentes unités épistémiques entre elles, de manière à y opérer éventuellement une certaine sélection.

Dans cette partie, je me restreins au cas suivant : pour être comparables, deux unités épistémiques $\epsilon = (\sigma, E, \mathbb{P}, D, M, d)$ et $\epsilon' = (\sigma', E', \mathbb{P}', D', M', d')$ doivent partager certaines caractéristiques minimales :

- Elles visent à s'appliquer aux mêmes ensembles de situations empiriques, soit $\sigma = \sigma'$. Une condition plus lâche peut être que leurs ensembles d'application respectifs doivent avoir une intersection non nulle, soit $\sigma \cap \sigma' \neq 0$.
- Elles partagent le même cadre empirique, soit $E = E'$ (pas nécessairement le même fond de connaissance).
- Leurs données respectives doivent avoir une intersection non nulle, soit $D \cap D' \neq 0$.

On peut alors énoncer le principe de scientificité (empirique) suivant :

Principe de corroboration maximale : *Étant donné un ensemble fini d'unités épistémiques comparables $\mathcal{E} = \{\epsilon_1, \epsilon_2, \dots\}$, l'unité ayant le degré de scientificité maximale est celle qui maximise le degré de corroboration informationnelle $C[\epsilon]$ pour $\epsilon \in \mathcal{E}$.*

4.4.2 Quelques critères épistémologiques retrouvés

Comme on l'a déjà évoqué, cela fait longtemps qu'un projet épistémologique strictement *fondationnaliste* semble voué à l'échec. Il semble impossible de faire reposer la justification de nos connaissances empiriques sur les seules lois de la logique. Mon parti pris est d'assumer ce fait, et de remplacer cet objectif par l'énoncé d'un *principe* fondamental de scientificité.

Ce principe, bien qu'il soit la traduction formelle d'une intuition assez consensuelle et bien qu'on puisse le défendre comme tel, ne pourra pas être justifié en lui-même à partir des seules lois de la logique.

Ici, je vais montrer que la perte de justification logique induite par le fait de rajouter "à la main" un tel principe peut être compensée par le fait que celui-ci semble contenir en lui-même un certain nombre de critères épistémologiques bien connus : l'adéquation empirique, l'opérationnalisation précise des variables de base, la progressivité des *problemshifts* lakatosiens, ainsi que, dans une certaine mesure, la parcimonie et la cohérence.

En d'autres termes, ces critères épistémologiques se trouvent être des cas particuliers de l'application du principe plus général de scientificité.

L'adéquation empirique

Comme on l'a vu, et Popper le reconnaît également, l'adéquation empirique, bien qu'elle ne soit pas une condition suffisante de scientificité, en reste une condition nécessaire. Il peut sembler étrange à première vue que le principe de scientificité énoncé plus haut ne porte que sur la maximisation du degré de corroboration informationnelle $\mathcal{C}[\epsilon]$, et non pas sur le degré d'adéquation $\alpha[\epsilon]$, auquel il est lié mais sans s'y réduire strictement.

En réalité, comme je vais le montrer maintenant, le degré de corroboration est construit de telle façon à ce que l'adéquation empirique se trouve être une conséquence logique de la maximisation de \mathcal{C} , bien que dans un cas relativement restreint.

Le raisonnement est le suivant. Soit ϵ et ϵ' égaux en tout point excepté que ϵ' est composé d'un modèle empirique différent $M' \neq M$ et plus adéquat empiriquement. On a donc $\alpha(\epsilon') < \alpha(\epsilon)$. Dans ce cas, puisque M est moins adéquat que M' vis-à-vis des mêmes données, la région, dans l'ensemble des données possibles $E^{|D|}$, compatible avec M est "plus grande" que la région compatible avec M' . On a donc : $\mathbb{P}(\mathcal{D}[\epsilon]) > \mathbb{P}(\mathcal{D}[\epsilon'])$ et ainsi $\mathcal{C}[\epsilon] < \mathcal{C}[\epsilon']$.

En d'autres termes, un α croissant (une moins bonne adéquation) implique donc un plus faible degré de corroboration informationnelle. Ainsi, à l'inverse, exiger un fort degré de corroboration implique nécessairement un niveau plus élevé d'adéquation empirique. Dans ce cas particulier, l'exigence d'adéquation empirique est *déjà incluse* dans la maximisation du degré de corroboration.

Le même phénomène survient dans le second cas particulier suivant. On suppose ici que ϵ et ϵ' sont deux unités épistémiques qui ne diffèrent que

dans les données empiriques D et D' sur lesquelles elles s'appliquent, tel que $D \subset D'$. Supposons de plus que $\alpha(\epsilon') > \alpha(\epsilon)$.

En d'autres termes, on compare le même modèle M à plus de données et il se trouve être globalement moins adéquat. La situation concrète à laquelle on pourrait penser est que de nouvelles données soient tout à coup disponibles et que l'on réalise que notre modèle n'est plus aussi bon que ce que l'on pensait.

Quoiqu'il en soit, la région des données qui sont aussi adéquates avec M que ce dernier l'est avec D est plus petite que la région de données aussi adéquates avec M que ce dernier l'est avec D' , ou encore : $\mathbb{P}(\mathcal{D}[\epsilon]) < \mathbb{P}(\mathcal{D}[\epsilon'])$, et donc $\mathcal{C}[\epsilon] > \mathcal{C}[\epsilon']$.

Ainsi, dans ce cas particulier également, un degré d'adéquation plus faible (un α croissant) implique un degré de corroboration lui aussi plus faible, et donc de nouveau, l'exigence d'une maximisation du degré de corroboration contient en elle-même l'exigence d'un degré d'adéquation élevé.

Il est commode d'illustrer ce phénomène avec un cas spécifique, afin d'en rendre compte d'une manière plus "visuelle". Prenons le cas où x et y sont deux variables quantitatives et continues, pour lesquelles sont faites n mesures sur une situation empirique donnée σ , c'est-à-dire que l'on a $D = \{(x_i, y_i)\}$.

On considère que les variables x et y ne peuvent prendre des valeurs que sur des intervalles bien définis, notés Δx et Δy respectivement. En d'autres termes, la mesure de probabilité \mathbb{P} sur $E = \mathbb{R}^2$, l'espace correspondant au cadre empirique (x, y) , est nulle hors du rectangle $\Delta x \times \Delta y$, et égale à la mesure de Lebesgue à l'intérieur, normalisée telle que $\mathbb{P}(\Delta x \times \Delta y) = 1$.

On prend alors un modèle empirique linéaire, c'est-à-dire : $M : y = kx$, avec k une constante dont on suppose que l'on connaît la valeur. Admettons que l'on obtienne une certaine adéquation empirique α mesurée à l'aide de la distance d usuelle sur l'espace E :

$$\alpha = d(M, D) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - kx_i)^2}. \quad (4.13)$$

On note $\epsilon = \{E, \mathbb{P}, D, M, d\}$ l'unité épistémique ainsi constituée.

Une manière de visualiser l'espace $\mathcal{D}[\epsilon]$ des données qui sont au moins aussi adéquates avec M que D est de changer d'espace de représentation. Au lieu de considérer les points de données comme des points dans l'espace à deux dimensions E défini par les variables x et y , on considère un nouvel

espace \mathbb{E}_n de dimension n (la taille des données), tel que chaque dimension représente un point de donnée particulier.

En d'autres termes, l'axe 1 représente l'ensemble des valeurs possibles pour la mesure de y sur le premier point de mesure, et ainsi de suite. L'ensemble des valeurs mesurées pour la variable y prend alors la forme d'un vecteur $\mathcal{Y} \in \mathbb{E}_n$, comme illustré sur la figure 4.4.

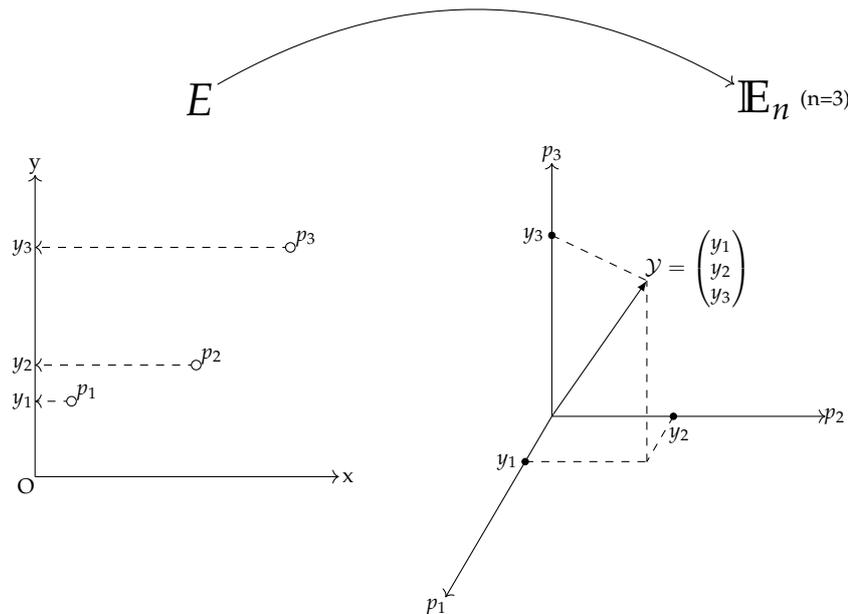


FIGURE 4.4 – Changement de perspective. À gauche : trois points de données dans l'espace défini par le cadre empirique $E = (x, y)$. À droite : une représentation géométrique de ces données comme un vecteur dans l'espace \mathbb{E}_n à trois dimensions.

Le modèle M , une fois données les n valeurs mesurées de la variable x , se représente lui aussi sous la forme d'un vecteur dans l'espace \mathbb{E}_n , que je note $\mathcal{Y}_M = \{y_{Mi}\}_i$ ici. Étant donné la valeur de l'adéquation empirique α entre M et D , l'ensemble des données dans \mathbb{E}_n qui sont au moins aussi adéquates avec M que D , c'est-à-dire $\mathcal{D}[\epsilon]$, est donc l'ensemble des vecteurs $\mathcal{Y} = \{y_i\}_i$ de \mathbb{E}_n tel que :

$$\frac{1}{n} \sum_{i=1}^n (y_i - y_{Mi})^2 \leq \alpha^2. \quad (4.14)$$

En d'autres termes, l'ensemble $\mathcal{D}[\epsilon]$ est une *boule* (en dimension n , donc une *hyperboule*) de centre \mathcal{Y}_M et de rayon α , comme représenté sur la figure 4.5.

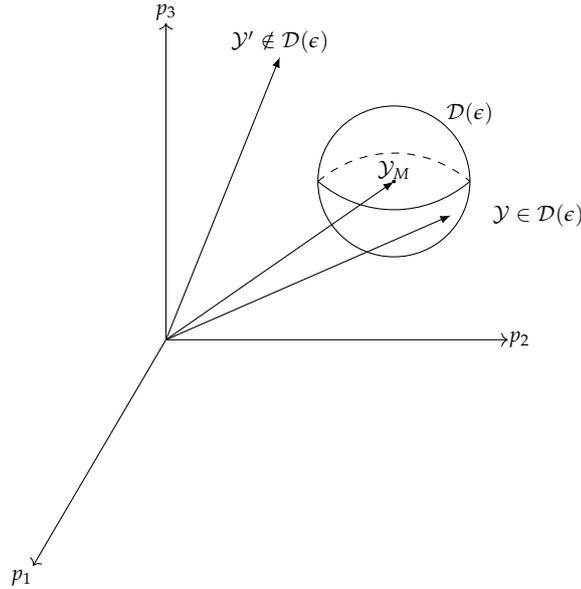


FIGURE 4.5 – Dans \mathbb{E}_3 , $\mathcal{D}(\epsilon)$ est représenté par une hyperbole de rayon α , centrée sur le vecteur \mathcal{Y}_M prédit par le modèle M . \mathcal{Y} représente un vecteur-données qui est encore plus compatible avec M que D , tandis que \mathcal{Y}' représente un vecteur-données qui est moins compatible avec M que D .

Le volume de cette hyperbole, noté $V^n(\alpha)$, est donné par la formule suivante :

$$V^n(\alpha) = \frac{\pi^{n/2} \alpha^n}{\Gamma(\frac{n}{2} + 1)}, \quad (4.15)$$

où Γ désigne la *fonction gamma*.⁷

Le volume $\mathbb{P}(\mathcal{D}[\epsilon])$, qui encode à quel point on gagne de l'information à savoir qu'un modèle est à ce point adéquat à certaines données, est alors proportionnel à $(\alpha/(\Delta y))^n$. Le degré de corroboration correspondant est alors égal à :

$$\mathcal{C}(\epsilon) = -n \log(\alpha) + n \log(\Delta y) + f(n), \quad (4.16)$$

où $f(n)$ est une quantité qui ne dépend que de n . À l'aide de l'équation (4.16), on voit alors directement que \mathcal{C} est une fonction décroissante de α , et donc l'exigence d'un degré \mathcal{C} élevé, toute chose étant égale par ailleurs, implique que α soit le plus faible possible, c'est-à-dire que le modèle soit le plus adéquat que possible.

7. C'est une fonction classique qui prolonge la fonction factorielle, initialement définie sur les nombres entiers, à l'ensemble des nombres complexes.

Le critère d'adéquation empirique, dans les cas particuliers décrits ci-dessus, est donc déjà encodé dans le principe fondamental de scientificité énoncé plus haut.

Opérationnalisation précise des variables de base

Un autre critère épistémologique que l'on peut également déduire de l'exigence d'une maximalisation du degré de corroboration informationnelle est le fait que les variables de base composant le cadre empirique E doivent être autant que faire se peut *précisément opérationnalisées*.

Bien entendu, l'opérationnalisation d'une variable, c'est-à-dire l'ensemble d'opérations concrètes qu'il faut faire pour lui donner une valeur, n'est pas un processus unique. Une même variable (comme par exemple, la distance en physique, ou bien le chômage en sociologie) peut être opérationnalisée de différentes façons, selon la question précise qui est posée au réel.

Pour autant, cette multiplicité n'empêche pas une certaine rigueur dans la définition opérationnelle que l'on donne. De plus, le parti pris ici est de considérer que toute variable, même définie de manière floue, est forcément tributaire d'une certaine opérationnalisation, même implicite.

Or, si cette opérationnalisation est trop floue, cela va induire de l'ambiguïté sur les valeurs des différentes variables une fois que l'on va les appliquer à des situations empiriques données. Cette ambiguïté va alors avoir pour conséquence que pour un même modèle empirique et une adéquation α donnée, davantage de données seront a priori compatible à un degré α avec le modèle empirique.

En d'autres termes, si des variables peu opérationnalisées augmentent le degré d'adéquation empirique d'un modèle correspondant, c'est en toute probabilité au désavantage du degré de corroboration.

Ainsi, demander un degré de corroboration maximal implique d'exiger des variables de base les plus précisément opérationnalisées que possible.

La progression des *problemshifts* lakatosiens

Comme on l'a rappelé dans le chapitre 2, Imre Lakatos a développé son épistémologie à partir d'une version sophistiquée du falsificationnisme. Selon Lakatos comme selon Popper, il est acceptable de modifier une ou plusieurs hypothèses auxiliaires pour "sauver" une théorie de la réfutation si et seulement si cette modification augmente le degré total de falsifiabilité de la théorie et finalement mène à de nouvelles découvertes.

Du point de vue développé ici, restreint aux modèles empiriques, un modèle M , ou l'unité épistémique qui le contient ϵ , rencontre une anomalie empirique (ce qu'on pourrait appeler naïvement une "réfutation") lorsque son degré d'adéquation $\alpha(\epsilon)$ descend en dessous d'un certain seuil critique α_{crit} . À partir de là, un ajustement est une modification d'un élément de ϵ en une unité ϵ' comparable tendant à résoudre l'anomalie, c'est-à-dire de telle sorte que $\alpha(\epsilon') < \alpha_{crit} < \alpha(\epsilon)$.

Cependant, ce que rejette l'épistémologie poppérienne ainsi que lakatosienne, ce sont les modifications *ad hoc*. Il s'agit de modifications qui, certes, rendent le modèle plus adéquat, mais trivialement, en quelque sorte, c'est-à-dire sans faire augmenter globalement le degré de corroboration.

Au contraire, l'exigence que doit satisfaire un tel *problemshift* $\epsilon \rightarrow \epsilon'$, selon Popper, Lakatos et le principe général de scientificité énoncé plus haut, est simplement que :

$$\mathcal{C}(\epsilon') > \mathcal{C}(\epsilon). \quad (4.17)$$

Dans les termes de Lakatos, un *problemshift* tel que :

$$\alpha(\epsilon') < \alpha(\epsilon) \text{ et } \mathcal{C}(\epsilon') \leq \mathcal{C}(\epsilon) \quad (4.18)$$

est un *problemshift dégénèrescent*. Ainsi, a priori, on peut modifier n'importe quel élément d'une unité épistémique ϵ face à une anomalie empirique, mais *pas de n'importe quelle façon*. Ce que je soutiens ici, c'est que la façon acceptable de faire est ainsi déjà encodée dans le principe de maximisation énoncé plus haut.

Au contraire, un *problemshift* tel que :

$$\alpha(\epsilon') < \alpha(\epsilon) \text{ et } \mathcal{C}(\epsilon') > \mathcal{C}(\epsilon) \quad (4.19)$$

est un *problemshift progressif* dans le langage de Lakatos, dans le sens où résoudre une anomalie mène à accroître le degré global de corroboration.

On peut donner une représentation graphique de cette dynamique dans un diagramme (α, \mathcal{C}) . Une représentation de *problemshifts* progressifs et dégénèrescents est donné dans la partie gauche de la figure 4.6.

Cette vision est bien entendu trop simpliste. En effet, un *problemshift* n'est usuellement pas composé uniquement de deux modèles, mais d'une suite de modèles (une suite de *théories* pour Lakatos). Un modèle, face à une anomalie empirique, peut être modifié par une hypothèse qui pourrait à première vue apparaître comme *ad hoc*, mais s'avérerait finalement féconde.

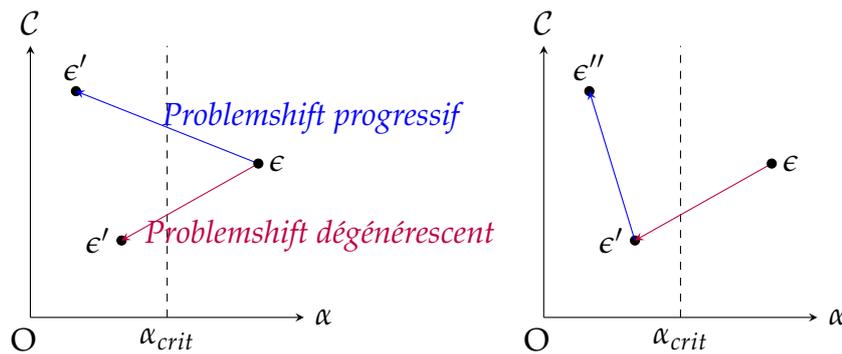


FIGURE 4.6 – Une représentation diagrammatique des *problemshifts* progressifs et dégénèrescents à la Lakatos. L'origine O du diagramme représente les régularités endogènes au cadre empirique E correspond : $\alpha = 0$ (elles ont une adéquation empirique parfaite) mais $C = 0$ également (elles ne contiennent aucune information). Diagramme de gauche : différence entre les deux types de *problemshifts*. Diagramme de droite : un *problemshift* qui semble au départ dégénèrescent mais qui est finalement progressif.

Par exemple, imaginons que ϵ' dans le cas dégénèrescent fasse une prédiction qui est ensuite empiriquement confirmée par de nouvelles données, de telle façon à ce que le degré de corroboration finisse par augmenter encore plus.

On pourrait schématiser cette situation comme :

$$\epsilon \xrightarrow{\alpha} \epsilon' \xrightarrow{C} \epsilon'' \quad (4.20)$$

Ici, on a réduit ϵ aux éléments (M, D) par abus de langage. Cet ajustement est diagrammatiquement représenté sur le coté droit de la figure 4.6.

Bien que toujours trop simpliste, cela permet toutefois de nous figurer la dynamique des modèles empiriques suite à la découverte de nouvelles données (évolution de D) ou à la suite d'autres modifications de l'unité épistémologique (modification des variables de base, c'est-à-dire de (E, \mathbb{P}) , du modèle empirique lui-même M , ou des méthodes statistiques pour évaluer l'adéquation empirique, traduite dans une modification de la distance d sur E , ...) comme des *trajectoires* dans un certain diagramme. Les critères épistémologiques apparaissent alors comme des contraintes sur ces trajectoires.

Pour conclure cette partie, on peut représenter schématiquement trois cas importants, dans la situation particulière d'un *problemshift* $\epsilon \rightarrow \epsilon'$ où l'on ne considère que des changements de données D ou de modèle empirique M .

En notant $\Delta\alpha = \alpha(\epsilon') - \alpha(\epsilon)$ et $\Delta C = C(\epsilon') - C(\epsilon)$, et en suivant le même raisonnement que plus haut, on en déduit que :

$$\Delta\alpha > 0 \text{ implique } \Delta C \leq 0. \quad (4.21)$$

Cela est dû au fait déjà mentionné que les degrés d'adéquation et de corroboration ne sont pas entièrement indépendants l'un de l'autre, et que si α augmente, alors, *ceteris paribus*, C diminue.

Il s'en suit logiquement, comme on l'a déjà noté également, que :

$$\Delta C > 0 \text{ implique } \Delta\alpha \leq 0. \quad (4.22)$$

Cela signifie, *ceteris paribus*, qu'un degré de corroboration croissant implique forcément une meilleure adéquation empirique.

Cependant, et c'est bien le fond du problème, le contraire n'est pas vrai : un degré d'adéquation croissant n'implique pas un degré de corroboration plus élevé. On peut très bien avoir :

$$\Delta\alpha \leq 0 \text{ et } \Delta C \leq 0. \quad (4.23)$$

Ce cas correspond à une situation de dégénérescence.

Les trois trajectoires correspondantes dans le diagramme (α, C) : réfutation (4.21), progression (4.22) et dégénérescence (4.23) sont représentées dans la figure 4.7. Bien entendu, elles doivent être considérées davantage comme des tendances que comme des trajectoires effectives.

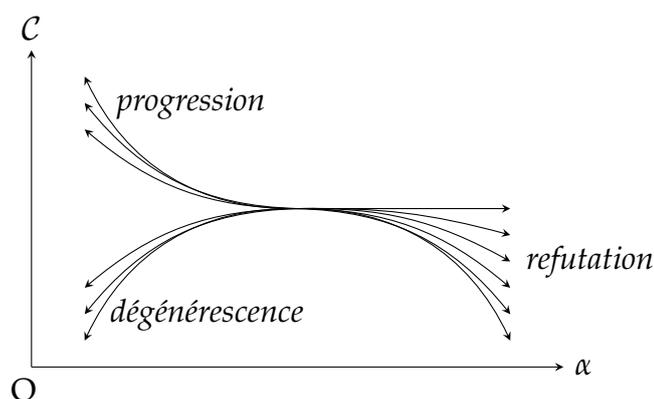


FIGURE 4.7 – Esquisse de trois familles possibles de trajectoires pour les modèles empiriques.

Finalement, il reste une question en suspend, que je n'ai pas eu l'occasion de traiter dans le présent travail. Il s'agit de savoir dans quelle mesure ce cadre d'analyse peut être adapté facilement au cas général pour lequel n'importe quel élément d'une unité épistémique peut être modifié.

Parcimonie et cohérence

Un critère de scientificité généralement reconnu comme pertinent est le principe de parcimonie, ou rasoir d'Occam. Comme précisé dans (SOBER, 2015), si sa pertinence est plutôt consensuelle dans l'histoire de la philosophie, et en particulier en épistémologie, la difficulté réside plutôt dans la justification que l'on peut donner de ce principe.

Ici, je propose, à la suite de Popper et d'autres travaux récents (FALK et MUTHUKRISHNA, 2021), de voir la parcimonie comme un cas particulier de la falsifiabilité. En d'autres termes, de ce point de vue, les théories les plus parcimonieuses sont meilleures parce qu'elles sont plus falsifiables.

Dans les termes du modèle que j'ai développé plus haut, les modèles empiriques les plus parcimonieux sont ceux qui reposent sur le moins de paramètres libres, et donc qui restreignent le plus l'espace des données. Ils sont donc, pour un degré d'adéquation fixé, compatibles avec "moins" de données. Ainsi, ils possèdent un degré de corroboration informationnelle supérieur.

La cohérence est parfois invoquée pour justifier la parcimonie : une théorie parcimonieuse en hypothèses est une théorie qui repose sur le moins de choses inexplicées, inconnues ou nouvelles, et donc qui est la plus cohérente avec ce qui est déjà considéré comme une connaissance établie.

Cependant, la cohérence peut également être vue comme un cas particulier du principe de maximisation du degré de corroboration. En effet, exiger plus de cohérence implique des contraintes supplémentaires sur l'espace des données, exactement comme l'application du principe de parcimonie. Ainsi, les modèles les plus cohérents sont ceux pour lesquels le plus de paramètres sont fixés (par cohérence avec ce que l'on sait par ailleurs). De ce fait, la région de l'espace des données compatibles, pour un α donné, avec un tel modèle est plus restreinte que pour un modèle similaire mais moins cohérent.

Prenons un exemple, déjà cité : l'invocation d'un principe de simplicité par Henri Poincaré pour justifier du choix d'une courbe plutôt qu'une autre lorsqu'il s'agit d'ajuster un ensemble de données, appliqué cette fois-ci au cas de la détermination de la masse de Jupiter :

On peut déduire la masse de Jupiter soit des mouvements de ses satellites, soit des perturbations des grosses planètes, soit de celles des petites planètes. Si l'on prend la moyenne des déterminations obtenues par ces trois méthodes, on trouve trois nombres très voisins mais différents. On pourrait interpréter ce résultat en

supposant que le coefficient de la gravitation n'est pas le même dans les trois cas; les observations seraient beaucoup mieux représentées. Pourquoi rejetons-nous cette interprétation? Ce n'est pas qu'elle soit absurde, c'est qu'elle est inutilement compliquée. (POINCARÉ, 1902, p. 162)

Dans ce passage, la simplicité est vue comme une convention qui guide l'enquête scientifique, sans autre justification. Parmi différentes fonctions pouvant s'ajuster à un ensemble de données, on choisit la "plus simple", par exemple une droite linéaire plutôt qu'une courbe faisant des zigzags compliqués.

Dans le cas de la masse de Jupiter, on interprète les trois résultats différents comme étant dû à des imprécisions d'observations et non pas à une constante gravitationnelle G qui dépendrait de la situation particulière.

D'un point de vue purement logique, et comme Poincaré le suggère, rien ne nous empêche a priori de choisir cette seconde option. Si on rejette cette possibilité, c'est en vertu d'un principe indépendant, que Poincaré appelle la "simplicité", afin d'écartier une solution "inutilement compliquée".

Du point de vue que j'ai développé dans ce chapitre, ce choix est justifié par le même principe : celui de la maximisation du degré de corroboration (et non nécessairement du degré d'adéquation). Un modèle reposant sur la même valeur de la constante G pour différentes situations expérimentales a un degré de corroboration informationnelle plus élevé qu'un modèle qui permet à G de varier d'une situation à l'autre. En effet, la région des données, toute chose étant égale par ailleurs, qui est compatible avec le premier modèle est tout simplement plus restreinte que celle qui est compatible avec le deuxième.⁸

Ces considérations à propos de la vertu épistémologique de la parcimonie ont aussi à voir avec la partie explicative des unités épistémiques, c'est-à-dire les modèles théoriques. On l'a volontairement laissée de côté ici en nous concentrant sur les modèles empiriques, mais elle forme précisément l'objet du chapitre suivant.

4.5 Conclusion

J'ai présenté dans ce chapitre une tentative de mise à jour des idées centrales du falsificationnisme en les adaptant à une vision de la connaissance

8. Je rappelle qu'on a pu donner un sens précis à la notion de région "plus restreinte" qu'une autre à l'aide de la mesure de probabilité \mathbb{P} que l'on s'est donnée sur E .

scientifique basée sur les modèles comme unités d'analyse.

D'un côté, plusieurs points positifs semblent ressortir de cette tentative.

Elle permet, tout d'abord, d'intégrer les aspects bénéfiques du falsificationnisme (détecter des régularités non triviales comme élément épistémique fondamental) tout en surmontant certaines difficultés épistémologiques que l'on a pu mettre en évidence dans les premiers chapitres.

En particulier, mon approche intègre dès le début des notions comme la charge théorique de l'observation, le holisme de la confirmation ou la sous-détermination théorique. Pour cela, j'ai construit une certaine quantité, le degré de corroboration informationnelle, et j'ai posé un principe de scientificité empirique exigeant sa maximisation.

La charge théorique de l'observation a été prise en compte dès le début par le fait d'explicitier le cadre empirique en jeu, et le problème épistémologique associé neutralisé en faisant de ce cadre empirique un élément variable comme un autre dans le processus de maximisation. La charge théorique n'apparaît alors plus comme un problème en soi, mais bien comme *une raison supplémentaire* de considérer la falsifiabilité des théories comme une vertu épistémologique centrale.

Pour intégrer l'aspect structurel complexe de notre connaissance scientifique sans pour autant adopter une unité d'analyse nécessairement de plus en plus globalisante comme ça a été la tendance ces cinquante dernières années, j'ai assumé dès le début reposer sur un certain (méta-)modèle de notre connaissance scientifique, et j'ai ensuite basé le reste de ma construction sur ce modèle.

De même, la sous-détermination théorique n'est plus spécialement un problème en soi, puisque ce n'est pas l'adéquation empirique qui compte, mais bien le degré de corroboration tel qu'on l'a défini. Ainsi, il semble exister un critère de choix entre différents modèles, quand bien même il existerait un grand nombre de modèles aussi adéquats les uns que les autres avec les données.

Deuxièmement, mon approche formelle prend les modèles empiriques comme unités de base, et non plus uniquement des "systèmes d'énoncés". Cela me permet de définir plus proprement certaines quantités d'essence poppérienne qui avaient déjà fait l'objet de tentatives à l'époque. Comme je l'ai noté en début de chapitre, on peut penser au degré de falsifiabilité basé sur les "classes de falsificateurs", ou encore à la "vérisimilitude", .

Les conditions pour que deux unités épistémiques soient comparables étant, dans mon cas, plus faibles que pour Popper, mon approche permet

de s'appliquer à plus de situations, et pas uniquement à des situations pour lesquelles les "classes de falsificateurs" doivent être sous-ensemble l'une de l'autre.

C'est le fait d'explicitier un cadre empirique, de se donner une mesure de probabilité sur cet espace, et de considérer comme élément principal les modèles empiriques comme des sous-espaces de cet espace, qui permet d'élargir la notion de degré de falsifiabilité à davantage de cas sans essuyer les mêmes limites que Popper – bien que l'idée fondamentale reste la même.

Un dernier point positif est que ce modèle ne dépend a priori pas de la discipline scientifique, donc de l'objet, en question. En effet, tout ce qu'on a supposé, c'est qu'il faut un certain cadre empirique, c'est-à-dire un certain nombre de variables assez bien définies, et que ce que l'on cherche à faire, c'est à détecter des régularités dans les données en définissant des modèles empiriques que l'on voit comme des (familles de) sous-espaces.

On a vu que ce (méta-)modèle s'applique a priori à tout un ensemble de cas concrets, à la fois en sciences physiques et en sciences sociales. En d'autres termes, rien dans la définition des différents objets que l'on a donnée ne repose sur le fait que les objets à l'étude sont des systèmes ou phénomènes spécifiquement physiques.

La vision globale qui se dégage de cette première étape dans l'élaboration d'un modèle unitaire de la scientificité est que l'on peut voir celle-ci comme un processus d'optimisation. Cependant, il ne faut pas se méprendre sur la portée de ce modèle : il ne suppose absolument pas que c'est ce que font les scientifiques au quotidien, ni même que c'est ce qu'ils doivent faire.

En réalité, cette optimisation n'intervient qu'au niveau global et collectif, restant hors de portée du travail scientifique individuel. Les scientifiques, dans leur quotidien, vont bien plus avoir recours à certaines *heuristiques*, comme des critères épistémologiques précis s'appliquant dans des cas particuliers : l'adéquation empirique, la parcimonie des hypothèses, la simplicité, et ainsi de suite.

Du point de vue développé dans cette thèse, ces critères gagnent à être interprétés comme des *proxy*,⁹ à l'échelle micro, de l'optimisation ayant cours à plus grande échelle. Les critères épistémologiques de sélection utilisés par les scientifiques sont donc en général des bons critères, mais ils ne garantissent jamais la sélection de la meilleure théorie, justement parce qu'ils ne sont que des *proxy*, des indices partiels, d'une dynamique plus globale de

9. En sciences sociales, un *proxy* est une grandeur mesurable supposée donner une certaine idée d'un phénomène non directement observable – par exemple, le PIB par habitant est parfois utilisé comme un *proxy* du niveau de vie dans un pays.

maximisation qui reste difficile (voire impossible) à considérer en tant que telle à l'échelle individuelle.

D'un autre côté, il y a (au moins) deux coûts majeurs à ma démarche. Le premier, qui ne semble pas lui être propre, c'est que définir la scientificité revient, il semble, à énoncer un principe auquel doivent se conformer les unités épistémiques *sans que l'on puisse en donner une justification a priori*.

Comme on l'a vu en fin de chapitre, il semble toutefois que la perte de justification logique peut être quelque peu compensée par le fait de retrouver "naturellement", à partir de ce principe, un certain nombre de critères épistémologiques bien connus et acceptés comme des cas particuliers. On offre ainsi une vision épistémologique unifiée dans le sens où des critères épistémologiques a priori de nature différentes se trouvent en réalité considérés comme différents aspects d'un même principe unique.

Le second défaut de mon approche est le niveau de formalisation qui peut sembler quelque peu démesuré. S'il permet de construire, au moins de manière abstraite, une version améliorée du degré de corroboration poppérien, il semble limiter toute application pratique de ces constructions.

En effet, comme je l'ai annoncé, mon approche n'a pas pour but de fournir des outils pouvant s'appliquer directement à des cas concrets, mais constitue une tentative de reconstruction formelle de ces critères épistémologiques, sans qu'il soit a priori garantie qu'une telle reconstruction soit possible.

De plus, cette première étape constructive se focalise principalement sur les modèles empiriques, c'est-à-dire des unités d'analyse "proches" (dans le tissu millien ou quinéen de notre connaissance) de l'expérience. Or, les modèles empiriques eux-mêmes s'articulent entre eux et peuvent être également déduits théoriquement, dans certains cas, d'autres types d'unités épistémiques que l'on va appeler tout naturellement des *modèles théoriques*.

Pour être complet dans notre analyse, il nous reste donc à étudier les conditions de scientificité de ces modèles théoriques : c'est ce que je vais aborder au chapitre suivant.

Chapitre 5

Épistémologie des modèles théoriques

5.1 Introduction

J'ai présenté, au chapitre 3, un ensemble d'arguments *théoriques* à propos de la nécessité ou non, pour les sciences sociales et en particulier la sociologie, de jouir d'un cadre épistémologique à part, c'est-à-dire d'une définition propre de la scientificité. Une question restée en suspend est cependant celle de *la possibilité effective* de produire de la connaissance scientifique sur le monde social sans devoir s'inscrire dans un cadre épistémologique singulier. En effet, bien qu'il ne semble pas y avoir de raisons de principe empêchant une telle production, un exemple concret augmenterait toutefois significativement la plausibilité de cette position.

Au chapitre 4, j'ai proposé une approche permettant de définir un certain degré de scientificité d'un modèle empirique, c'est-à-dire d'une unité cognitive ayant pour objectif d'extraire de l'information d'un ensemble de données empiriques. Cependant, les modèles en général ne sont pas uniquement constitués d'une partie empirique mais possèdent également des composantes *théoriques*.

Comme on l'a déjà évoqué dans la section 3.4 à propos des deux grands types de lois en science qu'il s'agit de ne pas confondre, cette distinction modèles théoriques/modèles empiriques est assez classique dans la littérature. Elle correspond par exemple à la dichotomie modèles théoriques/modèles méthodologiques (SKVORETZ, 2016), ou plus généralement la distinction entre "lois fondamentales" et "lois phénoménologiques", ou encore entre "lois théoriques" et "lois empiriques".

De surcroît, les modes de justification de ces deux types de modèles ne semblent pas coïncider. En effet, si la falsifiabilité d'un modèle empirique peut être assez facilement définie (comme je me suis efforcé de le montrer

dans le chapitre 4 de cette thèse), ce n'est pas le cas des principes fondamentaux qui, comme l'avait noté Lakatos (voir section 2.3), n'entretiennent pas le même rapport à la réfutation empirique.

Le présent chapitre a *deux* fonctions principales relativement à l'objet global de mon travail de thèse : présenter un contre-exemple concret aux thèses pluralistes, et partir de ce cas pour élaborer, plus généralement, un critère épistémologique portant sur la composante *théorique* des modèles.

Ce chapitre offre ainsi, d'une part, une présentation du courant dit "analytique" en sociologie qui constitue, comme on va le voir, un contre-exemple manifeste aux thèses pluralistes. La sociologie analytique prétend en effet produire une connaissance scientifique sur le monde social sans pour autant devoir dépendre d'une épistémologie alternative à celle ayant cours dans les autres sciences, par exemple en physique ou en biologie. Elle se donne pour but de mettre en exergue des phénomènes sociaux généraux à l'aide de variables mesurables bien définies, et propose des explications systématiques de ces phénomènes au sein de modèles théoriques abstraits, formels, mathématisés, et même parfois numériques. La section 5.2 sera consacrée à une présentation générale de ce courant, à la fois dans ses aspects historiques et théoriques.

Cependant, pour que la sociologie analytique constitue bel et bien un argument *par l'exemple* valide en faveur de la thèse moniste, encore faut-il lui adresser certaines questions épistémologiques critiques quant aux modes de justification de ses différents éléments. En effet, ce n'est pas parce que ce courant prétend produire une connaissance scientifique en adoptant les façons de faire existantes ailleurs que cet objectif est effectivement atteint.

En particulier, le cœur de l'approche analytique en sociologie est constitué de *théories de l'action* qui modélisent le comportement (micro) des individus pour expliquer, par agrégation de ces comportements, les phénomènes sociaux (macro). Or, comme on le verra en section 5.3, des débats importants existent à propos des critères épistémologiques pertinents à appliquer à ces modèles explicatifs. D'un côté, ces derniers semblent reposer très souvent sur des hypothèses hautement idéalisées, voire notoirement fausses – concernant le comportement humain, par exemple – ce qui semble miner quelque peu leur portée explicative. D'un autre côté, rendre ces modèles plus "réalistes" implique l'ajout d'hypothèses et peut alors vider l'explication de tout contenu : cette dernière peut en effet se retrouver sous-déterminée à un point tel qu'elle peut trivialement s'adapter à n'importe quelle situation.

Ce dilemme n'est pas propre aux théories de l'action en sciences sociales,

mais caractérise une situation épistémologique plus générale concernant les modèles théoriques en science. Je proposerai alors, en section 5.4, un critère épistémologique qui permettra à la fois de résoudre le dilemme décrit ci-dessus et de compléter l'élaboration d'un modèle de scientificité dans sa partie *théorique*.

Finalement, je mettrai à profit cette élaboration théorique générale afin d'évaluer à quel point les modèles théoriques mobilisés en sociologie analytique répondent effectivement à ce critère. Cela permettra de répondre à ma question initiale qui était de savoir dans quelle mesure ce courant de la sociologie constitue un contre-exemple robuste à la position pluraliste.

5.2 La sociologie analytique

La sociologie analytique, en plus de se présenter comme un contre-exemple éventuel à la thèse épistémologique pluraliste, est un champ de recherche relativement jeune dans l'histoire des sciences sociales, qui est toujours en développement et qui ne fait pas consensus, aujourd'hui, au sein de la sociologie.¹ Elle constitue, pour ces raisons, un cas d'étude intéressant à aborder dans le cadre de mon travail de thèse.

Si ce sont ses aspects épistémologiques qui m'intéressent principalement dans cette thèse, il me faut tout d'abord présenter ce qu'est la sociologie analytique. C'est ce à quoi est dédiée cette partie.

Après un bref historique, j'en décrirai les grands principes à travers l'effort programmatique constant dont ont fait preuve ses tenants. Je clôturerai cette présentation en exposant le cadre général de la *théorie du choix rationnel*. Il s'agit de la théorie de l'action la plus développée et discutée dans la littérature, et elle focalisera la discussion épistémologique sur laquelle porte le reste du chapitre.

5.2.1 Une brève histoire de la sociologie analytique

La sociologie analytique est un courant de la sociologie qui se développe depuis le milieu des années 1990. Elle cristallise l'héritage intellectuel de plusieurs sociologues de renom comme Robert Merton, Raymond Boudon, James Coleman, Jon Elster ou encore Peter Hedström.

1. Pour rappel, comme on l'a vu au chapitre 3, *aucun* cadre théorique ni épistémologique ne fait pour le moment consensus en sociologie.

On peut identifier le symposium sur le thème des mécanismes sociaux ayant eu lieu à Stockholm en 1996 ainsi que l'ouvrage *Social mechanisms – An analytical approach to Social Theory* (HEDSTRÖM et SWEDBERG, 1998) lui faisant suite comme des repères historiques importants de la naissance de ce courant.

L'ambition première d'un tel rassemblement était de réunir des chercheurs en sciences sociales, particulièrement en sociologie, autour d'un programme de recherche commun fondé principalement sur deux réquisits fondamentaux :

LA CLARTÉ CONCEPTUELLE : les concepts et les variables de base en jeu dans la mise en évidence d'un phénomène social doivent être définis le plus clairement possible. Ils doivent être, autant que faire se peut, *opérationnalisés*, c'est-à-dire donnés avec les méthodes explicites à utiliser pour les mesurer, le tout dans l'optique de tester des hypothèses bien identifiées.

L'EXPLICATION À BASE DE MÉCANISMES : les théories sociologiques doivent être développées en assumant leur portée explicative. Les éléments de base des explications sociologiques doivent être des mécanismes bien définis et identifiés, agissant notamment au niveau micro (individuel) et permettant de reproduire les phénomènes sociaux comme émergeant macroscopiquement de l'agrégation de l'effet de ces mécanismes.

Je reviendrai en détail sur ces différents points. Il faut cependant remarquer que l'émergence d'un tel programme de recherche correspond aussi à une certaine frustration collectivement ressentie face à l'état dans lequel se trouvaient (et se trouvent toujours en partie) les sciences sociales. Depuis la fin des années soixante, en effet, ces disciplines ont connu une "explosion de paradigmes" (CUIN, GRESLE et HERVOUET, 2017, Chap. VII), et une grande partie des travaux sociologiques s'est alors focalisé davantage sur des descriptions sociographiques de plus en plus précises, singulières et de portée restreinte. Ces champs de recherche ont peu à peu perdu leur ambition explicative, et se sont tournés davantage vers l'interprétation et la *compréhension* des actions des individus. Le qualificatif "analytique" est d'ailleurs le reflet d'une logique d'opposition à ces traditions sociologiques, dont ce courant cherche à se distinguer sur plusieurs points qu'il faudra préciser.

L'ouvrage *Dissecting the social* de Peter Hedström (HEDSTRÖM, 2005) marque un premier pas dans l'explicitation systématique des différents principes épistémologiques et méthodologiques sous-tendant la sociologie analytique – d'après (MANZO, 2014b, p. 1). C'est d'ailleurs avec cet ouvrage que l'expression *sociologie analytique* "rentre officiellement dans le vocabulaire sociologique". Il sera suivi de plusieurs ouvrages collectifs et autres *Handbooks* (HEDSTRÖM et BEARMAN, 2009 ; DEMEULENAERE, 2011 ; MANZO, 2014a ; MANZO, 2021) ayant pour objectifs à la fois de clarifier les différents aspects du programme de recherche analytique, de présenter un large panel de travaux sociologiques produits par des chercheurs qui s'en réclament, et de répondre aux diverses critiques qui émergent à mesure que ce programme se consolide et se structure.

Ces critiques touchent plusieurs aspects. Par exemple, le manque de définition claire du concept de "mécanisme social", ou encore à quel point la sociologie analytique se distingue réellement d'autres courants sociologiques comme la théorie (sociologique) du choix rationnel.

Parmi les débats importants qui ont traversé la sociologie analytique depuis les années 1990, celui qui m'intéresse plus particulièrement concerne les théories de l'action à l'œuvre au niveau micro afin d'expliquer les phénomènes macro. Comme on va le voir, l'héritage "théorie du choix rationnel" laisse place peu à peu à une plus grande diversité d'approches et de modèles, en partie en réaction à des critiques quant au statut épistémologique des principes sous-jacents.

Aujourd'hui, la sociologie analytique est un champ de recherche dynamique qui s'enrichit en particulier d'outils formels et numériques comme ceux des *réseaux complexes*, lui permettant à la fois de traiter un nombre considérable de données empiriques et de produire des modèles explicatifs à base d'agents au sein de simulations numériques (*agent-based modeling*, ou ABM). Sa diffusion est très hétérogène à travers le monde. En France, par exemple, la sociologie est relativement peu connue, adoptée ou même enseignée. À d'autres endroits, comme à l'université de Linköping, en Suède, elle semble davantage implantée. On peut en effet y trouver l'*Institute for Analytical Sociology* (IAS) fondé il y a une dizaine d'années par Peter Hedström.

5.2.2 Principes de base

Un premier ensemble de principes

Les principes au cœur de l'approche analytique en sociologie ont été explicités à mesure de son développement. Dans l'introduction de l'ouvrage séminal, l'approche "basée sur les mécanismes" en sociologie est caractérisée par quatre principes (HEDSTRÖM et SWEDBERG, 1998, p. 24) : l'action, la précision, l'abstraction et la réduction.

Le premier principe, l'action, précise qu'un mécanisme va plus loin qu'une simple association entre des variables. En effet, le cœur d'une explication sociologique, ce ne sont pas des variables mais bien des actions concrètes. Le mécanisme est précisément constitué de ces actions concrètes et de la façon avec laquelle elles s'agencent.

Ce principe rejoint l'*individualisme méthodologique* auquel ce programme adhère et qui stipule que l'on doit chercher à expliquer des phénomènes sociaux au niveau macro en les reproduisant comme le résultat de l'agrégation d'actions individuelles. On aura l'occasion d'y revenir plus loin dans ce chapitre en présentant le diagramme de Coleman.

Le second principe, la précision, restreint la portée des explications sociologiques à des phénomènes sociaux bien identifiés. Il ne s'agit donc pas de décrire des phénomènes trop généraux ou trop vaguement définis, ni d'élaborer des lois sociales universelles.

C'est un héritage direct des *middle range theories* de Robert Merton (MERTON, 1968, p. 61) qui enjoint à se focaliser sur des phénomènes bien circonscrits, "in between pure description and storytelling, on the one hand, and grand theorizing and universal social laws, on the other" (HEDSTRÖM et SWEDBERG, 1998, p. 2).

Le troisième principe, l'abstraction, invite à fonder les explications à base de mécanismes sur des *modèles*, c'est-à-dire sur une représentation simplifiée, abstraite, du phénomène étudié. Au sein de tels modèles, par définition, certains aspects sont considérés comme pertinents pour la situation en question, tandis que d'autres en sont (provisoirement) écartés. Les mécanismes sociaux sont alors rendus explicites et plus facilement traitables (*tractable*) sur le plan cognitif.

Enfin, le quatrième principe, la réduction, revient à chercher le plus possible à limiter les "*blackboxes*", c'est-à-dire à réduire les écarts qu'il pourrait y avoir entre ce que l'on cherche à expliquer et les mécanismes mobilisés pour l'expliquer.

Les auteurs donnent alors une définition plus formelle d'un *mécanisme social* :

A social mechanism is an integral part of an explanation which (1) adheres to the four core principles stated previously, and (2) is such that on the occurrence of the cause or input, *I*, it generates the effect or outcome, *O*. (HEDSTRÖM et SWEDBERG, 1998, p. 25)

Ils précisent toutefois que l'approche qu'ils prétendent développer ne repose pas autant sur une définition précise de ce que sont les mécanismes sociaux mais se caractérise davantage par un "style of theorizing". Ils ajoutent : "This style can be roughly characterized by a focus on middle-range puzzles or paradoxes for which precise, action-based, abstract, and fine-grained explanations are sought." (HEDSTRÖM et SWEDBERG, 1998, p. 25)

La question de la définition d'un mécanisme social a d'ailleurs ressurgi à plusieurs reprises le long du développement de la sociologie analytique. Peter Hedström et Petri Ylikoski ont par exemple fourni un tour d'horizon d'un certain nombre de définitions que l'on pouvait trouver dans la littérature, en 2010 (HEDSTRÖM et YLIKOSKI, 2010), ayant l'avantage d'en souligner la grande diversité.

Dans la section suivante, je vais présenter l'outil conceptuel central de la sociologie analytique : le diagramme de Coleman. Il illustre la forme globale des explications sociologiques poursuivies et développées par ce courant, au sein duquel on va pouvoir identifier la place qu'occupent les mécanismes.

L'individualisme structural et le diagramme de Coleman

Le diagramme de Coleman, aussi appelé "baignoire" de Coleman (*bath-tub*), est un outil conceptuel décrit en détail par James Coleman en 1987 (COLEMAN, 1987) ainsi que dans l'introduction de son ouvrage fondamental *Foundations of social theory* (COLEMAN, 1990). Si elle apparaît ici sous la forme explicite de diagramme, cette idée de structure générale des explications sociologiques se retrouve déjà chez les tenants de la "dutch sociology" prônant un "individualisme structural" (WIPPLER, 1978), ou encore chez Raymond Boudon (BOUDON, 1986, chapitre 2). Ce dernier la fait d'ailleurs remonter à la sociologie classique de tradition allemande avec Weber et Simmel, italienne avec Pareto et Mosca, ou encore américaine avec Parsons et Merton, sans compter bien sûr différents courants économiques.

L'idée fondamentale est que les phénomènes sociaux doivent être vus, décrits et expliqués comme la résultante d'actions individuelles élémentaires.

Ce paradigme épistémologique, appelé individualisme méthodologique, a été, en particulier en France, défendu par Raymond Boudon et les chercheurs qui s'en réclament – notamment les sociologues se reconnaissant dans l'approche analytique.

Une distinction importante à opérer pour éviter les quiproquos est qu'il peut exister dans la littérature deux types d'individualisme méthodologique. Le premier, la version "forte", stipule que l'explication d'un phénomène macro ne doit se réduire *qu'à* des considérations individuelles (les actions, motivations, objectifs poursuivis par les individus), sans intégrer dans l'*explicans* d'autres aspects macro.

La version "faible", la plus défendue en sociologie analytique, défend quant à elle que l'explication d'un phénomène social nécessite (ou peut nécessiter) de considérer certains faits, certaines situations ou structures sociales comme données, c'est-à-dire non expliquées. Cela rejoint complètement l'idée mer-tonienne de considérer principalement des théories de moyenne portée, qui ont un objectif explicatif bien identifié, et qui acceptent de considérer certaines choses comme étant données le temps de l'explication.

Considérant par principe que certains aspects ne sont pas à expliquer, cette vision des choses peut être perçue comme limitée par le fait qu'elle propose des explications systématiquement partielles, non totales, des phénomènes. Cependant, cette façon de procéder a également une vertu, à savoir précisément le fait de contrôler là où l'on s'arrête dans la suite potentiellement infinie de causes à mobiliser pour expliquer un phénomène donné.

Je qualifierai cette forme d'individualisme méthodologique de *structural*, à la suite par exemple de (WIPPLER, 1978) ou encore (MANZO, 2014a, p. 17), afin de le différencier de l'individualisme méthodologique dans son sens fort.

La baignoire de Coleman, représentée figure 5.1, est tout simplement une représentation diagrammatique de cette structure explicative.

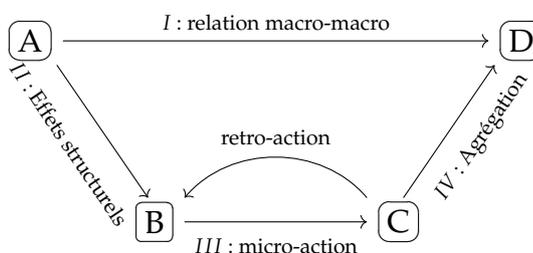


FIGURE 5.1 – Le diagramme de Coleman

Ce diagramme représente la façon de faire le lien entre des phénomènes sociaux observés au niveau *macro* (cases *A* et *D*) et des mécanismes décrits

par des relations entre différents niveaux *micro* (cases *B* et *C*).

Je vais maintenant expliciter ces différents éléments ainsi que les relations entre eux, illustré par un exemple classique en sociologie dont on a déjà parlé au chapitre 3, celui du différentiel social dans l'accès à la réussite scolaire ou académique.

$I : A \longrightarrow D$ représente une relation de causalité supposée entre deux variables au niveau macro. Il ne s'agit donc pas ici d'une "simple" association statistique (par exemple, une corrélation élevée entre des variables quantitatives, ou des odd-ratios dans le cas de variables catégorielles), mais bien d'une hypothèse authentiquement *inductive*, c'est-à-dire qui va *au-delà* des seules données empiriques disponibles. On pourrait imaginer que cette relation repose sur tout un ensemble de données pour lesquelles on a calculé la force de différentes associations statistiques *dans des conditions contrôlées*, la valeur d'autres variables ou d'autres facteurs étant fixée.

Dans le cas du différentiel social, les hypothèses empiriques classiques peuvent être celles d'un lien de causalité entre le niveau socio-économique des ménages (*A*) et diverses dimensions de la réussite scolaire ou académique des enfants (*D*). Par exemple, *ceteris paribus*, plus ce niveau est élevé, plus le niveau scolaire atteint est élevé également (BOURDIEU et PASSERON, 1970) et le temps passé dans les études supérieures est long (BOUDON, 1973). Il existe également un lien très fort entre le plus haut diplôme obtenu par les parents d'un ménage (*A*) et celui obtenu par les enfants (*D*) (MANZO, 2013a). Ou encore, on observait, dans les années 1990, que bien que le coût moyen de l'éducation n'avait fait que décroître depuis les années 1950, menant à ce que de plus en plus d'enfants aient accès à l'école (toutes catégories sociales confondues), les odd-ratios mesurant les inégalités entre les classes sociales quant à l'accès à l'éducation restaient plus ou moins constants (BREEN et GOLDTHORPE, 1997). En d'autres termes, le nombre de personnes ayant accès à des plus haut niveaux d'étude augmentait dans le temps dans chaque catégorie sociale, mais les catégories sociales supérieures restaient sur-représentées au sein des personnes ayant effectivement accès à ces niveaux d'études.

$II : A \longrightarrow B$ représente l'influence (causale) de l'état macro *A* sur l'état *micro* *B*. Ce dernier peut être modélisé comme l'ensemble restreint de choix qui se présentent à un agent donné, et cet ensemble de possibilités est, d'une façon explicitée dans le mécanisme en jeu, conditionné

par certains faits sociaux au niveau macro (A).

Dans notre exemple d'illustration, B pourrait représenter par exemple l'ensemble des choix possibles pour un ménage et ses enfants vis-à-vis de l'institution scolaire (rester dans le système scolaire ou le quitter pour une voie professionnelle), et les coûts et bénéfices perçus de ces choix. Un modèle simple suppose que le bénéfice lié au fait de rester dans le système scolaire plutôt que de le quitter dépend du bénéfice attendu ou espéré de la situation professionnelle que l'élève pourra atteindre en faisant cela, pondéré par la probabilité (perçue) de succès (BOUDON, 1973; BREEN et GOLDTHORPE, 1997). D'autre part, le coût de rester dans le système scolaire plutôt que de quitter est un coût principalement matériel (fournitures scolaires, logement, ...) La flèche $A \rightarrow B$ représente la façon avec laquelle ces bénéfices et coûts attendus et leur perception sont influencés par la catégorie socio-économique de laquelle provient l'agent (l'élève, le ménage, etc) en question, et peut ainsi modéliser facilement un élément essentiel des explications sociologiques, à savoir l'accès différencié aux ressources (économiques, sociales, culturelles, etc.)

III : $B \rightarrow C$ représente comment un certain état micro va influencer un autre état micro. Par exemple, comment la limitation de l'ensemble des possibilités d'actions ainsi que les coûts et bénéfices perçus associés à chacune d'entre elles vont se traduire dans le choix effectif d'une action par l'agent. Il s'agit donc ici de la *théorie de l'action* à l'œuvre. On peut aussi imaginer ici des cas où plusieurs agents interagissent, et donc intègrent dans leurs actions les actions (attendues ou effectives) des autres agents : on a alors une sorte de retro-action représentée par la flèche $C \rightarrow B$.

On peut imaginer un modèle de choix rationnel pour lequel les agents font *l'action qui maximise l'utilité espérée* qu'ils attribuent à chaque action possible (rester dans le système scolaire ou le quitter) (BOUDON, 1973). On peut aussi imaginer un modèle d'aversion au risque relatif, où les agents cherchent à simplement éviter de se retrouver dans une catégorie sociale plus "faible" de laquelle ils sont partis, en supposant que les familles "seek to ensure [...] that their children acquire a class position at least as advantageous as that from which they originate" (BREEN et GOLDTHORPE, 1997, p. 283). Des mécanismes additionnels peuvent bien sûr être implémentés, comme l'influence

éventuelle des pairs dans le choix d'une trajectoire scolaire ou académique comme dans (MANZO, 2013a), incarnant l'effet retro-actif d'une interaction sociale.

$IV : C \longrightarrow D$ représente finalement la façon avec laquelle les actions des agents au niveau micro s'agrègent les unes avec les autres et finissent par produire l'état macro D .

L'agrégation est produite en modélisant effectivement les actions supposées des individus sous les contraintes données par B tel que les exigences de la théorie de l'action $B \longrightarrow C$ soient respectées. On peut imaginer, par exemple, résoudre le modèle, si possible, mathématiquement, ou bien via une simulation numérique. L'état D est donné, par exemple, par les résultats scolaires effectifs des différents agents, ou bien par le diplôme le plus élevé obtenu en moyenne par les différents groupes étudiés.

Le diagramme de Coleman peut être considéré comme un (méta-)modèle, un patron épistémologique de ce à quoi devraient ressembler idéalement la mise en évidence des phénomènes sociaux et leur explication.

On peut également le trouver sous la forme d'une équation, chez Boudon par exemple (BOUDON, 1986, p. 30) :

$$M = M(m[S(M')]). \quad (5.1)$$

L'équation 5.1 condense l'idée que l'état macro M doit être vu comme une fonction des actions individuelles m , dépendantes elles-mêmes des situations S dans lesquels les acteurs étudiés prennent leurs décisions, situations qui dépendent elles-mêmes d'autres états macro M' .

Ainsi, il est bon d'écarter tout de suite la caricature qui consisterait à réduire l'individualisme méthodologique, dans cette version *structurale*, à un *atomisme*, où l'on tenterait de reconstruire les phénomènes macro comme agrégations d'actions d'individus *atomisés*, c'est-à-dire sans interaction les uns avec les autres et sans être plongés dans certaines structures sociales. Ici, au contraire, les effets éventuels des structures sociales ainsi que des interactions inter-agents sont intégrés dès le début dans le mode d'explication sociologique proposé.

Le diagramme de Coleman est également un bon outil pédagogique pour introduire à cette façon particulière de faire de la sociologie. Petri Ylikoski (YLIKOSKI, 2021), par exemple, l'introduit en expliquant que ce que l'on cherche à expliquer, c'est comment le fait de modifier une variable macro (A) en

modifie une autre (D), en introduisant pour ça des variables micro intermédiaires (B et C), qui sont explicatives en tant qu'elles s'articulent au sein d'un (ou de plusieurs) mécanisme(s) dont l'agrégation redonne la variation de D .

Ce schéma général peut également s'appliquer à la description d'un phénomène social dynamique, pour lequel les macro états A et D sont respectivement $M(t)$ et $M(t + 1)$, et où l'on cherche à expliquer l'état macro au temps $t + 1$ à l'aide de l'état macro au temps t .

Les *modèles de seuils* développés entre autre par Mark Granovetter (GRANOVETTER, 1978) pour modéliser certains phénomènes collectifs constituent un exemple parfait d'illustration. Dans de tels phénomènes, on considère un groupe d'individus pouvant prendre part à une action collective a . Il peut s'agir d'une émeute, d'une grève, de l'adoption d'une innovation, de la diffusion d'une rumeur, en tout cas une situation où l'on considère que le choix est binaire : y participer ou ne pas y participer.

Dans la version la plus simple de ces modèles de seuils, on considère que la condition pour que l'agent i participe, au temps t , à l'action collective est que le nombre d'autres agents qui y participent déjà soit supérieur à un certain *seuil* s_i , qui n'est pas forcément le même d'un agent à l'autre. Ce type de modèle est notamment utilisé pour montrer qu'une légère variation dans la distribution des seuils $\{s_i\}_i$ au sein de la population en question peut provoquer un changement macro conséquent.

Un exemple typique de la vie quotidienne est celui de la traversée d'un passage piéton "illégalement", c'est-à-dire lorsque le feu est rouge pour les piétons. Pour que ce phénomène puisse s'observer, il faut qu'il y ait initialement ($t = 0$), parmi les personnes qui attendent sur le trottoir, des personnes ayant un seuil égal à zéro, c'est-à-dire qui peuvent traverser même si elles se retrouvaient seules au milieu de la route.

Ces personnes traversent alors, admettons qu'il n'y en ait qu'une seule ($N(t = 1) = 1$). Ensuite, les personnes dont le seuil est égal à un, c'est-à-dire qui ont besoin, pour traverser, qu'il y ait au moins une personne pour traverser, traversent, mettons qu'il y en ait trois. Il y a maintenant $N(t = 2) = 4$ personnes en train de traverser, ce qui va permettre aux personnes ayant un seuil de 4 de traverser à leur tour, et ainsi de suite.

Ainsi, on peut observer que petit à petit une foule entière traverse alors que le feu est au rouge. Par contre, il aurait suffi d'un changement très léger, par exemple simplement qu'il n'y ait personne avec un seuil en dessous de quatre, pour que l'on observe une seule personne courir toute seule au milieu de la route au lieu de voir une foule entière traverser.

On peut modéliser mathématiquement ce phénomène à l'aide de la fonction "marche" que je note σ , qui est égale à zéro pour les nombres négatifs et égale à 1 pour les nombres positifs :

$$\sigma(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} \quad (5.2)$$

On peut alors écrire que le nombre de personnes engagées dans l'action au temps $t + 1$ est égal au nombre de personnes dont le seuil est supérieur au nombre de personnes engagées au temps t , soit :

$$N(t + 1) = \sum_i \sigma(s_i - N(t)), \quad (5.3)$$

où la somme \sum_i est effectuée sur tous les agents en présence. Par construction, (5.2) indique que pour tout i , $\sigma(s_i - N(t)) = 1$ si $s_i \geq N(t)$ et $\sigma(s_i - N(t)) = 0$ sinon.

L'équation (5.3) est ainsi une instanciation particulière de l'équation de Boudon (5.1) décrivant l'individualisme structural. On peut également la décrire à l'aide du diagramme de Coleman, comme représenté figure 5.2.

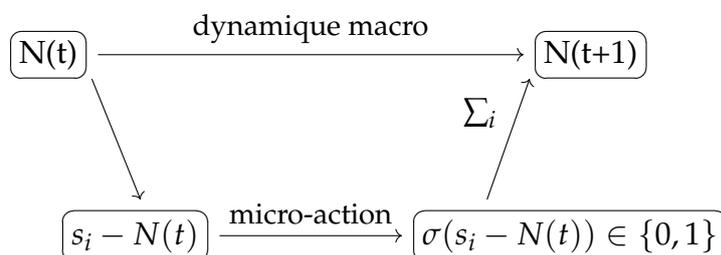


FIGURE 5.2 – Représentation diagrammatique des modèles de seuil granovetteriens.

Si les principes énoncés dans (HEDSTRÖM et SWEDBERG, 1998) et présentés ci-dessus, ainsi que le diagramme de Coleman, forment des éléments du noyau dur du programme de recherche analytique, d'autres tentatives d'explicitation programmatique se sont progressivement rajoutées, comme on va le voir dans la section suivante.

Une explicitation programmatique progressive

À la fin des années 2000, Peter Hedström et Peter Bearman ont poursuivi l'explicitation d'un programme de recherche clair pour la sociologie analytique. Ils ont notamment proposé, sous la forme d'une méthode générale,

une "stratégie explicative" à suivre pour aborder un phénomène social dans la tradition analytique (HEDSTRÖM et BEARMAN, 2009, p. 16) :

- (1) Commencer par délimiter clairement le fait social que l'on cherche à expliquer.
- (2) Formuler différentes hypothèses à propos des mécanismes, au niveau micro, qui pourraient être pertinents.
- (3) Traduire ces hypothèses en modèles computationnels.
- (4) Simuler le résultat du modèle pour dériver le type de fait social que chaque mécanisme micro produit.
- (5) Comparer les faits sociaux générés par chaque modèle avec les faits sociaux observés initialement.

Cette méthode a été, par exemple, suivie telle quelle par Gianluca Manzo dans son travail sur les choix éducatifs (MANZO, 2013a), qui se place dans la lignée des travaux de Boudon (BOUDON, 1973), ou encore de Breen et Goldthorpe (BREEN et GOLDTHORPE, 1997), en testant en plus l'effet des interactions sociales dans le différentiel éducatif observé.

Plus précisément, le fait social à expliquer consiste en l'observation de fortes associations statistiques entre le niveau éducatif le plus élevé atteint par des individus (dans un échantillon tiré d'une étude de l'INSEE du début des années 2000) et celui atteint par leurs parents, représenté classiquement par des *odds-ratios* généralisés significativement différents de 1.

Différentes hypothèses pour expliquer ce fait bien précis sont ensuite formulées : une différence de compétences (*ability*), un calcul coût/bénéfice différent, des perceptions subjectives de la probabilité de succès différentes, et l'effet des influences sociales exercées par les pairs.

Ces différentes hypothèses sont testées les unes après les autres dans des modèles computationnels (des simulations numériques, dans le style ABM) de plus en plus complexes. Plus précisément, un tel protocole permet de tester, en premier, l'influence de la différence de compétences seule, puis de rajouter à ce modèle l'influence d'un calcul rationnel différencié, et ainsi de suite.

La dernière étape consiste alors à comparer les résultats produits par ces modèles de degré croissant de complexité aux faits empiriques que l'on veut reproduire, pour pouvoir juger éventuellement de quelles hypothèses sont nécessaires pour expliquer les observations. Le point défendu par Manzo dans ce travail particulier est qu'un différentiel éducatif aussi élevé ne peut pas s'expliquer sans prendre en compte l'effet d'amplification induit par l'influence sociale exercée par les pairs.

Gianluca Manzo a lui-même poursuivi cet effort d'explicitation dans le premier chapitre du Handbook de 2014, où il éclaire plusieurs points de méthode qui font écho à ceux déjà cités (MANZO, 2014b, p. 28-54) : la clarté, la description, les modèles génératifs, l'individualisme méthodologique structural, les logiques de l'action, l'interdépendance structurelle, la modélisation basée sur les agents (ABM) et enfin le fait que les données empiriques restent le dernier juge de la pertinence de l'explication théorique. Cela lui permet également de répondre à un certain nombre de critiques qui ont pu émergé tout au long des années 2000.

On peut enfin citer l'article de Peter Hedström : *Past and future of analytical sociology* paru dans le dernier *Research Handbook on Analytical Sociology*, (HEDSTRÖM, 2021), ainsi que la conférence du même nom (HEDSTROM, 2021), qui dresse un portrait actuel du champ, ainsi que des idées de développement futurs.

Hedström revient notamment sur le statut des théories de l'action, qui forment la pierre de voute des mécanismes mobilisés dans les explications en sociologie analytique. En particulier, il prend une position plus radicale qu'à l'accoutumée en enjoignant non plus uniquement à un certain "réalisme" ou plausibilité des hypothèses (HEDSTROM, 2005) mais à fonder entièrement les mécanismes mobilisés sur les données empiriques.

En d'autres termes, il joint à ne plus se focaliser sur les intentions ou les états mentaux des agents dans les explications des phénomènes sociaux – à part si ceux-ci sont directement observables, ce qui n'est presque jamais le cas – et à leur préférer des explications purement causales, ne reposant que sur des données disponibles. Le but est, entre autre, d'éviter de se perdre dans des "just-so stories". On aura l'occasion de revenir sur ce point.

Quelques critiques basiques

Il me semble important de revenir sur certaines critiques qui peuvent selon moi être facilement écartées en ce qu'elles relèvent davantage de malentendus que de réelles difficultés théoriques ou conceptuelles.

La première est l'accusation éventuelle d'*atomisme* qui peut être fait à l'encontre du paradigme individualiste, qui reviendrait à tenter d'expliquer les phénomènes sociaux par l'action d'individus qui n'interagiraient pas entre eux ni ne seraient plongés dans certaines structures sociales. La forme particulière d'individualisme qui est convoquée en sociologie analytique n'est pas un atomisme *par construction* : si les phénomènes sociaux sont reconstruits comme agrégation d'actions individuelles, celles-ci ne supposent en

aucun cas que les acteurs sont libres de toute influence sociale, fût-elle inter-individuelle ou structurelle. Ce fait est modélisé en particulier par la relation macro-micro (flèche II) du diagramme de Coleman.

Un autre type de critique dirigée contre l'individualisme méthodologique énonce qu'un tel paradigme est insuffisant pour traiter de phénomènes qui ne seraient pas réductibles aux actions individuelles, pour lesquels un paradigme *holiste* serait alors préférable. Pourtant, le fait de voir les phénomènes sociaux comme le produit de l'agrégation d'actions individuelles n'empêche pas (au contraire, devrait-on dire) de traiter des phénomènes *émergents*, c'est-à-dire de traiter de l'existence de certaines propriétés au niveau macroscopique qui ne correspondent pas aux propriétés micro mais qui émergent, de manière hautement non triviale, de la composition des actions micro.

Le paradoxe de Condorcet (CONDORCET, 1785) est un exemple classique d'un tel phénomène, qui peut survenir lors d'un vote par suffrage universel. Imaginons que chaque individu doive choisir entre trois candidats possibles, mettons *A*, *B* et *C*, et que les préférences des individus soient transitives, c'est-à-dire telles que si *A* est préférée à *B* et *B* à *C*, alors on doit avoir que *A* est préférée à *C*. Alors, dans certains cas bien choisis mais qui surviennent fréquemment dans des élections réelles, il se peut que les préférences *agrégées*, elles, ne soient plus transitives. En d'autres termes, il se peut qu'une majorité de personnes préfère *A* à *B*, une majorité de personnes préfère *B* à *C* et une majorité de personnes préfère *C* à *A*.

Dans cet exemple, on voit donc qu'une propriété micro, la transitivité des préférences, disparaît à l'échelle macro. Ainsi, même si le niveau macro est conçu comme une agrégation d'actions individuelles, cela ne signifie pas que les propriétés du niveau macro se réduisent simplement à celles du niveau micro.

De la même façon, dans des modèles classiques d'actions collectives (OLSON, 1971), si on considère des agents rationnels qui : *ont intérêt* à faire une action *a* et *sont en mesure* de faire l'action *a*, alors individuellement, on s'attend à ce que n'importe quel individu, seul, fasse effectivement l'action *a*. Par contre, et cela correspond à des observations contrôlées comme quotidienne, dès l'instant où il y a plusieurs individus ensemble, alors le groupe n'agit plus du tout comme un agent rationnel : chaque individu, dans le choix de faire l'action *a*, va considérer qu'il est encore plus profitable pour lui d'attendre que d'autres fassent l'action, pour pouvoir profiter des bénéfices sans avoir à en payer les coûts. Le résultat attendu, au niveau macro, est que très

peu d'agents, voire aucun, ne fasse effectivement l'action – ce qui va à l'encontre direct de leur intérêt.

Comme dans l'exemple précédent, la mise en interaction de plusieurs agents rationnels fait émerger une certaine irrationalité qui n'est pas trivialement déductible des propriétés individuelles. Encore une fois, envisager les phénomènes macro comme agrégation d'actions micro n'implique pas une simple "somme" des actions individuelles, ni même une réductibilité triviale des propriétés d'un niveau à l'autre.

Enfin, pour revenir aux structures sociales, le paradigme individualisme n'implique pas de considérer des individus jouissant d'une distinction ontologique particulière. Dans certains cas, ils ne sont que le substrat des causes sociales qui agissent réellement à travers le réseau social, par exemple, dans lequel ils sont plongés.

Prenons par exemple le cas de la relation entre la centralité des universités médiévales et leur taux de cosmopolitisme (voir section 3.2.3). Cette relation, observée au niveau macro, s'explique en considérant les échanges d'information dans le réseau qu'elles constituent. Or, c'est bien ce réseau, et surtout sa structure particulière, qui forme la base de l'explication causale de la relation macro.

Les universités particulières dont il est question ne sont en fait pas considérées en elles-mêmes, mais simplement comme des nœuds de ce réseau. En d'autres termes, elles constituent des substrats concrets sur lequel agissent les "forces" sociales, ce qui permet de dépasser la simple métaphore (comme celle des "courants suicidogènes" de Durkheim) et de les incarner concrètement au sein de modèles bien définis.

Ainsi, le paradigme individualiste structural ne s'oppose pas, au contraire, à l'étude de phénomènes émergents, et peut même tout à fait concrétiser un certain type d'explication a priori considérée comme *holiste*.

5.2.3 La théorie du choix rationnel

La théorie de l'action que vais analyser épistémologiquement dans le reste de ce chapitre est la théorie du choix rationnel (notée TCR à partir de maintenant), aussi appelée *théorie de l'utilité espérée*. Ce choix particulier se justifie par le fait que cette théorie jouit du développement le plus ramifié et a été la plus discutée dans la littérature, en particulier dans le cadre de la sociologie

analytique. Cette partie est consacrée plus particulièrement à une présentation générale de la TCR, avant d’aborder ses aspects épistémologiques dans les parties suivantes.

Bien que cette théorie provienne en premier lieu de l’économie, elle est appliquée en sociologie depuis au moins les années soixante, sur des sujets aussi divers que la criminalité (BECKER, 1968), l’action collective (OLSON, 1971) ou la réussite scolaire (BOUDON, 1973). L’ouvrage déjà cité de James Coleman (COLEMAN, 1990) est une tentative majeure de synthèse et d’unification de l’application de la théorie du choix rationnel à tout un ensemble de phénomènes sociaux.

Dans le diagramme de Coleman représenté figure 5.1, une telle théorie intervient au niveau micro–micro (flèche III), pour modéliser comment un agent choisit entre différentes actions possibles qui s’offrent à lui, en fonction du contexte dans lequel il est plongé et des ressources auxquelles il a accès.

Pour illustrer la TCR, je vais prendre comme exemple introductif le modèle de choix éducatifs développé par Breen et Goldthorpe (BREEN et GOLDTHORPE, 1997), tel que reformulé par (TUTIĆ, 2017). Ces approches sont directement héritées du travail séminal de Raymond Boudon (BOUDON, 1973), appliquant la TCR au système éducatif et à la reproduction sociale, que j’ai déjà évoqué au chapitre 3.

Dans le modèle de Breen et Goldthorpe (noté BG à partir de maintenant), les agents sont supposés pouvoir provenir de trois classes sociales distinctes : la *service class* (S), la *working class* (W) et la *under class* (U). L’objectif premier de ce modèle est d’expliquer une observation empirique à leur disposition dans les années 1990 : le fait que bien que le coût de l’éducation n’ait cessé de décroître avec le temps, menant à l’augmentation de la proportion d’enfants persistant dans le système scolaire (que nous notons symboliquement $P(\textit{stay}|x)$, pour chaque classe sociale $x \in \{S, W, U\}$), le différentiel éducatif restait néanmoins élevé. Un tel différentiel est typiquement mesuré par des odd–ratios comme :

$$\frac{P(\textit{stay}|S)}{1 - P(\textit{stay}|S)} \bigg/ \frac{P(\textit{stay}|W)}{1 - P(\textit{stay}|W)}. \quad (5.4)$$

En d’autres termes, les odd–ratios (5.4) restaient donc élevés devant 1.

Au niveau micro, qui nous intéresse particulièrement ici, les agents sont supposés faire face à un ensemble d’actions \mathcal{A} possibles, que l’on peut également trouver sous le nom de choix, de préférences, de loteries, etc. selon le

contexte scientifique (sociologie, économie, théorie de la décision, ...) dans lequel on se trouve, menant à un ensemble donné de résultats Ω avec certaines probabilités.

En d'autres termes, $\omega \in \Omega$ est un résultat possible d'une action $a \in \mathcal{A}$ qui a une probabilité $p_a(\omega) \in [0, 1]$ de survenir en faisant l'action a .

Par exemple, le modèle BG se focalise sur le choix des agents des classes S et W . Ils sont supposés avoir deux actions possibles, soit rester dans le système scolaire (*stay*), soit le quitter (*leave*) : $\mathcal{A} = \{\text{stay}, \text{leave}\}$. Les résultats possibles de chaque action sont les différentes classes sociales dans lesquelles les étudiants peuvent terminer en faisant l'une ou l'autre action, soit $\Omega = \{U, S, W\}$. Il est, de plus, supposé que s'ils restent dans le système scolaire, ils peuvent *réussir* avec une probabilité de π ou *échouer* avec une probabilité $1 - \pi$.

Les résultats possibles de ce choix binaire et leurs probabilités associées sont représentés dans un arbre de choix figure 5.3.

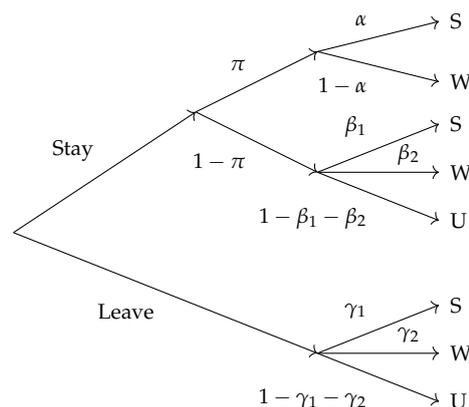


FIGURE 5.3 – Schéma des trajectoires (probabilistes) possibles dans le modèle BG.

Par exemple, $(1 - \pi)\beta_1$ est donc la probabilité de terminer dans la classe sociale S en ayant échoué après être resté dans le système scolaire, etc. De plus, les paramètres $\alpha, \beta_1, \beta_2, \gamma_1$, and γ_2 sont supposés satisfaire certaines relations censées refléter à quel point il est probable ou non de terminer dans telle ou telle classe sociale en suivant telle ou telle trajectoire (voir (BREEN et GOLDTHORPE, 1997, p. 282) pour plus de détails). Ces probabilités et l'arbre de choix fig. 5.3 dans son ensemble représentent ainsi les aspects *structurels* du problème.

Le cœur de la TCR est l'hypothèse que chaque agent attache une certaine *utilité* $u(\omega) \in \mathbb{R}$ à chaque résultat possible ω , et que le choix effectif de l'agent

dépend uniquement de l'utilité espérée (*expected utility*) rattachée à chaque action $a \in \mathcal{A}$:

$$\mathcal{U}(a) = \sum_{\omega \in \Omega} p_a(\omega)u(\omega). \quad (5.5)$$

Par exemple, comme cela est souvent postulé, on modélise les agents comme des maximisateurs d'utilité, choisissant l'action $a \in \mathcal{A}$ telle que :

$$\mathcal{U}(a) = \max_{a' \in \mathcal{A}} \mathcal{U}(a'). \quad (5.6)$$

Modéliser les choix éducatifs dans le modèle BG nécessite donc d'associer à chaque action (*stay* et *leave*) une certaine utilité espérée, qui dépend à la fois de l'utilité particulière attachée par l'agent au fait d'atteindre les différentes classes U , W et S et des différentes probabilités structurant l'arbre de choix ci-dessus.

Comme remarqué dans (TUTIĆ, 2017, p. 402), dans le modèle BG les utilités attachées aux différents résultats sont implicitement choisies comme suit :

$$u_S(S) = u_W(S) = 1 \text{ et } u_S(U) = u_W(U) = 0, \quad (5.7)$$

$$u_S(W) = 0 \text{ et } u_W(W) = 1, \quad (5.8)$$

où $u_x(y)$ est l'utilité attachée par un agent provenant de la classe sociale $x \in \{S, W\}$ au fait d'atteindre la classe sociale $y \in \{S, W, U\}$. L'hypothèse (5.8) est une façon d'implémenter l'*aversion au risque relatif*, c'est-à-dire le fait que les agents préfèrent éviter la mobilité sociale descendante, ce qui est justement le mécanisme central du modèle BG.

L'utilité espérée \mathcal{U}_S (\mathcal{U}_W) pour des agents provenant de la classe sociale S (W) s'écrit donc :

$$\begin{aligned} \mathcal{U}_S(\textit{stay}) &= \pi\alpha + (1 - \pi)\beta_1, \quad \mathcal{U}_S(\textit{leave}) = \gamma_1 \\ \mathcal{U}_W(\textit{stay}) &= \pi + (1 - \pi)(\beta_1 + \beta_2), \quad \mathcal{U}_W(\textit{leave}) = \gamma_1 + \gamma_2 \end{aligned}$$

Partant du principe de maximisation de l'utilité espérée, un agent provenant de la classe S reste dans le système éducatif si :

$$\pi\alpha + (1 - \pi)\beta_1 > \gamma_1, \quad (5.9)$$

et un agent de la classe W reste si :

$$\pi + (1 - \pi)(\beta_1 + \beta_2) > \gamma_1 + \gamma_2, \quad (5.10)$$

et ainsi de suite.²

Le modèle BG est aussi une illustration de la manière avec laquelle les structures sociales peuvent être facilement implémentées dans ce genre de modèles, comme on l'a déjà suggéré plus haut. Ici, c'est l'effet structurel (la flèche *II*) qui fait le travail : l'utilité attachée à chaque résultat possible (et donc, par ricochet, à chaque action possible) diffère pour des agents provenant de classes sociales différentes, prenant ainsi en considération le mécanisme central de ce modèle, à savoir l'aversion au risque relatif.

Plus généralement, l'effet des structures sociales (par exemple, les accès différenciés aux ressources économiques, culturelles ou sociales) dans lesquelles les agents sont plongés peut être facilement modélisé dans ce cadre théorique, en étant encodé dans l'utilité espérée attachée à chaque action possible qui peut différer selon l'état macro (la classe sociale d'origine par exemple) en considération. L'interaction sociale peut également être implémentée de la même manière, par exemple via des modèles de théorie des jeux dans lesquels l'utilité espérée associée au choix d'un agent dépend des actions éventuelles des autres agents.

Quoiqu'il en soit, l'utilisation d'une telle théorie au sein de la sociologie analytique pose des questions épistémologiques importantes. Par exemple, quels sont les critères de sélection des modèles d'action micro ? Plus précisément, dans quelles conditions est-il légitime d'utiliser des modèles reposant sur des hypothèses hautement idéalisées à propos du comportement humain afin d'expliquer les phénomènes sociaux ? C'est à cette discussion, dont la TCR n'est qu'un prétexte, que va être consacrée la partie suivante de ce chapitre.

5.3 Épistémologie des théories de l'action

5.3.1 Entre irréalisme et trivialité

Les critiques de la TCR en général et appliquée aux sciences sociales en particulier sont structurées le long d'une tension entre deux situations également insatisfaisantes : l'*irréalisme* de ses hypothèses de base et la *trivialité* de ses explications.

2. Voir (TUTIĆ, 2017; BECKER, 2022) pour un état de l'art relativement récent sur ce type de modèle.

D'une part, et en dépit d'un certain succès empirique, les hypothèses de la TCR sont souvent vues comme "irréalistes" vis-à-vis de ce qui est par ailleurs connu à propos du comportement humain. Comme le formule Margaret Marini :

The most obvious problem with the axiomatization of utility and probability as a theory of the way people behave in choice situations is that it assumes that people have a high level of knowledge and computational ability with which to determine and evaluate a set of available alternatives. It assumes knowledge of all the alternatives available, as well as the consequences that will follow from each of the alternatives. (COLEMAN et FARRARO, 1992, pp. 24-25)

Des champs de recherche entiers, comme l'économie comportementale (PAGE, 2022), se sont même constitués sur l'accumulation d'observations empiriques (DELLAVIGNA, 2009) pour lesquelles les axiomes de la TCR semblaient systématiquement contredits.

Il est important de noter que le "réalisme" des hypothèses qui est au centre de la discussion ici ne doit pas être confondu avec le réalisme dans le sens du *réalisme scientifique*. Ce dernier terme réfère à des questionnements portant sur l'existence ou non d'une réalité objective et indépendante (réalisme métaphysique), sur la question de savoir si le contenu d'une théorie, notamment les entités inobservables qu'elle suppose, doit être interprété littéralement (réalisme sémantique), ou bien si on peut déduire du succès empirique de nos meilleures théories que les entités inobservables qu'elle postule existent, d'une façon ou d'une autre (réalisme épistémique) (PSILLOS, 1999).

Bien que ces problèmes puissent être reliés, lorsqu'il est question du "réalisme" des hypothèses de base des modèles théoriques dans la littérature, cela réfère le plus souvent à leur "consistance" ou leur "cohérence" avec ce qui est connu par ailleurs (à propos du comportement humain, par exemple). Par exemple, comme je l'ai déjà mentionné, Hedström insistait dans les années 2000 sur la "plausibilité psychologique" (HEDSTROM, 2005, p. 35) des hypothèses de base comme un critère nécessaire (ou en tout à contenter autant que faire se peut) pour qu'un modèle théorique soit acceptable en tant qu'explication authentique d'un phénomène social.

Si l'utilisation extensive d'hypothèses qui semblent être "irréalistes", au sens où on l'a précisé ci-dessus, constitue un constat bien partagé, son aspect

problématique ne fait pas l'unanimité. On peut en effet trouver dans la littérature un ensemble d'arguments soutenant que ce type d'hypothèses est tout à fait recevable :

EXPLIQUER BEAUCOUP AVEC PEU : on peut défendre l'idée, comme (FRIEDMAN, 1953 ; JASSO, 1988), que puisque *expliquer* implique de dériver un grand nombre de faits et de phénomènes avec aussi peu d'hypothèses que possible, les meilleures hypothèses sont forcément "fausses" en ce qu'elles simplifient à outrance, par construction, ce qu'il se passe en réalité. C'est même l'argument de base de la vision "as-if" des explications en économie, que l'on doit en particulier à Milton Friedman.

DE-IDÉALISATION : la méthode de la dé-idéalisation ou de "l'abstraction décroissante", comme décrite par exemple dans (COLEMAN et FARRARO, 1992, Chap. 2) ou défendue plus récemment dans (PERUZZI et CEVOLANI, 2021), considère les hypothèses idéalisées comme des points de départ méthodologiques pour le développement de modèles qui gagnent éventuellement en complexité à mesure qu'on les confronte à davantage de faits empiriques. Ils forment alors, en économie ou plus généralement en sciences sociales, des modèles de plus en plus "réalistes" en ce qu'ils intègrent de plus en plus d'aspects de la réalité, tout en augmentant dans le même temps leur succès empirique.

ÉCHELLE D'ANALYSE : on peut considérer, comme (GOLDTHORPE, 1998 ; HECHTER et KANAZAWA, 2019), que l'objectif premier d'une science *sociale* est d'expliquer les phénomènes *macro* et non pas les comportements ou les actions individuels :

[There is a] common misconception about the nature of rational choice. The theory does not aim to explain what a rational person will do in a particular situation. That question lies firmly in the domain of decision theory. Genuine rational choice theories, by contrast, are concerned exclusively with social rather than individual outcomes. (HECHTER et KANAZAWA, 2019, p. 3)

Bien entendu, même dans ces cas-là, la TCR est utilisée pour modéliser le comportement des individus, mais comme un moyen de reproduire les phénomènes sociaux, et non pas comme une fin en soi – ce qui peut être le cas, par contre, en théorie de la décision.

LA LOI DES GRANDS NOMBRES : cet argument est relié au précédent. Il soutient que la TCR n'affirme *pas* que "all actors at all times act in an entirely rational way" mais seulement que les déviations éventuelles à

la rationalité se compensent et s'annulent dans l'agrégation des comportements individuels (flèche IV fig. 5.1), c'est-à-dire que "the law of large numbers' will (...) ensure that it is the rational tendency that dominates." (GOLDTHORPE, 1998, pp. 168-169). Ainsi, il est simplement équivalent, et donc plus commode, de supposer dès le début que tous les agents sont rationnels.

LA MEILLEURE THÉORIE JUSQU'ICI : comme l'argumente (OPP, 2013), bien que la TCR ait certains défauts ou limitations, elle reste le meilleur point de départ à disposition des sociologues pour la formulation de modèles rigoureux et féconds des phénomènes sociaux.

TOUT VA BIEN : d'autres auteurs, enfin, comme par exemple (FRIEDMAN, 1953) en économie, défendent qu'il n'y a aucun problème à utiliser des modèles "as-ifs", voire même (MOSCATI, 2023) que tous les modèles (mêmes les plus élaborés) restent de toute façon "as-if", et qu'il faut simplement comprendre le succès empirique de nos meilleures théories dans un paradigme philosophique adapté, à savoir un paradigme *anti-réaliste*.

Je reviendrai, dans ce chapitre, sur la possibilité d'utiliser des hypothèse "irréalistes" dans des modèles théoriques à visée explicative, en mobilisant, entre autres, certains des arguments cités ci-dessus.

D'autre part, il est possible de rester dans le cadre de la TCR et de complexifier les hypothèses de base pour les rendre plus "réalistes", par exemple dans la lignée des approches de type rationalité limitée (*bounded rationality*) du comportement humain (SIMON, 1957; WHEELER, 2020).

Il s'agit ici d'élargir la définition de rationalité pour rendre compte des déviations observées vis-à-vis d'une conception plus restreinte. Une possibilité est par exemple de remplacer le principe de maximisation stricte (5.6) par un principe *probabiliste* comme décrit par exemple dans (KRUIS et al., 2020), utilisé par exemple dans le modèle numérique de choix éducatif développé par Manzo (MANZO, 2013a, p. 58), ou encore dans la lignée de ce qui est connu en théorie des jeux sous le nom de *quantal response equilibrium* (MCKELVEY et PALFREY, 1995) : chaque agent a une probabilité $P(a)$ de faire l'action $a \in \mathcal{A}$, avec :

$$P(a) = \frac{e^{\beta U(a)}}{\sum_{a' \in \mathcal{A}} e^{\beta U(a')}} \quad (5.11)$$

où $U(a)$ est toujours l'utilité espérée de l'action $a \in \mathcal{A}$, et $\beta > 0$ est un

paramètre supplémentaire du modèle, et dans le cas où \mathcal{A} est un ensemble fini – la généralisation étant immédiate.

L'aspect théorique intéressant est que (5.11) se réduit au principe stricte de maximisation (5.6) pour $\beta \rightarrow \infty$ et constitue donc une généralisation directe de ce dernier, et P est uniforme pour $\beta = 0$. Dans ce dernier cas, l'agent agit de manière purement "aléatoire", ayant une même probabilité de choisir n'importe quelle action.

Le paramètre β peut être vu comme représentant un certain niveau d'information dont disposerait l'agent : plus β est élevé, et plus l'agent a une information fiable sur sa propre situation, ce qui fait que son comportement se rapproche du comportement d'un agent strictement rationnel. Au contraire, pour β tendant vers zéro, c'est-à-dire pour un agent ayant peu d'information sur sa situation, celui-ci tend à agir "au hasard".

La théorie des perspectives (*prospect theory*) (TVERSKY et KAHNEMAN, 1992; BARBERIS, 2013) est un autre exemple de complexification de ces hypothèses de base. Dans sa forme la plus simple, elle consiste à jouer sur les probabilités p_a , pour chaque action $a \in \mathcal{A}$, définies sur l'ensemble des résultats possibles Ω , afin de tenir compte à la fois des probabilités subjectives des agents (non pas la probabilité objective que telle action mène à tel résultat, mais la probabilité perçues par l'agent que telle action mène à tel résultat), et du fait que l'utilité est toujours définie vis-à-vis d'une certaine référence.

Dans cette théorie, la fonction d'utilité a alors une forme en S, concave au dessus du point de référence et convexe en dessous, avec une forme plus prononcée au dessus qu'en dessous. Cette forme particulière a pour but de modéliser la dépendance des préférences individuelles à une certaine référence, les "rendements décroissants" (*diminishing returns*) et l'aversion à la perte, qui sont des aspects comportementaux observés en psychologie ou en économie comportementale (PAGE, 2022).

Ce ne sont que quelques possibilités parmi d'autres formes, possiblement infinies, que l'on peut donner à ces "risky curves" (FRIEDMAN, 2017). Il est également possible de prendre en compte la dimension temporelle des préférences (et plus particulièrement l'inconsistance temporelle) en rajoutant des corrections hyperboliques à la fonction d'utilité standard (GINTIS, 2009), ou en tenant compte des croyances subjectives à l'aide de la théorie bayésienne de la décision (BINMORE, 2011), le tout "without modifying any of [RCT's] basic conceptual elements" (MANZO, 2013b, p. 369).

Cette démarche vise à gagner en "réalisme" mais est également souvent critiquée, pour plusieurs raisons. Selon Ivan Moscati (MOSCATI, 2023), par

exemple, ces modèles soit-disant alternatifs à la "théorie néo-classique" sont tout aussi *as-if* que les autres, bien qu'ils soient peut-être légèrement plus élaborés. Ils tombent donc sous le coup du même type de critiques, bien qu'ils cherchaient initialement à s'en départir.

Cependant, la critique principale qui m'intéresse ici au plus haut point est que ce type d'ajustements théoriques peut rapidement rendre l'explication sous-jacente *triviale*. En effet, moins il y a de contraintes sur la façon avec laquelle le principe de rationalité s'instancie, plus il est facile de produire des modèles qui sont empiriquement adéquats – dans ce cas, par contre, l'explication sera vidée de tout contenu. Par exemple, le modèle de *quantal response equilibrium* en théorie des jeux, cité plus haut, a été critiqué précisément pour cette raison (HAILE, HORTAÇSU et KOSENOK, 2008).

Quoiqu'il en soit, ce cas d'une TCR trop "lâche" qui explique tout mais de manière triviale apparaît comme une instance particulière de ce que Popper qualifie de théorie infalsifiable, ou ce que Imre Lakatos appelle un *problem-shift dégénéré*.

Ce problème a été identifié à de nombreuses reprises en sciences sociales. Par exemple (KRONEBERG et KALTER, 2012, p. 83) :

The wide version of RCT is able to assimilate almost any psychological concept or theory and translate it into more or less "soft" incentives or a more or less inaccurate belief.

Ou encore, comme le notent Hedström et Ylikoski (MANZO, 2014a, p. 59-60) :

From the point of view of the core assumptions of RCT, broadening the range of individuals' concerns to non-monetary goods and to the welfare of others is straightforward, although there is concern about whether this can be done in a non-arbitrary manner. (...) Finding a RCT model that fits a particular phenomenon becomes almost trivially easy as there are no real constraints on preferences and beliefs that can be attributed to the individuals in question.

Plus récemment, (HEDSTRÖM, 2021, p. 498) porte des réflexions du même type :

What Becker and his followers showed is that it is possible to come up with coherent rational narratives that fit even the most puzzling and seemingly irrational kinds of behavior, particularly if the narratives are free to ignore known empirical facts about the behavior of individuals.

Ces quelques exemples illustrent ainsi précisément le problème épistémologique que j'ai mis en exergue plus haut.

Le même type de critiques s'applique aux modèles de comportement humain "basé sur les raisons", dans la tradition sociologique boudonienne de "rationalité généralisée" (BOUDON, 1996). Cette dernière est basée sur l'idée que les agents dont on doit expliquer les comportements doivent être supposés avoir des *bonnes raisons* d'agir comme ils le font, de telle sorte que "these reasons can be in some circumstances of the cost-benefit comparison type, but in other circumstances of other types" (BOUDON, 1996, p. 147). La critique centrale adressée au modèle "cognitivist" de Boudon est qu'il semble toujours possible, a posteriori, de trouver des raisons qu'un acteur aurait de se comporter tel qu'il se comporte, ce qui amenuise la portée explicative d'un tel cadre si aucune contrainte n'est ajoutée.

Ce débat a pris une forme particulière au sein de la sociologie analytique. Comme on l'a vu, les premières prémisses d'application de l'individualisme structural, issues de la *dutch sociology* (WIPPLER, 1978) ou de l'ouvrage de Coleman (COLEMAN, 1990), dont la sociologie analytique est directement héritière, sont basées sur la théorie du choix rationnel.

Pendant, la position au sein de ce courant concernant les théories de l'action et en particulier la TCR a évolué depuis les années 1990 "toward a more and more explicit pluralistic claim" (MANZO, 2021, p. 32). En d'autres termes, la TCR a progressivement cessé d'être considérée comme le seul (si non comme un bon ou même un fécond) point de départ pour la description des comportements micro.

Peter Hedström, par exemple, a développé une théorie "desires-beliefs-opportunities" (DBO) de manière à pallier certains des défauts qui semblaient mener à un instrumentalisme épistémologiquement inacceptable (HEDSTROM, 2005). De son point de vue de l'époque, les théories de l'action doivent être, entre autres choses, "psychologically plausible" (HEDSTROM, 2005, p. 35), dans le sens d'une certaine consistance avec ce qui est par ailleurs connu du comportement humain, par exemple en psychologie, pour que les modèles des phénomènes sociaux puissent trouver une assise épistémologique acceptable.

Reconnaissant bien sûr que tout modèle scientifique repose sur des hypothèses qui, d'une façon ou d'une autre, offrent une approximation de la réalité, il mettait en exergue une différence cruciale entre des approximations et

des hypothèses fausses. Il résumait ainsi son inquiétude à propos d'une "unfortunate instrumentalist tendency" : "Knowingly accepting false assumptions because they lead to better predictions or to more elegant models threatens the explanatory value and the long-term viability of the rational-choice approach." (HEDSTROM, 2005, p. 9).

Depuis lors, Hedström est devenu "considerably more skeptical" (HEDSTRÖM, 2021, Part. 4) sur la nécessité et la viabilité d'implémenter les intentions et les motivations individuelles dans les modèles de mécanismes micro, s'il n'est pas donné aux modélisateurs d'avoir un accès empirique à ces aspects individuels. En d'autres termes, une telle implémentation, pour produire une explication authentique, "has to be firmly anchored in known facts about the acting individuals and their social settings." (HEDSTRÖM, 2021, p. 498). Expliquer un fait macro signifie alors le réduire à des aspects proprement *observables* des comportements individuels, et non pas à des modèles abstraits à propos de ce qui est supposé se passer dans leur esprit – à moins que cette information ne soit disponible au sociologue. Il qualifie en particulier de "just-so stories" (HEDSTRÖM, 2021, p. 490) les modèles de comportements micros reposant sur des hypothèses inobservables à propos des états mentaux individuels, et encourage donc promptement à éviter ce type de modèles pseudo-explicatifs.

Additionnellement à la théorie DBO, d'autres théories de l'action, comme la théorie de la sélection de cadre (*selection frame theory*) de Kroneberg (MANZO, 2014a, Chapter 4) ou le modèle de choix basé sur les raisons (*reasons-based choice*) de Franz Dietrich et Christian List (DIETRICH et LIST, 2013 ; DIETRICH et LIST, 2016) se sont développées durant ces dix dernières années.

Il s'avère ainsi que la TCR – mais cela vaut généralement pour toutes les théories de l'action dans un contexte d'individualisme structural – se retrouve dans une position épistémologique critique entre des modèles "as-if" et des "just-so stories". Les premiers peuvent avoir un certain succès empirique mais reposent sur des hypothèses irréalistes. Quant aux secondes, elles reposent sur des hypothèses plus précises en termes descriptifs – soit en étant plus réalistes ou bien en étant inobservables – mais de telle manière à ce que l'explication qui en découle soit, *in fine*, triviale.

Ce dilemme a lui-même été identifié depuis un certain temps, par exemple par Neil J. Smelser (SMELSER, 1992) cité dans (HERNES, 1992, p. 421), qui résume parfaitement la situation :

the postulates of rational choice are not realistic, as the assumptions (...) leave out key aspects of human behavior (...) [and] the responses to incorporate criticism of rational choice theory have led to a watering down and a movement in the direction of theoretical indeterminacy – more abstract, more truistic, more context free, incapable of falsification – which “produces the specter of an inclusive and universally applicable construct that simultaneously explains everything and therefore nothing.”

Ce dilemme se reflète également dans l'opposition que l'on peut trouver parfois entre des versions "narrow" (étroites) et "wide" (larges) de la TCR, cette dernière affaiblissant les postulats de rationalité vis-à-vis de la première (OPP, 2013 ; MANZO, 2013b). On peut aussi penser à l'opposition entre des modèles "thin" (fin) et "thick" (épais) de choix rationnel : les premiers n'assument rien à propos des motivations individuelles alors que les seconds prennent en considérations des propriétés psychologiques individuelles plus complexes (HECHTER et KANAZAWA, 2019).

Face à ce dilemme, la question qui m'anime est proprement épistémologique : elle a trait aux conditions de justification et plus généralement de sélection des modèles explicatifs. Comment faire un choix non arbitraire entre différents cadres théoriques, ou plus précisément ici, différentes théories de l'action, pour expliquer des phénomènes sociaux donnés ? Le "réalisme" des hypothèses de base est-il une condition nécessaire à l'adoption d'un modèle explicatif ? Si ce n'est pas le cas, alors selon quels critères juger qu'un modèle est meilleur qu'un autre ?

Je vais plus précisément aborder ces questions dans la suite de ce chapitre.

5.3.2 Une première précision : de quelles hypothèses parle-t-on ?

Afin de préciser le problème, une première question importante à se poser quand on discute du "réalisme" des hypothèses est précisément : *quelles sont les hypothèses dont on exige le réalisme, et dans quel sens ?* C'est à répondre à cette question qu'est dédié cette section.

Comme je l'ai annoncé en introduction de ce chapitre, on se focalise ici sur les modèles théoriques, c'est-à-dire les modèles qui ont comme objectif l'explication de certains faits mis en évidence à travers des modèles empiriques, que l'on considère comme donnés ici.

Je considère, plus précisément, que les modèles théoriques en question sont composés de trois éléments distincts : un cadre théorique \mathcal{T} composé, entre autre, d'un certain nombre de *principes fondamentaux*, un nombre fini d'hypothèses théoriques H_1, H_2, \dots , que l'on note génériquement \mathcal{H} , qui ont pour fonction d'instancier les principes fondamentaux de T , et des conditions spécifiques (ou initiales, ou contextuelles) que je note Γ . Cette vision d'un modèle théorique comme un triplet (T, H, Γ) est également assez classique, dans la lignée de la vision dite *hiérarchique* des modèles.

La littérature sur l'explication scientifique, puisque c'est de ça dont il s'agit, est abondante et de nombreux modèles d'explications co-existent. Ici, j'utiliserai un motif très général d'explication qui, me semble-t-il, fait relativement consensus. On considère qu'un fait F (mis en évidence à travers un modèle empirique particulier appliqué à des données empiriques), auquel sont attachées certaines conditions spécifiques données Γ , est expliqué au sein d'un cadre théorique \mathcal{T} s'il existe un ensemble fini \mathcal{H} d'hypothèses théoriques tel que F se déduit logiquement de la conjonction de \mathcal{T} et \mathcal{H} en utilisant Γ comme des "règles de correspondance" entre les entités abstraites du modèle $(\mathcal{T}, \mathcal{H})$ et des éléments concrets du fait F à expliquer.

On peut représenter ce pattern de la façon suivante :

$$\exists \mathcal{H} \mid \mathcal{T} \cdot \mathcal{H} \xrightarrow{\Gamma} F. \quad (5.12)$$

Remarquons que ce schéma porte sur la *structure* des explications, et non pas sur sa valeur épistémologique. En d'autres termes, la donnée de cette structure ne suppose pas que l'explication en elle-même soit une *bonne* explication. C'est d'ailleurs le point central de notre discussion : quelles sont les conditions (et sur quels éléments portent-elles ?) qui font qu'une explication est une bonne explication ?

Je vais expliciter les différents éléments de cette structure sur un exemple, pour bien fixer les idées. Prenons le cas, en physique, de la chute d'une bille sphérique lâchée d'une hauteur initiale h_0 dans un fluide donné (par exemple l'air) sans vitesse initiale. La trajectoire de la bille est représentée par une certaine fonction $t \rightarrow \mathbf{r}(t) = (x(t), y(t), z(t)) \in \mathbb{R}^3$.

Le cadre théorique \mathcal{T} correspond ici au cadre de la mécanique classique de Newton, et en particulier au principe fondamental de la dynamique stipulant la relation que $\mathbf{r}(t)$ doit satisfaire :

$$m\ddot{\mathbf{r}} = \sum_i \mathbf{F}_i, \quad (5.13)$$

où $\{\mathbf{F}_i\}_i$ est un ensemble de forces supposées s'appliquer sur le système physique – ici, sur la bille.

Les hypothèses théoriques $\mathcal{H} = \{H_1, H_2\}$ spécifient alors le type et la forme précise de ces forces. Dans ce cas, on considère la force de gravité terrestre :

$$H_1 : \mathbf{F}_1 = m\mathbf{g}, \quad (5.14)$$

et la force de frottement :

$$H_1 : \mathbf{F}_2 = -\alpha\dot{\mathbf{r}}. \quad (5.15)$$

Les conditions spécifiques Γ (ou "initiales" dirait-on ici) consistent en la donnée des valeurs initiales de $\mathbf{r}(t)$ et de sa dérivée $\dot{\mathbf{r}}(t)$: $\mathbf{r}(0) = (0, 0, h_0)$ et $\dot{\mathbf{r}}(0) = \mathbf{0}$. La mesure indépendante de la masse m peut aussi compter comme une telle hypothèse.

L'équation 5.13 peut alors être résolue, donnant comme solution une trajectoire $t \rightarrow z(t)$ (le modèle empirique) qui peut ensuite être comparée aux données.

Le cas d'un modèle de choix rationnel exhibe le même type de structure générale :

CADRE THÉORIQUE \mathcal{T} : il s'agit ici du cadre général décrit en section 5.2.3 et d'un principe reliant l'utilité espérée des différentes actions au choix effectif de l'agent, comme (5.6) ou encore (5.11).

HYPOTHÈSES THÉORIQUES \mathcal{H} : ces hypothèsesinstancient concrètement le principe fondamental et sont divisées en deux types :

- quels types d'utilité sont considérés : économique, sociale, cognitive, axiologiques, ...? Dans le modèle BG dans sa forme simplifiée présentée plus haut, on a considéré uniquement l'utilité économique, mais dans le modèle original les compétences cognitives sont également prises en compte. Dans des modèles plus élaborés comme celui de Manzo, les interactions sociales sont également implémentées.
- quelle est leur forme mathématique précise? Dans le cas du modèle BG, l'utilité économique est réduite à une valeur binaire (équations (5.7) et (5.8)). Prendre en compte les compétences cognitives implique de leur attribuer une certaine *distribution* dans la population considérée (par exemple, de type normale ou log-normale). Prendre en compte l'influence sociale se traduit également par l'ajout

d'un terme d'utilité ayant une certaine forme, comme dans (MANZO, 2013a, p. 62).

CONDITIONS SPÉCIFIQUES Γ : elles peuvent consister par exemple dans la spécification de la structure des choix ou des actions disponibles. Par exemple, dans le modèle BG, ces hypothèses portent sur le fait de ne considérer que trois catégories sociales, ainsi que la structure de choix dans son ensemble donnée dans l'arbre de décision, qui sont supposés correspondre à ce qui est effectivement offert aux individus dans la réalité.

Le dilemme classique présenté dans la section 5.3.1 précédente peut maintenant être reformulé ainsi : d'une part, si aucune contrainte n'est imposée à l'ensemble \mathcal{H} , alors (5.12) est trivialement vraie et l'explication est sans contenu ; d'autre part, une explication moins triviale implique donc plus de contraintes sur l'ensemble \mathcal{H} qui finit donc par être constitué d'hypothèses sur-simplifiées voire "irréalistes". Le même argument s'applique au cadre \mathcal{T} lui-même (et aux principes qu'il contient) : plus il est spécifique, moins la condition (5.12) est triviale mais moins il est réaliste, et vice versa.

Les hypothèses Γ , quant à elles, ont un sens empirique plus "direct" en ce qu'elles permettent de connecter les entités abstraites du modèle à des caractéristiques et des propriétés concrètes du système étudié. Dans d'autres termes, leur "réalisme" est moins questionnable, puisque par construction elles sont faites pour représenter des variables que l'on mesure lorsqu'on applique le modèle à un fait particulier.

Par exemple, il est évident que si la hauteur initiale d'une bille que l'on va lâcher peut être mesurée et est égale à h_0 , alors cette valeur doit être donnée à $z(t = 0)$ (la hauteur à $t = 0$) dans le modèle. Dans les sciences sociales, ces hypothèses sont utilisées pour connecter les modèles abstraits aux situations sociales ou politiques étudiées. Pour que le modèle BG puisse être appliqué à un système scolaire ou académique concret, la structure de choix, par exemple, doit être reliée d'une façon ou d'une autre aux possibilités réelles offertes aux individus.

Dans tous le reste de ce chapitre, nous nous focaliserons donc sur la question du "réalisme" des deux autres éléments d'un modèle théorique, à savoir \mathcal{T} et \mathcal{H} , et je considérerai les Γ comme non problématiques.

Dans le cas de la TCR (mais ces considérations touchent toutes les théories de l'action), les inquiétudes quant au "réalisme" de ces deux types d'hypothèses peuvent être formulées comme suit. D'une part, on peut se demander

si les individus, lorsqu'ils agissent, sont vraiment en train de calculer l'utilité espérée de chaque action possible pour ensuite choisir, par exemple, celle qui maximise cette quantité. D'autre part, si on suppose le cadre théorique \mathcal{T} , d'autres questions pourraient par exemple porter sur les différents types d'utilité à l'œuvre dans la situation considérée : sont-elles vraiment les seules en présence ? Est-ce que leur forme mathématique est bien celle supposée ? Et ainsi de suite.

La vision tripartite des modèles théoriques présentée plus haut est donc en partie méthodologiquement justifiée par le fait que ces questions sont effectivement de natures différentes et n'appellent pas le même type de réponses.

Ce qui va nous intéresser ici, c'est le "réalisme" du cadre théorique \mathcal{T} principalement. Il s'agira de nous demander, par exemple, dans quelle mesure le "réalisme" du principe de maximisation de la TCR est pertinent vis-à-vis de sa légitimité à être utilisé comme principe explicatif.

Je me focaliserai moins sur les questions portant sur \mathcal{H} en elle-même. En effet, elles semblent poser moins de difficultés épistémologiques : si on est d'accord pour utiliser des principes généraux abstraits et non "réalistes", alors on admettra sans difficulté que les hypothèses théoriques que l'on va leur adjoindre dans des cas particuliers (par exemple : les différents types d'utilité considérés dans un modèle de choix rationnel ou bien les différentes forces en présence dans un modèle de physique classique) seront celles supposées jouer un rôle pertinent dans la dérivation de l'*explanandum*.

En d'autres termes, il ne sera pas nécessaire, pour rendre le modèle acceptable, d'avoir implémenté tous les types possibles d'utilité ou de forces qui sont "vraiment" en présence dans le phénomène en question. Le modèle de chute libre (c'est-à-dire en ne prenant pas en compte les frottements), lorsqu'il est empiriquement valide, ne suppose pas qu'il n'y a pas de fluide (dans ce cas cela serait effectivement une hypothèse fautive), mais seulement qu'étant données les conditions en présence (décrites par les Γ), la présence du fluide n'a pas de conséquences observables sur le résultat, et donc n'est pas pris en compte.

La condition de "réalisme" des hypothèses de base peut sembler de prime abord assez naturelle, autant pour les hypothèses de type conditions spécifiques que pour celles constituant le cadre théorique \mathcal{T} ainsi que pour les hypothèses théoriques \mathcal{H} . En effet, comment un modèle peut-il prétendre *expliquer* quoique ce soit si les hypothèses qui le constituent sont elles-mêmes fausses ?

Cette question est à la base de toute une littérature en philosophie et en épistémologie des modèles scientifiques, que l'on a l'habitude d'appeler la *vision fictionnelle des modèles* (FRIGG, 2009; SUÁREZ, 2009; BOKULICH, 2009; BOKULICH, 2017). Le constat premier, et paradoxal, est précisément celui-ci : l'utilisation de modèles idéalisés voire d'hypothèses notoirement fausses est largement répandue et reconnue comme une pratique tout à fait acceptable en science. Dans la section suivante, je vais argumenter précisément dans ce sens en faisant un détour par la physique. Je m'attacherai à décrire plusieurs exemples concrets pour lesquels le "réalisme" des hypothèses de base ne semble pas être un critère pertinent de sélection théorique.

5.3.3 Le statut épistémologique des hypothèses : un détour par la physique

La première étape pour surmonter le dilemme épistémologique que j'ai décrit plus haut est de réaliser que le "réalisme", ou la consistance externe, des éléments du cadre théorique \mathcal{T} ainsi que des hypothèses théoriques \mathcal{H} est une condition qui ne semble pas *nécessaire*, dans bien des cas, pour garantir à un modèle théorique un degré élevé de justification. Trois cas généraux vont être abordés : premièrement, le cas d'hypothèses de base tout simplement notoirement fausses ou inconsistantes ; ensuite, l'existence de plusieurs formulations empiriquement et théoriquement équivalentes mais ontologiquement incompatibles ; enfin, l'intestabilité de certaines hypothèses ou principes fondamentaux. On va voir que dans les exemples présentés, aucune de ces propriétés n'est considérée comme problématique vis-à-vis de la légitimité des modèles théoriques correspondant à soutenir une authentique explication des phénomènes considérés.

Hypothèses fausses, "irréalistes" ou théoriquement inconsistantes

C'est un constat assez bien partagé et plutôt ancien (POINCARÉ, 2017; CARTWRIGHT, 1983; SHECH, 2023) qu'en physique certaines hypothèses de base sont à strictement parler fausses, ou bien hautement idéalisées et donc "irréalistes". D'autres peuvent même parfois être *interdites* par la théorie considérée elle-même, c'est-à-dire inconsistantes avec d'autres réquisits.

Par exemple, expliquer les trajectoires de planètes dans le système solaire au sein de la mécanique newtonienne implique de les décrire comme des objets hautement idéalisés, par exemple des points matériels. Or, les planètes,

on le sait, *ne sont pas des points*, mais plutôt des gros objets de forme approximativement sphérique; pourtant, un tel modèle, tant qu'il produit ensuite des prédictions falsifiables et empiriquement adéquates, est bien considéré comme le substrat d'une explication scientifique authentique.

On pourrait vouloir être "plus réalistes" et modéliser les planètes comme des sphères, mais pour une certaine gamme d'observations cette hypothèse "plus réaliste" n'aura absolument aucun effet sur le succès empirique du modèle, et donc sur son pouvoir explicatif.

Ajoutons que les points matériels, non seulement n'existent pas, mais sont bien interdits par la physique newtonienne elle-même, puisqu'ils supposent une densité d'énergie infinie. Cela suggère déjà un point que je développerai en détail plus loin : que ce type d'hypothèses n'a pas de contenu empirique substantiel mais ne sert qu'à soutenir une certaine structure explicative.

Un autre exemple est la théorie cinétique des gaz, de laquelle on peut par exemple dériver la loi des gaz parfaits déjà citée, reliant la pression P , le volume V , la température T et la quantité de matière n d'un gaz dans certaines conditions expérimentales :

$$PV = nRT, \quad (5.16)$$

où $R = 8,314 \text{ J mol}^{-1} \text{ K}^{-1}$ est la "constante des gaz parfaits". La relation (5.16) qui, je le rappelle, constitue un modèle empirique dans notre vocabulaire, peut être dérivée de considérations microscopiques, en modélisant le gaz comme un ensemble de N particules identiques (des points matériels de masse m) confinées dans un volume V . Les particules sont supposées interagir seulement avec les parois de la boîte dans laquelle elles sont confinées via des chocs élastiques, et pas entre elles.

Due au nombre de particule qu'un tel gaz contient effectivement (de l'ordre de $\approx 10^{23}$), une application stricte de la mécanique newtonienne est impossible : une approche statistique est nécessaire. Plus précisément, la vitesse des particules est supposée être distribuée selon une certaine loi statistique, la loi de Maxwell-Boltzman. Ensuite, la stratégie est de dériver la pression P comme l'effet macroscopique des chocs de particules sur les parois (Cf figure 5.4) en comptant le nombre de particules ayant une certaine vitesse (plus ou moins une quantité infinitésimale) qui atteint une surface infinitésimale de la paroi dans un temps lui-même infinitésimal.

Ces considérations mènent alors à la force totale qui s'exerce sur la paroi par unité de surface, et permet de retrouver la loi (5.16).

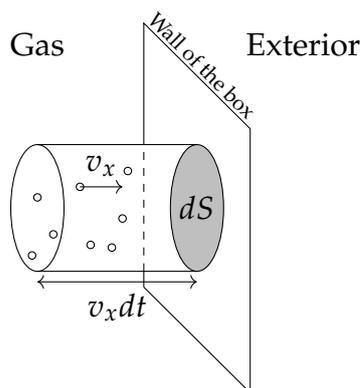


FIGURE 5.4 – Schéma de la modélisation d'un gaz. Combien de particules avec une vitesse selon x v_x dans l'intervalle $[v_x, v_x + dv_x]$ atteint la surface d'aire dS dans l'intervalle de temps dt ?

Dans ce modèle, on suppose premièrement que les molécules du gaz sont des points matériels qui n'interagissent pas entre eux, alors qu'en "réalité" elles sont probablement soumises à des milliards de chocs par seconde. Ce n'est donc pas une simple approximation, mais bien une hypothèse *strictement fausse*. Deuxièmement, leur vitesse est supposée être distribuée d'une certaine façon, mais cette hypothèse n'est pas directement testable.

Enfin, on sait par ailleurs que les atomes et les molécules sont en réalité des objets bien plus compliqués que des simples points matériels ou même des assemblages compliqués de sphères : en physique quantique, les systèmes fondamentaux sont décrits non pas comme des objets matériels comme des sphères, mais comme des fonctions d'onde ne fournissant de l'information qu'à propos de la probabilité du système de se trouver dans différents états.

Ainsi, les hypothèses de base de ces modèles ne sont pas uniquement invérifiables, mais probablement complètement fausses au sens strict du terme. Cependant, il est important d'observer que dans le cas de la théorie cinétique des gaz, il n'est pas nécessaire de modéliser les molécules comme en mécanique quantique. En effet, cela n'affaiblit pas le pouvoir explicatif du modèle statistique, alors qu'un modèle "plus réaliste", c'est-à-dire un modèle reposant sur une description quantique des objets de base plutôt que sur une description en termes de points matériels ne serait pas forcément un meilleur modèle – si, bien sûr, l'objectif est de dériver la loi des gaz parfaits.

Il est toujours possible de relier certains aspects d'une description simple à des considérations quantiques, par exemple en reliant l'hypothèse des atomes ou des molécules "comme des sphères" au modèle des harmoniques sphériques de l'atome en mécanique quantique. J'ajouterai également qu'il

est clair que cette consistance externe est une bonne propriété épistémologique.

Cependant, mon propos ici est que cette connexion n'est pas *nécessaire* pour qu'un modèle théorique soit acceptable.

Diverses formulations théoriques ontologiquement incompatibles

Une théorie physique donnée peut habituellement être formulée de plusieurs manières différentes, et ces formulations, même si elles sont théoriquement et empiriquement équivalentes, postulent différentes entités ou mécanismes.³

Un exemple bien connu est la mécanique newtonienne elle-même, qui peut être formulée en postulant soit un ensemble de forces $\{\mathbf{F}_i\}_i$ qui agissent sur des points matériels telles que leur trajectoire satisfait le principe fondamental de la dynamique (5.13), ou bien un lagrangien \mathcal{L} associé au point matériel et à la situation physique, tel que celui-ci suit la trajectoire γ qui minimise une certaine quantité, appelée *Action* : $\int_{\gamma} \mathcal{L} dt$, ou encore un hamiltonien source de la dynamique du système physique, vu comme un point (\mathbf{p}, \mathbf{q}) dans un espace des phases, satisfaisant les équations d'Hamilton : $\frac{dq}{dt} = \frac{\partial \mathcal{H}}{\partial p}$ et $\frac{dp}{dt} = -\frac{\partial \mathcal{H}}{\partial q}$.

De manière plus spécifique, on peut aussi mentionner la théorie *classique* de la gravitation qui peut être formulée dans le cadre newtonien en postulant la force bien connue en $1/r^2$, mais peut aussi être formulée dans le cadre géométrique de Newton–Cartan sans postuler une force mais en la décrivant comme une manifestation de la courbure d'un espace–temps (classique) sous-jacent.

Cette pluralité de formulations n'est pas une spécificité de la mécanique newtonienne, mais une propriété partagée par toutes les théories physiques. Par exemple, la mécanique quantique connaît au moins neuf formulations différentes (STYER et al., 2002) : la formulation de Heisenberg basée sur les matrices, celle de Schrödinger basée sur la fonction d'onde, celle de Feynman basée sur l'intégrale de chemin, celle de Wigner basé sur l'espace des phases, ou encore celle basée sur la matrice densité, la seconde quantification, la formulation variationnelle, celle de de Broglie-Bohm basé sur l'onde pilote ou finalement celle dite de Hamilton-Jacobi.

Il en va de même pour la relativité générale, qui peut également être formulée de différentes façons (GÖCKELER et SCHÜCKER, 2011 ; KRASNOV,

3. Voir par exemple (SUPPE, 2000) qui présente ce fait comme un argument allant à l'encontre de certaines positions soutenues par l'empirisme logique.

2020; ARNOWITT, DESER et MISNER, 2008) : en utilisant le calcul tensoriel sur des variétés différentielles, la géométrie de Cartan sur des fibrés principaux, en encodant la gravitation dans la torsion de l'espace-temps au lieu de l'encoder dans la courbure, ou en dérivant les équations de champs à partir d'un principe lagrangien ou même hamiltonien.

Toutes ces formulations sont équivalentes dans le sens où toutes les étapes du passage mathématique de l'une à l'autre sont parfaitement maîtrisées, et possèdent le même contenu empirique. Pourtant, les entités que ces différentes formulations postulent (des forces ou des lagrangiens, des états quantiques sous forme de vecteur ou bien de matrice densité, une torsion spatio-temporelle ou une courbure, etc.) ainsi que les principes et mécanismes de base qui sont au cœur de leur pouvoir explicatif n'ont parfois rien à voir les uns avec les autres.

Les images respectives de la réalité données par ces formulations sont donc profondément distinctes, même si celles-ci sont équivalentes entre elles et fournissent des modèles explicatifs hautement acceptables épistémologiquement.

De plus, parfois, certaines entités postulées par une théorie T ne sont plus postulées par une théorie T' qui l'englobe et la dépasse, c'est-à-dire que le succès empirique de T est strictement inclus dans celui de T' .

Un cas bien connu est la relativité générale dépassant la théorie newtonienne de la gravitation. Cependant, ce fait n'a jamais affaibli le pouvoir explicatif de la théorie newtonienne dans son domaine de validité. Ainsi, des principes fondamentaux "faux" ou en tout cas dépassés peuvent toujours jouir d'une haute valeur explicative. De la même façon qu'avec l'exemple des hypothèses de type quantiques pour dériver la loi des gaz parfaits, on pourrait même dire que pour toute une gamme de phénomènes (comme la chute d'objets sur Terre) la mécanique newtonienne est bien plus efficace (en tout cas bien mieux adaptée) que la relativité générale si l'on veut en rendre compte.

Les principes généraux et les hypothèses théoriques sont difficilement testables indépendamment

Plus généralement, on peut aussi soutenir que les principes généraux, éléments du cadre théorique \mathcal{T} et les hypothèses théoriques \mathcal{H} ne peuvent pas être directement testées, en tout cas de la même façon que les hypothèses recouvrant les conditions spécifiques Γ .

Plus précisément, et comme on l'a déjà mentionné, les principes généraux, sans plus de précision, rendent le schéma d'explication (5.12) trivialement vrai : quel que soit le fait empirique donné, il existe toujours, quitte à avoir un peu d'imagination, un ensemble d'hypothèses \mathcal{H} tel qu'il est possible d'expliquer ce fait empirique au sein du cadre théorique \mathcal{T} .

Considérons les principes fondamentaux de la dynamique newtonienne. Dans ce cadre théorique, expliquer la trajectoire $\mathbf{r}(t)$ d'un système signifie trouver un ensemble fini de forces $\{\mathbf{F}_i\}_i$ tel que (5.13) est satisfaite, c'est-à-dire tel que $m\ddot{\mathbf{r}} = \sum_i \mathbf{F}_i$. Cependant, à strictement parler, il est *toujours possible*, une fois donné une certaine trajectoire $\mathbf{r}(t)$, de trouver un certain ensemble de forces $\{\mathbf{F}_i\}_i$ tel que (5.13) soit satisfaite. Il s'agit du même écueil épistémologique rencontré avec la TCR : il est toujours possible, une fois donné un comportement humain, de trouver une fonction d'utilité sur un ensemble de choix possibles tel que ce comportement corresponde à la maximisation de cette fonction d'utilité.

De plus, à chaque fois qu'un modèle donné est apparemment réfuté, il est toujours possible d'ajouter ou de modifier une hypothèse de telle sorte que le modèle soit de nouveau empiriquement adéquat. La même remarque s'applique bien sûr à n'importe quel principe général en physique : une quantité (un ensemble de forces, un lagrangien, un hamiltonien, ...) est supposée gouverner la dynamique du phénomène que l'on cherche à expliquer au sein d'un certain cadre \mathcal{T} , mais sans plus de précision il est toujours trivialement possible de trouver la forme que cette quantité doit satisfaire afin de dériver les observations voulues à partir des éléments de \mathcal{T} .

Ainsi, la question de savoir si les hypothèses constituant le cadre \mathcal{T} sont "vraies" n'est pas une question empirique. Exiger qu'elles le soit, même "approximativement", condamne donc sans raison les modèles théoriques sur le plan épistémologique.

En ce qui concerne les hypothèses théoriques, il s'avère qu'il est impossible de les tester hors de tout cadre théorique \mathcal{T} . En d'autres termes, les hypothèses théoriques \mathcal{H} sont toujours *subordonnées* à des principes constitutifs de cadres théoriques \mathcal{T} .

Par exemple, se demander dans quelle mesure l'énergie cinétique E_c d'un point matériel de masse m et de vitesse v , en mécanique newtonienne, est "vraiment" égale à $\frac{1}{2}mv^2$ n'a aucun sens empirique (ou en tout cas pas celui qu'on lui donnerait à première vue), car l'hypothèse :

$$H : E_c = \frac{1}{2}mv^2 \quad (5.17)$$

n'a aucune signification en dehors de la mécanique newtonienne. C'est le cas d'ailleurs de n'importe quelle hypothèse de ce type, comme par exemple une force de frottement (5.15). En d'autres termes, il n'y a aucun moyen de mesurer indépendamment l'énergie cinétique de manière à la comparer à (5.17) pour "tester" celle-ci.

Bien entendu, les principes fondamentaux constitutifs d'un cadre théorique peuvent parfois être dérivés de principes plus généraux comme des cas particuliers. Par exemple, les équations classiques de la gravitation peuvent être dérivées des équations d'Einstein comme un cas limite, pour des potentiels gravitationnels Φ tels que $\Phi \ll c^2$ où c est la vitesse de la lumière.

Les hypothèses théoriques peuvent aussi être dérivées comme des cas limites d'autres hypothèses plus générales. Par exemple, la force de gravité terrestre (5.14) est un cas spécial de la force gravitationnelle $\mathbf{F} = -\frac{GmM}{r^2}\mathbf{e}_r$ pour $M = M_T$ (la masse de la Terre) et $r = R_T + z$ avec $z \ll R_T$, où R_T est le rayon de la Terre.

Cependant, dans tous ces cas, la dérivation repose de toute façon elle-même sur d'autres principes fondamentaux ou d'autres hypothèses théoriques – ainsi, épistémologiquement parlant, cela ne fait que déplacer le problème.

Ces différents constats suggèrent ainsi fortement que le "réalisme" des hypothèses de base n'est finalement pas un critère aussi pertinent qu'on pourrait le croire a priori dans l'acceptation des modèles théoriques. La consistance des hypothèses de bases dans les modèles théoriques en physique ne semble pas seulement superflu mais parfois même impossible à atteindre, sans que cela n'affecte réellement leur valeur épistémologique.

L'utilisation de telles hypothèses ne semble donc pas poser de problème dans l'explication des phénomènes physiques. La question se porte donc davantage sur la manière de *justifier* cette utilisation. Quel critère est à l'œuvre pour assurer une justification aux modèles théoriques?

La partie suivante va être consacrée à répondre à ces interrogations.

5.4 Critère structural et classification des phénomènes

5.4.1 Contraintes épistémologiques structurelles

La présente discussion est partie de la constatation d'un certain dilemme épistémologique concernant la mobilisation, à des fins explicatives, des modèles théoriques. Je suis partie du schéma explicatif suivant : un fait F est expliqué au sein d'un cadre théorique \mathcal{T} s'il existe des hypothèses théoriques \mathcal{H} tel que F se déduit logiquement de la conjonction de \mathcal{T} et de \mathcal{H} . Je l'ai illustré comme suit :

$$\exists \mathcal{H} \mid \mathcal{T} \cdot \mathcal{H} \xrightarrow{\Gamma} F. \quad (5.18)$$

Les "conditions spécifiques" Γ jouent le rôle des "conditions initiales" correspondant à la situation concrète étudiée. Elles permettent de faire le lien entre les grandeurs théoriques et certaines caractéristiques observables.

Comme je l'ai décrit plus haut, un modèle théorique se retrouve alors entre deux positions insatisfaisantes sur le plan épistémologique. D'une part, si aucune contrainte n'est imposée aux hypothèses théoriques permises dans le type d'explication défini par le cadre \mathcal{T} , alors la condition (5.18) est trivialement satisfaite, et l'explication est donc vidée de toute substance. D'autre part, une explication moins triviale implique davantage de contraintes sur les hypothèses \mathcal{H} permises, qui finissent donc par être très idéalisées puisque gagnant en généralité, et donc peuvent être vues comme "irréalistes".

Une réflexion épistémologique à propos de ce qui fonde une bonne explication théorique nécessite donc de se positionner face à ce dilemme.

L'argument que je défends ici s'inspire des théories physiques et peut se présenter comme suit. Éviter la trivialité des explications revient à imposer certaines contraintes sur les hypothèses théoriques \mathcal{H} acceptables. C'est sur *la structure de ces contraintes* que doivent porter les critères épistémologiques pertinents, et non pas sur les hypothèses théoriques elles-mêmes.

Prenons l'exemple de l'hypothèse de la force gravitationnelle en physique newtonienne : $\mathbf{F} = -\frac{GmM}{r^2}\mathbf{e}_r$. Ce qui rend cette hypothèse épistémologiquement acceptable, ce n'est pas le fait qu'elle soit "réaliste", ou tout simplement "vraie", pour toutes les raisons exposées dans la section précédente. Ce qui justifie son utilisation, c'est que cette hypothèse est *systématiquement associée* avec succès à un certain ensemble de situations empiriques. Le succès de cette

association peut être mesurée relativement à l'adéquation empirique non triviale qu'elle génère – ce que j'ai formalisé sous le concept de *corroboration informationnelle* au chapitre 4.

Ainsi, les contraintes imposées à l'ensemble des hypothèses théoriques acceptables exhibent à terme une certaine *structure* : une association systématique entre des phénomènes et des classes de modèles qui les expliquent non trivialement.

Un principe fondamental permet donc avant tout de générer des classes de modèles au sein desquels on va pouvoir catégoriser les phénomènes étudiés. En physique newtonienne, les classes de modèles sont générées à partir du principe fondamental de la dynamique en spécifiant la forme des forces considérées. Par exemple, l'hypothèse $\mathbf{F} = -\frac{GmM}{r^2}\mathbf{e}_r$ génère ainsi la classe des modèles de gravitation newtonienne, dont les chutes libres, les chutes avec frottement, etc. constituent des sous-classes.

Un autre exemple, toujours en mécanique newtonienne, est la classe des phénomènes *conservatifs*, définie par des forces de la forme :

$$\mathbf{F} = -\nabla U,$$

où U (le potentiel) est une application scalaire et ∇ représente l'opérateur gradient.

En anticipant un peu, notons que ce constat s'applique également aux théories de l'action en sciences sociales, et en particulier à la TCR. Au sein de cette théorie, dans une situation de choix binaire (où l'on note les deux choix possibles $\{-1, 1\}$), un terme d'*influence sociale* peut être ajouté à l'utilité totale de l'agent (BROCK et DURLAUF, 1999; DURLAUF, 2001) :

$$U_i^S(A, t + 1) = -\sum_j \omega_{ij}(A - A_j(t))^2.$$

Cela signifie que l'utilité sociale que l'agent i associe au temps $t + 1$ à l'action $A \in \{-1, 1\}$ dépend de l'action faite par les agents auxquels l'agent i est connecté, le terme ω_{ij} quantifiant l'influence entre i et j .

Un tel terme d'utilité définit ainsi une classe de modèles explicatifs au sein de la théorie du choix rationnel. Il peut ainsi être utilisé pour classer des situations ou phénomènes sociaux comme étant des phénomènes "avec influence sociale". Ce terme, très général, peut lui-même être spécifié davantage, ce qui crée ainsi des sous-classes de modèles, et ainsi de suite.

En physique, comme on l'a vu, les théories peuvent la plupart du temps être formulées de différentes manières, ontologiquement incompatibles entre elles. Pourtant, les classes de modèles générées par les différentes formulations forment des structures *isomorphes*. Pour reprendre l'exemple de la force de gravitation, le terme "proportionnel à l'inverse du carré de la distance" est valable dans la formulation standard, où les entités explicatives de base sont des *forces*. Dans la formulation lagrangienne ou hamiltonienne, ce terme n'a tout simplement plus de sens. Cependant, dans chacune de ces formulations, il existe un terme bien précis encodant le même phénomène de gravitation. Ainsi, les autres formulations génèrent in fine les mêmes classes de modèles théoriques.

La *structure* qui émerge de l'application systématique d'un certain type d'hypothèses à un certain type de phénomènes se retrouve donc être indépendante de la formulation particulière du cadre et des hypothèses théoriques mobilisés. C'est sur cette structure qu'un critère de choix théorique doit s'appliquer. Ma proposition est que *la classification des phénomènes générées par l'application d'un cadre théorique \mathcal{T} donné doit reposer sur le moins de classes possibles tout en générant un degré moyen de corroboration informationnel élevé*.

La figure 5.5 illustre l'idée d'une telle classification. À partir d'un cadre théorique \mathcal{T} , on construit des classes de modèles en lui adjoignant des ensembles d'hypothèses théoriques $\mathcal{H}_I, \mathcal{H}_{II}, \dots$. On construit alors une application, que je note Φ , qui associe un fait empirique à une classe de modèles théoriques si ce fait peut être expliqué par un modèle de cette classe dans le sens où on l'a décrit plus haut – c'est-à-dire en satisfaisant (5.18), tout en générant un certain degré moyen de corroboration informationnelle (une adéquation empirique non triviale).

À l'aide de ce critère permettant de caractériser la valeur épistémologique d'un cadre théorique et des hypothèses théoriques associées, je vais montrer dans la section suivante comment il permet de se positionner face au dilemme au fondement de cette discussion.

5.4.2 Conséquences épistémologiques

Le "réalisme" des hypothèses de base n'étant plus un critère de sélection théorique pertinent, le dilemme épistémologique initial peut alors être naturellement dépassé. En effet, si les hypothèses de type "conditions spécifiques", que j'ai notées Γ , ont un sens empirique direct, ce n'est pas le cas des

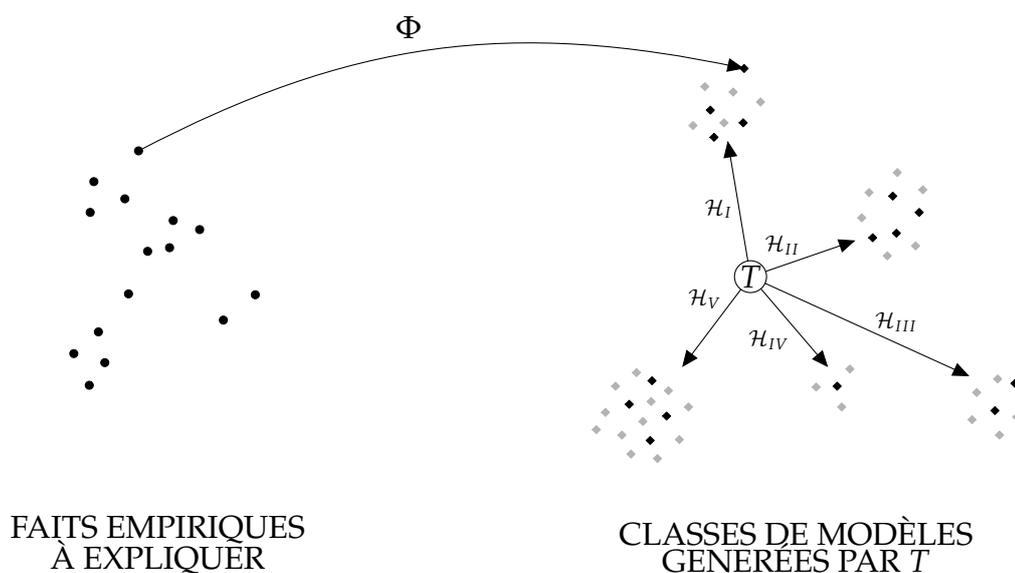


FIGURE 5.5 – Les principes constitutifs du cadre théorique \mathcal{T} vus comme des applications classificatoires. Les cercles noirs pleins représentent les faits empiriques à expliquer. Les carrés noirs pleins représentent les modèles générés à partir de \mathcal{T} et de l'adjonction d'un ensemble fini d'hypothèses théoriques, notés $\mathcal{H}_I, \mathcal{H}_{II}, \dots$. Les carrés gris pleins représentent les modèles qui ne sont pas (ou pas encore) associés à un fait empirique mais qui appartiennent tout de même à l'espace des modèles défini par \mathcal{T} . Φ associe les faits empiriques aux modèles théoriques qui les expliquent d'une manière non triviale.

hypothèses contenues dans \mathcal{T} ou \mathcal{H} . La confusion semble provenir principalement du fait que ces dernières sont souvent identifiées comme étant de même nature que les hypothèses du type Γ . Certaines hypothèses, comme "les agents calculent l'utilité espérée de chaque action possible et choisissent celle qui la maximise", ou bien "les particules suivent la trajectoire qui minimise l'intégrale du lagrangien" semblent effectivement *décrire une situation réelle*, au même titre que les Γ .

Cependant, je propose ici de les voir comme des *substrats de modélisation*, de la même façon que la matière dans laquelle est réalisée la maquette d'un avion, par exemple, n'est pas une caractéristique fidèle de l'avion réel ainsi modélisé, mais simplement un substrat (ici, physique) de la représentation. Si la maquette est en plastique, on comprendra en la manipulant qu'elle ne suggère pas que l'avion réel représenté (la "cible" de la représentation) est en plastique lui aussi. Elle permet *tout de même* de représenter d'autres choses fidèlement, comme la taille relative des différents éléments de l'avion. Ainsi, une telle maquette permet d'incarner une certaine *structure* (un ensemble de relations entre des éléments), et le plastique n'est que le substrat physique utilisé pour donner à cette structure une forme manipulable.

De la même façon, ce point de vue implique de voir les hypothèses des composantes \mathcal{T} et \mathcal{H} d'un modèle théorique comme une façon d'incarner une certaine structure, qui se trouve ici être une classification des phénomènes qui ne dépend pas in fine de la façon avec laquelle elle est représentée.

Ainsi, il semble pas que l'important soit le "as-if" mais bien le "how-if". L'important, ce n'est pas de savoir si les forces newtoniennes existent et agissent vraiment sur des points matériels, ou bien si les agents sont réellement en train de calculer des utilités espérées. Ce qui compte, c'est plutôt *comment* ils sont supposés le faire : quel type de force, ou d'utilité, et ainsi de suite, est utilisé dans quel type de situations ?

La réponse à cette question doit mettre en lumière une classification des phénomènes au sein d'un nombre fini de classes de modèles, et cette classification ne doit pas dépendre de la formulation spécifique utilisée.

Il n'en reste pas moins que le statut épistémologique des forces, des utilités ou de tout autre quantité non directement observable est une question légitime d'un point de vue philosophique. Mon propos est que si le critère ci-dessus est satisfait, alors ces modèles peuvent déjà jouir d'une certaine acceptabilité épistémologique, qui ne dépend pas de la réponse que l'on donne à la question ontologique – même si cette dernière s'avère souvent féconde.

Les *cartes* offrent des analogies éclairantes à ce genre de discussions (GIERE,

1999, p. 25, 81-88, 214). Prenons par exemple une carte de métro. Elle peut être vue, en quelque sorte, comme un modèle (ou en tout cas une "représentation épistémique" (CONTESSA, 2007)) du métro réel qu'elle vise à représenter.

Certaines règles générales, équivalentes des principes d'un cadre théorique \mathcal{T} , sont données pour tracer cette carte : le réseau du métro est représenté comme un ensemble de points reliés entre eux par une ligne droite. Des conditions spécifiques Γ font le lien entre les éléments de la carte et les éléments concrets du métro réel, comme par exemple le fait que les points représentent des stations de métro et sont nommés en accord avec les vraies stations de métro. Finalement, certaines hypothèses théoriques donnent une certaine forme à la carte : par exemple, un certain ordre entre les stations, ou encore leur point d'intersection – en d'autres termes, *la structure d'adjacence* du réseau.

À partir de ce modèle, on peut ensuite faire certaines prédictions non triviales : si on part de la station S , alors on peut sortir à la station S' en prenant telle ligne dans telle direction, en changeant à telle station, et ainsi de suite.

Cependant, la carte du métro, dans le but d'être rendue opérante (utilisable), repose aussi, dans sa réalisation, sur certaines "hypothèses" qui ne correspondent à aucune caractéristique du métro réel. Il s'agit de ce que Mary Hesse appelle les "analogies négatives", des propriétés que nous savons appartenir au modèle mais pas au système modélisé, en opposition aux "analogies positives", des propriétés que nous savons appartenir aux deux, et aux "analogies négatives", des propriétés pour lesquelles nous ne savons pas (encore) si elles appartiennent aux deux ou bien au modèle seul. Comme elle l'énonce (HESSE, 1963, p. 9-10) : "When we consider a theory based on a model as an explanation for a set of phenomena, we are considering the positive and neutral analogies, not the negative analogy, which we already know we can discard."

Dans le cas de notre carte de métro, la forme des lignes entre les stations est complètement arbitraire, tant que l'ordre (la structure d'adjacence) des stations est préservé. En plus de ça, il peut très bien exister plusieurs représentations distinctes d'un métro donné, avec différentes formes de lignes, couleurs ou autre caractéristiques du genre.

Critiquer le "réalisme" des hypothèses théoriques dans les modèles (par exemple, l'hypothèse de maximisation dans les modèles de choix rationnel) revient à critiquer le fait de représenter les lignes entre les stations comme ayant une forme qui n'est pas la bonne, par exemple une ligne droite. En d'autres termes, certains éléments d'une carte de métro (comme par exemple

la forme des lignes entre les stations) ne sont pas des énoncés empiriques à propos du métro réel mais simplement une manière relativement commode de représenter une certaine structure, et c'est cette structure qui correspond à l'engagement empirique authentique du modèle.

On pourrait même aller plus loin et dire qu'il n'est même pas nécessaire de modéliser avec exactitude la façon avec laquelle une rame de métro fonctionne effectivement pour que la carte atteigne son objectif épistémologique. C'est pour cela qu'une théorie de l'action "réaliste" ou "psychologiquement plausible" pour modéliser les phénomènes sociaux (macro) semble être un critère épistémologique excessivement exigeant. Il revient en quelque sorte, si l'on file la métaphore, à exiger une description précise de la façon avec laquelle un métro fonctionne effectivement pour pouvoir tracer une carte de métro qui aurait pour objectif de permettre à son utilisateur de circuler efficacement entre les stations.

Il est sans doute *plus aisé* de tracer une carte de métro si on sait qu'une rame de métro est une sorte de train circulant sur des rails de telle ou telle façon plutôt que si l'on croit qu'il s'agit d'une monstrueuse créature des profondeurs qui erre sans but dans les sous-sols urbains. De la même façon, il est probablement plus facile de construire des modèles sociologiques si on a à disposition des modèles robustes du comportement humain. Encore une fois, ce n'est pas un critère *nécessaire* d'un point de vue épistémologique.

De même que les lignes de métro doivent être représentées ayant une certaine forme (par exemple, comme des lignes droites), les agents dans les modèles de choix rationnels doivent être représentés comme agissant d'une certaine façon (par exemple, en maximisant une certaine fonction).

Comme dans les théories physiques, la TCR peut être formulée de différentes façons, en supposant différents genres d'entités et de mécanismes. Soit A un ensemble d'actions possibles menant à un ensemble Ω avec certaines probabilités associées. Si une relation de préférence \leq_A est attribuée à l'ensemble A de telle façon à ce qu'elle satisfasse certaines propriétés de bases comme la complétude, la transitivité, la continuité et l'indépendance (voir par exemple (WHEELER, 2020, p. 4-5)), alors il est possible de montrer, et c'est d'ailleurs un résultat fondamental de la théorie de la décision (NEUMANN et MORGENSTERN, 2007), qu'il est alors toujours possible d'associer à chaque action $a \in A$ une utilité espérée $U : A \rightarrow \mathbb{R}$ telle que la relation de préférence \leq_A soit représentée par la relation d'ordre usuelle sur les nombres

réels, c'est-à-dire telle que pour tout $a, b \in A$,

$$a \leq_A b \Leftrightarrow U(a) \leq U(b).$$

Ainsi, la théorie de l'utilité espérée est simplement une représentation particulière d'un certain principe de rationalité. Elle peut également être formulée de façon plus "géométrique" (RYAN, 2008 ; DURAND et al., 2015). Dans ce cas, les actions sont représentées par des vecteurs d'un espace de dimension $n = |\Omega|$, plus précisément des vecteurs du simplexe unité de \mathbb{R}^n . L'utilité assignée à chaque résultat possible (sur les différentes dimensions de l'espace) est alors représentée par un unique vecteur, et l'utilité espérée d'une action se retrouve être naturellement le produit scalaire de cette action et du vecteur utilité.

Une fois encore, à la lumière des réflexions précédentes, ces objets particuliers ne doivent pas être interprétés trop littéralement, mais simplement comme des *représentations* possibles d'un certain principe. Ces représentations n'ont pas besoin d'être réalistes, ou même consistantes avec ce qui est connu par ailleurs du comportement humain. Ce qui compte réellement, si on reprend l'exemple de la TCR, est quels types d'utilité sont pris en considération dans le processus d'optimisation, et si ces types d'utilité sont déjà utilisés dans d'autres contextes similaires – et non pas si ce processus est "réaliste" ou pas.

J'insiste sur un point qui me semble central : les hypothèses de base des modèles théoriques (celles de type \mathcal{T} ou \mathcal{H} , pas celles de types Γ) n'ont pas besoin d'être "réalistes" car, malgré leur forme, elles ne sont pas des énoncés empiriques. Du point de vue que je développe ici, elles sont simplement une façon d'incarner un certain principe classificatoire qui est la proposition empirique vraiment testée dans le modèle.

Je me place donc en désaccord avec les positions épistémologiques qui exigent, pour qu'un modèle théorique soit acceptable, que ses hypothèses de base soit "réalistes", comme défendu par (HEDSTROM, 2005). Cependant, je suis également en désaccord avec des positions purement instrumentalistes comme (FRIEDMAN, 1953) pour lesquelles l'adéquation empirique est un critère primordial et central.

En effet, de mon point de vue, le succès empirique n'est pas une condition suffisante pour qu'un modèle théorique soit acceptable si ce succès n'est pas relié à la mise en évidence d'une classification non triviale des phénomènes étudiés qui soit indépendante de la formulation du modèle théorique.

Concrètement, dans le cadre de la TCR, "the relevant question to ask" (FRIEDMAN, 1953, p. 15) pour justifier l'utilisation, par exemple, d'une certaine forme d'utilité, est dans quelle mesure cette forme particulière est utilisée dans des situations similaires, c'est-à-dire pour des faits ou des situations empiriques appartenant à la même *classe*.

Cependant, il est important de ne pas perdre de vue qu'il s'agit d'un processus *dynamique*. Les contraintes imposées aux hypothèses théoriques, c'est-à-dire les réponses aux questions du type : "quelles hypothèses doit-on ou peut-on utiliser dans telles situations ?" ne sont que partiellement données a priori. Je vais examiner ces aspects dans la section suivante.

5.4.3 Homogénéité des phénomènes et aspects dynamiques

Un dernier aspect important reste à éclaircir. Avant toute classification théorique, les phénomènes ou situations empiriques jouissent déjà d'une certaine classification, aussi intuitive ou "pré-scientifique" soit-elle. Avec ou sans théorie scientifique, on "range" en effet les objets et les phénomènes du monde dans un nombre limité de catégories.

Comme l'explique (EGRÉ et O'MADAGAIN, 2019) dans leur formalisation de la notion de *concept utility* (voir aussi la section 2.9.3 de cette thèse), les catégories dans lesquelles on classe un certain ensemble d'objets doivent être à la fois inclusives et homogènes. En d'autres termes, elles doivent regrouper chacune le plus d'objets possibles, mais tout en veillant à ce que l'homogénéité des différentes catégories reste élevée. Si les objets dans une même catégorie sont assez homogènes, on peut en effet inférer des propriétés connues des uns aux autres avec une certaine assurance.

Cette façon de formaliser la conceptualisation peut être facilement adaptée au propos qui m'occupe ici. Dans mon cas, il ne s'agit pas de concepts à proprement parler, mais de classes de modèles théoriques. On cherche à classer un certain ensemble d'objets (ici : des phénomènes ou faits empiriques divers) dans des classes de modèles théoriques (qui jouent le rôle des concepts pour la *concept utility*). Pour pouvoir appliquer avec une certaine confiance un modèle théorique donné à un nouveau phénomène, il faut donc que les classes de phénomènes produites par un tel procédé jouissent également d'une certaine homogénéité.

Cependant, il n'y a aucune raison a priori que les meilleures catégorisations au sens où je l'ai défini plus haut maximisent aussi l'homogénéité des classes de phénomènes correspondantes. Il y a en réalité une relation

assez complexe entre les classifications pré-théoriques (basée sur une similarité/homogénéité qualitative) et théorique (basée sur l'inscription du phénomène dans une classe de modèle explicatif).

Partant d'une classification pré-théorique, on associe dans un premier temps des phénomènes similaires à des modèles similaires. En faisant cela, il se peut toutefois que des phénomènes considérés a priori comme hétérogènes se retrouvent finalement appartenir à la même classe – par exemple, s'ils sont expliqués efficacement par le même type de modèles théoriques. Dans ces cas-là, *la notion même d'homogénéité sera alors révisée* pour augmenter celle-ci au sein des différentes classes.

Soyons plus précis. Admettons que l'on définisse l'homogénéité à partir de la présence ou l'absence d'un certain nombre de caractères $\mathcal{F} = \{F_1, \dots, F_N\}$, considérant que ces caractères sont pertinents pour classer les phénomènes en question. Il se peut que l'homogénéité des classes de phénomènes diminue si, par exemple, on parvient à expliquer des phénomènes a priori perçus comme hétérogènes au sein de la même classe de modèles.

Dans ce cas, une manière d'augmenter de nouveau l'homogénéité pourra être de modifier l'ensemble \mathcal{F} en un nouvel ensemble \mathcal{F}' de caractères qui seront alors perçus comme plus pertinents.

Illustrons cette dynamique avec des exemples historiques bien connus.

On peut penser à l'unification des phénomènes gravitationnels par Newton, qui interprète certains phénomènes terrestres (comme les chutes d'objets) et des phénomènes célestes (comme la trajectoire des planètes ou de la lune), a priori hétérogènes, comme faisant en réalité partie de la même classe de phénomènes, c'est-à-dire pouvant être expliqués par des modèles théoriques du même type.

Les qualificatifs "terrestres" et "célestes" sont alors rendus caducs par la classification théorique, et de ce fait l'homogénéité de ces classes de phénomènes augmente à nouveau. En d'autres termes, ces phénomènes ont bien sûr toujours les caractères "terrestres" ou "célestes", c'est simplement qu'ils ont disparu de l'ensemble fini de caractères considérés comme pertinents pour savoir dans quelle classe de modèles on inscrit les phénomènes.

L'unification opérée par la théorie électromagnétique de Maxwell au dix-neuvième siècle est du même type. Des phénomènes a priori très hétérogènes sont finalement rangés dans une classe commune pour des raisons théoriques. L'hétérogénéité interne à ces classes diminue alors en cessant de considérer certains caractères de ces phénomènes comme pertinents.

Un autre exemple important en physique est la détection d'une anomalie empirique et sa résolution dans un nouveau cadre. Prenons le cas de l'avancée du périhélie de Mercure dans un contexte pré-relativiste, par exemple à la fin du dix-neuvième siècle. Si cette observation est considérée comme une anomalie, c'est parce qu'elle semble constituer a priori un phénomène gravitationnel newtonien tout à fait classique mais que les astronomes de l'époque ne parviennent pas à l'expliquer au sein d'un modèle de ce type.

Plus précisément, en utilisant mon cadre analytique, on peut résumer la chose comme suit : 1. le degré moyen de corroboration informationnelle diminue lorsqu'on observe mieux la trajectoire de Mercure et qu'on lui applique un certain modèle (modèle à 8 planètes) ; 2. on élabore un modèle avec une nouvelle planète, nommée Vulcain, pas encore observée, qui augmente le degré d'adéquation mais toujours pas le degré de corroboration ; 3. on n'observe pas Vulcain, ce qui le cas échéant aurait bien augmenté de nouveau le degré moyen de corroboration informationnelle.

Lorsque la théorie de la relativité générale parvient finalement à expliquer "naturellement" la trajectoire de Mercure (c'est-à-dire sans avoir à postuler l'existence d'une nouvelle planète), c'est parce qu'elle constitue un nouveau cadre théorique qui propose une nouvelle classification des phénomènes.

La trajectoire de Mercure apparaît alors comme n'appartenant pas à la même classe de phénomènes que la trajectoire des autres planètes du système solaire, se trouvant dans un champ de gravitation trop intense pour cela. Le fait d'être une planète tournant autour du Soleil n'est alors plus considéré comme un caractère pertinent pour classer les phénomènes gravitationnels et doit être remplacé par un nouveau – ici, l'intensité du champ de gravitation dans laquelle l'objet se trouve.

En biologie, la théorie de l'évolution a produit exactement le même type de changement de critères classificatoires. Des êtres vivants a priori hétérogènes (comme les dauphins et les loups) se trouvent en réalité appartenir à la même classe explicative (ici, les mammifères) malgré leur hétérogénéité apparente, là où d'autres (comme les dauphins et les requins), pourtant a priori plus semblables, finissent par ne plus appartenir au même ensemble.

C'est exactement la même dynamique qui est à l'œuvre : on commence à travailler à partir de certains caractères qui nous semblent pertinents, on élabore des théories qui nous permettent de classer les êtres vivants, puis on modifie, éventuellement, l'ensemble de caractères de base qu'on s'était donné a priori.

J'ai formulé ce critère *classificatoire* en m'inspirant principalement de ce

qu'il se passe en physique. J'ai suggéré à plusieurs reprises, dans les sections précédentes, que l'on pouvait l'appliquer à d'autres disciplines, et en particulier aux sciences sociales. Pour être totalement complet dans mon analyse, il me reste donc maintenant à évaluer dans quelle mesure la sociologie analytique se plie également à ce critère, afin de savoir si elle constitue, comme je l'ai supposé, un contre-exemple robuste à la thèse épistémologique pluraliste.

5.4.4 Retour sur les théories de l'action en sociologie analytique

En ce qui concerne la sociologie analytique et sa constitution satisfaisante en tant que contre-exemple effectif de la thèse épistémologique pluraliste, il reste à évaluer si les théories de l'action mobilisées – qu'il s'agisse de la théorie du choix rationnel ou d'autres théories – répondent bien au critère que j'ai élaboré dans ce chapitre. En d'autres termes, il reste à examiner dans quelle mesure ces théories produisent effectivement une classification des phénomènes sociaux au sein d'un nombre restreint de classes de modèles explicatifs.

Une chose semble déjà sûre : si une telle classification existe, elle est de prime abord beaucoup moins explicite qu'en physique.⁴ Cependant, certains aspects du développement de la sociologie analytique font pourtant bien penser qu'une telle sorte de classification est non seulement possible et pertinente, mais qu'elle est également déjà en cours de formation. C'est ce que je voudrais soutenir dans cette dernière partie.

Pour rendre compte d'une telle classification dans le cadre de la sociologie analytique, il faut d'abord prendre conscience que si les théories de l'action en constituent le noyau explicatif, c'est dans la mesure où celles-ci s'incarnent concrètement dans tout un ensemble de *mécanismes générateurs*. Ces derniers acquièrent leur caractère explicatif en ce qu'ils s'articulent au sein de théories de moyenne portée (*middle-range theories*). Ce concept, forgé initialement par Robert Merton (MERTON, 1968, p. 61), constitue, comme on l'a vu dans ce chapitre, un élément central du programme de recherche de la sociologie analytique. Les théories de moyenne portée, ou "modèles de portée intermédiaire" (BOUVIER, 2008), se focalisent sur des phénomènes ou des situations sociales circonscrites, et font intervenir un nombre restreint d'éléments à la

4. Voir qu'en biologie si l'on considère la taxonomie des espèces comme correspondant à une classification des phénomènes.

fois dans les aspects empiriques mis en évidence et dans les mécanismes explicatifs mobilisés.

J'ai déjà cité un certain nombre de ces mécanismes, comme la prophétie auto-réalisatrice (MERTON, 1948) ou le phénomène d'effet de seuil (GRANOVETTER, 1978). On pourrait également penser aux mécanismes de ségrégation modélisés par Thomas Schelling (SCHELLING, 1978), ou encore à ceux qui sous-tendent les phénomènes de diffusion d'innovation (COLEMAN, KATZ et MENZEL, 1957; MANZO et al., 2018).

De toute évidence, un mécanisme, pour être effectif, doit quand même reposer sur des règles générales de fonctionnement, et les théories de l'action (comme la théorie du choix rationnel) constituent précisément les cadres théoriques qui donnent et structurent ces règles. Même si cela demanderait une réflexion approfondie, il semble d'ailleurs assez direct que les quatre types de mécanismes cités au paragraphe précédent peuvent tous se formuler, dans une forme ou une autre, au sein de la théorie du choix rationnel.

Pour autant, ce sont avant tout les mécanismes proposés qui jouent le rôle d'éléments explicatifs, les lois sous-jacentes n'étant pas, de fait, systématiquement explicitées. C'est d'ailleurs ce que défend Alban Bouvier (BOUVIER, 2008; BOUVIER, 2023) relativement à la réception qu'a pu avoir l'ouvrage *Foundations of social theory* de James Coleman (COLEMAN, 1990). Si cet ouvrage a pu, en effet, être perçu comme l'application systématique de la théorie du choix rationnel dans une visée explicative générale des phénomènes sociaux, son apport essentiel, selon Bouvier, est qu'il propose de structurer un ensemble de familles de mécanismes explicatifs.

C'est précisément là que ces notions – à savoir : les théories de l'action, les mécanismes générateurs et les théories de moyenne portée – s'intègrent à la structure des modèles théoriques présentée dans ce chapitre. Les théories de l'action, comme je l'ai déjà suggéré, fonctionnent comme les cadres théoriques au sein desquels les modèles théoriques explicatifs sont formés. Ce que j'ai nommé "modèles théoriques" s'identifie alors avec les théories de moyenne portée. Enfin, les mécanismes générateurs jouent le rôle des hypothèses théoriques, celles-ci ayant pour but d'instancier des principes généraux sous une forme concrète.

Ainsi, la question de la justification des principes fondamentaux se déporte ici sur celle de l'utilisation de ces mécanismes. En d'autres termes, les mécanismes générateurs, en tant qu'ils permettent d'incarner concrètement des principes plus généraux (même si ces derniers ne sont pas, encore une fois, directement utilisés tels quels en sociologie analytique), jouent le même

rôle classificatoire en rendant compte d'une grande variété de phénomènes à l'aide d'un nombre relativement restreint de mécanismes de base, seuls ou combinés avec d'autres. Les "classes de modèles théoriques", comme je l'ai formulé, pourraient donc être remplacées par des "classes de mécanismes générateurs".

Ce qui compte, épistémologiquement parlant, c'est donc à quel point le même type de mécanismes est appliqué à des phénomènes sociaux similaires, et à quel point cette application systématique fini également par faire rentrer dans la même catégorie des phénomènes a priori perçu comme étant différents. C'est précisément ce processus qui semble être à l'œuvre en sociologie analytique, comme le remarque par exemple Bouvier à propos des théories (de moyenne portée) mobilisées pour expliquer la survenue des révolutions :

Thus the theory of rising expectations (Birnton, 1965) does not only explain revolutions, but also suicides, as Emile Durkheim (1897) brilliantly demonstrated. Similarly, the theory of relative deprivation (Runciman, 1966) is not only a theory that could explain the origin of revolutions (p. 475), but also the relative frustration of officers in the Air Force, where promotion was more rapid than in the military police (Stouffer and al. 1949). (BOUVIER, 2023, p. 88)

La force d'un mécanisme explicatif, donc, n'est pas nécessairement de représenter vraiment ce qu'il se passe, mais de permettre une classification systématique de phénomènes a priori très différents.

Ce point de vue rejoint ainsi l'image développée par Bouvier d'une "théorie sociologique générale comme système hiérarchisé de modèles de portée intermédiaire" (BOUVIER, 2008). Ce "système hiérarchisé" correspond précisément, me semble-t-il, à la "classification saillante des phénomènes" que développée dans ce chapitre comme un critère épistémologique fondamental de justification des explications scientifiques – et auquel la sociologie analytique semble avoir toutes les chances de se conformer.

Plus généralement, cette discussion recoupe ce que l'on disait au chapitre 3 à propos de la diversité du monde social et de la difficulté voire de l'impossibilité supposée à détecter des régularités empiriques. Comme on l'a vu, les régularités empiriques dépendent aussi de la façon avec laquelle on regarde le monde. Ce constat, encore une fois, n'est pas propre aux sciences sociales et reste pertinent dans n'importe quelle discipline.

Ainsi, développer un cadre théorique permettant de classer les phénomènes sociaux comme décrit ci-dessus impliquerait de regrouper dans la même classe de phénomènes des situations jugées a priori très différentes. Comme je viens de le suggérer, même si le processus est bien moins développé en sciences sociales qu'en physique ou en biologie, il semble toutefois être en cours.

À mesure que ce phénomène progresse, la manière même avec laquelle on observe le monde social évolue, et certains éléments qui nous paraissent pertinents aujourd'hui pour classer les phénomènes sociaux finiront peut-être par être rendus caducs par une classification provenant de cadres théoriques unificateurs. On peut donc s'attendre à ce que notre vision actuelle, faisant apparaître des phénomènes comme essentiellement distincts, laisse place peu à peu à une nouvelle catégorisation des phénomènes d'où émergera alors une nouvelle définition de l'homogénéité qui pourra tout à fait dépasser l'intuition a priori que l'on en avait.

5.5 Conclusion

Ce chapitre a permis de remplir deux objectifs.

Premièrement, j'ai pu exposer un argument *par l'exemple* à la thèse épistémologique pluraliste, en présentant la sociologie analytique, une branche de la sociologie qui se développe depuis les années 1990. L'intérêt d'étudier un tel exemple est que ce courant de la sociologie ne semble pas nécessiter une épistémologie *propre*, c'est-à-dire une définition nouvelle de ce qu'est une théorie scientifique.

Deuxièmement, pour évaluer à quel point la sociologie analytique constitue effectivement un contre-exemple à la thèse pluraliste, il m'a fallu examiner ses fondements épistémologiques. En particulier, je me suis focalisé sur les principes fondamentaux au cœur des explications qu'elle propose, à savoir les théories de l'action mobilisées au niveau individuel pour reproduire les phénomènes collectifs comme agrégation. Cela m'a permis ensuite de traiter le sujet de l'épistémologie des modèles théoriques en général, pour compléter l'élaboration d'un modèle unitaire de la scientificité, initié au chapitre 4 au niveau des modèles empiriques.

Concernant les modèles théoriques, il est apparu un dilemme fondamental entre, d'une part, une tendance à produire des explications triviales (si

aucune contrainte n'est imposée au type d'hypothèses que l'on peut mobiliser dans l'explication) et, d'autre part, une tendance à reposer sur des hypothèses hyper-idéalisées, voire notoirement fausses. On peut sortir de ce dilemme, qui est à la base de la vision dite "fictionnelle" des modèles, en relaxant la condition de "réalisme" des hypothèses de base, pour la remplacer par un autre type de contraintes.

Le point de vue que j'adopte ici est que les hypothèses à la base des modèles explicatifs n'ont pas besoin d'être interprétées littéralement, car elles ne sont rien d'autres que des substrats de modélisation permettant d'instancier concrètement un principe fondamental. Corrélativement, ce qui est exigé de ces hypothèses n'est pas d'être "réalistes", mais de générer des classes de modèles théoriques qui vont permettre de classer d'une manière non triviale les phénomènes sur lesquels ces modèles portent.

Comme je l'ai précisé, cette classification est un processus dynamique dont le résultat à un instant donné n'est pas forcément accessible au scientifique individuel. Comme dans le cas de la maximisation du degré de corroboration informationnelle que j'ai décrit au chapitre précédent, les scientifiques, à leur échelle, utilisent des critères épistémologiques "locaux", que l'on peut interpréter comme des *proxy*, c'est-à-dire des *indicateurs*, de la classification plus générale ayant cours à un niveau supérieur.

Par exemple, le principe de *parcimonie* enjoint à préférer une explication reposant sur des entités ou des mécanismes déjà connus plutôt que d'en introduire des nouveaux (SOBER, 2015). On peut l'interpréter naturellement dans le sens du critère que j'ai développé dans ce chapitre. Toute chose étant égale par ailleurs, créer une nouvelle classe de modèles est coûteux, puisque cela rend la classification globalement un peu plus triviale. En effet, plus on a de classes de modèles, plus on va pouvoir classer facilement les phénomènes, mais moins on va gagner en information. Ainsi, le principe de parcimonie est un principe "local" qui assure, dans certaines conditions, que la classification dans son ensemble reste la moins triviale possible. Les erreurs commises parfois par les scientifiques en utilisant le principe de parcimonie s'expliquent en partie par le fait qu'ils raisonnent localement, sans avoir nécessairement accès à la vision globale, c'est-à-dire à la classification dans son ensemble.

Finalement, l'évaluation de la sociologie analytique en tant que contre-exemple valable de la thèse pluraliste reste alors à approfondir, en particulier dans sa capacité à mettre en évidence une classification des phénomènes au

sein de familles de *mécanismes* de base. Au vu des arguments avancés dans ce chapitre, il y a toutefois de bonnes raisons de considérer ce courant comme étant le meilleur candidat pour produire des connaissances *fondamentales* sur le monde social. En d'autres termes, il semble être le meilleur programme de recherche pour développer la sociologie dans sa dimension *cognitive*, au sens où je l'ai défini dans l'introduction de cette thèse.

En effet, d'une part, comme on vient de le voir, l'application systématique d'un petit nombre de mécanismes de base à un nombre croissant de phénomènes sociaux a priori différents est un bon signe de la possibilité classificatoire du paradigme analytique. D'autre part, enfin, il semble que les obstacles que ce courant doit surmonter ne soient pas d'une nature différente que ceux que rencontrent habituellement les disciplines scientifiques dans leur développement.

Quoiqu'il en soit, et au vu des arguments avancés dans ce chapitre en complément à ceux présentés au chapitre 3, cette situation ne semble pas justifier l'élaboration et la défense d'un concept de scientificité qui soit propre aux sciences sociales.

Chapitre 6

Conclusion

Ayant soutenu dans cette thèse la possibilité et la pertinence d'une définition générale et unique du concept de science – ou, en d'autres termes, d'un modèle unitaire de la scientificité – ce chapitre de conclusion va me permettre de faire un point sur mon cheminement global. Après avoir évalué dans quelle mesure les diverses constructions et réflexions théoriques que j'ai élaborées permettent de répondre à mes questionnements initiaux, j'identifierai certaines pistes de développement et d'amélioration. Enfin, j'esquisserai des perspectives de recherches qui pourraient être déployées à partir du travail présenté dans cette thèse.

6.1 Une proposition de modèle unitaire de la scientificité

Le point de départ de toute ma réflexion a été le constat d'une pluralité de conceptions différentes et parfois contradictoires de la *scientificité*. D'une part, le problème global de la démarcation scientifique, bien qu'il ait donné lieu à de nombreuses propositions au cours du siècle passé, n'a pas encore trouvé de solution générale satisfaisante. D'autre part, dans des champs spécifiques comme les sciences sociales, il semble que des débats épistémologiques profonds, c'est-à-dire portant sur ce que peut ou doit être une connaissance des phénomènes sociaux, soient toujours en cours. À cette instabilité conceptuelle apparente, j'oppose dans cette thèse des arguments soutenant la possibilité et la pertinence d'une définition générale et unique du concept de science, ou en d'autres termes, d'un modèle unitaire de la scientificité.

Dans les tentatives contemporaines de définition de la scientificité, la tendance est à la prise en considération croissante de la forte dépendance d'une telle définition à des éléments du contexte socio-historique. Cela se traduit par l'adoption d'unités d'analyse de plus en plus englobantes et hétérogènes,

comme les champs épistémiques bungéens qui contiennent à la fois des éléments purement cognitifs et des éléments socio-politiques. Certains courants des *science studies* se caractérisent même par l'abolition de toute distinction pertinente entre les différents niveaux d'analyse.

Il est indéniable que des éléments sociaux, historiques et même politiques interviennent dans la classification effective d'une théorie en science, non science ou pseudo-science. Cependant, cela m'est apparu comme *une raison supplémentaire* de bien identifier l'échelle d'analyse à laquelle on se place, pour justement être en mesure de distinguer entre ce qui relève de contingences socio-historiques et ce qui n'en relève pas. C'est ce qui a guidé mon choix vers l'adoption d'un niveau d'analyse bien précis, à savoir ce que j'ai appelé les *unités épistémiques*, que j'ai identifiées aux divers types de *modèles* habituellement utilisés par les scientifiques.

Toujours est-il qu'il existe bien une tension entre d'un côté, la volonté de définir le caractère scientifique d'une unité cognitive en elle-même et, de l'autre, le fait de reconnaître l'intervention inévitable d'éléments socio-historiques dans cette définition. Mon idée centrale pour résoudre ce dilemme a été de considérer la scientificité non pas comme une qualité intrinsèque appartenant en propre à une unité épistémique donnée, mais comme le résultat d'un processus (collectif) d'optimisation – plus précisément, comme la maximisation d'une certaine quantité. Une partie de mon travail de thèse a consisté à définir précisément cette quantité à maximiser, afin de donner du contenu à cette proposition.

Ce point de vue a l'avantage d'intégrer le fait que la qualification *effective* d'une unité en science ou pseudo-science dépend de certains aspects socio-historiques, en les interprétant comme des contraintes contingentes qui pèsent sur le processus d'optimisation, tout en considérant que la quantité à maximiser, elle, n'en dépend pas.

Afin de mener à bien cette tâche, j'ai commencé par me focaliser sur une unité d'analyse encore plus restreinte, les *modèles empiriques*, et j'ai pu définir sur celles-ci un *degré de corroboration informationnelle*. Il s'agit d'une version améliorée du degré de corroboration poppérien, appliqué aux modèles et non plus uniquement aux énoncés ou systèmes d'énoncés. Ce changement d'unité d'analyse permet de définir ce degré de corroboration de telle sorte qu'il incorpore les idées fondamentales du falsificationnisme tout en neutralisant certaines limites.

La scientificité d'un tel modèle empirique est alors reliée *par définition* à la maximisation de ce degré de corroboration informationnelle sur tous les

modèles concurrents, à données empiriques égales. Ce principe de maximisation permet alors de retrouver des critères épistémologiques connus (l'adéquation empirique, la progression des *problemshifts*, ou encore la simplicité et la parcimonie) comme des instanciations particulières de ce principe.

Il ne s'agissait pas pour moi d'appliquer "le" modèle de scientificité des sciences physiques directement aux sciences sociales, puisque la question de définir la scientificité en général, et en particulier en physique, n'est elle-même pas résolue définitivement. C'est pourquoi mon approche se revendique davantage comme étant *unificatrice*, c'est-à-dire qu'elle ne cherche pas à imposer un modèle de scientificité déjà existant, mais bien à proposer un modèle qui puisse ensuite s'appliquer aux différentes disciplines, indépendamment de la nature de leurs objets.

Cependant, la question de la scientificité des sciences sociales s'est tout de même posée, en particulier concernant la revendication de certains courants à jouir d'une épistémologie à part – c'est-à-dire, selon moi, d'une définition propre de la scientificité. Après avoir présenté quatre familles principales d'arguments "pluralistes", je leur ai opposé un certain nombre de contre-arguments classiques. Il est alors apparu qu'il ne semblait pas y avoir de raisons de principe de considérer comme nécessaire l'élaboration d'une épistémologie à part pour les sciences sociales, et en particulier la sociologie.

Pour appuyer mon propos, j'ai mobilisé un argument "par l'exemple", en examinant un cas concret et manifeste d'un courant de la sociologie qui ne semble pas avoir besoin d'une épistémologie alternative, à savoir : la sociologie analytique. Cependant, pour qu'elle constitue effectivement un contre-exemple valable, il m'a fallu éprouver les conditions de justifications des principes fondamentaux logés au cœur des modèles explicatifs qu'elle propose, à savoir les théories de l'action. Ces théories sont utilisées pour modéliser le comportement d'agents individuels afin de reproduire, par leur agrégation, les phénomènes collectifs.

Cela m'a alors mené à la deuxième étape de mon élaboration : un degré de scientificité appliqué au niveau des *modèles théoriques* en général, c'est-à-dire pas uniquement en sociologie analytique mais également dans les sciences physiques. Cette élaboration m'a alors naturellement amené à la question de savoir dans quelle mesure et selon quels critères un modèle constitué d'hypothèses idéalisées – voire, pouvant être considérées comme notoirement fausses – peut produire une explication satisfaisante des phénomènes (sociaux ou physiques) étudiés.

A ce niveau, après avoir passé en revue diverses réactions possibles à ce

constat, j'ai défini un critère épistémologique *structural*. Selon ce critère, ce qui fait la scientificité d'un principe fondamental, ce n'est pas à quel point il est "réaliste", mais à quel point il permet de générer une classification des phénomènes à l'aide de classes de modèles théoriques qui les expliquent, de manière à ce que cette classification soit in fine indépendante de la formulation particulière des modèles en question. En d'autres termes, le degré de scientificité d'un cadre théorique est mesuré à l'aune de la qualité de la classification des phénomènes qu'il permet.

Afin de synthétiser mon point de vue, je propose en figure 6.1 un modèle général des unités épistémiques considérées dans cette thèse – en d'autres termes, un *méta-modèle* des modèles scientifiques.

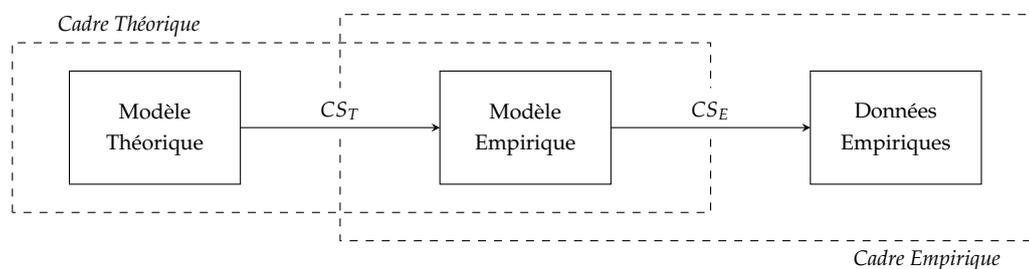


FIGURE 6.1 – Un méta-modèle des unités épistémiques élémentaires

Certains éléments de ce schéma sont aisément identifiables vis-à-vis des différentes constructions théoriques présentées dans cette thèse.

Les *données empiriques*, tout d'abord, sont produites au sein d'un certain *cadre empirique* que j'ai représenté comme un rectangle en pointillé. Le *modèle empirique*, en lui adjoignant des hypothèses supplémentaires précisant les conditions spécifiques de l'observation et notées CS_E , peut être comparé aux données empiriques. L'objectif n'est pas simplement qu'il soit *adéquat*, mais qu'il le soit de la manière la moins triviale possible, c'est-à-dire en maximisant le degré de corroboration informationnelle.

Si j'ai rendu le cadre empirique explicite dans ce schéma, c'est pour illustrer le fait que celui-ci constitue un élément potentiellement *variable* dans le processus de maximisation. Changer de cadre est non seulement permis mais parfois nécessaire. En tout cas, un changement de cadre implique une modification de la façon avec lequel le monde nous apparaît, c'est-à-dire une modification dans les données produites, comme en témoignent les exemples donnés en section 2.4.1.

La maximisation de ce degré de corroboration est donc le moteur d'une *co-évolution* de ce que l'on dit du monde *et* du langage utilisé pour le dire. Ce point de vue permet de neutraliser un certain nombre d'obstacles épistémologiques fondés sur la difficulté apparente, pour les sciences sociales, à détecter des régularités empiriques aussi stables qu'en sciences physiques. En effet, la régularité apparente du monde n'apparaît pas uniquement comme une propriété qui lui est intrinsèque mais dépend également des *lunettes* (le cadre empirique) que l'on porte pour l'observer.

Apparaissent également dans le schéma de la figure 6.1 le *modèle théorique* ainsi que le *cadre théorique* dans lequel il s'inscrit. Un modèle théorique permet d'*expliquer* un fait empirique F en re-dérivant le modèle empirique associé à F à l'aide d'hypothèses supplémentaires notées CS_T dans le schéma ci-dessus. Le rôle global d'un cadre théorique est alors de générer des familles de modèles théoriques de telle sorte à offrir une classification non triviale des phénomènes.

Le cadre théorique est également représenté sous la forme d'un rectangle en pointillé, recouvrant en partie le cadre empirique. Ce recouvrement est volontaire et signifie que les modèles empiriques sont écrits dans un langage qui est commun à la fois aux données empiriques qu'ils visent à retrouver, et aux modèles théoriques qui visent à les expliquer.

Par exemple, la variable t dans le modèle de la chute libre a à la fois un sens empirique comme faisant partie du cadre empirique (la façon avec laquelle cette variable est effectivement mesurée) et une signification théorique en tant que paramètre mathématique de la trajectoire du point matériel dans un repère donné. Dans un modèle de physique statistique visant à retrouver la loi des gaz parfaits, la variable de température T est, de même, commune aux deux cadres. Elle correspond à une certaine mesure bien définie (par exemple, via l'utilisation d'un thermomètre), dans le cadre empirique, et reflète l'énergie cinétique moyenne des particules dans le cadre théorique.

Cette identification est ce qui connecte les termes abstraits et formels du modèle théorique aux données empiriques. C'est une réminiscence des fameuses "règles de correspondance" des empiristes logiques. La différence cruciale ici est que les modèles théoriques comme les modèles empiriques, ainsi que leur recouvrement, ne sont pas fixés une fois pour toute : ce sont des présupposés hypothétiques comme les autres, que rien n'empêche a priori d'évoluer.

Les cadres théoriques et empiriques ne sont toutefois pas totalement identifiés. D'une part, certains termes théoriques n'ont pas de signification empiriques directe. Par exemple, les forces newtoniennes ne peuvent pas être directement mesurées, c'est-à-dire mesurées indépendamment du cadre théorique qui leur donne un sens. D'autre part, certains termes empiriques rentrant en compte au sein d'un modèle donné sont définis indépendamment du cadre théorique de ce modèle. Par exemple, dans le cas de la chute libre, les distances peuvent être mesurées indépendamment du cadre théorique newtonien – ou en tout cas en reposant sur des hypothèses vraiment minimales.

La scientificité d'une unité épistémique telle que représentée ci-dessus est alors définie par un processus d'optimisation global consistant à générer un fort degré moyen de corroboration informationnel tout en faisant en sorte que l'ensemble des modèles empiriques se structurent de manière non triviale au sein de classes de modèles théoriques.

C'est cette vision précise qui constitue ma proposition de modèle unitaire de la scientificité. Ce dernier est avant tout *théorique*. Cela signifie que son objectif premier n'est pas de pouvoir s'appliquer directement à des cas concrets pour savoir s'ils sont scientifiques ou pas, bien qu'il puisse être utilisé qualitativement dans ce but. De plus, même si la scientificité est perçue ici comme un processus d'optimisation, cela ne veut pas dire que cela donne une méthode systématique que l'on peut appliquer concrètement dans telle ou telle situation. Ainsi, ce modèle unitaire ne prétend pas décrire ce que font effectivement les scientifiques au quotidien, ni même ce qu'ils doivent faire.

Cependant, il permet d'interpréter d'une manière unifiée l'utilisation de critères "locaux" de scientificité, comme l'adéquation empirique ou la parcimonie des hypothèses. En effet, comme je l'ai montré à plusieurs reprises dans cette thèse, ces critères classiques apparaissent comme des *proxy* du processus d'optimisation qui s'opère à une échelle qui est le plus souvent inaccessible au scientifique individuel. Ce point de vue unifié constitue l'apport essentiel de ce travail de thèse.

Il existe sans aucun doute des moyens d'améliorer les différentes constructions que j'ai proposées dans cette thèse. La section suivante est précisément consacrée à identifier certaines limitations et à esquisser des pistes possibles d'amélioration de mon travail. Il est cependant important de garder à l'esprit que celui-ci contient *deux* propos en partie autonomes. D'une part, l'idée que l'on peut voir la scientificité comme un processus d'optimisation. D'autre part, la proposition d'une définition effective de ce qui est optimisé. Ainsi,

des critiques visant la définition particulière que j'ai proposée n'affaibliraient pas nécessairement l'idée sous-jacente – et plus importante selon moi – de la scientificité comme optimisation.

6.2 Limites

La première limite que l'on peut percevoir à mon travail est en lien direct avec son aspect théorique. Comme je l'ai expliqué ci-dessus, il ne s'agissait pas pour moi ici de fournir un nouvel ensemble de critères pratiques permettant de distinguer concrètement les disciplines ou contenus scientifiques des non-scientifiques ou pseudo-scientifiques. Mon objectif était d'élaborer une vision unifiée de la scientificité d'où l'on peut déduire certains critères épistémologiques classiques comme des cas particuliers, clarifiant par là même leur statut logique. Ainsi, les diverses constructions que j'ai proposées dans cette thèse pourront probablement sembler inutiles ou difficilement utilisables concrètement, puisque ce n'était pas leur but initial. Toutefois, une perspective possible d'application de ce point de vue à l'épistémologie sociale sera donnée dans la troisième partie de cette conclusion.

En lien avec cet aspect théorique, l'utilisation d'outils formels avait pour objectif principal de définir clairement et de manière la moins équivoque possible les objets étudiés. En particulier, j'ai mis un point d'honneur à fournir des détails structurels concernant les différentes unités d'analyse sur lesquelles j'ai basé mes réflexions. Cependant, j'admets que cette formalisation peut parfois alourdir le propos plutôt que l'éclaircir, même si je me suis efforcé autant que possible d'explicitier qualitativement les différentes constructions que j'ai menées à bien.

Une autre limite de mon travail est l'inégale élaboration des différentes constructions. En effet, si ma proposition de degré de corroboration informationnel sur les modèles empiriques est approfondie, il n'en est pas de même pour les constructions portant sur les modèles théoriques, à savoir l'évaluation quantitative de la qualité de la classification. Ma proposition représente plutôt un point de départ qualitative pour une édification future plus précise et complète.

6.3 Perspectives

Dans une perspective de recherche future, il serait intéressant d'appliquer le même cadre d'analyse conceptuelle et les mêmes questionnements

à d'autres disciplines que je n'ai pas abordées ici, comme les sciences biologiques ou psychologiques. En particulier, on pourrait se demander si le modèle unitaire de la scientificité que j'ai proposé peut permettre d'éclaircir certaines discussions épistémologiques en leur sein. Comme je l'ai suggéré, en effet, des questions du même genre, c'est-à-dire concernant la possibilité de concevoir la scientificité de ces disciplines dans un cadre unitaire, sont également vivement débattues depuis longtemps.

En philosophie de la biologie, par exemple, on peut trouver chez Ernst Mayr l'idée que l'historicité de cette discipline l'éloigne, à l'inverse des sciences physiques, de la recherche de lois, et que ses explications sont alors davantage basée sur des *narratifs* :

Evolutionary biology, in contrast with physics and chemistry, is a historical science — the evolutionist attempts to explain events and processes that have already taken place. Laws and experiments are inappropriate techniques for the explication of such events and processes. Instead one constructs a historical narrative, consisting of a tentative reconstruction of the particular scenario that led to the events one is trying to explain. (MAYR, 2000, p. 80)

L'historicité des phénomènes évolutifs pourrait donc justifier, comme en sciences sociales, la défense d'une épistémologie à part. En particulier, Mayr défend l'idée que la construction des théories en biologie est davantage basée sur des concepts "such as competition, female choice, selection, succession and dominance" (MAYR, 2000, p. 81), que sur des lois.

À cela, on pourrait répondre que ces concepts font en réalité référence à des *mécanismes*, lesquels sont supposés agir en permanence et donc ne pas dépendre d'un instant ou d'une époque donnés. De plus, le succès de la théorie de l'évolution tient justement au fait de pouvoir expliquer la grande diversité des formes de vie et de leur adaptation mutuelle en recourant à un nombre relativement restreint de mécanismes de base.

Ainsi, à première vue, le même type de schémas argumentatifs que l'on a déjà identifiés en épistémologie des sciences sociales semble se dessiner. On pourrait donc avoir de bonnes raisons d'appliquer mon modèle unitaire aux sciences biologiques afin d'éclairer certains débats philosophiques et épistémologiques.

Un autre point que j'aurais pu développer est l'application du modèle unitaire de la scientificité que je défends au problème de la démarcation scientifique à proprement parler. Si mes questionnements initiaux, comme

je l'ai mentionné en introduction, portaient bien sur la démarcation entre science, non-science ou encore pseudo-science, le corps de mon travail a surtout été focalisé sur une possible définition de la scientificité. En effet, je ne suis pas revenu sur la question d'une possible caractérisation des systèmes scientifiques dans l'optique de les démarquer des systèmes non-scientifiques ou pseudo-scientifiques.

Une façon de faire pourrait être la suivante. Partant de mon modèle de la scientificité qui la voit comme le résultat d'un processus d'une optimisation, alors les systèmes non scientifiques pourraient être vus comme des unités cognitives n'ayant pas vocation à maximiser quoique ce soit, voire même n'ayant pas une forme leur permettant de définir la quantité à maximiser. Les pseudo-sciences pourraient alors être vues de deux manières : soit comme des non-sciences (des systèmes non maximisateurs) que l'on présente comme étant scientifiques ; soit comme des unités cognitives sur lesquelles la maximisation pourrait être définie, mais qui sont simplement sous-optimaux – tout en étant présentés comme optimaux. Il s'agit bien sûr ici simplement d'une *esquisse* d'une possible application de mon modèle de scientificité au problème de la démarcation à proprement parler, esquisse qui pourrait être approfondie dans un travail futur.

Enfin, on pourrait également défendre l'application des idées développées dans cette thèse à l'*épistémologie sociale*. Comme je l'ai répété à plusieurs reprises, la qualification effective d'une théorie ou d'un modèle comme étant scientifique, non scientifique ou pseudo-scientifique est un processus collectif et *social* en ce qu'il fait intervenir des éléments socio-historiques contingents, comme par exemple les données empiriques disponibles au moment du choix théorique en question.

L'approche que j'ai développée dans cette thèse suggère alors un processus collectif d'*optimisation*, bien qu'il ne soit pas précisé ici comment cette optimisation se réalise concrètement. On pourrait alors imaginer développer des modèles formels, dans la même veine par exemple que le modèle de Krause-Hegselmann (HEGSELMANN et KRAUSE, 2002 ; HEGSELMANN et KRAUSE, 2006 ; HEGSELMANN et KRAUSE, 2009) ou d'autres modèles d'épistémologie formelle ou de dynamique d'opinions. Dans de tels modèles, des agents forment une croyance à la fois par interaction directe avec le monde et sous l'influence de leurs pairs. Il s'agit de cadres formels très généraux qui peuvent être appliqués à beaucoup de cas concrets différents.

Dans le cas de la scientificité vue comme un processus d'optimisation, on

peut imaginer un groupe d'individus ayant chacun accès à des données différentes et échangeant éventuellement de l'information, qui recherchent collectivement le maximum d'une certaine fonction – qu'il s'agirait de définir. Or, se basant sur des données distinctes, ils ne percevraient pas forcément le maximum au même endroit. Ils pourraient être également guidés, dans leur recherche d'optimum, par des éléments non épistémiques comme le maintien d'une position de pouvoir ou l'influence d'une idéologie sous-jacente diffuse dans la société donnée. On pourrait imaginer également qu'un excès de conservatisme, par exemple, pourrait les bloquer collectivement sur un maximum local, les empêchant de découvrir un meilleur optimum.

Quoiqu'il en soit, cela pourrait être une manière simple et éclairante de modéliser les *controverses scientifiques*. Une question qui semble cruciale et qui traverse la philosophie, la sociologie et l'histoire des sciences est notamment celle de la résolution de ces controverses. En particulier, une question centrale consiste à déterminer dans quelle mesure une controverse est résolue par l'exercice d'une rationalité collective outrepassant les différents biais culturels et individuels, ou bien si toute résolution de controverse n'est que le reflet des rapports de pouvoir présents à un moment donné dans la communauté scientifique en question.

Cette perspective serait une bonne occasion d'utiliser la vision épistémologique théorique que j'ai développée dans cette thèse à des fins pratiques, dans des modèles plus complexes faisant intervenir des processus collectifs, en lien étroit avec des questions et des méthodes fondamentales de la sociologie et de l'histoire des sciences.

Bibliographie

- ADORNO, Theodor W. (1976). *The Positivist dispute in German sociology*. Heinenmann, p. 307.
- (1979). *De Vienne à Francfort. la querelle allemande des sciences sociales*. Complexe, p. 278.
- AKAIKE, Hirotogu (1998). « Information Theory and an Extension of the Maximum Likelihood Principle ». In : *Selected Papers of Hirotugu Akaike*. Springer New York, p. 199-213. ISBN : 9781461216940. DOI : 10.1007/978-1-4612-1694-0_15.
- AMALDI, Edoardo (1984). « From the discovery of the neutron to the discovery of nuclear fission ». In : *Physics Reports* 111.1, p. 1 -331.
- AMENDOLA, Luca et al. (avr. 2018). « Cosmology and fundamental physics with the Euclid satellite ». In : *Living Reviews in Relativity* 21.1. ISSN : 1433-8351. DOI : 10.1007/s41114-017-0010-3.
- ARNOWITT, Richard, Stanley DESER et Charles W. MISNER (2008). « Republication of : The dynamics of general relativity ». In : *General Relativity and Gravitation* 40.9, p. 1997-2027. DOI : 10.1007/s10714-008-0661-1.
- BARBERIS, Nicholas C (2013). « Thirty Years of Prospect Theory in Economics : A Review and Assessment ». In : *Journal of Economic Perspectives* 27.1, p. 173-196. DOI : 10.1257/jep.27.1.173.
- BECKER, Gary S. (1968). « Crime and Punishment : An Economic Approach ». In : *Journal of Political Economy* 76.2, p. 169-217.
- BECKER, Rolf (2022). « Explaining educational differentials revisited : an evaluation of rigorous theoretical foundations and empirical findings ». In : *Handbook of Sociological Science*. Edward Elgar Publishing, p. 356-371. DOI : 10.4337/9781789909432.00029.
- BERTHELOT, Jean-Michel (1990). *L' intelligence du social. le pluralisme explicatif en sociologie*. Presses universitaires de France, p. 249. ISBN : 2130431844.
- BIGO, Didier et al. (déc. 2021). « Mimetic Rivalry : A Matrix of the War on Terror and its Discursive Strategies? » In : *Cultures & conflits* 123-124, p. 21-35. ISSN : 1777-5345. DOI : 10.4000/conflits.23159.
- BINMORE, Ken (2011). *Rational Decisions*. Princeton University Press. ISBN : 9780691149899.

- BLANCHARD, Marianne et Joanie CAYOUILLE-REMBLIÈRE (fév. 2017). « Penser les inégalités scolaires : quelques travaux contemporains en sociologie ». In : *Idées économiques et sociales* N° 187.1, p. 6-16. ISSN : 2257-5111. DOI : 10.3917/idee.187.0006.
- BLOOR, David (1976). *Knowledge and Social Imagery*. Chicago : University of Chicago Press.
- BOKULICH, Alisa (2009). « How scientific models can explain ». In : *Synthese* 180.1, p. 33-45. DOI : 10.1007/s11229-009-9565-1.
- (2017). « Models and Explanation ». In : *Springer Handbook of Model-Based Science*.
- BOUDON, Raymond (1973). *L'inégalité des chances*. Armand Colin.
- BOUDON, Raymond (1986). *Theories of social change. a critical appraisal*. University of California. ISBN : 0520057597.
- (oct. 1990). « Les causes de l'inégalité des chances scolaires ». In : *Commentaire* Numéro 51.3, p. 533-542. ISSN : 0180-8214. DOI : 10.3917/comm.051.0533.
- (1996). « The 'cognitivist model' : a generalized 'rational-choice model' ». In : *Rationality and Society* 8.2, p. 123-150. DOI : 10.1177/104346396008002001.
- (juin 2002). « À quoi sert la sociologie ? » In : *Cités* n° 10.2, p. 133-156. ISSN : 1299-5495. DOI : 10.3917/cite.010.0133.
- BOURDIEU, Pierre (1980). *Le sens pratique*. Editions de Minuit, p. 475. ISBN : 2707302988.
- BOURDIEU, Pierre et Jean-Claude PASSERON (1970). *La reproduction*. Les éditions de minuit.
- (1984). *Les Héritiers. Les étudiants et la culture*. Les Editions de Minuit, p. 192. ISBN : 9782707300812.
- BOUVIER, Alban (fév. 2008). « La théorie sociologique générale comme système hiérarchisé de modèles de portée intermédiaire ». In : *Revue européenne des sciences sociales* XLVI-140, p. 87-106. ISSN : 1663-4446. DOI : 10.4000/ress.171.
- (2023). « Middle-range Theories and the unification problem in social science. » In : sous la dir. de Stephen Turner ROBERT LEROUX Thierry Martin. *The Future of Sociology*, Routledge, p. 91-115.
- BOYD, Nora Mills et James BOGEN (2021). « Theory and Observation in Science ». In : *The Stanford Encyclopedia of Philosophy*. Sous la dir. d'Edward N. ZALTA. Winter 2021. Metaphysics Research Lab, Stanford University.

- BRADFORD HILL, Austin (1965). « The Environment and Disease : Association or Causation? » In : *Proceedings of the Royal Society of Medicine* 58, p. 295-300.
- BREEN, Richard et John H. GOLDTHORPE (1997). « Explaining Educational Differentials ». In : *Rationality and Society* 9.3, p. 275-305. DOI : 10.1177/104346397009003002.
- BRIGANDT, Ingo (juill. 2009). « The epistemic goal of a concept : accounting for the rationality of semantic change and variation ». In : *Synthese* 177.1, p. 19-40. ISSN : 1573-0964. DOI : 10.1007/s11229-009-9623-8.
- BROCK, William A. et Steven N. DURLAUF (1999). « A Formal Model of Theory Choice in Science ». In : *Economic Theory* 14.1, p. 113-130. ISSN : 09382259, 14320479.
- BUNGE, Mario (1983a). *Treatise on Basic Philosophy, vol. 5 : Epistemology and Methodology I : exploring the world*. Sous la dir. de D. Reidel : DORDRECHT.
- (1991). « What is science? Does it matter to distinguish it from pseudoscience? A reply to my commentators ». In : *New Ideas in Psychology* 9.2, p. 245-283. ISSN : 0732-118X. DOI : [https://doi.org/10.1016/0732-118X\(91\)90030-P](https://doi.org/10.1016/0732-118X(91)90030-P). URL : <https://www.sciencedirect.com/science/article/pii/0732118X9190030P>.
- BUXTON, Richard B. (2010). « Interpreting oxygenation-based neuroimaging signals : the importance and the challenge of understanding brain oxygen metabolism ». In : *Frontiers in Neuroenergetics*. ISSN : 1662-6427. DOI : 10.3389/fnene.2010.00008.
- CAMPBELL, Donald T. (avr. 1990). « The Meehilan Corroboration-Verisimilitude Theory of Science ». In : *Psychological Inquiry* 1.2, p. 142-147. ISSN : 1532-7965. DOI : 10.1207/s15327965pli0102_2.
- CARRANANTE, Géraldine (2023). « How to deal with a plurality of scientific concepts? » In : *Séminaire Jeunes Chercheurs de Philosophie Argumentative*. Lyon, France. URL : <https://hal.science/hal-04405389>.
- CARTWRIGHT, Nancy (1983). *How the Laws of Physics Lie*. Clarendon paperbacks. Clarendon Press. ISBN : 9780198247043.
- CHANG, Hasok (2004). *Inventing Temperature : Measurement and Scientific Progress*. Oup Usa.
- CHATTERJEE, Arnab, Marija MITROVIĆ et Santo FORTUNATO (2013). « Universality in voting behavior : an empirical analysis ». In : *Scientific Reports* 3.1. DOI : 10.1038/srep01049.
- COBB, Charles W. et Paul H. DOUGLAS (1928). « A Theory of Production ». In : *The American Economic Review* 18.1, p. 139-165.

- COLEMAN, James, Elihu KATZ et Herbert MENZEL (déc. 1957). « The Diffusion of an Innovation Among Physicians ». In : *Sociometry* 20.4, p. 253. ISSN : 0038-0431. DOI : 10.2307/2785979.
- COLEMAN, James S. (1987). « Microfoundations and Macrosocial Behavior ». In : *The micro-macro link*. Sous la dir. de Jeffrey C. ALEXANDER et al. University of California Press, p. 153-176.
- (1990). *Foundations of social theory*. Harvard University Press.
- COLEMAN, James S. et Thomas J. FARRARO (1992). *Rational Choice Theory. Advocacy and Critique (Key Issues in Sociological Theory)*. Sage Publications, Inc, p. 256. ISBN : 9780803947610.
- COLLARD-WEXLER, Simon, Costantino PISCHEDDA et Michael G. SMITH (2014). « Do Foreign Occupations Cause Suicide Attacks? » In : *The Journal of Conflict Resolution* 58.4, p. 625-657. (Visité le 19/02/2024).
- COMTE, Auguste (1830). *Cours de philosophie positive*. Sous la dir. de Paris ROUEN FRÈRES Bachelier.
- CONDORCET, Jean-Antoine-Nicolas de Caritat marquis de (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Sous la dir. d'Imprimerie royale (PARIS).
- CONTESSA, Gabriele (jan. 2007). « Scientific Representation, Interpretation, and Surrogate Reasoning ». In : *Philosophy of Science* 74.1, p. 48-68. ISSN : 1539-767X. DOI : 10.1086/519478.
- COWAN, Clyde L. et al. (1956). « Detection of the Free Neutrino : a Confirmation ». In : *Science* 124.3212, p. 103-104.
- CUIN, Charles-Henry (2000). *Ce que (ne) font (pas) les sociologues. petit essai d'épistémologie critique*. Droz, p. 214. ISBN : 9782600004268.
- CUIN, Charles-Henry, François GRESLE et Ronan HERVOUET (2017). *Histoire de la sociologie*. La Découverte. ISBN : 9782707179166. DOI : 10.3917/dec.cuinc.2017.01.
- CZERNY, Božena et al. (jan. 2018). « Astronomical Distance Determination in the Space Age : Secondary Distance Indicators ». In : *Space Science Reviews* 214.1. ISSN : 1572-9672. DOI : 10.1007/s11214-018-0466-9.
- DANJON, André (1946). « Le centenaire de la découverte de Neptune ». In : *Ciel et Terre* 62, p. 369.
- DELLAVIGNA, Stefano (2009). « Psychology and Economics : Evidence from the Field ». In : *Journal of Economic Literature* 47.2, p. 315-372. DOI : 10.1257/jel.47.2.315.
- DELORI, Mathias et al. (déc. 2021). « Les interventions militaires sont-elles une cause du « terrorisme » ? : De l'utilité des analyses quantitatives pour

- les études critiques de la sécurité ». In : *Cultures & conflits* 123–124, p. 37-65. ISSN : 1777-5345. DOI : 10.4000/conflits.23183.
- DEMEULENAERE, Pierre (2011). *Analytical Sociology and Social Mechanisms*. Cambridge University Press. ISBN : 9781139497961.
- DESROSIÈRES, Alain (2014). *Prouver et gouverner. une analyse politique des statistiques publiques*. La Découverte, p. 284. ISBN : 9782707178954.
- DIETRICH, Franz et Christian LIST (2013). « A Reason-Based Theory of Rational Choice ». In : *Noûs* 47.1, p. 104-134. ISSN : 00294624, 14680068. URL : <http://www.jstor.org/stable/43828819> (visité le 27/06/2023).
- (2016). « Reason-Based Choice and Context-dependence : an Explanatory Framework ». In : *Economics and Philosophy* 32.2, p. 175-229. DOI : 10.1017/s0266267115000474.
- DILTHEY, Wilhelm (1989). *Introduction to the human sciences*. Princeton University Press, p. 524. ISBN : 0691073074.
- DUBOIS, Michel (2001). *La nouvelle sociologie des sciences*. Presses universitaires de France, p. 256. ISBN : 2130514324.
- DUHEM, Pierre (1906). *La théorie physique : son objet, sa structure*. Chevalier et Rivière.
- DURAND, François et al. (nov. 2015). « Geometry on the Utility Space ». In : p. 189-204.
- DURLAUF, Steven N. (2001). « A Framework For The Study of Individual Behavior and Social Interactions ». In : *Sociological Methodology* 31.1, p. 47-87. DOI : 10.1111/0081-1750.00089.
- EGRÉ, Paul et Cathal O'MADAGAIN (2019). « Concept Utility ». In : *The Journal of Philosophy* 116.10, p. 525-554. ISSN : 0022-362X. DOI : 10.5840/jphil20191161034.
- FABIANI, Jean-Louis (2013). « Faire son choix théorique en sciences sociales ». In : *Devenir chercheur*. Éditions de l'École des hautes études en sciences sociales, p. 47-62. ISBN : 9782713225963. DOI : 10.4000/books.editionsehess.1291.
- FALK, Carl F. et Michael MUTHUKRISHNA (2021). « Parsimony in model selection : Tools for assessing fit propensity. » In : URL : <https://doi.org/10.1037/met0000422>.
- FEIGL, Herbert (1970). « The "orthodox" view of theories : remarks in defense as well as critique ». In : *Analyses of theories and methods of physics and psychology. Minnesota studies in the philosophy of science*. Sous la dir. d'University of MINNESOTA PRESS. T. 4, p. 3-16.

- FELIPE, Jesus et John S.L. MCCOMBIE (2013). *The Aggregate Production Function and the Measurement of Technical Change : 'not Even Wrong'*. Edward Elgar.
- FERNANDEZ-BEANATO, Damian (2020). « The Multicriterial Approach to the Problem of Demarcation ». In : *Journal for General Philosophy of Science / Zeitschrift für Allgemeine Wissenschaftstheorie* 51.3, p. 375-390. DOI : 10 . 1007/s10838-020-09507-5.
- FERNANDEZ-BEANATO, Damian (2021). « Feng Shui and the Demarcation Project ». In : *Sci & Educ*, p. 1333-1351.
- (2022). « A working scientific demarcation ». Thèse de doct. University of Bristol.
- FEYERABEND, Paul (1975). *Against Method*. London : New Left Books, p. 106. DOI : 10 . 2307/2184355.
- FORTUNATO, Santo et Claudio CASTELLANO (2007). « Scaling and Universality in Proportional Elections ». In : *Physical Review Letters* 99.13, p. 138701. DOI : 10 . 1103/physrevlett . 99 . 138701.
- FORUM, World Economic (2022). *Global Gender Gap Report 2022*. Geneva, Switzerland.
- FRAASSEN, Bas C. van (1980). *The Scientific Image*. Clarendon Library of Logic and Philosophy. Clarendon Press.
- FRIEDMAN, Milton (1953). « The Methodology of Positive Economics ». In : *Essays in Positive Economics*. Sous la dir. de Milton FRIEDMAN. University of Chicago Press, p. 3-43.
- FRIEDMAN, Daniel (2017). *Risky Curves On the Empirical Failure of Expected Utility. On the Empirical Failure of Expected Utility*. Taylor & Francis Group, p. 152. ISBN : 9781138096462.
- FRIGG, Roman (2009). « Models and fiction ». In : *Synthese* 172.2, p. 251-268. DOI : 10 . 1007/s11229-009-9505-0.
- GABAIX, Xavier (1999). « Zipf's Law for Cities : An Explanation ». In : *The Quarterly Journal of Economics* 114.3, p. 739-767.
- GIERE, Ronald N. (1999). *Science without laws*. University of Chicago Press, p. 285. ISBN : 0226292088.
- GINTIS, Herbert (2009). *The bounds of reason. game theory and the unification of the behavioral sciences*. Princeton University Press. ISBN : 9780691140520.
- GIRARD, Alain (1970). *Population et l'enseignement*. PUF.
- GOLDTHORPE, John H. (sept. 1996). « Class analysis and the reorientation of class theory : the case of persisting differentials in educational attainment ». In : *The british journal of sociology* 47.3.

- (1998). « Rational Action Theory for Sociology ». In : *The British Journal of Sociology* 49.2, p. 167. DOI : 10.2307/591308.
- GRANOVETTER, Mark (mai 1978). « Threshold Models of Collective Behavior ». In : *American Journal of Sociology* 83.6, p. 1420-1443. ISSN : 1537-5390. DOI : 10.1086/226707.
- GÖCKELER, Meinulf et Thomas SCHÜCKER (2011). *Differential Geometry, Gauge Theories, and Gravity*. Cambridge University Press. ISBN : 9780511628818.
- HAILE, Philip A, Ali HORTAÇSU et Grigory KOSENOK (2008). « On the Empirical Content of Quantal Response Equilibrium ». In : *American Economic Review* 98.1, p. 180-200. DOI : 10.1257/aer.98.1.180.
- HANSON, Norwood (1958). *Patterns of discovery*. Cambridge University Press.
- HANSSON, Sven Ove (2021). « Science and pseudo-science ». In : sous la dir. de Stanford Encyclopedia of PHILOSOPHY.
- HARDING, Sandra (1975). *Can Theories Be Refuted? : Essays on the Duhem-Quine Thesis*. Reidel.
- HECHTER, Michael et Satoshi KANAZAWA (2019). « Sociological Rational Choice Theory ». In : *Rational Choice Sociology*. Edward Elgar Publishing. DOI : 10.4337/9781789903256.00007.
- HEDSTROM, Peter (2005). *Dissecting the Social*. Cambridge University Press. DOI : 10.1017/cbo9780511488801.
- (2021). « The past and future of analytical sociology ». In : Institute for Analytical Sociology. Linköping, Sweden. URL : https://www.youtube.com/watch?v=U9aXennCykQ&ab_channel=InstituteforAnalyticalSociology.
- HEDSTRÖM, Peter (2021). « Coda - The past and future of analytical sociology ». In : *Research Handbook on Analytical Sociology*. Edward Elgar Publishing. DOI : 10.4337/9781789906851.00035.
- HEDSTRÖM, Peter et P. BEARMAN (2009). *The Oxford handbook of analytical sociology*. Oxford University Press.
- HEDSTRÖM, Peter et Richard SWEDBERG (1998). *Social Mechanisms. An Analytical Approach to Social Theory (Studies in Rationality and Social Change)*. Cambridge University Press, p. 352. ISBN : 9780521593199.
- HEDSTRÖM, Peter et Petri YLIKOSKI (juin 2010). « Causal Mechanisms in the Social Sciences ». In : *Annual Review of Sociology* 36.1, p. 49-67. ISSN : 1545-2115. DOI : 10.1146/annurev.soc.012809.102632.
- HEGSELMANN, Rainer et Ulrich KRAUSE (2002). « Opinion Dynamics and Bounded Confidence, Models, Analysis and Simulation ». In : *Journal of*

- Artificial Societies and Social Simulation* 5.3, p. 2. URL : <http://jasss.soc.surrey.ac.uk/5/3/2.html>.
- HEGSELMANN, Rainer et Ulrich KRAUSE (2006). « Truth and Cognitive Division of Labour : First Steps Towards a Computer Aided Social Epistemology ». In : *Journal of Artificial Societies and Social Simulation* 9.3, p. 1-10. URL : <https://ideas.repec.org/a/jas/jasssj/2005-45-2.html>.
- (2009). « Deliberative Exchange, Truth, and Cognitive Division of Labour : A Low-Resolution Modeling Approach ». In : *Episteme* 6.2, p. 130-144. DOI : 10.3366/e1742360009000604.
- HEMPEL, Carl G. (1951). « The Concept of Cognitive Significance : A Reconsideration ». In : *Proceedings of the American Academy of Arts and Sciences* 80.1, p. 61. ISSN : 0199-9818. DOI : 10.2307/20023635.
- HERNES, Gudmund (1992). « We Are Smarter than We Think ». In : *Rationality and Society* 4.4, p. 421-436. DOI : 10.1177/1043463192004004005.
- HESSE, Mary B. (1963). *Models and Analogies in Science*. University of Notre Dame Press.
- HUBER, Franz (2008). « The Plausibility-Informativeness Theory ». In : *New Waves in Epistemology*. Sous la dir. de Vincent HENDRICKS. Palgrave-Macmillan.
- HUME, David (1748). *Enquiry Concerning Human Understanding*. London : A. Millar.
- JASSO, Guillermina (1988). « Principles of Theoretical Analysis ». In : *Sociological Theory* 6.1, p. 1. DOI : 10.2307/201910.
- KRASNOV, Kirill (2020). *Formulations of General Relativity Gravity, Spinors and Differential Forms. Gravity, Spinors and Differential Forms*. Cambridge University Press. ISBN : 9781108674652.
- KRONEBERG, Clemens et Frank KALTER (2012). « Rational Choice Theory and Empirical Research : Methodological and Theoretical Contributions in Europe ». In : *Annual Review of Sociology* 38.1, p. 73-92. DOI : 10.1146/annurev-soc-071811-145441.
- KRUIS, Joost et al. (2020). « Deviations of rational choice : an integrative explanation of the endowment and several context effects ». In : *Scientific Reports* 10.1. DOI : 10.1038/s41598-020-73181-2.
- KUHN, Thomas (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
- (1977). *The Essential Tension : Selected Studies in Scientific Tradition and Change*. Philosophy of science. University of Chicago Press.

- LAHIRE, Bernard (2004). *À quoi sert la sociologie ?* La Découverte, p. 193. ISBN : 2707144215.
- (2023). *Les structures fondamentales des sociétés humaines*. Sciences sociales du vivant. La Découverte. ISBN : 9782348077722. URL : <https://books.google.fr/books?id=7rbIEAAAQBAJ>.
- LAKATOS, Imre (mai 1978). *The Methodology of Scientific Research Programmes : Philosophical Papers*. Sous la dir. de John WORRALL et Gregory CURRIE. Cambridge University Press. ISBN : 9780511621123.
- LATOURET, Bruno et Steve WOOLGAR (2006). *La vie de laboratoire*. LA DECOUVERTE, p. 308. ISBN : 9782707148483.
- LAUDAN, Larry (1983). « The Demise of the Demarcation Problem ». In : *Physics, Philosophy and Psychoanalysis : Essays in Honour of Adolf Grünbaum*. Sous la dir. de R. S. COHEN et L. LAUDAN. Dordrecht : Springer Netherlands, p. 111-127.
- LE VERRIER, Urbain J. (oct. 1846). « Recherches sur les mouvements d'Uranus par U. J. Le Verrier (Fortsetzung) ». In : *Astronomische Nachrichten* 25, p. 65.
- LEVENSON, Thomas (2015). *The Hunt for Vulcan : How Albert Einstein Destroyed a Planet and Deciphered the Universe*. Head of Zeus. ISBN : 9781784973964.
- LEWIS, David (1973). *Counterfactuals*. Blackwell.
- (1983). « New work for a theory of universals ». In : *Australasian Journal of Philosophy* 61.4, p. 343-377.
- (1994). « Humean Supervenience Debugged ». In : *Mind* 103.412, p. 473-490.
- LIDDLE, Andrew R. (2000). *Cosmological inflation and large-scale structure*. Cambridge University Press, p. 400. ISBN : 052166022X.
- LONGINO, Helen E. (mars 1979). « Evidence and Hypothesis : An Analysis of Evidential Relations ». In : *Philosophy of Science* 46.1, p. 35-56. ISSN : 1539-767X. DOI : 10.1086/288849.
- MAHNER, Martin (déc. 2007). « Demarcating Science from Non-Science ». In : *General Philosophy of Science : Focal Issues*, p. 515-575. DOI : 10.1016/B978-044451548-3/50011-2.
- (2013). « Science and Pseudoscience ». In : *Philosophy of Pseudoscience*. University of Chicago Press, p. 29-44. ISBN : 9780226051826. DOI : 10.7208/chicago/9780226051826.003.0003.
- MANZO, Gianluca (2013a). « Educational Choices and Social Interactions : A Formal Model and a Computational Test ». In : *Comparative Social Research*. Emerald Group Publishing Limited, p. 47-100. DOI : 10.1108/s0195-6310(2013)0000030007.

- MANZO, Gianluca (2013b). « Is rational choice theory still a rational choice of theory? A response to Opp ». In : *Social Science Information* 52.3, p. 361-382. DOI : 10.1177/0539018413488477.
- (2014a). *Analytical Sociology. Actions and Networks*. Wiley series in computational et quantitative social science.
- (2014b). *Data, generative models, and mechanisms : More on the principles of analytical sociology*. DOI : 10.1002/9781118762707.ch01.
- (2021). *Research Handbook in Analytical Sociology*. Edward Edgar Publishing.
- MANZO, Gianluca et al. (2018). « Complex Contagions and the Diffusion of Innovations : Evidence from a Small-N Study ». In : *Journal of Archaeological Method and Theory* 25.4, p. 1109-1154. DOI : 10.1007/s10816-018-9393-z.
- MAYO, Deborah G. (1996). *Error and the growth of experimental knowledge*. University of Chicago Press, p. 493. ISBN : 0226511979.
- MAYR, Ernst (2000). « Darwin's Influence on Modern Thought ». In : *Scientific American* 283.1, p. 78-83.
- MCKELVEY, Richard D. et Thomas R. PALFREY (1995). « Quantal Response Equilibria for Normal Form Games ». In : *Games and Economic Behavior* 10.1, p. 6-38. DOI : 10.1006/game.1995.1023.
- MEEHL, Paul E. (1990). « Appraising and Amending Theories : The Strategy of Lakatosian Defense and Two Principles That Warrant It ». In : *Psychological Inquiry* 1.2, p. 108-141.
- MERRITT, David (2017). « Cosmology and convention ». In : *Studies in History and Philosophy of Science Part B : Studies in History and Philosophy of Modern Physics* 57, p. 41 -52.
- MERTON, Robert K. (1948). « The Self-Fulfilling Prophecy ». In : *The Antioch Review* 8.2, p. 193. ISSN : 0003-5769. DOI : 10.2307/4609267.
- (1968). *Social Theory and Social Structure*. Free Press, p. 702. ISBN : 9780029211304.
- MILLER, Dale T. (oct. 2023). « A century of pluralistic ignorance : what we have learned about its origins, forms, and consequences ». In : *Frontiers in Social Psychology* 1. ISSN : 2813-7876. DOI : 10.3389/frsps.2023.1260896.
- MILL, John Stuart (1866). *Système de logique déductive et inductive : exposé des principes de la preuve et des méthodes de recherche scientifique*. vol. 1. Librairie Philosophique de Ladrangé.
- MITCHELL, Peter D. et al. (2013). « How well can we really estimate the stellar masses of galaxies from broad-band photometry? » In : *Monthly Notices of the Royal Astronomical Society* 435.1, p. 87-114.

- MOSCATI, Ivan (2023). « Behavioural and heuristic models are as-if models too – and that’s ok ». In : *Economics and Philosophy*, p. 1-31. DOI : 10 . 1017/s0266267123000093.
- MUCCHIELLI, Laurent (juin 2008). « Le « nouveau management de la sécurité » à l’épreuve : délinquance et activité policière sous le ministère Sarkozy (2002-2007) ». In : *Champ pénal* Vol. V. ISSN : 1777-5272. DOI : 10 . 4000/champpenal . 3663.
- NEUMANN, John Von et Oskar MORGENSTERN (2007). *Theory of Games and Economic Behavior (Commemorative Edition) (Princeton Classic Editions)*. Princeton University Press, p. 776. ISBN : 9780691130613.
- OLSON, Mancur (1971). *The logic of collective action. public goods and the theory of groups*. Harvard University Press, p. 186. ISBN : 0674537513.
- OPP, Karl-Dieter (2013). « What is Analytical Sociology ? Strengths and weaknesses of a new sociological research program ». In : *Social Science Information* 52.3, p. 329-360. DOI : 10 . 1177/0539018413483939.
- PAGE, Lionel (2022). *Optimally Irrational*. Cambridge University Press. DOI : 10.1017/9781009209175.
- PASSERON, Jean-Claude (1991). *Le raisonnement sociologique. L’espace non-poppérien du raisonnement naturel*. Paris, Nathan.
- (2006). *Le raisonnement sociologique un espace non poppérien de l’argumentation. un espace non poppérien de l’argumentation*. Albin Michel, p. 666. ISBN : 9782226158895.
- PERUZZI, Edoardo et Gustavo CEVOLANI (2021). « Defending De-idealization in Economic Modeling : A Case Study ». In : *Philosophy of the Social Sciences* 52.1-2, p. 25-52. DOI : 10 . 1177/004839312111049759.
- PESTRE, Dominique (juin 2006). *Introduction aux Science Studies*. La Découverte. ISBN : 9782707145963. DOI : 10 . 3917/dec . pestr . 2006 . 01.
- PIGLIUCCI, Massimo et Maarten BOUDRY (2013). *Philosophy of Pseudoscience : Reconsidering the Demarcation Problem*. University of Chicago Press.
- POINCARÉ, Henri (1902). *La science et l’hypothèse*. Flammarion.
- (2017). *Science and Hypothesis*. Bloomsbury Publishing.
- POPPER, Karl (1934). *Logik der Forschung*. Julius Springer.
- (1959). *The Logic of Scientific Discovery*. Hutchinson et Co.
- (1962). *Conjectures and Refutations : The Growth of Scientific Knowledge*. London, England : Routledge.
- (1972). *Objective Knowledge*. Clarendon Press - Oxford.
- PSILLOS, Stathis (1999). *Scientific realism. how science tracks truth*. Routledge, p. 341. ISBN : 0415208181.

- QUINE, Willard V. O. (1951). « Two Dogmas of Empiricism ». In : *Philosophical Review* 60, p. 20-43.
- (1975). « On Empirically Equivalent Systems of the World ». In : *Erkenntnis* (1975-) 9.3, p. 313-328.
- RAYNAUD, Dominique (2005). « Le raisonnement expérimental en physique et en sociologie ». In : URL : <https://shs.hal.science/halshs-00005915>.
- (2006). *La Sociologie et sa vocation scientifique*. Hermann.
- (2018). *Sociologie des controverses scientifiques*. Editions Matériologiques. ISBN : 9782373611359. DOI : 10.3917/edmat.rayna.2018.01.
- (2019). « Le raisonnement expérimental en sociologie ». In : *Philosophia Scientiæ* 23.2, 19–45.
- (2021). *Sociologie fondamentale*. Sous la dir. d'Éditions MATERIOLOGIQUES.
- REICHENBACH, Hans (1949). « The Philosophical Significance of the Theory of Relativity ». In : *Defending Einstein : Hans Reichenbach's Writings on Space, Time and Motion*. Sous la dir. de Steven GIMBEL et Anke WALZ. Cambridge University Press, 95–160.
- RICHARDSON, Sarah S. et al. (2020). « Is There a Gender-Equality Paradox in Science, Technology, Engineering, and Math (STEM)? Commentary on the Study by Stoet and Geary (2018) ». In : *Psychological Science* 31.3, p. 338-341.
- RIESS, Adam (oct. 2020). « New Measurements of the Expansion Rate of the Universe ». In : *Astrofisica UC*. URL : https://www.youtube.com/watch?v=pSQ8tY7YNlU&ab_channel=AstrofisicaUC.
- RUSSO, Jorge G., Leonard SUSSKIND et Lárus THORLACIUS (oct. 1992). « Black hole evaporation in 1+1 dimensions ». In : *Physics Letters B* 292.1–2, p. 13-18. ISSN : 0370-2693. DOI : 10.1016/0370-2693(92)90601-y.
- RYAN, Matthew J. (2008). « Generalizations of SEU : a geometric tour of some non-standard models ». In : *Oxford Economic Papers* 61.2, p. 327-354. DOI : 10.1093/oep/gpn027.
- SALMON, Wesley C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press, p. 305. ISBN : 0691072930.
- SHELLING, Thomas C. (1978). *Schelling Micromotives and Macrobehavior (Cloth) (Fels lectures on public policy analysis)*. W W Norton Co Ltd, p. 252. ISBN : 9780393057010.
- SHECH, Elay (2023). *Idealizations in Physics*. Cambridge University Press. DOI : 10.1017/9781108946742.

- SIGNOL, Arnaud (2011). « La psychanalyse est-elle une science ? » In : *Le Journal des psychologues* 291.8, p. 54-58.
- SIMON, Herbert A. (1957). *Models of Man : Social and Rational ; Mathematical Essays on Rational Human Behavior in Society Setting*. Continuity in administrative science. Wiley. URL : https://books.google.fr/books?id=_wlgAAAAIAAJ.
- SIZZI, Francesco (1611). *Dianoia astronomica, optica, physica, qua Syderei Nuncij rumor de quatuor planetis à Galilaeo Galilaeo mathematico celeberrimo recens perspicillì cuiusdam ope conspectis, vanus redditur*. URL : <https://archive.org/details/ita-bnc-mag-00000877-001/mode/2up>.
- SKVORETZ, John (avr. 2016). « All for one and one for all : Theoretical models, sociological theory, and mathematical sociology ». In : *The Journal of Mathematical Sociology* 40.2, p. 71-79. ISSN : 1545-5874. DOI : 10.1080/0022250x.2016.1147446.
- SMELSER, Neil J. (1992). « The Rational Choice Perspective ». In : *Rationality and Society* 4.4, p. 381-410. DOI : 10.1177/1043463192004004003.
- SMOLIN, Lee (2007). *Rien ne va plus en physique ! l'échec de la théorie des cordes*. Dunod, p. 486. ISBN : 9782100507023.
- SOBER, Elliott (2015). *Ockham's Razors : A User's Manual*. Cambridge University Press.
- STEGENGA, Jacob (2018). *Medical Nihilism*. Oxford University Press, p. 256. ISBN : 9780198747048.
- STOET, Gijsbert et David C. GEARY (2018). « The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education ». In : *Psychological Science* 29.4, p. 581-593.
- STYER, Daniel F. et al. (2002). « Nine formulations of quantum mechanics ». In : *American Journal of Physics* 70.3, p. 288-297. DOI : 10.1119/1.1445404.
- SULLIVAN, Gail M. et Richard FEINN (2012). « Using Effect Size-or Why the P Value Is Not Enough. » In : *J Grad Med Educ*. 4.3, p. 279-82.
- SUPPE, Frederick (1974). *The Structure of Scientific Theories*. Urbana, University of Illinois Press.
- (2000). « Theory Identity ». In : *A Companion to the Philosophy of Science*. Sous la dir. de William H. NEWTON-SMITH. Wiley-Blackwell, p. 525-527.
- SUÁREZ, Mauricio (2009). *Fictions in science. philosophical essays on modeling and idealization*. Routledge. ISBN : 9780415990356.
- TESTART, Alain (1991). *Pour les sciences sociales. essai d'epistemologie*. C. Bourgois, p. 173. ISBN : 2267007134.

- TUTIĆ, Andreas (2017). « Revisiting the Breen–Goldthorpe Model of educational stratification ». In : *Rationality and Society* 29.4, p. 389-407. DOI : 10.1177/1043463117734177.
- TVERSKY, Amos et Daniel KAHNEMAN (1992). « Advances in prospect theory : Cumulative representation of uncertainty ». In : *Journal of Risk and Uncertainty* 5.4, p. 297-323. DOI : 10.1007/bf00122574.
- WHEELER, Billy (2018). *Idealization and the Laws of Nature*. Switzerland : Springer.
- (déc. 2019). « Compressibility and the Algorithmic Theory of Laws ». In : *Principia : an international journal of epistemology* 23.3, p. 461-485. ISSN : 1414-4247. DOI : 10.5007/1808-1711.2019v23n3p461.
- WHEELER, Gregory (2020). « Bounded Rationality ». In : *The Stanford Encyclopedia of Philosophy*. Sous la dir. d'Edward N. ZALTA. Fall 2020. Metaphysics Research Lab, Stanford University.
- WIPPLER, Reinhard (1978). « The Structural Individualistic Approach in Dutch Sociology. Towards an Explanatory Social Science ». In : *Netherlands (The) Journal of Sociology and Sociologia Neerlandica Amsterdam* 14.2, p. 135-155.
- WOIT, Peter (2007). *Not Even Wrong*. Vintage Books, p. 304. ISBN : 9780099488644.
- YLIKOSKI, Petri K (2021). « Understanding the Coleman boat ». In : *Research Handbook on Analytical Sociology*. Sous la dir. de Gianluca MANZO. Edward Elgar, p. 49-63.
- ZWICKY, Fritz (1933). « Die Rotverschiebung von extragalaktischen Nebeln ». In : *Helvetica Physica Acta* 6, p. 110-127.