



HAL
open science

La valeur de l'incertitude : l'évaluation de la précision des mesures physiques et les limites de la connaissance expérimentale

Fabien Grégis

► To cite this version:

Fabien Grégis. La valeur de l'incertitude : l'évaluation de la précision des mesures physiques et les limites de la connaissance expérimentale. Histoire, Philosophie et Sociologie des sciences. Université Paris Diderot (Paris 7), Sorbonne Paris Cité, 2016. Français. NNT : . tel-01553086v2

HAL Id: tel-01553086

<https://theses.hal.science/tel-01553086v2>

Submitted on 13 Jun 2018 (v2), last revised 10 Feb 2021 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

UNIVERSITÉ SORBONNE PARIS CITÉ
UNIVERSITÉ PARIS.DIDEROT (PARIS 7)

ÉCOLE DOCTORALE: Savoirs Scientifiques (ED 400)
Laboratoire SPHERE, UMR 7219

DOCTORAT
Histoire et philosophie des sciences

Fabien GRÉGIS

**La valeur de l'incertitude:
l'évaluation de la précision des mesures physiques et
les limites de la connaissance expérimentale**

Thèse dirigée par Nadine DE COURTENAY et Olivier DARRIGOL

Soutenue le vendredi 25 mars 2016

JURY

Mme Anouk BARBEROUSSE (Présidente, rapporteur)
Professeur à l'Université Paris 4

M. Michael HEIDELBERGER (Rapporteur)
Professeur émérite à l'Université de Tübingen

M. Marc HIMBERT
Professeur au Conservatoire national des arts et métiers

M. Giora HON
Professeur à l'Université de Haïfa

M. Luca MARI
Professeur à l'Université Carlo Cattaneo, Milan

Mme Nadine DE COURTENAY (Directrice de thèse)
Maître de conférences à l'Université Paris 7

M. Olivier DARRIGOL (Directeur de thèse)
Directeur de recherche au CNRS

**La valeur de l'incertitude:
l'évaluation de la précision des mesures physiques et
les limites de la connaissance expérimentale**

Fabien GRÉGIS

Table des matières

| | |
|---|------------|
| Table des matières | 5 |
| Introduction générale | 19 |
| 1 Préambule : erreur et incertitude | 31 |
| 1.1 Le problème de l'erreur | 31 |
| 1.2 L'erreur de mesure | 34 |
| 1.3 L'incertitude de mesure | 38 |
| 1.4 Modélisation de la mesure et fonction de mesure | 42 |
| 1.5 Mesure et métrologie | 44 |
| I L'interprétation probabiliste de l'incertitude de mesure | 47 |
| Vue d'ensemble de la partie I | 49 |
| 2 Fondements généraux de l'analyse d'incertitude | 51 |
| 2.1 La dichotomie entre erreurs aléatoires et systématiques | 52 |
| 2.2 Répétabilité et reproductibilité | 57 |
| 2.3 Typologie des erreurs et analyse d'erreur | 60 |
| 2.4 Erreurs de mesure et méthodes d'analyse d'incertitude | 62 |
| 3 L'approche fréquentiste de la mesure | 67 |
| 3.1 Fondement fréquentiste du traitement statistique des « erreurs aléatoires » | 70 |
| 3.2 Seconde hypothèse du modèle fréquentiste | 76 |
| 3.3 Estimateurs statistiques et troisième hypothèse de l'analyse fréquentiste | 82 |
| 3.4 Niveau de confiance et intervalle de confiance | 87 |
| 3.5 Traitement fréquentiste des erreurs systématiques | 94 |
| 3.6 Formulation du résultat final et limites de l'approche traditionnelle | 97 |
| 4 Des probabilités épistémiques à une modélisation bayésienne | 105 |
| 4.1 Les probabilités épistémiques | 107 |
| 4.2 L'approche du GUM et l'introduction des probabilités épistémiques | 110 |
| 4.3 Les méthodes de « type B » dans l'approche du GUM | 112 |

| | | |
|------------|---|------------|
| 4.4 | « Intervalles élargis » et critiques de l'approche du GUM | 116 |
| 4.5 | Vers une approche pleinement bayésienne de la mesure | 122 |
| 5 | Discussion : les ramifications philosophiques du débat statistique | 129 |
| 5.1 | Les deux modèles statistiques amènent-ils à des conclusions similaires? | 131 |
| 5.2 | Un tournant épistémique en métrologie | 136 |
| 5.3 | Objectivité et subjectivité de la mesure physique | 141 |
| 5.4 | De l'exactitude au taux de succès | 146 |
| 5.5 | Conclusion | 151 |
| | Conclusions de la première partie | 153 |
| II | La « valeur vraie » d'une grandeur | 157 |
| | Vue d'ensemble de la partie II | 159 |
| 6 | L'argument d'« inconnaissabilité » de la valeur vraie | 161 |
| 6.1 | Grandeur physique et valeur d'une grandeur | 163 |
| 6.2 | La critique de la valeur vraie dans le VIM et le GUM | 167 |
| 6.3 | L'usage de la valeur vraie dans la mesure | 172 |
| 6.4 | Questionnement philosophique : valeur « vraie » et réalisme scientifique . . . | 176 |
| 6.5 | La valeur vraie à la croisée de questionnements scientifiques et philosophiques | 187 |
| 7 | Valeur vraie : non unicité et incertitude définitionnelle | 195 |
| 7.1 | Non-unicité de la valeur vraie d'une grandeur | 196 |
| 7.2 | L'« incertitude définitionnelle » | 205 |
| 7.3 | « Non-unicité » et grandeurs physiques | 214 |
| 7.4 | Synthèse | 231 |
| | Conclusions de la seconde partie | 233 |
| III | Les ajustements des constantes de la physique | 235 |
| | Vue d'ensemble de la partie III | 237 |
| 8 | Combinaison des résultats et ajustements aux moindres carrés | 241 |
| 8.1 | La « combinaison des observations » | 242 |
| 8.2 | Mayer et Euler face à deux problèmes d'astronomie | 243 |
| 8.3 | Surdétermination des systèmes d'équations | 245 |
| 8.4 | Propagation des erreurs | 246 |
| 8.5 | Conditions d'observation et reproductibilité des mesures | 248 |
| 8.6 | Conclusion : la méthode des moindres carrés | 250 |

| | |
|---|------------|
| 9 L'ajustement des constantes de la physique : l'initiative de Birge | 255 |
| 9.1 L'idée initiale de Birge | 256 |
| 9.2 L'ajustement : principes généraux | 261 |
| 9.3 Procédure d'ajustement : l'approche de Birge | 265 |
| 9.4 Incertitude de mesure et évaluation d'un résultat | 271 |
| 10 Le traitement des données discordantes dans les ajustements | 275 |
| 10.1 Une tradition d'ajustements | 276 |
| 10.2 La critique des ajustements | 279 |
| 10.3 Deux attitudes épistémiques : précision contre exactitude | 286 |
| 10.4 Conclusion sur les ajustements | 294 |
| 11 Incertitude et exactitude | 299 |
| 11.1 Deux horizons insatisfaisants | 300 |
| 11.2 Ajustements et perspective objectiviste | 302 |
| 11.3 Ajustements et perspective épistémique | 306 |
| 11.4 Les neutrinos « supraluminiques » | 309 |
| 11.5 Perspectives de recherche et perspectives appliquées | 313 |
| Conclusions de la troisième partie | 315 |
| IV Épilogue : la genèse du GUM | 317 |
| 12 Épilogue : la genèse du GUM | 319 |
| 12.1 De la lettre d'Ambler à la « recommandation INC-1 » | 320 |
| 12.2 L'ISO prend la main | 324 |
| 12.3 La création du JCGM | 327 |
| 12.4 Assise institutionnelle de l'incertitude de mesure | 328 |
| Conclusion générale | 335 |
| Annexes | 341 |
| A Développements techniques sur les méthodes statistiques | 343 |
| A.1 Estimateurs statistiques | 343 |
| A.2 Intervalles de confiance | 345 |
| A.3 Traitement bayésien des méthodes de type A | 347 |
| A.4 Randomisation des erreurs systématiques : un exemple | 351 |
| A.5 Combinaison de plusieurs résultats | 356 |
| Bibliographie | 359 |
| Index des noms propres | 379 |

Table des figures

| | | |
|-----|--|-----|
| 1.1 | Principe de l'étude de Bruno Suchaut sur la notation des copies au baccalauréat. | 32 |
| 1.2 | Métaphore des cibles | 37 |
| 1.3 | Illustration du principe de fonctionnement du système de technologie sur la ligne de but | 40 |
| 1.4 | Extrait du tableau récapitulatif des valeurs recommandées de 2010 des constantes de la physique publiées par le CODATA | 41 |
| 1.5 | Mesure de la longueur d'un calibre à bouts | 42 |
| 2.1 | Illustration de la dichotomie entre erreurs aléatoires et systématiques à partir de la métaphore des cibles | 54 |
| 2.2 | Exemple de diagramme dit d'Ishikawa | 62 |
| 2.3 | Métaphore des cibles et analyse des sources d'incertitude | 64 |
| 3.1 | Illustration d'une distribution de probabilité de la variable aléatoire représentant un processus de mesure | 77 |
| 3.2 | Illustration de l'erreur affectant une donnée expérimentale | 79 |
| 3.3 | Illustration de l'interprétation fréquentiste d'un intervalle de confiance. | 92 |
| 4.1 | Illustration de la différence entre probabilité fréquentiste et probabilité épistémique sur un modèle d'urnes. | 108 |
| 4.2 | Distribution de probabilité associée à une variable aléatoire décrivant l'état de connaissance disponible à propos d'une grandeur | 114 |
| 4.3 | Deux exemples d'intervalles élargis à 68% pour une même loi normale centrée réduite. | 117 |
| 4.4 | Produit de convolution d'une fonction rectangulaire et d'une fonction gaussienne | 126 |
| 4.5 | Exemple d'intervalle de crédibilité et d'incertitude élargie | 127 |
| 5.1 | Illustration sur un exemple de l'influence du choix de la distribution de probabilité <i>a priori</i> sur la distribution <i>a posteriori</i> obtenue, dans le cas des méthodes bayésiennes de type A. | 133 |
| 5.2 | Illustration du « taux de succès » d'une procédure. | 147 |
| 7.1 | Figure reproduite à partir du GUM (figure D.2, p.55), intitulée « Illustration graphique des valeurs, de l'erreur et de l'incertitude » | 212 |

| | | |
|------|--|-----|
| 7.2 | Modèle du pendule simple | 220 |
| 7.3 | Interprétation de l'incertitude définitionnelle à l'aide de la métaphore des cibles | 229 |
| 8.1 | Illustration de la méthode des moindres carrés sur un exemple fictif arbitraire | 253 |
| 9.1 | Mesures de la vitesse de la lumière dans le vide c recensées par Birge en 1929 | 267 |
| 9.2 | Mesures de la constante de Planck h recensées par Birge en 1929 | 269 |
| 10.1 | Reproduction de deux graphes tirés de Taylor <i>et al.</i> (1969) illustrant l'évolution de différentes constantes physiques depuis 1950 | 284 |
| 10.2 | Mesures de la constante de structure fine en 1965 | 293 |
| 10.3 | Mesures du rayon du proton à partir de principes de mesure orientés autour de l'hydrogène d'une part et de l'« hydrogène muonique » d'autre part, recensées lors de l'ajustement du CODATA de 2014 | 296 |
| A.1 | Mesure ordinaire de la longueur d'un objet à l'aide d'une règle graduée | 352 |
| A.2 | Méthode « randomisée » d'une mesure de longueur à l'aide d'une règle graduée | 353 |
| A.3 | Distribution de probabilité d'une erreur systématique « randomisée » | 354 |

Liste des tableaux

| | | |
|-----|---|-----|
| 2.1 | Différentes définitions d'« erreur aléatoire » et « erreur systématique » dans les trois éditions du <i>Vocabulaire International de Métrologie (VIM)</i> | 56 |
| 2.2 | Définitions de « Conditions de répétabilité » et « conditions de reproductibilité » dans la troisième édition du <i>Vocabulaire International de Métrologie (VIM)</i> | 57 |
| 3.1 | Trois paramètres statistiques remarquables d'une variable aléatoire : espérance, variance et écart-type. | 77 |
| 5.1 | Différentes définitions d'« incertitude de mesure » dans les trois éditions du <i>Vocabulaire International de Métrologie (VIM)</i> | 140 |
| 6.1 | Circularité soulignée par Ehrlich <i>et al.</i> , caractéristique de l'argument d'« inconnaisabilité ». | 169 |
| 6.2 | Différentes définitions de « valeur conventionnellement vraie » dans les deux premières éditions du <i>Vocabulaire International de Métrologie (VIM)</i> et de « valeur conventionnelle » dans la troisième édition du VIM. | 185 |
| 6.3 | « Erreur de mesure » dans la troisième édition du <i>Vocabulaire International de Métrologie (VIM)</i> | 186 |
| 6.4 | « Valeur de référence » dans la troisième édition du <i>Vocabulaire International de Métrologie (VIM)</i> | 186 |
| 7.1 | Différentes définitions de « mesurande » dans les trois éditions du <i>Vocabulaire International de Métrologie (VIM)</i> | 197 |
| 7.2 | Exemples de mesurandes incomplètement définis, ainsi que des facteurs possibles à l'origine de l'incomplétude de ces définitions, et des exemples de mesurandes définis de de façon plus précise | 201 |
| 7.3 | Essai de classification des grandeurs selon un critère double. | 219 |
| 8.1 | Équations de conditions rencontrées par Euler et Mayer | 245 |
| A.1 | Tableau de valeur des quantiles de la distribution de Student | 348 |

Remerciements

Mes premières pensées vont à ma directrice de thèse, Nadine de Courtenay, qui m'a accompagné et guidé dans mon travail depuis maintenant sept ans – en incluant les travaux effectués dans le cadre du M2 LOPHISS du département d'Histoire et philosophie des sciences de l'Université Paris 7. Ce fut un voyage à ses côtés que j'ai grandement apprécié sur le plan professionnel tout autant que sur le plan humain. Nadine a fait preuve de beaucoup de patience à mon égard, mais également de beaucoup de bienveillance, d'exigence et d'abnégation, autant de qualités qui en font une formidable directrice, et qui m'ont permis d'achever cette thèse avec un sentiment de satisfaction. J'espère avoir pu lui rendre ne serait-ce qu'une fraction de ce que son suivi, son aide, ses conseils et ses remarques avisées m'ont apporté durant toutes ces années. Je remercie également Olivier Darrigol d'avoir accepté de co-diriger mon travail ; ses remarques ont toujours visé particulièrement juste et m'ont offert la possibilité de renouveler sans cesse ma compréhension de mon propre sujet de recherche.

Je remercie Anouk Barberousse, Michael Heidelberger, Marc Himbert, Giora Hon et Luca Mari d'avoir accepté de faire partie de mon jury de thèse.

Je remercie le laboratoire SPHERE de m'avoir accueilli en son sein et de m'avoir fourni un support logistique inestimable – l'« open space » des doctorants permettant de mettre en place une véritable communauté épistémique qui stimule la curiosité et la réflexion beaucoup plus que ne le permettrait le fait de travailler seul. Je suis reconnaissant également à l'École Normale Supérieure de Cachan qui m'a permis d'obtenir le financement du Ministère de l'Enseignement Supérieur et de la Recherche, lequel m'a couvert durant les trois premières années de cette thèse ; ainsi que l'Université Paris Diderot qui m'a recruté comme moniteur puis ATER durant les nombreuses années de mon travail.

Mes pensées vont de plus à l'équipe administrative du département HPS et du laboratoire SPHERE : Françoise Contamina, Morgane Paquereau, Sandrine Pellé, Virginie Maouchi, Patricia Philippe. Je remercie tout particulièrement Nad Fachard pour sa présence, son aide et son soutien durant toutes ces années.

Bien sûr, je souhaite également témoigner d'une affection très profonde envers mes camarades de fortune ou d'infortune (les deux facettes de la même pièce qu'est le doctorat) qui ont partagé avec moi un statut – celui de doctorant – et un lieu de travail – l'open space du bâtiment Condorcet – durant toutes ces années. Je pense aux co-organisateurs du groupe de travail des doctorants en histoire et philosophie de la physique, organisé de 2012 à 2015 : Vincent Dau-

don, Jonathan Regier et Philippe Stamenkovic. Ma gratitude va également à Aurélien Féron et Quentin Lade, qui ont longuement travaillé avec moi sur la préparation de cours divers, ce qui a été l'occasion d'un très grand enrichissement intellectuel et surtout d'un grand plaisir partagé sur des thématiques qui se sont révélées passionnantes. Je pense aussi à tous les autres collègues doctorants ou post-doctorants du laboratoire dont j'ai pu faire la connaissance : Nacera Bensaou, Pascal Bertin (que je remercie tout particulièrement pour son aide concernant la rédaction de mon introduction), Sylvain Cabanacq, Christine Cachot, Pierre Chaigneau, João Cortese, Simon Decaens, Fernando Galvez, Roberto Hastenreiter-Cruz, Emmylou Haffner, Zeinab Karimian, Ramzi Kebaili, Julien Page, Mattia Petrolo, Eleonora Sammarchi, Sergio Valencia, Xiaofei Wang, Federico Zalamea, Felix Zheng, et tous ceux que j'aurais malheureusement oubliés dans le flux et le reflux des effectifs... Mes pensées se tournent enfin vers mes compagnons de correction des copies de médecine, lesquelles ont été pendant cinq années consécutives l'occasion de grands moments de convivialité – compagnons que je ne peux pas nommer ici tant ils sont nombreux.

Ces années de travail ont été l'occasion de rencontres très enrichissantes avec des chercheurs dont l'éclairage m'a beaucoup profité. Je remercie l'ensemble des chercheurs du laboratoire SPHERE de m'avoir fait une place dans ce milieu professionnel et intellectuel très stimulant.

Je remercie Michèle Désenfant, Nicolas Fischer et Marc Priel, du Laboratoire National de Métrologie et d'Essais, de leur intérêt envers mon travail et de leur contact très chaleureux. Je remercie également François Nez et toute l'équipe du Laboratoire Kastler Brossel qui m'ont ouvert à plusieurs occasions les portes de leur laboratoire et m'ont fait profiter de leur expertise. J'exprime ma gratitude envers Walter Bich, Charles Ehrlich, Barry Taylor et Wolfgang Wöger pour avoir sacrifié une partie de leur temps afin de m'accorder des entretiens. Je remercie David Newell et Claudine Thomas de nous avoir accueillis à plusieurs reprises, Nadine de Courtenay et moi-même, aux réunions du CODATA, ce qui m'a permis d'observer de l'intérieur une thématique que la littérature seule ne permettait d'aborder que de façon plus abstraite. De même, je remercie Franck Bielsa et Patrick Juncar, qui m'ont très aimablement fait visiter les locaux du LNE à Trappes et ont pris le temps de m'expliquer le fonctionnement de leur « balance du Watt ». Mes remerciements vont également à Séverine Démeyer pour m'avoir accueilli à une réunion du groupe de travail BaBayes dans les locaux de l'INRA, à Jouy en Josas, alors que j'en étais encore à mes premiers balbutiements sur le sujet de l'emploi des statistiques bayésiennes en métrologie.

J'adresse de plus mes remerciements à Giora Hon et Luca Mari pour les multiples discussions que nous avons eues sur la thématique de la mesure et des erreurs de mesure, et qui ont grandement nourri mon travail. Je remercie Clément Maisch de m'avoir fait part de ses remarques et d'avoir discuté avec moi de son travail en didactique de la physique, lorsque je débute ma thèse. Je suis reconnaissant envers Daniel Andler, Anouk Barberousse, Isabelle Drouet, Sara Franceschelli et Cyrille Imbert de m'avoir reçu comme étudiant « indépendant » dans leurs cours respectifs. Je remercie Alexandre Mallard de l'intérêt qu'il a accordé à mon travail, et des conseils qu'il m'a fournis, lesquels m'ont permis d'entrouvrir des horizons nouveaux encore largement inexplorés pour l'instant. Merci également à Oliver Schlaudt de son

soutien et de ses remarques très pertinentes, que ce soit quant au fond ou à la forme, qui m'ont permis de prendre du recul lors des premières étapes de la rédaction de ma thèse. Je remercie pour finir tous ceux, très nombreux, que je n'ai pas cités mais qui m'ont également beaucoup apporté que ce soit sur le plan intellectuel, personnel ou professionnel.

Merci à Jean Deveaux pour sa relecture attentive d'une version précédente de ce manuscrit, à Helmy Chekir pour une relecture de l'introduction, ainsi qu'à Émile Gayoso et Quentin Lade pour leur aide en vue de la préparation de la soutenance.

Mes dernières pensées vont à mes amis et à ma famille ; mes frères Vincent, Pascal et Nicolas ; mes parents, mes grands-parents, cousins et autres, dont la présence a toujours été inestimable. J'embrasse ma femme Agnès, qui a supporté sans broncher tous mes états d'âme et m'a apporté un soutien indéfectible ; qu'importe la thèse – je suis plus que jamais heureux, Agnès, que nous ayons fait le choix de partager nos vies.

Mes ultimes remerciements, bien plus lointains, vont au quatuor Takács, dont la fabuleuse interprétation des derniers quatuors à corde de Ludwig van Beethoven est un trésor précieux.

Avertissements sur les traductions et la mise en forme

Traductions

Lorsque les ouvrages cités ont fait l'objet d'une publication traduite en langue française, c'est à cette traduction que je me réfère, sauf mention contraire explicite. Les numéros de pages renvoient alors à la version française, comme l'indiquera d'ailleurs systématiquement l'appel bibliographique correspondant.

Lorsque les ouvrages cités n'ont pas fait l'objet d'une publication traduite, la traduction proposée dans cette thèse est alors la mienne, sauf mention contraire explicite.

Mise en forme

La version électronique de cette thèse contient des figures en couleur. Par conséquent, un soin particulier a été accordé à ce que les figures restent compréhensibles pour les versions imprimées en noir et blanc. Les légendes, en particulier, sont rédigées de façon à ce que le renvoi à des éléments d'une figure ne fasse pas seulement appel à sa couleur mais également à un autre signe distinctif permettant de le repérer quelle que soit la version du document.

Introduction générale

Si tous les philosophes et scientifiques n'ont pas nécessairement la « conviction empiriste selon laquelle l'unique source de toutes nos connaissances est l'expérience sensible »¹, il est communément admis que, dans les sciences empiriques, la validité des théories est tributaire de leur accord avec l'expérience. La mesure, en particulier, constitue l'un des modes les plus puissants d'accès au donné empirique en vue du test des théories. L'amélioration toujours plus spectaculaire de la précision des mesures physiques est régulièrement évoquée pour illustrer le succès de l'entreprise scientifique. Dans ce contexte, les acteurs du domaine scientifique discutent depuis plusieurs siècles d'une notion essentielle de leur pratique que l'on désigne aujourd'hui en métrologie – la science de la mesure – par le terme d'« incertitude de mesure », et qui a trait à l'évaluation de la qualité et de la fiabilité d'un résultat de mesure. Une telle évaluation est une part intégrante et fondamentale de l'activité de mesure, et touche un très vaste ensemble de secteurs qui ne sont pas seulement scientifiques, mais aussi techniques ou économiques. Comme le fait remarquer Marc Himbert, « (l)a société s'appuie sur une infrastructure qui n'est pas mais pourrait être familière, souvent invisible, de services, de denrées, de réseaux de transport et de communication dont le bon fonctionnement est essentiel à la vie quotidienne [...] C'est par la métrologie qu'on peut assurer que les volumes des marchandises qui font l'objet de transactions commerciales sont mesurés de façon correcte, comme le pétrole brut, le gaz naturel, l'eau » ou « qu'on peut fabriquer de façon efficace les composants d'objets aussi variés que les lecteurs de disques numériques » ou encore garantir le fonctionnement des systèmes de télécommunication². Bien que l'incertitude de mesure soit une notion capitale en métrologie, elle n'a jusqu'à présent guère recueilli l'attention des philosophes des sciences, à l'exception notable de quelques travaux récents qui révèlent tout l'intérêt de cette thématique³. Cette thèse s'inscrit dans le sillage de tels travaux et se propose d'explorer les questions philosophiques que l'on peut soulever à propos de la notion d'incertitude de mesure. Quelle est la « valeur » de l'incertitude ? Quelle est sa valeur numérique : comment la calcule-t-on ? Quelle est sa valeur épistémique : comment peut-on et doit-on interpréter un résultat de mesure dans les sciences contemporaines ? Une première idée consiste à avancer que l'incertitude de mesure quantifie l'imprécision d'un résultat de mesure. Mais le sens commun de « précision » n'est pas univoque, et le terme renvoie à autant d'acceptions spécifiques qui n'ont pas toutes la même signification. Ce constat motive le choix du titre de cette thèse : s'interroger sur l'incertitude

1. Barberousse, Kistler et Ludwig (2000), p.10

2. Himbert (2001), p.2.

3. Tal (2012), Boumans, Hon et Petersen (2014)

de mesure, c'est explorer la multiplicité des modes par lesquels l'idée générale de « précision » est appréhendée par les scientifiques dans leurs sphères spécialisées. Cela nous amène dans le même temps à réfléchir à la façon dont l'imprécision des mesures physiques se rapporte à des limites possibles de la connaissance expérimentale.

Philosophie des sciences et philosophie de l'expérimentation

Dans leur très grande majorité, les conceptions philosophiques de la démarche scientifique s'accordent sur l'idée que les théories scientifiques sont, d'une façon ou d'une autre, évaluées à l'aune des faits expérimentaux. La façon dont les théories scientifiques sont confrontées aux données empiriques et dont on peut tirer les conclusions d'une telle confrontation constitue l'une des thématiques les plus classiques de la philosophie des sciences, qui a engendré des positionnements très variés. Ainsi, dans la première moitié du XX^e siècle, se sont opposés les thèses des positivistes logiques, portées par exemple par Rudolf Carnap, et le falsificationnisme de Karl Popper qui souhaitait contourner l'inéluctable « problème de l'induction » faisant obstacle à toute démarche empiriste⁴. Ces raisonnements fondés sur une analyse logique de la structure scientifique traitaient principalement de la relation entre des énoncés scientifiques et des énoncés d'observation, et ne considéraient pas comme problématique la question de l'observation elle-même.

Analysant la portée de la déduction mathématique à l'intérieur des théories scientifiques, Pierre Duhem avait pourtant déjà mis en avant le fait que le rôle des mathématiques en physique est inévitablement dépendant de la façon dont les « faits pratiques » sont « traduits » en « faits théoriques », ce qui introduit une réflexion sur l'erreur expérimentale⁵. Environ un demi-siècle plus tard, en réaction au positivisme logique et à l'empirisme en général, des philosophes et historiens comme Norwood Hanson, Thomas Kuhn, ou encore Paul Feyerabend développaient à leur tour l'idée de « charge théorique des observations »⁶, soulignant que les théories scientifiques n'étaient pas évaluées en regard de données « pures », mais que l'observation devait elle-même être soumise à la critique. Néanmoins, ces auteurs adoptaient une approche très théorique, avec une attention modérée envers le détail de la façon dont les expériences sont menées chez les scientifiques. Chez Kuhn par exemple, la thèse de la charge théorique des observations s'articule essentiellement autour d'aspects concernant la psychologie de la perception : l'observateur est soumis à certaines illusions, en conséquence de quoi ses préconceptions théoriques influencent ce qu'il observe.

Vers la fin du XX^e siècle, d'autres philosophes et historiens, à la suite de Ian Hacking, Peter Galison, ou encore Allan Franklin, ont souligné le manque d'attention que l'histoire et la philosophie des sciences a porté à l'expérimentation⁷. Ils ont également critiqué la façon dont il est fait appel aux cas historiques pour défendre des thèses philosophiques, Franklin allant jusqu'à dénoncer des exemples de traitement « mythifié »⁸ des expériences. Ces philosophes et histo-

4. Carnap (1936); Popper (1935)

5. Duhem (1906)

6. Hanson (1961); Kuhn (1983); Feyerabend (1975)

7. Hacking (1983); Galison (1987); Franklin (1989). Précisons toutefois que l'expérimentation avait déjà été abordée par un courant de pensée allemand, voir par exemple Janich (1978); à ce sujet, voir Schlaudt (2014).

8. Franklin (1989), pp.1-3. Franklin illustre cet aspect par trois exemples d'épisodes traités parfois avec trop de

riens, que Robert Ackermann a désigné sous le nom de « nouveaux expérimentalistes »⁹, ont manifesté l'importance qu'il y a d'effectuer une étude rigoureuse et spécifique de l'expérimentation, sans la subordonner à la question de la construction des théories, et sans centrer son objet sur l'analyse des théories scientifiques. Le mouvement des nouveaux expérimentalistes a notamment défendu la thèse de l'autonomie de l'expérience par rapport à la théorie, montrant que « l'expérimentation a une vie propre »¹⁰ et que l'organisation du travail scientifique se structure en cultures et sous-cultures partiellement étanches¹¹, dont en particulier celles des expérimentateurs et celles des théoriciens¹². Ces auteurs ont, de plus, fortement insisté sur la différence entre perception sensorielle et expérimentation, défendant la spécificité du rôle de l'instrumentation dans le travail scientifique. Dudley Shapere soutient ainsi que, contrairement à ce que laissent penser les post-positivistes, une enquête expérimentale ne fait appel que de façon très minimale à la perception même de l'expérimentateur, lequel se repose sur des instruments et des théories qui participent au travail d'objectivation de l'observation effectuée ; il n'y a pas alors de véritable lien entre perception et observation, la seconde étant un acte beaucoup plus complexe¹³. Hacking rappelle de plus qu'une expérimentation est une construction active et ne se limite pas à observer passivement des faits expérimentaux qui nous sont donnés. Ces considérations amènent notamment à critiquer la thèse de la « charge théorique » des observations¹⁴.

Le mouvement des nouveaux expérimentalistes a contribué à un renouvellement des travaux sur l'histoire et la philosophie de l'expérimentation. Un autre apport notable de ce mouvement tient au fait qu'il a dégagé les lignes d'une distinction entre justification et fiabilité. L'examen du test des théories ne peut pas se résumer à analyser le rapport *logique* de justification qui relie les données empiriques aux théories : ainsi Galison rappelle-t-il que « les démonstrations des expérimentateurs sur la réalité [...] n'auront jamais la forme close d'un argument déductif »¹⁵. Par conséquent, il faut aussi étudier ce qui rend l'activité scientifique *raisonnable*, et en particulier les raisons pour lesquelles on peut considérer les données expérimentales comme fiables – c'est-à-dire non pas comme vraies, mais comme dignes de confiance. Galison a cherché à comprendre comment les scientifiques en viennent à considérer qu'ils peuvent accepter avec *confiance* les conclusions tirées d'une expérience donnée. Il note qu'il n'existe pas un instant précis durant lequel l'expérience devient une preuve, mais que c'est aux acteurs scientifiques de prendre la décision de considérer qu'il n'y a plus lieu de douter suffisamment pour en rejeter les

légèreté : la prétendue expérience de la chute des corps par Galilée depuis le haut de la Tour de Pise ; l'expérience des fentes doubles de Young mettant en évidence le caractère ondulatoire de la lumière ; et l'expérience de Michelson et Morley visant à mesurer la vitesse de la lumière et à tester l'hypothèse de l'éther.

9. Ackermann (1989)

10. Hacking (1983), p.150.

11. Galison (1997)

12. Galison (1987), p.13.

13. « La science en est venue de plus en plus à exclure autant que possible la perception par les sens de tout rôle à jouer pour l'acquisition de preuves observationnelles », Shapere (1982), p.508.

14. C'est ce qu'a également fait Michael Heidelberger en montrant que l'expérience pouvait, au moins dans certains cas, être indépendante de la théorie (Heidelberger, 2003).

15. Galison (2002), p.2.

conclusions – c'est ce que Galison appelle « achever les expériences »¹⁶. Franklin décrit à son tour un certain nombre de « stratégies épistémiques » qui permettent d'attester avec fiabilité de l'observation d'un phénomène ou de la validité d'un résultat expérimental¹⁷ sans passer par la démonstration complète du résultat en question¹⁸. Dans la lignée des travaux précédents, Bogen et Woodward ont quant à eux montré que l'étude de phénomènes physiques en vue de leur mise en relation avec des conclusions théoriques fait appel à l'analyse de grandes collections de données expérimentales qu'il est impossible d'analyser une à une¹⁹. Ils font valoir qu'il est impossible de prédire, d'expliquer et de justifier individuellement la valeur de chaque donnée expérimentale. En revanche, le travail des scientifiques consiste en partie à s'assurer de la *fiabilité* des données par un raisonnement qui ne dépend pas des théories impliquées dans l'explication du phénomène concerné. C'est une fois que la fiabilité des données expérimentales est attestée que celles-ci offrent la possibilité de mettre en évidence l'observation d'un phénomène physique, puis d'*inférer* quelque chose sur les théories que l'expérience vise à tester.

La distinction soulevée par les nouveaux expérimentalistes, entre justification des théories par l'expérience d'une part, et analyse de la fiabilité des données d'expérience d'autre part, trouve un intérêt tout particulier en ce qu'elle renvoie à la pratique effective des scientifiques, lesquels sont extrêmement attentifs aux questionnements liés à la fiabilité de leurs mesures. Les scientifiques ne se contentent pas de collecter des résultats expérimentaux et de conclure si ces résultats viennent vérifier ou infirmer leurs théories. Une portion substantielle de leur travail réside dans l'examen de la fiabilité et de la crédibilité de ces résultats expérimentaux. De fait, les scientifiques n'ont pas eu besoin de la philosophie des sciences pour aborder à leur façon la question du test des théories et du rapport entre théorie et expérience; celle-ci fait partie intégrante de leurs différents champs disciplinaires. Cette question prend donc, dans ce cadre, une forme non pas philosophique mais d'abord pleinement scientifique, bien que les thématiques liées au caractère conclusif des résultats de mesure soient ensuite également pour les scientifiques l'occasion d'un mouvement réflexif qui les amène à élargir les questionnements soulevés à propos de leurs pratiques ainsi que des formalismes qu'ils mobilisent.

En s'interrogeant sur la fiabilité des observations, en particulier des observations astronomiques au XVIII^e siècle, les savants en sont venus à développer une véritable théorie du rapport entre théorie et expérience, que Johann Einrich Lambert a appelé « théorie de l'erreur »²⁰. Par ce mouvement, les scientifiques se sont emparés d'une question qui semble épistémologique par nature, mais qu'ils abordent de façon technique et qui prend la forme d'une théorie mathématisée qui relève du calcul des probabilités. Ce faisant, ils ne thématisent l'angle philosophique de ce questionnement que de façon implicite, sans le développer ouvertement. En examinant la façon dont les scientifiques se sont emparés de cette thématique, nous souhaitons quant à nous

16. Galison (1987)

17. Franklin (1989), Franklin (1990). Voir également Franklin et Perovic (2015), section 1.1.1. Il peut simplement s'agir d'étalonner un instrument en l'utilisant sur un phénomène déjà connu et maîtrisé, ou encore de neutraliser certains artefacts en les reproduisant en l'absence du phénomène étudié afin de montrer qu'ils sont causés par des éléments du protocole expérimental.

18. Franklin précise d'ailleurs que « la validité d'un résultat expérimental ne garantit pas que le résultat soit correct », Franklin (1990), p.99, note n° 1.

19. Bogen et Woodward (1988)

20. Armatte (2010), p.2.

en tirer un objet d'étude épistémologique. Cela nous amène à l'objet central de cette thèse, à savoir le concept d'« incertitude de mesure » autour duquel, nous le montrerons, les scientifiques articulent leurs raisonnements. Notre parti pris consiste ainsi à prendre pour point de départ non pas ce que les philosophes disent de l'erreur et de l'incertitude en science, mais la façon dont les scientifiques eux-mêmes ont été amenés à traiter ces concepts dans leurs pratiques.

Suite au mouvement initié dans les années 1980 qui a poussé les philosophes à accorder une attention plus détaillée à l'expérimentation, la mesure devient à son tour peu à peu, en particulier depuis les années 2000, l'objet d'un intérêt spécifique²¹ ; nous situons notre étude dans la continuité de ce sujet, encore très neuf et relativement peu exploré par les philosophes aujourd'hui.

Mesure, problème de l'erreur et incertitude de mesure

La mesure est une activité essentielle de la démarche scientifique moderne. Elle est au fondement de l'évaluation des hypothèses, des modèles et des théories scientifiques, et est considérée comme le moyen par excellence d'accéder à un matériau empirique objectif. La mesure est fondée sur une architecture des grandeurs physiques, chimiques, biologiques ou encore sociales, qui conditionne la mathématisation de la science. Mesurer une grandeur, c'est considérer que l'on peut représenter mathématiquement les propriétés du monde physique, c'est-à-dire que l'on peut quantifier les grandeurs physiques par des nombres, pour peu que l'on définisse des unités de mesure appropriées.

Suivant les métrologues, on considère généralement que « mesurer, c'est comparer ; c'est comparer une grandeur physique inconnue avec une grandeur de même nature prise comme référence, à l'aide d'un instrument. C'est exprimer le résultat de cette comparaison à l'aide d'une valeur numérique, associée à une unité qui rappelle la nature de la référence, et assortie d'une incertitude qui dépend à la fois des qualités de l'expérience effectuée et de la connaissance que l'on a de la référence et de ses conditions d'utilisation »²². Cependant, les métrologues rappellent immédiatement que ce n'est pas là une caractérisation complète de ce qu'est mesurer. En effet, il n'est jamais possible de garantir qu'un résultat est exact. Par conséquent, un résultat de mesure acceptable ne peut pas se présenter sous la forme d'une unique valeur numérique, mais doit être accompagné d'une évaluation de l'« incertitude de mesure » qui lui correspond²³. Ce mouvement a deux conséquences notables. D'une part, il permet de communiquer le résultat

21. Voir par exemple Chang (2004), van Fraassen (2008), Tal (2015). L'ouvrage collectif édité par Hans Radder (Radder, 2003) sur la philosophie de l'expérimentation incluait déjà deux chapitres consacrés aux instruments de mesure (Harré, 2003; Heidelberger, 2003). Il est à noter que si l'intérêt des philosophes pour ce sujet n'est qu'assez récent, la métrologie et l'instrumentation étaient soumises à l'examen des sociologues et des historiens depuis davantage de temps, en témoignent par exemple les travaux de Simon Schaffer (voir par exemple Schaffer, 1992); on peut également se référer à Wise (1997) ou encore à Bourguet, Licoppe et Sibum (2002).

22. Site web de l'association « Metrodiff », http://www.metrodiff.org/cmsms/index.php/metrologie-contemporaine/cssmenu_horizontal.html (consulté le 19 février 2016).

23. Comme l'exprime Marc Himbert : « (s)euls l'établissement non ambigu de la traçabilité du résultat et l'évaluation convenable de son incertitude peuvent inspirer une confiance suffisante pour utiliser effectivement la donnée de sortie de la démarche expérimentale engagée, afin de prendre des décisions, sur le plan scientifique, technique, environnemental, réglementaire, diagnostique, thérapeutique, sportif, etc. », Himbert (2001).

et de le rendre utilisable par d'autres. De fait, cette évaluation fait partie du processus qui consiste à quantifier à quel point le résultat est digne de confiance. Un résultat de mesure est produit dans un contexte particulier, celui d'une expérience donnée, et l'incertitude de mesure est indispensable si l'on souhaite extraire ce résultat de son contexte et l'employer dans d'autres situations. D'autre part, l'évaluation de l'incertitude de mesure joue un rôle primordial dans le test des théories. En effet, c'est elle qui permet d'évaluer en quoi un résultat de mesure est significatif ou non : si l'on note une différence entre une valeur théorique et une valeur mesurée d'une même grandeur, le désaccord constaté ne peut être considéré comme sérieux que s'il est sensiblement supérieur à l'incertitude de mesure associée au résultat expérimental.

L'idée qu'un résultat de mesure n'est jamais exact n'est pas simplement l'expression d'un scepticisme très général, voire métaphysique. Elle découle de constats très concrets : observations expérimentales en désaccord les unes avec les autres alors qu'elles portent sur un même phénomène, théories bien trop solidement attestées pour que l'on puisse croire sans réserve un résultat de mesure récent qui semble aller dans un sens contraire, etc. Cela mène à introduire le concept d'erreur de mesure, compris comme l'écart entre un résultat expérimental et le résultat qui aurait dû être obtenu si la mesure était exacte. L'ensemble des questionnements épistémologiques qui découlent de l'introduction du concept d'erreur en science entrent dans ce que Giora Hon a désigné comme le « problème de l'erreur »²⁴ : peut-on identifier la présence d'erreurs expérimentales, expliquer cette présence, et apprendre de l'erreur ? À quelles conditions un résultat expérimental, bien qu'inexact, peut-il être un outil pertinent et performant pour le test d'une théorie ? L'erreur de mesure et l'incertitude de mesure sont deux des concepts par lesquels les scientifiques eux-mêmes ont intégré le problème de l'erreur dans leurs propres pratiques et à l'intérieur des formalismes qu'ils utilisent. Avec l'incertitude de mesure, les scientifiques s'efforcent de quantifier un doute et s'acquittent à leur façon d'une tâche à laquelle les philosophes n'ont accordé que très peu d'attention jusqu'à présent, et qu'il nous semble essentiel d'analyser plus en détail.

Le constat effectué jusqu'ici nous amène à conclure qu'un résultat de mesure n'est pas constitué d'une valeur unique, mais est un *ensemble* de valeurs, par exemple un intervalle dont la largeur est l'incertitude de mesure. Pour autant, l'incertitude de mesure est un concept complexe et difficile à appréhender – ne serait-ce déjà que dans son rapport à l'erreur de mesure : *les deux concepts sont différents*, et nous nous attacherons à le montrer. Le terme d'« erreur de mesure » a connu de nombreuses acceptions, parfois très différentes les unes des autres. Si l'analyse statistique des erreurs de mesure a émergé dès la fin du XVIII^e siècle (d'abord en astronomie et en géodésie), le traitement rigoureux de l'incertitude reste en constante évolution et sujet à de nombreux débats. Il n'y a pas encore consensus dans la communauté scientifique sur la façon dont il faut l'interpréter, la calculer et l'intégrer au formalisme scientifique.

À l'intérieur de la métrologie, ces débats demeurent ouverts malgré la volonté des institutions internationales de trouver un consensus sur la question depuis la fin des années 1970. La métrologie est en effet une discipline visant à établir les fondements de la mesure et remplit en particulier une tâche de régulation des méthodes et des pratiques des expérimentateurs, ainsi que du vocabulaire employé pour en rendre compte. Elle est elle-même en évolution et devient

24. Hon (1987), Hon (1989b), Hon (1998), Hon, Schickore et Steinle (2009)

peu à peu une discipline normative dont l'actualité est particulièrement riche. À l'initiative de plusieurs organisations, le Bureau International des Poids et Mesures a rassemblé des experts pour parvenir à la production d'un document, le *Guide pour l'expression de l'incertitude de mesure* (GUM), en vue de fournir des recommandations communes à l'ensemble des praticiens dans un large spectre de disciplines : « un consensus mondial sur l'évaluation et l'expression de l'incertitude de mesure permettrait à la signification d'un vaste éventail de résultats de mesure en science, ingénierie, commerce, industrie, et de réglementation d'être aisément comprises et correctement interprétées »²⁵. À la publication de ce document s'ajoute celle d'un lexique, le *Vocabulaire International de Métrologie* (VIM), publié et réédité depuis 1984 sous trois versions successives, la dernière d'entre elles datant de 2008. Notre travail accordera une attention toute particulière aux dernières avancées de la métrologie et aux documents à nature normative que sont le GUM et le VIM. L'analyse de ces documents sera au centre d'une première partie, et certains des débats philosophiques et conceptuels supplémentaires qu'ils suscitent seront abordés dans une seconde partie. Une troisième partie visera à mettre les enseignements des deux premières parties en regard d'une pratique spécifique, celle des physiciens impliqués dans le travail de l'organisation scientifique internationale qu'est le CODATA (Committee on Data for Science and Technology).

Plan de l'argumentation

Cette thèse est structurée en trois parties, précédées par un bref chapitre d'exposition, le chapitre 1. Ce préambule présente la nature de notre objet d'étude en introduisant les concepts qui reviendront de façon récurrente dans les différentes parties : erreur et incertitude de mesure, grandeur physique et « mesurande », modèle et fonction de mesure. Ce chapitre explicite également le contexte dans lequel nous nous plaçons, celui des développements récents que la métrologie a connus sous l'impulsion des institutions nationales et internationales de métrologie.

La première partie est consacrée au traitement probabiliste de l'incertitude de mesure. Elle se situe d'emblée dans une perspective contemporaine et sous l'angle des fondements techniques de l'analyse d'incertitude. Les développements techniques des métrologues font peu à peu émerger des problématiques philosophiques. La seconde partie se place dans la continuité de la première en examinant l'élaboration philosophique des métrologues dans leurs ouvrages techniques, élaboration que nous remettons en question. Elle est centrée autour d'un concept particulièrement discuté : celui de « valeur vraie » d'une grandeur physique. La troisième partie se tourne vers un usage spécifique de l'incertitude de mesure. Ce faisant, elle sort du domaine exclusif de la métrologie pour aborder la question de l'incertitude de mesure sous l'angle de la physique de précision, au travers de l'activité des ajustements des constantes de la physique. Ce changement de perspective nous permettra d'enrichir la façon dont on peut répondre aux problématiques développées jusqu'alors.

Un épilogue retracera brièvement, dans une perspective plus historique, la genèse du *Guide pour l'expression de l'incertitude de mesure* (GUM) et, à travers elle, les questions institution-

25. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.vii.

nelles et conceptuelles qui se sont posées aux métrologues dans le courant des années 1970, 1980 et 1990.

Première partie : l'interprétation probabiliste de l'incertitude de mesure

La première partie vise à aborder le concept d'incertitude de mesure et ses ramifications philosophiques, à partir de la façon dont le GUM tente de le rationaliser. L'objectif est d'inspecter la structure technique du concept, pour identifier la façon dont certains questionnements techniques font émerger des problématiques d'ordre philosophique. Notre étude se focalisera sur la nature et la fonction des modèles statistiques d'analyse d'incertitude. Nous verrons en particulier que la métrologie contemporaine est traversée par des questionnements très ouverts sur la nature des probabilités à employer dans ces modèles, et que l'adhésion traditionnelle à une interprétation fréquentiste des probabilités est sérieusement remise en question depuis le début des années 1970. Les textes contemporains de métrologie proposent ainsi de s'appuyer sur une interprétation spécifique des probabilités, dite « épistémique ». En parallèle, un nombre grandissant d'acteurs proposent de développer une approche pleinement bayésienne de la mesure. Or, ce mouvement s'accompagne d'une évolution notable de la conception de la mesure, certains métrologues soulignant tout particulièrement la subjectivité attachée à l'activité de mesure, ce qui va à contre-courant de la conception classique de la mesure comme outil objectif d'enquête par excellence, et n'est donc pas sans engendrer quelques difficultés. Nous chercherons à comprendre les raisons pour lesquelles les métrologues en sont venus à critiquer la conception de la mesure qui était en vigueur au milieu du XX^e siècle et à chercher à développer des alternatives qui leurs semblent préférables.

Avant d'entrer dans le cœur des questions probabilistes, le chapitre 2 présente les éléments communs qui servent de fondement à l'introduction de modèles statistiques. Est discutée en particulier la dichotomie entre erreurs dites « aléatoires » et erreurs « systématiques » de mesure.

Cela nous permet d'attaquer ensuite de front le modèle fréquentiste de l'analyse d'incertitude, objet central du chapitre 3. L'un des points essentiels de notre analyse consiste à identifier quel est l'objet des probabilités au sein du modèle, et à comprendre sur quelles hypothèses fondamentales ce modèle s'appuie. Nous verrons alors que le calcul fréquentiste des incertitudes de mesure ne permet pas de rendre compte de toutes les composantes d'erreur de mesure, ce qui a été perçu comme une limite sérieuse par les métrologues à la fin des années 1970.

L'utilisation de probabilités épistémiques en métrologie vise à pallier ces limites, et pousse au développement d'une conceptualisation bayésienne de la mesure. C'est ce que nous décrivons en deux temps dans le chapitre 4. Nous montrons d'abord comment les probabilités épistémiques viennent répondre aux problèmes posés par l'approche traditionnelle, puis nous expliquons comment certains métrologues et statisticiens ont ensuite proposé d'étendre cette solution de façon à aboutir à une approche entièrement bayésienne de la mesure, dont certaines formulations assument explicitement leur caractère subjectif.

Nous constatons alors que les deux approches, fréquentiste et bayésienne, amènent à interpréter l'incertitude de mesure de plusieurs façons différentes, une première insistant sur l'objectivité de la mesure, une autre insistant au contraire sur la subjectivité de cette activité ; cela

engendre de nombreuses discussions dans la littérature spécialisée. Nous explorons ainsi dans le chapitre 5 les conséquences conceptuelles et philosophiques de l'évolution qui se dessine en métrologie, que nous qualifions de « tournant épistémique ». Cela nous amène à soulever deux questions particulières. La première porte sur le statut de la valeur vraie d'une grandeur dans l'approche contemporaine de la mesure. La seconde porte sur le dialogue entre deux concepts essentiels de la mesure, celui d'« incertitude de mesure » d'une part, et celui d'« exactitude de mesure » d'autre part, ce dernier traduisant la tendance d'un processus de mesure à produire des erreurs faibles. Ces deux questions annoncent les deux parties suivantes de notre travail.

Seconde partie : la « valeur vraie » d'une grandeur

La seconde partie est consacrée à la notion de « valeur vraie » d'une grandeur physique, intuitivement définie comme la valeur que l'on obtiendrait d'une mesure parfaite, c'est-à-dire d'une mesure qui n'est affectée d'aucune erreur. Cette définition fait cependant apparaître une circularité, puisque l'erreur de mesure est justement définie comme une déviation à la valeur vraie. La valeur d'une grandeur est un élément de représentation. Or, l'idée de valeur vraie peut aboutir trop vite à une conception naïve de la mesure où la nature numérique du monde physique précède sa représentation, et où la mesure ne serait que le dévoilement du monde physique dans son essence mathématique. Les métrologues ont à plusieurs reprises critiqué le concept de valeur vraie d'une grandeur, et ont plus récemment traduit ces critiques dans certains textes en tâchant de faire disparaître le concept de valeur vraie du formalisme de l'analyse d'incertitude. L'idée sous-jacente consiste à affirmer que l'évitement du concept de valeur vraie vient épurer les conceptions des métrologues et des scientifiques en général, à la fois sur le plan philosophique et technique. Nous distinguons deux types d'arguments que nous étudions tour à tour : un argument d'« inconnaissabilité » et un argument de « non-unicité ». Malgré les difficultés réelles qu'introduisent ces deux arguments, nous défendons qu'il n'y a pas lieu pour autant d'abandonner le concept lui-même.

Dans le chapitre 6, nous abordons l'argument d'« inconnaissabilité » de la valeur vraie, qui consiste à voir dans la valeur vraie d'une grandeur quelque chose d'à jamais inconnaissable, qui constitue dès lors un objectif illusoire dont il est préférable de se passer. Contre ces prétentions, nous objectons que malgré les aménagements que les métrologues ont effectués pour se débarrasser dans leurs textes de la valeur vraie, ces derniers continuent à en faire l'usage dans leur formalisme. Par conséquent, s'il y a une critique à formuler concernant la valeur vraie d'une grandeur, ce n'est pas sur l'usage du terme mais son interprétation. Après une discussion sur les interprétations possibles que l'on peut faire de ce terme théorique, nous décrivons la critique des métrologues comme une position essentiellement anti-métaphysique, qui n'est pas pour autant anti-réaliste. Nous défendons alors que l'argument d'« inconnaissabilité » a des limites et que l'attachement à la valeur vraie présente des vertus épistémiques, parmi lesquelles celle de maintenir ouvert un processus permanent de correction qui guide le progrès expérimental.

Le chapitre 7 est tourné vers le second argument, l'argument de non-unicité de la valeur vraie. Pour un mesurande donné, il n'est que très rarement possible de concevoir une valeur vraie unique, parce que la définition du mesurande n'est elle-même pas assez précise pour cela. Si une valeur n'est pas unique, peut-on encore véritablement parler de valeur « vraie » ? Peut-

on vraiment penser que les phénomènes physiques peuvent exactement être capturés sous la forme de nombres uniques ? Notre réponse croise des réflexions qui touchent au réductionnisme en science et à l'épistémologie de l'approximation. Nous défendons que la « valeur vraie » d'une grandeur conserve une légitimité, à condition de parler de valeur « approximativement vraie ». Dans ce cadre, l'« incertitude définitionnelle », que les métrologues ont introduite pour rendre compte de la non-unicité des valeurs des grandeurs physiques, apparaît comme une estimation de la limite de précision des modèles physiques dans lesquels les grandeurs que l'on souhaite mesurer sont mobilisées.

Troisième partie : les ajustements des constantes de la physique

La troisième partie de ce travail vise à donner une perspective supplémentaire à notre étude, en cherchant dans la physique de précision des réponses nouvelles aux questions qui ont été posées jusqu'ici. Nous proposons une étude de la pratique dite d'« ajustements des constantes de la physique », qui prend forme en 1929 avec les travaux pionniers de Raymond Birge, et qui perdure encore aujourd'hui sous l'égide d'une institution internationale, le CODATA (Committee on Data for Science and Technology). Notre réflexion dans cette partie s'appuiera sur deux questions initiales : comment peut-on combiner entre eux différents résultats de mesure obtenus dans des conditions différentes ? Comment les scientifiques peuvent-ils s'accorder sur la valeur d'une grandeur ? Cela nous amènera alors à nous interroger sur l'usage qui est fait de l'incertitude de mesure au sein de cette pratique spécifique, et à revenir sur la question essentielle du rapport entre incertitude de mesure et exactitude de mesure.

Le chapitre 8 commence par rappeler un épisode de l'histoire de l'astronomie au XVIII^e siècle lors duquel les savants ont été amenés à la question de la « combinaison des observations », c'est-à-dire de la façon dont on peut utiliser conjointement des résultats de mesure obtenus par diverses personnes dans des conditions différentes. À travers cet épisode, nous pouvons apercevoir les fondements de l'idée de *reproductibilité* d'une mesure, attachée à un examen attentif des sources d'erreurs, ce qui suggère en filigrane le début d'un calcul d'incertitude.

Dans le chapitre 9, nous attaquons le sujet des ajustements proprement dits, avec une étude du travail de Birge qui illustre le rôle de l'incertitude de mesure dans la comparaison et la combinaison des résultats de mesure. L'incertitude de mesure apparaît comme un outil central pour quantifier et traiter mathématiquement le désaccord ou l'accord à l'intérieur d'une collection de résultats expérimentaux caractérisant des grandeurs reliées théoriquement entre elles par un système d'équations.

Nous continuons l'examen des ajustements au chapitre 10 en nous intéressant à la période de trente ans qui suit l'initiative de Birge, et en nous attardant en particulier sur une discussion menée en 1970 par des physiciens, des métrologues et des statisticiens, autour de la question du traitement des données discordantes. Celle-ci fait apparaître différentes positions des acteurs quant au statut de l'incertitude de mesure. Une approche « conservative » défend qu'il faut faire en sorte de s'assurer que les résultats de mesure soient les plus exacts possibles à tout instant, même si cela doit se faire au prix d'une incertitude plus grande. Une seconde position, défendue par des acteurs très influents, considère qu'il est au contraire préférable de proposer des résultats qui soient les plus précis possibles, quitte à ce que ces derniers se révèlent inexacts, et ce parce que cela permet à *terme* le progrès scientifique par l'identification et la correction

des erreurs de mesure.

Le chapitre 11 viendra alors conclure la dernière partie de ce travail en revenant de façon plus systématique sur le rapport entre incertitude de mesure et exactitude de mesure. Nous verrons que l'approche défendue par les physiciens et décrite au chapitre 10 correspond à une perspective de recherche, et voit l'exactitude de mesure comme un concept non pas statique mais dynamique, tourné vers un progrès futur. Une seconde perspective concerne les acteurs de domaines appliqués (industrie, ingénierie, santé publique, etc.), qui sont amenés à réfléchir en termes de risque et pour lesquels l'exactitude devient un enjeu plus immédiat.

Chapitre 1

Préambule : erreur et incertitude

Dans ce chapitre, nous souhaitons introduire l'objet central de notre étude : l'*incertitude de mesure*. Comme nous le montrerons dans le cours de l'exposition, l'incertitude de mesure est le corrélat d'une opération essentielle à la mesure, à savoir l'identification et la correction des *erreurs de mesure*. C'est pourquoi nous expliciterons au préalable ce qu'est l'erreur de mesure avant de préciser la nature et la fonction de l'incertitude de mesure elle-même.

Nous entamons notre présentation par une brève introduction au « problème de l'erreur » (section 1.1) avant de développer ce qu'est une « erreur de mesure » en tant que concept quantitatif (section 1.2), puis ce qu'est une « incertitude de mesure » (section 1.3). Nous explicitons ensuite un élément essentiel de la modélisation de la mesure, à savoir la « fonction de mesure » qui relie la grandeur cible de la mesure, appelée « mesurande », aux grandeurs et paramètres à partir duquel le mesurande est déterminé (section 1.4). Enfin, nous précisons le cadre disciplinaire qui sera celui d'une grande partie de notre travail, à savoir celui de la métrologie, et nous présenterons deux familles de guides internationaux de métrologie sur lesquels nous fonderons notre étude (section 1.5).

1.1 Le problème de l'erreur

En 2008, Bruno Suchaut publiait une étude dénonçant la « loterie des notes au bac »¹. Deux groupes de trois copies de sciences économiques et sociales des sessions de juin 2006 et juin 2007 du baccalauréat étaient soumises à l'évaluation indépendante d'environ trente correcteurs. Les résultats de l'étude montraient une grande dispersion des notes attribuées à une même copie par les différents correcteurs, révélant ainsi « l'incertitude de la notation des élèves »² (voir figure 1.1). L'étude s'attelaient alors à analyser plusieurs causes possibles de variabilité, et montrait en particulier que l'écart entre les notes ne pouvait pas être expliqué par une différence de sévérité entre les correcteurs, puisqu'aucun d'entre eux ne montrait de tendance à systématiquement donner de moins bonnes ou de meilleures notes que la moyenne des autres correcteurs. L'auteur concluait qu'en raison de l'impossibilité matérielle de minimiser les aléas de correction, il était préférable d'abandonner cette pratique évaluative.

1. Suchaut (2008)

2. Suchaut (2008), p.6.

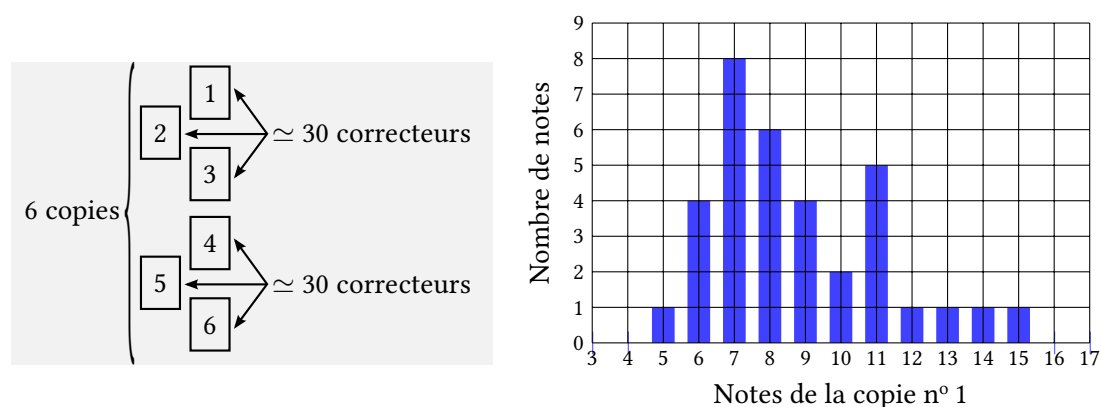


FIGURE 1.1 – Principe de l'étude de Bruno Suchaut sur la notation des copies au baccalauréat. Deux jeux de six copies sont corrigés chacun par environ trente correcteurs, de façon indépendante. Les notes données par les différents correcteurs sont alors analysées de façon à faire apparaître la variabilité inhérente au mode d'évaluation. Le graphe représente l'histogramme des notes attribuées à la copie n° 1; les autres copies présentent une dispersion similaire des résultats.

En laissant de côté les considérations pédagogiques et politiques qui motivent l'étude de Suchaut, il nous apparaît qu'elle est un exemple de la façon dont émerge l'idée d'« erreur » à partir d'un constat d'incohérence. Les différents correcteurs étaient chargés de l'évaluation d'une *même* copie avec des instructions de correction identiques pour tous; par conséquent, il était souhaitable que tous aboutissent aux mêmes conclusions. Les contradictions observées sont la marque d'une possible erreur – sans que l'on puisse encore vraiment identifier la nature de cette erreur. L'erreur pourrait provenir de la façon dont les correcteurs évaluent la copie; suivant cette suggestion, certains correcteurs auraient une notation juste tandis que d'autres surévalueraient ou sous-évalueraient certaines copies. Cette suggestion engage l'hypothèse selon laquelle la copie présenterait une qualité intrinsèque indépendante du correcteur, le « niveau » de la copie, que le correcteur chercherait justement à évaluer. Mais l'erreur pourrait provenir de l'idée même de chercher à évaluer le niveau d'une copie – et, à travers lui, celui de l'étudiant qui l'a rédigée. Peut-on tirer quelque chose d'utile de la pratique évaluative de notation d'une copie d'examen, qui induit en apparence d'importantes incohérences quant au procédé de correction? Répondre à cette question, c'est déjà réfléchir sur le concept d'« erreur ».

Considérons un second exemple. Le 20 mars 2012, le cycliste chilien Luis Mansilla était contrôlé positif à l'Érythropoïétine, dite EPO, une hormone classée parmi les produits dopants³. Il demandait alors l'analyse de l'« échantillon B » – un échantillon qui a été relevé auprès de l'athlète lors du contrôle antidopage et qui n'a pas été analysé, mais conservé intact pour per-

3. « Dopage - Luis Mansilla contrôlé positif à l'EPO », url = <http://www.cyclismactu.net/lire.php?titre=dopage-luis-mansilla-controle-positif-a-l-epo&id=23099> (page consultée le 22 février 2016).

mettre une éventuelle contre-analyse. L'analyse de l'échantillon B de Luis Mansilla se révélait négative, et l'athlète était alors disculpé⁴ – car le doute lui était accordé. Cette procédure est un autre exemple de la façon dont l'idée d'erreur est déduite d'un constat d'incohérence. Dans le cas présent, les échantillons proviennent du corps d'un même athlète et ont été relevés conjointement ; l'on s'attend donc à ce qu'ils présentent les mêmes propriétés. Suivant cette hypothèse, si les résultats des tests sont contradictoires, c'est que l'une des analyses (au moins) est affectée d'une *erreur*. L'origine de cette erreur est alors à déterminer. Ce peut être une erreur instrumentale ; le désaccord pourrait également provenir d'une dégradation du second échantillon, si par exemple celui-ci a été conservé dans de mauvaises conditions. Trancher entre les différentes options et remonter à la source de l'erreur constitue l'une des facettes de l'analyse d'erreur en science.

Les exemples choisis ici empruntent volontairement à des domaines plus ou moins éloignés des sciences physiques, lesquelles seront au cœur de notre étude dans la suite de ce travail. Néanmoins, la façon dont le problème de l'erreur apparaît est similaire à ce que les physiciens rencontrent au cours de leurs activités, que ce soit pour mesurer la masse d'une particule, la densité d'un fluide, etc. Pour résumer, une « erreur » se manifeste de la façon suivante. Lorsque deux observateurs étudient un même phénomène et tirent des conclusions différentes, c'est le témoignage d'une contradiction apparente. Pour résoudre cette contradiction, plusieurs hypothèses sont envisageables. Les observateurs ont-ils vraiment observé le *même* phénomène, dans les mêmes conditions ? Est-il possible que ce phénomène soit apparu différemment à chacun ? Doit-on plutôt conclure que l'un des observateurs (ou les deux) est *dans l'erreur* ? L'« erreur » apparaît comme une hypothèse explicative que rien ne justifie *a priori* d'introduire. Introduire l'hypothèse de l'erreur, puis chercher à l'exploiter de façon bénéfique constitue ce que Giora Hon, en particulier, a appelé le « problème de l'erreur » :

L'erreur est un phénomène épistémologique multiple qui consiste essentiellement en une séparation entre éléments qui peuvent être soit concrets soit abstraits. Dans le premier cas, des objets matériels que des prédictions font coïncider sont en fait éloignés ; dans le dernier cas, une incohérence produit une brèche entre des propositions qui sont supposées s'accorder l'une avec l'autre. [...] L'erreur est l'expression d'une divergence dont la marque est la contradiction – une contradiction qui émerge d'une procédure d'évaluation par rapport à un critère choisi. La nature de cette contradiction, la raison de son occurrence, comment la traiter et ce qu'il peut en être appris une fois qu'elle a été perçue et comprise, constituent le vaste sujet du problème de l'erreur.⁵

L'hypothèse de l'erreur découle d'un constat d'incohérence, mais l'explication qu'elle propose de ce constat introduit une dimension supplémentaire, celle de l'*exactitude* des propositions et des conclusions formulées. Parler d'erreur, c'est laisser entendre que certaines propositions peuvent être vraies ou correctes, tandis que d'autres sont fausses ou incorrectes, et que l'incohérence entre deux propositions est la marque du fait que l'une des deux (au moins) est fausse.

4. "Mansilla cleared of EPO doping charges negative B sample", url = <http://www.cyclingnews.com/news/mansilla-cleared-of-epo-doping-charges-negative-b-sample/> (page consultée le 22 février 2016).

5. Hon (1998), p.466.

À l'inverse, la convergence des observations, par exemple des résultats de mesure d'une même grandeur physique, est considéré comme un signe positif, montrant que le résultat est reproductible dans le temps et dans l'espace, ce qui témoigne du fait qu'il n'est pas simplement un artefact lié aux contingences des conditions d'observation.

1.2 L'erreur de mesure

L'« erreur de mesure », telle qu'elle est introduite dans la théorie statistique des erreurs que les scientifiques mettent à contribution depuis la fin du XVIII^e siècle, est une instance spécifique et *mathématisée* de l'idée générale d'erreur. Comment la définir ? Pour répondre à cette question, il nous faut d'abord expliciter les conditions dans lesquelles et pour lesquelles on parle d'erreur de mesure.

L'un des témoins les plus évidents et les plus classiques de la présence d'erreurs de mesure est la variabilité des résultats de mesure. Lorsqu'un expérimentateur effectue plusieurs mesurages⁶ d'une même grandeur dans des conditions qu'il juge identiques, il arrive fréquemment que les indications données par les instruments de mesure ne soient pas toutes identiques : l'expérimentateur obtient une collection de données dispersées. De façon similaire, lorsque différents expérimentateurs effectuent des mesurages d'une même grandeur avec des principes de mesure différents et dans des conditions différentes, il arrive fréquemment que leurs résultats divergent, et ce avec une amplitude qui peut être plus ou moins forte. Un exemple de ce second cas est donné par la mesure du « facteur de Landé » de l'électron, une grandeur physique sans dimension, notée g , reliant les propriétés cinétiques et magnétiques de l'électron en orbite autour d'un noyau dans un atome, que Peter Galison a documenté⁷. En 1915, Barnett obtenait une valeur proche de 2 alors qu'Einstein et de Haas, de leur côté, aboutissaient à une valeur proche de 1. Si l'on souhaitait concilier ces deux résultats, il fallait enquêter du côté des possibles erreurs qui étaient susceptibles de les affecter. La reproduction des expériences dans la décennie suivante par Barnett, De Haas et d'autres physiciens semblant confirmer la thèse d'une valeur de g proche de 2, de Haas reconnaissait sans peine au congrès Solvay de 1921 que la mesure de g qu'il avait menée avec Einstein était susceptible d'être affectée d'un biais expérimental significatif, expliquant le désaccord avec les résultats suivants⁸.

Cet exemple illustre que dès lors que l'on pense mesurer la même grandeur, aboutir à des résultats variés apparaît comme une anomalie. Toutefois, l'association entre *variabilité* des résultats de mesure et *erreur* n'est ni immédiate ni systématique. La variabilité des résultats de mesure n'est qu'un témoin possible de l'erreur de mesure ; en particulier, elle n'est ni une condition nécessaire ni une condition suffisante pour conclure à la présence d'erreur.

Pourquoi la variabilité des résultats des mesure ne constitue-t-elle aucunement une condi-

6. Dans le vocabulaire de la métrologie, le terme spécifique de « mesurage » est employé pour désigner l'action de mesurer, à laquelle ne se limite pas le terme plus général de « mesure ». Voir [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.xiv.

7. Galison (2002), p.66–75.

8. Galison (2002), p.71.

tion *nécessaire* à la présence d'erreurs de mesure ? On peut le comprendre à l'aide d'un exemple simple, comme la mesure de la longueur d'une feuille de papier rectangulaire à l'aide d'un double-décimètre gradué au millimètre. Dans ce cas, une mesure répétée est fortement susceptible de renvoyer toujours au même résultat : l'écart entre deux graduations – la résolution de la règle – n'est vraisemblablement pas assez petit pour que l'on puisse observer une variation entre les résultats de plusieurs mesurages. Or, dans de telles situations, où la résolution des instruments est trop faible pour observer une variabilité, il faut bien sûr se garder de conclure trop rapidement à l'absence d'erreur de mesure. De toute évidence, la limite de résolution de la règle graduée est une source immédiate d'erreur de mesure, puisque la règle ne permet pas de discriminer entre longueurs différentes qui tombent entre deux mêmes graduations. Ce n'est d'ailleurs pas la seule source possible : l'expérimentateur peut positionner la règle imparfaitement ; celle-ci peut également présenter un défaut d'étalonnage ; la mesure de longueur peut encore être influencée par des phénomènes physiques perturbateurs (dilatation thermique, etc.). On comprend ainsi que lors de la mesure répétée d'une même grandeur avec un même dispositif de mesure, l'erreur expérimentale ne se traduit pas toujours par une dispersion des données, mais que cela dépend des conditions de mesure et du dispositif choisi. Il est d'ailleurs notable que ce ne soit qu'avec l'amélioration de la précision (et donc de la résolution) des instruments d'observation que les astronomes et les géodésiens furent confrontés à ce problème de façon de plus en plus fréquente à partir du XVIII^e siècle ⁹.

Pourquoi la variabilité n'est-elle pas non plus, à l'inverse, une condition *suffisante* pour conclure à la présence d'erreurs de mesure ? Pour répondre à cette seconde question, il nous faut revenir sur les raisons pour lesquelles nous sommes amenés à établir un lien entre variabilité et erreur de mesure. Ce faisant, nous serons donc également amenés à répondre à la question de départ : pourquoi, et à quelles conditions, la dispersion des résultats de mesure peut effectivement être considérée comme la marque d'une erreur de mesure ? Il nous faut d'abord prendre la suite de Kyburg, qui a insisté avec raison sur le fait que le lien établi entre variabilité et erreur n'est valable que si on le suspend à des considérations théoriques :

Il n'y a pas de raison – pas de raison a priori, irréfutable – de supposer que l'une de nos mesures soit erronée. Supposons que nous mesurons [une] table cinq fois et que nous obtenons des résultats différents ? Eh bien il n'y a pas deux mesures qui aient été effectuées au même moment ; par conséquent, qu'est-ce qui dit que la table n'a pas changé de longueur au fil du temps de telle façon à rendre toutes ces mesures exactes à 100% ? Qu'est-ce qui le dit ? Cela est parfaitement clair, une fois que nous avons posé la question. C'est notre vague et générale [...] théorie des objets physiques, ou notre précise et technique théorie des objets physiques, si nous en avons une. Nous *savons*, et la physique nous soutient en cela, que les objets comme la table ne viennent pas à changer de longueur sans raison. Sans raison ? Enfin, sauf en réponse à de telles choses comme des changements de température, d'humidité, des contraintes physiques, et ainsi de suite. ¹⁰

L'*hypothèse* selon laquelle une mesure peut être entachée d'une erreur de mesure provient de

9. [Armatte \(2004\)](#), p.143.

10. [Kyburg \(1992a\)](#), p.77 (Kyburg souligne).

la conception que l'expérimentateur se fait de la quantité mesurée : il y a erreur s'il y a une différence entre ce qu'attend l'expérimentateur et ce qu'il observe ; en particulier, si l'expérimentateur s'attend à ce que la grandeur soit constante et que celle-ci varie. Suivant cette idée, l'hypothèse de l'erreur de mesure n'a de sens que s'il y a une bonne raison de penser que la grandeur visée soit *censée* être stable durant la répétition de l'expérience. L'hypothèse de l'erreur de mesure n'est qu'une hypothèse parmi d'autres possibles, et la dispersion des résultats d'une mesure répétée d'une même grandeur n'est pas une condition suffisante pour conclure à la présence d'erreurs de mesure : elle n'en est une condition suffisante que si l'on a supposé par ailleurs la fixité de la grandeur visée. Suivant l'exemple de Kyburg, si l'on mesure la longueur d'une table, et que la grandeur visée est « la longueur de *cette* table au moment de la mesure », alors la dilatation thermique n'est pas un phénomène extérieur venant perturber la mesure, mais un phénomène inhérent au comportement de la grandeur visée (quand bien même ce comportement dépend des conditions extérieures) ; dans ce cas, la variabilité des résultats de mesure est une information positive qui ne peut pas être traitée comme une erreur de mesure. Mais si la grandeur visée est la longueur de la table à *une température bien définie*, la variabilité des résultats de mesure constituera un indice de ce que les conditions expérimentales réelles ne correspondent pas tout à fait à celles spécifiées – et donc, que le résultat est entaché d'erreur.

Il découle des considérations précédentes qu'il y a un caractère *théorique* dans le mouvement qui consiste à voir dans la variabilité des résultats expérimentaux la trace d'erreurs de mesure. Nadine de Courtenay a poursuivi cette idée, et, en s'appuyant sur l'apport de Norman Campbell à la réflexion sur les fondements axiomatiques de la mesure, a insisté sur le caractère théorique, subordonné à des lois, de la notion de grandeur physique¹¹, dont l'axiomatisation passe par une hypothèse, qui vise à « expliquer les désaccords observés » entre ce qui est obtenu empiriquement et les propriétés attendues des grandeurs étudiées « en faisant appel à l'idée d'erreur »¹². Ce que révèlent les considérations précédentes, c'est qu'il ne peut y avoir d'erreur de mesure que si la mesure vise une *cible* qui aura été définie préalablement : la notion de mesure est une notion *intentionnelle*, attachée à un objectif. Cette cible, les métrologues la désignent par le terme de « mesurande », défini comme la « grandeur que l'on souhaite mesurer »¹³. L'attention que la métrologie a apporté depuis quelques décennies au rôle du « mesurande » reflète l'idée qu'il est nécessaire de *définir* la grandeur que l'on *souhaite* mesurer, c'est-à-dire que la cible de la mesure n'est pas quelque chose qui est *donné* mais qui est *posé*¹⁴.

Il va de soi que les observations que l'on tire d'une expérimentation ne sont pas erronées en elles-mêmes ; ce sont les conclusions que l'on tire de ces observations qui peuvent être entachées d'une erreur. Dans ce cas, le geste théorique consistant à supposer l'existence d'erreurs de mesure consiste à attribuer la cause de la variabilité observée à des paramètres extérieurs à la cible elle-même : imperfections des instruments, mauvaises manipulations de l'expérimenta-

11. « En tout état de cause, les réflexions de Campbell suggèrent que [...] il est nécessaire d'établir, entre les qualités sensibles et les nombres, une relation non pas directe mais indirecte qui passe par la formation de grandeurs physiques obéissant à des axiomes constituant une protothéorie de l'expérience. », de Courtenay (2008), p.235.

12. de Courtenay (2008), p.233

13. Joint Committee for Guides in Metrology (JCGM) (2012), p.17.

14. Je remercie Luca Mari pour avoir attiré mon attention sur ce point particulier.



FIGURE 1.2 – Métaphore des cibles (figure a) : un mesurage est comparé à un exercice de tir sur une cible dont le centre (point central, en rouge) représente la « valeur vraie » du mesurande. L'éloignement des données expérimentales (représentées par les croix, en bleu) au centre de la cible représente l'erreur de mesure. Cette métaphore n'est toutefois pas parfaite. Une difficulté de cette métaphore réside dans le fait que, puisque la cible est représentée à un endroit précis, cela peut donner l'impression que l'expérimentateur *connaît* la valeur vraie, et qu'il peut alors déterminer *grâce à cela* les erreurs commises. Or, le problème est en fait l'inverse : on cherche à déterminer la position du centre de la cible à partir de celles des données expérimentales, et l'expérimentateur doit *pour cela* évaluer l'erreur commise. La situation qui apparaît à l'expérimentateur correspond donc à la représentation de la figure (b) plutôt qu'à celle de la figure (a).

teur, influence de phénomènes physiques faisant fluctuer les indications de mesure, etc. L'écart entre ce qui était attendu et ce qui a été obtenu est donc attribué à des perturbations continues, et est traité comme une anomalie : métaphoriquement, la mesure a manqué sa cible¹⁵. L'erreur de mesure traduit le fait qu'une mesure a « manqué la cible » en exprimant que le résultat de mesure diffère de ce qu'il *aurait dû être* si la mesure *avait été correcte*. Autrement dit, la valeur obtenue par mesurage diffère de la « valeur vraie » de la grandeur visée. Ainsi, pour une grandeur physique X , on écrira pour le résultat x d'un mesurage donné :

$$x = \chi + \epsilon \quad \begin{cases} x = \text{donnée expérimentale (valeur obtenue par la mesure)} \\ \chi = \text{valeur vraie de la grandeur } X \\ \epsilon = \text{erreur de mesure} \end{cases} \quad (1.1)$$

Les manuels scientifiques illustrent souvent cette définition par la « métaphore des cibles », dans laquelle l'expérimentateur vise une cible dont il n'atteint jamais parfaitement le centre (figure 1.2)¹⁶.

Laissant pour l'instant de côté le problème conceptuel que ne manque pas de poser l'introduction d'un concept tel que celui de « valeur vraie » d'une grandeur, nous notons qu'étant donnée sa définition, l'erreur de mesure est d'emblée une expression numérique, un terme théorique quantifié. Les scientifiques prennent bien soin de marquer la distinction entre ce type bien

15. « Manquer la cible est le plus concret et, historiquement, le plus ancien de tous les types d'erreur », Hon (1989b), p.466.

16. Voir par exemple Taylor (1997), p.95.

précis d'erreur qu'est l'erreur quantitative de mesure, et d'autres types d'erreurs telles des erreurs humaines, des erreurs de réglage, des erreurs de raisonnement, etc.¹⁷. De fait, l'« erreur de mesure » n'est pas un concept général visant à souligner les fautes de l'expérimentateur, mais un concept mathématisé visant à être introduit dans une théorie, autrefois baptisée *théorie des erreurs*, qui vise à donner un sens à des résultats de mesure en présence du problème de l'erreur. Cette théorie des erreurs n'a pas directement pour but la représentation et l'explication de phénomènes naturels ou sociaux. Elle vise à déployer une méthodologie, transversale et commune aux différentes disciplines scientifiques, permettant l'exploitation à bon escient des résultats de mesure. C'est cette théorie des erreurs, devenue aujourd'hui l'« analyse d'incertitude », que nous étudierons plus en détail dans les prochains chapitres sous l'angle de la méthode statistique.

Postuler l'erreur de mesure, c'est ouvrir la voie à son *traitement* par plusieurs moyens. Il se peut que la cause matérielle de l'erreur soit identifiée à l'intérieur de la chaîne de mesure, auquel cas cette cause sera (si possible) éliminée par un aménagement adapté du dispositif de mesure, afin d'effectuer une nouvelle mesure, plus exacte. Il se peut que l'amplitude numérique de l'erreur fasse l'objet d'une estimation d'une façon ou d'une autre, ce qui mène à une possible correction : le résultat de mesure est modifié *a posteriori* de façon à éliminer non pas la cause de l'erreur, mais ses conséquences sur la valeur obtenue. Il existe une troisième voie, classique, et qui sera traitée en détail dans les développements futurs des prochains chapitres, mais que nous ne pouvons mentionner que brièvement ici : l'erreur de mesure peut également être « réduite » statistiquement par répétition des mesurages et application d'une méthode statistique adaptée visant à diminuer les conséquences de la présence d'erreur lorsque celle-ci ne peut pas être éliminée à la source et ne peut pas non plus être corrigée.

C'est un des fondements de la méthode scientifique que de considérer que l'on peut progresser par élimination ou correction successive des erreurs. Mais il faut rappeler alors qu'il subsiste toujours un doute quant à l'exactitude des résultats de mesure. De là découle que la formulation d'un résultat de mesure vient avec une *incertitude* que l'on cherchera à caractériser de façon quantitative par ce qui est désormais communément désigné sous le nom d'« incertitude de mesure ».

1.3 L'incertitude de mesure

L'erreur de mesure fait état d'une différence entre ce qui est visé et ce qui est obtenu ; c'est un concept que l'on pourrait qualifier de « normatif » en ce qu'il décrit une différence entre un résultat effectif et le résultat qui *aurait dû* être obtenu si la mesure était exacte. De ce fait, l'erreur de mesure qui affecte un résultat expérimental est quelque chose que l'on souhaite corriger. À cet égard, on peut envisager deux possibilités principales. Il se peut que les phénomènes qui causent une erreur soient inconnus, ou encore n'aient pas été identifiés ; dans ce cas, la correction est impossible. À l'inverse, quelle que soit l'origine d'une erreur de mesure,

17. « Il convient de ne pas confondre l'erreur de mesure avec une erreur de production ou une erreur humaine. », *Joint Committee for Guides in Metrology (JCGM) (2012)*, p.22. « En science, le mot *erreur* [*error*] ne porte pas les connotations usuelles des termes *faute* [*mistake*] ou *maladresse* [*blunder*] », *Taylor (1997)*, p.2.

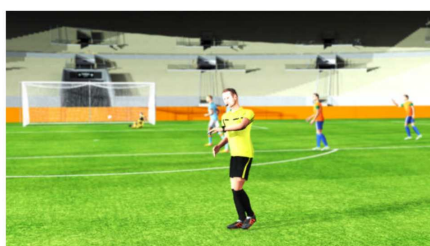
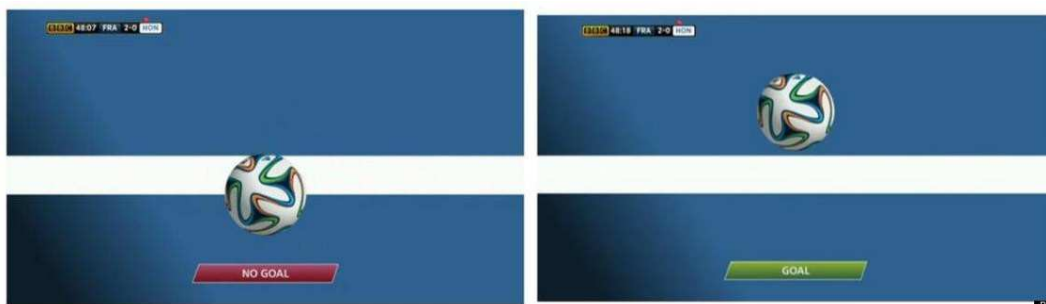
et quelle que soit la façon dont cette erreur se manifeste, il est possible d'amorcer un processus de *correction* dès lors que l'on sait expliquer ou prédire l'erreur en question en la reliant par un modèle à l'évolution de paramètres d'influence qui peuvent eux-même être mesurés ou évalués. Une telle correction présente toutefois des limites : même lorsque les causes d'une erreur ont été identifiées, l'erreur peut jamais être parfaitement quantifiée. Toute limite de connaissance à propos des erreurs de mesure est à l'origine d'une incertitude de mesure qu'il faudra alors se mettre en quête de déterminer. C'est pourquoi à toute composante d'erreur est associée une composante d'incertitude de mesure, et la détermination de cette incertitude est tributaire de la façon dont on peut connaître ou non l'erreur en question, et donc de la façon dont on peut corriger cette dernière. L'incertitude de mesure caractérise en définitive un résultat de mesure qui a été obtenu expérimentalement pour évaluer la façon dont celui-ci permet de dire quelque chose à propos de la grandeur que l'expérience menée vise à déterminer.

Ces remarques font apparaître l'incertitude de mesure comme le corrélat de l'élimination, de la correction, et de la réduction statistique des erreurs de mesure. C'est pourquoi erreur et incertitude de mesure ne se situent pas à un même niveau : l'erreur de mesure traduit l'imperfection d'une mesure ; l'incertitude de mesure traduit l'imperfection de l'analyse de l'erreur de mesure. En particulier, l'amplitude de l'erreur commise ne préjuge en rien de l'incertitude résultante. Eran Tal présente l'exemple de l'étalon primaire de fréquence des États-Unis, le NIST-F1, qui sert de référence pour les mesures du temps au laboratoires américains¹⁸. Cet étalon diverge des indications des horloges à césium, qui réalisent la définition de la seconde ; il présente donc un biais ; pourtant, l'incertitude relative à cet étalon est très faible car l'erreur en question est prédictible avec une très faible incertitude¹⁹. Cette remarque est absolument fondamentale pour bien situer la différence de nature entre les concepts d'erreur et d'incertitude de mesure.

Par suite, l'incertitude de mesure renvoie à la *précision* des mesures, sans qu'il soit pour nous possible, à ce stade, de déterminer de façon rigoureuse la signification de ce terme, dont l'usage courant est d'ailleurs très varié. Le rôle de l'« incertitude de mesure » est de répondre à trois préoccupations majeures. (i) Elle traduit l'impossibilité d'attribuer à une grandeur mesurée une valeur unique, en raison des incohérences constatées. Ces incohérences ne peuvent être résolues qu'en faisant l'hypothèse de l'erreur de mesure. L'incertitude de mesure provient alors du fait que les erreurs de mesure commises ne sont pas parfaitement connues. (ii) Elle sert comme outil, afin de permettre le test des théories et la validation de produits en regards d'exigences formulées en termes de normes légales, de tolérances, etc. En 2014, à l'occasion de la coupe du monde de football organisée au Brésil, les instances internationales ont expérimenté l'utilisation d'une technologie de détection, appelée « technologie sur la ligne de but », visant à déterminer lorsqu'un ballon franchit la ligne de but avec une précision satisfaisante. La technologie employée au Brésil a été développée par la firme allemande « GoalControl »,

18. Tal (2012), pp.166–167.

19. Les documents de métrologie insistent également sur ce point : « (l)'incertitude d'une correction appliquée à un résultat de mesure pour compenser un effet systématique *n'est pas* l'erreur systématique due à cet effet », [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.5 (souligné dans le document).



Accuracy of goal detection: a few mm

GoalControl-4D is the result of restless development and countless hours of testing the system's accurate ball detection and tracking function under all weather conditions. Tests that included many types of footballs, different pitch surfaces, various types of goal nets and game situations.

FIGURE 1.3 – Illustration du principe de fonctionnement du système de technologie sur la ligne de but (figure du haut, source site web de la BBC, consulté en septembre 2014). La compagnie allemande GoalControl revendique une précision de l'ordre de quelques millimètres (figure du dessous, tirée de la page d'informations médias du site web de la compagnie, url = http://www.goalcontrol.de/en/assets/content/presse/goalcontrol_system_facts.pdf, consulté en septembre 2014).

qui revendique une « exactitude de quelques millimètres » (figure 1.3). Sans l'évaluation de cette exactitude, il est impossible de juger ce que permet de conclure une telle technologie : la technologie sur la ligne de but n'est utile qu'à la condition d'être suffisamment précise.

Un troisième rôle de l'incertitude de mesure est (iii) de permettre la communication des résultats de mesure d'un laboratoire à un autre, c'est-à-dire les fixer en tant que *connaissances* réutilisables par tous. C'est ce qu'illustrent typiquement les « tables de valeurs » des constantes physiques publiées tous les quatre ans par le CODATA (figure 1.4). Ces tables de valeurs illustrent de plus l'une des conclusions essentielles de la conception moderne de la mesure : *un résultat n'est jamais une valeur ponctuelle mais un ensemble de valeurs dont la largeur se rapporte à l'incertitude de mesure.*

L'incertitude de mesure est un concept quantifié, intégré à une théorie de l'erreur. Ce constat suscite trois questions. Comment la calcule-t-on? Comment l'interprète-t-on? Comment l'utilise-t-on, c'est-à-dire comment articule-t-on la théorie de l'incertitude de mesure dans la structure mathématique des modèles et théories scientifiques? L'examen du calcul, de l'interprétation et de l'usage de l'incertitude de mesure révèle certains contours de la façon dont les scientifiques expriment leur rapport à la connaissance à l'intérieur même des échafaudages formels vers lesquels leur travail est tourné. Nous verrons qu'il n'y a pas un accord unanime

TABLE XLI. The CODATA recommended values of the fundamental constants of physics and chemistry based on the 2010 adjustment.

| Quantity | Symbol | Numerical value | Unit | Relative std. uncert. u_r |
|---|-------------------|---------------------------------------|---|-----------------------------|
| Newtonian constant of gravitation | G | $6.673\,84(80) \times 10^{-11}$ | $\text{m}^3 \text{kg}^{-1} \text{s}^{-2}$ | 1.2×10^{-4} |
| | $G/\hbar c$ | $6.708\,37(80) \times 10^{-39}$ | $(\text{GeV}/c^2)^{-2}$ | 1.2×10^{-4} |
| Planck constant | h | $6.626\,069\,57(29) \times 10^{-34}$ | J s | 4.4×10^{-8} |
| | $h/2\pi$ | $4.135\,667\,516(91) \times 10^{-15}$ | eV s | 2.2×10^{-8} |
| | \hbar | $1.054\,571\,726(47) \times 10^{-34}$ | J s | 4.4×10^{-8} |
| | $\hbar c$ | $6.582\,119\,28(15) \times 10^{-16}$ | eV s | 2.2×10^{-8} |
| Planck mass $(\hbar c/G)^{1/2}$ energy equivalent | m_{P} | $197.326\,9718(44)$ | MeV fm | 2.2×10^{-8} |
| | $m_{\text{P}}c^2$ | $2.176\,51(13) \times 10^{-8}$ | kg | 6.0×10^{-5} |
| Planck temperature $(\hbar c^5/G)^{1/2}/k$ | T_{P} | $1.220\,932(73) \times 10^{19}$ | GeV | 6.0×10^{-5} |
| Planck length $\hbar/m_{\text{P}}c = (\hbar G/c^3)^{1/2}$ | l_{P} | $1.416\,833(85) \times 10^{32}$ | K | 6.0×10^{-5} |
| Planck time $t_{\text{P}}/c = (\hbar G/c^5)^{1/2}$ | t_{P} | $1.616\,199(97) \times 10^{-35}$ | m | 6.0×10^{-5} |
| | | $5.391\,06(32) \times 10^{-44}$ | s | 6.0×10^{-5} |

FIGURE 1.4 – Extrait du tableau récapitulatif des valeurs recommandées de 2010 des constantes de la physique publiées par le CODATA, tiré de [Mohr, Taylor et Newell \(2012\)](#), p.1587. Les valeurs recommandées ne consistent pas seulement en une unique valeur numérique, mais sont accompagnées d'une incertitude (incertitude relative, colonne de droite) qui représente la demi-largeur des valeurs que l'on peut raisonnablement attribuer à la grandeur concernée.

sur ces questions.

Précisons un élément important : l'incertitude de mesure est autant l'expression d'un doute que celui d'une connaissance. De fait, parler d'« erreur » et d'« incertitude » peut sembler renvoyer à un constat négatif, voire à un constat d'échec²⁰ ; la rhétorique du doute et de l'incertitude est effectivement l'une des armes préférées des mouvements souvent considérés comme « pseudoscientifiques », tel le créationnisme ou le climatoscepticisme²¹. Mais s'il y a expression d'un doute, c'est celui d'un doute rationnel, justifié sur des bases rigoureuses ; et si ces termes doivent être maniés avec soin, il n'en reste pas moins qu'ils traduisent d'abord un constat positif, selon lequel il *reste possible* d'affirmer quelque chose à propos du monde *malgré* l'imperfection de nos méthodes de mesure. L'incertitude de mesure est un concept dont l'introduction répond à l'objectif d'intégrer aux résultats de mesure une information sur les limites raisonnables qui affectent ce résultat, et ces limites sont celles qui ont été identifiées par l'enquête scientifique : on parle de « source d'incertitude ». Le travail le plus fondamental de l'analyse d'incertitude est celui qui consiste à *identifier* les différentes sources d'incertitude de mesure affectant une mesure donnée. Il s'agit d'une véritable enquête mêlant des examens théoriques et expérimentaux, qui constitue l'une des tâches essentielles de l'expérimentateur dans son travail de spécialiste²². Ce sera l'objet du chapitre 2.

20. Ainsi peut-on lire dans le document d'accompagnement des professeurs du secondaire de la rentrée 2002 la remarque suivante : « (l)le terme 'incertitude' a une connotation négative, à cause de son contraire 'certitude'. Les termes 'variabilité' et 'dispersion' ne possèdent pas cette connotation et sont en réalité plus fondamentaux, car ils renvoient au caractère aléatoire de la mesure », [Ministère de la Jeunesse, de l'Éducation nationale et de la Recherche \(2002\)](#), p.89. Cette remarque ne nous semble pas vraiment fondée : la variabilité est une cause possible d'incertitude mais ne suffit pas à elle seule pour comprendre cette dernière.

21. [Oreskes et Conway \(2012\)](#)

22. C'est ce que souligne Hon : « L'expérimentateur [...] se heurte constamment [...] à des divergences entre les théories et les résultats d'observation ; de fait, une partie de sa routine quotidienne [...] consiste, comme M. Polanyi laisse entendre, à expliquer [*explain away*] ces divergences. », [Hon \(1989b\)](#), p.473.

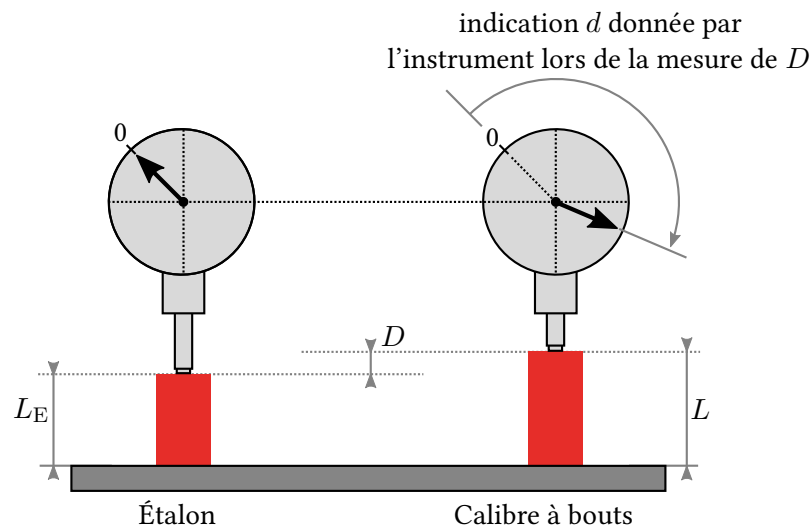


FIGURE 1.5 – Mesure de la longueur L d'un calibre à bouts par comparaison avec un étalon de longueur L_E . La mesure s'effectue en deux temps. (i) Dans un premier temps, le comparateur est calé sur l'étalon, et on règle alors l'indication qu'il donne sur le zéro. (ii) Le comparateur est ensuite calé sur le calibre. Il indique alors par d la différence de longueur qu'il mesure entre le calibre et l'étalon.

Au préalable, il nous faut toutefois expliciter un élément essentiel de la modélisation de la mesure dans l'analyse d'incertitude, qui servira de base à l'ensemble de nos discussions futures : il s'agit du « modèle de mesure », c'est-à-dire la construction d'un modèle de l'expérience qui permet de relier les indications données par les instruments de mesure utilisés, ainsi que les informations extérieures (comme les données théoriques), avec la grandeur que l'on cherche à mesurer.

1.4 Modélisation de la mesure et fonction de mesure

La métrologie contemporaine fonde l'analyse d'incertitude sur une modélisation physique et mathématique du dispositif de mesure. Supposons que l'on cherche à mesurer la longueur L d'un « calibre à bouts » par comparaison avec un étalon de longueur L_E , au moyen d'un comparateur (figure 1.5)²³. La mesure est ici *indirecte* : la longueur du calibre à bouts est calculée à partir de la grandeur D que mesure le comparateur, laquelle est une *différence* de longueurs. C'est pourquoi il faut distinguer la grandeur *visée*, ici la longueur du calibre à bouts, de la grandeur effectivement mesurée (ou *des* grandeurs mesurées : dans nombre de dispositifs plus complexes, la grandeur visée est calculée à partir de la mesure de plusieurs autres grandeurs). Nous avons vu que les métrologues ont appelé « mesurande » la grandeur visée ; c'est là un

23. Cet exemple est repris de l'annexe H du GUM : [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), pp. 82-88, et est développé plus en détail dans [Grégis et de Courtenay \(2016\)](#).

concept essentiel, que nous utiliserons à de nombreuses reprises (employant parfois « grandeur visée » comme synonyme).

L'objectif de la modélisation du dispositif de mesure consiste à établir le lien qui relie le mesurande avec la ou les grandeurs effectivement mesurées. Ainsi, dans le cas présent, la relation entre L , L_E et D est donnée par l'équation suivante :

$$L = L_E + D \quad (1.2)$$

Cette relation exprime une loi, laquelle gouverne le *principe* de la mesure. Mais cette loi est idéalisée : elle ne tient pas compte de l'ensemble des phénomènes qui, en pratique, peuvent venir perturber la mesure. Ainsi, dans le cas présent, il peut se révéler nécessaire de tenir compte des phénomènes de dilatation thermique qui affectent à la fois l'étalon et le calibre à bouts. Il faut alors préciser la température T_0 à laquelle on définit le mesurande, c'est-à-dire la longueur du calibre à bouts que l'on cherche à mesurer ; ensuite, une modélisation plus fine de l'ensemble du dispositif de mesure mène à une équation plus complexe. Dans notre exemple, nous aurons ainsi :

$$L = \frac{L_E (1 + \alpha_E (T - T_0)) + D}{1 + \alpha (T - T_0)} = f(D, L_E, \alpha, \alpha_E, T - T_0) \quad (1.3)$$

où T est la température à laquelle la mesure est effectuée dans le contexte concret de l'expérience, et α et α_E sont les coefficients respectifs de dilatation thermique du calibre à bouts et de l'étalon. Les métrologues appellent « fonction de mesure » la fonction f , qui joue un rôle central dans l'analyse d'incertitude. À chaque fois que l'expérimentateur identifie la présence d'un nouvel effet physique influençant la mesure et susceptible d'être une source d'erreur de mesure, il peut intégrer une modélisation de ce phénomène dans le modèle de mesure, ce qui l'amène en définitive à modifier la fonction de mesure de façon adaptée.

Dans le cas général, le mesurande, que l'on désignera de façon générique par le symbole Y , est calculé à partir des valeurs de p autres grandeurs, que l'on appellera X_i . Parmi les grandeurs d'entrée X_i , certaines sont évaluées par une mesure au sein de l'expérience considérée ; mais ce n'est pas nécessairement le cas de chacune d'entre elles. En addition à ces mesures, l'expérimentateur dispose également d'une somme d'informations qualitatives ou quantitatives qui concerneront entre autres les grandeurs qu'il ne mesure pas directement lui-même : résultats d'autres expériences menées par d'autres scientifiques, valeurs recommandées dans des tables, notices des instruments utilisés. Dans certains cas, il devient nécessaire de faire appel à un jugement, par exemple l'avis d'un expert sur un instrument de mesure. Cela peut aller jusqu'à un jugement dit « subjectif », quand l'expérimentateur lui-même ne peut tout au plus que proposer une estimation d'un paramètre²⁴.

24. Dans la littérature, cela est parfois désigné sous le nom de « supposition éclairée » (“educated guess” dans les textes anglophones). Voir par exemple Bich et Kool (2012), pp.6-7 ; ou encore Lautrup et Zinkernagel (1999), p.88. Müller parle de “guesstimates”, Müller (1979), p.244. Eisenhart a insisté sur les limites de ce type d'évaluation : « Une approche réaliste du problème de l'attribution des ‘meilleures’ valeurs aux constantes fondamentales de la physique avec des indications valides de leurs [incertitudes] exige que nous arrêtions de supposer des valeurs pour les composantes d'erreur que nous n'avons jamais mesurées et, au moyen d'expérimentations dans un cadre adéquat, que nous découvrons ce que les faits sont réellement. », Eisenhart (1971), p.517 ; sa remarque témoigne de la position objectiviste qui est caractéristique de l'approche fréquentiste, comme nous le verrons dans les prochains chapitres.

En fin de compte, la procédure globale d'évaluation de Y amène l'expérimentateur à porter d'abord son attention sur chacune des grandeurs d'entrée prise individuellement, et c'est alors par l'entremise de la fonction de mesure qu'il peut en déduire la valeur du mesurande. La fonction de mesure est le plus souvent écrite de la façon suivante :

$$Y = f(X_1, \dots, X_p) \quad (1.4)$$

Nous reprendrons cette écriture à de multiples reprises dans l'ensemble de ce travail²⁵. Le mesurande Y est également appelé « grandeur de sortie », et les différentes grandeurs X_i sont dénommées « grandeurs d'entrée »²⁶; nous utiliserons également à de multiples reprises ce terme de « grandeurs d'entrée ».

La « fonction de mesure » permet de calculer la valeur du mesurande à partir des données disponibles sur les grandeurs d'entrée. Mais, en présentant l'expérience sous une forme décomposée, elle permet également d'appréhender de façon détaillée les sources d'incertitude de mesure, et sert ainsi de base à la mathématisation de l'incertitude de mesure. On peut distinguer alors deux problèmes spécifiques. Le premier problème est relatif à l'évaluation individuelle de chaque grandeur d'entrée : étant données les informations dont un expérimentateur dispose à propos d'une grandeur d'entrée X_i , quel est le spectre de « valeurs raisonnables » qu'il peut attribuer à X_i ? On peut reformuler ce problème en considérant qu'il s'agit de déterminer la meilleure valeur à attribuer à X_i ainsi que l'incertitude associée. Le second problème est relatif à l'évaluation de Y à partir des X_i , connaissant la fonction de mesure; ce problème est connu, dans le milieu de la métrologie, sous le nom de « propagation des incertitudes ». Comment peut-on « propager » toutes les informations disponibles à propos des grandeurs d'entrée de façon à ce qu'elles soient intégrées à notre connaissance globale de Y ? Nous montrerons dans la partie I que ces deux problèmes font intervenir des méthodes statistiques et qu'il n'en existe pas une conceptualisation unique.

1.5 Mesure et métrologie : le *Guide pour l'expression de l'incertitude de mesure* (GUM)

Si les méthodes d'analyse d'erreur ont fait l'objet de nombreux essais depuis déjà plusieurs siècles²⁷, il ne serait pas possible d'en extraire une conceptualisation commune; par conséquent, notre approche consistera à nous restreindre pour l'essentiel au point de vue exprimé ces dernières décennies dans la sphère scientifique. Nous nous focaliserons tout particulièrement sur l'étude d'une discipline spécifique, la science de la mesure, ou « métrologie ». La

25. Notons que dans certains cas, il n'est pas possible de développer un modèle de mesure, car il n'existe pas de relation mathématique connue entre les grandeurs d'entrée et le mesurande. Ce type de cas n'est pas forcément très fréquent dans la recherche scientifique, mais il est plus commun dans le domaine de l'ingénierie ou de l'industrie. Dans ce cas, les métrologues appliquent une méthode dite de « comparaison interlaboratoire » que nous ne discuterons pas ici. Cette méthode a été décrite par Michèle Désenfant et Marc Priel ([Désenfant et Priel, 2006](#)).

26. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.9.

27. On peut ainsi remonter au traité de Gauss sur le traitement probabiliste de l'erreur de mesure, [Gauss \(1823\)](#).

métrologie est une discipline qui cherche à établir, de façon transversale aux disciplines scientifiques, les fondements de l'activité de mesure. Elle développe de ce fait un travail en grande partie méthodologique, qui permet de coordonner des différents acteurs scientifiques, techniques, économiques, politiques, etc. dans ce qui s'apparente nettement à ce que Hacking a décrit comme une « technologie de l'intersubjectivité »²⁸. Elle indique en particulier quelles sont les bases d'une bonne mesure, et fournit aux scientifiques des outils pour exprimer les résultats de leurs mesures de façon à ce que ces derniers soient communicables et utilisables par d'autres de façon fiable. La métrologie est de ce fait une « infrastructure cachée »²⁹ de la science et de la technologie : elle établit les conditions de possibilité des mesures de précision. Elle profite du progrès scientifique et technique autant qu'elle y participe, des derniers développements de la physique fondamentale jusqu'aux systèmes hautement technologiques qui sont intégrés aux instruments de mesure.

La façon dont la métrologie intègre les questions liées à l'incertitude a constamment évolué et a tout particulièrement connu des transformations notables à la fin du XX^e siècle. En effet, la métrologie s'est engagée dans un vaste programme de réforme qui comprend, notamment, un profond remaniement du Système International d'unités (SI), et qui est accompagné par la publication de deux guides internationaux destinés à harmoniser les pratiques de mesure et la formulation des résultats de mesure. Ces guides sont le *Vocabulaire International de Métrologie* (VIM), publié pour la première fois en 1984³⁰, et le *Guide pour l'expression de l'incertitude de mesure* (GUM), publié en 1993³¹. Ces documents sont le produit d'un travail conjoint de nombreuses organisations nationales et internationales. En particulier, il est publié au nom de sept organisations : Bureau international des poids et mesures (BIPM), Commission électronique internationale (CEI), Fédération internationale de chimie clinique (FICC), Organisation internationale de normalisation (ISO), Organisation internationale de métrologie légale (OIML), Union internationale de chimie pure et appliquée (UICPA), et Union internationale de physique pure et appliquée (UIPPA).

28. Hacking (1992), p.152.

29. Quinn (2002), p.1.

30. Joint Committee for Guides in Metrology (JCGM) (2012), http://www.bipm.org/utis/common/documents/jcgm/JCGM_200_2012.pdf

Le VIM a connu trois éditions différentes, en 1984, 1993 puis 2008. En 2012, le document a connu une révision mineure, pour corriger quelques erreurs de forme; la version de 2012 demeure donc la troisième édition du VIM. Celle-ci est couramment appelée « VIM3 ». Par la suite, nous nous référerons systématiquement à la version de 2012 lorsque nous mentionnerons le contenu du VIM3. Les deux premières éditions seront également désignées par les termes respectifs de VIM1 et VIM2.

31. Joint Committee for Guides in Metrology (JCGM) (2008d), http://www.bipm.org/utis/common/documents/jcgm/JCGM_100_2008_F.pdf

La première version du GUM a été publiée par l'ISO (Organisation Internationale de Normalisation) en 1993, puis a été republiée sous une forme très légèrement remaniée en 1995. En 1997, l'ISO a transféré au BIPM (Bureau International des Poids et Mesures) la charge du maintien et de la mise à jour du document. Le BIPM a publié en 2008 une version librement accessible, ne contenant que des modifications mineures par rapport à la version de 1995. Il n'y a donc à ce jour qu'une unique édition du GUM. Dans cet article, nous nous référerons exclusivement au document de 2008. Le VIM et le GUM sont publiés conjointement en français et en anglais. Nous citerons par la suite les versions françaises de ces documents. Le chapitre 12 de cette thèse propose un aperçu plus détaillé de l'histoire de ce document.

Le VIM et le GUM constituent des documents essentiels pour étudier la façon dont est aujourd'hui comprise la notion d'incertitude de mesure en métrologie, ainsi que ses méthodes de calcul. L'analyse du GUM se révèle d'autant plus instructive que les enseignements que l'on en tire ne se limitent pas au contenu du texte lui-même, mais également aux débats nourris que la publication du document a engendrés. De fait, le GUM est le produit d'une volonté commune, de la part des métrologues, d'aboutir à un consensus concernant les méthodes d'évaluation des résultats de mesure et de l'incertitude qui leur est associée, dans le but d'uniformiser les méthodes employées par les scientifiques dans un large spectre de disciplines scientifiques, et à l'échelle internationale. Or, le GUM n'a pas totalement rencontré le succès escompté. De fait, loin de clore les débats concernant les méthodes d'analyse d'incertitude, sa publication a au contraire suscité de nombreuses réactions critiques chez des métrologues et statisticiens, mais également chez les utilisateurs auxquels il est destiné, parmi lesquels scientifiques et industriels. Dans le courant de ces discussions, associées initialement à des préoccupations méthodologiques, liées à des objectifs en grande partie pratiques, métrologues et scientifiques en sont venus à aborder des questions pleinement épistémologiques. Nous souhaitons mettre en évidence ces questionnements dans notre description des approches probabilistes de l'incertitude de mesure, que nous étudierons en prenant le GUM et le VIM comme point de départ.

Première partie

L'interprétation probabiliste de l'incertitude de mesure

Vue d'ensemble de la partie I

Cette partie vise à étudier les fondements probabilistes de deux modèles statistiques qui sont employés en métrologie pour effectuer l'analyse de l'incertitude attachée aux résultats de mesure. On compte d'une part l'approche souvent appelée « traditionnelle » de la mesure, attachée à un idéal d'objectivité de la science, et qui s'appuie sur une interprétation fréquentiste des probabilités ; on trouve d'autre part une approche bayésienne, bien plus récente, qui tend aujourd'hui à prendre le pas sur le modèle fréquentiste, et fait appel à une interprétation épistémique des probabilités tout en revendiquant le caractère irrémédiablement subjectif de l'activité de mesure.

Nous décrirons tour à tour ces deux modèles – qui à eux deux ne couvrent pas l'ensemble des pratiques des métrologues et des scientifiques – en nous plaçant du point de vue de la métrologie contemporaine. Nous nous appuierons donc en particulier sur les guides de métrologie, en particulier le *Guide pour l'expression de l'incertitude de mesure* (GUM, 1993), mais aussi ses suppléments, ainsi que le *Vocabulaire International de Métrologie* (VIM, 1984, 1993, 2008). À ces références s'ajoutent la littérature contemporaine de métrologie, suscitée en partie en réaction à la publication de ces guides, ainsi que des textes dont la publication remonte à des dates antérieures du XX^e siècle, et qui nous permettront d'avoir un meilleur aperçu des conceptions traditionnelles de la mesure.

Notre description ne se veut pas simplement un résumé des deux approches. Plutôt, nous souhaitons mettre de l'ordre dans les différents principes épistémologiques et techniques sur lesquels chacune des approches est fondée. Ces principes apparaissent à la lecture de la littérature spécialisée, mais, bien qu'ils soient parfois discutés de façon explicite par les acteurs du monde scientifique, il n'en reste pas moins difficile d'en avoir un aperçu global. Notre analyse s'attache à répondre à cette difficulté en formulant une synthèse critique et structurée des deux approches, qui permet de mettre en évidence un certain nombre de questionnements ignorés ou traités de façon périphérique dans la littérature spécialisée. Nous montrons en particulier que les deux modèles ne se distinguent pas tant par les modes de calcul mais par l'*objet* des probabilités : l'approche fréquentiste attribue des probabilités à des événements physiques là où l'interprétation épistémique concerne des propositions.

Néanmoins, cette partie n'est pas non plus simplement consacrée à une description comparative des deux approches. Elle vise également à expliquer pourquoi et comment l'approche traditionnelle de la mesure a été progressivement remise en question jusqu'au point où certains métrologues ont considéré nécessaire de faire appel à un modèle alternatif, puisant à des conceptions sensiblement différentes, et accompagné d'une philosophie explicitement subjek-

tiviste. Nous montrons en particulier que certains avantages et certaines des limites les plus remarquables de l'approche traditionnelle – et, de manière générale, des deux approches – sont relatives à des objectifs bien précis, eux-mêmes adossés aux besoins des usagers.

Notre démarche n'est ni une démarche strictement historique, ni une pure reconstruction logique. D'un côté, nous ne chercherons pas à rendre fidèlement compte de la multiplicité des conceptions et des pratiques que l'on peut observer tout au long du XX^e siècle. Nous prendrons le parti de décrire les modèles selon le point de vue de la métrologie contemporaine. Cependant, ce point de vue est tantôt lacunaire, tantôt partial est il nous sera nécessaire de puiser dans l'histoire de la métrologie la signification de certains concepts et le contenu de certaines pratiques qui ont parfois été perdus dans la présentation que proposent les textes contemporains.

Le chapitre 2 est consacré à l'exposition de certains fondements généraux de l'analyse d'incertitude. Il expose en particulier la distinction classique que les scientifiques opèrent entre erreurs aléatoires et systématiques, et le rôle de cette distinction pour guider les méthodes d'analyse d'incertitude.

Dans le chapitre 3, nous décrivons l'approche fréquentiste de la mesure. Nous montrons que les probabilités y sont employées pour représenter un *processus de mesure* et non pas des résultats de mesure en eux-mêmes. Nous isolons les différentes hypothèses sur lesquelles est fondé le modèle, en montrant en quoi ces hypothèses permettent l'emploi même de probabilités et guident les conclusions que l'on peut en tirer. Nous montrons ensuite que le choix d'une interprétation fréquentiste des probabilités vient avec une contrepartie, l'impossibilité de traiter certaines composantes d'erreur – les erreurs dites « systématiques » – de façon probabiliste, et nous expliquerons alors pourquoi cela a été perçu comme une limite du modèle.

Nous exposons dans le chapitre 4 la conceptualisation bayésienne de la mesure. Cela est fait en deux temps. Nous montrons d'abord comment l'approche classique peut être aménagée au moyen de l'interprétation épistémique des probabilités, laquelle permet de déterminer l'incertitude associée aux erreurs systématiques selon un calcul probabiliste. puis nous présentons la façon dont cette solution a ensuite été étendue aux erreurs aléatoires dans un modèle explicitement bayésien de données, qui indique lui-même la direction d'une approche *entièrement* bayésienne.

Enfin, dans le chapitre 5, nous revenons sur les questionnements conceptuels et philosophiques qu'engendre la confrontation des deux modèles statistiques à propos de la nature et de la fonction de la mesure en science. De fait, les discussions engagées dans le champ métrologique prennent un tour en partie philosophique, qui mobilise en particulier une réflexion sur le caractère objectif ou subjectif de la mesure et sur le rapport entre connaissance et description d'une réalité physique. Cela nous amènera à examiner la relation entre deux concepts métrologiques auxquels nous apportons une attention particulière, l'« incertitude de mesure » et l'« exactitude de mesure ».

Chapitre 2

Fondements généraux de l'analyse d'incertitude

Dans notre exposition générale de l'erreur et l'incertitude de mesure au chapitre 1, nous avons évoqué le fait que l'analyse d'incertitude procède en deux étapes successives. Toutes deux s'appuient sur un élément de mathématisation de l'expérience, la « fonction de mesure », que l'on écrit de façon générique :

$$Y = f(X_1, \dots, X_p) \quad (2.1)$$

Rappelons que cette fonction explicite la façon dont la valeur du mesurande Y est déterminée à partir des grandeurs « d'entrée » X_i . Elle guide également l'analyse d'incertitude. En effet, une première étape concerne l'*identification* des sources d'erreur portant sur chacune des grandeurs d'entrée X_i ; celle-ci mobilise un travail de spécialiste, spécifique aux phénomènes étudiés et aux méthodes expérimentales employées. Pour chaque source d'erreur, il faut alors déterminer l'incertitude associée, qui résulte de l'impossibilité d'éliminer, de corriger, ou de réduire entièrement la source d'erreur en question. Ce faisant, on entre dans une seconde étape de l'analyse d'incertitude, qui porte sur la quantification de l'incertitude de mesure elle-même. Cette quantification mêle deux problèmes, à savoir la détermination de l'incertitude associée à chaque grandeur d'entrée de la fonction de mesure d'une part, et la « propagation » de l'incertitude d'autre part, c'est-à-dire l'opération par laquelle on peut déterminer l'incertitude sur le mesurande Y à partir de l'incertitude associée aux grandeurs d'entrée.

Nous comprenons donc que l'identification des erreurs de mesure, lesquelles sont autant de sources d'incertitude de mesure, et la quantification des composantes d'incertitude de mesure associées, sont deux opérations bien distinctes. Si la première fait avant tout appel au savoir spécialisé et demande une réflexion au cas par cas sur les possibles erreurs de mesure qui affectent l'expérience, la seconde est une opération qui fait appel à une méthode mathématique que l'on peut chercher à généraliser à des disciplines variées. En particulier, les méthodes de quantification de l'incertitude de mesure s'enracinent dans une classification des erreurs de mesure qui prend la forme d'une dichotomie entre erreurs aléatoires et systématiques. Cette séparation est intimement liée au traitement probabiliste de l'incertitude de mesure, et ce dès les

début de la « théorie des erreurs », dont elle a justement permis de structurer les fondements¹. Dans ce chapitre, nous nous proposons d'expliciter la nature de la distinction entre erreurs aléatoires et systématiques de mesure, afin de comprendre en quoi cette distinction permet d'envisager le traitement quantitatif de l'incertitude de mesure.

Nous montrerons dans la section 2.1 que la classification en erreurs aléatoires et systématiques se fait selon deux critères distincts, l'un temporel, l'autre épistémique.

Dans la section 2.2, nous expliquerons alors que c'est une réflexion sur les conditions dans lesquelles sont menées la mesure qui permet de réunir les deux critères précédents de façon à construire les catégories d'erreur aléatoire et d'erreur systématique de façon univoque.

Nous insisterons alors, dans la section 2.3, sur l'importance de distinguer le travail d'identification et d'analyse des *causes* d'erreur de celui de classification des erreurs selon la catégorie aléatoire ou systématique. En effet, le premier porte sur la nature des erreurs elles-mêmes (et de leurs causes), alors que le second engage une réflexion sur la façon dont ces erreurs *seront ensuite* traitées quantitativement de façon à déterminer l'incertitude associée.

Nous concluons alors dans la section 2.4 sur les fondements généraux de l'analyse d'incertitude et nous relierons cela aux différentes méthodes statistiques que l'on peut envisager pour le calcul de l'incertitude de mesure.

2.1 La dichotomie entre erreurs aléatoires et systématiques

Nous avons vu dans le chapitre 1 que l'incertitude de mesure apparaît comme le corrélat d'un processus de correction et de réduction des erreurs de mesure. La détermination d'une incertitude de mesure dépend ainsi de la façon dont on peut appréhender les différentes erreurs de mesure affectant un dispositif donné. Or, ces erreurs se manifestent de diverses façons, et les scientifiques ont depuis longtemps introduit, comme support méthodologique à l'analyse d'incertitude, une dichotomie séparant deux types d'erreurs de mesure : les erreurs dites « aléatoires » ou « statistiques »² ; et les erreurs dites « systématiques ». La dichotomie entre erreurs aléatoires et systématiques mélange deux grilles d'analyse qui ne sont pas *a priori* du même type.

D'un côté, elle caractérise les erreurs sous un angle *temporel*. Certaines erreurs ont des causes constantes, et induisent donc toujours (« systématiquement ») le même décalage sur les

1. On retrouve cette distinction dès les travaux de Gauss, [Gauss \(1823\)](#).

2. La notion a connu de nombreuses dénominations différentes, pas toutes équivalentes. On parle parfois d'erreurs « accidentelles », ce qui est moins satisfaisant car sous-entend que les erreurs de ce type ne peuvent être évitées, et que, par opposition, les autres erreurs peuvent l'être. Cela rejoint la critique que formule Hon et que nous reprendrons plus en détail par la suite, voir [Hon \(1989b\)](#), pp.475–476. En 1978, le Bureau National de Métrologie (BNM, aujourd'hui Laboratoire National de métrologie et d'Essais) proposait quant à lui « erreur fortuite », annexe III, p.2. Dans le même document, le Physikalisch-Technische Bundesanstalt (PTB) proposait également le terme très approprié d'« erreur de répétition » ([Kaarls, 1980](#), annexe IV, p.3.). Cette dernière dénomination présente l'intérêt de présenter ce type d'erreur pour ce qu'elle est de prime abord, à savoir une erreur constatée lorsqu'une mesure est répétée. Aujourd'hui, le terme d'« erreur aléatoire » s'est largement imposé, plutôt à juste titre, mais cela présente le défaut notable d'anticiper sur le traitement statistique réservé à ce type d'erreur – la dénomination découle du traitement statistique appliqué à l'erreur, non à la nature de l'erreur elle-même.

données expérimentales. Par conséquent, il n'est pas possible de les mettre en évidence simplement en répétant la mesure un grand nombre de fois. Ces erreurs sont rangées dans la catégorie des erreurs dites « systématiques ». Il peut s'agir de la conséquence d'un phénomène physique comme la dilatation thermique d'un matériau, qui influe sur le résultat de la mesure de sa longueur d'une façon constante si la température à laquelle la mesure est effectuée est elle-même constante. D'autres erreurs ont des causes variables et présentent des fluctuations parfois très rapides d'où une variabilité des données expérimentales ; ce sont les erreurs « aléatoires ». Elles sont la cause d'une dispersion des résultats de mesure et présentent ainsi des caractères que l'on peut observer directement dans la forme des échantillons de données issus de mesurages répétées d'une grandeur. Le critère temporel de séparation entre erreurs aléatoires et systématiques est très souvent illustré en faisant appel à la métaphore des cibles (figure 2.1). Notons que le caractère aléatoire ou systématique d'une erreur n'est pas lié à la nature de la cause de l'erreur elle-même. En effet, pour un même phénomène, une erreur peut être constante ou variable selon les conditions de l'expérience. Ainsi, la dilatation thermique peut être à l'origine d'erreurs aléatoires si la température à laquelle la mesure est effectuée fluctue.

La dichotomie entre erreurs aléatoires et systématiques ne se contente pas de classifier les erreurs selon le critère temporel que nous venons de décrire. D'un autre côté, elle déploie également un critère *épistémique*. En effet, il s'agit de distinguer les erreurs de mesure selon la façon dont on peut les connaître.

On compte en premier lieu les erreurs qui n'échappent pas totalement à notre compréhension et qui peuvent être en partie expliquées, ou, à défaut, prédites. Ici encore, la dilatation thermique d'un matériau peut être à l'origine de ce type d'erreurs. Celle-ci influe sur le résultat de la mesure de la longueur du matériau d'une façon dont on peut rendre compte par une loi physique, et que l'on peut donc corriger, pour peu que l'on connaisse la température du matériau au moment de la mesure – et que l'on dispose du modèle phénoménologique adéquat. On peut également imaginer l'observation d'une dérive constante ou d'une fluctuation périodique d'une grandeur physique qui pourrait être facilement prédictible sans pour autant que son origine physique soit expliquée. Ici encore, ce type d'erreur entre dans la catégorie des erreurs « systématiques ».

Mais à l'inverse, certaines erreurs de mesure ne peuvent pas être corrigées. C'est bien sûr le cas lorsque celles-ci n'ont pas été identifiées, et demeurent de fait inconnues. Mais il existe des cas où l'expérimentateur sait qu'il est en présence d'erreurs de mesure, sans pour autant pouvoir en esquisser une correction. Le terme d'erreur « aléatoire » renvoie à cette catégorie d'erreur. Ici, « aléatoire » n'est pas à interpréter comme le signe d'une fluctuation ou d'une variabilité, mais est à prendre de façon littérale : l'erreur est aléatoire car elle est imprédictible. Ainsi, les erreurs dites « statistiques » ou « aléatoires » au sens temporel de la classification, qui sont associées à la variabilité des résultats d'une mesure répétée, sont bel et bien des erreurs « aléatoires » au sens épistémique de la classification, dès lors que cette variabilité ne peut pas être expliquée ou prédite, puis, par extension, corrigée. C'est cette variabilité, celle que l'on estime impossible à comprendre ou à prédire, dans un état donné de connaissances théoriques et expérimentales, et étant donnés les moyens techniques à disposition – et cette variabilité seulement – que les métrologues ont modélisé au travers du terme consacré d'« erreurs aléa-

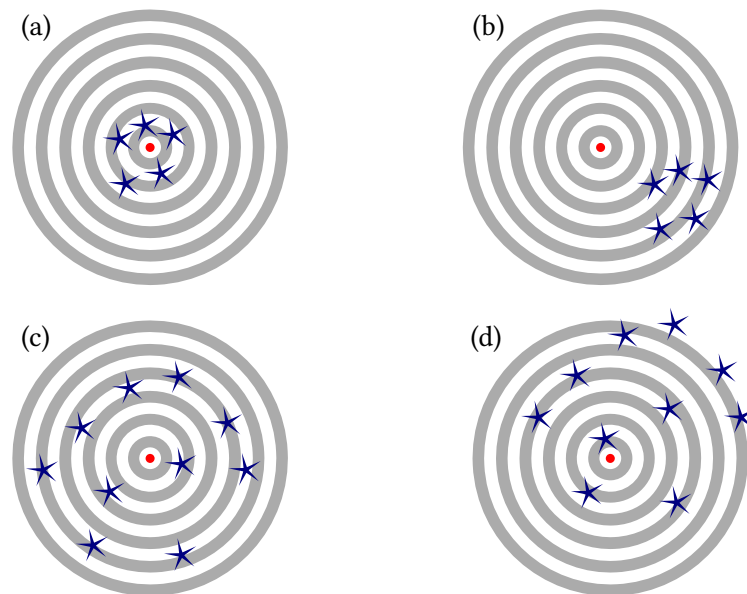


FIGURE 2.1 – Illustration de la dichotomie entre erreurs aléatoires et systématiques à partir de la métaphore des cibles. Des données expérimentales peuvent être resserrées autour de la valeur vraie du mesurande (figure a), auquel cas les composantes aléatoire et systématique de l'erreur sont faibles; la mesure possède une bonne exactitude. Elles peuvent être resserrées autour d'une valeur qui n'est *pas* la valeur vraie (figure b), auquel cas l'erreur systématique est plus élevée que les erreurs aléatoires; la mesure est dite « fidèle » ([Joint Committee for Guides in Metrology \(JCGM\), 2012, p.22](#)) mais elle n'est pas « juste » ([Joint Committee for Guides in Metrology \(JCGM\), 2012, p.21](#)). On peut encore avoir une erreur systématique faible et une erreur aléatoire élevée (figure c) ou des erreurs systématiques et aléatoires toutes deux significatives (figure d). Cette représentation, cependant, n'est valable qu'à condition qu'on puisse identifier effectivement la dispersion des données expérimentales aux erreurs aléatoires; c'est là qu'entre en jeu un second critère de distinction des erreurs.

toires ». Il peut s'agir généralement d'un hasard épistémique³, où l'imprédictibilité des erreurs aléatoires n'est due qu'à un manque de connaissance et aux limites des capacités théoriques et calculatoires, ou encore parfois d'un hasard physique⁴, ce que résume par exemple cette formulation du GUM : « (l)erreur aléatoire provient probablement de variations temporelles et spatiales non prévisibles ou stochastiques de grandeurs d'influence »⁵. Comme nous aurons l'occasion de le préciser dans la prochaine section, puis dans le chapitre suivant, le caractère aléatoire des erreurs du même nom est un attribut qui leur est accordé par une *hypothèse* dans un *modèle* statistique ayant pour but de rendre possible leur traitement. Dans ce cas, la variabilité est effectivement une source bien spécifique d'incertitude de mesure. Nous verrons au prochain chapitre que le calcul de cette incertitude est à la racine des développements probabilistes et statistiques aujourd'hui constitutifs de l'analyse des résultats de mesure en métrologie.

Les deux versants, temporel et épistémique, de la dichotomie entre erreurs aléatoires et systématiques, sont tous deux présents dans les différentes définitions qui ont évolué au gré des éditions du VIM. Celles-ci sont mises en regard dans le tableau 2.1 (p.56). En 1984, le VIM1 définit l'erreur aléatoire par rapport à son caractère *imprévisible*, et à l'impossibilité d'une correction – c'est le versant épistémique de la définition. Mais en 1993, le VIM2 en donne une définition statistique liée à la variabilité temporelle : l'erreur aléatoire est la différence entre le résultat obtenu pour un mesurage individuel et la moyenne d'un échantillon de résultats de taille infinie. Enfin, en 2008, le VIM3 renoue avec la définition épistémique du VIM1, tout en conservant la caractérisation mathématique du VIM2, qui est alors reléguée en note. L'histoire de la définition de l'erreur systématique suit le même mouvement.

Le mélange des genres proposé dans la définition du VIM3 semble de fait produire une synthèse des deux façons différentes de considérer le concept, selon un critère temporel et un critère épistémologique. Pour comprendre cette synthèse, il faut remarquer au préalable que le statut et le traitement de la variabilité des résultats de mesure est intimement liée aux conditions de mesure. Par conséquent, il faut expliciter une distinction classique de la pratique métrologique : la distinction entre conditions de répétabilité et conditions de reproductibilité. Cette distinction permet de mieux comprendre les fondements mathématiques et épistémolo-

3. C'est la position que défend Armatte : « L'analyse a priori et a posteriori des conditions de l'observation par un professionnel de l'astronomie doit lui permettre d'identifier certaines causes d'erreur systématiques, que celles-ci soient dues à l'observateur, aux instruments, à leur interaction (la lecture d'une division) ou à d'autres facteurs (par exemple la parallaxe). En supprimant ou contrôlant ces causes bien identifiées, on peut éliminer leurs effets et donc réduire l'erreur à sa partie « non expliquée ». C'est cela le grand partage qui permet d'extraire le minerai de l'erreur fortuite de sa gangue métaphysique. Mais comme tout dans le monde a une cause – c'est le principe de raison suffisante invoqué par Jacques Bernoulli, Leibniz et John Locke – il est clair que les disparités d'observations qui demeurent après ce nettoyage ont elles aussi des causes, qui nous sont tout simplement inconnues. L'erreur qui demeure après cette « réduction » est attribuée à des causes dite accidentelles, ou encore « au hasard », si ce mot est bien le cache sexe de notre ignorance, de notre méconnaissance partielle de ce qui a troublé l'observation. Dans sa nature métaphysique, l'erreur accidentelle est une erreur épistémique. », [Armatte \(2010\)](#), pp.2–3.

4. Certaines théories physiques cependant des composantes de hasard physique – en particulier, notoirement, la physique quantique. Par conséquent, il n'est pas possible d'affirmer que le hasard associé aux erreurs de mesure est *exclusivement* épistémique. Dans ce cas, la question est alors de savoir avec quelle amplitude ces effets aléatoires influent sur la variabilité globale du système. La question de la nature du hasard dans la caractérisation des erreurs aléatoires n'influe pas, pour l'essentiel, sur notre propos.

5. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.5.

| | |
|----------------------------------|--|
| VIM1 (1984) (p.16) | <p>Erreur aléatoire : composante de l'erreur de mesure qui, lors de plusieurs mesurages du même mesurande, varie d'une façon imprévisible</p> <p>NOTE : on ne peut pas tenir compte d'une erreur aléatoire par application d'une correction.</p> |
| VIM2 (1993) (p.26) | <p>Erreur aléatoire : résultat d'un mesurage moins la moyenne d'un nombre infini de mesurages du même mesurande, effectués dans les conditions de répétabilité</p> <p>NOTE 1 : L'erreur aléatoire est égale à l'erreur moins l'erreur systématique.</p> <p>NOTE 2 : Comme on ne peut faire qu'un nombre fini de mesurages, il est seulement possible de déterminer une estimation de l'erreur aléatoire.</p> |
| VIM3 (2008) (p.23) | <p>Erreur aléatoire : composante de l'erreur de mesure qui, dans des mesurages répétés, varie de façon imprévisible</p> <p>NOTE 1 : La valeur de référence pour une erreur aléatoire est la moyenne qui résulterait d'un nombre infini de mesurages répétés du même mesurande.</p> <p>NOTE 2 : Les erreurs aléatoires d'un ensemble de mesurages répétés forment une distribution qui peut être résumée par son espérance mathématique, généralement supposée nulle, et par sa variance.</p> <p>NOTE 3 : L'erreur aléatoire est égale à la différence entre l'erreur de mesure et l'erreur systématique.</p> |
| VIM1 (1984) (p.16) | <p>Erreur systématique : composante de l'erreur de mesure qui, lors de plusieurs mesurages du même mesurande, reste constante ou varie d'une façon prévisible.</p> <p>NOTES : 1. Les erreurs systématiques et leurs causes peuvent être connues ou inconnues. [...]</p> |
| VIM2 (1993) (p.27) | <p>Erreur systématique : moyenne qui résulterait d'un nombre infini de mesurages du même mesurande, effectués dans les conditions de répétabilité, moins une valeur vraie du mesurande.</p> <p>NOTE 1 : L'erreur systématique est égale à l'erreur moins l'erreur aléatoire.</p> <p>NOTE 2 : Comme la valeur vraie, l'erreur systématique et ses causes ne peuvent pas être connues complètement.</p> |
| VIM3 (2008) (pp. 22-23) | <p>Erreur systématique : composante de l'erreur de mesure qui, dans des mesurages répétés, demeure constante ou varie de façon prévisible.</p> <p>NOTE 1 : La valeur de référence pour une erreur systématique est une valeur vraie, une valeur mesurée d'un étalon dont l'incertitude de mesure est négligeable, ou une valeur conventionnelle.</p> <p>NOTE 2 : l'erreur systématique et ses causes peuvent être connues ou inconnues. On peut appliquer une correction pour compenser une erreur systématique connue.</p> <p>NOTE 3 : l'erreur systématique est égale à la différence entre l'erreur de mesure et l'erreur aléatoire.</p> |

TABLEAU 2.1 – Différentes définitions d'« erreur aléatoire » et « erreur systématique » dans les trois éditions du *Vocabulaire International de Métrologie* (VIM).

| | |
|------------------------------|---|
| VIM3 (2008) (p.23) | Conditions de répétabilité : condition de mesurage dans un ensemble de conditions qui comprennent la même procédure de mesure, les mêmes opérateurs, le même système de mesure, les mêmes conditions de fonctionnement et le même lieu, ainsi que des mesurages répétés sur le même objet ou des objets similaires pendant une courte période de temps |
| VIM3 (2008) (p.24) | Conditions de reproductibilité : condition de mesurage dans un ensemble de conditions qui comprennent des lieux, des opérateurs et des systèmes de mesure différents, ainsi que des mesurages répétés sur le même objet ou des objets similaires |

TABLEAU 2.2 – Définitions de « Conditions de répétabilité » et « conditions de reproductibilité » dans la troisième édition du *Vocabulaire International de Métrologie* (VIM).

giques de la distinction entre erreurs « systématiques » et erreurs « aléatoires », et, de plus, nous permettra de préciser la façon dont on peut alors donner un cadre probabiliste au traitement des erreurs aléatoires, et éventuellement au traitement des erreurs systématiques (ce qui se révélera une tâche plus délicate).

2.2 Répétabilité et reproductibilité

La double nature de la dichotomie entre erreurs aléatoires et systématiques est délicate en raison du fait qu'une erreur dite « systématique » n'est pas toujours constante. La définition du VIM3 indique qu'elle peut « varier de façon prévisible ». Cette définition n'est pas la première à prendre acte de cet état de fait, et certains scientifiques ont déjà expliqué à différentes occasions que ce qui est souvent appelé « erreur systématique » pouvait être effectivement variable⁶. De fait, le caractère constant ou variable d'une erreur n'est pas lié à la nature de l'erreur, mais aux conditions dans lesquelles la mesure est répétée. Selon les conditions de mesure, un même phénomène physique peut avoir un effet constant ou variable. C'est pourquoi, pour comprendre la façon dont les versants épistémique et temporel de la dichotomie entre erreurs aléatoires et systématiques sont intriqués, nous devons faire appel à la typologie que les métrologues ont eux-mêmes développée à propos des *conditions* d'expérience. Les métrologues distinguent ainsi les « conditions de répétabilité » des « conditions de reproductibilité » lorsqu'ils traitent de la répétition des mesures d'une même grandeur – c'est-à-dire les mesures dans lesquelles le mesurande est (par définition) le même. Les définitions de ces deux termes, telles qu'elles apparaissent dans le VIM3, sont reportées dans le tableau 2.2.

6. « L'erreur systématique d'un processus de mesure aura ordinairement à la fois des composantes constante et variable », Eisenhart (1963), p.170. « Les incertitudes peuvent être divisées en un certain nombre de groupes différents. L'un d'eux comprend des effets systématiques qui varient d'une observation à la suivante, et de cette façon sont en mesure de contribuer à la dispersion des observations. », Petley (1988), p.291.

L'hypothèse centrale commune aux deux conditions est l'*identité du mesurande* : la même grandeur est visée à chaque fois. Or, dans le cadre de la conception contemporaine de la mesure, la grandeur visée – le mesurande – est l'objet d'une définition. La définition des conditions « de reproductibilité » pourrait paraître triviale : il y a reproductibilité dès lors que le mesurande est *par définition* le même, et ce quelles que soient les conditions effectives de mesure. Cet aspect est moins anodin qu'il ne peut y paraître : l'idée de « conditions de reproductibilité » consiste essentiellement, sur le plan épistémologique, à revendiquer qu'il est possible de viser une même grandeur par des moyens différents, à l'opposé d'une posture opérationnaliste, pour laquelle « la signification d'un concept est entièrement spécifiée par sa méthode de mesure »⁷, et pour laquelle à chaque méthode de mesure correspond un concept différent.

Dans les conditions dites de répétabilité, les conditions de mesure sont censées être très proches – bien qu'il soit impossible de garantir leur parfaite identité. Par suite, il est attendu qu'il n'y ait que de faibles perturbations du dispositif de mesure, et une faible variabilité du mesurande lui-même (si celui-ci a été défini de telle façon qu'il puisse varier). Spécifier, pour un expérimentateur, que l'expérience est menée dans des conditions de répétabilité sert de marqueur épistémologique, dans le sens où l'expérimentateur précise ainsi qu'il considère qu'il ne voit aucune raison *bien identifiée* pour que les résultats expérimentaux varient. Dès lors, toute variabilité des résultats de mesure est la trace d'une erreur de mesure ; mais, de plus, puisqu'il n'existe aucune raison bien identifiée à cette variabilité, c'est qu'elle ne peut pas non plus être corrigée. Par conséquent, lorsqu'un expérimentateur stipule qu'il est dans des conditions de répétabilité, il indique qu'il ne pourra pas – ou ne souhaitera pas – proposer un modèle physique aux variations des résultats de mesure qu'il observera, s'il en observe⁸.

Les conditions de répétabilité se définissent comme les conditions dans lesquelles l'expérimentateur affirme ne plus pouvoir être capable ou ne plus souhaiter faire de différence entre les conditions dans lesquelles ont lieu l'expérience⁹. Il n'y a pas alors d'autre solution pour lui que de constater son incapacité à identifier individuellement les sources d'erreur. Il s'en remet donc à un traitement statistique : à défaut de pouvoir connaître le mécanisme physique à l'origine de chaque erreur de mesure, il considère le processus de mesure comme une « boîte noire » dont il serait vain de chercher à révéler la dynamique interne, mais dont il peut analyser les

7. Chang et Cartwright (2008), p.367.

8. Comme le précise Willink, « l'idée de ne pas savoir ou de ne pas se soucier de la provenance de l'erreur est associé au terme de 'pure erreur' parfois utilisée en statistique », Willink (2013), p.24.

9. « Il est clair ensuite que la solution mathématique ne peut intervenir qu'une fois reconnues deux propriétés de l'erreur, à savoir son irréductibilité par des moyens physiques, et son indépendance de toute cause identifiable. Il faut "nettoyer" la notion d'erreur où se mêlent des considérations physiques et psychologiques, où entre en jeu l'infinité des conditions de l'observation, pour que la mathématique puisse en faire l'objet d'une unique théorie. Exit donc de celle-ci les erreurs systématiques assignables à des "causes constantes". Les mathématiciens ne traiteront que de "mesures faites dans les mêmes circonstances, et avec un égal soin" pour prendre une formulation ultérieure de Gauss. Il y a des conditions à la mise en équivalence qui précède ou sous-tend un calcul, fut-ce celui d'une moyenne, et dont Jean Bernoulli ne dit rien, bien que cette construction même du concept d'erreur aléatoire ait été obtenu par une lente décantation de la prénotion commune d'erreur dont l'on trouve trace dans les carnets de relevés et les compte-rendu minutieux des praticiens de l'astronomie », Armatte (2004), p.147.

relations entrée-sortie¹⁰. Cela sous-tend un traitement *collectif* des erreurs de mesure – visant à leur réduction statistique – plutôt qu’un traitement *individuel* visant à leur correction¹¹.

C’est pourquoi l’idée de conditions de répétabilité permet de comprendre les fondements d’une idée d’erreur « aléatoire » telle qu’elle est conçue dans les modèles statistiques de la mesure. Une variation des données expérimentales, qui apparaît dans des conditions de répétabilité, est quelque chose qu’il est par définition impossible de prédire autrement que par son comportement statistique (moyenne, amplitude des erreurs, etc.). C’est à ces conditions qu’il est justifié de supposer que l’opération de mesure est une opération aléatoire¹² – et qu’il en est de même pour les erreurs commises – ce qui permet d’engager un travail statistique.

La définition des conditions de répétabilité permet finalement de concilier les deux façons différentes de concevoir la dichotomie entre « erreurs aléatoires » et « erreurs systématiques » que les définitions du VIM3 superposent – à savoir selon un critère temporel et selon un critère épistémique. Ainsi, les conditions de répétabilité tracent la frontière entre ce que l’expérimentateur affirme pouvoir expliquer et ce qu’il laisse au traitement statistique. Par conséquent, si les erreurs systématiques correspondent aux erreurs « prévisibles », alors, *dans les conditions de répétabilité*, celles-ci ne peuvent être que constantes. Réciproquement, sous ces mêmes conditions, toute erreur variable est une erreur « aléatoire ». Il nous faut donc insister sur ce point précis : *il n’y a pas de différence de nature entre les phénomènes qui causent les erreurs aléatoires et ceux qui causent les erreurs systématiques*; la différence provient des conditions de mesure.

L’intérêt de la définition des conditions de répétabilité et de la distinction entre erreurs aléatoires et systématiques réside, en fin de compte, dans le mode de traitement qu’ils suggèrent pour appréhender chaque type d’erreur. Nous accordons à la répétabilité l’idée d’une immédiateté du traitement statistique et, dans ces conditions, la variabilité présente un statut

10. Cela rejoint les remarques de Bogen et Woodward sur le caractère idiosyncratique des données d’expérience, [Bogen et Woodward \(1988\)](#). Ces auteurs font remarquer que les données qui permettent d’attester l’existence des phénomènes ont lieu dans des conditions très précises, et dépendent du protocole expérimental mis en place. Elles sont le produit d’une conjonction *bien spécifique* de causes variées. L’un des objectifs de l’expérimentateur est de remonter la chaîne causale pour garantir que les données obtenues mettent effectivement en évidence des caractéristiques du phénomène qu’il souhaite étudier; il faut en quelque sorte extraire le phénomène des données.

11. Kyburg a beaucoup insisté sur le caractère collectif du traitement des erreurs de mesure : « (l)e seul sens dans lequel nous pouvons considérer que nous *mesurons* des erreurs de mesure est collectivement – en déterminant, dans les unités données, la distribution statistique des erreurs produite par une certaine procédure de mesure. Le point ici est que nous ne mesurons pas, et ne pouvons pas mesurer, une “erreur” seule. Il y a des exceptions à cela – par exemple, lorsque nous étalonnons un instrument de mesure [...] mais cela n’est pas le cas ordinaire de traitement de l’erreur. », [Kyburg \(1992a\)](#), p.76. Cette remarque de Kyburg s’insère dans une critique de la théorie représentationnelle de la mesure qui voit dans la structure mathématique des grandeurs physiques un moyen de représenter la structure empirique des propriétés physiques : « (m)a position est qu’un point de vue représentationnel sur la mesure rend le traitement quantitatif des erreurs de mesure impossible. », p.76. Les remarques de Kyburg sont tout à fait valables, si l’on précise toutefois qu’elles s’appliquent au cas des erreurs aléatoires. La façon dont Kyburg écarte un peu rapidement l’évaluation individuelle des erreurs systématiques pourrait donner à croire que celle-ci est accessoire ou périphérique, alors qu’il s’agit d’un élément central de l’analyse d’erreur.

12. C’est ainsi que l’opération de mesure est présentée aux enseignants du secondaire : « (f)aire une mesure, c’est réaliser une expérience qui comporte une part d’aléatoire : chaque mesure est un tirage dans l’ensemble (en général infini) des résultats possibles; il y a fluctuation d’échantillonnage d’une série de mesure à une autre. », [Ministère de la Jeunesse, de l’Éducation nationale et de la Recherche \(2002\)](#), p.87.

un peu privilégié du fait qu'elle fournit de façon immédiate la trace d'une erreur de mesure, comme le souligne Colclough :

Bien que la nature exacte de la distinction entre erreurs systématiques et aléatoires soit souvent l'objet de confusions, l'intention pratique qu'il y a derrière est suffisamment claire. Elle résulte de la perception du fait que certaines erreurs, les erreurs 'aléatoires', peuvent être traitées statistiquement et en principe réduites à n'importe quel niveau souhaité *seulement sur la base des résultats*, alors que d'autres, en raison d'une tendance à agir dans une direction particulière, ne le peuvent pas.¹³

À l'inverse, les erreurs systématiques ne se manifestent pas directement, de façon observable, dans les données expérimentales. En revanche, si elles peuvent être prédites, elles font alors l'objet d'une correction au cas par cas qui appelle une analyse d'incertitude qui, en principe, devrait différer de celle des erreurs aléatoires.

Ces considérations annoncent les différents traitements statistiques qui ont été proposés pour les erreurs aléatoires et les erreurs systématiques, traitements que nous expliciterons dans les prochains chapitres. Notons en passant que, si le traitement de la variabilité a historiquement été à l'origine d'un développement statistique qui s'est révélé particulièrement fécond, la variabilité n'est pas pour autant la racine unique de la notion d'incertitude de mesure sur le plan épistémologique. Il serait donc particulièrement erroné d'identifier le *calcul d'incertitude* avec le seul traitement statistique des erreurs dites « aléatoires »¹⁴.

2.3 Typologie des erreurs et analyse d'erreur

Giora Hon a proposé un examen critique de la question de la classification des erreurs expérimentales¹⁵, dans un article divisé en deux parties. La première est une critique de la façon dont philosophes et scientifiques envisagent la question de l'erreur expérimentale, et conclut sur un plaidoyer envers une classification de l'erreur expérimentale qui s'appuie sur des critères épistémologiques et non mathématiques. Dans la seconde partie, Hon propose sa propre classification des erreurs. Selon Hon, la classification que les philosophes et les scientifiques proposent usuellement ne s'attache pas à la *source* des erreurs. Au contraire, en appliquant un critère purement mathématique, elle fait de l'analyse d'erreur une routine qui, de ce fait,

13. Colclough (1987), pp.171-172 (nous soulignons).

14. L'importance accordée aux erreurs dites « statistiques » dans les manuels peut mener à l'illusion selon laquelle le traitement de ces erreurs constitue l'essentiel du travail d'analyse d'incertitude. Hon formule une critique dans ce sens : « La plupart des auteurs dans ce domaine se satisfont de quelques remarques concernant l'origine et le traitement des erreurs systématiques. Ayant brièvement examiné les erreurs systématiques, ces auteurs procèdent à une analyse détaillée de la théorie des erreurs qui fournit l'outil mathématique pour traiter les erreurs aléatoires », Hon (1989b), pp.477-478. Il n'est pas surprenant que le traitement des erreurs aléatoires, auquel correspond une méthode systématique, soit plus détaillé que celui des erreurs systématiques, qui se traitent au cas par cas. Cependant, il n'est pas pour autant justifié de réduire l'analyse d'erreur à l'analyse des erreurs aléatoires.

15. Hon (1989b)

n'apporte rien à la question épistémologique du « problème de l'erreur » en science¹⁶. Hon propose en retour une classification de l'erreur fondée sur des critères épistémologiques, et structurée autour d'un découpage d'une expérimentation en quatre étapes, auxquelles correspondent quatre origines générales de l'erreur expérimentale¹⁷.

La critique que Hon formule dans la première partie de son article présente l'intérêt de rappeler que la question de l'erreur expérimentale n'est pas, en reprenant la citation qu'il tire de Mellor, « une excroissance fastidieuse, mais triviale, de la pure structure déductive de la science »¹⁸. Cependant, nous pouvons à notre tour nuancer cette critique. En effet, Hon laisse entendre que la dichotomie entre erreurs aléatoires et systématiques serait considérée comme le point de départ de tout travail d'analyse d'erreur et, ce faisant, se substituerait à tout travail préalable d'identification des sources d'erreurs, qu'elle rendrait inutile. Or, l'analyse des sources d'erreur de mesure et le traitement de ces erreurs sont deux étapes *bien différentes* de l'analyse d'erreur, qui ne font pas appel aux mêmes connaissances ni aux mêmes méthodes. Hon note à raison que la classification en erreurs aléatoires et systématiques est celle que l'on retrouve le plus couramment chez les philosophes et les scientifiques¹⁹. Effectivement, dans nombre de présentations classiques de manuels ou de documents techniques, la dichotomie entre erreurs aléatoires et systématiques est souvent déployée dès les premiers chapitres, et joue un rôle central. Mais cela est dû au fait que la majeure partie de ces textes ont pour objet même l'étape de quantification de l'incertitude, et non l'étape d'identification des sources d'incertitude, car, comme nous l'avons expliqué à la section précédente, c'est la première qui s'appuie sur des méthodes transversales, communes aux différentes disciplines, alors que l'étape précédente d'identification des sources d'incertitude fait partie intégrante d'un travail de spécialiste. La dichotomie elle-même a pour but de guider le traitement statistique des erreurs de mesure, et non pas leur identification préalable. Il nous semble ainsi que la critique de Hon ne tient pas compte du fait que les scientifiques considèrent bel et bien comme une partie essentielle de leur travail que de raisonner sur les sources possibles d'erreur de mesure, leurs causes et les phénomènes qui sont à leur origine, avant même d'entrer dans le questionnement d'ordre statistique qui, lui, s'appuie effectivement sur une classification mathématique des erreurs de mesure. Les scientifiques ont d'ailleurs eux-mêmes proposé des classifications qui s'apparentent, avec des

16. « Deux objections étroitement liées peuvent être soulevées contre cette classification en erreurs systématiques et aléatoires. Premièrement, comme les protagonistes de cette classification sont intéressés dans l'erreur résultante et non pas tant dans sa source, ils sont voués à classer ensemble des phénomènes qui peuvent en effet agir comme s'ils étaient du même type d'erreur, par exemple l'erreur aléatoire, mais qui se distinguent pourtant quant à leurs causes. [...] Pour le dire autrement, la dichotomie entre erreurs systématiques et aléatoires ne se concentre pas sur la source de l'erreur; plutôt, elle examine la nature de l'erreur en appliquant un critère mathématique. [...] Le fait que le critère qui sous-tend cette dichotomie est mathématique constitue la seconde objection » Hon (1989b), p.477.

17. Les quatre origines en question sont (1) la théorie d'arrière-plan, (2) les hypothèses concernant le protocole expérimental utilisé et son fonctionnement, (3) les rapports d'observation et (4) les conclusions théoriques. Voir en particulier le tableau synthétique, Hon (1989b), p.480.

18. Hon (1989b), p.472, tirée de Mellor (1967), p.6.

19. « La classification la plus courante des erreurs expérimentales est la classification qui distingue entre deux catégories de l'erreur : erreurs systématiques et aléatoires », Hon (1989b), p.474. Hon indique ensuite : « De nombreux textes et manuels sur la conception des expériences et le traitement des données peuvent attester de l'acceptation générale de cette classification qui est même parfois considérée comme exhaustive », p.475.

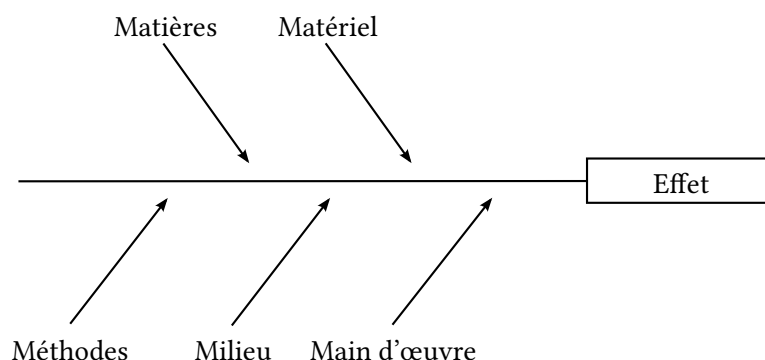


FIGURE 2.2 – Exemple de diagramme dit d'Ishikawa (conçu à partir d'un modèle tiré de wikipedia, https://fr.wikipedia.org/wiki/Diagramme_de_causes_et_effets). Ce diagramme a pour but de recenser par catégories un ensemble de causes amenant à un effet donné. Il offre donc en particulier une structure pour rechercher des causes d'erreur dans un processus de mesure. Il est utilisé en particulier dans la gestion de la qualité dans l'industrie, en ingénierie, ou dans l'entreprise.

vertus différentes, à celle que propose Hon dans la seconde moitié de son article. Parmi les plus populaires, on compte celle qu'a proposé à la fin des années 1960 l'ingénieur et chimiste japonais Kaoru Ishikawa, dans son travail sur la gestion de la qualité, et que l'on peut représenter par un diagramme dit « d'Ishikawa », ou « diagramme de causes et effets »²⁰ (voir figure 2.2). Ces diagrammes sont effectivement utilisés, dans des milieux plutôt liés à l'ingénierie, afin de remonter aux racines des erreurs expérimentales²¹. La critique de Hon est donc valable avant tout si elle est tournée vers l'état des travaux de philosophie des sciences à l'époque de sa publication, ou aux présentations parfois trop simplistes que l'on retrouve dans les manuels scientifiques ; appliquée aux *pratiques* des scientifiques, elle ne fonctionne que partiellement. La dichotomie entre erreurs aléatoires et systématiques n'a pas pour but d'expliquer comment les scientifiques doivent chercher à identifier les sources d'erreur et les sources d'incertitude, mais permet de structurer le traitement quantitatif des erreurs de l'incertitude *une fois* ces sources identifiées.

2.4 Erreurs de mesure et méthodes d'analyse d'incertitude

À la lumière des développements précédents, nous pouvons désormais résumer la façon dont on envisage, de façon générale, l'analyse de l'incertitude associée à la mesure d'une grandeur. Nous avons vu que dans une expérience complexe visant à déterminer une grandeur physique donnée (par exemple une constante de la physique, ou encore une caractéristique d'un produit manufacturé que l'on soumet à un test de qualité), on distingue la grandeur visée par

20. Voir Ishikawa (1991). Je remercie Nadine de Courtenay d'avoir attiré mon attention sur ces travaux.

21. Pour une présentation didactique, voir par exemple Barsalou (2014), chapitre 2, ainsi que les figures données p.70.

l'expérience elle-même – le mesurande – des grandeurs par l'entremise desquelles on accède au mesurande le long de la chaîne expérimentale. Ainsi, la mesure d'une longueur peut-elle exploiter à la fois les indications d'appareils de mesure de longueur et des thermomètres (afin d'effectuer les corrections d'effets dus à la dilatation thermique). C'est précisément cela que la fonction de mesure $Y = f(X_1, \dots, X_n)$ représente mathématiquement. La grandeur visée est le mesurande, grandeur *de sortie* Y de la fonction de mesure.

L'analyse d'incertitude fonctionne selon plusieurs étapes. La première d'entre elles consiste à affiner le plus possible la fonction de mesure $Y = f(X_1, \dots, X_n)$ en incorporant une modélisation de tous les biais de mesure que l'on a pu identifier.

Une seconde étape consiste à considérer chaque grandeur d'entrée individuellement. Certaines grandeurs d'entrée sont l'objet d'une mesure au sein du protocole expérimental. Dans ce cas, la mesure peut être affectée par des effets physiques qui induisent des erreurs de mesure. Il faut donc rechercher les possibles sources d'erreurs dites « systématiques », et les corriger au mieux, le cas échéant. À chaque correction est associée une incertitude de mesure qui rend compte du fait que la correction effectuée n'est pas parfaite. Ensuite, en spécifiant les conditions dans lesquelles la mesure de la grandeur d'entrée considérée est répétée, on peut alors circonscrire la variabilité résiduelle des données expérimentales et traiter cette variabilité par le modèle des « erreurs aléatoires » et déterminer une incertitude qui lui est associée. Cela est illustré une fois encore par la métaphore des cibles : on comprend quel est le travail de l'expérimentateur si l'on est conscient que la cible *n'est pas connue* (figure 2.3).

D'autres grandeurs d'entrée sont évaluées à partir d'indications *extérieures* : tables de valeurs obtenues à partir d'expériences différentes menées par d'autres scientifiques avec d'autres moyens et selon d'autres principes de mesure, prédictions théoriques, jugements d'expert, etc. Toutes les erreurs de mesure affectant les indications extérieures que l'expérimentateur exploite se répercuteront sur l'évaluation finale du mesurande. Les métrologues ont à plusieurs occasions proposé de parler à ce propos d'erreurs « fossilisées »²² car, quelle que soit la source de l'erreur, et que l'erreur soit variable ou constante dans l'expérience qui a été menée pour évaluer la grandeur d'entrée correspondante, l'erreur se retrouve fixée une fois le résultat communiqué et exploité par un expérimentateur dans une autre expérience. De ce fait, ces erreurs entrent dans la catégorie des erreurs systématiques²³. En fin de compte, il est essentiel de rendre compte des possibles sources d'erreur affectant les grandeurs d'entrée correspondantes. C'est pourquoi les indications extérieures doivent contenir des informations permettant de déterminer l'incertitude de mesure associée aux valeurs communiquées. Cette catégorie d'informations rappelle que la mesure présente un caractère collectif et qu'un expérimentateur dépend le plus souvent de façon très importante du travail mené par d'autres scientifiques en d'autres lieux et d'autres instants. Les tables de valeurs dans lesquelles on vient chercher une connaissance de certaines grandeurs d'entrée ont précisément pour but d'économiser à l'expérimentateur le besoin de reproduire l'ensemble des travaux qui ont été effectuées *par d'autres*, en d'autres lieux

22. Moffat (1988), p.7; Kirkup et Frenkel (2006), p.45

23. Comme le faisait remarquer le physicien Paul Vigoureux : « (i)l faut se rappeler que certaines erreurs sont aléatoires pour une personne et systématiques pour une autre. [...] Si, dans l'élaboration de mon résultat, je dois utiliser le résultat de [quelqu'un d'autre], alors j'appelle toute erreur possible dans ce résultat une erreur systématique car elle était avec moi tout le temps », Langenberg et Taylor (1971a), p.524.

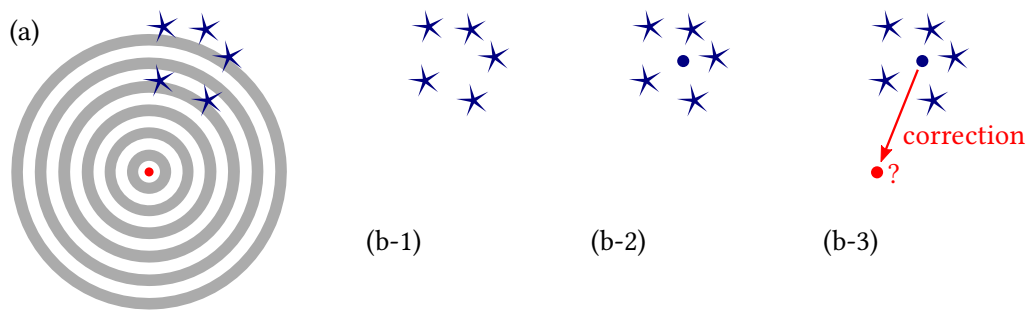


FIGURE 2.3 – Métaphore des cibles et analyse des sources d'incertitude. La situation d'un expérimentateur n'est pas le cas (a) où la cible est connue mais le cas (b-1) où la position de la cible doit être retrouvée à partir des données expérimentales. L'expérimentateur souhaite s'appuyer sur les données expérimentales pour localiser le centre de la cible. Il considérera typiquement la moyenne arithmétique des données, prélude à un traitement statistique *collectif* des erreurs aléatoires (réduction des erreurs, figure b-2). Cependant, cette moyenne ne localise pas le centre de la cible s'il y a présence d'erreurs systématiques. Il faut donc corriger *individuellement* chaque erreur systématique qui a pu être identifiée (figure b-3). Chacune de ces opérations présente des imperfections et l'estimation de ces imperfections mène à quantifier une incertitude de mesure correspondante. Notons que l'expérimentateur aurait tout à fait pu chercher à corriger individuellement chaque donnée expérimentale (mais n'y serait peut-être pas parvenu) : tout dépend de la façon dont il définit les conditions de répétabilité. Dans les conditions de répétabilité, les données variables sont considérées *par définition* comme imprévisibles, et ne peuvent donc être traitées que collectivement.

et avec d'autres moyens (parfois plus performants) – travaux dont les résultats ont justement été collectés et résumés dans des tables. L'expérimentateur aurait pu envisager de reproduire lui-même ces travaux mais cela aurait été à la fois fastidieux et peu utile. Au contraire, la métrologie est une activité qui vise justement à réguler et maintenir l'architecture sociale de la connaissance scientifique et, ce faisant, celle-ci permet à un scientifique de fonder son analyse sur une expérimentation de taille réduite qui tire parti de la somme de connaissances déjà acquises.

Ce faisant, nous avons décrit les deux premières étapes du travail d'analyse d'incertitude. Il reste alors à effectuer une troisième étape, dont nous avons déjà brièvement évoqué l'existence, qui consiste à « propager » les incertitudes relatives aux grandeurs d'entrée de façon à déterminer l'incertitude résultante associée au mesurande Y qui constitue la cible finale de la mesure. Cette question fait appel à un outillage mathématique parfois très développée, comme nous le verrons dans les prochains chapitres, et peut se révéler complexe d'un point de vue technique – de fait, nous verrons qu'elle constitue un point nodal dans l'évolution récente des méthodes statistiques employées en métrologie.

L'approche décrite jusqu'ici est très générale, et il reste désormais à expliciter la façon

dont, concrètement, une incertitude de mesure peut être calculée pour chaque source d'erreur et pour chaque grandeur d'entrée. C'est à ce niveau que la distinction entre erreurs aléatoires et systématiques joue un rôle essentiel. En effet, elle permet de guider l'expérimentateur vers le choix d'une méthode bien déterminée d'analyse d'incertitude, adaptée à chaque type d'erreur. Comme nous le verrons dans les prochains chapitres, différentes méthodes ont été envisagées à la fois pour le traitement des erreurs aléatoires et pour celui des erreurs systématiques.

À la fin des années 1970, les métrologues ont souhaité mettre de l'ordre dans la diversité des pratiques observées selon les laboratoires, les disciplines et les traditions nationales. Le Bureau International des Poids et Mesures (BIPM) a alors réuni un groupe de travail visant à formuler un consensus sur la question de l'analyse d'incertitude. Cette négociation a abouti à la publication par le BIPM, en 1980, d'une recommandation internationale, la « recommandation INC-1 », qui jette les premiers fondements, encore très incomplets, de l'analyse contemporaine d'incertitude. Ceux-ci seront ensuite développés dans le *Guide pour l'expression de l'incertitude de mesure* (GUM), publié environ dix ans plus tard. L'un des mouvements les plus notables de cette recommandation suit une requête exprimée par les métrologues du National Physical Laboratory (NPL) au Royaume-Uni²⁴, qui appelle à distinguer deux catégories de méthode d'analyse d'incertitude, baptisées dans la recommandation INC-1 méthodes « de type A » et méthodes « type B » :

L'incertitude d'un résultat de mesure comprend généralement plusieurs composantes qui peuvent être groupées en deux catégories d'après la méthode utilisée pour estimer leur valeur numérique :

A. celles qui sont évaluées à l'aide de méthodes statistiques,

B. celles qui sont évaluées par d'autres moyens.²⁵

Il est ici entendu par « méthodes statistiques » les méthodes qui s'appuient sur l'exploitation d'échantillons de données numériques – c'est le cas, en particulier, du traitement des erreurs aléatoires. À l'inverse, les méthodes de « type B » sont nécessaires dans les situations où de tels échantillons ne sont *pas* disponibles – c'est le cas des erreurs systématiques lorsqu'on travaille dans des conditions de répétabilité²⁶. L'identification du type d'erreur, aléatoire ou systématique, est importante car elle renvoie aux types d'informations disponibles à propos de la grandeur considérée, et, par suite, à la méthode de mesure employée²⁷.

24. Kaarls (1980), annexe V, p.3.

25. Joint Committee for Guides in Metrology (JCGM) (2008d), p.viii. La recommandation INC-1 est également reproduite, mais pas tout à fait à l'identique, p.30 du même document. Elle fut initialement publiée dans les rapports du CIPM, ainsi que dans *Metrologia*, voir Giacomo (1981), p.73.

26. Il est bien sûr possible d'analyser les erreurs systématiques par une étude statistique sur un échantillon de résultats de mesure, lorsque ces mesures sont effectuées dans des conditions variées dans lesquelles les erreurs dites « systématiques » sont susceptibles de varier elles aussi. Mais ce travail ne peut pas être effectué au sein d'une expérience particulière, où certaines conditions de mesure peuvent varier mais où il est rarement, voire jamais, possible de faire varier l'ensemble des paramètres expérimentaux, et où la mesure est dépendante d'une unique méthode de mesure – changer de méthode de mesure reviendrait tout simplement à changer d'expérience.

27. C'est également ainsi que Willink interprète la recommandation. Willink propose de calquer une classification des erreurs sur celle des méthodes d'évaluation ; dans ce cas, « L'amplitude d'une erreur de type A est estimée à partir des données expérimentales, par exemple l'écart-type d'un échantillon, mais l'amplitude d'une erreur de type B est décidée à partir d'autres considérations, telles que la déclaration d'exactitude d'un fabricant d'équipement », Willink (2013), p.30.

Cependant, rien, dans toutes ces considérations, ne détermine la nature exacte de la méthode qui sera employée pour déterminer l'incertitude de mesure associée à une source d'erreur. La recommandation INC-1 est en fait très brève et les contenus des méthodes de « type A » et de « type B » ne sont pas explicités. De fait, les intentions des rédacteurs de la recommandation n'étaient pas de désigner la méthode adéquate pour chaque type d'évaluation. Plutôt, les deux catégories étaient d'abord des emplacements [*“placeholders”*]²⁸ pour lesquels différents choix étaient envisageables. C'est justement en cherchant à déterminer le contenu des méthodes de type A et de type B d'analyse d'incertitude que l'on est amené à entrer dans le détail des conceptualisations probabilistes de la mesure et des approches statistiques qu'elles sous-tendent.

Dans les prochains chapitres de cette partie, nous décrirons les fondements conceptuels de deux conceptions majeures de la mesure en métrologie, c'est-à-dire l'approche fréquentiste dite « traditionnelle », et une approche bayésienne qui a principalement été développée lors des deux dernières décennies (bien qu'elle s'appuie elle-même sur des avancées générales des statistiques qui remontent pour certains à plusieurs siècles). Dans le chapitre 3, nous nous attacherons en particulier à montrer pourquoi l'approche traditionnelle, qui a apporté une interprétation probabiliste de l'erreur de mesure longtemps considérée comme satisfaisante, a fini par apparaître lacunaire, ce qui a stimulé le développement d'approches complémentaires ou alternatives. Nous montrerons alors, dans le chapitre 4, comment l'interprétation épistémique des probabilités a été mobilisée pour pallier les lacunes attestées de l'approche traditionnelle, et comment cela a finalement abouti au développement d'une conception qui se veut entièrement bayésienne, de la mesure.

28. Notons d'ailleurs que l'intention initiale n'était pas de perpétuer les désignations « type A » et « type B » mais plutôt de trouver un accord temporaire en vue d'un développement ultérieur plus abouti. La terminologie a finalement été conservée jusque dans la première version du GUM. D'Agostini rapporte un témoignage selon lequel la typologie adoptée révèle la difficulté de l'ISO d'atteindre un consensus : « Les noms 'A' et 'B' ne présentent pas beaucoup de fantaisie, mais selon Klaus Weise [communication privée] ce fut là le seul accord auquel le comité de l'ISO put aboutir », D'Agostini (2003), p.182.

Chapitre 3

L'approche fréquentiste de la mesure

L'utilisation de méthodes statistiques est aujourd'hui omniprésente dans les problèmes liés à l'analyse des résultats de mesure, et ce dans toutes les disciplines scientifiques. L'origine de ces méthodes statistiques remonte aux XVIII^e et XIX^e siècles, pendant lesquels se sont succédées de nombreuses évolutions conceptuelles, ainsi que l'ont montré Stigler¹ ou encore Armatte². Stigler, en particulier, a soigneusement décrit le rôle qu'ont joué les probabilités dans le développement des théories de l'erreur. Il a montré que si certaines méthodes *ad hoc* – comme le moyennage, ou, plus tard, le principe des « moindres carrés » – ont été construites et exploitées avant même toute théorisation probabiliste, les probabilités se sont révélées nécessaires pour *justifier* ces méthodes sur la base de principes robustes. Ainsi, la moyenne arithmétique a d'abord connu un usage en astronomie et en géodésie que nous pourrions qualifier de « pré-statistique », fondé sur des principes heuristiques. Par la suite, cet usage s'est progressivement diffusé dans différentes disciplines des sciences de la nature, en particulier en physique, dans lesquelles elle s'est vue constamment perfectionnée. Stigler affirme que, dès lors, c'est la volonté de propager les outils et méthodes de l'analyse d'erreurs depuis l'astronomie vers les sciences sociales, ces dernières étant soucieuses d'asseoir leur scientificité sur les mêmes fondements mathématiques que les sciences dites « exactes », qui a mené notamment Galton, Edgeworth, Pearson et Yule à fonder les statistiques comme discipline au tournant du XX^e siècle³. Le développement des probabilités et des statistiques est donc intimement lié à leurs applications dans les différentes disciplines scientifiques dans lesquelles elles jouaient un rôle transversal, dévoué à l'analyse des données – la théorie des erreurs de mesure étant l'un des supports de ce rôle transversal.

L'approche fréquentiste que nous souhaitons expliciter dans ce chapitre est un produit de cet héritage historique complexe, au cours duquel il n'y a pas eu une interprétation unique et parfaitement homogène des probabilités. Ainsi, parler d'approche « fréquentiste » de la mesure semble surtout valoir pour désigner non pas la théorie des erreurs depuis sa genèse, mais plutôt la généralisation d'un certain nombre d'approches qui se sont imposées durant la première

1. Stigler (1986)

2. Armatte (2010), Armatte (2011), Armatte (2004)

3. Stigler (1986), pp.7–8.

moitié du XX^e siècle jusqu'au début des années 1970 environ. C'est pourquoi nous l'appelons, nous inspirant en cela de la littérature métrologique actuelle, d'approche « traditionnelle » de la mesure en métrologie⁴. L'un de nos objectifs est de montrer comment cette approche traditionnelle s'est retrouvée progressivement critiquée au profit de conceptions de la mesure fondées sur des interprétations épistémiques et subjectives des probabilités – jusqu'à ce que soient promues en métrologie des méthodes bayésiennes. Le terme d'approche « traditionnelle » ne doit pas donner l'impression – erronée – d'une théorie des erreurs exclusivement fréquentiste dès ses premiers développements, seulement remplacée au tournant des années 1970 par une conception radicalement nouvelle, fondée sur l'interprétation épistémique des probabilités. Les historiens ont montré que l'interprétation classique des probabilités, culminant lors de la première moitié du XIX^e siècle, mélange conceptions objectives et subjectives des probabilités⁵. Ainsi, la probabilité a présenté dès son émergence un versant objectif et un versant épistémique, comme l'a signalé Hacking⁶. De fait, on retrouve régulièrement des traces des deux facettes dans la littérature métrologique, même si l'on peut arguer que l'interprétation fréquentiste s'est imposée durant une grande partie du XX^e siècle dans les sciences de la nature, et en particulier en métrologie. Le tournant initié au début des années 1970 ne constitue pas pour autant une découverte des approches bayésiennes – elle en est plutôt un renouveau.

Nous avons identifié, au 1.4, deux problèmes relatifs à la détermination de l'incertitude de mesure sur une grandeur Y calculée indirectement à partir de la mesure de p grandeurs d'entrée X_i , auxquelles Y est reliée par une fonction de mesure $Y = f(X_1, \dots, X_p)$. Rappelons que les informations exploitées ne portent pas directement sur le mesurande Y , grandeur de sortie de la fonction de mesure $Y = f(X_1, \dots, X_p)$ mais sur les grandeurs d'entrée de cette fonction prises individuellement. Le premier problème consiste à déterminer l'incertitude de mesure relative à chaque grandeur d'entrée. Nous avons exposé au 2.4 que la façon dont cela est effectué dépend du type d'information à disposition. Nous expliciterons tour à tour la façon dont sont traités ces deux types d'information.

En reprenant l'analyse de la notion de répétabilité que nous avons proposée au 2.2, nous constatons que l'interprétation fréquentiste des probabilités apparaît particulièrement adaptée lorsqu'il s'agit de traiter des collections de données expérimentales issues de la mesure répétée d'une même grandeur dans des conditions considérées identiques – c'est-à-dire lorsqu'il s'agit de s'attaquer aux erreurs dites aléatoires. Cela correspond à ce que la recommandation INC-1 appelle les méthodes de « type A » d'évaluation de l'incertitude. Nous montrerons donc, dans une première partie de ce chapitre, comment l'approche fréquentiste traditionnelle effectue ce traitement probabiliste, en nous appuyant principalement sur le modèle proposé dans le GUM. Le modèle fréquentiste est déployé en trois hypothèses successives : les sections 3.1 à 3.4 y sont consacrées.

Cependant, le traitement proposé pour les erreurs aléatoires n'est pas adapté pour les ana-

4. Comme le rapportent par exemple les métrologues Willink et White, « (p)endant de nombreuses années, le seul type de statistiques qui était enseigné dans les lycées et les universités était fréquentiste. L'approche bayésienne a connu un renouveau ces dernières années, et les statistiques fréquentistes sont par conséquent désormais désignées comme des "statistiques classiques" », Willink et White (2012), p.3.

5. Hacking (1992); Daston (1995); Daston (1989) (en particulier pp.720-722).

6. Hacking (1975)

lyses d'incertitude relatives aux grandeurs dont la connaissance est obtenue par l'exploitation d'informations externes à l'expérimentation elle-même – c'est-à-dire pour ce que l'on regroupe sous le nom d'erreurs systématiques. Pour ce type d'erreurs, il faut envisager une seconde méthode. Nous montrerons donc, dans une seconde partie de ce chapitre, comment l'approche traditionnelle propose de traiter ce type de cas, et nous verrons qu'il s'agit d'une solution sensiblement différente de la solution statistique adoptée pour les erreurs aléatoires. Cela fait l'objet de la section 3.5.

En couvrant le traitement des erreurs aléatoires et des erreurs systématiques, les sections 3.1 à 3.5 répondent au premier problème recensé, à savoir déterminer l'incertitude de mesure relative à chaque grandeur d'entrée de la fonction de mesure. Un second problème consiste à « propager » les composantes d'incertitudes ainsi déterminées pour obtenir l'incertitude résultante relative au mesurande Y , déterminé à partir des grandeurs d'entrée et qui constitue l'objectif final de la mesure. C'est ainsi que, dans un troisième temps, nous porterons notre attention sur le traitement de l'incertitude de mesure finale et de la façon dont ce traitement fait usage des déterminations explicitées dans les sections précédentes. Dans la section 3.6, nous présentons ainsi la façon dont un résultat de mesure est exprimé dans sa forme finale dans une approche fréquentiste stricte. En particulier, nous constatons que cela mène à écrire l'incertitude de mesure selon deux composantes séparées de natures bien distinctes. Nous verrons alors en quoi cela a été interprété comme une lacune sérieuse, et montrerons que l'exposition de cette lacune est l'un des principaux moteurs sous-jacents au développement de conceptions nouvelles, parmi lesquelles l'approche bayésienne que nous développerons à son tour dans le chapitre suivant.

Dans ce chapitre, ainsi que les deux chapitres suivants, notre point de repère principal sera le *Guide pour l'expression de l'incertitude de mesure*. Cependant, celui-ci ne suffit pas à comprendre les questionnements épistémologiques et techniques qui caractérisent la tension entre les approches fréquentiste et bayésienne de la mesure. En particulier, bien que le VIM et le GUM fassent fréquemment référence à une approche traditionnelle de la mesure, il est difficile de cerner de quoi il retourne précisément à la seule lecture de ces documents. En effet, ceux-ci font la synthèse des idées nouvelles qui sont justement venues réformer l'approche traditionnelle en profondeur. Ils portent sur l'approche traditionnelle une vision rétrospective qui entretient parfois le flou quant au contenu effectif de cette approche⁷. Pour véritablement cerner l'approche traditionnelle de la mesure, il est donc également nécessaire de s'en remettre à la littérature de la période qui précède le travail sur le GUM. Cette dernière contient de nombreuses expositions des idées constitutives de l'approche traditionnelle ; cependant, elle n'est pas présentée d'une façon bien homogène qui laisserait entrevoir une unique approche bien définie et utilisée de la même façon par tous. Or, notre description de l'approche fréquentiste ne se veut pas une reconstitution exacte de la pluralité de conceptions du XX^e siècle qui entrent dans ce que l'on peut appeler une approche traditionnelle de la mesure. L'objectif est de regrouper et expliciter les idées charnières communes qui sont constitutives des approches fréquentistes en général, sans pour autant être rigoureusement fidèles à ce que ces approches étaient à l'époque des pu-

7. En ce sens, nous sommes plutôt d'accord avec la position de Willink et White selon laquelle « il y a un degré auquel l'histoire semble être réécrite », [Willink et White \(2012\)](#), p.2.

blications étudiées. Nous choisissons ici de ne pas nous restreindre quant au vocabulaire et au formalisme employé – le GUM demeure notre point de repère principal. Cela implique parfois de faire dialoguer le vocabulaire, les concepts, les idées et les arguments tirés de la compréhension contemporaine de l'approche fréquentiste avec ceux tirés d'approches plus anciennes.

Par la suite, nous considérerons le plus souvent un cas simplifié ne présentant que deux grandeurs d'entrée X et Z , où X est l'objet d'un mesurage au sein de l'expérience considérée, et où Z est connue au moyen d'informations extérieures. Nous considérons de plus que la fonction de mesure se réduit à une simple addition⁸. Dans ce cas, on considère la fonction de mesure simplifiée :

$$Y = X + Z \quad (3.1)$$

En particulier, cette notation simplifiée fait l'économie d'indices pour désigner les différentes grandeurs d'entrée, ce qui permet d'éviter toute ambiguïté avec l'information expérimentale disponible à propos de X . Celle-ci se présente sous la forme d'un *échantillon de données expérimentales* $\{x_1, \dots, x_n\}$, que nous désignerons par la suite par $\{x_i\}$. Nous utiliserons cette notation dans l'ensemble du présent chapitre et dans le chapitre suivant.

3.1 Fondement fréquentiste du traitement statistique des « erreurs aléatoires »

Le traitement fréquentiste des erreurs de mesure est fondé sur l'idée qu'à défaut de pouvoir *corriger* ces erreurs, il demeure possible d'en effectuer une *réduction statistique* en répétant la mesure pour collecter un maximum de données expérimentales. Par exemple, l'on considérera généralement que la moyenne d'un ensemble de résultats de mesure est affectée d'une erreur bien moindre que chacune des données individuelles – une propriété classique que nous nous attacherons à démontrer par la suite dans ce chapitre. C'est dans le but de justifier et d'expliquer les opérations menées dans le traitement statistique des données expérimentales, comme l'opération de moyennage, qu'est introduit un *modèle probabiliste*. Dans ce chapitre, nous nous attachons à décrire de façon structurée un modèle probabiliste qui s'appuie sur une interprétation fréquentiste des probabilités. Ce modèle n'est pas la reproduction d'une méthode bien précise utilisée par les métrologues ; elle est plutôt une reconstruction qui vise à mettre en évidence les différentes hypothèses sous-jacentes qui ne sont pas toujours rendues explicites dans les descriptions classiques que l'on retrouve dans la littérature. Nous souhaitons insister sur les conditions qui rendent possible un traitement fréquentiste des données expérimentales, et sur la signification qu'ont alors les résultats obtenus. Ce faisant, les quatre premières sections de ce chapitre s'attachent à mettre en évidence les hypothèses constitutives du modèle fréquentiste, et à montrer comment celui-ci permet la détermination d'une incertitude de mesure associée à la variabilité des résultats expérimentaux. En particulier, dans cette section, nous souhaitons

8. Bien entendu, dans ce cas, cela simplifie nettement un certain nombre de problèmes techniques classiques. Entrer dans le détail mathématique des calculs relatifs à différentes fonctions de mesure plus ou moins complexes n'aurait pas ici de grand intérêt, et l'exemple simplifié considéré ici suffit à faire émerger les problèmes épistémologiques principaux que nous souhaitons mettre en évidence.

expliciter le cadre bien spécifique dans lequel il est possible d'introduire des probabilités de type fréquentiste. Cela nous mènera à restreindre le traitement statistique à celui des seules erreurs dites « aléatoires » de mesure, et à insister sur un caractère fondamental du modèle fréquentiste de la mesure, à savoir qu'ils prend pour objet non pas la cible de la mesure – la « valeur vraie » de la grandeur ciblée – mais le *processus de mesure*.

Puisque nous souhaitons développer un modèle fréquentiste des erreurs de mesure, il nous faut au préalable rappeler ce qu'est l'interprétation dite « fréquentiste » des probabilités. Les différentes interprétations des probabilités convergent le plus souvent quant à l'idée générale que « les événements les plus probables se produisent plus fréquemment que les événements les moins probables »⁹. De ce fait, « les fréquences relatives entretiennent une relation intime avec les probabilités »¹⁰ – la fréquence relative d'un événement A étant le rapport du nombre d'occurrences de l'événement A sur le nombre d'essais susceptibles de faire survenir l'événement A. Cette idée prend forme en particulier au travers de la « loi des grands nombres »¹¹, qui établit que pour des événements aléatoires, les fréquences relatives ont tendance à se stabiliser et à tendre vers les probabilités correspondantes lorsque le nombre d'essais augmente. L'interprétation fréquentiste des probabilités va cependant plus loin que de simplement souligner qu'il existe une relation entre probabilité et fréquence ; de fait, elle postule que les probabilités s'identifient¹² aux fréquences relatives. Ainsi, lors du lancer « à pile ou face » d'une pièce de monnaie, la probabilité que la pièce retombe du côté « face » visible est de 1/2 si cette issue survient en moyenne une fois sur deux – la fréquence relative n'est pas une conséquence de la probabilité, mais sa mesure même. Dans le détail, l'identité entre probabilité et fréquence relative peut prendre plusieurs formes, et il existe non pas une unique interprétation fréquentiste mais une classe d'interprétations qui présentent ce même postulat général en commun (par exemple, des différences affleurent selon que l'on identifie probabilité et fréquence *limite* ou probabilité et fréquence effective sur un nombre fini de tirages). Cependant, dans tous les cas, l'interprétation présente une caractéristique remarquable, à savoir que dès lors que les probabilités sont liées aux fréquences d'occurrence des événements, elles sont indépendantes du sujet¹³ qui effectue les essais et évalue les probabilités ; c'est la raison pour laquelle ces probabilités sont considérées comme « objectives »¹⁴. Si l'on estime par exemple qu'une pièce de

9. Hájek (2012), section 2. Entrée consultée le 22 décembre 2015.

10. Hájek (2012), section 3.4.

11. Voir Stigler (1986), pp.65-68 pour une présentation historique, et Mayo (1996), pp.165-170, pour une défense des probabilités fréquentiste articulée autour de la loi des grands nombres. La loi des des grands nombres ne se limite pas à la seule propriété énoncée ici, mais c'est cette dernière qui est utile dans le cadre présent.

12. Hájek (2012), section 3.4.

13. Notons que la notion de « sujet » peut être comprise dans un sens large ; Bienvenu parle d'« évaluateur » qu'il caractérise ainsi : « (n)ous entendons par 'évaluateur' l'instance formant les jugements ou énoncés de probabilité. Selon les théories, il peut être défini de façon différente : il peut s'agir d'un sujet abstrait à valeur universelle (Reichenbach), d'un individu empirique (De Finetti, Ramsey), d'un système de règles (Carnap), d'une méthodologie, d'un automate, etc. », Bienvenu (2007), p.12.

14. Il n'existe pas de consensus sur la juste terminologie à employer. Plusieurs interprétation sensiblement différentes des probabilités peuvent prétendre au qualificatif d'« objectives » ; c'est le cas par exemple du bayésianisme objectif défendu par Jaynes (Jaynes et Bretthorst, 2003) et d'autres, ou encore du propensionnisme initié par Popper. Les probabilités objectives ne se restreignent donc pas au fréquentisme ; comme le précise Alexis Bienvenu, « il existe en effet des objectivistes non fréquentistes, des subjectivistes non bayésiens, ainsi que des bayésiens non sub-

monnaie a autant de chance de tomber du côté « pile » que du côté « face », et donc que la probabilité de chacune de ces issues est $1/2$, c'est parce qu'on pense que *l'état matériel* de la pièce (et du dispositif de lancer) rend ces issues équiprobables. Les probabilités fréquentistes présentent donc deux caractéristiques essentielles : elles sont mesurées par des fréquences, mais surtout, elles prennent pour objet *des événements du monde physique*.

De par leur nature, les probabilités fréquentistes ne peuvent porter que sur les issues possibles d'une expérience aléatoire que l'on est amené à répéter. L'objet des probabilités ne peut pas être la valeur visée – la « valeur vraie » de la grandeur – qui est une cible fixe, et qui n'est pas un événement du monde physique mais une proposition sur le monde physique. Dans un cadre fréquentiste, il n'y a donc pas de sens à parler de la probabilité qu'aurait la valeur vraie d'une grandeur à être égale à telle ou telle valeur. C'est en revanche l'issue des mesurages répétés de la grandeur qui se prête à une description probabiliste en termes de fréquences. Comme nous l'avons déjà décrit au chapitre précédent, de tels mesurages répétés peuvent présenter une variabilité en raison des contingences inhérentes à l'opération concrète de mesure et aux fluctuations des conditions d'expérience. Nous convergions ici vers une idée sur laquelle nous souhaitons tout particulièrement insister : dans les modèles fréquentistes de la mesure, l'objet des probabilités n'est pas la valeur visée ou le résultat final, mais le *processus de mesure*.

L'idée que les erreurs de mesure peuvent être comprises comme les éléments de sortie d'une chaîne de production a été formalisée par Walter Shewhart (1891-1967), physicien, statisticien et ingénieur américain, réputé pour son apport dans la théorisation de la gestion de la qualité des produits industriels et de leurs processus de production¹⁵. En 1922, dans un article consacré à l'analyse de la « mesure d'une quantité physique dont la grandeur est influencée par des causes primaires au-delà du contrôle de l'observateur »¹⁶, Shewhart compare la mesure d'une grandeur à un processus industriel de production où les produits seraient les résultats de mesure – ou, alternativement, les erreurs de mesure. C'est ce processus de mesure que le modèle fréquentiste propose de considérer comme un mécanisme aléatoire. La probabilité d'une issue possible du processus de mesure, c'est-à-dire la probabilité d'obtenir une valeur mesurée donnée, est alors identifiée à la fréquence relative avec laquelle cette valeur est produite par le processus. L'une des formes les plus classiques de l'interprétation fréquentiste, qui est la forme la plus courante en métrologie, conçoit la probabilité d'un événement comme la fréquence relative *limite* d'occurrence de cet événement lorsqu'une situation est répétée *à l'identique une infinité de fois*¹⁷. Cependant, pour qu'une telle probabilité soit définie, il faut au préalable s'as-

jectivistes », [Bienvenu \(2007\)](#), p.10. En revanche, il n'existe pas à notre connaissance de fréquentisme subjectiviste. Voir encore [Gillies \(2000\)](#), pp.18-22 pour une discussion de ce débat.

15. [Shewhart \(1939\)](#)

16. [Shewhart \(1922\)](#), p.248. Notons que le titre de l'article de Shewhart rappelle une fois encore que la motivation de la description statistique des données expérimentales provient du fait que l'on ne peut pas avoir une compréhension physique complète de la variabilité de ces données.

17. Il existe différentes façons de concevoir et de définir la probabilité selon une acception fréquentiste. En particulier, toutes les conceptions fréquentistes ne s'accordent pas sur le fait que les probabilités sont des fréquences *limite* définies sur des échantillons *infinis* de données – une telle définition pose des problèmes, en particulier sur le caractère accessible ou non d'une telle probabilité. Ainsi, le fréquentisme fini, proposé par Venn, évite cet écueil en ne s'appuyant que sur des échantillons réels (voir [Hájek, 2012](#), section 3.4) – au prix d'autres inconvénients. Les multiples conceptions fréquentistes des probabilités peuvent différer significativement de l'une à l'autre, mais

surer qu'une telle fréquence limite *existe*, ce que rien ne garantit *a priori*. Ainsi, la définition de la probabilité comme fréquence limite ne peut-elle être acceptable qu'à la condition que les fréquences limites d'occurrence de l'événement décrit *convergent* vers une valeur stable lorsque le nombre de tirages tend vers l'infini. En particulier, le simple fait qu'un processus soit (supposé) aléatoire ne suffit pas à rendre possible la quantification par des probabilités fréquentistes. Eisenhart, à la suite de Shewhart, a considéré qu'en métrologie, étudier les erreurs de mesure à l'aune de l'hypothèse d'un tirage aléatoire des résultats de mesure est ainsi suspendu à la bonne *stabilité* du processus de mesure. Il parle ainsi d'un « état de contrôle statistique », qu'il place comme pré-requis à la possibilité d'une modélisation statistique des erreurs de mesure :

(U)ne cohérence ou une stabilité statistique d'un type très spécial est requise : pour être considérée comme un processus de mesure, une opération de mesure doit avoir atteint ce qui est connu dans le langage du contrôle de qualité industrielle comme un état de contrôle statistique. Jusqu'à ce qu'une opération ait été "déboguée" dans la mesure ou elle a atteint un état de contrôle statistique, elle ne peut pas en toute logique être considérée comme mesurant quoi que ce soit.¹⁸

L'obtention et le contrôle d'un tel état de « contrôle statistique » renvoie à un ensemble de procédures que Shewhart, en particulier, a développé en détail et qu'Eisenhart expose brièvement dans la suite de son article. Cependant, ce n'est pas la présentation exhaustive des procédures elles-mêmes qui attirent notre attention, mais l'idée générale que Shewhart et Eisenhart ont soulignée, et qui renvoie, en termes contemporains, au fait que l'analyse statistique des données expérimentales est d'abord pensée dans le cadre de conditions dites « de répétabilité ». Dans le chapitre 2, nous avons distingué deux types d'erreurs de mesure, les erreurs dites « aléatoires » et les erreurs dites « systématiques », et nous avons montré que la classification des erreurs de mesure dans l'une ou l'autre des catégories dépend des conditions dans lesquelles les mesures sont supposées avoir été effectuées. En particulier, les erreurs qui varient dans les conditions de répétabilité sont (par définition) les erreurs aléatoires. En effet, l'erreur dite « aléatoire » est un modèle d'erreur de mesure associé à une composante de variabilité que l'on considère ne pas savoir expliquer ou prédire – ou que l'on ne souhaite pas s'embarrasser à traiter théoriquement. En postulant l'existence d'erreurs de type « aléatoire », on remplace en fin de compte une description physique des phénomènes à l'origine des erreurs de mesure par

ce n'est pas l'objet de cette section que de rentrer dans le détail de la construction des conceptions fréquentistes de la probabilité. Nous adoptons donc une définition classique en termes de fréquences limites, parce qu'elle est celle à laquelle ont majoritairement tendance à se rallier les métrologues, comme en témoigne la littérature sur le sujet depuis un siècle. Citons par exemple cette remarque de Cohen et DuMond : « (l)'autre façon de faire, bien plus utile, est bien sûr de considérer le signe plus-ou-moins [dans un résultat de mesure] comme une estimation d'un paramètre de largeur d'une certaine distribution statistique de valeurs observées *qui seraient obtenues si la mesure était répliquée un grand nombre de fois* », Cohen et DuMond (1965), p.540 (nous soulignons). La présence du « si » témoigne du caractère contrefactuel de l'interprétation des auteurs, caractère qui est typique d'une interprétation de la probabilité comme fréquence relative *limite*.

18. Eisenhart (1963), p.162. Churchill Eisenhart a travaillé jusqu'en 1983 comme statisticien au National Institute of Standards and Technology (NIST, le laboratoire national de métrologie des États-Unis). Il a produit une synthèse minutieuse de l'approche fréquentiste dans un article publié en 1963, et peut, dans une certaine mesure, être considéré comme l'un de ceux qui a formulé de la façon la plus claire une approche fréquentiste que l'on pourrait qualifier d'orthodoxe.

une description statistique des erreurs de mesure elles-mêmes¹⁹. Le traitement fréquentiste des erreurs de mesure est donc d'abord un traitement statistique des erreurs *aléatoires* de mesure. Cela ne signifie pas qu'il soit rigoureusement impossible d'employer des méthodes statistiques pour étudier les erreurs systématiques, mais nous verrons par la suite que cela a effectivement posé des problèmes aux métrologues.

En reprenant le modèle de mesure simplifié représenté par l'équation (3.1) : $Y = X + Z$, proposé dans le préambule de ce chapitre, nous considérerons par la suite la mesure de la grandeur d'entrée X dans des conditions de répétabilité. La situation peut être résumée comme suit : l'expérimentateur dispose, à propos de la grandeur X étudiée, d'un échantillon de n données expérimentales $\{x_1, \dots, x_n\}$. Selon la définition d'une erreur de mesure, on peut écrire pour chaque donnée expérimentale x_i que²⁰ :

$$x_i = \chi + \epsilon_i \quad \begin{cases} x_i = \text{donnée expérimentale (valeur obtenue en pratique)} \\ \chi = \text{valeur vraie de la grandeur } X \\ \epsilon_i = \text{erreur de mesure} \end{cases} \quad (3.2)$$

Nous pouvons dès lors formaliser le modèle fréquentiste des erreurs aléatoires en exprimant une première hypothèse qui en est constitutive, selon laquelle le résultat effectif x_i d'une mesure individuelle est le produit du tirage aléatoire d'une valeur parmi une collection de résultats possibles. Cette collection, parfois appelée « population statistique », est déterminée par l'état de la grandeur mesurée et par les conditions de répétabilité dans lesquelles l'expérience est menée. Nous retrouvons chez Student, dès 1908, une formulation de ce type :

Toute expérience peut être considérée comme formant un individu parmi une “population” d'expériences qui pourraient être effectuées dans les mêmes conditions.

Une série d'expériences est un échantillon tiré de cette population.²¹

Les statisticiens et les métrologues reprendront et enrichiront cette idée, en introduisant la notion de « variable aléatoire » sur lequel la suite de notre exposition s'appuiera extensivement ; cela apparaît par exemple chez Eisenhart :

19. Campbell a eu à ce sujet une remarque très pertinente, que nous sortons cependant quelque peu de son contexte (Campbell désirant avant tout discuter du caractère objectif de la probabilité) : « (d)ans toutes les discussions sur les chances et la probabilité, un paradoxe caractéristique apparaît toujours, à savoir que nous semblons déduire une connaissance du fait même que nous n'avons pas de connaissance », Campbell (1920), p.161. C'est effectivement latent dans l'analyse statistique des erreurs aléatoires, celle-ci revenant à tirer une connaissance d'une collection de données, en partant de l'idée que l'on n'est pas capable d'expliquer physiquement le comportement de ces données. Bien entendu, la connaissance que l'on tire de cette analyse statistique n'émerge pas de rien, mais elle est conditionnée à différentes hypothèses, que nous exposerons dans la suite de ce chapitre. En particulier, elle découle de l'idée commune selon laquelle les erreurs aléatoires « s'annulent en moyenne ». Autrement dit, notre connaissance des erreurs aléatoires est tributaire de la connaissance collectée en amont du modèle ; c'est cette connaissance que l'on importe en tant qu'hypothèses fondatrices du modèle fréquentiste.

20. Nous choisissons de désigner par χ la valeur vraie de la grandeur X . De manière générale, nous nous efforcerons de nous en tenir aux notations suivantes : la lettre grecque minuscule désigne la valeur vraie de la grandeur correspondante. Le mesurande final Y obtenu via la fonction de mesure fera exception : nous utiliserons le symbole v .

21. Student (1908), p.1.

[...] les n premières mesures d'une quantité donnée, générées par un processus particulier de mesure, fournissent une base logique pour prédire le comportement d'autres mesures de la même quantité, par le même processus de mesure, si et seulement si ces n mesures peuvent être considérées comme un *tirage aléatoire* d'une "population" ou d'un "univers" de toutes les mesures concevables de la quantité étudiée par le processus de mesure concerné : c'est-à-dire, dans le langage des statistiques mathématiques, si et seulement si les n mesures en main peuvent être considérées comme des "valeurs observées" d'une séquence de variables aléatoires caractérisées par une distribution de probabilité identifiée avec le processus de mesure concerné, et reliée par les valeurs d'un ou plus de ses paramètres à l'amplitude de la quantité mesurée.²²

Une variable aléatoire est une variable qui peut prendre un ensemble de valeurs différentes, chacune avec une probabilité spécifique. La variable aléatoire est caractérisée par sa « distribution de probabilité » (ou loi de probabilité), une fonction mathématique qui associe à chaque valeur possible de la variable aléatoire sa probabilité correspondante²³. Le fondement du modèle fréquentiste peut dès lors être résumé à l'aide de cette notion, sous la forme d'une hypothèse (H1) énoncée comme suit :

(H1) Le processus de mesure de la grandeur X peut être représenté mathématiquement par une variable aléatoire \hat{x} , dont la distribution de probabilité associée aux valeurs possibles du résultat d'un mesurage individuel, représentées par x , les fréquences relatives limites $f_{\hat{x}}(x)$ d'occurrence de ces valeurs.

L'équation (3.2) rappelle par ailleurs que le résultat x_i d'un mesurage et l'erreur ϵ_i affectant ce résultat sont reliés à une constante près – la valeur vraie de la grandeur X mesurée. Par conséquent, l'on peut immédiatement définir une variable aléatoire $\hat{\epsilon}$ à partir de \hat{x} , par la relation :

$$\hat{x} = \hat{\epsilon} + \chi \quad (3.3)$$

Dès lors, (H1) est encore équivalente à l'hypothèse (H1') suivante :

(H1') Le processus de production d'une erreur ϵ_i lors de la mesure de X peut être représenté mathématiquement par une variable aléatoire $\hat{\epsilon}$ dont la distribution de probabilité prend pour valeurs les fréquences relatives limites d'occurrence $f_{\hat{\epsilon}}$ des erreurs ϵ_i .

Selon l'hypothèse (H1), les résultats de mesure x_i sont tous caractérisés par la même loi de probabilité; on parle pour $f_{\hat{x}}$ et $f_{\hat{\epsilon}}$ des distributions probabilistes « parentes »²⁴. De plus, les résultats de mesurages différents sont *indépendants* entre eux; bien sûr, ils dépendent de la

22. Eisenhart (1963), p.166 (Eisenhart souligne).

23. Lorsque la variable aléatoire est continue à valeurs réelles, les valeurs prises par la distribution de probabilité ne sont pas des probabilités, mais des densités de probabilité. En effet, les probabilités correspondantes sont infinitésimales, puisque le domaine de la fonction couvre une infinité de valeurs possibles. Notons que les densités de probabilité ne sont pas des nombres purs mais des grandeurs dimensionnées. Nous n'accorderons pas à cet aspect d'importance particulière, bien que nous nous efforcerons de maintenir la distinction à chaque fois que celle-ci sera nécessaire.

24. Deming et Birge (1934), p.123; Willink (2013), p.44.

même distribution parente, mais le résultat d'un mesurage donné n'est en rien déterminé par celui du mesurage précédent.

Dès lors que \hat{x} et $\hat{\epsilon}$ sont définies, on peut les caractériser par certains de leurs paramètres statistiques, dont les plus remarquables sont leur espérance et leur variance, qui représentent respectivement la moyenne et la dispersion du processus de mesure. Nous en donnons la définition au tableau 3.1. À quoi mène l'expression de l'hypothèse (H1)? Comme l'exprime Eisenhart dans la citation précédente, modéliser le processus de mesure par un mécanisme aléatoire ne présente un intérêt que s'il est possible, par ailleurs, de relier les paramètres caractéristiques de la variable aléatoire \hat{x} à l'état de la grandeur mesurée elle-même. À la figure 3.1, nous représentons un exemple de distribution de probabilité pour la variable aléatoire \hat{x} , ainsi que son espérance, sa variance, et qu'un exemple (arbitraire) d'histogramme de valeurs numériques que l'on peut obtenir par tirage aléatoire.

Il nous faut donc en fin de compte distinguer trois niveaux : le niveau de la grandeur X mesurée, représenté par sa valeur vraie (inconnue); le niveau du processus de mesure, représenté par la variable aléatoire \hat{x} de l'hypothèse (H1), et le niveau des données expérimentales $\{x_i\}$. L'objectif d'un modèle probabiliste de la mesure est de remonter des données expérimentales jusqu'aux paramètres principaux (inconnus) de \hat{x} , puis de remonter de ces paramètres jusqu'à la valeur vraie χ de la grandeur X visée. Pour cela, il est nécessaire d'ajouter plusieurs hypothèses supplémentaires au modèle, ce que nous explorons dans les deux prochaines sections.

3.2 Seconde hypothèse du modèle fréquentiste : lien entre espérance statistique et « valeur vraie » de la grandeur

Le modèle fréquentiste considère les erreurs de mesure comme le produit d'un mécanisme aléatoire. Il n'exclut pas pour autant l'idée que le processus de mesure est déterminé par une conjonction de phénomènes physiques partiellement ou entièrement déterministes²⁵. L'élément essentiel est que les résultats obtenus dépendent de l'état de la grandeur visée, sans quoi le fait même de mesurer cette grandeur ne mènerait à rien. C'est parce que les résultats obtenus dépendent de l'état de la grandeur visée que le modèle fréquentiste peut servir à effectuer une réduction des erreurs aléatoires de mesure par répétition des mesurages puis application d'une méthode purement statistique.

L'hypothèse (H1), cependant, ne permet pas à elle seule de répondre au problème posé, à savoir : comment remonter de l'échantillon $\{x_i\}$ à la valeur vraie χ de X ? Il est nécessaire pour cela d'introduire deux hypothèses supplémentaires; l'une pour faire le lien entre $\{x_i\}$

25. La question du déterminisme n'a pas véritablement d'importance ici : que le phénomène à l'origine des erreurs soit déterministe ou stochastique, le même outil probabiliste sera mis à contribution pour modéliser les erreurs aléatoires. Lorsque la théorie physique mise à l'essai dans une expérience présente une part de hasard, comme en mécanique quantique, ce hasard est la cible de l'expérimentation et l'on cherche à mesurer les probabilités correspondantes. Dans l'analyse statistique des erreurs aléatoires, l'objectif est d'utiliser les probabilités comme outil pour réduire le bruit occasionné par ces erreurs. Ce sont donc deux activités différentes. Par la suite, nous ignorerons volontairement l'étude spécifique des phénomènes considérés aujourd'hui comme intrinsèquement aléatoires, comme les désintégrations radioactives ou les mesures en physique quantique, puisque cela sort de notre cadre d'étude.

| | \hat{x} | $\hat{\epsilon}$ | Relations |
|------------|--|--|---|
| Espérance | $\mu_{\hat{x}} = \int_{-\infty}^{+\infty} x f_{\hat{x}}(x) dx$ | $\mu_{\hat{\epsilon}} = \int_{-\infty}^{+\infty} \epsilon f_{\hat{\epsilon}}(\epsilon) d\epsilon$ | $\mu_{\hat{x}} = \mu_{\hat{\epsilon}} + \chi$ |
| Variance | $V_{\hat{x}} = \int_{-\infty}^{+\infty} (x - \mu_{\hat{x}})^2 f_{\hat{x}}(x) dx$ | $V_{\hat{\epsilon}} = \int_{-\infty}^{+\infty} (\epsilon - \mu_{\hat{\epsilon}})^2 f_{\hat{\epsilon}}(\epsilon) d\epsilon$ | $V_{\hat{x}} = V_{\hat{\epsilon}}$ |
| Écart-type | $\sigma_{\hat{x}} = \sqrt{V_{\hat{x}}}$ | $\sigma_{\hat{\epsilon}} = \sqrt{V_{\hat{\epsilon}}}$ | $\sigma_{\hat{x}} = \sigma_{\hat{\epsilon}}$ |

TABLEAU 3.1 – Trois paramètres statistiques remarquables d’une variable aléatoire : espérance, variance et écart-type. L’espérance est définie comme la moyenne des valeurs obtenues sur un nombre infini de tirages aléatoires de la variable. L’écart-type d’une variable aléatoire est la dispersion quadratique moyenne de sa distribution de probabilité, et est défini comme la racine carré de la variance. L’équation (3.3) reliant les variables aléatoires \hat{x} et $\hat{\epsilon}$ à une constante près permet facilement de retrouver les relations simples entre les paramètres statistiques des deux variables aléatoires.

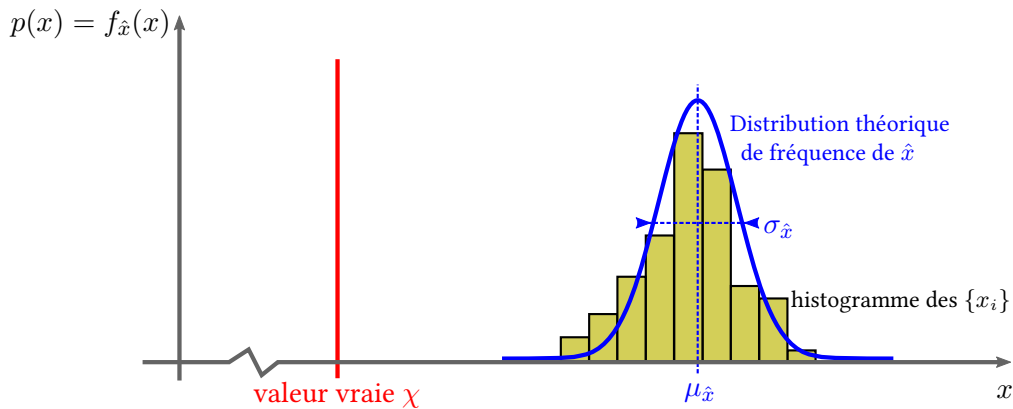


FIGURE 3.1 – Illustration d’une distribution de probabilité de la variable aléatoire \hat{x} , et d’un histogramme de données expérimentales correspondantes. Les formes de la distribution et de l’histogramme ont été choisies arbitrairement, pour l’exemple; la distribution est ici gaussienne. Cette figure, comme les suivantes du même type, est inspirée des représentations graphiques proposées dans Ehrlich, Dybkaer et Wöger (2007).

et les paramètres principaux de \hat{x} , et l'autre pour faire le lien entre ces derniers et χ . Dans cette section, nous nous intéressons à la relation entre \hat{x} et χ . Celle-ci fait l'objet d'une seconde hypothèse, que nous exposerons après être revenu sur la dichotomie entre erreurs aléatoires et systématiques.

3.2.1 Erreurs aléatoires et erreurs systématique : retour sur la dichotomie

Nous avons montré au chapitre 2 que la dichotomie entre erreurs aléatoires et systématiques provient d'une conjonction de deux critères. Le premier critère est épistémique : une erreur systématique est une erreur que l'expérimentateur considère pouvoir connaître individuellement ; une erreur aléatoire est une erreur qu'il ne peut ou ne veut pas chercher à prédire par un modèle physique et qu'il se contentera de chercher à réduire statistiquement. Le second critère est temporel : une erreur systématique est une erreur qui demeure constante lorsque la mesure est répétée ; une erreur aléatoire présente une variabilité. Nous avons défendu l'idée que c'est la stipulation des conditions de répétabilité qui permet d'articuler ensemble ces deux critères. Hors des conditions de répétabilité, les erreurs systématiques peuvent tout à fait être variables. Ces deux critères sont indépendants de tout modèle probabiliste ; de fait, ils donnent un cadre aux modèles probabilistes que l'on souhaiterait *ensuite* appliquer. Cependant, une fois explicitées les conditions de répétabilité et formulée l'hypothèse (H1), nous pouvons reconsidérer la classification en erreurs aléatoires et systématiques. Puisqu'une erreur systématique sur X est considérée comme constante dans les conditions de répétabilité, elle contribue uniquement à l'espérance de la variable aléatoire \hat{x} . Il est donc logique de la représenter mathématiquement de la façon suivante :

$$\epsilon_{\text{sys}} = \mu_{\hat{x}} - \chi \quad (3.4)$$

Dans ce cas, le terme ϵ_{sys} représente l'erreur systématique *totale* qui affecte le processus de mesure. Cette erreur ne caractérise pas un résultat de mesurage en particulier, mais l'ensemble des données expérimentales $\{x_i\}$.

La distinction entre erreurs aléatoires et systématiques constitue une décomposition de l'erreur totale, de telle sorte que la somme de l'erreur systématique et de l'erreur aléatoire est égale à l'erreur totale ; on peut donc écrire l'erreur aléatoire de la façon suivante :

$$\epsilon_{\text{al},i} = \epsilon_i - \epsilon_{\text{sys}} = x_i - \mu_{\hat{x}} \quad (3.5)$$

Il est également possible de réécrire ces définitions à partir de la variable aléatoire $\hat{\epsilon}$ définie par l'équation 3.3. Par propriété de l'espérance d'une variable aléatoire :

$$\mu_{\hat{x}} = \mu_{\hat{\epsilon}} + \chi \quad (3.6)$$

Par conséquent, on peut écrire les relations suivantes (représentées graphiquement à la figure 3.2) :

$$\begin{cases} \epsilon_{\text{sys}} = \mu_{\hat{x}} - \chi = \mu_{\hat{\epsilon}} \\ \epsilon_{\text{al},i} = \epsilon_i - \epsilon_{\text{sys}} = x_i - \mu_{\hat{x}} \end{cases} \quad (3.7)$$

C'est bien ainsi que les deux termes sont définis mathématiquement dans le VIM²⁶.

26. [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), pp.22–23. Nous renvoyons également à la discussion

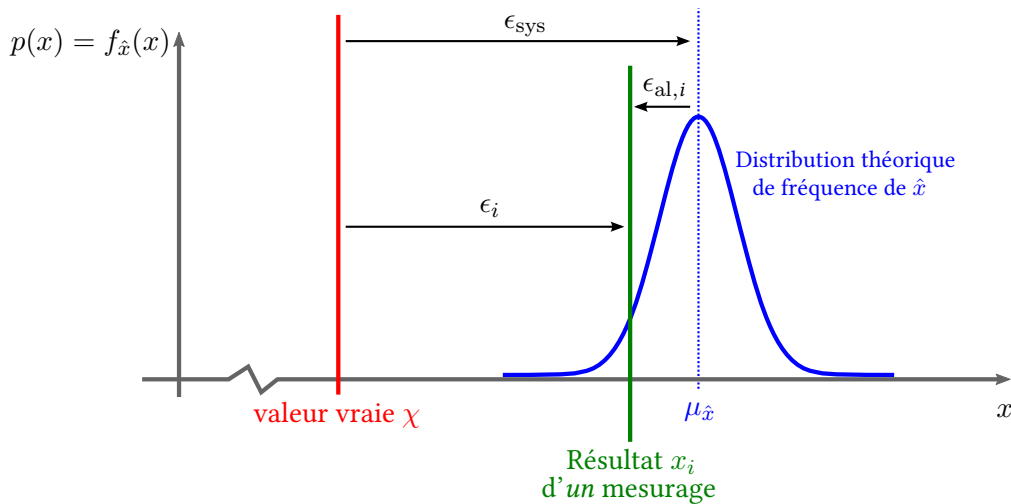


FIGURE 3.2 – Illustration de l’erreur ϵ_i affectant une donnée expérimentale x_i , de l’erreur aléatoire correspondante $\epsilon_{\text{al},i}$ et de l’erreur systématique ϵ_{sys} affectant l’ensemble du processus de mesure. L’erreur systématique est la différence entre la valeur vraie χ de la grandeur et l’espérance $\mu_{\hat{x}}$ de la distribution théorique (inconnue) de fréquence des données expérimentales. L’erreur aléatoire est la différence entre le résultat x_i d’un mesurage spécifique et $\mu_{\hat{x}}$. La somme de l’erreur aléatoire et de l’erreur systématique est égale à l’erreur finale affectant x_i . La distribution théorique de fréquence des données expérimentales est ici arbitrairement assimilée à une loi normale, et ce uniquement pour l’exemple.

3.2.2 Seconde hypothèse du modèle fréquentiste

La composante systématique de l’erreur de mesure est une agrégation des différents biais constants qui affectent le résultat de mesure. Certains de ces biais sont connus et sont donc corrigés ; cependant, après correction, une erreur est susceptible de *subsister* car la correction d’un biais n’est jamais parfaite. L’erreur qui subsiste *n’est pas* le biais lui-même, puisque celui-ci a été corrigé ; elle est une composante *inconnue* d’erreur. Certains autres biais échappent à la connaissance de l’expérimentateur ; dans ce cas, l’amplitude de l’erreur systématique correspondante

que nous avons menée dans le chapitre précédent, et au tableau 2.1 p.56. Notons que dans la littérature spécialisée, on trouve parfois une autre façon de définir l’erreur aléatoire et l’erreur systématique. Plutôt que de définir l’erreur systématique comme l’écart de l’espérance à la valeur vraie comme dans l’équation (3.7), l’erreur systématique est définie comme l’écart à la valeur moyenne de l’échantillon : $\epsilon_{\text{sys}} = \bar{x} - \chi$. Dans ce cas, l’erreur aléatoire est $\epsilon_{\text{al}} = x_i - \bar{x}$. Cette possibilité est présentée dans Ehrlich, Dybkaer et Wöger (2007), pp.204–205. Elle est motivée par le fait que de telles définitions rendent la composante d’erreur aléatoire calculable à partir des données expérimentales, alors que $\mu_{\hat{x}}$ n’est pas connue. Cependant, l’intérêt d’une telle définition est très contestable, parce que l’erreur systématique n’est plus une propriété du processus de mesure seule, mais également de l’échantillon de données expérimentales ; en particulier, selon une telle définition, l’erreur systématique varierait à chaque nouvelle mesure. Dès lors, il devient plus difficile de raisonner sur l’erreur systématique et de formuler les hypothèses supplémentaires que nous exprimons dans la suite de cette section. Dans le présent travail, nous nous en tenons donc à la définition donnée dans l’équation (3.7).

est bien évidemment elle-même inconnue. Par conséquent, *après correction*, la composante systématique de l'erreur est inconnue, au même titre que l'erreur totale, et pour la même raison : si celle-ci était connue, elle pourrait être corrigée, et, une fois corrigée, ne serait plus une composante d'erreur. Puisque l'erreur systématique affectant un résultat de mesure est inconnue (au moment de la mesure²⁷), l'hypothèse la plus raisonnable consiste à la supposer comme nulle. C'est ainsi que le GUM présente la chose :

On suppose qu'après correction l'espérance mathématique de l'erreur qui provient d'un effet systématique est égale à zéro.²⁸

Nous en venons de ce fait à la seconde hypothèse du modèle fréquentiste :

(H2) Une fois toutes les corrections effectuées, l'erreur systématique est considérée nulle : $\epsilon_{\text{sys}} = 0$, c'est-à-dire que l'espérance de \hat{x} s'identifie à la valeur vraie : $\mu_{\hat{x}} = \chi$.

Étant donnée l'équation (3.5), l'hypothèse (H2) est encore équivalente à l'hypothèse (H2') suivante :

(H2') Une fois toutes les corrections effectuées, l'espérance des erreurs aléatoires est considérée comme nulle.

L'hypothèse (H2') revient à dire, de façon plus courante, que les erreurs de mesure « s'annulent en moyenne »²⁹, un principe qui a été proposé dès 1750 par Thomas Simpson³⁰. Dès lors, prendre la moyenne des résultats de mesure revient à sommer des erreurs de mesure qui ont tendance à se compenser mutuellement. C'est pourquoi il est préférable de s'intéresser à la moyenne des valeurs obtenues qu'aux données prises individuellement. Cependant, cela ne vaut qu'à la condition de la validité de l'hypothèse (H2), c'est-à-dire à la condition que les erreurs systématiques aient été correctement corrigées, ce qui ne peut jamais être parfaitement garanti.

Avec l'hypothèse (H2), le problème peut être réduit en un problème uniquement statistique, et nous verrons à la section suivante que ce problème trouve une réponse très classique dans la démarche fréquentiste. Ce faisant, la formulation de l'hypothèse (H2) peut donner l'illusion que l'analyse d'incertitude est uniquement une question statistique, et qu'ainsi, l'expérimentateur se débarrasse de la question pourtant essentielle des erreurs systématiques. De fait, Hon critique certaines démarches qui réduisent l'analyse d'erreur à une étude simplement statistique des erreurs dites aléatoires³¹. On retrouve également dans le GUM lui-même une remarque visant à prévenir contre toute tentation de réduire l'analyse d'incertitude à une routine purement

27. Notons que, comprise ainsi, l'erreur systématique n'est certainement pas vouée à rester éternellement inconnue. C'est au moment où la mesure est analysée qu'il faut la considérer inconnue ; il est plus logique, par ailleurs, d'incorporer les biais de mesure connus à la fonction de mesure, au travers de termes de correction.

28. Joint Committee for Guides in Metrology (JCGM) (2008d), p.5.

29. Joint Committee for Guides in Metrology (JCGM) (2008d), p.5.

30. Armatte (2011), p.30.

31. Dans un premier temps, Hon critique ce qu'il considère comme un discours ordinaire sur les erreurs systématiques : « il est affirmé que l'expérimentateur qualifié peut éliminer toutes les erreurs systématiques. [...] Ainsi, l'expérimentateur est coupable quand il ou elle commet, pour ainsi dire, des erreurs systématiques : l'expérimentateur doit être blâmé pour avoir échoué à éliminer les erreurs systématiques ou à en avoir rendu compte, mais ne peut pas être blâmé pour l'occurrence d'erreurs aléatoires ; ces dernières sont inévitables et l'expérimentateur n'est pas responsable de leur apparition », Hon (1989b), p.475. Il ajoute ensuite : « (e)n somme, le point de vue standard est que, en dehors des erreurs aléatoires, toutes les erreurs expérimentales peuvent être éliminées, et que la distribution des erreurs aléatoires peut être capturée par une loi simple, à savoir la loi normale qui, dit-on, a

statistique³². Cependant, l'hypothèse (H2) ne doit pas être comprise comme une occultation du problème de l'erreur systématique. Elle consiste seulement à affirmer que, *une fois la traque des biais expérimentaux considérée comme (temporairement) achevée*, il n'y a pas de raison de supposer autre chose qu'une absence d'erreur systématique, bien qu'il soit, dans le même temps, largement raisonnable de penser que les corrections effectuées ne sont pas parfaites et qu'il subsiste toujours des erreurs de mesure inconnues qui n'ont pas été corrigées. Par conséquent, l'hypothèse (H2) n'exempte en aucun cas l'expérimentateur d'effectuer un travail expérimental et théorique poussé dans lequel il chercherait à identifier un maximum de biais expérimentaux. Au contraire, il s'agit même là du cœur de son travail de spécialiste ! Seulement, cette étape du travail expérimental ne se situe pas au niveau de l'exploitation statistique des données mais lui est soit antérieure (dans la construction du modèle de mesure) soit postérieure (si les nouvelles mesures effectuées révèlent des incohérences qui ne peuvent être expliquées que par l'existence d'un biais expérimental).

En opérant un lien entre valeur vraie de la grandeur soumise à la mesure et espérance de la variable aléatoire décrivant le processus de mesure, l'hypothèse (H2) permet de relier le modèle physique et le modèle statistique de la grandeur étudiée. Une fois cette hypothèse formulée, il est plus facile d'entrevoir une solution au problème posé, qui a été réduit à une question statistique assez classique.

été établie de manière empirique. », [Hon \(1989b\)](#), p.476. Il est vrai que certains textes spécialisés ont trop souvent tendance à réduire l'analyse d'incertitude à un travail purement statistique sur les collections de données expérimentales. Ainsi, le manuel de John Taylor, par ailleurs de très bonne qualité, ne traite que très brièvement des « incertitudes systématiques », qui plus est dans une sous-partie du chapitre « analyse statistique des incertitudes aléatoires », alors qu'un chapitre entier est consacré à la loi normale (voir [Taylor, 2000](#)). Cependant, si l'analyse de Hon est peut-être pertinente quant aux travaux de philosophie des sciences portés sur la question de l'erreur, il nous semble en tout cas qu'elle ne s'applique pas aux métrologues qui ont quant à eux parfaitement conscience du rôle central des erreurs systématiques dans la question de l'erreur. Toujours chez Taylor, nous trouvons d'ailleurs une remarque, à destination d'étudiants en science, qui nous semble illustrer que, dans la communauté scientifique tout du moins, un expérimentateur n'est pas « coupable » de commettre une erreur systématique : « (à) l'évidence, les incertitudes systématiques sont particulièrement difficiles à repérer et ont défié les efforts de grands scientifiques. En tout état de cause, vous ne serez donc pas jugé trop sévèrement si vous n'y parveniez pas vous-même. Vous devez néanmoins comprendre le sujet et admettre en toute modestie que vos mesures présentent fort probablement des incertitudes systématiques que vous êtes incapables d'identifier », [Taylor \(2000\)](#), p.102. La critique de Hon ne justifie pas non plus d'abandonner la dichotomie entre erreurs aléatoires et systématiques, qui est cruciale pour développer un modèle statistique fonctionnel.

32. « Bien que ce *Guide* fournisse un cadre pour l'estimation de l'incertitude, il ne peut remplacer ni la réflexion critique ni l'honnêteté intellectuelle ni la compétence professionnelle. L'évaluation de l'incertitude n'est jamais une tâche de routine ni une opération purement mathématique ; elle dépend de la connaissance détaillée de la nature du mesurande et du mesurage. La qualité et l'utilité de l'incertitude fournie pour le résultat d'un mesurage dépendent, en fin de compte, de la compréhension, de l'analyse critique et de l'intégrité de ceux qui contribuent à son évaluation. », [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.8.

3.3 Estimateurs statistiques et troisième hypothèse de l'analyse fréquentiste

Les hypothèses (H1) et (H2), définissent un cadre mathématique précis dans lequel opérer. Dès lors que la valeur vraie χ visée est identifiée à l'espérance $\mu_{\hat{x}}$ de la variable aléatoire \hat{x} , le problème devient purement statistique. C'est un problème d'échantillonnage, il se résume à évaluer l'espérance d'une variable aléatoire à partir d'un échantillon de réalisations de cette variable aléatoire, et sa résolution n'est pas spécifique au domaine de la métrologie. Nous présentons ici une résolution du problème qui passe par l'introduction d'« estimateurs statistiques » et qui est reconstruite à partir de la présentation qui est en faite dans le GUM.

La cible de l'inférence statistique est l'espérance $\mu_{\hat{x}}$ inconnue de la variable aléatoire \hat{x} . L'approche fréquentiste traditionnelle fonctionne par la construction d'un *estimateur statistique* de l'espérance $\mu_{\hat{x}}$ visée. Un estimateur statistique de $\mu_{\hat{x}}$ est une grandeur statistique que l'on calcule à partir de l'échantillon de données expérimentales et qui fournit une valeur approchée satisfaisante de l'espérance $\mu_{\hat{x}}$ visée. L'estimateur statistique de $\mu_{\hat{x}}$ le plus classique est la moyenne arithmétique \bar{x} de l'échantillon $\{x_i\}$, définie par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.8)$$

La moyenne arithmétique est un estimateur statistique de l'espérance $\mu_{\hat{x}}$ car sa valeur tend vers $\mu_{\hat{x}}$ quand la taille de l'échantillon tend vers l'infini (on parle plus précisément d'estimateur « sans biais »³³). Toutefois, \bar{x} ne coïncide pas avec $\mu_{\hat{x}}$ – ou, du moins, cela est très improbable. On ne peut donc pas se contenter que \bar{x} soit un estimateur de $\mu_{\hat{x}}$; il est également nécessaire de savoir en quoi \bar{x} peut être considérée comme une « bonne » valeur approchée de $\mu_{\hat{x}}$.

Cette question est délicate, et engage une distinction subtile entre estimation en tant que résultat et estimation en tant que processus. De prime abord, l'on pourrait être tenté de chercher la réponse dans la différence $|\mu_{\hat{x}} - \bar{x}|$ entre les valeurs de l'espérance visée $\mu_{\hat{x}}$ et de son estimateur \bar{x} . Cette différence pourrait être appelée, en s'inspirant du vocabulaire de la métrologie, l'erreur ou l'« inexactitude » de l'estimateur : la moyenne arithmétique \bar{x} est un estimateur d'autant plus inexact de $\mu_{\hat{x}}$ que $|\mu_{\hat{x}} - \bar{x}|$ est grande. L'objectif étant d'estimer $\mu_{\hat{x}}$ par \bar{x} , il est souhaitable que cette différence soit aussi petite que possible. De ce point de vue, l'on peut considérer que \bar{x} est une valeur approchée de $\mu_{\hat{x}}$ d'autant meilleure que $|\mu_{\hat{x}} - \bar{x}|$ est faible. Or, cela est problématique car cette différence est *inconnue* au même titre que $\mu_{\hat{x}}$, et le critère proposé ne permet donc pas de savoir si, *effectivement*, \bar{x} est un bon estimateur de $\mu_{\hat{x}}$. De ce fait, l'écart $|\mu_{\hat{x}} - \bar{x}|$ n'est pas un paramètre satisfaisant pour déterminer si l'estimateur \bar{x} est ou non une bonne valeur approchée de sa cible $\mu_{\hat{x}}$. L'origine du problème réside dans le fait que la moyenne arithmétique d'un échantillon de données est fixe, et est déterminée une fois pour toutes dès lors que les mesurages ont été effectués. Ainsi, \bar{x} peut être plus ou moins inexacte sans que rien, du point de vue de l'expérimentateur, ne lui permette de le déterminer. On comprend ainsi que ce n'est pas la valeur effective de \bar{x} elle-même qui a de l'importance ici, mais la *tendance* qu'a le *processus* à *produire* des valeurs de \bar{x} plus ou moins exactes.

33. Saporta (2006), p.290.

Ces remarques rejoignent les travaux de Ramsey sur l'épistémologie de l'approximation. Ramsey affirme qu'il ne faut pas réduire l'évaluation d'une approximation à la seule question de l'amplitude du désaccord numérique entre valeurs théorique et expérimentale³⁴. Selon Ramsey, de telles positions, courantes chez de nombreux auteurs, présentent l'inconvénient de restreindre l'approximation à une opération de copie, la création d'un objet *similaire* à celui qui doit être approximé³⁵; cela occulte un certain nombre des caractéristiques et des vertus épistémologiques de l'approximation. Ramsey propose de prendre en compte des critères supplémentaires qui se focalisent sur l'opération d'approximation en tant que *processus* et qui garantissent la qualité du processus d'approximation *indépendamment* de l'exactitude effective d'une valeur approximée donnée³⁶. Bien que l'inférence fréquentiste étudiée ici ne soit pas rigoureusement une approximation, la question de la qualité de l'estimateur choisi recoupe tout à fait la problématique que soulève Ramsey. Puisqu'il n'existe aucun moyen de connaître exactement $|\mu_{\hat{x}} - \bar{x}|$, il faut s'assurer plutôt que le *processus* de production d'un estimateur de $\mu_{\hat{x}}$ est effectivement fiable, c'est-à-dire qu'il a *tendance* à produire des estimateurs adéquats. Bien sûr, conformément à ce que souligne également Ramsey à propos des approximations, l'écart $|\mu_{\hat{x}} - \bar{x}|$ reste déterminant, dans le sens où il constitue l'objectif de l'inférence statistique; cependant, cet écart n'est pas directement mis à contribution dans l'évaluation de l'inférence effectuée. Il ne faut donc pas raisonner sur l'estimateur \bar{x} effectivement obtenu dans un cas particulier, mais sur la méthode générale par laquelle est obtenue \bar{x} . Cette focalisation sur le processus et non sur le résultat lui-même est une caractéristique essentielle des modèles fréquentistes de la mesure³⁷. De fait, on aboutit en particulier au résultat contre-intuitif selon lequel \bar{x} peut être un bon estimateur de $\mu_{\hat{x}}$ tout en étant fortement inexact : le processus peut avoir tendance à fournir régulièrement des estimateurs raisonnablement exacts tout en aboutissant parfois à des estimateurs très éloignés de la valeur visée.

Nous renvoyons à l'annexe A.1 pour une description plus complète de l'analyse statistique effectuée dans le GUM. Nous n'exposerons ici que les résultats essentiels. Un résultat classique

34. « Pour évaluer une approximation, la plupart des auteurs placent une limite à l'écart admissible entre les valeurs théoriques et expérimentales. Dans cet article, je soutiens qu'un tel écart est une condition nécessaire mais en aucun cas suffisante pour juger de la validité d'une approximation », Ramsey (1992), p.154. Barberousse a souligné en quoi la position de Ramsey se démarque d'un certain nombre d'autres travaux, par ailleurs tout à fait féconds, qui « mettent de côté les caractéristiques sémantiques, et parfois métaphysiques, de la notion d'approximation, mais laissent de côté son versant épistémologique », Barberousse (2008), p.55.

35. « Je fais aussi valoir que les auteurs qui se reposent sur l'amplitude du désaccord comme critère d'évaluation le font parce qu'ils caractérisent l'approximation en grande partie comme une ressemblance avec quelque chose d'autre », Ramsey (1992), p.154.

36. « Je suis Cartwright et Laymon en adoptant une caractérisation qui souligne le processus d'approximation. [...] Une approximation est un acte ou un processus et pas seulement une relation », Ramsey (1992), p.157 (nous soulignons). « En somme, lorsqu'on fait l'éloge ou qu'on blâme un résultat approximé, il n'est pas suffisant de considérer uniquement l'amplitude de l'écart entre les résultats théorique et expérimental (ou observationnel). La façon dont nous sommes arrivés au résultat est tout aussi importante que le fait que nous soyons arrivés très près de l'endroit où nous voulions aller. Seule une discussion sur le processus d'approximation nous fournira les moyens de déterminer si la stratégie d'approximation a été employée de façon légitime ou non. », Ramsey (1992), p.162.

37. On retrouve cela de façon très nette chez Eisenhart; ainsi, « (l)'erreur systématique, la précision, et l'exactitude sont des caractéristiques inhérentes à un processus de mesure et pas à une mesure particulière générée par le processus », Eisenhart (1963), p.162.

montre que la moyenne arithmétique \bar{x} de l'échantillon a tendance à être d'autant plus proche de $\mu_{\hat{x}}$ que :

- (i) la tendance à la dispersion des données expérimentales est faible. Celle-ci est caractérisée par l'écart-type σ de la variable aléatoire \hat{x} – ou, alternativement, par sa variance $V = \sigma^2$.
- (ii) la taille de l'échantillon, c'est-à-dire le nombre n de mesurages effectués, est grande. En effet, la qualité du résultat final est d'autant meilleure que la mesure est répétée, puisque cela fait croître la quantité d'informations disponibles à propos de la cible.

Cependant, le point (i) n'est utile que s'il est possible, par ailleurs, de connaître la variance V de la variable aléatoire \hat{x} . Or, V , au même titre que $\mu_{\hat{x}}$, est une propriété *inconnue* de la variable aléatoire \hat{x} . Seule est connue la dispersion effective des données expérimentales collectées; mais ces dernières peuvent être, par le fruit du hasard, plus ou moins dispersées que ce que le processus de mesure a tendance à produire. Par conséquent, la méthode fréquentiste ne s'appuie pas sur la construction d'un seul mais de deux estimateurs statistiques, l'un pour évaluer la valeur inconnue $\mu_{\hat{x}}$, l'autre pour évaluer la variance V de \hat{x} . Ce second estimateur est appelé « variance empirique », est noté s^2 , et est défini de la façon suivante :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.9)$$

On montre que, comme \bar{x} , s^2 est un estimateur « sans biais », c'est-à-dire que s^2 tend vers V lorsque la taille de l'échantillon tend vers l'infini³⁸. En combinant les deux points (i) et (ii), on montre alors que l'on peut construire un paramètre d'incertitude $u(x)$, par :

$$u(x) = \sqrt{\frac{s^2}{n}} \quad (3.10)$$

Le terme $u(x)$ a longtemps été désigné, dans la littérature anglophone, sous le nom de “standard error of the mean”³⁹; dans le GUM, il lui est préféré l'appellation « écart-type expérimental de la moyenne »⁴⁰. Le GUM marque l'idée que le terme $u(x)$ n'est pas simplement un terme statistique de dispersion mais peut être compris comme un paramètre d'incertitude, en introduisant un nouveau concept, celui d'« incertitude-type »⁴¹, auquel $u(x)$ vient *ici* s'identifier. L'incertitude-type est un paramètre de dispersion issu des propriétés de l'échantillon de données; elle est d'ailleurs définie dans le VIM3 comme une « incertitude de mesure exprimée sous la forme d'un écart-type »⁴²; cependant, comme nous le verrons par la suite, les paramètres d'incertitude-type ne sont pas toujours calculés à partir des équations (3.9) et (3.10).

38. On parle parfois de « variance empirique corrigée » pour marquer la différence avec la grandeur statistique $(1/n) \sum (x_i - \bar{x})^2$, qui n'est pas un estimateur sans biais (Saporta, 2006, p.290). Par la suite, nous n'apporterons aucune importance à cette distinction et nous appellerons « variance empirique » la grandeur s^2 .

39. Voir à ce sujet la remarque de [Campion, Burns et Williams \(1973\)](#), p.1.

40. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.11. Le terme d'erreur de la moyenne est d'ailleurs critiqué : « (l)'écart-type expérimental de la moyenne est parfois appelé à tort erreur de la moyenne », [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.38.

41. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.3. Bien que le terme « incertitude-type » soit défini dans le GUM dès 1993, il est absent de l'édition du VIM de la même année (VIM2, 1993) et n'apparaît dans le VIM qu'en 2008 avec la troisième édition.

42. [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.26.

Pourquoi $u(x)$ peut-il être compris comme un paramètre d'*incertitude*? Il faut pour cela comprendre deux points spécifiques. Pour commencer, $u(x)$ n'est pas une erreur de mesure. L'erreur absolue sur la cible $|\bar{x} - \mu_{\hat{x}}|$ demeure inconnue, et l'incertitude-type $u(x)$ ne s'y identifie certainement pas. Rappelons d'ailleurs que si une erreur de mesure est connue, celle-ci peut alors être corrigée, et, par conséquent, n'est à l'origine d'aucune incertitude de mesure. L'incertitude-type traduit un double constat : d'une part, il existe une erreur de mesure, et l'amplitude exacte de cette erreur n'est *pas* connue (constat négatif) ; mais d'autre part, l'existence de cette erreur est attestée et suffisamment de connaissances ont été acquises à son propos pour que l'on puisse formuler une proposition quant à la grandeur mesurée (constat positif). Par conséquent, si l'on peut effectivement parler d'« incertitude » pour un tel paramètre, c'est parce qu'il indique l'amplitude avec laquelle la variabilité du résultat de mesure *masque la connaissance* de la cible $\mu_{\hat{x}}$. L'incertitude-type $u(x)$ est calculée à partir de la mesure d'une variabilité physique, mais elle ne vise pas tant à représenter la variabilité elle-même qu'à caractériser les conséquences de cette variabilité sur la façon dont la mesure permet de connaître la cible ; il y a donc un constat épistémique derrière l'outil statistique.

L'expression « incertitude-type » est la contraction des deux notions dont le terme dérive : incertitude de mesure et écart-type statistique. De fait, le raisonnement employé jusqu'ici ne s'appuie que sur des calculs d'espérance et de variance, c'est-à-dire sur des calculs en termes de *limites* asymptotiques uniquement. Par conséquent, l'incertitude-type présente un intérêt pragmatique tout particulier en ce qu'elle est aisée à appréhender mathématiquement. D'une part, elle se calcule très facilement, et ce calcul, en particulier, ne nécessite aucune hypothèse sur la *forme* de la distribution de probabilité de la variable aléatoire \hat{x} (nous reviendrons sur cet aspect dans la prochaine section). D'autre part, elle est ensuite plutôt simple à manipuler. Si l'on mesure par des mesures répétées et indépendantes deux grandeurs d'entrée différentes X et X' d'un modèle de mesure $Y = X + X'$, les variances des distributions \hat{x} et \hat{x}' sont telles que :

$$V(\hat{y} = \hat{x} + \hat{x}') = V(\hat{x}) + V(\hat{x}') \quad (3.11)$$

ce qui justifie la formule dite de « propagation des incertitudes », longtemps appelée loi de « propagation des erreurs »⁴³ qui stipule que l'on peut combiner les incertitudes-types de façon quadratique⁴⁴ :

$$u_c^2(y) = u^2(x) + u^2(x') \quad (3.12)$$

43. Le nom de loi de « propagation des erreurs » est désormais jugé inapproprié à double titre. Il est d'abord considéré comme obsolète, car il fait écho à une période durant laquelle les termes d'erreur et d'incertitude étaient considérés comme synonymes, ce qui n'est plus le cas aujourd'hui. De plus, le GUM émet une deuxième objection, en renvoyant cette dénomination à une autre pratique de propagation qu'il considère « sécuritaire », fondée sur une somme non pas quadratique mais linéaire. Cette pratique de propagation était et est toujours employée dans les modèles non probabilistes, comme celui que nous décrivons plus loin, dans la section 3.5. Voir [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), annexe E, pp.56–62, en particulier les points E.3 et E.5 pour une discussion de ces pratiques « sécuritaires » du point de vue contemporain.

44. En toute rigueur, il faudrait également montrer que la relation précédente est également valable pour les variances empiriques, puisque, rappelons-le, les incertitudes-types ne correspondent pas aux variances des distributions mais bien à leurs estimateurs.

et qui a été utilisée extensivement depuis les débuts de la théorie des erreurs⁴⁵. Par sa simplicité, l'incertitude-type se révèle satisfaisante pour de nombreux usages. Cette simplicité, cependant, peut être vue simultanément comme un avantage et comme une limite.

L'avantage réside dans une propriété essentielle de l'incertitude-type : son calcul ne demande aucune hypothèse sur la forme de la distribution de \hat{x} – en effet, elle n'en dépend pas. Cela octroie à l'incertitude-type une grande généralité, et c'est précisément cette généralité qui en fait un outil prisé dans les pratiques scientifiques. Au cours du XX^e siècle, les scientifiques ont rappelé avec une insistance croissante l'intérêt que présentent des calculs sur les écarts-types; cette attitude procède d'une volonté de se détacher de toute hypothèse sur la forme des distributions probabilistes employées, et en particulier de s'émanciper de l'hypothèse gaussienne. Cela est particulièrement visible dans l'histoire des ajustements des constantes de la physique (voir la partie III où nous revenons plus en détail sur cette activité). Au début du XX^e siècle, c'est notamment le concept d'« erreur probable » qui est utilisée de façon très répandue⁴⁶. L'erreur probable correspond à une amplitude que l'erreur finale est susceptible de dépasser avec une probabilité 1/2. Cela renvoie donc à un constat probabiliste, et il n'est possible de la calculer que lorsqu'on connaît la distribution des erreurs. Dans la pratique des scientifiques du début du XX^e siècle, l'erreur probable est attachée à l'hypothèse gaussienne. Mais Jesse W. DuMond et E. Richard Cohen soulignent dès 1948 que l'usage de l'écart-type est préférable à celui de l'erreur probable. Cette préférence est émise précisément parce que l'écart-type permet de s'affranchir de toute hypothèse sur la forme de la distribution des erreurs⁴⁷. Une telle préférence ne se limite d'ailleurs pas au domaine de la mesure de précision des constantes de la physique, elle s'étend à des sphères plus larges dans la seconde moitié du XX^e siècle. Ainsi, Gié et Moreau ont également mis en garde contre l'excès de confiance en certaines hypothèses probabilistes, et souligné la bonne robustesse de l'écart-type⁴⁸.

45. Notons également que cette formule de propagation n'est valable que dans les cas où les variables statistiques ne sont pas corrélées. Si la valeur de X dépend de celle de X' , ou inversement, alors on ne peut plus considérer que les erreurs ont tendance à « se compenser » lorsqu'elles sont sommées.

46. On note également l'usage de l'« erreur moyenne » ; les deux concepts sont équivalents à un facteur près. Comme l'indique Birge en 1929, « les scientifiques allemands utilisent souvent le terme d'« erreur moyenne. » C'est une quantité qui, lorsqu'elle est multipliée par 0,674 5, donne l'erreur probable », Birge (1929b), p.4.

47. « Dans la théorie générale des ajustements aux moindres carrés de plusieurs variables, l'écart-type de chaque variable peut être défini d'une façon simple, qui implique seulement les quantités observées et est indépendante des hypothèses détaillées quant à la nature de la distribution dont les données ont été tirées. L'erreur probable, d'un autre côté, est plus difficile à déterminer et exige une connaissance ou au moins une supposition de la forme des fonctions de distribution », DuMond et Cohen (1948), p.87. DuMond et surtout Cohen répéteront à plusieurs occasions que le champ d'application des méthodes d'analyses peut aisément être élargi aux cas non gaussiens, à condition de travailler sur les écarts-types et non sur les erreurs probables. Voir par exemple Cohen et DuMond (1965), pp.540–541.

48. « Un tournant s'est amorcé, il y a quelques années, donnant à la théorie probabiliste une place de choix. Mais on risquait, a contrario, un engouement injustifié pour le “tout probabiliste”. En effet, il est nécessaire de garder à l'esprit qu'on ne dispose, en général, que d'un nombre limité de résultats de mesures, et qu'en outre, les lois statistiques supposées gouverner les erreurs de mesures sont, a priori, inconnues, du moins dans un grand nombre de cas. Il importe donc d'utiliser un outil d'évaluation s'accommodant autant que possible de cette ignorance. », Gié et Moreau (1987), p.159. L'incertitude-type répond exactement à cette dernière injonction : elle permet de rendre compte de l'incertitude sur une mesure *sans* avoir à effectuer d'hypothèse sur la forme de la distribution, donc sans savoir quelle est la distribution effective des erreurs.

La généralité et la commodité technique de l'incertitude-type vient avec une contrepartie. Cette contrepartie est l'impossibilité de raisonner en termes d'intervalles probabilistes. En particulier, si le calcul d'une incertitude-type ne dépend pas de la forme de la distribution de \hat{x} , c'est qu'elle ignore cette information et ne fournit alors qu'une description synthétique et simplifiée. Par conséquent, en toute rigueur, l'information fournie par une incertitude-type est incomplète, et un modèle complet devrait viser une formulation plus aboutie encore par la production d'intervalles d'incertitude *probabilistes*, c'est-à-dire des intervalles d'incertitude auxquels sont attachées des interprétations formulées en termes de probabilités. Sur le plan conceptuel, l'incertitude-type n'est donc pas entièrement satisfaisante, et, pour aller jusqu'au bout de la démarche, il se révèle nécessaire d'introduire une hypothèse supplémentaire, et de mener des calculs probabilistes plus poussés. Dans le domaine fréquentiste, les intervalles d'incertitude probabilistes sont des « intervalles de confiance », associés à un « niveau de confiance » matérialisé par une probabilité p . On parle encore de « limites de confiance ». Dans la section suivante, nous précisons la nature des concepts d'« intervalle de confiance » et de « niveau de confiance », et nous montrons qu'il est nécessaire d'introduire une nouvelle hypothèse du modèle fréquentiste afin de pouvoir les calculer. Cette hypothèse permet de faire le lien entre incertitude-type et incertitude probabiliste.

3.4 Niveau de confiance et intervalle de confiance

Nous venons de conclure sur le fait que l'analyse précédente, construite autour des estimateurs statistiques, est incomplète pour certains usages. Elle ne permet pas de conclure par une interprétation probabiliste de l'incertitude de mesure. Bien entendu, le modèle fréquentiste utilisé pour calculer l'« incertitude-type » est un modèle probabiliste, et la manipulation des variables aléatoires qu'il introduit fait appel à la théorie des probabilités. Cependant, l'incertitude-type elle-même ne peut pas être *interprétée* en termes probabilistes : il n'est pas possible d'y associer une quelconque probabilité qui décrirait d'une façon ou d'une autre le niveau de doute – ou, inversement, de confiance – associé à cette incertitude. « Niveau de confiance » est justement le terme utilisé par les statisticiens : c'est une probabilité p qui caractérise un intervalle d'incertitude donné. Le niveau de confiance p est choisi arbitrairement par l'expérimentateur. En effet, il exprime une exigence : l'expérimentateur souhaite formuler un intervalle de confiance qui satisfait le niveau de confiance demandé.

Dans cette section, nous décrivons la façon dont est construit un intervalle de confiance pour un niveau de confiance p choisi au préalable, et nous montrons comment interpréter ledit niveau de confiance. Nous verrons ainsi que les intervalles d'incertitude fréquentiste sont des « intervalles de confiance » dont nous expliciterons la signification. Par la suite, nous désignerons un tel intervalle de confiance par le symbole I_p . En développant la façon dont on construit un intervalle d'incertitude associé à un niveau de confiance p donné, nous serons amenés à discuter l'une des spécificités les plus saillantes de l'approche fréquentiste : il est impossible de formuler une proposition probabiliste qui porte directement sur le paramètre cible $\mu_{\hat{x}}$. En effet, $\mu_{\hat{x}}$ est un paramètre *fixe*, et, par conséquent, il ne peut être l'objet d'une description probabiliste si les probabilités employées décrivent des fréquences relatives limites. Nous verrons ainsi, dans la continuité de ce que nous avons montré à la section précédente, que le traitement

fréquentiste ne peut porter, en toute rigueur, que sur le processus de production des données expérimentales – et, de façon plus large, sur le processus de mesure – et non sur le résultat lui-même.

Nous avons vu que les propriétés des estimateurs statistiques employés usuellement en métrologie sont indépendantes de la forme de la distribution de probabilité de la variable aléatoire \hat{x} . Dès lors, l'information véhiculée par ces estimateurs statistiques est incomplète. Si l'on souhaite aller plus loin (et uniquement si cela est souhaité), il est nécessaire de faire appel à une hypothèse supplémentaire, que nous nommerons (H3), et qui stipule la forme de la distribution de probabilité $f_{\hat{x}}$ de la variable aléatoire \hat{x} . Par « forme » de la distribution de probabilité, il n'est pas entendu l'expression complète de la fonction $f_{\hat{x}}$. La fonction de densité de probabilité contient des informations inconnues, comme l'espérance $\mu_{\hat{x}}$ (qui, rappelons-le, a été identifiée à la valeur vraie de la grandeur visée), ainsi que la variance V . La forme de la distribution correspond à une classe de fonctions qui représente un comportement général des erreurs de mesure dans certaines conditions elles-mêmes assez générales, les paramètres $\mu_{\hat{x}}$ et V étant laissés libres. L'hypothèse consiste donc à supposer que les erreurs de mesure ont tendance (sous certaines conditions) à avoir un comportement *général* bien identifié, indépendamment des conditions particulières (valeur vraie de la grandeur visée et variance du processus de mesure). L'hypothèse la plus classique est de considérer que \hat{x} est distribuée selon une loi normale, ou gaussienne :

(H3.1) \hat{x} est distribuée selon une loi normale :

$$f_{\hat{x}}(t) \propto \exp\left(-\frac{1}{2} \left(\frac{t - \mu_{\hat{x}}}{\sigma}\right)^2\right)$$

L'hypothèse gaussienne est le cadre de travail le plus classique, et ce dès les travaux séminaux de Laplace puis Gauss sur les fondements probabilistes de la théorie des erreurs. Elle est particulièrement classique en métrologie, et demeure d'ailleurs l'hypothèse de travail du GUM⁴⁹. Son utilisation est justifiée par des considérations à la fois empiriques et théoriques. Elle semble parfois considérée comme la distribution la plus « naturelle », et, sur le plan empirique, on en retrouve la forme dans de nombreuses distributions statistiques mesurées⁵⁰. D'un point de vue théorique, elle apparaît comme une conséquence directe du « théorème central limite » sous certaines conditions. Comme l'explique Gilbert Saporta, ce théorème établit que « dans certaines conditions la somme, et donc la moyenne, de variables indépendantes et de même loi est asymptotiquement une loi normale »⁵¹. Saporta conclut :

Ce phénomène est souvent exprimé de la manière suivante : si une variable est la résultante d'un grand nombre de causes, petites, à effet additif, cette variable suit

49. [Joint Committee for Guides in Metrology \(JCGM\) \(2009\)](#), p.11.

50. Cela a mené Quetelet à considérer la loi normale comme naturelle, une hypothèse que Stigler a qualifiée d'« hypnotique » : « (l)a renommée de Quetelet dans l'histoire des statistiques repose essentiellement sur deux idées hypnotiques : le concept de l'homme moyen, et l'idée que toutes les distributions naturelles de données proprement collectées et triées suivent une courbe normale », [Stigler \(1986\)](#), p.201.

51. [Saporta \(2006\)](#), p.43.

une loi de Gauss. On peut y voir la justification de l'emploi abondant et souvent abusif de la loi de Laplace-Gauss comme modèle.⁵²

Si l'on conçoit une erreur de mesure comme l'accumulation d'une somme de petites erreurs de causes indépendantes, on comprend que le théorème central limite justifie d'adopter l'hypothèse gaussienne. Même dans les cas où l'hypothèse gaussienne ne convient pas pour les résultats de mesurages individuelles, la distribution de la moyenne arithmétique de ces mesurages tend vers une loi normale lorsque le nombre de mesurages augmente. On retrouve par exemple ces arguments chez Birge et Deming, en 1934⁵³. Mayo qualifie ce théorème d'« étonnant » et de « remarquable » et estime qu'il structure les modélisations fréquentistes de l'expérience⁵⁴. La loi normale trouve donc une double justification, à la fois théorique et empirique. Cette double justification a pu conférer à la loi normale un statut usurpé d'universalité, et a suscité la critique⁵⁵. Mais il y a encore d'autres raisons pour lesquelles la distribution a prospéré. Cette dernière répond également à un impératif pragmatique, car elle se révèle particulièrement simple à l'usage, dans les calculs mathématiques en particulier. La combinaison des ces arguments empirique, théorique, et pragmatique explique pourquoi la loi normale est généralement employée « par défaut » dans les raisonnements statistiques de métrologie⁵⁶ – et dans le GUM. Le nom de « loi normale » est souvent attribué à Karl Pearson⁵⁷, qui regretta cependant ce choix :

Il y a de nombreuses années, j'ai appelé la courbe de Laplace-Gauss la courbe *normale*, dont le nom, alors qu'il évite une question internationale de priorité, a le désavantage de conduire les gens à croire que toutes les autres distributions de fréquence sont en un sens ou un autre "anormales"⁵⁸

Des critiques ont régulièrement été formulées à l'encontre de la trop grande confiance accordée au théorème central limite et l'emploi sans recul de la loi normale⁵⁹.

52. Saporta (2006), p.67.

53. Deming et Birge (1934), pp.123–125.

54. Mayo (1996), pp.171–172.

55. Poincaré a émis des réserves quant au caractère gaussien de la théorie des erreurs, et a rapporté ce mot, devenu fameux, de Lippmann, de la façon suivante : « [La loi des erreurs] ne s'obtient pas par des déductions rigoureuses ; plus d'une démonstration qu'on a voulu en donner est grossière, entre autres celle qui s'appuie sur l'affirmation que la probabilité des écarts est proportionnelle aux écarts. Tout le monde y croit cependant, me disait un jour M.Lippmann, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental. », Poincaré (1896), p.149. Pour plus de détails quant aux travaux de Poincaré sur les probabilités et sur la loi des erreurs, voir Mazliak (2015).

56. Ainsi Student écrit-il : « si notre échantillon est petit, [...] [il] n'est pas suffisamment grand pour déterminer quelle est la loi de distribution des individus. Il est courant, toutefois, de supposer une distribution normale car, dans un très grand nombre de cas, cela donne une approximation si proche qu'un petit échantillon ne nous donnera aucune information réelle quant à la façon dont la population dévie de la normalité : puisqu'une loi de distribution doit être supposée, il est préférable de travailler avec une courbe dont l'aire et les ordonnées sont tabulées, et dont les propriétés sont bien connues », Student (1908), p.1. On voit, dans cette citation, que ce n'est pas sur la base d'un seul argument, mais sur une combinaison d'une remarque sur la justesse de la loi normale et d'une remarque sur sa praticité. La remarque de Student montre de plus que la loi normale n'est pas considérée comme automatiquement valable en toutes circonstances.

57. Voir Stigler (1999), p.406.

58. Pearson (1920), p.25 (Pearson souligne).

59. Dans le domaine de la métrologie, on peut constater que la critique de la loi normale remonte à des périodes

L'hypothèse gaussienne demeure particulièrement utile en métrologie, mais, bien qu'elle soit historiquement attachée à l'approche traditionnelle, elle n'est pas en principe une condition nécessaire à son bon déroulement. On peut envisager également, à titre d'exemple, une loi rectangulaire :

(H3.2) \hat{x} est distribuée selon une loi rectangulaire :

$$\begin{cases} f_{\hat{x}}(t) \propto 1 & \text{si } t \in [a, b] \\ f_{\hat{x}}(t) = 0 & \text{sinon} \end{cases}$$

De manière plus générale, les distributions les plus intuitives et les plus courantes partagent certaines propriétés d'ensemble. Comme l'a montré Armatte,

(T)rois types de contraintes [...] vont longtemps circonscrire la famille des lois possibles, à savoir la symétrie qui rend équiprobable les erreurs de même grandeur absolue, la décroissance qui rend les petites erreurs plus probables que les grandes, et enfin l'hypothèse d'un borne supérieure ou limite d'erreur qui dépend à la fois de l'instrument et de l'observateur.⁶⁰

Ces propriétés restent aujourd'hui des propriétés considérées comme intuitives et le plus souvent vérifiées, sans qu'elles ne soient pour autant des conditions nécessaires. La troisième n'est cependant pas toujours prise en compte, en particulier du fait de la prédominance de la loi normale, qui n'est pas bornée (elle associe une densité de probabilité non nulle à toute amplitude d'erreur possible). Cependant, c'est là plus un artefact mathématique lié à l'emploi de la loi normale que le résultat d'un constat empirique.

Selon la façon dont est définie la fonction $f_{\hat{x}}$, la complexité du problème est circonscrite à un nombre fini de paramètres. Dans le cas de l'hypothèse gaussienne, $f_{\hat{x}}$ est entièrement définie dès que $\mu_{\hat{x}}$ et σ sont spécifiés. L'hypothèse (H3) est introduite afin de pallier la lacune que comporte le travail sur les estimateurs statistiques, à savoir l'impossibilité de caractériser l'incertitude-type par un niveau de confiance p . Elle permet un calcul mathématique à partir duquel on peut établir un intervalle d'incertitude appelé « intervalle de confiance » I_p associé au niveau de confiance p , tel que :

$$I_p = \left[\bar{x} - k_p u(x); \bar{x} + k_p u(x) \right] \quad (3.13)$$

où $u(x)$ est l'incertitude-type, et k_p un facteur qui dépend, une fois p choisi, de la forme de la distribution de probabilité de \hat{x} .

Pourquoi un tel intervalle peut-il être associé au niveau de confiance p ? Le calcul correspondant est développé plus en détail dans l'annexe A.2, et nous n'indiquons ici que les principes

assez lointaines. En 1934, Birge et Deming témoignent déjà d'un ensemble de critiques contre l'usage généralisé de la loi normale, même s'ils cherchent quant à eux à en défendre la légitimité en répondant à certaines de ces critiques (Deming et Birge, 1934, pp.123–125). L'article de 1965 de Cohen et DuMond montre qu'en aucun cas les physiciens ne considéraient à cette époque la loi normale comme absolument justifiée : « La distribution gaussienne d'erreur est fréquemment supposée être la forme typique applicable à toutes les mesures physiques. Cela, bien sûr, n'est pas vrai du tout », Cohen et DuMond (1965), p.540. De manière générale, nous renvoyons par exemple à Stigler (1999), pp.403–430 pour une discussion sur les origines et les critiques de la loi normale et de sa dénomination.

60. Armatte (2004), p.149.

généraux qui le sous-tendent. Pour interpréter un intervalle de ce type de façon pleinement fréquentiste, il nous faut une fois encore insister sur la différence entre le résultat proprement dit et le processus qui a amené au résultat. Il est essentiel de constamment marquer une différence entre deux catégories générales selon lesquelles on peut considérer les grandeurs statistiques employées. Celles-ci peuvent être vues (i) simplement comme un résultat, c'est-à-dire une valeur (ou un intervalle) *fixe* qui *a été obtenue* par l'expérimentateur lorsqu'il a mené l'opération de mesure. Mais elles peuvent également être comprises (ii) comme l'issue d'un *processus*, que le modèle fréquentiste suppose aléatoire, et que l'on peut dès lors caractériser en termes probabilistes, en le décrivant au moyen des fréquences relatives limites des issues qui *auraient* pu être obtenues par l'expérimentateur s'il avait répété la même opération une infinité de fois⁶¹. Ce n'est que selon le point de vue (ii) que l'on peut réellement comprendre la signification *probabiliste* des grandeurs statistiques utilisées – cela nécessite d'ailleurs, en toute rigueur, l'emploi d'un formalisme plus poussé que nous réservons pour l'annexe A.2 consacrée au développement technique de ces calculs.

L'intervalle I_p et le paramètre $\mu_{\hat{x}}$ sont tous deux bien définis et *fixes*. Par conséquent, il n'est pas possible de formuler une proposition telle que « il y a une probabilité p que l'intervalle I_p contienne le paramètre $\mu_{\hat{x}}$ »⁶². Plutôt, toute approche *fréquentiste* de la mesure ne peut porter que sur la *procédure dans son ensemble* (c'est-à-dire selon le point de vue ii exposé ci-dessus), et non sur le résultat final effectivement obtenu (point de vue i). C'est là la différence fondamentale avec le modèle bayésien que nous développerons à son tour, dans lequel les probabilités traduisent un niveau de connaissance et peuvent donc porter sur des grandeurs fixes mais inconnues. La relation entre $\mu_{\hat{x}}$ et I_p est clairement définie, et seuls deux cas sont possibles :

- (i) $\mu_{\hat{x}}$ est incluse dans I_p .
- (ii) $\mu_{\hat{x}}$ n'est pas incluse dans I_p .

Le terme $\mu_{\hat{x}}$ étant la cible de l'inférence, l'objectif est d'aboutir au cas (i), et donc de construire un intervalle de confiance qui contienne effectivement $\mu_{\hat{x}}$. On peut alors parler, le cas échéant, de « succès ». Dans le cas contraire, on pourra parler d'« échec ». Or, ce que le calcul probabiliste mené à partir du modèle fréquentiste établi sur les hypothèses (H1), (H2) et (H3) permet d'affirmer, c'est que le *processus* de production des données expérimentales a tendance à mener

61. La langue anglaise a intégré à son vocabulaire la différence entre les acceptions (i) et (ii), désignées respectivement par “estimate” (le résultat de l'estimation) et “estimator” (le processus d'estimation). Voir par exemple Willink (2013), p.59.

62. C'est une raison qui pousse Lecoutre à considérer cette interprétation comme contre-intuitive, et à plaider pour une préférence envers les statistiques bayésiennes. « (L)es interprétations spontanées des résultats des procédures statistiques traditionnelles (seuils de signification, intervalles de confiance), même par des utilisateurs “avertis”, sont le plus souvent en termes de probabilités sur les paramètres, qui sont en fait les probabilités naturelles : “celles qui vont du connu vers l'inconnu”. Ainsi, dans un ouvrage récent d'introduction à la statistique, appartenant à une collection destinée au grand public, dont l'objectif est de permettre au lecteur d'“accéder aux intuitions profondes du domaine”, on trouve l'interprétation suivante de l'intervalle de confiance (ou “fourchette”) pour une proportion : “si dans un sondage de taille 1000, on trouve P [la proportion observée] = 0.613, la proportion π_1 à estimer a une probabilité 0.95 de se trouver dans la fourchette : [0.58,0.64]” (Claudine Robert, 1995, page 221). Si vous n'êtes pas (encore) bayésien et si votre intuition profonde est que cette interprétation est, soit correcte, soit peut-être incorrecte mais en tout cas souhaitable, vous devez sérieusement vous demander si vous n'êtes pas un bayésien “qui s'ignore”. », Lecoutre (2005), p.92.

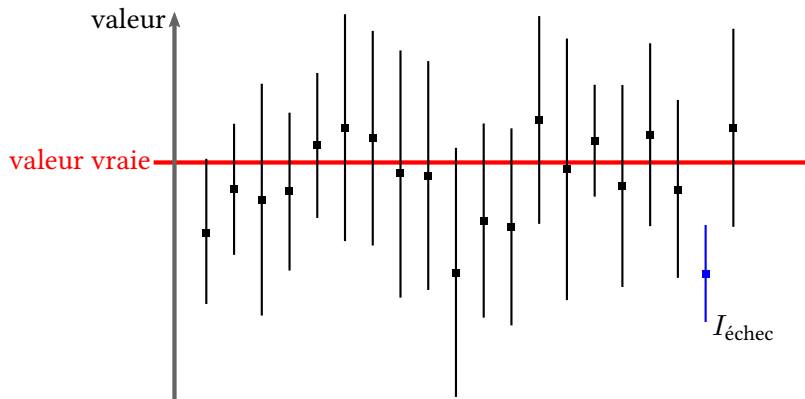


FIGURE 3.3 – Illustration de l'interprétation fréquentiste d'un intervalle de confiance. Sur cette figure sont représentés vingt intervalles de confiance à 95% calculés à partir d'échantillons de données de mêmes tailles, et indépendants les uns des autres. L'interprétation probabiliste de l'intervalle de confiance ne porte pas sur un intervalle particulier, mais sur la procédure qui amène à l'obtention de l'intervalle, qui aboutit à des succès avec une fréquence limite de 95%. Dans cet exemple, l'un des intervalles ($I_{\text{échec}}$, en bleu) ne contient pas la valeur vraie de la grandeur. Au total, 95% des intervalles contiennent la valeur vraie, ce qui est cohérent avec le niveau de confiance associé aux intervalles. Bien entendu, la fréquence effective d'intervalles de confiance contenant la valeur vraie n'est pas toujours exactement égale au niveau de confiance; c'est à la limite, pour une infinité d'intervalles, que le taux d'intervalles contenant la valeur vraie tend vers le niveau de confiance.

à un succès (cas i : $\mu_{\hat{x}} \in I_p$) avec une fréquence relative limite égale à p . Si l'expérimentateur répétait la même opération une infinité de fois, il constaterait que $100p\%$ des intervalles de confiance I_p obtenus contiennent effectivement $\mu_{\hat{x}}$. C'est sur la foi de ce résultat que l'on peut accorder à I_p une interprétation probabiliste, et parler à ce titre d'intervalle d'incertitude associé à un niveau de confiance p . On peut alors affirmer :

Si l'on répète la procédure consistant à collecter un échantillon de n données à partir duquel on calcule un intervalle de confiance I_p , alors la fréquence limite d'occurrence des cas pour lesquels l'intervalle I_p contiendra effectivement le paramètre $\mu_{\hat{x}}$ sera le niveau de confiance p .

Une illustration de cette proposition est fournie à la figure 3.3 où sont représentés plusieurs intervalles de confiance calculés à partir d'un tirage aléatoire d'échantillon de données de même taille.

Il est également possible de revenir à une formulation en termes d'« erreurs de mesure ». Tout écart $e = \bar{x} - \mu_{\hat{x}}$ entre l'estimateur \bar{x} obtenu et le paramètre statistique $\mu_{\hat{x}}$ visé aura pour répercussion une composante d'erreur de même amplitude dans le calcul final de la valeur du mesurande : il s'agit de la composante de l'erreur finale due aux « erreurs aléatoires ». Dans ce cas, on peut affirmer :

Si l'expérimentateur adopte pour valeur de X l'estimateur \bar{x} , son résultat sera entaché d'une erreur (inconnue) $e = \bar{x} - \mu_{\hat{x}}$. La procédure tend à commettre une erreur absolue $|e|$ qui sera, dans 100p% des cas, inférieure à la demi-largeur $U_p(x) = k_p u(x)$ de l'intervalle de confiance I_p .

Puisque la procédure tend à aboutir à des intervalles qui sont des succès selon un niveau de confiance prédéterminé, on peut considérer, en raisonnant à rebours, que l'on a une bonne raison de penser que l'intervalle nous permet de localiser l'espérance $\mu_{\hat{x}}$ visée. L'intervalle I_p est donc interprété comme le spectre de valeurs dans lequel on peut raisonnablement penser que $\mu_{\hat{x}}$ se trouve, avec le niveau de confiance p . On considérera que l'on prend un risque, puisque l'intervalle ne contient pas toujours $\mu_{\hat{x}}$, mais ce risque est connu, et le niveau de risque peut être adapté aux circonstances en choisissant le niveau de confiance avec lequel on détermine l'intervalle de confiance. C'est pourquoi I_p est relié à un constat d'incertitude : il permet de localiser l'espérance visée – et, en remontant au mesurande par le jeu des corrections, la valeur vraie de la grandeur visée – selon une procédure qui fonctionne avec une probabilité connue. Dans la continuité, la demi-largeur $U_p(x) = k_p u(x)$ de l'intervalle I_p est interprétée comme l'incertitude de mesure qu'il est raisonnable d'associer au résultat obtenu, avec le niveau de confiance p .

On se retrouve ici avec un terme d'incertitude, $U_p(x)$, d'une nature bien différente de l'incertitude-type $u(x)$ qui avait quant à elle été obtenue à partir d'un simple calcul de variance. La détermination de $U_p(x)$ passe par un calcul probabiliste qui dépend d'une hypothèse – l'hypothèse (H3) – qui stipule la forme de la distribution de probabilité de la variable aléatoire \hat{x} . L'incertitude $U_p(x)$ n'est donc pas un paramètre de dispersion, même si son calcul passe par la détermination préalable de $u(x)$. Par conséquent, contrairement à l'incertitude-type $u(x)$, l'incertitude $U_p(x)$ ne peut être proprement interprétée que si l'on indique le niveau de confiance auquel elle est associée. Le calcul et la manipulation de $U_p(x)$ se révèlent beaucoup plus complexes que ceux de l'incertitude-type, mais en contrepartie, le concept trouve une interprétation précise, contrairement à celle de l'incertitude-type qui laisse la place à l'ambiguïté.

Nous avons vu à la section 3.3 que durant la première moitié du XX^e siècle, les constats d'incertitude sont généralement formulés en termes d'« erreur probable », qui désigne l'amplitude que l'erreur finale est susceptible de dépasser avec une probabilité 1/2 – et qui est donc également susceptible d'être supérieure à l'erreur finale avec une probabilité 1/2. Il s'agit donc, selon les termes utilisés ici, d'une incertitude associée à un niveau de confiance $p = 0,5$. Dans le GUM et dans le VIM, $U_p(x)$ sera par la suite appelée « incertitude élargie à 100p% »⁶³, k_p étant le « facteur d'élargissement »⁶⁴. Nous verrons au chapitre suivant que cette dénomination s'insère dans une méthode d'analyse d'incertitude qui adopte une interprétation hybride des probabilités.

Le paramètre $U_p(x)$ calculé ici est le paramètre d'incertitude *lié à la variabilité du résultat* de la mesure sur X . Il ne constitue qu'une composante de l'incertitude totale de mesure, d'une

63. Joint Committee for Guides in Metrology (JCGM) (2012), p.27. Joint Committee for Guides in Metrology (JCGM) (2008d), p.24.

64. Joint Committee for Guides in Metrology (JCGM) (2012), p.28.

part parce que la variabilité n'est pas la seule source d'incertitude sur une grandeur d'entrée, et d'autre part parce qu'il faut également tenir compte de l'incertitude liée aux autres grandeurs d'entrée de la fonction de mesure. Dans la section suivante, nous passons à la caractérisation de l'incertitude de mesure pour les grandeurs à propos desquelles sont fournies des informations qui ne proviennent pas d'une mesure répétée. Nous en viendrons alors à la question de la combinaison des incertitudes, et nous montrons en quoi cette problématique s'est révélée être le pivot d'une critique de l'approche fréquentiste qui a abouti au développement d'une approche bayésienne fondée sur une interprétation épistémique des probabilités.

3.5 Traitement fréquentiste des erreurs systématiques

Les sections précédentes ont été consacrées à l'analyse des collections de données expérimentales obtenues dans les conditions dites « de répétabilité », c'est-à-dire au traitement statistique des erreurs aléatoires. Dans cette section, nous souhaitons désormais attaquer la question de la détermination des composantes d'incertitudes relatives aux cas où l'information disponible ne provient pas des résultats de l'expérimentation elle-même, mais sont importés par l'expérimentateur depuis différentes sources – ce que Stigler appelle une « évaluation externe de la précision »⁶⁵. Nous pourrions parler à ce sujet de traitement fréquentiste des erreurs systématiques. En effet, cela concerne typiquement la correction des effets constants (dans les conditions de répétabilité) qui entrent en jeu dans la modélisation générale du dispositif de mesure. Il convient de noter toutefois qu'il s'agit d'une façon bien spécifique de traiter des erreurs systématiques, *dans le cadre d'un dispositif spécifique de mesure*. Il existe d'autres façons de s'attaquer aux erreurs systématiques, par exemple en multipliant les mesures dans des conditions de reproductibilité, plus larges que les conditions de répétabilité, dans lesquelles les erreurs systématiques sont alors susceptibles de varier. C'est ce que font par exemple les physiciens lorsqu'ils mesurent des constantes physiques par des dispositifs séparés, parfois basés sur des principes différents, puis qu'ils comparent en retour les valeurs obtenues. Nous décrirons ce procédé à la partie III de cette thèse, consacrée à l'« ajustement » des constantes de la physique. Une autre façon de traiter les erreurs systématiques est d'aménager le dispositif de mesure de façon à les « randomiser », c'est-à-dire à les rendre variables et à les considérer ensuite comme le résultat d'un tirage aléatoire susceptible d'être décrit par des probabilités. Nous en donnons un exemple à l'annexe A.4. Ce type d'aménagement peut néanmoins se révéler très coûteux, et n'est pas toujours préférable à une modélisation théorique plus directe.

Nous reprenons la fonction de mesure simplifiée que nous avons proposée au début de ce chapitre, donnée par l'équation (3.1) qui stipule que $Y = X + Z$. Dans cette section, la grandeur d'intérêt est la grandeur d'entrée Z , que l'on suppose ne pas être l'objet d'une mesure spécifique dans le dispositif employé par l'expérimentateur : celui-ci exploite des informations qui lui proviennent de l'extérieur. Les informations exploitées par l'expérimentateur peuvent être des résultats d'expérience obtenus et transmis par d'autres scientifiques, comme ceux que l'on retrouve dans les tables de valeurs des constantes tenues par la communauté scientifique. Elles peuvent être des spécifications lues dans les notices des instruments utilisés, fournies par

65. Stigler (1986), p.6.

les constructeurs. L'on peut envisager également des cas où l'information provient de l'avis d'un expert sur un dispositif expérimental, ou encore d'une modélisation proposée par l'expérimentateur fondée sur des valeurs théoriques. Quoiqu'il en soit, l'expérimentateur ne dispose pas d'un échantillon de valeurs expérimentales issues de mesurages répétés de la grandeur.

Les informations disponibles à propos de la grandeur étudiée peuvent être communiquées de diverses façons. Il peut s'agir d'une valeur accompagnée d'une incertitude-type, ou encore accompagnée d'une incertitude « élargie » à un niveau de confiance spécifié. Si les expérimentateurs qui ont mesuré la grandeur Z ont eux-même suivi la méthodologie fréquentiste, alors l'information communiquée à propos de Z prendra logiquement la forme d'un intervalle de confiance, ou de tout autre résumé du résultat qui permet d'appréhender Z par des probabilités fréquentistes. Dans ces situations, le travail de l'expérimentateur est assez aisé, car les informations prennent directement la forme sous laquelle il souhaite les utiliser. L'on peut alors considérer qu'il s'agit simplement d'un cas de partage du travail expérimental et que l'expérimentateur peut utiliser les informations disponibles comme s'il les avait obtenues lui-même.

Les informations disponibles à propos de Z ne prennent toutefois pas toujours la forme d'une expression probabiliste. Une façon possible de rendre compte d'une erreur systématique est de la caractériser par un encadrement. Par exemple, le laboratoire qui a effectué l'étalonnage d'un étalon de travail peut communiquer le résultat de l'étalonnage de la façon suivante :

$$z_0 - a < z < z_0 + a \quad (3.14)$$

Un tel encadrement signifie que l'on considère raisonnable de penser que la valeur vraie ζ de la grandeur d'entrée Z est comprise entre les deux bornes :

$$z_0 - a < \zeta < z_0 + a \quad (3.15)$$

L'erreur de mesure ϵ_Z sur la grandeur Z , est, rappelons-le, définie par l'écart entre la valeur finale (adoptée, attribuée à Z) et la valeur vraie. Ici, la valeur finale n'est pas définie mais il est légitime de prendre le milieu de l'intervalle, soit z_0 . Dans ce cas, au vu de l'équation (3.15), il est raisonnable de penser que l'erreur de mesure est comprise dans l'encadrement suivant :

$$-a < \epsilon_Z < a \quad (3.16)$$

L'équation (3.16) fournit de ce fait un encadrement de l'erreur. Bien entendu, cet encadrement n'établit pas des limites certaines, mais il correspond à ce que le laboratoire d'étalonnage considère légitime de croire : c'est pourquoi on parle de « limites crédibles de l'erreur »⁶⁶. À une évaluation de ce type correspond une propagation des erreurs de type linéaire. En effet, si pour deux grandeurs Z et Z' , on a respectivement :

$$\begin{cases} -a < \epsilon_Z < a \\ -a' < \epsilon_{Z'} < a' \end{cases} \quad (3.17)$$

On peut alors affirmer que :

$$-(a + a') < \epsilon_{Z+Z'} < a + a' \quad (3.18)$$

66. Eisenhart (1963), p.181.

En cela, le calcul des « limites d'erreur » se différencie nettement de celui des intervalles statistiques produits par l'analyse probabiliste des erreurs dites aléatoires. Armatte fait remonter aux débuts de la théorie des erreurs la coexistence des deux traditions⁶⁷, l'une s'appuyant sur le calcul différentiel (comme le fait Roger Cotes en 1722) et pour lesquelles les limites d'erreur sont cumulatives, et l'autre probabiliste (défendue par exemple par Pierre Bouguer en 1749 ou Joseph Fourier en 1829) et pour laquelle les erreurs ont tendance à se compenser lorsqu'elles sont sommées les unes avec les autres, selon une intuition qui a émergé dès la première moitié du XVIII^e siècle⁶⁸. On retrouve une discussion similaire des deux modes de propagation chez Birge⁶⁹.

Aucune des deux traditions n'est *a priori* invalide si l'on demeure conscient que chacune d'entre elles est adaptée à un type particulier d'information disponible. Cependant, on entrevoit ici une difficulté majeure qui poussera les métrologues à critiquer l'approche traditionnelle : celle-ci oblige à pratiquer en parallèle deux méthodes mathématiques substantiellement différentes.

En soi, l'encadrement par des « limites crédibles de l'erreur » peut être interprété comme un constat d'incertitude, très similaire à celui qu'exprime une incertitude-type ou une incertitude « élargie », puisque dans les deux cas, l'objectif est de transcrire de façon quantitative les *conséquences* d'une variabilité physique inexplicée sur la façon dont on peut connaître la valeur de la grandeur mesurée. Cependant, à la différence du constat d'incertitude établi dans la section précédente, celui-ci ne se présente pas sous une forme probabiliste. Une telle forme est même rigoureusement exclue dans le cas présent, puisque les probabilités employées, fréquentistes, ne s'appliquent qu'à des résultats potentiels d'une procédure qui est *répétée*. Or, comme nous allons le voir, l'impossibilité de formuler des énoncés probabilistes à propos de ce type de grandeur est perçue, dans certaines situations, comme une limite sérieuse de l'approche fréquentiste⁷⁰. Dans la section suivante, nous explicitons les critiques qui ont été opposées à l'approche traditionnelle après avoir explicité la façon dont celle-ci se propose de formuler un résultat de mesure.

67. Armatte (2010), p.4.

68. « Lentement se dégage vers 1730 l'idée que l'erreur moyenne espérée diffère de l'erreur maximale à craindre. », Armatte (2004), p.148. Armatte parle encore à ce sujet d'une opposition entre « erreur du physicien » et « erreur du praticien ».

69. « Si chaque erreur peut être soit positive soit négative, l'erreur probable finale est obtenue par la racine carrée de la somme des carrés de toutes les erreurs probables, incluant les erreurs accidentelles au moindres carrés. Si, d'un autre côté, la supposée "limite d'erreur" est donnée pour chaque source d'erreur, alors de toute évidence la "limite d'erreur" finale doit être prise comme la *moyenne arithmétique* des composantes d'erreur limite », Birge (1929b), pp.6-7.

70. Il faut préciser toutefois que dès le début du siècle, l'estimation par des limites d'erreur ne faisait pas l'unanimité, en particulier chez les physiciens qui s'attelaient à la mesure de précision des constantes physiques. On trouve par exemple des critiques notoires chez Birge (1929b), Bearden et Thomsen (1957) (voir p.272), ou encore chez Cohen et DuMond (1965) (voir en particulier p.540). Ces critiques sont annonciatrices des débats que l'on retrouve plus tard, à la fin des années 1970, mais il est à noter qu'elles n'étaient pas alors orientées directement contre les principes généraux de l'approche traditionnelle dans son ensemble.

3.6 Formulation du résultat final et limites de l'approche traditionnelle

Dans les sections 3.1 à 3.4, nous avons montré comment l'approche fréquentiste effectue le traitement statistique des erreurs dites « aléatoires ». Dans la section 3.5, nous avons exposé ce qui peut être dit à propos d'une erreur dite « systématique ». Ce faisant, nous avons répondu au premier des deux problèmes mathématiques présentés au 1.4 : *comment évaluer l'incertitude de mesure pour les grandeurs d'entrée ?* L'étape suivante consiste à répondre au second de ces deux problèmes, à savoir la question de la combinaison des constats formulés sur chacune des grandeurs d'entrée – ce qui est appelé aujourd'hui la « propagation » des incertitudes – afin d'aboutir à une incertitude résultante, relative au mesurande Z . Cette étape n'est pas immédiate, d'autant que les composantes d'incertitudes exprimées sur les grandeurs d'entrée sont de nature différente selon la méthode qui a été utilisée pour les déterminer, et donc, en creux, selon le type d'erreur analysé.

3.6.1 Formulation traditionnelle du résultat final

Il nous faut ouvrir au préalable une parenthèse terminologique et historique, et revenir sur un élément essentiel, que nous avons déjà mentionné, mais qu'il est indispensable de rappeler pour remettre en perspective l'usage que nous avons réservé jusqu'ici à l'expression « incertitude de mesure ». En effet, le langage utilisé dans ce chapitre est quelque peu anachronique : il regroupe sous une même acception, « incertitude », une variété de concepts similaires mais pas toujours rigoureusement équivalents. Il est à noter que le terme d'« incertitude de mesure » lui-même n'apparaît qu'assez rarement, voire pas du tout, dans la littérature métrologique et scientifique du début du XX^e siècle. Les physiciens emploient alors majoritairement l'« erreur probable »⁷¹ dont l'usage disparaît dans les années 1960 au profit, justement, de celui d'incertitude de mesure. Un statisticien comme Eisenhart met en avant les notions de « précision » (auquel la terminologie standardisée de la langue française préfère « fidélité »⁷², afin d'éviter l'ambiguïté que procure la trop grande proximité de « précision » avec les usages du langage courant) et « exactitude » de mesure. Son article de 1963, qui peut être considéré comme un aboutissement de l'approche fréquentiste traditionnelle ne mentionne jamais le terme « incertitude ». Ces remarques éclairent l'une des citations introductives du GUM qui établit que :

Le concept d'*incertitude* comme attribut quantifiable est relativement nouveau dans l'histoire de la mesure bien que l'*erreur* et l'*analyse des erreurs* soient des concepts depuis longtemps pratiqués dans la science de la mesure, c'est-à-dire en métrologie.⁷³

Nous rejoignons tout à fait ce constat ; en revanche, la suite de cette citation laisse un peu perplexe :

On reconnaît maintenant largement que, lorsqu'on a évalué la totalité des composantes de l'erreur connues ou soupçonnées et que les corrections appropriées

71. Voir par exemple Birge (1932a).

72. Joint Committee for Guides in Metrology (JCGM) (2012), p.22.

73. Joint Committee for Guides in Metrology (JCGM) (2008d), p.vii.

ont été appliquées, il subsiste encore une incertitude sur la validité du résultat exprimé, c'est-à-dire un doute sur la manière dont le résultat de mesure représente correctement la valeur de la grandeur mesurée.⁷⁴

Il est bien peu vraisemblable que les scientifiques aient attendu si longtemps avant de prendre conscience que, même après correction, un résultat n'est pas certain. Ce que nous constatons surtout, c'est une évolution dans les méthodes et dans le vocabulaire employés, ainsi que dans l'interprétation des différents concepts employés pour caractériser un résultat de mesure, dont ceux d'erreur et d'incertitude de mesure.

Dans ce chapitre, nous avons choisi de nous en tenir le plus souvent à la terminologie contemporaine de la métrologie, en particulier celle utilisée dans le GUM. C'est le cas en particulier des sections 3.2 à 3.4. Cela nous a amené à utiliser de façon très générique le terme d'« incertitude de mesure ». Cependant, si l'on souhaite expliquer comment l'approche traditionnelle se propose de formuler le résultat final, il faut s'extraire du compte-rendu du GUM qui *n'est pas* une exposition de l'approche traditionnelle, puisqu'il a justement été rédigé pour venir amender cette dernière sur certains points, en particulier le traitement des erreurs systématiques et la formulation des résultats de mesure. Pour trouver une image cohérente et complète de la façon dont l'approche traditionnelle conçoit l'expression d'un résultat de mesure, nous avons choisi de nous référer une fois de plus au texte phare d'Eisenhart⁷⁵. Comme nous l'avons mentionné, Eisenhart s'appuie sur deux notions clés, qui apparaissent dès le titre : *précision et exactitude*.

Avant d'entrer plus en détail dans la façon dont Eisenhart explicite ces deux termes, il nous faut ici émettre une remarque essentielle pour la suite de notre analyse. Le terme d'« incertitude de mesure » peut être compris de deux façons. Il est d'abord un terme *générique*, qui vient désigner la façon dont on évalue un résultat de mesure, et qui peut correspondre au concept mobilisé dans l'approche traditionnelle que nous décrivons ici, ou dans l'approche bayésienne que nous décrivons par la suite. Mais cette acception générique doit être distinguée de l'interprétation contemporaine qui en est fait depuis quelques décennies. Bien que le terme d'incertitude ait été utilisé çà et là dans les textes scientifiques Le terme d'incertitude de mesure avant les années 1970, il n'était pas pour autant un terme consacré, et le déploiement généralisé du terme a coïncidé avec le développement d'une interprétation épistémique, comme représentation d'un état de connaissance, que nous décrivons dans le prochain chapitre. Par conséquent, parler d'« incertitude de mesure » introduit une ambiguïté, en ce que l'on pourrait se rapporter tout aussi bien au terme générique qu'à l'acception contemporaine. Jusqu'ici, nous avons employé le terme dans son sens générique, parce que notre étude s'est appuyée sur des travaux récents, en particulier le GUM, qui font usage de ce terme, et ce même lorsqu'il s'agit de désigner des concepts de l'approche traditionnelle. Ce mélange des genres n'était pas souhaitable en principe, mais nous avons préféré nous y tenir pour ne pas introduire trop de distance avec l'usage actuel des métrologues.

Chez Eisenhart, le terme d'incertitude n'apparaît pas. Ce dernier décrit un processus de mesure de la façon suivante :

74. Joint Committee for Guides in Metrology (JCGM) (2008d), p.vii.

75. Eisenhart (1963)

Un processus de mesure est la réalisation d'une méthode de mesure en termes d'appareils et d'équipements particuliers d'un type prescrit, de conditions particulières qui, aux mieux, ne font qu'approximer les conditions prescrites, et de personnes particulières comme opérateurs et observateurs.⁷⁶

Dès lors, lorsqu'une mesure est menée et qu'elle est considérée comme utilisable,

(l) peut encore demeurer la question de savoir s'il est fidèle à la méthode de mesure dont il est destiné à être la réalisation.⁷⁷

Une première notion, la *précision*, apporte une information à ce propos :

La précision, ou, plus correctement, l'imprécision d'un processus de mesure est habituellement résumée par l'écart-type du processus, qui exprime le désaccord caractéristique entre des mesures répétées d'une quantité donnée par le processus en question, et qui, de ce fait, sert à indiquer à quel point une mesure particulière est susceptible de différer d'autres valeurs que le même processus de mesure aurait pu avoir fournies dans ce cas précis, ou pourrait donner, à une autre occasion, lors d'une nouvelle mesure de la même quantité.⁷⁸

L'analyse que donne Eisenhart de la précision du processus de mesure, calculée à partir d'une variance statistique, peut être ramenée sans difficulté à l'analyse des erreurs aléatoires que nous avons proposée aux sections 3.1 à 3.4. Toutefois, spécifier la précision du processus de mesure ne suffit pas : pour une description complète des limites de ce dernier, il est souhaitable de le caractériser plutôt par un constat plus complet d'*exactitude*. Or, on aboutit ici à une difficulté car, contrairement à la précision (désormais appelée fidélité dans le vocabulaire de la métrologie⁷⁹), l'exactitude ne peut pas être exprimée sous une forme simple – typiquement, par une valeur unique. Comme l'exprime Eisenhart,

Malheureusement, il n'existe pas de mesure complète de l'exactitude (ou *inexactitude*) d'un processus de mesure, analogue à la mesure de son imprécision par l'écart-type.⁸⁰

Par conséquent, un énoncé d'exactitude nécessite au moins deux composantes bien distinctes :

(L)'exactitude des [...] résultats finaux peut ordinairement être caractérisée de façon satisfaisante en indiquant (a) leur imprécision, telle qu'elle est exprimée par leurs écarts-types et (b) l'erreur systématique du processus par lequel elles ont été obtenues.⁸¹

Les composantes (a) et (b) renvoient respectivement au traitement des erreurs aléatoires et à celui des erreurs systématiques. Il y a donc deux composantes différentes, auxquelles correspondent deux modes de calcul différents. Un peu plus loin, Eisenhart insiste de nouveau :

76. Eisenhart (1963), p.161.

77. Eisenhart (1963), p.162.

78. Eisenhart (1963), p.162 (Eisenhart souligne).

79. Joint Committee for Guides in Metrology (JCGM) (2012), p.22.

80. Eisenhart (1963), p.162 (Eisenhart souligne).

81. Eisenhart (1963), p.162.

(O)n est amené par la force de la nécessité à la conclusion inévitable qu'ordinairement (au moins) deux nombres sont nécessaires pour caractériser adéquatement l'exactitude d'un processus de mesure. Et cela a été reconnu par l'American Society for Testing and Materials dans leurs recommandations récentes⁸²

L'affirmation d'Eisenhart repose sur un constat classique et bien connu des métrologues à cette époque : les composantes d'exactitude relatives aux erreurs aléatoires et systématiques ne sont pas de même nature mathématique, et il n'est pas possible de les réunir en une seule composante d'ensemble. Jusqu'au début des années 1970, ce constat est assumé sans réserve par un certain nombre de métrologues, et la position d'Eisenhart n'est pas isolée : on la retrouve par exemple dans les recommandations émises par le NPL (National Physical Laboratory, Londres) dans un fascicule destiné à la communauté des métrologues :

(I)l est nécessaire de conserver les deux catégories d'incertitude séparées, non seulement dans la combinaison progressive des incertitudes tout au long d'une expérience complexe, mais aussi dans la formulation du résultat final.⁸³

Dans l'approche traditionnelle, le caractère composite et irréductible d'un constat d'exactitude est quelque chose d'inévitable, mais ce n'est pas pour autant problématique. À l'usage, cependant, cela peut devenir délicat. C'est précisément à ce sujet que vont être émises certaines des objections les plus notables.

3.6.2 Limites de l'approche fréquentiste traditionnelle

Comme le montre notre analyse, l'expression fréquentiste traditionnelle d'un résultat de mesure comporte deux composantes bien distinctes qui ne peuvent pas être interprétées de la même façon – l'une est probabiliste, l'autre non – et ne peuvent pas plus être combinées en une résultante unique. Cela se révèle être à la source de l'une des critiques les plus notoires qui lui a été adressée. Dans les années 1970, l'obligation de séparer deux composantes dans l'évaluation de l'exactitude d'un résultat de mesure apparaît de plus en plus comme une limite à la fois conceptuelle et pratique. Conceptuelle, car il n'est pas très satisfaisant de devoir juxtaposer deux termes d'exactitude calculés et interprétés différemment, et auxquels correspondent deux modes différents de « propagation » des erreurs, alors qu'ils portent sur deux instances d'un même dispositif de mesure. Pratique, car elle est liée à l'*usage* auquel est destiné le résultat de mesure. De fait, la question de la combinaison, ou propagation, des incertitudes, met en évidence deux perspectives différentes sur le résultat de mesure qui correspondent à deux exigences qui peuvent se révéler antagonistes.

Selon une première conception, que l'on pourrait appeler « point de vue de l'expérimentateur » (ou du fournisseur), un résultat de mesure est d'autant plus riche, compréhensible, utilisable, et fidèle aux conditions d'expérimentation, qu'il contient un maximum de détails

82. Eisenhart (1963), p.180.

83. *Campion, Burns et Williams* (1973), p.2. Notons que *Campion et al.* se rallient à l'usage grandissant du terme « incertitude », alors même que leur conception reste très proche de l'approche traditionnelle. Ici, « incertitude » doit être compris dans son acception générique, et peut directement se rapporter à ce qu'Eisenhart désigne par « exactitude ».

quant aux conditions de prise de mesure, aux différentes sources d'erreur identifiées et corrigées, aux méthodes mathématiques utilisées pour les décrire, etc. Selon cette perspective, transmettre un résultat le plus complet possible, et donc le plus exhaustif possible, permet de limiter la perte d'information qu'entraîne automatiquement le fait de transmettre un résultat le long d'une chaîne de communication. Exiger d'adjoindre un maximum d'informations au résultat de mesure garantit que celui-ci puisse être utilisé dans les meilleures conditions. Dès lors, le caractère composite de l'incertitude de mesure, décrit précédemment et énoncé par exemple par Eisenhart ou dans le fascicule du NPL, n'est pas une limite de l'approche fréquentiste⁸⁴. Il est même conforme aux attentes et aux pratiques de la catégorie d'utilisateurs concernée, des scientifiques le plus souvent.

À l'inverse, cette décomposition est préjudiciable pour d'autres catégories d'utilisateurs, ingénieurs et industriels par exemple, qui souhaitent faire usage du résultat tel un « produit fini » sans avoir à revenir de façon critique sur les conditions de son obtention. La communication d'un résultat sous une forme adaptée à leurs besoins répond à des impératifs techniques et économiques en particulier. En effet, selon une seconde perspective que l'on pourrait appeler « point de vue de l'utilisateur » (ou du client), un résultat de mesure est un produit que l'on peut utiliser de façon simple et immédiate selon une procédure standardisée. La valeur indiquée peut être comparée à un seuil, par exemple une norme légale ou une exigence industrielle. Mais cette comparaison doit prendre en compte l'incertitude de mesure, par exemple sous forme d'un test statistique ; et il faut donc, pour que l'opération puisse être effectuée, que l'incertitude de mesure se présente sous la forme d'une composante unique. L'obtention d'une incertitude unique nécessite de fonder un tel calcul d'incertitude sur des bases statistiques solides. Or, c'est précisément cet aspect qui se révèle le plus difficile à obtenir, car il est le plus inhomogène dans les pratiques des scientifiques. Dès lors, la rigidité du modèle fréquentiste apparaît dans ce cadre – et dans ce cadre uniquement – comme un écueil sérieux. On voit ici que c'est relativement à un *usage* auquel est destiné le résultat que sont émises les exigences concernant l'expression des résultats de mesure⁸⁵.

Ce n'est qu'à la fin des années 1970 que la question des limites de l'approche traditionnelle

84. Il est néanmoins à noter que, dès 1929, Birge plaidait pour une formulation exclusivement probabiliste des composantes d'erreur. « Dans la plupart des travaux expérimentaux, en addition à l'erreur probable due aux erreurs purement accidentelles, pour lesquelles les moindres carrés fonctionnent correctement, il y a de nombreuses autres sources d'erreur, constantes ou systématiques. Certaines d'entre elles peuvent être connues, d'autres peuvent être tout à fait insoupçonnées. L'enquêteur tente de faire une estimation de l'amplitude de chacune de ces erreurs, et dans ce cas aussi, il semble grandement préférable de donner ce qui semble être l'erreur probable », Birge (1929b), p.6. Cela témoigne du fait qu'il est impossible d'identifier alors une même approche qui soit communément utilisée par tous les scientifiques. La multiplicité des pratiques métrologiques a renforcé la motivation des métrologues pour les réformes amorcées à la fin du XX^e siècle.

85. En témoigne cet extrait du GUM : « en pratique, la quantité d'information nécessaire pour documenter un résultat de mesure dépend de l'usage prévu », Joint Committee for Guides in Metrology (JCGM) (2008d), p.26. En 1980, le BIPM indiquait aussi : « Il semble qu'un accord peut être atteint à l'égard des mesures de très haute précision, pour lesquelles il est toujours désirable de donner autant d'informations que possible à propos de l'estimation des incertitudes. D'un autre côté, il faut souvent, dans l'usage courant, caractériser l'incertitude ou la tolérance par un paramètre unique (ou si besoin par un très petit nombre de paramètres). C'est sur ce problème que nous espérons progresser. », Kaarls (1980), annexe 1, p.1.

devient débattue de façon pressante⁸⁶. Les métrologues émettent le souhait d'uniformiser les pratiques d'analyse d'incertitude à l'échelle internationale, et le BIPM réunit à cet effet un groupe de travail dont le rôle est de formuler un consensus quant aux principes généraux à adopter. Le groupe de travail envoie un questionnaire aux laboratoires nationaux de métrologie afin de consulter leurs avis et de se renseigner sur la diversité des pratiques existantes. Ce questionnaire fait l'objet d'un rapport, où l'on trouve le commentaire suivant :

Commentaires du BIPM : La croyance auparavant très populaire que toute combinaison des types [d'erreur] doit être évitée semble changer. L'attitude nouvelle et plus réaliste est sans nul doute inspirée par les besoins pratiques. Bien qu'il existe désormais probablement une majorité qui est prête à accepter une règle pour la combinaison des incertitudes aléatoires et systématiques, aucun consensus sur la façon dont cela devrait être accompli en détail n'est encore clairement visible.⁸⁷

De fait, le groupe de travail s'accorde sur une exigence commune : il est souhaitable que les techniques statistiques d'analyse d'incertitude soient compatibles avec l'expression naturelle d'une incertitude prenant la forme d'une composante unique. Cette exigence est officialisée dans une recommandation internationale très succincte, la recommandation « INC-1 »⁸⁸, émise par le BIPM en 1980. Sous l'impulsion de cette recommandation, la question de la combinaison des composantes d'incertitudes, déjà activement discutée depuis le début de la décennie, devient un problème essentiel – et la conception fréquentiste traditionnelle n'est alors plus satisfaisante. Le GUM sera ensuite le produit de ce choix commun : il viendra compléter les fondements jetés dans la recommandation INC-1 pour fournir une réponse détaillée à l'exigence d'une résultante unique d'incertitude. Nous renvoyons au chapitre 12 pour une description de l'histoire de la conception du GUM, et de quelques enjeux épistémologiques et institutionnels qui y sont attachés.

Des principes *ad hoc* ont été proposés pour résoudre l'impossibilité de principe de combiner les composantes d'incertitude. En 1953, Cohen et DuMond convertissaient les « limites d'erreurs » en variances statistiques par une méthode loin d'être exempte d'arbitraire⁸⁹. L'usage des termes hybrides « incertitude aléatoire » et « incertitude systématique » s'est un temps im-

86. Il est intéressant de voir que selon les métrologues eux-mêmes, la question de l'évaluation des erreurs semblait considérée comme majoritairement réglée à la fin des années 1960. Ainsi peut-on lire dans le compte-rendu du BIPM de 1980 la remarque suivante : « Il n'y a pas beaucoup de doute que le sujet général de la façon d'exprimer les incertitudes expérimentales est passé au centre de l'attention au cours de ces quelques dernières années. Ce regain d'intérêt est frappant pour ceux qui se souviennent de l'état des choses il y a disons dix ans, lorsque tout semblait avoir été réglé pour toujours », Kaarls (1980), p.12. Nous renvoyons au chapitre 12 pour une narration de la façon dont les différents écueils de l'approche traditionnels ont été perçus par les métrologues à la fin des années 1970, et comment cela a fini par aboutir à la création du GUM.

87. Kaarls (1980), p.7.

88. Kaarls (1980), version anglaise p.13 et version française p.14. Voir aussi Giacomo (1981), pp.73–74, ainsi que Joint Committee for Guides in Metrology (JCGM) (2008d), p.30.

89. Ce cas se présente lorsque Cohen et DuMond recensent un résultat formulé en termes de limites d'erreurs : « (Q)uand les expérimentateurs rapportent des "limites d'erreurs", nous avons suivi la règle consistant à diviser celles-ci par deux pour obtenir l'erreur type », DuMond et Cohen (1953), p.699. Les auteurs ne donnent pas de justification précise à ce choix mais font comprendre qu'une telle opération leur est indispensable, et qu'ils n'ont pas de meilleur choix à opposer parce que l'auteur de l'article source a péché par manque d'information.

posé⁹⁰ dans le courant des années 1970. Dans ce cas, un résultat de mesure était formulé de la façon suivante :

$$X = x \pm u_{\text{al}}(x) \pm u_{\text{sys}}(x) [X] \quad (3.19)$$

où $[X]$ est l'unité de mesure employée. Dans la continuité, des méthodes ad hoc suggéraient alors une combinaison quadratique des différentes composantes :

$$u_{\text{totale}}^2 = u_{\text{al}}^2 + u_{\text{sys}}^2 \quad (3.20)$$

Nous pouvons illustrer cette étape par un exemple récent, qui révèle d'ailleurs que cette pratique n'est pas totalement abandonnée. En 2011, le laboratoire national du Gran Sasso (Italie) menait une expérience, l'expérience OPERA, dont les premiers résultats semblaient démontrer l'existence de « neutrinos supraluminiques », c'est-à-dire dont la vitesse est supérieure à la célérité de la lumière dans le vide. Ce résultat a immédiatement trouvé un écho assez retentissant, en particulier dans la presse généraliste, parce qu'il venait contredire la théorie de la relativité (nous revenons sur cette exemple au chapitre 11). Dans l'article publié initialement par l'équipe, le résultat final est présenté de la façon suivante :

La différence relative de la vitesse du neutrino muonique avec la vitesse de la lumière est :

$$(v - c)/c = \delta t / (TOF'c - \delta t) = (2.48 \pm 0.28(\text{stat.}) \pm 0.30(\text{sys.})) \times 10^{-5}.$$

avec une significativité globale de 6.0σ .⁹¹

On retrouve les deux composantes d'incertitude, statistique (aléatoire) et systématique, exprimées séparément ; mais la « signification statistique globale de 6 sigma » n'a pu être calculée qu'en combinant ces deux types d'incertitude : la propagation quadratique $0,28^2 + 0,30^2$ des deux composantes d'incertitude aboutit à une incertitude relative résultante de 0,41 – effectivement 6,04 fois inférieure à l'amplitude relative de l'effet, qui a été estimée à 2,48.

Cependant, comme l'ont souligné à leur tour différents métrologues dont Eisenhart⁹² puis Campion, Burns et Williams⁹³, cette opération ne trouve pas de justification formelle. Pour

90. L'expression a été suggérée par Campion, Burns et Williams en 1973 dans un article de la revue *Metrologia* (Burns, Campion et Williams, 1973) et a été reprise par les auteurs dans le fascicule qu'ils ont rédigé pour le NPL (Campion, Burns et Williams, 1973). Nous renvoyons à Protassov (2002), pp.115-116, pour une explicitation plus détaillée de certains aspects plus techniques relatifs à cette terminologie. Cette double appellation semble avoir été surtout utilisée pendant une décennie environ, de 1970 à 1980, et a été régulièrement critiquée comme étant impropre, le caractère aléatoire ou systématique étant relatif aux erreurs correspondantes, non aux incertitudes elles-mêmes (Kaarls, 1980, annexe V, p.3 ; Taylor et Kuyatt, 1994, p.16.). Burns, Campion et Williams reconnaissent d'ailleurs dès le départ le « léger illogisme sémantique » de l'expression, Burns, Campion et Williams (1973), p.101.

91. Adam, T. *et al.* (2011), p.22.

92. « Je réalise que certaines des incertitudes estimées subjectivement et tabulées par les expérimentateurs sont destinées, lorsqu'elles sont combinées “en quadrature” avec l'erreur type expérimentale de la moyenne [*statistical standard error of the mean*], de fournir une “erreur type” amplifiée de la moyenne qui peut être utilisée pour “fournir une indication de la proximité avec laquelle le résultat de cette expérience est susceptible de s'accorder avec d'autres résultats qui pourraient être obtenus si l'expérience entière était réalisée de nouveau ‘depuis le début’” ; mais l'expérience montre que c'est une illusion : il n'y a pas de substitut à une répétition complète et réaliste de l'expérience entière ! », Eisenhart (1971), p.516.

93. « La combinaison avec l'incertitude systématique pour donner l'incertitude totale est obsolète », Campion, Burns et Williams (1973), p.12.

dépasser les limites que proposaient l'approche fréquentiste traditionnelle en regard de l'exigence d'une composante unique, guidée par la pratique, certains métrologues ont fait progressivement émerger l'idée que les probabilités *épistémiques*, comprises en tant que mesures de degrés de croyances attachés à un sujet connaissant, étaient les plus à même de servir d'outil mathématique le plus adapté aux besoins de la métrologie. Cette émergence s'est accompagnée d'une remise en question des principes épistémologiques qui guidaient la pratique de la mesure depuis plus d'un demi-siècle. Le caractère subjectif d'un résultat de mesure a ainsi été signalé, à l'encontre de l'intuition classique selon laquelle elle constitue l'une des sources les plus objectives de données sur le monde physique. Le GUM, publié à la suite de la recommandation INC-1, est le réceptacle de ces nouvelles méthodes, que nous décrivons dans le chapitre suivant.

Chapitre 4

Des probabilités épistémiques à une modélisation bayésienne de la mesure

Dans le chapitre précédent, nous avons vu comment construire une approche fréquentiste fonctionnelle de la mesure. Nous avons montré l'usage qui y est fait des probabilités, et insisté sur le fait que celles-ci portent non pas sur le résultat de mesure mais sur le processus de mesure. Les probabilités prennent alors pour objet les caractéristiques d'un dispositif physique, ce qui correspond à l'interprétation objectiviste qu'on leur octroie généralement. Notre description du modèle fréquentiste des erreurs aléatoires nous a permis de comprendre comment, en l'absence d'informations permettant de corriger les erreurs de mesure commises, une analyse statistique demeure possible pour réduire ces erreurs et agréger de la meilleure façon possible l'information contenue dans les résultats de mesurages répétés de la grandeur dans les conditions dites de répétabilité. Cependant, nous avons également vu que l'évolution plutôt récente des objectifs et des usages dans la métrologie de la seconde moitié du XX^e siècle a mené les métrologues à pointer du doigt certaines limites de l'approche fréquentiste, la principale d'entre elles étant l'impossibilité de formuler une incertitude de mesure au travers d'une unique valeur. En effet, l'approche fréquentiste traditionnelle ne permet pas de combiner les différentes composantes d'incertitude autrement que par des méthodes *ad hoc* mal justifiées et peu satisfaisantes sur le plan théorique. Cela est dû au fait que la nature des composantes diffère de l'une à l'autre, certaines étant probabilistes, les autres non. L'approche fréquentiste avait jusqu'alors prospéré car elle répondait de façon cohérente à une difficulté épistémologique et méthodologique sérieuse, celle posée par la variabilité des résultats d'une mesure répétée. En ce sens, l'histoire de l'approche fréquentiste de la mesure est d'abord l'histoire d'un succès majeur.

Cependant, les progrès scientifiques, ceux des statistiques entre autres choses, ont poussé les métrologues à chercher à raffiner encore plus les méthodes employées. De fait, les métrologues ont finalement perçu l'impossibilité de combiner les composantes d'incertitude comme une limite suffisamment critique de l'approche fréquentiste traditionnelle pour chercher activement à en trouver des méthodes de remplacement. Ils ont trouvé dans les probabilités épistémiques, puis dans les statistiques bayésiennes, une alternative qui n'a, depuis, cessé de prendre

de l'ampleur dans la communauté métrologique, jusqu'à devenir l'un des fondements du GUM, pour peut-être, dans un avenir proche, devenir le principal mode d'analyse d'incertitude en métrologie.

Notre objectif, dans ce chapitre, est de décrire la façon dont les probabilités épistémiques permettent de rendre compte de l'incertitude de mesure, puis de montrer comment les statistiques bayésiennes deviennent par la suite un ressort du calcul de l'incertitude. Nous nous appuyons sur les développements proposés dans le GUM et dans ses suppléments, ainsi que sur l'abondance littéraire qui a été proposée dans ce domaine, en particulier depuis le début des années 1990. Nous verrons en particulier que ces approches se caractérisent par un changement d'objet des probabilités, celles-ci ne portant plus sur un dispositif physique, le processus de mesure, mais sur un état de connaissance relatif au sujet connaissant – les probabilités étant alors rangées dans la catégorie des probabilités subjectives.

Nous commencerons à la section 4.1 par une rapide description de ce qu'est l'interprétation épistémique des probabilités, ce qui nous permettra de clarifier le cadre conceptuel qui est le nôtre.

Nous reviendrons, dans un second temps, sur les limites de l'approche fréquentiste, et la façon dont ces limites ont mené à la mise en avant d'une interprétation épistémique des probabilités. C'est l'objet de la section 4.2.

Dans la section 4.3, nous décrirons ensuite la façon dont sont développées les méthodes dites « de type B » dans le GUM, c'est-à-dire les méthodes consacrées à l'analyse d'incertitude dans les cas où l'information ne provient pas du résultat de l'expérience menée elle-même, mais de sources extérieures. Nous verrons que ces méthodes peuvent être interprétées comme un traitement bayésien des erreurs systématiques de mesure, même si elles ne sont pas présentées comme telles dans le GUM lui-même.

La section 4.4 conclut sur l'approche du GUM en montrant comment celui-ci opère la propagation des incertitudes issues des différentes méthodes, de « type A » et de « type B », et résout ainsi le problème central de l'approche fréquentiste, à savoir l'impossibilité de formuler une incertitude de mesure sous la forme d'une composante unique. Cependant, cette résolution vient elle-même avec son lot de difficultés et nous verrons que la résolution du GUM n'est pas non plus pleinement satisfaisante sur le plan conceptuel.

Cela nous amènera alors à la section 4.5 où nous développerons une méthode, absente du GUM, mais exposée dans l'un de ses suppléments, lui-même publié en 2008, où l'emploi des probabilités épistémiques est étendue aux méthodes de « type A », c'est-à-dire au traitement de la variabilité des résultats de mesure – donc des erreurs aléatoires. Cette méthode attribue un rôle de pivot au « théorème de Bayes »¹ et peut, de ce fait, être considérée comme une

1. Le théorème de Bayes est une propriété générale des probabilités dites « conditionnelles », qui se démontre à partir des axiomes des probabilités et dont la validité mathématique est indépendante de l'interprétation accordée aux probabilités. Il tire son nom du pasteur presbytérien Thomas Bayes (1702–1761) qui en a développé une première forme dans son essai posthume *Essais sur la manière de résoudre un problème dans la doctrine des risques* (Bayes, 1763), avant que Laplace ne le redémontre indépendamment en 1812 (Laplace, 1812). Bien que sa validité soit très générale, le théorème constitue plus spécifiquement l'un des fondements des statistiques bayésiennes en ce qu'il permet d'exprimer une probabilité sur un paramètre fixe inconnu (au moyen de ce qui prend la forme d'une « inversion de probabilités », voir à ce sujet Stigler, 1986, chapitre 3 : « Inverse Probability, pp.99–138), et qu'il permet par

ébauche d'approche pleinement bayésienne de la mesure.

4.1 Les probabilités épistémiques

Selon l'interprétation épistémique des probabilités, les probabilités ne décrivent pas un état du monde physique, mais un degré de croyance entretenu par un agent à propos du monde physique. L'exemple classique du tirage « à pile ou face » vient illustrer la comparaison de cette interprétation avec l'interprétation fréquentiste. Selon l'interprétation fréquentiste, la pièce présente une tendance à tomber côté pile (respectivement côté face) avec une certaine fréquence limite, qui dépend de l'état physique de la pièce ainsi que du dispositif de lancer. La probabilité de l'issue « pile » s'identifie alors à la fréquence limite correspondante. Le cas classique est celui de la pièce équilibrée : si les résultats « pile » et « face » surviennent dans les mêmes proportions, leur probabilité commune est $1/2$. Selon l'interprétation épistémique des probabilités, les probabilités ne décrivent pas des fréquences limites mais des états de croyance. La question n'est donc pas d'évaluer comment le dispositif a tendance à se comporter, mais de décrire ce que sait l'agent vis-à-vis des différentes issues possibles. Ainsi, un agent attribuera des probabilités égales à « pile » et « face » s'il ne dispose d'aucune information lui permettant de privilégier une issue sur l'autre.

Du fait de leur signification différente, les probabilités épistémiques et fréquentistes ne concordent pas toujours en toute situation. Un second exemple vient illustrer ce fait. Supposons que l'on dispose de deux urnes, l'une contenant deux boules bleues et une boule rouge, l'autre contenant deux boules rouges et une boule bleue (figure 4.1). Si l'on procède au tirage aléatoire non biaisé d'une boule dans la première urne, la probabilité fréquentiste d'obtenir une boule bleue est de $2/3$. Si l'agent connaît la composition de l'urne, la probabilité épistémique d'obtenir une boule bleue est également de $2/3$. Le même raisonnement, inversé, vaut pour la seconde urne. Supposons maintenant que l'on procède de la façon suivante. On tire d'abord au hasard une urne, *sans savoir laquelle*, que l'on conserve pour la suite de l'expérience. On procède alors au tirage aléatoire répété d'une boule dans cette même urne, en remplaçant à chaque fois la boule dans l'urne après tirage. Dans ce cas, on voit apparaître une différence concrète entre les deux types de probabilités. En effet, si l'on adopte une interprétation fréquentiste des probabilités, la fréquence limite de tirage d'une boule bleue dépendra de l'urne choisie, et pourra être soit $1/3$, soit $2/3$, sans autre option possible. La probabilité est alors déterminée par un état physique, à savoir l'identité de l'urne qui a été sélectionnée au hasard ; et cette probabilité est *indépendante* de l'état de connaissance de l'expérimentateur. Selon l'interprétation épistémique des probabilités, la probabilité doit quantifier le degré de croyance de l'expérimentateur relativement à la possibilité de tirer l'une des boules. Or, ne sachant pas quelle urne a été préalablement tirée, l'expérimentateur ne dispose d'aucune connaissance lui permettant de privilégier une issue sur l'autre. Ainsi, la probabilité de tirer une boule bleue est de $1/2$, *indépendamment* de l'identité effective de l'urne. Cette probabilité est susceptible de varier

suite d'exprimer la probabilité comme un degré de croyance que l'on peut réviser à chaque nouvelle information disponible à propos du paramètre étudié. Nous aurons par la suite l'occasion de développer plus en détails la façon dont ce théorème fonctionne.

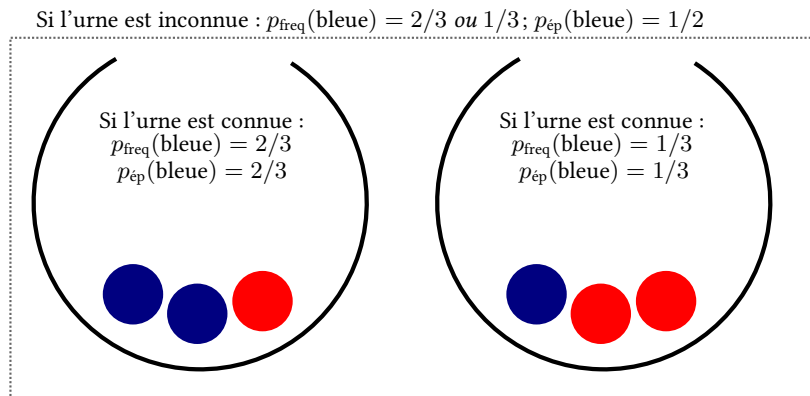


FIGURE 4.1 – Illustration de la différence entre probabilité fréquentiste et probabilité épistémique sur un modèle d'urnes. On dispose de deux urnes, l'une contenant deux boules rouges (claires) et une boule bleue (foncée), l'autre contenant deux boules bleues et une boule rouge. On sélectionne d'abord au hasard l'une des urnes, *sans que l'on sache de laquelle il s'agit*. Ensuite, une fois l'urne désignée, on procède au tirage aléatoire répétée d'une boule *en conservant à chaque fois la même urne*, et en remplaçant la boule dans l'urne après tirage. Selon l'interprétation fréquentiste des probabilités, la probabilité de tirer une boule rouge est alors soit $2/3$, soit $1/3$, selon l'urne choisie, mais ce ne peut être que l'une ou l'autre. La probabilité fréquentiste est ainsi déterminée par un état physique (l'urne qui a été tirée au sort) et est entièrement indépendante de l'état de connaissance dont on dispose sur le système. Notons que cette probabilité demeure inconnue tant que l'on n'a pas identifié l'urne qui a été préalablement choisie, ou que l'on n'a pas procédé à sa mesure expérimentale (en répétant les tirages un grand nombre de fois, ou en regardant le contenu de l'urne). Selon l'interprétation épistémique des probabilités, la probabilité de tirer une boule rouge, lors du premier tirage, est de $1/2$: en effet, on ne sait pas quelle urne a été choisie au préalable et rien, dans les connaissances disponibles, ne permet de favoriser une issue sur l'autre. La probabilité épistémique est de ce fait déterminée par un état de connaissance, et – pour ce qui est de la probabilité avant le premier tirage – est *indépendante* de l'identité de l'urne qui a été effectivement choisie. Cette probabilité peut ensuite être révisée au fur et à mesure des tirages, par révision bayésienne des connaissances (en utilisant le théorème de Bayes, décrit plus loin dans ce chapitre). La probabilité épistémique tendra d'ailleurs vers la probabilité fréquentiste à mesure que le nombre de tirages tend vers l'infini car l'analyse des résultats du tirage permettra de déterminer avec de plus en plus de certitude l'identité de l'urne qui a été sélectionnée au départ.

si la connaissance de l'expérimentateur varie, par exemple si l'identité de l'urne est dévoilée, mais aussi si l'expérimentateur cherche à inférer l'identité de l'urne à partir d'une analyse des données expérimentales (par exemple en employant un processus de révision bayésienne des connaissances, que nous décrivons plus loin, à la section 4.5).

Dans sa formulation classique, cette vision de la probabilité renvoie au « démon » de Laplace : un être omniscient et doué d'une capacité illimitée de calcul pourrait, en vertu du déterminisme, prédire avec certitude l'issue de chaque lancer de pièces et n'aurait aucune raison de parler en termes probabilistes. C'est seulement son manque de connaissance qui rend nécessaire une description probabiliste. Les formulations modernes de l'interprétation épistémique visent à formaliser la notion de « degré de croyance » et à exprimer rigoureusement leur rapport aux probabilités. Celles-ci sont généralement regroupées sous le nom d'approches bayésiennes. Comme l'indique Isabelle Drouet², le bayésianisme est fondé sur plusieurs thèses dont les deux suivantes : d'une part, les croyances viennent par degrés, c'est-à-dire qu'elles sont quantifiables ; d'autre part, pour un agent *rationnel*, les degrés de croyances doivent vérifier les principes mathématiques des probabilités, c'est-à-dire l'axiomatique de Kolmogorov. D'autres thèses supplémentaires viennent s'ajouter, régissant en particulier la dynamique des croyances et de leur révision³ (typiquement par le théorème de Bayes). C'est autour de l'idée de pari que l'on peut alors chercher à quantifier les degrés de croyance d'un agent, en les mesurant par sa propension à s'engager ou non dans un pari en fonction des gains possibles. Ainsi, si un agent rationnel est prêt à parier 1 euro sur le fait qu'une pièce, après lancer, retombera côté « face » visible, pour un gain net de 1 euro en cas de succès, c'est que son degré de croyance envers cette issue est au moins égal à 50%. L'idée que les degrés rationnels de croyance doivent respecter les axiomes des probabilités repose alors sur le fait que si un agent tient des croyances qui ne suivent pas de tels axiomes, il devient possible de l'engager dans des paris qu'il est certain de perdre – et qui ont été baptisés « paris hollandais »⁴ ; dans ce cas, la croyance ne peut pas légitimement être tenue pour rationnelle.

On parle souvent dans ce cas de probabilités « subjectives » puisque ces probabilités sont relatives au sujet connaissant – par opposition aux probabilités « objectives » (par exemple fréquentistes) dont le domaine est celui des phénomènes physiques. Cependant, le terme de « probabilités subjectives » est délicat, plus encore que celui de « probabilités objectives ». Dire d'une probabilité qu'elle est « subjective », c'est considérer que celle-ci est relative à un individu particulier, et c'est donc indiquer que chaque sujet connaissant est susceptible de tenir, à propos d'une proposition donnée, une probabilité qui lui est propre. On conçoit aisément en quoi une conception de la probabilité comme degré de croyance pourrait être immédiatement associée à une nature subjective (et donc individuelle) ; cependant, une telle association ne doit pas être automatique. En effet, il demeure possible de concevoir un état de connaissance autrement que comme une expression purement subjective ; c'est le cas en particulier si l'on considère qu'il n'y a qu'une seule expression *rationnelle* d'une somme d'informations accessibles à une communauté, et qu'il n'y a alors pas de subjectivité à l'œuvre dans la formulation d'une connaissance.

2. Drouet (2016), à paraître.

3. Drouet (2016).

4. Drouet (2016). Hájek (2012), section 3.3.2 (entrée consultée le 5 janvier 2016).

En particulier, certaines conceptions de la probabilité comme état de connaissance se veulent objectives, et s'inspirent par exemple de la théorie de Keynes, articulée autour d'un « principe d'indifférence »⁵, ou de celle de Jaynes, mettant en jeu le principe d'« entropie maximum »⁶ (qui peut être vu comme une généralisation du premier⁷). Pour éviter l'écueil qu'il y aurait à décrire toutes ces interprétations sous un même qualificatif de « subjectif », Hacking a proposé une terminologie qui semble avoir atteint le consensus, celle de « probabilités épistémiques », laquelle recouvre des acceptions subjectivistes ainsi que des conceptions intersubjectives ou objectives de la probabilité. Pour résumer, une probabilité épistémique est une probabilité qui est susceptible de varier lorsqu'un état de connaissance varie, alors même que l'état du système étudié lui-même reste identique – la probabilité dépend de l'état des connaissances, et décrit cet état de connaissance. Une probabilité subjective est une probabilité qui est susceptible de varier d'un individu à l'autre. Les deux acceptions ne se recouvrent pas automatiquement. Pour la suite de notre analyse, le terme de « probabilité épistémique » est celui que nous conserverons car il présente une plus grande généralité.

4.2 L'approche du GUM et l'introduction des probabilités épistémiques

Le succès des probabilités fréquentistes dans les méthodes d'analyse des collections de données expérimentales tient au fait qu'elles semblent parfaitement adaptées au traitement de la variabilité : les échantillons de résultats de mesure qui sont analysés par le spectre des probabilités fréquentistes fournissent une façon naturelle de réfléchir en termes de populations statistiques. Mais en contrepartie, ces probabilités n'offrent pas la possibilité de saisir les composantes dites « systématiques » d'incertitude. C'est là le principal point d'achoppement de l'approche traditionnelle, qui devient critique si l'on se reporte aux exigences formulées dans la recommandation INC-1. Les critiques de l'approche n'ont pas manqué de le faire remarquer ; ainsi Bich note-t-il que :

Les erreurs aléatoires sont considérées comme une population (typiquement gaussienne), à laquelle les outils statistiques classiques peuvent être appliqués. Par exemple, la variance V d'une population est généralement évaluée par la variance s^2 d'un échantillon. Cette procédure ne peut pas être utilisée pour les incertitudes liées à des biais, car il n'y a pas de population derrière une erreur systématique et, par conséquent, il n'y a pas d'échantillon derrière un biais.⁸

C'est en grande partie pour répondre à ce problème que le BIPM, au début des années 1980, a décidé la rédaction d'un document de référence, devenu plus tard le *Guide pour l'expression de l'incertitude de mesure* (GUM). La conception du GUM est orientée autour d'un objectif central, celui de traiter toutes les composantes d'incertitude de façon probabiliste, de façon à permettre la propagation naturelle de toutes les composantes en une résultante unique. Cela a été un peu

5. Gillies (2000), pp.25–49.

6. Jaynes (1957)

7. Hájek (2012), section 3.1 (entrée consultée le 5 janvier 2016).

8. Bich (2012a), p.2156.

rapidement assimilé au fait de traiter de façon *identique* les différentes composantes d'incertitude. Ainsi peut-on lire dans le GUM la remarque suivante :

(C)e Guide traite exactement de la même façon les composantes de l'incertitude provenant d'effets aléatoires et celles provenant des corrections estimées d'effets systématiques, lorsqu'on évalue l'incertitude du résultat d'un mesurage. [...] (T)outes les composantes de l'incertitude sont de même nature et doivent être traitées de manière identique.⁹

Malgré l'affirmation similaire d'Ehrlich, Dybkaer et Wöger : « les erreurs systématiques et aléatoires sont traitées sur la même base probabiliste »¹⁰, il convient en fait de nuancer ces prétentions. Comme nous le verrons, il n'est pas possible de dire, à la suite du GUM, que les composantes d'incertitude sont de « même nature » et qu'elles sont alors traitées sur la « même base ». Tout au plus sont-elles toutes deux formulées en termes probabilistes, ce qui constitue effectivement une évolution majeure par rapport à l'approche fréquentiste traditionnelle. Cependant, les probabilités employées sont, comme nous allons le voir, de natures différentes.

Les principes généraux du GUM sont directement nourris des choix émis en 1980 dans la recommandation INC-1. En particulier, le GUM reprend la catégorisation des méthodes d'évaluation d'incertitude selon deux types, les méthodes de « type A » et de « type B » (voir l'extrait de la recommandation INC-1 cité dans le chapitre précédent, à la section 2.4 p.65). La recommandation INC-1 n'impose aucune technique spécifique pour les méthodes de type B¹¹. Le choix effectué dans le GUM consiste essentiellement à reprendre l'approche traditionnelle pour le traitement des résultats répétés de mesure (méthodes de « type A »), et à implémenter une méthode fondée sur l'emploi des probabilités épistémiques pour ce qui est du traitement des autres composantes d'incertitude (méthodes de « type B »). Les soubassements interprétatifs de ce choix ne sont pas délimités très explicitement dans le corps du document. En effet, la question de l'interprétation des probabilités n'est véritablement introduite qu'en annexe :

(P)our certains traitements traditionnels de l'incertitude de mesure, [...] le concept de probabilité est considéré comme s'appliquant *seulement* aux événements qui peuvent être répétés un grand nombre de fois, essentiellement dans les mêmes conditions, la probabilité p d'un événement ($0 \leq p \leq 1$) indiquant la *fréquence* avec laquelle se produit l'événement. Par contraste avec ce point de vue de la probabilité fondée sur la fréquence, un autre point de vue aussi valable est que la probabilité est une mesure du *degré de croyance* en ce qu'un événement se produise. [...] La recommandation INC-1 (1980) sur laquelle est fondé ce *Guide* adopte implicitement une telle approche de la probabilité [...]¹²

9. Joint Committee for Guides in Metrology (JCGM) (2008d), p.57.

10. Ehrlich, Dybkaer et Wöger (2007), p.205.

11. La recommandation mentionne que « (l)es composantes de la catégorie B devraient être caractérisées par les variances estimées u_j^2 , qui peuvent être considérées comme des approximations des variances correspondantes dont on admet l'existence. », ce qui, sans donner pour autant une indication très claire de la façon dont ces variances sont calculées, laisse à penser qu'une méthode était déjà pressentie – en l'occurrence, il semble que nous pouvons assez raisonnablement supposer qu'il s'agit de celle qui a ensuite été développée dans le GUM.

12. Joint Committee for Guides in Metrology (JCGM) (2008d), p.59 (souligné dans le document).

Le GUM revendique ainsi une certaine neutralité interprétationnelle : aucune interprétation n'est considérée comme meilleure. Mais il entretient pourtant l'ambiguïté, d'autant qu'il exprime explicitement sa préférence envers *l'usage* des probabilités épistémiques. Cette ambiguïté nous semble être pour partie à la racine des discussions qui abondent à ce sujet depuis la publication du document. Par la suite, le mouvement opéré dans le GUM sera interprété par certains comme une ouverture vers l'adoption de statistiques bayésiennes, bien que le document lui-même n'y fasse jamais référence. Ainsi en atteste cette remarque de Lira et Wöger :

[Un effet systématique] constitue un élément pour lequel le traitement statistique classique échoue complètement. C'est principalement pour cette raison que le GUM a été écrit. Le GUM recommande d'évaluer les effets systématiques conformément à la procédure d'évaluation qu'on appelle « type B ». Bien que ce ne soit pas mentionné explicitement dans le corps principal du GUM, cette procédure est de nature essentiellement bayésienne.¹³

Dans la section suivante, nous explicitons les principes des méthodes dites de « type B » et la façon dont la résolution d'un problème pratique, l'expression de l'incertitude de mesure en une composante unique, a engendré l'adoption d'une interprétation épistémique des probabilités. Nous montrerons dans le même temps comment ce mouvement a été prolongé par une défense de l'usage des statistiques bayésiennes en métrologie.

4.3 Les méthodes de « type B » dans l'approche du GUM

Dans la recommandation INC-1, les méthodes de type B sont définies par la négative : en réordonnant la façon dont celles-ci sont introduites, on aboutit au fait qu'est de type B toute « méthode utilisée pour estimer [la] valeur numérique [des composantes d'incertitudes] qui sont évaluées par d'autres moyens [que statistiques]. »¹⁴. À la section 2.4, nous avons anticipé sur la catégorisation de la recommandation INC-1, qui est aussi celle du GUM; et, à la suite de certains métrologues, nous avons insisté sur l'importance de la distinction entre deux types d'*informations* disponibles à propos de la grandeur soumise à l'analyse d'incertitude. Par « d'autres moyens que statistiques », la recommandation INC-1 désigne de ce fait les cas où il ne s'agit pas d'exploiter directement les données expérimentales, mais où il faut évaluer certains paramètres, par exemple des termes correctifs, en s'appuyant sur des modèles théoriques, en exploitant des résultats effectués par d'autres, etc.

Le problème posé est le même que celui décrit dans la section 3.5 du chapitre 3, de laquelle nous reprenons les notations et le cadre général d'étude. On considère ainsi le cas où un laboratoire d'étalonnage fournit les informations relatives à un étalon correspondant à une grandeur d'entrée Z sous forme d'un encadrement de sa valeur :

$$z_0 - a < z < z_0 + a \quad (4.1)$$

13. Lira et Wöger (2006), p.S253.

14. Joint Committee for Guides in Metrology (JCGM) (2008d), p.30.

L'approche traditionnelle suggère de s'en tenir à un tel encadrement des limites crédibles d'erreur. L'objectif énoncé dans la recommandation INC-1 est de pouvoir associer à z une incertitude de mesure comprise en termes probabilistes – afin qu'à l'instar de l'« incertitude-type » décrite à la section 3.3, elle prenne la forme d'un *paramètre de dispersion*, ce qui garantirait alors la possibilité de combiner les différentes composantes d'incertitude. Ainsi est-il écrit que :

Les composantes de la catégorie B devraient être caractérisées par les variances estimées u_j^2 [...] Les termes u_j^2 peuvent être traités comme des variances et les termes u_j comme des écarts-types. Le cas échéant, les covariances doivent être traitées de façon analogue.¹⁵

Le GUM met en application la recommandation INC-1 en proposant une méthode de type B qui soit compatible avec les exigences qui y sont formulées. La méthode développée dans le GUM n'est pas une méthode systématique mais un traitement au cas par cas de différentes situations possibles, exposé à travers plusieurs exemples qui visent à broser un large spectre de cas¹⁶. Bien que cela ne soit qu'implicite dans le GUM, l'approche qui y est proposée peut être interprétée comme consistant à introduire une variable aléatoire, que nous nommerons ici $\hat{\zeta}$, qui porte sur la valeur *fixe* de Z et exprime le degré de croyance de l'expérimentateur quant à cette valeur fixe. Dans ce cas, il n'est plus question de décrire la variabilité d'un processus de mesure par des probabilités fréquentistes. Ces dernières, qui portent sur des événements physiques, sont abandonnées au profit de l'usage de probabilités épistémiques, dont l'objet est tout autre : dans les méthodes de type B, les probabilités visent à décrire un état de connaissance ou de croyance à propos d'une *proposition*¹⁷. Elles décrivent une caractéristique du *sujet* connaissant (l'expérimentateur) et non de l'objet de connaissance (le processus de mesure et la grandeur mesurée). Si ces probabilités sont choisies dans le cas présent, c'est parce qu'elles présentent l'intérêt d'être opérantes dans les cas où l'approche traditionnelle montre ses limites. C'est sur ce mode que Weise et Wöger défendent l'emploi de telles probabilités :

En interprétant la probabilité comme une description numérique d'un état de connaissance incomplète, sur la base de toute forme d'information rationnelle et pertinente possible, les statistiques bayésiennes offrent un énoncé de probabilité portant sur la vraie valeur, non aléatoire, inconnue, d'une quantité.¹⁸

La distribution de probabilité de la variable aléatoire ζ est *construite* à partir des informations disponibles. Dans le cas de l'encadrement strict de l'équation 4.1, les valeurs extérieures à l'intervalle sont en principe exclues : il leur est associé une probabilité nulle. Les autres valeurs sont également crédibles, puisque aucune information ne vient spécifier le contraire¹⁹. C'est

15. Joint Committee for Guides in Metrology (JCGM) (2008d), p.viii.

16. Joint Committee for Guides in Metrology (JCGM) (2008d), pp.12–15.

17. Ce changement apparaît nettement dans cette formule d'Estler : « [A probability] describes not reality in itself but only one's knowledge about reality. », Estler (1999), p.618.

18. Weise et Wöger (1993), p.2. L'article de Weise et Wöger ne visait pas à défendre l'emploi des probabilités épistémiques dans le GUM, mais dans une approche entièrement bayésienne qui serait venue se substituer aux méthodes en vigueur, et en particulier à celle que le GUM prévoyait de recommander. L'argumentaire concernant cette question précise est également valable dans le cas du GUM.

19. Ce cas est traité au point 4.3.7 du GUM, p.13.

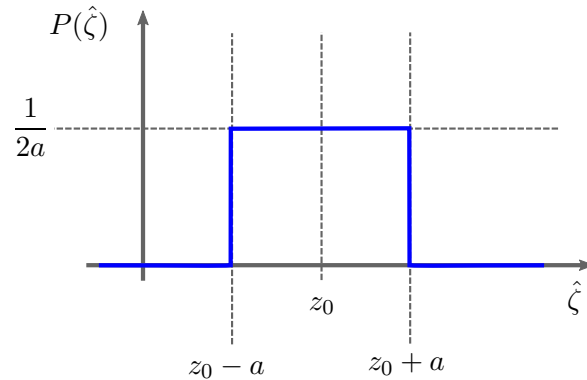


FIGURE 4.2 – Distribution de probabilité associée à la variable aléatoire $\hat{\zeta}$ décrivant l'état de connaissance disponible à propos de la grandeur Z . Lorsque la valeur de la grandeur est connue par un encadrement, le GUM préconise l'emploi d'une distribution « rectangulaire » attribuant une même probabilité à chaque valeur comprise à l'intérieur de l'intervalle d'encadrement, et une probabilité nulle aux valeurs situées en dehors de l'intervalle.

pourquoi il est raisonnable d'attribuer à $\hat{\zeta}$ une distribution de type « rectangulaire » centrée sur z_0 (figure 4.2). Nous verrons que cette procédure a par la suite été assimilée à la construction d'une probabilité bayésienne *a priori*; cependant, une telle notion de probabilité *a priori* est absente du GUM lui-même.

Dès lors que la variable aléatoire $\hat{\zeta}$ a été construite, on peut en calculer la *variance* – on a là le paramètre de dispersion initialement souhaité dans la recommandation INC-1 :

$$u(z) = \sqrt{V(\hat{\zeta})} \quad (4.2)$$

La variance de $\hat{\zeta}$ est une composante d'incertitude de mesure, puisqu'elle mesure la largeur d'une distribution qui porte directement sur l'état de connaissance de l'expérimentateur. Par analogie avec l'analyse fréquentiste de la variabilité expérimentale, elle est appelée « incertitude-type » sur Z . Il est possible de distinguer les deux types d'incertitude-type en parlant séparément d'« incertitude-type A » et d'« incertitude-type B ». Or, les mathématiques des probabilités ne varient pas d'une interprétation à l'autre²⁰. Toutes les règles de l'algèbre des variables aléatoires s'appliquent à $\hat{\zeta}$. C'est pourquoi il est possible de combiner entre elles des

20. On considère généralement que les différentes interprétations sont regroupées sous l'axiomatique formalisée par Kolmogorov en 1933 dans son monographe (Kolmogorov, 1933). Cet aspect des probabilités est précisément ce qui génère une telle discussion sur les interprétations des probabilités, comme l'explique Daston : « Lorsque Kolmogorov réussit à axiomatiser [le] calcul [des probabilités], il a défini une théorie mathématique : son essence se trouve dans la validité formelle des axiomes, définitions et théorèmes qui permettent une foule d'interprétations. Ce système formel ne nous oblige en aucune manière à choisir telle ou telle interprétation de la probabilité : au point de vue de la théorie mathématique, on ne peut pas choisir entre probabilité comme fréquence et probabilité comme degré de certitude ou entre bien d'autres interprétations possibles. », Daston (1989), p.715. Hacking, quant à lui, rappelle que la seule concordance mathématique ne justifie pas *a priori* d'employer la même dénomination pour un concept qui prend des interprétations sensiblement différentes : « Considérons une analogie : [...] une fois

incertitudes-type A et des incertitudes-type B. Cette idée est au fondement de la conception élargie de la *propagation* des incertitudes, une étape essentielle pour aboutir à une composante unique d'incertitude. Si l'on reprend la fonction de mesure simplifiée $Y = X + Z$ proposée à l'équation 3.1, on pourra obtenir l'incertitude-type sur Y à partir des incertitudes-types déterminées sur X (au moyen de ce que le GUM appelle méthodes de type A, c'est-à-dire le traitement fréquentiste des erreurs aléatoires décrit au chapitre précédent) et Z (au moyen des méthodes de type B présentées dans cette section). En vertu des propriétés mathématiques des variables aléatoires, on peut écrire que ²¹ :

$$V\left(\frac{\hat{x}}{n} + \hat{\zeta}\right) = V\left(\frac{\hat{x}}{n}\right) + V\left(\hat{\zeta}\right) \quad (4.3)$$

En remplaçant $V(\hat{x}/n)$, qui est inconnue, par son *estimateur* $u^2(x)$, cela permet ensuite d'énoncer la formule de *propagation des incertitudes* :

$$u_c^2(y) = u^2(x) + u^2(z) \quad (4.4)$$

Cette équation est similaire à l'équation (3.20) déjà présentée dans le cas fréquentiste, mais qui avait alors un caractère *ad hoc*. L'équation (4.4) vient répondre à l'objectif même pour lequel le GUM a été conçu : rendre possible le calcul d'une résultante *unique* $u(y)$ d'incertitude de mesure à propos du mesurande Y . La résultante $u_c(y)$ est appelée « incertitude-type composée ».

La propagation des incertitudes de tous types constitue une avancée notable du GUM. Cependant, la démarche du GUM occulte volontairement une difficulté majeure, qui a été à la source de nombreuses critiques postérieures au GUM : puisque les distributions \hat{x} et $\hat{\zeta}$ sont interprétées avec des probabilités de natures différentes, l'opération qui consiste à les combiner n'est pas si immédiate qu'elle peut en avoir l'air de prime abord. Vient ainsi un obstacle conceptuel majeur : si l'opération de propagation des incertitudes consiste à combiner probabilités fréquentistes et probabilités épistémiques, quelle est la nature de la probabilité associée à l'incertitude finale ²²? Sans surprise, cette question oppose les défenseurs d'une approche fréquentistes et les promoteurs de la méthode bayésienne. Pour les premiers, il faut trouver une solution à ce problème conceptuel en développant une méthode de type B qui s'appuie sur des fondements fréquentistes. Pour les seconds, il faut au contraire embrasser une perspective pleinement épistémique. Dans tous les cas, cela impliquerait de réviser le GUM.

que les principes de Newton ont été donnés en termes distincts, personne ne voulait parler de poids actifs et passifs. Personne n'aurait dit que puisque le poids et la masse satisfont tous deux les axiomes de la théorie de la mesure, nous avons besoin du même mot pour les deux. Pourtant cela a été utilisé comme argument pour garder le même mot pour les notions aléatoires et épistémiques de la probabilité », [Hacking \(1975\)](#) p.13. Hájek note que le terme d'« interprétation » des probabilités peut être trompeur : il ne s'agit pas d'interprétations différentes, mais bien de *concepts* distincts ; voir [Hájek \(2012\)](#), en particulier l'introduction (entrée consultée le 3 juillet 2015).

21. Rappelons qu'en prenant la moyenne des observations individuelles, on a réduit la variabilité d'un facteur n , c'est pourquoi dans l'équation on considère non pas la variance de \hat{x} mais celle de \hat{x}/n .

22. En 1980, le NPL formulait déjà un avertissement à ce sujet : « des traditions différentes peuvent demander des types différents d'énoncés et des méthodes diverses de dérivation [pour le traitement des incertitudes]. Mais si, dans un énoncé d'incertitude, la *signification* même de l'énoncé est incertaine, alors l'énoncé n'a aucune valeur », [Kaarls \(1980\)](#), annexe V, p.4 (souligné par les auteurs).

Pour mieux comprendre en quoi le problème conceptuel énoncé ici se pose de façon très concrète aux métrologues, il faut s'intéresser à la façon dont le GUM propose de passer d'une incertitude-type, qui, rappelons-le, n'est d'abord qu'un paramètre de dispersion sans interprétation probabiliste précise, à un intervalle d'incertitude véritablement probabiliste. Nous développons cet aspect dans la prochaine section.

4.4 « Intervalles élargis » et critiques de l'approche du GUM

4.4.1 Au-delà de la propagation des incertitudes-types : les incertitudes « élargies »

Dans le chapitre consacré à l'approche traditionnelle, nous avons expliqué que les « incertitudes-types », qui ne sont en définitive que des paramètres de dispersion, ne sont pas suffisantes pour appréhender le résultat obtenu de façon pleinement probabiliste. Dans le cas traditionnel, nous avons abouti à la construction d'un intervalle de confiance, associé à un niveau de confiance p . Le constat est identique en ce qui concerne les méthodes de type B du GUM : l'incertitude-type B n'est elle aussi qu'un paramètre de dispersion. Dès lors que l'on veut obtenir une incertitude associée à une probabilité bien déterminée, il faut engager un raisonnement supplémentaire qui, à l'instar de la démarche opérée dans l'approche traditionnelle, nécessite de se référer à la forme de la distribution de la variable aléatoire $\hat{\zeta}$. Les intervalles de confiance ne sont toutefois opérants que dans le cadre d'un modèle fréquentiste de données. Dans le cas des méthodes de type B du GUM, la méthode consistant à associer une probabilité p à un intervalle donné est particulièrement directe, puisqu'on a construit la variable aléatoire visant à caractériser les valeurs possibles de la grandeur. Pour la grandeur Z des sections précédents, cela revient simplement à un calcul d'aire, à savoir l'aire contenue sous la courbe représentant la distribution de probabilité de $\hat{\zeta}$ entre les bornes de l'intervalle choisi²³ (voir figure 4.3).

Récapitulons : si l'on se limite au cas de la seule analyse des erreurs aléatoires – les méthodes de type A, fréquentistes, du GUM – l'expression probabiliste du résultat de mesure fait appel aux intervalles de confiance et, à défaut d'être simple, la méthode pour y aboutir est bien déterminée. Le constat est similaire pour les méthodes de type B, mais les intervalles produits ne sont pas de même nature. L'on peut reprendre une fois encore la fonction de mesure simplifiée $Y = X + Z$ où X est déterminée à l'aide de mesures répétées, dont les résultats sont analysés par des méthodes de type A, et Y est déterminée à l'aide d'informations extérieures, analysées par des méthodes de type B. Un intervalle de confiance vient caractériser X ; un autre intervalle probabiliste vient caractériser Z . Mais qu'en est-il pour Y ? Rappelons que la mission du GUM est justement de rendre possible l'expression conjointe de toutes les composantes d'incertitude ; et donc, par suite, de faire en sorte que l'on puisse formuler un intervalle probabiliste qui caractérise non pas X ou Z pris séparément, mais le mesurande final Y . La synthèse que le GUM se propose d'opérer s'appuie sur l'introduction d'un nouveau terme, l'« incertitude

23. Nous laissons ici de côté les questionnements portant sur le fait qu'à une probabilité donnée peuvent correspondre plusieurs, voire une infinité, d'intervalles élargis (illustrée à la figure 4.3). Ces questions sont discutées dans la littérature spécialisée mais n'ont pas de réelle conséquence sur la discussion présente.

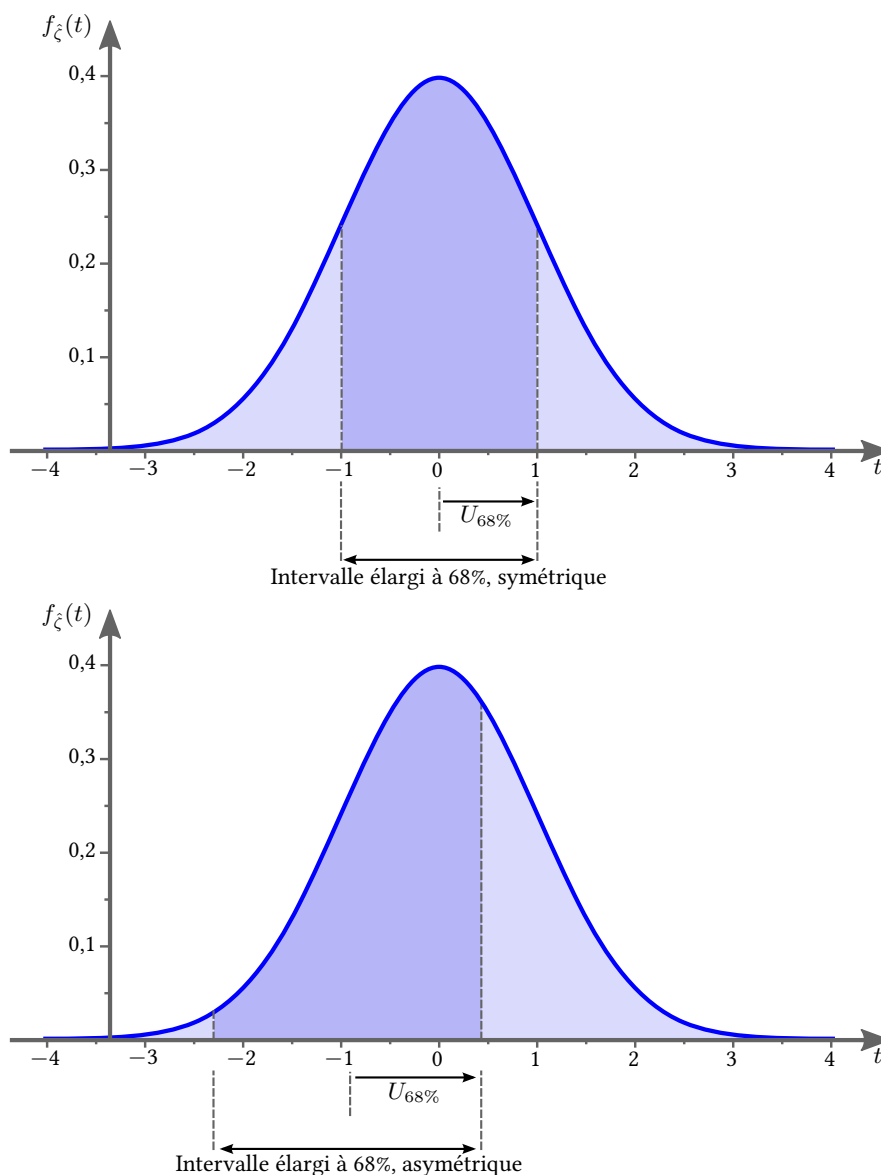


FIGURE 4.3 – Deux exemples d'intervalles élargis à 68% pour une même loi normale centrée réduite. L'intervalle est défini de telle façon qu'il contienne 68% de l'aire contenu sous la courbe. La probabilité de 0,68 qui lui est associée est appelée « probabilité de couverture ». L'« incertitude élargie » est la demi-largeur de l'intervalle. Comme la figure le montre, un tel intervalle n'est pas défini de façon unique. En général, on privilégie en métrologie les intervalles symétriques autour de la valeur centrale de la distribution (eux-mêmes sont souvent les intervalles les plus étroits qu'on puisse obtenir), mais cela n'est pas *a priori* nécessaire.

élargie », d'abord définie comme suit :

Grandeur définissant un intervalle, autour du résultat d'un mesurage, dont on puisse s'attendre à ce qu'il comprenne une fraction élevée de la distribution des valeurs qui pourraient être attribuées raisonnablement au mesurande.²⁴

La définition du GUM laisse encore place au vague (« fraction élevée ») mais le VIM3 est ensuite venu compléter l'exploration de cette notion. L'incertitude élargie se rapporte à un intervalle, l'« intervalle élargi », dont elle est la demi-largeur, et à une probabilité, la « probabilité de couverture », qui représente la « fraction élevée » en question²⁵. Dans un cas uniquement fréquentiste (méthodes de type A du GUM), probabilité de couverture et intervalle élargi s'identifient respectivement au niveau de confiance et à l'intervalle de confiance. Pour les méthodes de type B du GUM, on se rapporte à la méthode décrite précédemment (voir figure 4.3).

Un problème apparaît lorsqu'on s'intéresse à la détermination d'une incertitude « élargie » à partir d'une conjonction d'incertitude fréquentistes et d'incertitudes de type B. Rappelons que la combinaison des composantes d'incertitude – ce qui est appelé généralement la *propagation* des incertitudes – est en apparence résolue dans le GUM, mais *pour les incertitudes-types seulement* ; c'est ce que nous avons vu à la section précédente, où nous avons montré que :

$$u^2(y) = u^2(x) + u^2(z)$$

Cependant, rien n'indique que l'on puisse légitimement procéder à la même opération pour les incertitudes *élargies*. Si la propagation des incertitudes-types décrite est immédiate – ou, tout du moins, semble l'être – c'est parce que le raisonnement qui s'appuie sur les propriétés des variances des variables aléatoires permet de court-circuiter tout raisonnement sur la forme des distributions de probabilité, et, par suite, sur la signification de ces distributions. De ce fait, la propagation des incertitudes « élargies » se révèle plus complexe.

Comment calculer l'« incertitude élargie » associée à une « probabilité de couverture » p donnée ? Il semble qu'il n'y ait que deux voies possibles. La première consiste à adopter un raisonnement entièrement fréquentiste et à trouver un moyen de traduire la composante d'incertitude de type B en une composante fréquentiste. L'on peut recenser plusieurs tentatives de ce genre. Eisenhart proposait déjà en 1963 des façons de considérer l'erreur systématique sous un angle probabiliste²⁶, mais en nuancit immédiatement sa portée²⁷, et ne considérait pas

24. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.3.

25. Notons que la définition d'« incertitude élargie » du VIM3 n'est pas particulièrement plus spécifique : la notion est comprise comme le « produit d'une incertitude-type composée et d'un facteur supérieur au nombre un » ([Joint Committee for Guides in Metrology \(JCGM\), 2012](#), p.27) sans que ne soit explicitée la façon dont est déterminé le facteur en question. Cette définition est solidaire de celles d'« intervalle élargi » et de « probabilité de couverture », elles aussi données dans le VIM3 ; ce n'est vraiment que par la conjonction des trois définitions que l'on peut vraiment comprendre la signification accordée à ces concepts. Les définitions vont comme suit : « intervalle élargi : intervalle contenant l'ensemble des valeurs vraies d'un mesurande avec une probabilité déterminée, fondé sur l'information disponible », [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.27 ; « probabilité de couverture : probabilité que l'ensemble des valeurs vraies d'un mesurande soit contenu dans un intervalle élargi spécifié », [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.28. Chacun de ces concepts peut être compris selon l'interprétation fréquentiste ou épistémique des probabilités, même si la formulation qui est choisie utilise plutôt le vocabulaire de la seconde.

26. [Eisenhart \(1963\)](#), pp.181–185.

27. « En toute rigueur, les analyses qui précèdent et qui emploient la théorie des probabilités sont [...] inappli-

cela suffisant pour pouvoir combiner les composantes d'incertitude. Plus récemment, Willink a proposé une solution qui consiste à considérer les erreurs systématiques dans une perspective dans laquelle celles-ci peuvent présenter une variabilité expérimentale. Cela revient à sortir des conditions strictes de répétabilité, et à se placer dans des conditions élargies – c'est-à-dire en fait des conditions de reproductibilité dans lesquelles la méthode de mesure, les instruments, les lieux, etc. peuvent varier – dans lesquelles on peut envisager les erreurs systématiques comme étant elles aussi le résultat d'un tirage aléatoire²⁸. La sous-section suivante montrera que le choix adopté dans le GUM est un entre-deux, qui consiste à utiliser des probabilités épistémiques pour traiter les erreurs systématiques, puis interpréter ces dernières *comme si* il s'agissait de probabilités fréquentistes. Du moins est-ce ainsi que de nombreux métrologues l'ont interprété ; et nous verrons que ces derniers ne l'ont pas accueilli très favorablement. Une seconde voie consiste, à l'inverse, à adopter un raisonnement entièrement fondé sur les probabilités épistémiques. C'est la voie que préconisent les défenseurs d'une approche bayésienne de la mesure, et c'est également celle qui a été choisie dans le premier supplément du GUM, publié en 2008 ; nous décrirons cette solution dans la section suivante.

4.4.2 Formule de Welch-Satterthwaite et interprétation fréquentiste du GUM

Dans le GUM, la solution adoptée pour effectuer la propagation des incertitudes « élargies » consiste à se ramener au cas fréquentiste standard (associé à l'hypothèse gaussienne) par un modèle *ad hoc*, caractérisé en particulier par une formule mathématique approchée nommée « loi de Welch-Satterthwaite », valable sous certaines conditions²⁹. La loi de Welch-Satterthwaite trouve son origine dans des travaux de statistiques fréquentistes³⁰, et est d'ailleurs majoritairement promue par des défenseurs de l'approche traditionnelle³¹. Il ne serait pas utile de la détailler sur le plan technique. Retenons simplement que pour l'interpréter de façon pleinement satisfaisante, il est nécessaire en principe que les méthodes de type B soient conçues en termes fréquentiste³². Or, nous l'avons vu, cela n'est pas le cas dans le GUM, qui fait usage de probabilités épistémiques. Pourtant, le GUM emploie tout de même la formule : cela passe en fait par une *réinterprétation* fréquentiste des méthodes de type B, qui affleure dans l'extrait suivant :

Pour une estimation x_i d'une grandeur d'entrée X_i qui n'a pas été obtenue à partir d'observations répétées, la variance estimée associée $u^2(X_i)$ ou l'incertitude-type $u(x_i)$ est évaluée par un jugement scientifique fondé sur toutes les informations disponibles au sujet de la *variabilité possible* de X_i .³³

Parler de « variabilité possible » de la grandeur, c'est replacer l'analyse dans un cadre fictif où

cables au problème à portée de main », Eisenhart (1963), p.184.

28. Willink (2013). Nous ne rentrons pas dans le détail de l'approche de Willink ; pour une description détaillée, voir Grégis et de Courtenay (2016).

29. Joint Committee for Guides in Metrology (JCGM) (2008d), pp.76–77. Une justification en est donnée dans Kirkup et Frenkel (2006), pp.174–179.

30. Cela est rappelé par Willink et Hall (2001), p.10. Voir aussi l'un des textes d'origine, Welch (1947).

31. Citons par exemple Grabe (1987) et Willink et Hall (2001).

32. Willink et Hall (2001).

33. Joint Committee for Guides in Metrology (JCGM) (2008d), p.12 (nous soulignons).

l'on *pourrait* modéliser l'erreur correspondante comme le résultat d'un tirage aléatoire parmi une population parente ; il s'agit d'une reconstruction fréquentiste. C'est précisément la raison pour laquelle cette équation ne satisfait pas les défenseurs de l'approche bayésienne. Ces derniers font remarquer qu'elle transforme en simple artifice l'introduction de probabilités épistémiques dans le GUM. Dans la continuité, ils notent une importante différence entre ce que le GUM annonce – une préférence envers l'usage des probabilités épistémiques – et ce que le GUM fait réellement – considérer pourtant les incertitudes issues des méthodes de type B *comme si* elles étaient des fréquences limites³⁴. C'est pourquoi les critiques de la loi de Welch-Satterthwaite parlent de « fréquentialisation » à son sujet³⁵, et la considèrent comme une limite conceptuelle majeure du GUM³⁶, en raison de laquelle, selon leur point de vue, le GUM reste essentiellement fréquentiste. Un tel mouvement ne peut pas les satisfaire, et ils défendent un mouvement inverse, qui consiste à unifier l'approche du GUM au moyen des seules probabilités épistémiques. On retrouve cette idée chez Bich et Kool :

Pour atténuer l'incohérence induite par le mélange de variances d'échantillons et de variances de distributions de probabilités, dans le GUM ces dernières sont artificiellement réduites à des estimateurs en leur attachant des degrés de liberté. [...] De cette façon, les deux points de vue existant dans le GUM sont réconciliés en orientant la position bayésienne vers la fréquentiste. On verra dans un moment que l'orientation opposée est désormais préférée.³⁷

Nous avons vu que la question de l'interprétation des probabilités n'est abordée que tangentiellement dans le GUM, l'objectif étant de souligner les bénéfiques mathématiques de la solution qui y est adoptée. Or, nous constatons, à la lumière de la discussion présente, que la question de l'interprétation des probabilités n'est pas une simple subtilité philosophique ; et l'on comprend que la solution apportée dans le GUM n'est pas aussi naturelle qu'elle y est présentée – au contraire, elle apparaît même très artificielle. Loin des prétentions initiales de traiter variabilité et informations externes « sur la même base probabiliste »³⁸ ou de produire des composantes d'incertitude « de même nature » qui « doivent être traitées de manière identique »³⁹, nous constatons que les fondements probabilistes des méthodes employées *diffèrent* et que les composantes d'incertitude ne peuvent pas être interprétées de la même façon. L'approche du GUM repose sur un artifice qui tient au fait que les mathématiques des probabilités sont identiques pour les différentes interprétations. Dans le GUM, l'accent est porté sur l'apparente cohérence mathématique de la méthode adoptée, et sur les problèmes techniques qu'elle résout. Les difficultés inhérentes à la question de l'interprétation des probabilités y sont donc masquées par l'absence de questionnement philosophique et critique sur la nature des probabi-

34. Comme l'exprime Lira, « Cette équation est obtenue en forçant la distribution de probabilité à être interprétée comme une distribution de fréquence », Lira (2002), pp.116-117.

35. « Cette procédure, que l'on pourrait appeler "fréquentialisation" à défaut d'un meilleur terme, conduit souvent à des résultats acceptables en pratique. Néanmoins, elle n'apparaît pas convaincante à tous les égards », Lira et Wöger (2001), p.1174.

36. « Bien des raisons font de la formule de Welch-Satterthwaite la recommandation probablement la plus controversée du GUM », Lira et Wöger (2006), p.S254.

37. Bich et Kool (2012), p.9.

38. Ehrlich, Dybkaer et Wöger, cités à la section 4.2.

39. GUM, cité à la section 4.2.

lités elles-mêmes. Comme la question philosophique n'est pas appréhendée de façon claire, cela conduit à un large spectre d'interprétations différentes, ce qui ne semble réellement satisfaire personne – que ce soient ceux qui considèrent intenable le subjectivisme bayésien ou ceux qui, au contraire, veulent prolonger le mouvement initié avec les méthodes de type B. Cette lacune a été immédiatement repérée par des métrologues plus enclins à s'attarder sur la question des fondements, à qui le GUM est apparu comme un compromis insatisfaisant⁴⁰.

Au vu des critiques formulées contre le GUM, on comprend que certains métrologues aient milité pour une réforme plus profonde encore des fondements probabilistes de la mesure. Par exemple, Kacker et Jones ont proposé une extension de l'emploi des probabilités épistémiques aux méthodes dites de type A ; ce qui revient, selon eux, à placer l'ensemble (type A et type B) sous un même cadre bayésien unifié.

[Le GUM] recommande les statistiques classiques (fréquentistes) pour l'évaluation des composantes d'incertitudes de Type A ; mais il interprète l'incertitude composée d'un point de vue bayésien. Cela est incohérent. Pour surmonter cette incohérence, nous suggérons que toutes les incertitudes de Type A soient évaluées au travers d'une approche bayésienne.⁴¹

Cette proposition est celle qui retient le plus notre attention, car elle correspond dans son esprit à l'approche développée dans le premier supplément du GUM⁴², publié en 2008. Ce document est le premier document original à avoir été publié par le premier groupe de travail du JCGM, l'organisme consacré à la maintenance et la révision du GUM et du VIM. Sa publication vise à couvrir un ensemble de cas dans lesquels le GUM est inopérant, par exemple lorsque la fonction de mesure est non-linéaire, ou lorsque l'hypothèse gaussienne pour les méthodes de type A n'est pas valable. Il permet également d'introduire une méthode de calcul d'incertitude fondé sur des simulations numériques, et ce afin d'exploiter à leur maximum les outils de calcul moderne. Depuis 2008, d'autres suppléments sont venus s'ajouter à ce premier supplément pour constituer ainsi la « famille de documents du GUM »⁴³ Le supplément 1 du GUM est peut-être le premier document officiel des institutions métrologiques internationales qui puisse ouverte-

40. À cela, nous pourrions ajouter l'hypothèse selon laquelle que le GUM a pu être immédiatement perçu comme un document en partie daté, du fait qu'il est avant tout un produit des années 1980, et même du début de cette décennie (la recommandation INC-1 datant de 1980) ; le document reste donc très conservateur, même si sa publication se situe dans une période de transformation conceptuelle. La remarque est d'autant plus valable aujourd'hui que la re-publication du GUM par le JCGM en 2008 peut donner l'illusion que le document est récent.

41. Kacker et Jones (2003), p.235.

42. Joint Committee for Guides in Metrology (JCGM) (2008b)

43. Cette expression a été utilisée lors du colloque *GUM : past, present, and future* organisé les 7 et 8 novembre 2013 au NPL. Le groupe de travail du JCGM consacré à la maintenance du GUM (JCGM WG1) a décidé d'adopter une stratégie de publication, consistant à rédiger plusieurs suppléments qui viennent compléter et affiner le GUM. En effet, il a jugé que cette stratégie était préférable au fait de réviser le GUM d'un seul bloc en publiant une toute nouvelle édition du document principal, alors même que celle-ci commençait à peine à être digérée par ses utilisateurs. Ce faisant, le JCGM a reconnu la difficulté de produire un document unique adapté à tous les usages et tous les buts. Chacun des suppléments vient répondre à un besoin précis auquel le GUM ne répond pas, ou répond trop rapidement. À ce titre, le supplément 1 du GUM constitue l'un des deux premiers documents *originaux* produits par le JCGM, en l'occurrence par le groupe de travail WG1 du JCGM – l'autre document étant le VIM3, rédigé par le groupe de travail WG2 du JCGM.

ment être qualifié de bayésien⁴⁴. Nous décrivons la démarche suivie dans la prochaine section.

4.5 Vers une approche pleinement bayésienne de la mesure

Précisons d'abord qu'il n'a fallu attendre ni 2008 et la publication du supplément 1 du GUM, ni 1993 et celle du GUM, pour voir apparaître des théories ouvertement bayésiennes en métrologie. Notons pour commencer que Birge et Deming, en 1934, mentionnaient déjà une « méthode du *posterior* »⁴⁵ qui à de nombreux égards, apparaît comme une formulation bayésienne de l'analyse d'erreur. Müller, en 1979, esquissait également une discussion sur l'opposition entre probabilités objectives et subjectives, et mentionnait le théorème de Bayes, bien que de façon assez brève⁴⁶. Plus récemment, Weise et Wöger avaient proposé une approche bayésienne dès 1992⁴⁷, insatisfaits de la tournure que prenait la rédaction du GUM. Si le formalisme et la méthode statistique utilisés dans leur article n'ont pas connu un succès remarqué⁴⁸, l'idée même de fonder une analyse d'incertitude sur une conception bayésienne, en revanche, a eu un fort impact sur la communauté, si bien que leur article est souvent considéré comme l'un des textes fondateurs de l'application des statistiques bayésiennes en métrologie⁴⁹. Dans la suite de cette section, nous choisissons plutôt de nous intéresser à la démarche qui est développée dans le supplément 1 du GUM.

4.5.1 L'approche du Supplément 1 du GUM : un traitement bayésien des méthodes de « type A »

Rappelons la situation générique dans laquelle nous nous sommes régulièrement placés jusqu'ici. La grandeur visée est ici la grandeur d'entrée X de la fonction de mesure $Y = X + Z$. Un échantillon $\{x_i\}$ de n données expérimentales a été obtenu après répétition de la mesure de X dans des conditions considérées comme identiques (conditions dites « de répétabilité »). Nous avons désigné par χ la « valeur vraie » de la grandeur X .

Comment procéder à un traitement bayésien des collections de données expérimentales variables sur X ? L'approche bayésienne consiste à considérer les résultats expérimentaux comme des informations qui permettent de mettre à jour les connaissances de l'expérimentateur à propos de la grandeur mesurée. La première étape est très semblable à l'opération effectuée dans les méthodes de type B développées dans le GUM, où l'introduction des probabilités épistémiques mène à ce que nous avons interprété comme étant la construction d'une variable aléatoire décrivant l'état de connaissance de l'expérimentateur – exprimé par des degrés de croyance – à

44. Elster, Wöger et Cox (2007)

45. Deming et Birge (1934), pp.149–157.

46. Müller (1979), voir en particulier p.246.

47. Weise et Wöger (1993). La mention des probabilités

48. Comme le fait remarquer D'Agostini : « (b)ien que les auteurs montrent comment [l'approche bayésienne] est puissante dans de nombreuses applications, l'utilisation du principe d'entropie maximum est, à mon avis, un point faible qui empêche la théorie d'être aussi générale que revendiquée », D'Agostini (1996), p.4. D'Agostini, comme d'autres ensuite, proposeront de fonder la conception bayésienne de la mesure sur le *théorème de Bayes*.

49. Demeyer (2011), p.21.

propos de la valeur la valeur de la grandeur visée. Est donc également introduite ici une variable aléatoire qui exprimant la connaissance de l'expérimentateur quant à la valeur de la grandeur X . Pour distinguer cette variable aléatoire de la variable \hat{x} introduite dans le cas traditionnel pour représenter le processus de mesure, nous la nommerons ici $\hat{\chi}$. Pour construire cette variable aléatoire, l'expérimentateur réunit d'abord les informations dont il dispose *avant* d'effectuer les mesures – on parle de *distribution a priori* (ou *prior*). À la différence des méthodes de type B du GUM, l'introduction d'une telle variable aléatoire est explicitement revendiquée dans le supplément 1. Dans le cas des méthodes des type B du GUM, nous pouvons en fin de compte *réinterpréter* la démarche adoptée comme étant elle aussi la construction d'une distribution *a priori* sur la grandeur visée, à la différence que, dans ce cas précis, l'opération s'arrête alors immédiatement : aucune mesure supplémentaire n'étant effectuée, l'incertitude de mesure est directement calculée sur la base du *prior* obtenu.

Dans l'approche bayésienne de la mesure, lorsque des résultats expérimentaux viennent s'ajouter à la connaissance *a priori* sur la grandeur visée, ce qui est par définition le cas pour les méthodes de type A, ceux-ci viennent enclencher un processus d'actualisation des connaissances : la distribution *a priori* est révisée à la lumière des données expérimentales obtenues. Les résultats de mesure sont exploités au moyen du théorème de Bayes qui fournit le mode de calcul d'une probabilité *a posteriori* représentant la connaissance de l'expérimentateur *une fois effectuées* les mesures expérimentales. Dans le cas étudié ici, le théorème de Bayes se présente sous la forme suivante :

$$p(\chi | \{x_i\}) \propto L(\chi | \{x_i\}) \times p(\chi) \quad (4.5)$$

$$\text{où } \begin{cases} p(\chi) \text{ est la probabilité } a \text{ priori} \\ p(\chi | \{x_i\}) \text{ est la probabilité } a \text{ posteriori} \\ L(\chi | \{x_i\}) \text{ est la fonction de « vraisemblance »} \end{cases}$$

Afin d'éclaircir la signification de l'équation (4.5), expliquons d'abord ce qu'est le théorème de Bayes. Le théorème de Bayes tire son origine des propriétés des probabilités conditionnelles. Si l'on considère deux propositions A et B , la probabilité conditionnelle de A sachant B est la probabilité de A dès lors que B est vraie, et est définie par⁵⁰ :

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (4.6)$$

où $P(A \cap B)$ est la probabilité de A et B , c'est-à-dire la probabilité d'avoir à la fois A et B vrais. Par symétrie, on a donc aussi :

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \quad (4.7)$$

Comme $P(B \cap A) = P(A \cap B)$, on trouve alors

$$P(A/B) \times P(B) = P(B/A) \times P(A) \quad (4.8)$$

50. Saporta (2006), p.7.

ou encore, en reformulant d'une façon similaire à l'équation (4.5) :

$$P(A/B) \propto P(B/A) \times P(A) \quad (4.9)$$

Comme l'indique Saporta, « le théorème de Bayes est souvent appelé théorème sur la “probabilité des causes” »⁵¹. En effet, si l'on considère A comme une cause possible de B , alors la connaissance de la probabilité de B sachant A permet de remonter à la probabilité de la cause A lorsque B est vérifié. De ce fait, on parle aussi d'« inversion de probabilité »⁵², car on inverse les rôles de A et de B . Dans le cas du traitement des erreurs aléatoires, la « cause » est la valeur vraie χ de la grandeur X mesurée, et la conséquence est l'échantillon $\{x_i\}$ de données expérimentales. Cela suppose, cependant, que l'on puisse parler de « probabilité de la cause » : l'équation (4.8) n'est valable que s'il existe en premier lieu une probabilité sur A et sur B . C'est là précisément un point de désaccord entre fréquentistes et bayésiens, sur lequel nous avons déjà insisté à plusieurs reprises.

Les probabilités $p(\chi)$ et $p(\chi|\{x_i\})$ sont donc respectivement la distribution de probabilité *a priori*, déterminée avant la mesure, et la distribution de probabilité *a posteriori*, déterminée après avoir effectué les mesures. Pour appliquer l'équation (4.5), il faut au préalable déterminer la fonction $L(\chi | \{x_i\})$. Sa détermination est fondée sur le raisonnement suivant : en fonction de la valeur vraie χ de la grandeur X mesurée, certaines données expérimentales sont plus vraisemblables que d'autres. On peut ainsi écrire la probabilité d'obtenir, pour une donnée expérimentale x_i , une valeur quelconque t , à partir d'une distribution de probabilité similaire à celle formulée dans l'hypothèse (H3) du modèle fréquentiste :

$$p(x_i = t | \chi) = f_\chi(t) \quad (4.10)$$

Par extension, on peut écrire la probabilité d'obtenir un échantillon complet $\{x_i\}$ de n données expérimentales (si les données sont indépendantes les unes des autres, cela se fait simplement en montant la fonction f à la puissance n) :

$$p(\{x_i\} = \{t_i\} | \chi) = g_\chi(\{t_i\}) \quad (4.11)$$

La fonction g est alors une fonction sur les valeurs possibles de $\{x_i\}$, et χ est un paramètre fixé. Dans l'approche bayésienne, la perspective est inversée. En effet, x_i est connue et il n'y a donc pas de sens à la décrire par une probabilité. En revanche, χ est inconnue : l'on aimerait donc obtenir une fonction sur les valeurs possibles de χ , où x_i serait le paramètre fixe. Il suffit pour cela de réécrire la fonction (4.11) en inversant les rôles de $\{x_i\}$ et χ . On obtient ainsi la fonction suivante :

$$L(\chi | \{x_i\}) = p(\{x_i\} | \chi) \quad (4.12)$$

La fonction L n'est pas une distribution de probabilité. En particulier, la somme de cette fonction sur tous les $\{x_i\}$ possibles n'est pas égale à 1. Cela signifie encore que la valeur de $L(\chi | \{x_i\})$ ne donne pas la probabilité de χ en fonction des données obtenues. On parle de fonction de vraisemblance. La fonction de vraisemblance ne dépend pas de la distribution de probabilité

51. Saporta (2006), p.10.

52. Voir Stigler (1986), chapitre 3 : “Inverse probability”, pp.99–138.

a priori, et c'est combinée avec cette distribution qu'elle permet justement de calculer ensuite la probabilité de χ sachant $\{x_i\}$, c'est-à-dire la probabilité *a posteriori*. Le théorème de Bayes permet donc une mise à jour des connaissances qui se traduit mathématiquement par le passage d'une distribution de probabilité *a priori* $\hat{\chi}$ avant mesure, à une distribution de probabilité *a posteriori* $p(\chi | \{x_i\})$ après mesure. Un exemple de calcul est donné à l'annexe A.3.

La distribution de probabilité de $\hat{\chi}$ est de même nature que la distribution de $\hat{\zeta}$ utilisée pour décrire la méthode de type B du GUM : toutes deux expriment l'état de connaissance de l'expérimentateur à propos de la valeur de la grandeur concernée, en tenant compte des informations dont il dispose avant l'expérience et, le cas échéant, de celles que l'expérience lui fournit. Dès lors, on peut immédiatement se rapporter aux calculs menés dans les méthodes de type B du GUM pour déterminer l'incertitude-type et l'incertitude élargie correspondantes. Par suite, la détermination d'une incertitude composée, par propagation des incertitudes, est elle aussi simplifiée ; c'est ce que nous décrivons dans la sous-section à venir.

4.5.2 Propagation des incertitudes et formulation du résultat final

Dans la conception bayésienne du supplément 1, il n'est plus nécessaire de faire la distinction entre méthodes de type A et méthodes de type B : en effet, n'y a plus qu'une seule méthode bayésienne unifiée. Si l'on considère une fonction de mesure générique $Y = f(X_1, \dots, X_p)$, toutes les grandeurs d'entrée X_i doivent faire l'objet préalable d'une description au moyen d'une distribution de probabilité *a priori* construite avec les informations disponibles avant toute mesure. Ensuite, seules certaines grandeurs d'entrée sont concernées par le processus d'actualisation des connaissances réalisé par la formule de Bayes : ce sont les grandeurs d'entrées qui font l'objet d'une mesure expérimentale effective. Toutes les distributions de probabilité ainsi construites font appel à une interprétation épistémique des probabilités et sont interprétées comme la mesure d'un état de connaissance. C'est exactement ainsi que Kacker et Jones interprètent la méthode :

[La fonction de mesure] $Y = f(X_1, \dots, X_n)$ est bien définie si l'on suit le point de vue bayésien suivant. Les évaluations de Type A sont des paramètres de distributions bayésiennes *a posteriori* et les évaluations de Type B sont des paramètres de distributions bayésiennes *a priori*. Les deux ont la même interprétation probabiliste.⁵³

Reste alors à s'attaquer à la question de la « propagation des incertitudes ». Dans le GUM, celle-ci avait été résolue (en apparence) pour les incertitudes-types (section 4.3) mais restait extrêmement problématique pour les incertitudes « élargies » (section 4.4). Dans le supplément 1 du GUM, elle est surmontée au travers d'une méthode plus générale, appelée « propagation des distributions »⁵⁴. Cette méthode stipule qu'il est préférable de raisonner d'abord sur les distributions de probabilité avant de chercher à en tirer une incertitude « élargie » finale. Si l'on considère de nouveau la fonction de mesure générique $Y = f(X_1, \dots, X_p)$, la méthode procède comme suit. À chaque grandeur d'entrée est associée une variable aléatoire bien définie, de

53. Kacker et Jones (2003), p.243.

54. Joint Committee for Guides in Metrology (JCGM) (2008b), p.5.

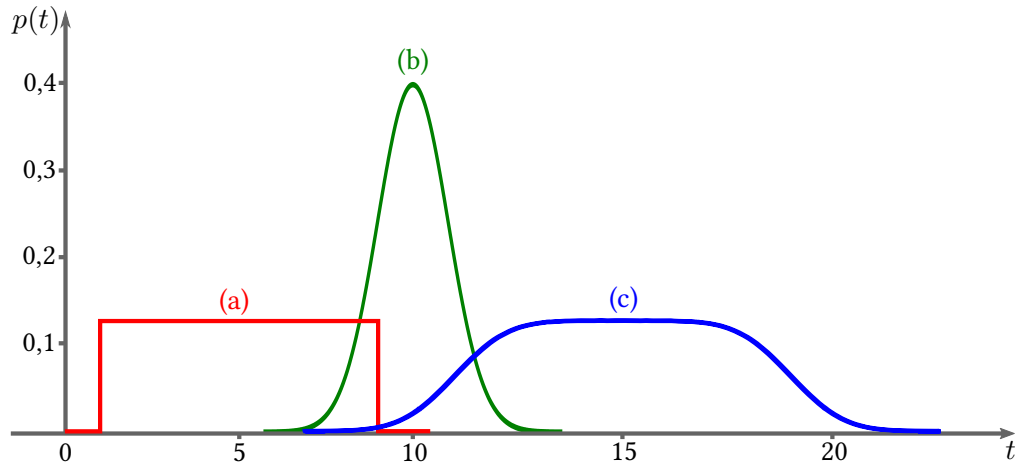


FIGURE 4.4 – Produit de convolution d’une fonction rectangulaire (courbe (a), en rouge) et d’une fonction gaussienne (courbe (b) en vert). La fonction obtenue est représentée par la courbe (c), en bleu. L’espérance et la variance du produit de convolution sont respectivement la somme des espérances et des variances des deux fonctions d’origine. Si la gaussienne est la distribution *a posteriori* de la variable aléatoire $\hat{\chi}$ associée à la grandeur X et la fonction rectangulaire est la distribution *a priori* de la variable aléatoire $\hat{\zeta}$ associée à la grandeur Z , alors le produit de convolution est la fonction que l’on obtient par « propagation des distributions » pour la grandeur $Y = X + Z$. L’incertitude élargie sur Y est alors bien déterminée à partir de la connaissance de \hat{y} .

sorte que toutes les variables aléatoires sont de même nature. Dès lors, on peut également associer une variable aléatoire \hat{y} au mesurande Y , et cette variable aléatoire est directement définie par la relation :

$$\hat{y} = f(\hat{\chi}_1, \dots, \hat{\chi}_2) \quad (4.13)$$

Avec une telle méthode, il n’est plus nécessaire de chercher à déterminer comment une incertitude élargie peut être calculée à partir de la combinaison des incertitudes élargies relatives à chacune des grandeurs d’entrée, ce qui était le mode opératoire de la formule controversée de Welch-Satterthwaite.

Dans le cas de la fonction de mesure simplifiée $Y = X + Z$ que nous avons considérée tout au long des deux derniers chapitres, la distribution de probabilité de la variable aléatoire $\hat{y} = \hat{\chi} + \hat{\zeta}$ est le produit de convolution⁵⁵ des distributions de probabilité des variables aléatoires $\hat{\chi}$ et $\hat{\zeta}$ (voir figure 4.4). Notons que, de manière générale, une telle opération de « propagation des distributions » peut se révéler mathématiquement très complexe, voire impossible à résoudre analytiquement. Cela explique la grande importance accordée dans le supplément 1 aux méthodes de calcul numériques, rendues possibles par l’accroissement récent de la puissance des ordinateurs. La question de la propagation des incertitudes élargies est alors elle-même résolue : il ne reste plus qu’à calculer l’incertitude élargie à partir de la distribution de probabilité

55. Joint Committee for Guides in Metrology (JCGM) (2009), p.7.

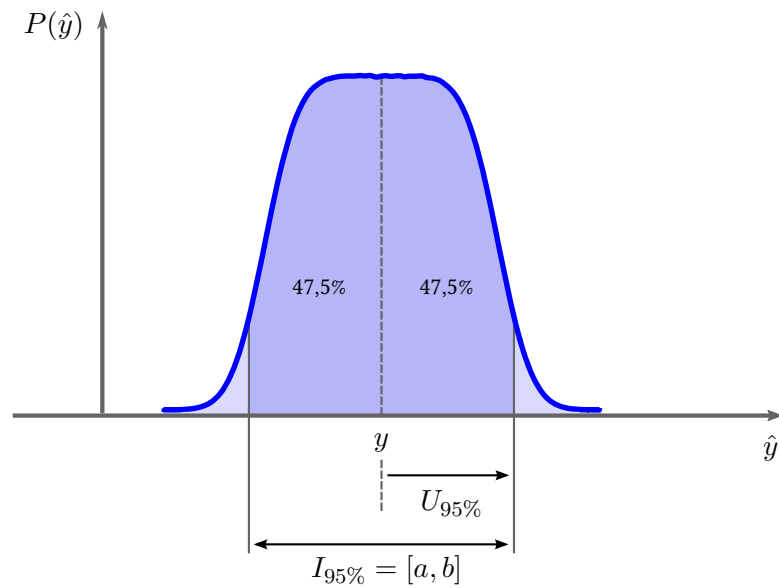


FIGURE 4.5 – Intervalle de crédibilité et incertitude élargie à 95% obtenus pour une fonction de de mesure $Y = X + Z$, où $\hat{\chi}$ est une distribution de Student et $\hat{\zeta}$ une fonction rectangulaire. La courbe a été produite avec l'aide du logiciel « Uncertainty Machine » fourni par le National Institute of Standards and Technology (NIST, URL : <http://www.nist.gov/itl/sed/gsg/uncertainty.cfm>) en nous appuyant sur des valeurs numériques données dans un exemple développé dans le GUM à la section H.1, et adapté au traitement bayésien du supplément 1. Nous ne présentons ici que l'allure de la courbe.

\hat{y} associée à Y .

Nous en arrivons finalement au constat suivant : dans l'approche bayésienne du supplément 1 du GUM, le résultat de mesure ne se réduit plus à un couple {valeur estimée, incertitude (élargie) de mesure}; plutôt, le résultat *est* une distribution de probabilité, qu'il est ensuite possible de *résumer* par un intervalle ou, de façon équivalente, par une valeur estimée et une incertitude de mesure. L'intervalle d'incertitude bayésien est un « intervalle de crédibilité », caractérisé par une probabilité p (voir figure 4.5). La probabilité p décrit le degré de croyance que tient l'expérimentateur quant à la proposition énonçant que la valeur du mesurande est contenue dans l'intervalle en question.

Chapitre 5

Discussion : les ramifications philosophiques du débat statistique

Nous venons de montrer que la métrologie a été le lieu d'une transformation substantielle dans la façon dont les acteurs interprètent les probabilités qui sont mises à contribution dans les modèles de mesure. On observe une transition d'une approche fréquentiste du traitement des erreurs de mesure, dite approche « traditionnelle » de la mesure en métrologie, vers une approche épistémique, qui emprunte au bayésianisme à partir du début des années 1990. L'approche fréquentiste de la mesure introduit les probabilités pour décrire les *processus* de mesure par leur tendance à aboutir à tel ou tel résultat. Les probabilités prennent pour objet non le résultat de mesure lui-même, qui est fixe, mais le processus de mesure, qui présente une variabilité. Dans l'approche traditionnelle, c'est d'abord l'erreur de mesure qui joue un rôle central, en relation avec un objectif fort : s'approcher au plus près de la « valeur vraie » de la grandeur visée. Ce que l'on désigne aujourd'hui sous le terme générique d'« incertitude de mesure » doit alors être compris comme une *exactitude de mesure*, qui est définie comme la proximité du résultat avec la valeur vraie. L'une des caractéristiques notables de l'approche traditionnelle est que les erreurs systématiques ne peuvent pas être décrites au moyen de probabilités, et que, par conséquent, un constat d'exactitude fait intervenir au minimum deux composantes différentes, l'une obtenue au moyen du traitement statistique des erreurs aléatoires, l'autre au moyen d'un encadrement des limites des erreurs systématiques. Ce point a été perçu comme une limite de l'approche traditionnelle, et l'approche bayésienne récemment promue dans la littérature métrologique propose une résolution du problème en instituant l'emploi de probabilités épistémiques qui permettent le traitement unifié des erreurs aléatoires et systématiques. L'approche épistémique est centrée autour d'une acception épistémique de l'incertitude de mesure, qui diffère de l'exactitude de mesure en ce qu'elle caractérise un niveau de croyance (ou de connaissance) porté *par le sujet connaissant* sur la grandeur qu'il mesure. L'erreur de mesure, et par extension l'exactitude de mesure, sont mises en retrait : comme elles sont inconnues, elles ne peuvent pas faire partie du vocabulaire avec lequel le résultat final est formulé.

Les deux modèles se distinguent donc d'abord par leur mode de *représentation*. Les probabilités qui y sont employées représentent des objets différents, qui ne font pas partie du même

domaine – d'un côté, des événements physiques, de l'autre côté, des degrés de croyance. Mais la différence ne tient pas au seul mode de représentation, et l'adhésion à l'une ou l'autre des approches vient le plus souvent avec un positionnement philosophique bien précis, qui implique différentes conceptions de la fonction de la mesure, de la nature des résultats de mesure, de l'interprétation de l'incertitude de mesure, et de la perception de l'activité scientifique en général. Les ramifications philosophiques de ce débat, qui se déroule essentiellement à l'intérieur de la sphère métrologique, que nous souhaitons développer dans le présent chapitre. Le point de départ de notre questionnement se trouve dans cette question simple : quels sont les enjeux attachés à la résolution du débat et à l'engagement envers l'une ou l'autre des méthodes ?

Nous montrons d'abord que si les modèles statistiques diffèrent par les concepts employés, ils peuvent également se distinguer par leurs conséquences concrètes. De ce fait, le débat ne peut pas être réduit à une simple querelle de spécialistes. En revanche, la façon dont les conséquences des modèles diffèrent engage déjà des questionnements d'ordre philosophique, parce que les zones de désaccord ont des racines épistémologiques visibles. En particulier, la question du rapport à l'objectivité et à la subjectivité transparait immédiatement dès lors que l'on s'intéresse à un élément pivot du modèle bayésien, à savoir la distribution *a priori* de probabilité. L'ensemble de ces remarques fait l'objet de la section 5.1.

Dans un second temps, en interrogeant les conséquences philosophiques du débat, nous constatons dans la section 5.2 qu'il accompagne un mouvement de fond qui traverse la métrologie au cours du XX^e siècle, et que nous qualifions de « tournant épistémique ». Celui-ci se traduit en particulier par un déplacement de l'objectif de la mesure, qui perd son statut de mode d'analyse *objectif* du monde physique, et qui devient une entreprise de connaissance, tournée vers la prise de décision rationnelle en accord avec des buts spécifiques à chaque activité. Nous étudions ce tournant épistémique à travers le prisme de l'évolution du statut de l'erreur de mesure et de l'incertitude de mesure, en regard de la conception traditionnelle d'exactitude de mesure, et de l'acception *épistémique* de l'incertitude de mesure dans la métrologie bayésienne.

Dans la section 5.3, nous nous appliquons alors à délimiter les contours d'une source majeure de discordance entre les tenants de chacune des approches : le caractère prétendument objectif ou subjectif de la mesure, et, par extension, d'un résultat de mesure. Nous explicitons d'abord très brièvement le rapport entre bayésianisme et subjectivité en métrologie, puis nous développons plusieurs points qui prêtent le flanc à une critique de la part des défenseurs d'un fréquentisme objectif, ainsi que les réponses que les bayésiens apportent le plus souvent à ces critiques. Cela nous amène à discuter du concept hautement controversé de distribution *a priori* de probabilité, qui constitue un sujet de discussion très classique de la philosophie des probabilités et des statistiques, bien au-delà de la seule sphère métrologique. Un deuxième point porte sur le rapport des probabilités bayésiennes à la vérité : selon les fréquentistes, du fait qu'elles formulent une proposition sur des états mentaux des expérimentateurs, les probabilités bayésiennes ne peuvent pas être l'objet d'un test empirique, et, par conséquent, ne peuvent pas être mises en *erreur*. L'attachement à l'erreur apparaît en fin de compte comme l'un des nœuds de l'opposition entre fréquentisme et bayésianisme.

Nous aboutissons, dans la continuité, à une discussion de la façon dont les fréquentistes conçoivent quant à eux le test empirique des procédures fréquentistes. Certains fréquentistes

en sont venus à redéfinir la philosophie de l'approche fréquentiste en réaction aux critiques qui ont été portées de la part des bayésiens. Ce faisant, ils ont repris l'idéal d'exactitude, caractéristique de l'approche traditionnelle, en le reformulant à partir du concept de « taux de succès » d'une procédure, qui fournit un critère de performance qu'ils estiment testable. Ce critère est exposé à la section 5.4. Le débat métrologique fait alors émerger une caractéristique remarquable de la position fréquentiste, à savoir qu'elle engage elle aussi des croyances, mais qu'elle ne les introduit pas au même niveau que le fait le modèle bayésien. Ce dernier point permet de mieux comprendre en quoi les fréquentistes s'estiment capables de démêler des difficultés apparemment insurmontables, liées en particulier au caractère idéal de l'objectivité en science. Ces remarques dévoilent de plus une dernière différence entre les deux approches, en ce que le modèle fréquentiste est une approche tournée vers un test *futur*, alors que la conception bayésienne est une approche tournée vers l'exposition d'un état *présent* de connaissances.

5.1 Les deux modèles statistiques amènent-ils à des conclusions similaires ?

Les approches fréquentiste et bayésienne fournissent deux méthodes différentes pour analyser les données expérimentales et les traduire en un résultat de mesure. Nous avons vu qu'elles s'appuient sur des interprétations différentes des probabilités, et que, par conséquent, elles ne mobilisent pas le même vocabulaire, et ne mènent pas à une même interprétation ni à une même formulation des résultats de mesure. Comme le montre le chapitre 3, l'approche traditionnelle ne permet pas de formuler le résultat au travers d'une composante unique d'incertitude ; elle rend compte des erreurs aléatoires et des erreurs systématiques de façon séparée. À l'inverse, le chapitre 4 montre comment l'approche bayésienne propose de rendre compte des deux types d'erreur de façon conjointe. Or, jusqu'ici, les différences revendiquées se limitent à des questions de langage ou d'interprétation. S'il s'agit en quelque sorte de dire la même chose avec des mots différents, l'on pourrait être amené à penser que l'opposition entre les deux approches se réduit à une simple querelle philosophique sans grande conséquence pour la pratique elle-même. De ce fait, nous sommes amenés à nous demander si ce sont là les seuls éléments de démarcation entre les deux approches, ou si celles-ci peuvent amener à des conclusions différentes lorsqu'elles sont appliquées à une même situation. En particulier, si l'on considère une même expérience, et si, par conséquent, les données expérimentales et les valeurs théoriques des paramètres du modèle sont les mêmes, la valeur finale attribuée au mesurande et l'incertitude finale associée sont-elles identiques dans les deux cas ? Autrement dit, les intervalles d'incertitude obtenus sont-ils les mêmes dans les deux cas ?

Il est difficile de répondre à cette question sans revenir une fois encore aux questions d'interprétation et de formulation. De fait, les résultats des deux approches sont formulés différemment, ce qui rend difficile leur confrontation : un énoncé d'exactitude, qui inclut un intervalle de confiance fréquentiste auquel s'ajoute une seconde composante, typiquement un encadrement de l'erreur systématique, ne peut pas aisément être comparé à un intervalle de crédibilité bayésien. Les résultats *sont* différents, non pas tant parce que les conclusions qu'ils entraînent sont dissemblables, mais plutôt parce qu'ils ne sont pas conceptualisés de la même manière. Nous aimerions pourtant aller un peu plus loin et nous interroger sur les conséquences concrètes

qu'engage l'adoption de l'une ou l'autre des méthodes. Pourrait-on être amené à conclure différemment, ou à prendre des décisions contraires, en fonction de l'approche employée, pour une même situation de départ ? Deux points en particulier peuvent être creusés de façon plus aboutie. Si l'on se restreint au traitement des erreurs aléatoires – c'est-à-dire des méthodes de « type A » du GUM – on peut assez facilement comparer les résultats. Par ailleurs, les méthodes employées ont une autre conséquence pratique, qui porte sur la façon dont les incertitudes de mesure sont « propagées » d'une grandeur à une autre. Dans la suite de cette section, nous traitons successivement ces deux points.

Si l'on s'intéresse au traitement des erreurs aléatoires, on peut montrer que les approches n'aboutissent pas automatiquement aux mêmes résultats, mais que sous certaines conditions, les deux approches concordent. Notre description du traitement bayésien des erreurs aléatoires à la section 4.5.1 montre que celui-ci fait intervenir le choix d'une distribution *a priori* de probabilité décrivant l'état initial de connaissance de l'expérimentateur à propos de la valeur de la grandeur visée. Sans surprise, le résultat final dépend du choix de cette distribution de probabilité *a priori* (voir figure 5.1 p.133). L'on peut alors se demander (i) s'il existe un *prior* particulier pour lequel les résultats fréquentiste et bayésien coïncident et (ii) si ce *prior*, le cas échéant, constitue le seul choix acceptable. À la première question, Kacker et Jones répondent par l'affirmative :

Il s'avère que les estimations tirées d'une analyse statistique classique sont soit égales soit approximativement égales aux estimations correspondantes tirées d'une analyse bayésienne avec des distributions de probabilité a priori non-informatives¹

Pour des *prior* non informatifs, les résultats fréquentiste et bayésien convergent. À l'inverse, pour des *prior* informatifs, il n'y a plus coïncidence entre les résultats. Faut-il alors s'en tenir à

1. Kacker et Jones (2003), p.235. Ce constat peut être illustré en confrontant les résultats classiques de métrologie que nous retrouvons dans les annexes A.2 et A.3. Dans l'annexe A.3, nous développons sur un exemple les calculs qu'impliquent le traitement bayésien des méthodes de type A. Nous montrons que pour des *prior* non informatifs, c'est-à-dire uniformes sur l'ensemble des valeurs possibles de la grandeur, la distribution de probabilité *a posteriori* est une distribution de Student à $n - 1$ degrés de liberté (où n est le nombre de mesurages effectués), centrée sur la moyenne arithmétique des données expérimentales. Par ailleurs, nous présentons à l'annexe A.2 l'équivalent fréquentiste de ce calcul d'incertitude et nous montrons que l'on aboutit également à représenter le dispositif de mesure par une variable aléatoire dont la distribution de probabilité est une distribution de Student à $n - 1$ degrés de liberté, centrée sur la moyenne arithmétique des données expérimentales. Rappelons une fois encore que les variables aléatoires du cas fréquentiste et du cas bayésien ne sont pas interprétées de la même façon, et, en particulier, ne portent pas sur les mêmes objets. Cependant, en fin de compte, les distributions obtenues étant identiques – elles sont décrites par la même fonction mathématique – l'intervalle d'incertitude qui leur correspond sera également identique dans les deux cas. Cela rejoint le constat de Kacker et Jones. Notons toutefois que les incertitudes-types obtenues selon les deux méthodes ne sont pas identiques, comme le soulignent Kacker et Jones eux-mêmes (Kacker et Jones, 2003, p.241) ou encore Bich (Bich, 2012a, p.2157). Nous renvoyons aux équations (A.6) et (A.25) de l'annexe pour confronter les valeurs respectives des incertitudes-types produites par les deux méthodes. Cette différence n'est pas sans conséquence, en particulier pour les laboratoires qui appliquent les méthodes recommandées de façon routinière, et qui verront leurs incertitudes-types augmenter s'ils optent dans le futur pour l'approche bayésienne. Elle est plus anodine sur le plan conceptuel, car elle apparaît en définitive comme un artifice mathématique plus que comme un élément conceptuellement significatif. En effet, la différence entre les incertitudes-types déterminées selon les deux approches provient du fait que ce n'est pas le même écart-type qui est calculé dans les deux cas ; pour autant, l'incertitude élargie obtenue en fin de compte est quant à elle invariante par changement de méthode.

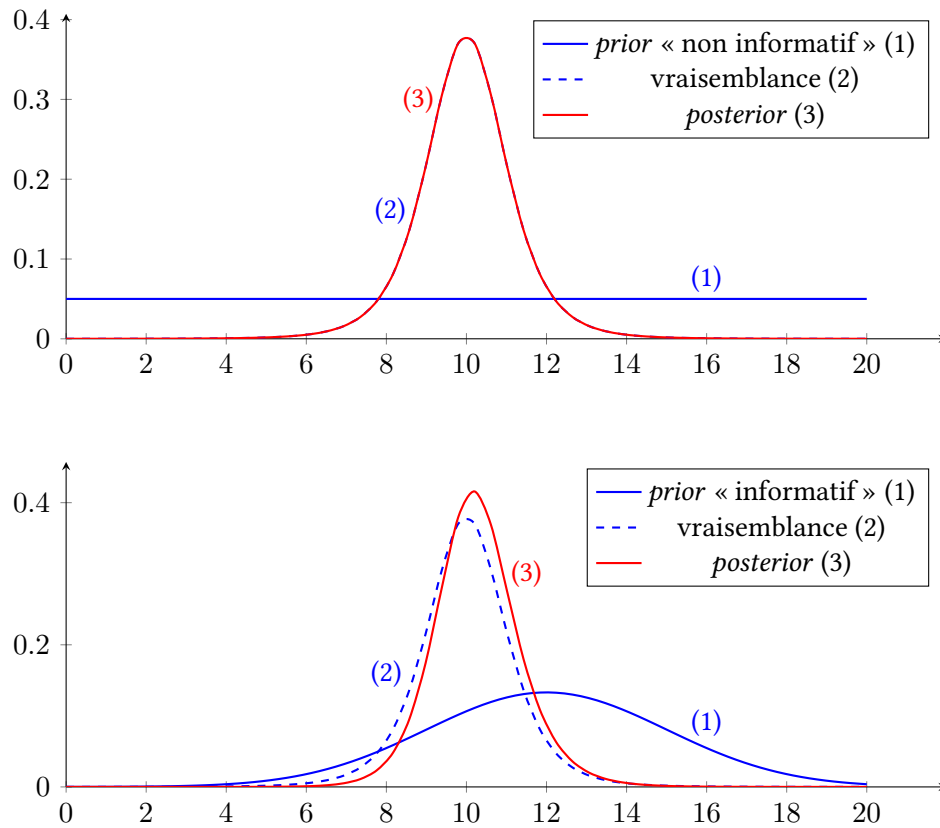


FIGURE 5.1 – Illustration sur un exemple de l’influence du choix de la distribution de probabilité *a priori* sur la distribution *a posteriori* obtenue, dans le cas des méthodes bayésiennes de type A. Les courbes représentées correspondent à l’exemple traité à l’annexe A.3. Dans les deux cas représentés, la fonction de vraisemblance est la même, mais les deux *prior* sont différents : l’un est « non informatif » (uniforme; figure du haut), l’autre est informatif (ici gaussien, pour l’exemple; figure du bas). Dans le cas du *prior non informatif*, le *posterior* s’identifie à la vraisemblance (les courbes (1) et (2) sont superposées, figure du haut). On peut observer l’influence du prior dans le second cas, où la distribution de probabilité *a posteriori* est légèrement décalée vers la droite; elle diffère de la vraisemblance, et, par conséquent, du *posterior* obtenu dans le premier cas : le résultat final est donc dépendant de la forme du *prior*, ce qui était attendu.

ce seul type de *prior* ? Pour Kacker et Jones, qui défendent le point de vue bayésien, ce constat montre d'abord que les méthodes fréquentistes restent acceptables à la condition d'être interprétées selon la perspective bayésienne². Selon le point de vue inverse, on peut interpréter ce constat comme étant le signe que les méthodes bayésiennes ne sont acceptables que si le *prior* est le moins spécifique possible, c'est-à-dire, selon une interprétation courante, le plus *objectif* possible. C'est une position qui est partagée par certains bayésiens en métrologie, et qui revient à adhérer à la forme objective du bayésianisme : nous reviendrons sur ce point à la section 5.3.

La remarque précédente établit que, à condition que soit respectée une hypothèse bien spécifique, celle d'un *prior* non informatif, les résultats des deux approches sont identiques ou approximativement égaux. Ce constat spécifique est valable quel que soit le nombre de mesurages effectués. On peut à l'inverse montrer que les résultats fréquentiste et bayésien convergent, sans hypothèse sur la nature du *prior*, mais à condition que le nombre de mesurages effectués soit grand ; c'est là la répercussion d'une propriété mathématique classique des statistiques bayésiennes³. Par conséquent, dans le cas où le *prior* est informatif, les deux approches se distinguent essentiellement pour un faible nombre de mesures, c'est-à-dire quand on dispose de peu d'informations empiriques, et quand le poids des connaissances préalables est, du même coup, susceptible d'être plus important. Or, de façon notable, l'approche fréquentiste n'est pas très satisfaisante lorsqu'il n'est pas possible de multiplier les mesures⁴. Bien entendu, dans ces situations, la méthode la plus efficace consiste à augmenter le nombre de mesurages de façon à pallier ce déficit d'information, mais cela n'est pas toujours possible, car la répétition d'une mesure présente un coût (par exemple dans le domaine industriel) ou peut être limitée par des contraintes temporelles (on peut ainsi penser à l'expédition menée par Arthur Eddington à Sao-Tomé-et-Principe, en mai 1919, où le groupe devait observer les positions des étoiles durant une éclipse de Soleil et par mauvais temps, ce qui offrait une fenêtre temporelle très brève). C'est pourquoi les défenseurs de l'approche bayésienne affirment que celle-ci se distingue par son adaptabilité, et que sa robustesse dans le cas d'un faible de nombre de mesures constitue un argument de poids en sa faveur. À l'inverse, les fréquentistes estiment que cela ne fait que rendre le résultat plus dépendant encore du choix du *prior*, et donc de la subjectivité de l'expérimentateur. La discussion nous renvoie donc invariablement à la question de la subjectivité de l'approche bayésienne, sur laquelle nous reviendrons à la section 5.3.

La discussion précédente montre que le choix du *prior* influence la façon dont les deux méthodes sont amenées ou non à aboutir à des résultats identiques. Ce constat est tout à fait logique si l'on se rappelle que les deux interprétations des probabilités sont assises sur les mêmes axiomes, et donc sur un même calcul des probabilités, et que l'existence d'une distribu-

2. « Ainsi, les estimations classiques (fréquentistes) peuvent être utilisées, à conditions qu'elles soient interprétées du point de vue bayésien », Kacker et Jones (2003), p.235.

3. Voir par exemple Demeyer (2011), p.52. Demeyer conclut ainsi : « le mécanisme bayésien permet de s'adapter aux différents cas rencontrés en pratique, en jouant sur la variabilité de l'information *a priori*. En particulier, un petit nombre d'observations peut être compensé par une croyance plus forte dans l'*a priori*. Réciproquement, l'information *a priori* s'efface en présence d'un grand nombre de données suffisamment précises », p.53.

4. « Un exemple de la difficulté de la théorie fréquentiste de l'inférence dans l'approche du GUM est que la théorie fréquentiste ne peut pas être utilisée pour attester de l'incertitude d'une valeur mesurée unique en utilisant un instrument de mesure tel qu'un voltmètre », Ehrlich, Dybkaer et Wöger (2007), p.213.

tion *a priori* dans le cas bayésien introduit une différence entre les deux approches sur le plan technique. Il ne s'agit pas pour autant de la seule différence. Nous pouvons identifier un second point technique en vertu duquel le choix de l'approche a une influence notable sur le résultat final. Pour cela, il faut se pencher sur la façon dont chaque approche effectue la « propagation » des incertitudes de mesure. Considérons une fonction de mesure générique :

$$Y = f(X_1, \dots, X_p) \quad (5.1)$$

Rappelons que la propagation des incertitudes est l'opération consistant à calculer l'incertitude sur la grandeur de sortie Y à partir des incertitudes respectives des grandeurs d'entrée X_i . Dans le cas fréquentiste, la méthode consiste à opérer en deux temps, en appliquant séparément la fonction (5.1) aux valeurs des grandeurs et à leurs incertitudes. Ainsi, on applique d'abord la fonction de mesure aux valeurs attribuées aux grandeurs :

$$y = f(x_1, x_2, \dots, x_p) \quad (5.2)$$

où les x_i sont les valeurs attribuées respectivement aux X_i , une fois leurs mesures respectives effectuées, et y est la valeur attribuée à Y , obtenue par calcul à partir des x_i . L'équation (5.2) est la première étape du calcul traditionnel. Reste alors à déterminer l'incertitude sur y . Pour cela, on peut appliquer la loi dite de « propagation des incertitudes », dont nous avons donné une forme simplifiée à l'équation (3.12) du chapitre 3, et qui s'écrit de façon plus générale⁵ :

$$u(y)^2 = \sum_{i=1}^p \left(\frac{\partial f}{\partial X_i}(x_i) \right)^2 \times u(x_i)^2 \quad (5.3)$$

L'on aboutit ainsi séparément à y et $u(y)$. Dans le cas bayésien, le calcul est effectué en une étape unique, en faisant appel à ce qui est appelé la « propagation des distributions » (décrite à la section 4.5.2) : la fonction f est directement appliquée aux *variables aléatoires* $\hat{\chi}_i$ du modèle bayésien, pour calculer la variable aléatoire \hat{y} :

$$\hat{y} = f(\hat{\chi}_1, \dots, \hat{\chi}_p) \quad (5.4)$$

Comme nous l'avons établi au chapitre 4, le résultat de mesure prend alors la forme d'une distribution de probabilité. On peut malgré tout le résumer par une valeur centrale et une incertitude, en les dérivant toutes deux de la distribution obtenue.

Du fait des deux modes de calcul différents, il n'est pas acquis que l'incertitude obtenue soit identique dans les deux cas. Willink, en particulier, a souligné avec insistance le fait que les résultats obtenus divergent dès lors que la fonction de mesure n'est pas linéaire⁶. Supposons par exemple qu'une grandeur Y est définie comme le carré d'une grandeur d'entrée X :

$$Y = X^2 \quad (5.5)$$

5. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.57. La formule que nous donnons ici est une version simplifiée de celle proposée dans le GUM. En effet, elle n'est valable qu'en l'absence de corrélation entre les différentes grandeurs d'entrée. Nous faisons le choix de ne pas prendre en compte de telles corrélations car, bien que ce soit une question importante pour les métrologues, cela n'affecte pas l'allure générale de notre analyse.

6. [Willink \(2010a\)](#)

où X est l'objet de n mesurages répétés. L'approche fréquentiste attribue à Y la valeur suivante :

$$y_{\text{freq}} = \bar{x}^2 \quad (5.6)$$

Le cas bayésien est différent. Typiquement, le traitement bayésien de la grandeur X aboutit à lui associer une variable aléatoire $\hat{\chi}$ centrée sur \bar{x} , c'est-à-dire que l'espérance de $\hat{\chi}$ vaut :

$$E(\hat{\chi}) = \bar{x} \quad (5.7)$$

On calcule ensuite la variable aléatoire décrivant la connaissance sur y :

$$\hat{y} = \hat{\chi}^2 \quad (5.8)$$

La valeur attribuée à Y est l'espérance de \hat{y} :

$$y_{\text{bayes}} = E(\hat{y}) \quad (5.9)$$

Or, on sait d'après les propriétés de l'algèbre des variables aléatoires, que :

$$E(\hat{y}) = E(\hat{\chi}^2) = V(\hat{\chi}) + E^2(\hat{\chi}) = V(\hat{\chi}) + \bar{x}^2 \quad (5.10)$$

C'est-à-dire en fait que :

$$y_{\text{bayes}} = y_{\text{freq}} + V(\hat{\chi}) > y_{\text{freq}} \quad (5.11)$$

L'équation (5.11) montre que dans cet exemple, la valeur centrale obtenue diffère selon la méthode employée. Selon Willink, cela constitue une difficulté notable de l'approche bayésienne : elle est parfois en contradiction avec les approches plus traditionnelles⁷. Il voit dans ce problème la limite d'une conception qui amène à représenter *les grandeurs elles-mêmes* par des variables aléatoires, et à appliquer la fonction de mesure à ces variables aléatoires sans recul critique, alors que celle-ci représente d'abord une relation physique entre grandeurs. C'est une critique de ce que Willink appelle « le concept de mesurande distribué [*the distributed-mesurand concept*] »⁸, à savoir le fait d'associer au *mesurande*, et non au processus de mesure, une distribution de probabilité – ce que fait l'approche bayésienne. La critique de Willink est plus large que ce que le seul exemple étudié ici permet d'évoquer. Elle revient en particulier sur la subjectivité de l'approche bayésienne et met en avant l'impossibilité de tester cette méthode sur un critère de performance concret. Nous développons ces points dans les sections suivantes.

5.2 Un tournant épistémique en métrologie

Nous avons insisté de façon répétée sur le fait que le passage des probabilités fréquentistes aux probabilités épistémiques en métrologie engage un changement dans l'objet des probabilités employées. Cela accompagne une transformation plus générale de la conception de la

7. Willink (2010a), p.8. Willink expose d'abord l'exemple repris ici, puis conclut : « nous suggérons qu'il y a un certain nombre de difficultés fondamentales avec le concept de mesurande distribué. Cela inclut l'idée [...] que [...] le concept de mesurande distribué n'est pas entièrement cohérent avec les autres concepts traditionnellement acceptés ».

8. Willink (2010a), p.2.

mesure en métrologie, qui voit se superposer à ce changement d'objet un changement d'*objectif*, lequel s'observe par exemple au travers de l'évolution de la définition de « mesurage » dans les différentes éditions du VIM. Ainsi le VIM1 voit-il un mesurage comme un « ensemble d'opérations ayant pour but de déterminer la valeur d'une grandeur »⁹ ; dans le VIM3, il s'agit d'un « processus consistant à obtenir expérimentalement une ou plusieurs valeurs que l'on peut raisonnablement attribuer à une grandeur »¹⁰. La mesure n'a plus pour but de décrire « ce qui est », c'est-à-dire de fournir un support objectif à la description d'un monde physique – à partir d'une proposition sur la valeur vraie de la grandeur¹¹. Plutôt, elle ne fait que donner à un agent une base rationnelle solide pour croire en certaines propositions formulées à propos du monde physique. L'évolution de la façon dont est comprise la visée de la mesure dépasse la simple adhésion à une interprétation épistémique des probabilités, qui n'est qu'un outil bien adapté à une telle conception : elle permet de réaliser naturellement l'objectif de décrire un état de connaissance, en quantifiant les croyances rationnelles par des probabilités. Nous proposons de qualifier de « tournant épistémique » l'évolution observée en métrologie durant la seconde moitié du XX^e siècle. L'approche bayésienne s'inscrit dans ce tournant épistémique.

Luca Mari a proposé une description de ce tournant épistémique dans son esquisse d'une histoire des différentes conceptions de la mesure ayant eu cours durant les derniers siècles. Il estime que cette dernière peut être décomposée en trois grandes périodes. Vient d'abord ce qu'il appelle la « période métaphysique », que Mari fait remonter à Galilée et Kepler (et qui puise dans la tradition pythagoricienne), dans laquelle la position caractéristique est que « les mesures sont des propriétés inhérentes des objets mesurés »¹². Suivent ensuite une « période anti-réaliste » puis une « période relativiste »¹³. Dans cette dernière, la mesure est avant tout comprise comme une opération *utile* en réponse à un objectif donné : « les mesures sont les résultats d'opérations reconnues comme adéquates par rapport à leur objectif, à savoir d'obtenir des informations sur les objets mesurés »¹⁴. Selon Mari, cela manifeste une tension entre deux conceptions de la science : la science considérée comme idéal de représentation d'une réalité indépendante de l'esprit, et la science considérée comme activité instrumentale orientée vers une application ou un but concret donné : « l'ontologie et le critère de vérité ont été remplacés par l'information et un critère d'adéquation »¹⁵. Mari et son collègue Alessandro Giordani situent les causes de ce mouvement dans des thèses désormais classiques, issues de la philosophie et de la sociologie des sciences¹⁶, qui remettent en question la notion de « donnée pure » – telle la thèse de la charge théorique des observations. Pour autant, il semble à tout le moins qu'il ne faille pas négliger les enjeux pratiques liés à la mesure. Durant le XIX^e siècle puis le XX^e siècle, celle-ci devient intimement attachée aux progrès technologique et industriel qui entraînent des conséquences économiques, politiques et sociales substantielles, à l'origine d'une intense vague

9. Organisation internationale de normalisation (ISO) (1984), p.11.

10. Joint Committee for Guides in Metrology (JCGM) (2012), p.16.

11. Comme l'exprime et le regrette Grabe, « (l)e GUM ne soulève aucune affirmation quant à la localisation des valeurs vraies des mesurandes », Grabe (2007), p.2.

12. Mari (2003), p.21.

13. Mari (2003), pp.21–24.

14. Mari (2003), p.24.

15. Mari (2003), p.25.

16. Mari et Giordani (2012), p.2145.

d'institutionnalisation et de standardisation des pratiques scientifiques, ce qui provoque *en retour* une démarche réflexive sur les fondements philosophiques des concepts et des méthodes employées. L'histoire du développement des probabilités bayésiennes en métrologie montre que c'est un impératif pratique qui a déclenché un mouvement de fond, lequel a été justifié en passant par une adhésion à une forme de subjectivisme. Même si ce dernier était déjà latent dans les décennies précédentes, l'adhésion assez massive des métrologues nous semble être la conséquence plutôt que la cause des transformations constatées.

Le tournant épistémique se matérialise dans le choix des probabilités employées, et ses conséquences ont des ramifications qui s'expriment tout particulièrement dans la comparaison entre deux façons différentes d'interpréter le concept générique d'« incertitude de mesure » : soit en termes d'exactitude de mesure, soit selon l'acception épistémique qu'en propose l'approche bayésienne. Dans l'approche traditionnelle de la mesure, le résultat est caractérisé par son erreur de mesure (inconnue, que l'on cherche à évaluer) et le processus de mesure est caractérisé par son exactitude, qui traduit sa tendance à produire des erreurs de mesure plus ou moins faibles. Erreur et exactitude sont deux faces de la même idée, qui renvoie à la concordance entre le résultat obtenu et l'état physique de la cible de la mesure. La formulation contemporaine du VIM donne de l'exactitude de mesure la définition suivante : « étroitesse de l'accord entre une valeur mesurée et une valeur vraie d'un mesurande »¹⁷. Une telle définition n'est pas totalement satisfaisante, car elle fait de l'exactitude une caractéristique du résultat de mesure lui-même. Il faut remonter plus loin pour trouver une définition d'« exactitude de mesure » plus en accord avec l'approche traditionnelle. On peut encore une fois se référer à Eisenhart, qui avait particulièrement insisté sur le fait que :

L'erreur systématique, la précision, et l'exactitude sont des caractéristiques inhérentes à un processus de mesure et pas à une mesure particulière générée par le processus.¹⁸

Indépendamment de cette subtilité, le point commun entre la définition du VIM et l'acception traditionnelle de l'exactitude de mesure réside en un point sur lequel Eisenhart insiste à plusieurs reprises, et qui nous semble être au fondement de l'approche traditionnelle : « l'« exactitude » est une question de proximité avec la vérité »¹⁹. Il s'agit donc de dire si le processus de mesure tend ou non à produire des valeurs correctes, en regard de l'état réel de la grandeur mesurée. On voit ainsi que le discours d'Eisenhart – tout comme celui du VIM – donne à la notion d'exactitude une portée métaphysique assez forte et, de façon remarquable, en fait un attribut *indépendant* de la connaissance de l'expérimentateur.

Dans l'approche bayésienne de la mesure, les deux concepts d'erreur et d'exactitude ne sont pas nécessairement tenus pour conceptuellement invalides, mais ils sont *a minima* considérés comme inutiles. En effet, puisqu'ils sont indépendants de la connaissance de l'expérimentateur, c'est qu'on ne peut pas les connaître eux-mêmes. C'est pourquoi l'approche bayésienne contemporaine met l'accent sur le concept d'« incertitude de mesure », compris comme un concept

17. Joint Committee for Guides in Metrology (JCGM) (2012), p.21. Les définitions du VIM1 et VIM2 sont sensiblement similaires.

18. Eisenhart (1963), p.162.

19. Eisenhart (1963), p.172.

pleinement épistémique, qui caractérise non pas l'objet visé lui-même, mais la connaissance que l'expérimentateur a de cet objet. Comme l'exprime Bich, l'incertitude de mesure « peut être vue comme la réciproque logique de l'état de connaissance »²⁰. Exactitude de mesure et incertitude de mesure diffèrent donc par ce qu'elles signifient. La différence entre les deux conceptions de la mesure, traditionnelle et épistémique, se reflète dans les définitions du terme depuis 1970. Ainsi, le National Physical Laboratory formule l'idée d'incertitude de mesure en 1973 de la façon suivante :

[Le signe +/-] indique la plage de valeurs du résultat final dans laquelle on croit que la vraie valeur de la quantité mesurée se trouve.²¹

De façon plus frappante encore, le VIM1 définit « incertitude de mesure » en 1984 comme :

Incertainitude de mesure : estimation caractérisant l'étendue des valeurs dans laquelle se situe la valeur vraie d'une grandeur mesurée.²²

Ces deux formulations contrastent nettement avec la tournure employée en 2009 dans l'un des suppléments du GUM :

L'incertitude de mesure peut ainsi être décrite comme la mesure de la façon dont on croit que l'on connaît la valeur vraie essentiellement unique du mesurande.²³

Ehrlich indique d'ailleurs que les métrologues envisagent de reprendre la formulation du GUM comme base pour définir le terme d'« incertitude de mesure » dans la prochaine édition du VIM (VIM4, en cours de développement)²⁴. Le tableau 5.1 expose les différentes définitions du terme depuis la première édition du VIM jusqu'à la possible prochaine définition que lui donnera le VIM4; il montre en particulier comment la notion de valeur vraie a disparu de la définition pour favoriser une description d'un état de croyance (ou de connaissance).

Les deux notions – l'exactitude de mesure d'une part, et l'acceptation épistémique de l'incertitude de mesure d'autre part – se démarquent donc par ce qu'elles représentent. De façon liée, elles se démarquent également par un autre trait essentiel : leur rapport à la connaissance. L'erreur finale sur une mesure est par définition inconnue : autrement, il suffirait de la corriger. En particulier, *rien* ne permet d'être certain que toutes les erreurs de mesure ont été identifiées et analysées; il se peut tout à fait qu'une erreur inconnue n'ait pas été prise en compte dans l'analyse de données. Dans ce cas, l'exactitude de mesure n'a pas été correctement estimée et, selon l'amplitude de l'erreur en question, la valeur obtenue est susceptible d'être nettement éloignée de la valeur vraie de la grandeur mesurée. Par conséquent, l'exactitude d'une mesure ne peut jamais être garantie. C'est pourquoi erreurs de mesure et exactitude font l'objet d'une *estimation* qui ne peut jamais être qu'imparfaite. En revanche, l'incertitude de mesure décrivant l'état de croyance de l'expérimentateur, peut être exactement *déterminée*. C'est pourquoi on doit distinguer en toute rigueur l'estimation d'une exactitude inconnue et approchée, de la détermination exacte d'une incertitude de mesure, comme le défend en particulier Willink²⁵.

20. Bich (2012a), p.2155.

21. Campion, Burns et Williams (1973), p.1.

22. Organisation internationale de normalisation (ISO) (1984), p.16.

23. Joint Committee for Guides in Metrology (JCGM) (2009), p.3.

24. Ehrlich (2014), p.S150.

25. Willink (2013), p.xvii.

| | |
|----------------------------------|--|
| VIM1 (1984) (p.16) | Incertitude de mesure : estimation caractérisant l'étendue des valeurs dans laquelle se situe la valeur vraie d'une grandeur mesurée. |
|----------------------------------|--|

| | |
|----------------------------------|---|
| VIM2 (1993) (p.25) | Incertitude de mesure : paramètre, associé au résultat d'un mesurage, qui caractérise la dispersion des valeurs qui pourraient raisonnablement être attribuées au mesurande. |
|----------------------------------|---|

| | |
|----------------------------------|--|
| VIM3 (2008) (p.25) | Incertitude de mesure : paramètre non négatif qui caractérise la dispersion des valeurs attribuées à un mesurande, à partir des informations utilisées. |
|----------------------------------|--|

| | |
|------------------------------|--|
| VIM4? (non publié) | Incertitude de mesure : paramètre décrivant la façon dont la valeur vraie (essentiellement unique) du mesurande est tenue pour être connue [<i>is believed to be known</i>] |
|------------------------------|--|

TABLEAU 5.1 – Différentes définitions d'« incertitude de mesure » dans les trois éditions du *Vocabulaire International de Métrologie* (VIM) et définition pressentie pour la prochaine édition, en cours de développement (« VIM4 »). La définition de 1984 fait appel à la notion de valeur vraie et décrit l'incertitude comme étant relative à l'*estimation* de cette valeur vraie. Ce sont là deux traits qui en font une notion encore totalement ancrée dans l'approche traditionnelle. De fait, la notion d'incertitude de mesure du VIM1 est de même nature que l'acception traditionnelle de l'« exactitude de mesure » : elle traduit une proximité avec la vérité, pour reprendre les termes d'Eisenhart. De plus, dans cette définition, l'incertitude de mesure caractérise *la grandeur* elle-même. La présence du terme « valeur vraie » disparaît dès la seconde édition, où l'accent est mis sur l'attribution de valeurs par l'expérimentateur; le conditionnel, cependant, laisse entendre que l'incertitude de mesure n'est pas totalement du domaine de la représentation. Dans la troisième édition du VIM, l'incertitude de mesure décrit exclusivement la façon dont l'expérimentateur représente la grandeur, indépendamment du statut de vérité des valeurs attribuées.

L'une des vertus proclamées de la conception épistémique de la mesure est qu'elle permet de contourner le principal écueil conceptuel, voire philosophique, de la conception traditionnelle : cette dernière rend nécessaire l'évaluation de concepts qu'il est raisonnable de considérer comme à jamais inconnaissables, comme la valeur vraie, l'erreur finale de mesure, ou l'exactitude de mesure. Pour les bayésiens, cet écueil est le signe que les défenseurs de l'approche traditionnelle se trompent de cible en se focalisant sur un idéal illusoire, celui de « vérité », alors que la mesure permet avant tout de produire des connaissances et de prendre des décisions en rapport avec un objectif précis. Ces remarques feront l'objet d'une attention toute particulière dans la prochaine partie, dans laquelle nous défendrons la notion de « valeur vraie » d'une grandeur contre deux critiques portées par les métrologues à son encontre.

En contrepartie vient une difficulté qu'il ne faut pas négliger : oublier l'exigence d'exactitude induit un risque, celui de la remplacer par un simple objectif de consensus, d'accord intersubjectif qui serait alors suffisant pour la prise de décision. Or, il est évident qu'un accord peut être trouvé à propos d'une proposition erronée, et certains utilisateurs, que ce soit dans le domaine légal, industriel, technique, mais aussi dans la recherche scientifique, ne sauraient s'en satisfaire. Comme l'explique Mari : « la mesure est-elle alors juste une évaluation dont les résultats sont subjectivement adéquats à des objectifs donnés ? »²⁶. Si l'on tient l'exactitude de mesure pour une notion importante, il est utile de s'interroger sur les conditions auxquelles on peut considérer que le résultat obtenu, que ce soit selon l'approche fréquentiste ou selon l'approche bayésienne, peut être tenu pour exact. Ce questionnement ne peut pas être résolu en s'intéressant individuellement à un résultat de mesure. Il ne prend corps que si l'on observe la façon dont le résultat en question vient s'insérer dans l'architecture des connaissances et des théories. Nous laissons ici la question en suspens. En effet, la description que nous avons proposée dans cette partie des méthodes d'analyse des données expérimentales en métrologie se limite à la question d'une mesure spécifique d'une grandeur dans des conditions particulières. À la partie III, en prenant pour objet d'analyse les ajustements des constantes de la physique, nous nous intéresserons à la confrontation de résultats issus de mesures différentes d'une même grandeur, menées dans des conditions distinctes et au moyen de principes de mesures variés. Cela nous permettra d'élargir la perspective qui est pour l'instant la nôtre et de revenir sur la question du rapport entre incertitude et exactitude de façon plus élaborée. Nous souhaitons pour l'instant revenir sur un second aspect de la confrontation entre les approches fréquentiste et bayésienne, à savoir l'opposition entre l'idéal objectif prôné par l'approche traditionnelle et le subjectivisme explicitement assumé par de nombreux défenseurs de l'application des méthodes bayésiennes en métrologie.

5.3 Conséquence du tournant épistémique : objectivité et subjectivité de la mesure physique

Nous avons déjà répété à plusieurs occasions que l'un des attributs majeurs de la transition d'une approche fréquentiste à une approche bayésienne de la mesure réside dans le changement d'objet sur lequel portent les probabilités. Le centre d'attention se déplace : alors qu'il

26. Mari (2003), p.25.

était l'*objet* visé dans la conception traditionnelle, il devient, dans l'approche contemporaine, le *sujet* connaissant. L'évolution constatée engage donc de façon directe un questionnement sur la subjectivité ou l'objectivité des résultats de mesure. Elle porte sur la façon dont les scientifiques ont décidé d'intérioriser la subjectivité d'une opération de mesure à l'intérieur même du processus d'analyse et d'expression d'un résultat de mesure. Les subjectivistes arguent ainsi qu'il est préférable, au vu de l'impossibilité de garantir l'exactitude d'un résultat, de parler directement de notre croyance plutôt que de l'exactitude elle-même – c'est cela qu'exprime avant tout la formulation épistémique de l'« incertitude de mesure » que nous avons mise en lumière à la section précédente.

Notons toutefois que ni l'interprétation épistémique des probabilités ni le bayésianisme ne sont automatiquement subjectivistes. Dans sa version la plus radicale, le bayésianisme subjectif stipule qu'il y a autant de probabilités que d'agents²⁷ : les degrés de croyance des agents doivent respecter les axiomes des probabilités, mais les agents ne sont pas tenus de croire la même chose. L'interprétation des probabilités est alors subjective dans le sens où elle renvoie au degré de croyance individuel de chaque agent. Une forme plus modérée de bayésianisme se place au niveau intersubjectif, et les probabilités décrivent alors le degré de croyance commun d'un groupe ayant atteint le consensus²⁸. Cependant, les agents restent libres de croire ou non en une proposition et d'atteindre le consensus. L'on arrive alors au bayésianisme objectif, qui affirme que différents agents, s'ils disposent des mêmes informations, doivent tenir les mêmes croyances, et donc adopter les mêmes probabilités. Ainsi le degré de croyance de l'agent est-il entièrement contraint par les informations dont il dispose ; il s'agit toujours d'une interprétation *épistémique* (elle décrit un degré de croyance justifié par un état de connaissance) mais elle n'est plus *subjective* – il y a en fait une sorte de retour à une conception logique des probabilités, telle que défendue en son temps par Laplace, ou plus tard par Keynes.

Parmi les défenseurs de l'approche bayésienne en métrologie, on peut observer un spectre assez large de positions concernant le degré de subjectivisme à adopter. Dans le GUM, la subjectivité des méthodes de type B est revendiquée sans détour : « pour une évaluation de Type B de l'incertitude-type, [l'incertitude] est une grandeur subjective »²⁹. Certains métrologues semblent embrasser une perspective assez explicitement subjectiviste³⁰. D'autres ne se satisfont pas d'une conception individuelle de la probabilité³¹. Pour autant, tous s'accordent sur le

27. Hájek (2012), section 3.3.1 (entrée consultée le 13 janvier 2016).

28. Gillies (2000), p.179.

29. Joint Committee for Guides in Metrology (JCGM) (2008d), p.77.

30. Bich et Kool écrivent ainsi : « dans les comparaisons de mesures, pour lesquelles différents expérimentateurs donnent pour un même mesurande non seulement des estimations différentes, mais aussi des incertitudes différentes, quelles interprétations peuvent-elles être données à ces dernières ? De l'avis de l'auteur, la réponse réside dans la subjectivité du concept. Chaque expérimentateur déclare à travers l'incertitude *son état personnel de connaissance* sur le mesurande, recueilli au moyen de l'expérience spécifique qu'il a réalisée », Bich et Kool (2012), p.9.

31. Lors de l'entretien qu'il m'a accordé, Wolfgang Wöger, qui avait publié avec Weise l'un des articles fondateurs de l'emploi des statistiques bayésiennes en métrologie, a énoncé sa méfiance envers des conceptions trop individuelles des probabilités, comme celle promue par Anthony O'Hagan (voir par exemple O'Hagan *et al.*, 2006). Pour Wöger, la mesure est effectivement une activité subjective, mais il n'y a aucune raison pour que deux expérimentateurs différents disposant des mêmes informations aboutissent à des probabilités différentes. Les probabilités servent à énoncer un accord intersubjectif et la subjectivité traduit simplement le fait que la mesure est adossée à des hypothèses et à des modèles que l'on ne peut pas prétendre parfaitement objectifs. (Wöger [2014])

fait que l'exigence d'objectivité stricte de la mesure n'est pas tenable, et qu'il est impossible d'éviter un degré minimum de subjectivité. Qu'ils soient attachés à une interprétation individuelle ou collective des probabilités, les bayésiens revendiquent que l'activité de mesure est une activité subjective du simple fait qu'il s'agit d'une activité humaine. La mesure est fondée sur des modèles, qui engagent des hypothèses qui sont eux-mêmes une marque de subjectivité. L'approche fréquentiste ne déroge pas à cette règle : ainsi, par exemple, la représentation du comportement des erreurs aléatoires par une distribution gaussienne n'est pas entièrement déterminé par l'état physique du dispositif mais également par des *choix* pragmatiques et techniques (simplicité, etc.)³².

S'il est difficile de contester que la mesure présente des limites du simple fait qu'il s'agisse d'une activité humaine, les fréquentistes estiment que ce n'est pas suffisant pour basculer entièrement dans une perspective subjectiviste. Le développement des statistiques bayésiennes en métrologie étant récent, on ne retrouvera aucune discussion à ce sujet dans la littérature du fréquentisme traditionnel, par exemple chez Eisenhart ; on trouvera plutôt de telles discussions dans la littérature plus contemporaine, en réaction au mouvement bayésien. Dans la sphère métrologique, le défenseur le plus déterminé de l'approche objective est peut-être le métrologue et statisticien Robin Willink. Si l'on cherche hors de la métrologie, le débat qui oppose fréquentistes et bayésiens a une histoire beaucoup plus riche. Parmi les défenseurs du fréquentisme, on retrouve en particulier la philosophe et statisticienne Deborah Mayo. Les critiques des approches bayésiennes reviennent invariablement sur ce point précis : l'approche épistémique de la mesure est séduisante, mais le subjectivisme qu'elle promeut presque invariablement n'est pas acceptable. Pour eux, l'objectif d'une mesure est bien de fournir des informations sur le monde physique, non sur l'état mental d'un individu. L'objection des fréquentistes à l'égard de la position subjectiviste prend forme sous deux arguments en particulier. Le premier argument porte sur une brique essentielle, hautement controversée, des modèles bayésiens : les distributions de probabilités *a priori*. Le second est orienté vers l'interprétation des probabilités elles-mêmes : parce que celles-ci ne prennent pas pour objet le monde physique, défendent les fréquentistes, elles ne peuvent pas être l'objet d'un test, et donc ne disent rien d'utile.

Parmi les cibles des critiques du subjectivisme bayésien, la notion de probabilité *a priori* se situe en première ligne. Ces probabilités sont construites sur la base des informations préalables dont dispose l'expérimentateur *avant* la mesure à propos de la grandeur visée. Elles sont requises pour enclencher ensuite le processus de révision bayésienne des connaissances une fois la mesure effectuée. Se pose alors la question suivante : comment ces probabilités sont-elles déterminées ? En particulier, existe-t-il une méthode systématique qui guide une telle détermination ? Si la détermination du *prior* est un *choix* laissé à l'appréciation de l'expérimentateur, ce choix est alors effectivement une source de subjectivité. En particulier, les probabilités sont alors spécifiques à un agent particulier, et sont susceptibles de varier d'une personne à l'autre. Comme nous l'avons vu, certains métrologues bayésiens ne se satisfont pas d'une conception personnelle de la probabilité épistémique et envisagent plutôt une conception intersubjective.

32. Wöger [2014]

Wöger et Weise, par exemple, dans leur article séminal de 1992³³, empruntent largement au versant objectiviste du bayésianisme en articulant leur démarche autour du « principe d'entropie maximum ». Ce principe, promu en particulier par Jaynes³⁴ revient, en résumant à grand traits, à choisir le *prior* qui soit à la fois cohérent avec les informations disponibles et qui soit le moins spécifique possible. Par exemple, si l'on sait qu'une valeur est contenue entre les bornes 0 et 1, le principe d'entropie maximum indique qu'il faut choisir un *prior* uniforme sur cet intervalle; si l'on tire à pile ou face sans avoir aucune information sur la pièce, il faut attribuer aux deux issues la même probabilité 1/2. Le principe d'entropie maximum est alors interprété comme la procédure de détermination la plus objective possible, qui ne crée par artificiellement d'information supplémentaire. Il est ainsi une généralisation du « principe d'indifférence » proposé par Keynes, lui-même inspiré de la conception laplacienne de la probabilité comme mesure de l'ignorance. On retrouve également ce principe dans le supplément 1 du GUM, où il est toutefois mobilisé de façon plus périphérique que dans l'article de Wöger et Weise – c'est au théorème de Bayes qu'est accordé un rôle central.

Cependant, le principe d'entropie maximum donne prise à plusieurs critiques. Certains ont relevé des difficultés notables dans sa mise en application, en particulier en vertu des « paradoxes de Bertrand »³⁵ qui montrent que l'on peut aboutir à différents *prior* par un simple changement de présentation du problème. De plus, nombre de bayésiens estiment qu'il n'est pas nécessaire de faire appel à un tel principe visant à rendre objectifs les *prior*. Ils contestent la nécessité d'imposer une telle contrainte et défendent l'intérêt d'une expression certes rationnelle, mais libre, des *prior*³⁶. En fin de compte, Lira et Wöger répondent aux critiques adressées par les fréquentistes en insistant sur le fait qu'il ne faut pas confondre subjectivité et arbitraire :

La détermination du *prior* est peut-être le point le plus controversé des statistiques bayésiennes. Les fréquentistes défendent que cette attribution, étant 'subjective', assujettit l'ensemble du processus d'inférence aux caprices de la personne qui effectue l'évaluation. Cela serait en effet le cas si cette attribution était effectuée arbitrairement.³⁷

Dans la conception de la plupart des métrologues défenseurs de l'approche bayésienne, le subjectivisme que cette approche engage ne revient pas à dire que l'expression d'un résultat de mesure est laissée à la fantaisie de l'expérimentateur. De fait, la défense du subjectivisme en métrologie n'exprime pas l'idée selon laquelle il serait philosophiquement souhaitable d'être subjectiviste, mais plutôt le fait que reconnaître le caractère subjectif des jugements que nous formulons à partir des mesures est en définitive la seule option possible³⁸; la subjectivité est considérée comme inhérente aux résultats de mesure. Dans la sphère métrologique, la subjectivité est souvent associée à la nécessité de faire appel au « jugement » de l'expérimentateur³⁹

33. Weise et Wöger (1993)

34. Jaynes (1968)

35. Voir Hájek (2012), section 3.1 (entrée consultée le 14 janvier 2016).

36. Voir Howson et Urbach (1993), pp.276-288 ou encore D'Agostini (2003), p.81.

37. Lira et Wöger (2006), p.S255.

38. « L'attitude [objectiviste] est basée sur l'illusion que la subjectivité peut être totalement évitée dans la mesure, alors qu'elle en imprègne une grande partie. », Bich (2012a), pp.2155-2156.

39. Voir par exemple cette remarque du GUM : « Les incertitudes déterminées à partir d'observations répétées sont souvent opposées à celles qui sont évaluées par d'autres moyens comme étant "objectives", "statistiquement

Les bayésiens font remarquer que l'approche fréquentiste incorpore elle-même de nombreux éléments de subjectivité à tous les niveaux de l'expérience, tant au niveau du choix des hypothèses que de la construction des « limites crédibles » d'erreur, mais que cette subjectivité est masquée⁴⁰. L'approche bayésienne présente alors, à leurs yeux, l'intérêt d'intégrer ces éléments de subjectivité directement dans le modèle et de pouvoir en rendre compte explicitement. Ainsi Possolo, Toman et Estler écrivent-ils que :

(L)e meilleur plan d'action est d'adopter une approche bayésienne qui *révèle ouvertement, de façon naturelle, toutes les hypothèses et croyances mises en jeu*, et les rend toutes *accessibles à l'examen et à la critique*, comme il sied à toute procédure scientifique.⁴¹

Les auteurs font ici valoir un argument de transparence que l'on retrouve couramment dans les discussions plus générales sur les vertus du bayésianisme, bien au-delà de la seule sphère métrologique⁴².

Ce faisant, nous aboutissons à un second argument qu'opposent les fréquentistes au subjectivisme bayésien. Les fréquentistes ne prétendent nullement qu'il n'y a pas de subjectivité dans l'activité de mesure ; de ce point de vue, ils rejoignent les conclusions des bayésiens. En revanche, ils ne voient pas pourquoi il faudrait embrasser une perspective subjectiviste en abandonnant toute prétention à décrire le plus objectivement possible le monde physique. Si l'objectivité peut effectivement sembler un objectif illusoire, c'est parce qu'elle est avant tout un *idéal*, et il reste possible de mettre en avant des critères d'objectivité, comme le fait de pouvoir procéder à des *tests*. Or, selon Willink, l'intervalle de crédibilité bayésien ne peut pas être confronté à l'expérience, puisque les probabilités impliquées portent sur l'état de croyance de l'expérimentateur et non sur l'objet mesuré lui-même :

Un modèle probabiliste invoqué par un statisticien classique est analogue à l'affirmation qu'une distribution de fréquence approxime les issues possibles de la répétition d'une expérience donnée, *et en principe cette affirmation peut être testée*. Ce modèle fondé sur la fréquence diffère des [modèles subjectifs] [...] qui ne sont des modèles de rien d'autre que d'un état mental de l'orateur, qui à son tour n'a que peu de pertinence pour la question scientifique.⁴³

Une probabilité épistémique peut être soumise à un test, mais ce test ne dira rien sur le monde. En effet, une telle probabilité décrit le degré de croyance d'un agent relativement à une proposition, qui elle-même peut être vraie ou fausse. Or, le statut de vérité de la proposition n'influe en rien sur la légitimité pour un agent de lui attribuer telle ou telle probabilité, puisque cette

rigoureuses», etc. Cela implique à tort qu'elles peuvent être évaluées simplement en appliquant des formules statistiques aux observations et que leur évaluation *ne nécessite pas l'application de jugement*. », [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.63 (nous soulignons).

40. Voir la remarque de Mayo citant Good : « Les bayésiens objecteront qu'il est impossible d'exclure les opinions, et qu'au moins mettent-ils ces dernières en évidence au lieu de les 'cacher sous le tapis', pour paraphraser I. J. Good », [Mayo \(1996\)](#), p.82. Voir aussi D'Agostini : « (c)es méthodes qui sont annoncées comme étant 'objectives' tendent en réalité à cacher les hypothèses sur lesquelles elles sont fondées », [D'Agostini \(2003\)](#), p.79.

41. [Possolo, Toman et Estler \(2009\)](#), p.L1 (nous soulignons).

42. [Drouet \(2016\)](#).

43. [Willink \(2013\)](#), p.38 (nous soulignons).

probabilité décrit *ce que croit* l'agent. Ainsi, un agent attribuant une probabilité de 95% à la proposition que les licornes existent ne peut pas être mis en tort sur la base de l'existence effective ou non des licornes ; seul peut être testé le fait que cela correspond ou non à son degré de croyance effectif, en vérifiant par exemple si l'agent est prêt à parier sur une telle hypothèse. On retrouve une discussion similaire chez Mayo, dans une section intitulée « la pernicieuse subjectivité des probabilités *a priori* »⁴⁴. Mayo pose la question suivante : lorsque différents agents tiennent différentes probabilités, « y a-t-il un moyen de dire qui a raison ? »⁴⁵. Elle cite le bayésien Denis Lindley, qui récuse l'emploi même du terme « avoir raison » et répond que « la théorie bayésienne porte sur la *cohérence*, pas sur le fait d'avoir raison ou d'avoir tort »⁴⁶. Effectivement, constate Mayo,

(I)l n'y a aucune raison de supposer qu'il existe un degré de croyance correct à entretenir. Mes opinions sont mes opinions et les vôtres sont les vôtres. [...] Si "avoir raison" n'a pas de signification, comment puis-je dire que vous êtes dans l'erreur ?⁴⁷

Et Mayo d'aboutir à un point essentiel, en citant Kyburg : « voici ce qui est presque la pierre de touche de l'objectivité : la possibilité d'une erreur »⁴⁸. C'est là exactement ce qui guide les défenseurs de l'approche traditionnelle en métrologie. Si l'impossibilité de confronter une probabilité à l'expérience n'est pas nécessairement problématique pour tous les usages de probabilités, on peut comprendre qu'il pose problème dans le domaine de la métrologie, qui attache traditionnellement la mesure physique à une recherche de la détermination de l'état objectif du monde physique. C'est parce que les fréquentistes sont attachés à la notion d'erreur qu'ils tiennent l'exactitude de mesure comme un critère essentiel d'évaluation d'un processus de mesure – et qu'on concept épistémique comme celui d'incertitude de mesure – dans son acception contemporaine – leur semble insuffisant. Pour répondre à l'évolution récente du vocabulaire de la métrologie, et au remplacement progressif de l'exactitude de mesure par l'incertitude de mesure, certains fréquentistes ont cherché à développer un critère de performance qui vise à donner un ancrage empirique à l'exactitude de mesure. Nous développons ce critère dans la section suivante.

5.4 De l'exactitude au taux de succès : une reformulation des objectifs de l'approche fréquentiste

Prenant acte des critiques formulées à l'encontre de l'idée d'exactitude, en particulier du fait qu'il n'est jamais possible de connaître avec certitude l'erreur de mesure affectant un résultat donné, certains défenseurs de l'approche fréquentistes ont récemment mis en avant un nouvel indicateur de performance, le « taux de succès » d'une procédure⁴⁹. Il s'agit d'un concept statistique qui caractérise la tendance d'une procédure de mesure à évaluer avec succès la valeur

44. Mayo (1996), p.83.

45. Mayo (1996), p.83.

46. Lindley cité par Mayo, Mayo (1996), p.83.

47. Mayo (1996), p.83

48. Mayo citant Kyburg, Mayo (1996), p.83. La citation est tirée de Kyburg (1992b), p.147.

49. Voir par exemple Willink (2010a), p.6.

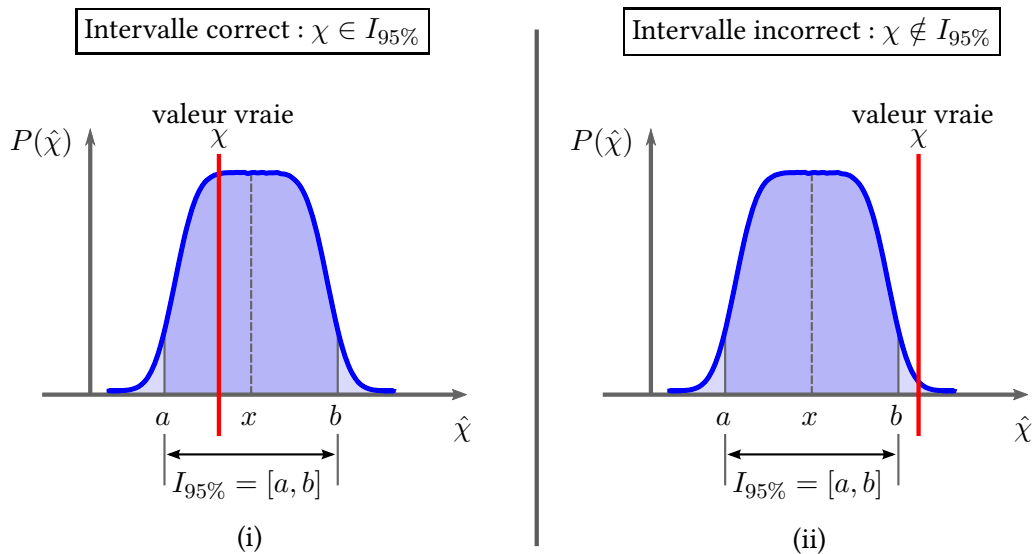


FIGURE 5.2 – Illustration du « taux de succès » d'une procédure. Pour un intervalle d'incertitude donné, deux options sont possibles. (i) L'intervalle contient la valeur vraie de la grandeur visée. On parle dans ce cas d'un « succès ». (ii) L'intervalle ne contient pas la valeur vraie de la grandeur visée. On parle alors d'un « échec ». Un processus de mesure consiste à répéter n mesurages d'une grandeur, et à construire l'intervalle d'incertitude associé aux données expérimentales obtenues. Si ce processus de mesure est lui-même répété, les intervalles d'incertitude obtenus différeront en fonction des cas. Par conséquent, un même processus de mesure peut aboutir parfois à des succès et parfois à des échecs. Le « taux de succès » du processus de mesure est défini par la fréquence limite avec laquelle celui-ci aboutirait à un succès s'il était répété à l'infini. (Précisons que les intervalles d'incertitude présentés ici sont, pour l'exemple, des intervalles de crédibilité bayésiens).

vraie de la grandeur visée. Nous avons vu que l'expression d'un résultat de mesure sous forme probabiliste passe par l'écriture d'un intervalle d'incertitude, qui peut prendre plusieurs formes. Cela peut être un intervalle de confiance fréquentiste, un intervalle de crédibilité bayésien, ou encore l'intervalle dit « élargi » qui est présenté dans le GUM et qui prend une forme hybride empruntant aux deux intervalles précédents. Quelle que soit la nature de l'intervalle, deux options – et deux options seulement – sont possibles. Il se peut (i) que l'intervalle contienne la valeur vraie de la grandeur visée. Dans ce cas, l'intervalle est correct : on parle d'un « succès ». Il se peut autrement (ii) que l'intervalle ne contienne pas la valeur vraie de la grandeur visée. Dans ce cas, on parle d'un échec (voir figure 5.2). Un même processus de mesure, lorsqu'il est répété, ne produit pas à chaque fois le même intervalle d'incertitude, en raison de la variabilité des résultats expérimentaux. Par conséquent, il est susceptible d'aboutir parfois à un succès et d'autres fois à un échec. Dès lors, on peut définir le taux de succès d'un processus de mesure donné de la façon suivante : il s'agit de la fréquence limite avec laquelle le processus aboutirait à un succès s'il était répété à l'infini. De par la définition même du taux de succès et son appel à des fréquences *limites*, celui-ci apparaît immédiatement comme un concept fréquentiste ; de

par son appel aux valeurs vraies, il s'inscrit également dans une tradition objectiviste.

Rappelons, sur la base de ce que les deux chapitres précédents ont montré, que l'intervalle de probabilité effectivement produit est associé à une probabilité p , cette probabilité pouvant être un niveau de confiance ou de crédibilité, ou une « probabilité élargie », selon l'approche adoptée. Ce qu'affirme Willink, c'est que le taux de succès d'une procédure de mesure caractérise d'une part la performance *effective* de la procédure, et que la probabilité associée à l'intervalle d'incertitude correspond à la performance *attendue* de la procédure⁵⁰. Willink propose alors le critère d'évaluation suivant : il n'est acceptable d'associer une probabilité p à l'intervalle d'incertitude concerné que si celle-ci est au moins égale au taux de succès de la procédure par laquelle l'intervalle a été obtenu⁵¹. Il s'agit là d'un *test* des méthodes statistiques de détermination des intervalles d'incertitude : si une méthode statistique produit des intervalles auxquels est associée une probabilité p , alors même que le taux de succès de la méthode est inférieure à cette probabilité, c'est que la méthode statistique employée n'est pas satisfaisante.

Les fréquentistes reconnaissent bien entendu qu'il n'est pas possible, en général, de déterminer si un intervalle est un succès ou un échec, pour la raison maintes fois rappelée qu'il faudrait pour cela connaître la valeur vraie de la grandeur visée, alors même que celle-ci est inconnue, puisque la mesure effectuée a précisément pour but de l'évaluer. Il est donc impossible, en général, de déterminer le taux de succès effectif d'une procédure. Cependant, il demeure possible de procéder par étalonnage, en testant les différentes méthodes sur des grandeurs considérées comme bien connues ; il est également possible d'évaluer ces méthodes au moyen de simulations numériques, en les testant sur des cas fictifs dans lesquels la valeur vraie de la grandeur mesure est fixée à l'avance. Ces simulations visant au test du taux de succès des procédures fréquentiste et bayésiennes ont fait l'objet de nombreuses discussions, en particulier dans les pages de la revue *Metrologia* où certains défenseurs de l'approche fréquentiste, parmi lesquels Willink, ont cherché à dévoiler des exemples de situations où l'approche bayésienne – en particulier celle développée dans le supplément 1 du GUM – est mise en échec⁵², ce à quoi

50. « Supposons qu'un client reçoive un intervalle spécifique $[a, b]$ comme un intervalle d'incertitude à 95% pour un mesurande Y . [...] le client est muni de la proposition " $a \leq Y \leq b$ " et de l'affirmation selon laquelle cette proposition a une probabilité 95% d'être vraie. Une conséquence naturelle de cela est que le client s'attendra à ce que dans 95% d'occasions indépendantes à laquelle ils recevrait une telle proposition, la proposition en question serait vraie. », Willink (2010a), p.6.

51. Il nous faut ici ouvrir une parenthèse, et une parenthèse seulement, à propos d'un point qui, en toute rigueur, ne relève pas simplement du détail. En effet, dans notre description de l'approche traditionnelle, nous avons répété à l'envi que le modèle fréquentiste ne permet *pas* de formuler un résultat par un seul constat probabiliste, car il faut y adjoindre un traitement différent des erreurs systématiques. Par conséquent, la discussion que nous menons ici semble impossible à appliquer à l'approche traditionnelle. De fait, Willink s'est efforcé de moderniser l'approche traditionnelle de façon à la rendre compatible avec l'exigence contemporaine d'une composante unique d'incertitude, régie par un modèle entièrement probabiliste. Il procède pour cela à une extension des conditions d'application du fréquentisme, de façon à concevoir les erreurs systématiques elles-aussi comme le produit d'un tirage aléatoire parmi une population statistique parente dont on pourrait décrire la distribution de probabilité. L'extension que propose Willink, à défaut d'être effectivement appliquée en métrologie, conduit à prendre en compte la nature collective de l'activité scientifique, et permet d'intégrer à l'analyse la composante *sociale* de la mesure. Nous ne pouvons pas développer ici ce point, même si nous y reviendrons, à partir d'un objet différent, à la partie III. Nous renvoyons à Grégis et de Courtenay [2016] (à paraître) pour une discussion détaillée de ce sujet.

52. Hall (2008); Wang et Iyer (2009); Willink (2010b); Willink (2010c).

leurs homologues bayésiens se sont attachés à répondre⁵³. Les fréquentistes qui défendent le concept de taux de succès ne prétendent nullement qu'il est possible d'évaluer celui-ci par une confrontation directe entre le résultat de mesure et la valeur vraie de la grandeur visée – confrontation qui est illusoire. En revanche, ils affirment que l'intérêt de ce critère réside dans le fait que l'on peut contrôler de façon indépendante la fiabilité de la procédure, en la testant en amont sur d'autres cibles que celle qu'elle servira ensuite à mesurer.

Nous pouvons voir dans l'attachement au taux de succès dans la littérature métrologique contemporaine, une reformulation de l'objectif d'exactitude du fréquentisme traditionnel. Certes, l'erreur d'une mesure est inconnue, et avec elle l'exactitude du processus de mesure. Cependant, si l'on sait que les méthodes employées présentent un taux de succès qui a été préalablement attesté et que l'on peut garantir avec une certaine sûreté, il est alors justifié de prétendre que la largeur de l'intervalle probabiliste déterminé par la méthode est une bonne évaluation de l'exactitude de mesure. Le taux de succès d'une procédure est ainsi le biais par lequel on peut garantir la confiance que l'on accorde à un résultat de mesure. La probabilité qui accompagne le résultat de mesure présente un intérêt non pas parce qu'elle rend compte de la croyance de l'expérimentateur mais parce qu'elle témoigne de la performance de la mesure. Selon les fréquentistes qui défendent l'importance du critère du taux de succès, celui-ci est le vecteur par lequel on s'assure de l'objectivité de l'entreprise de mesure.

Dans la description de son attachement au taux de succès, Willink commente la façon dont il perçoit l'interprétation épistémique des probabilités. L'édifice bayésien classique admet une thèse relative au rapport entre croyance et probabilité⁵⁴, qui peut prendre une forme faible et une forme forte. Dans sa forme faible, la thèse affirme que certaines probabilités sont des degrés de croyance. Dans sa forme forte, elle affirme que *toutes* les probabilités sont des degrés de croyance. Willink s'élève contre ces deux thèses en affirmant que « bien qu'une probabilité implique un degré de croyance, un degré de croyance n'implique pas une probabilité »⁵⁵. Willink ne remet pas en cause l'existence d'un lien fort entre probabilité et degré de croyance, mais il estime que ce lien ne prend pas la forme que lui donnent les bayésiens. Ce qu'il réfute, c'est que la probabilité soit *identifiée* à un degré de croyance. Certes, il est possible de chercher à quantifier les degrés de croyance (rationnels) à partir de la propension des agents à s'engager rationnellement dans un pari. Cependant, Willink défend que seule la notion d'espérance statistique permet de comprendre pourquoi cette quantification a un sens : si le degré de croyance de l'agent correspond effectivement à la fréquence relative limite de succès du pari, alors l'espérance de gain de l'agent est positive, ce qui le justifie dans sa croyance. Or, Willink rattache l'espérance statistique à l'interprétation fréquentiste au travers de la loi faible des grands nombres. C'est pourquoi l'interprétation épistémique lui semble en définitive *dépendre* de l'interprétation fréquentiste⁵⁶. Si un agent est prêt à s'engager rationnellement dans un pari, c'est qu'il pense que son espérance de gain est positive. Selon l'interprétation de Willink, le degré de croyance de l'agent *n'est pas* la probabilité (épistémique) attachée au pari, mais est ce que

53. Lira (2008) ; Possolo, Toman et Estler (2009) ; Lira (2009)

54. Drouet (2016)

55. Willink (2013), p.38.

56. Willink (2010c), p.344.

l'agent *croit être* la probabilité (fréquentiste) de l'issue du pari qui lui est favorable. Il y a ainsi une différence notable à entretenir entre une probabilité épistémique qui décrit directement le degré de croyance d'un agent, et le fait de croire qu'une probabilité objective vaut tant ou tant. On distingue ainsi les deux propositions suivantes, qui correspondent à deux interprétations différentes des probabilités⁵⁷ :

- (*épistémique*) « La probabilité p décrit le degré avec lequel je crois que l'événement E va survenir. »
 (*objectif*) « Je crois que la probabilité de l'événement E est p . »

Willink conclut ainsi : « ce n'est pas le concept de croyance qui distingue les statistiques bayésiennes des statistiques fréquentistes. C'est plutôt la nature des objets sur lesquels porte cette croyance »⁵⁸. La conception du fréquentisme que défend Willink prend acte de l'importance de la notion de croyance ; seulement, elle n'admet pas cette dernière pour objet. Chez Willink, la confiance est *justifiée* par des constats probabilistes. Willink distingue ainsi une « assurance pré-mesure » d'une « assurance post-mesure », à partir de l'idée suivante :

La confiance que nous pouvons placer, après la mesure, dans l'idée que nous *avons réussi* est la même que celle que nous avons placée, avant la mesure, dans l'idée que nous *allions réussir*.⁵⁹

Ce faisant, Willink attire notre attention sur une différence de perspective très intéressante entre les deux approches probabilistes, telles qu'elles sont conceptualisées en métrologie. L'insistance de Willink sur la notion de taux de succès, ainsi que sa distinction entre assurance pré-mesure et assurance post-mesure révèle un autre type de préoccupation. Elle montre que l'approche fréquentiste est d'abord tournée vers un succès *futur*. Si rien ne permet de garantir à l'instant présent qu'un résultat est exact, l'approche fréquentiste conserve pour objectif de produire un maximum de résultats les plus exacts possibles. Si les résultats produits se révèlent à terme inexacts, c'est que la méthode employée n'était pas bonne. De fait, les fréquentistes s'orientent en visant une *cible* virtuelle, qui guide leur entreprise scientifique – d'où leur insistance sur la notion de « valeur vraie » d'une grandeur, sur laquelle nous reviendrons dans la prochaine partie. Cette cible constitue un critère *d'action*.

L'approche bayésienne vise quant à elle à faire la synthèse d'un état de connaissance présent à partir des informations (expérimentales et théoriques) disponibles à l'instant où la mesure est effectuée. À chaque nouvelle information, les connaissances seront à leur tour actualisées. L'ensemble constitue une structure qui permet de supporter rationnellement les décisions que devront prendre des agents, à l'instant présent. L'approche bayésienne vise donc la meilleure utilisation possible des informations disponibles *dans le présent*. C'est pour cela, une fois encore, que certains fréquentistes – dont Willink – estiment qu'elle n'est pas falsifiable : si la connaissance vient à changer, cela ne remet pas pour autant en cause les décisions antérieures des agents, qui demeurent justifiées sur la base des informations disponibles au moment où

57. Je remercie Isabelle Drouet d'avoir attiré mon attention sur cette distinction lors de ses cours menés à la Sorbonne à l'hiver 2011 dans le cadre du master LoPhiSC (Université Paris 1 Panthéon-Sorbonne & Université Paris 4 Paris-Sorbonne).

58. Willink (2010c), p.344, note 1.

59. Willink (2013), p.51 (Willink souligne).

elles ont été prises. Dans la conception bayésienne, obtenir des résultats inexacts n'est pas un critère pour critiquer la méthode : bien qu'inexactes, l'important est que ces résultats reflètent effectivement l'état de connaissance à l'instant où ils ont été formulés. De ce fait, rien à l'intérieur du modèle bayésien n'explique pourquoi il demeure nécessaire de continuer l'entreprise scientifique au-delà de ce qu'elle apporte à l'instant présent. Les bayésiens ne disposent pas en métrologie de critère d'action qui provienne de l'intérieur de leur modèle ; il leur faut l'imposer de l'extérieur.

5.5 Conclusion

Notre examen des ramifications conceptuelles et philosophiques du débat qui a lieu en métrologie à propos des méthodes statistiques montre que ce débat ne se limite pas à la seule question de l'interprétation des probabilités. Les interprétations fréquentiste et épistémique des probabilités sont chacune mobilisées dans des conceptions générales de la mesure physique qui engagent des modes de représentation différents et sont guidées par des objectifs distincts. Une même interprétation des probabilités peut d'ailleurs être mise à contribution dans plusieurs conceptions différentes de la mesure. Le spectre de positionnements épistémologiques couverts va de l'empirisme à un réalisme assez marqué mais tout à fait ordinaire chez des scientifiques, et les conceptions se démarquent par le degré de subjectivité qu'elles accordent à l'activité de mesure. Deux approches ressortent plus particulièrement : la première est une approche « traditionnelle » plutôt caractéristique du milieu du XX^e siècle, et la seconde une approche épistémique dont la popularité va grandissante depuis quelques décennies.

Les traits essentiels de l'approche traditionnelle de la mesure, qui fait appel aux probabilités fréquentistes, peuvent être résumés de la façon suivante. Le processus d'évaluation de la fiabilité d'un résultat de mesure aboutit à l'estimation d'une *exactitude* de mesure. Le mécanisme de ce processus gravite autour du concept d'erreur de mesure, et est de ce fait une approche tournée vers la possibilité d'un test *futur* qui révélera si l'intervalle d'exactitude obtenu est correct ou non – c'est-à-dire vrai ou faux. La fréquence avec laquelle les intervalles d'exactitude fournis se révèlent corrects est le taux de succès du processus et mesure la performance de ce dernier.

L'approche épistémique de la mesure, qui mobilise des statistiques bayésiennes, produit quant à elle une évaluation de la fiabilité du résultat de mesure qui est formulée par une *incertitude* de mesure comprise comme un terme épistémique qui prend pour objet un état de connaissance. Ce n'est plus l'erreur de mesure mais la connaissance et la croyance des agents qui sont au centre du modèle. Par suite, l'approche bayésienne est tournée vers l'établissement d'un état *présent* de connaissance fondé sur une recension exhaustive et rationnelle des informations disponibles. La question du test ultérieur du résultat de mesure n'a pas d'importance, car ce qui compte est qu'un acteur prenne à un instant donné des décisions *rationnelles* en regard des informations dont il dispose.

Conclusions de la première partie

Nous avons choisi de nous focaliser sur les modèles fréquentiste et bayésien de la mesure car le premier correspond au mode dominant du courant du XX^e siècle et que le second est promu par un grand nombre de métrologues aujourd'hui. De plus, ils sont remarquables de par leur antagonisme quand à la thématique de l'objectivité et de la subjectivité dans la mesure. Cependant, il nous faut préciser deux choses. D'une part, il n'existe cependant pas *une* méthode fréquentiste unique ni *une* méthode bayésienne unique, mais un spectre d'approches différentes qui croisent entre elles des concepts et des termes provenant des deux modèles généraux que nous avons présentés, de sorte que l'on a en fait « un mélange de concepts et de terminologies d'une approche à l'autre »⁶⁰. Certains statisticiens font d'ailleurs remarquer que l'approche développée dans le supplément 1 du GUM ne peut pas être qualifiée de pleinement bayésienne, et proposent des aménagements visant à aller encore plus loin dans cette direction⁶¹. D'autre part, les approches qui ont été proposées pour répondre aux besoins liés à l'activité d'analyse d'incertitude ne se réduisent pas toutes à des variations autour des thèmes fréquentistes et bayésiens. Certains proposent par exemple d'intégrer les questionnements probabilistes à des raisonnements plus généraux orientés autour de la logique floue⁶². Il est à noter, de plus, que les méthodes de calcul numérique dites de « Monte-Carlo »⁶³ ont pris une ampleur particulière ces dernières décennies, et font d'ailleurs l'objet d'un traitement particulier dans le supplément 1 du GUM. Le développement de ces méthodes, qui présentent des avantages remarquables sur les techniques traditionnelles du GUM depuis que l'accroissement de la puissance de calcul des ordinateurs a permis leur démocratisation, pourrait amener les métrologues à de nouveau redéfinir le cadre d'étude standard de l'analyse d'incertitude.

Notre exposition des modèles statistiques de l'analyse d'incertitude en métrologie illustre comment des débats stimulés par des questionnements en rapport avec la pratique amènent les acteurs à soulever des problématiques conceptuelles qui prennent un tour philosophique. Le fait que les métrologues engagent ces problématiques dans la durée et avec sérieux montrent que celles-ci ne sont pas seulement anecdotiques mais concernent le cœur de leur activité. Tou-

60. Ehrlich, Dybkaer et Wöger (2007), p. 201.

61. Elster (2014)

62. Voir Zadeh (2005) pour une théorie générale, ou encore Urbanski et Wasowski (2003) pour un exemple de mise en application.

63. « Les simulations MC (Monte Carlo) sont des algorithmes implémentés sur ordinateur qui utilisent le hasard pour calculer les propriétés de modèles mathématiques, les modèles étudiés ne présentant pas eux-mêmes de caractère aléatoire », Winsberg (2015), section 2.4.

tefois, il serait inexact de prétendre que le positionnement philosophique adopté par les acteurs conditionne entièrement la résolution de ces débats techniques, tant il apparaît qu'en métrologie comme ailleurs, la préférence accordée à une méthode plutôt qu'une autre est suspendue à des considérations très concrètes liées à la faisabilité et à la performance des méthodes utilisées, comme l'a notamment fait remarquer Humphreys⁶⁴. De fait, les positions philosophiques des acteurs scientifiques sont au moins autant la *conséquence* que la cause de la méthodologie que ces acteurs valorisent, car c'est dans l'attitude réflexive visant à s'interroger sur la portée des méthodes qu'ils utilisent que ces acteurs en viennent à s'interroger sur les fondements épistémologiques qui les sous-tendent. Notre analyse montre que parmi les approches apparemment viables, les avantages et des limites sont *relatifs à des objectifs précis*, et que les désaccords entre métrologues sur l'approche à adopter prennent directement leurs racines dans la diversité de façons dont ils conçoivent les objectifs de la mesure. De ce fait, s'il a certainement un rôle à jouer à l'intérieur même de ces débats, le philosophe ne se situe *en aucun cas* dans une situation *privilegiée* pour faire valoir la légitimité des différentes approches. En effet, sa propension à trancher en faveur de l'une ou l'autre des approches marquerait plus son adhésion à un idéal et un objectif donnés que sa capacité à faire le tri objectif entre les différentes possibilités que les métrologues discutent.

Il faut donc évacuer la tentation qu'il y aurait à chercher à arbitrer entre les différentes approches que proposent les métrologues. Cela revient à renoncer, au moins en partie, à placer la philosophie des sciences dans une démarche normative déjà maintes fois critiquée. Le riche débat qui serait à même de faire éventuellement émerger un vainqueur légitime a déjà lieu chez les métrologues eux-mêmes, et il serait malvenu de prétendre pouvoir se substituer à eux en cherchant à légiférer au moyen d'une démarche philosophique qui se prétendrait en surplomb ; le philosophe ne peut que chercher à enrichir le débat en se plaçant au même niveau que les acteurs du monde scientifique. Il ne s'agit nullement d'adopter une perspective relativiste et de prétendre que toutes les approches se valent automatiquement, mais de rappeler que l'on ne peut pas verrouiller des débats scientifiques sur la base de principes philosophiques *a priori*. Il ne s'agit pas non plus de valoriser un *statu quo*. Les métrologues ont bel et bien besoin de trancher quant à la conceptualisation de la mesure qui leur apparaît comme la meilleure en regard de leurs besoins, dès lors qu'ils considèrent plus que jamais nécessaire de s'accorder sur leurs méthodes. C'est pourquoi un document comme le GUM vient réguler l'analyse d'incertitude, et permet aux métrologues d'étalonner leurs pratiques sur une base commune, sans pour autant que la méthode qui y est préconisée ne puisse être considérée comme universelle.

Les débats présentés ici à propos des méthodes statistiques nous orientent vers deux questionnements essentiels, que nous développerons dans les parties à venir, et qui s'appuient sur un même problème de base : dès lors qu'il est impossible de garantir l'exactitude d'une mesure, que peut-on conclure d'un résultat expérimental ? Un premier questionnement réside dans le fait que si l'on admet le caractère subjectif de l'activité de mesure, doit-on en venir à remettre en question le concept même de « valeur vraie » d'une grandeur sur lequel se fonde l'approche traditionnelle ? Le modèle bayésien semble l'avoir conservé, mais certains métrologues ont émis

64. [Humphreys \(2004\)](#)

l'idée qu'il pourrait être possible de faire l'économie du concept. La partie II est consacrée à l'examen de deux de leurs arguments, l'un lié à l'« inconnaisabilité » de la valeur vraie d'une grandeur, l'autre à l'impossibilité de déterminer une telle valeur de façon unique pour une grandeur donnée. Cela engage dans la continuité un second questionnement quant au rapport entre incertitude et exactitude de mesure. Cela est développé dans la partie III, qui s'articule autour de l'étude des ajustements des constantes de la physique. Ceux-ci mettent en lumière la façon dont les physiciens concilient l'impossibilité de garantir l'exactitude d'une mesure avec le maintien de l'idée d'un progrès expérimental.

Deuxième partie

La « valeur vraie » d'une grandeur

Vue d'ensemble de la partie II

L'examen des modèles statistiques de l'incertitude de mesure révèle que les métrologues adhèrent de plus en plus explicitement à une conception subjectiviste de la mesure, qui évacue l'objectif de représentation d'un monde physique tel qu'il est pour porter son attention sur la description de la connaissance des acteurs. Cette évolution s'observe sur fond de « tournant épistémique » en métrologie. L'un des éléments de discordance des métrologues concerne le statut de l'exactitude de mesure. S'il est acquis pour tous qu'il est impossible de *garantir* l'exactitude d'un résultat de mesure donné à un instant donné, les scientifiques ne s'accordent pas tous sur l'importance qu'il faut accorder à un tel concept. C'est là un point d'opposition majeur entre fréquentistes et bayésiens. À travers ces questionnements, c'est la pérennité de la notion de « valeur vraie » d'une grandeur qui est également en jeu.

Cette partie est consacrée à l'examen du statut de la valeur vraie d'une grandeur physique. La valeur vraie d'une grandeur est ordinairement considérée comme étant la valeur que l'on obtient par une mesure parfaite, c'est-à-dire par une mesure qui n'est affectée d'aucune erreur de mesure. Cependant, cette définition, en un sens, ne dit rien sur le concept puisqu'elle apparaît comme circulaire, l'erreur de mesure étant définie par rapport à la valeur vraie. Le caractère inaccessible et idéal de la valeur vraie a été compris depuis longtemps par les scientifiques, mais ce n'est qu'assez récemment que les textes de métrologie ont amorcé un virage visant à rendre le concept le plus silencieux possible, voire à le supprimer entièrement du formalisme de l'analyse d'incertitude. Les arguments opposés à la valeur vraie dans les textes de métrologie sont parfois un peu déçous. Nous défendons l'idée qu'il est nécessaire de séparer deux arguments que nous jugeons distincts, bien qu'ils soient parfois mélangés dans la littérature spécialisée. Nous étudions alors ces arguments tour à tour, et nous défendons l'idée que ni l'un ni l'autre n'amènent nécessairement à abandonner le concept de valeur vraie d'une grandeur. Nous faisons apparaître l'attachement à ce concept comme un attachement au caractère évaluatif de la mesure, tournée vers une cible.

Le premier argument porte sur l'« inconnaisabilité » de la valeur vraie, et est examiné dans le chapitre 6. La valeur vraie étant ce que l'on cherche à connaître, elle est par nature inconnue ; de plus, puisque tout résultat de mesure est susceptible d'être entaché d'une erreur de mesure, et que l'exactitude d'un résultat est impossible à garantir, alors la valeur vraie d'une grandeur est non seulement inconnue mais aussi *à jamais* inconnue, c'est-à-dire inconnaisable. Partant de cet argument, les métrologues soutiennent qu'il est nécessaire d'adopter une position, qu'ils appellent « opérationnelle », et qui ne fait appel qu'aux concepts qui se rapportent à ce que

l'on peut connaître. Il est alors souhaitable de ne plus faire mention de l'erreur de mesure et de la valeur vraie, pour n'employer que les termes ayant trait au résultat de mesure lui-même : valeur « attribuée », valeurs « raisonnable » et incertitude de mesure. Nous commentons cet argument selon deux axes. Pour commencer, nous soutenons que l'aménagement proposé par les métrologues ne change rien au niveau du formalisme, et qu'un terme théorique similaire au concept traditionnel de valeur vraie continue à être mobilisé dans les calculs d'incertitude. D'autre part, quittant l'aspect technique, nous cherchons à comprendre ce que signifie le terme de « valeur vraie » sur le plan philosophique, en interrogeant en particulier la signification du qualificatif « vrai ». Nous mettons alors en balance une interprétation du terme dans lequel le qualificatif « vrai » n'est qu'une dénomination sans rapport avec la vérité, avec une interprétation littérale du terme. Nous constatons que l'interprétation littérale du terme peut faire appel à deux catégories distinctes de vérité, l'une métaphysique, l'autre épistémique. Nous défendons que le concept de valeur vraie que critiquent les métrologues se rapporte à une interprétation littérale et métaphysique du terme, et nous affirmons que la critique du concept participe d'un mouvement de méfiance envers le contenu métaphysique des concepts employés. Nous prétendons alors que, contrairement à ce que l'on pourrait être amené à soupçonner, la critique de la valeur vraie n'est pas pour autant une critique du réalisme scientifique de la part des métrologues. Nous nuancions pour terminer l'argument d'inconnaissabilité et affirmons qu'à l'intérieur d'une position réaliste, le concept de valeur vraie présente l'intérêt d'induire un processus permanent de correction qui guide le progrès expérimental.

Le second argument, qui fait l'objet du chapitre 7, s'appuie sur le fait que la valeur vraie d'une grandeur physique, si elle existe, ne peut que très rarement être considérée comme unique. En effet, dans de nombreux cas, la définition du mesurande est insuffisamment précise, de telle sorte qu'il est possible de réaliser ce mesurande de différentes façons, en raison de quoi c'est un spectre continu de « valeurs vraies » que l'on peut légitimement associer au mesurande. Pour rendre compte de cet aspect, les métrologues ont introduit un concept supplémentaire, celui d'« incertitude définitionnelle ». Nous discutons de la portée de ce terme en arguant qu'il est effectivement dépendant d'un état de connaissance et peut légitimement être considérée comme une incertitude ; cependant, nous affirmons que ce n'est pas un terme classique d'incertitude, et qu'en particulier, il n'est pas possible de l'intégrer au bilan global d'incertitude comme une composante quantitative ordinaire. Nous cherchons alors à montrer comme le concept de valeur vraie est compatible avec le constat de non-unicité, en développant une approche qui fait appel à une forme de réductionnisme. La multiplicité des valeurs vraies d'une grandeur est alors reliée au caractère phénoménologique, donc non fondamental, de la grandeur considérée – c'est-à-dire que cela traduit la présence d'une sous-structure. Nous montrons alors qu'il est nécessaire de prendre en compte le caractère approché de la connaissance et de relier l'usage des grandeurs physiques aux modèles dans lesquelles celles-ci sont employées, en rapport avec un objectif bien déterminé qui conditionne la précision avec laquelle les mesurandes doivent être définis. Nous concluons que l'incertitude définitionnelle n'est pas tant une composante supplémentaire d'incertitude mais une estimation de la précision attendue de nos modèles qui indique le seuil de détail à partir duquel il devient nécessaire de les affiner.

Chapitre 6

L'argument d'« inconnaisabilité » de la valeur vraie

Notre étude des approches probabilistes de la mesure est venue renforcer un constat déjà pressenti par certains métrologues et philosophes, à savoir que la métrologie a connu une évolution notable lors du XX^e siècle qui se traduit par une transformation des objectifs et des modes de représentation associés à la mesure physique. Nous avons proposé d'appeler « tournant épistémique » ce mouvement général, et nous avons vu que l'inflexion des méthodes statistiques vers une approche bayésienne, fondée sur l'interprétation épistémique des probabilités, vient consolider ce tournant épistémique en l'intégrant au formalisme lui-même. L'évolution des méthodes statistiques employées en métrologie implique d'abord une transformation des modes de représentation, puisque le résultat de mesure ne porte plus sur l'objet physique lui-même mais sur un degré de croyance personnel ou collectif. Mais elle ne se limite pas à ce que représentent les résultats de mesure : elle vient également toucher ce que la mesure elle-même vise à atteindre. Ainsi, l'exigence traditionnelle d'exactitude s'efface derrière une visée descriptive : faire l'état des connaissances à un instant donné. En résumé, l'évolution de la conception de la mesure que font valoir le VIM et le GUM est caractérisée par deux aspects principaux :

- Un changement dans l'objet représenté. Avec le développement des statistiques bayésiennes en métrologie, un résultat de mesure ne représente plus l'état physique de la grandeur mesurée. Désormais, le résultat est la quantification de l'état de connaissance d'un personne ou d'un groupe, au moyen de probabilités épistémiques.
- Un changement de cible. La recherche d'une « exactitude » métaphysique, attachée aux idéaux inaccessibles que sont l'« erreur » et la « valeur vraie », est remplacée par une formulation centrée sur une « incertitude » comprise comme une notion épistémique.

Dans la continuité de ce cheminement conceptuel et méthodologique, et emportés par leur volonté de n'inclure dans les concepts de la mesure que ce qui est effectivement accessible par l'expérience, les métrologues en sont venus à questionner tout particulièrement un concept spécifique de la mesure : celui de « valeur vraie » d'une grandeur physique. Cette critique apparaît parfois en creux, parfois plus explicitement, dans nombre de textes récents de la littérature métrologique. On la retrouve en particulier dans le GUM, puis dans la troisième édition

du VIM. Dans les deux cas, la teneur exacte de la critique nous semble par moments assez confuse. Toutefois, nous pouvons isoler le rôle central d'un argument bien spécifique, l'argument d'« inconnaisabilité », qui prend la forme suivante : la valeur vraie d'une grandeur est un concept problématique car cela renvoie à une valeur à jamais inconnaisable. Dans ce chapitre, nous souhaitons explorer les ressorts de l'argument d'inconnaisabilité de la valeur vraie, afin de comprendre ce que celui-ci engage sur un plan philosophique – est-il par exemple le signe d'un désengagement envers le réalisme scientifique de la part des métrologues ? – et sur un plan conceptuel – signifie-t-il qu'il faille abandonner l'usage du concept ?

Au préalable, nous reviendrons sur la notion générale de grandeur physique, et nous chercherons à identifier ce qu'est la « valeur d'une grandeur », et, par suite, la « valeur vraie » d'une grandeur physique. Ce point est traité à la section 6.1.

Dans un second temps, nous chercherons à mieux définir les contours de la critique de la valeur vraie en métrologie. Ce faisant, notre objectif principal sera de recenser et d'explicitier les différents arguments qui sont mobilisés, dans la littérature métrologique et scientifique, pour remettre en question le rôle de la notion de « valeur vraie » pour caractériser les grandeurs physiques et les résultats de mesure. C'est l'objet de la section 6.2. Nous aboutirons à deux questions principales. Constate-t-on un abandon de l'usage du concept ? Cet abandon est-il nécessaire ? Si tel n'est pas le cas, doit-on alors réformer la signification de la « valeur vraie » d'une grandeur ?

Nous répondrons ensuite brièvement à la première question en montrant que les méthodes d'analyse de données les plus courantes continuent de faire usage d'un terme théorique que nous serions tentés d'appeler « valeur vraie ». C'est l'objet de la section 6.3. Nous concluons sur la nécessité d'interroger la signification et la portée de ce terme théorique pour comprendre l'argument d'inconnaisabilité.

La quatrième section de ce chapitre est consacrée à la question de l'interprétation du concept de « valeur vraie ». Le problème devient alors éminemment philosophique, et à certains égards métaphysiques : si l'on appelle ce terme « valeur vraie », que signifie le qualificatif « vrai » lui-même ? Cela nous incite à revenir sur des questionnements de philosophie générale des sciences, parmi lesquels le dialogue classique entre empirisme et réalisme scientifique, ainsi que des réflexions sur la nature de la vérité. Cet aspect est développé dans la section 6.4. Cela nous portera à conclure que c'est avant tout la teneur métaphysique du concept de valeur vraie qui rebute les métrologues.

La section 6.5 conclut ce chapitre sur une tentative d'évaluation de la portée de l'argument d'inconnaisabilité. Nous affirmerons alors qu'il est possible d'adopter une perspective différente de celle des métrologues pour concevoir la notion de valeur vraie sous un angle plus positif, en insistant sur la façon dont celle-ci guide l'entreprise scientifique en fournissant une cible, laquelle permet d'engager un processus de *correction* des erreurs de mesure qui constitue l'un des moteurs du *progrès* scientifique.

6.1 Grandeur physique et valeur d'une grandeur

Avant d'aborder l'argument d'« inconnaissabilité » de la valeur vraie, nous devons d'abord introduire plus formellement ce qu'est la « valeur vraie » d'une grandeur. Pour cela, il nous faut au préalable expliciter deux notions plus fondamentales encore, celle de « grandeur physique » et de « valeur d'une grandeur ».

En 1887, Hermann von Helmholtz (1821–1894) a proposé dans son ouvrage *Zählen und Messen*¹ un examen des conditions de mathématisation des grandeurs physiques². La démarche de Helmholtz consiste à considérer « mathématiques et physique [...] comme des champs distincts : pour Helmholtz, les équations de la physique mathématique apparaissent comme des *analogies* des relations physiques »³. Les nombres et leur structure mathématique sont donc introduits de façon indépendante de l'expérience sensible, puis des « conditions de mesurabilité » viennent définir la possibilité de transposer au domaine empirique les opérations et les déductions effectuées dans le domaine mathématique. Les conditions de mesurabilité assurent que les grandeurs et les relations physiques partagent des propriétés formelles communes avec les nombres et relations mathématiques, de telle façon que l'on peut mettre en correspondance leurs structures, et que l'on peut alors caractériser une grandeur physique par des nombres, et des opérations physiques entre grandeurs par des opérations mathématiques entre les nombres. Cette position est ordinairement considérée comme le point de départ de l'approche représentationnelle de la mesure⁴, qui voit la mathématisation des grandeurs physiques comme la mise en correspondance d'une structure numérique et d'une structure empirique, en vue d'une *représentation* de la seconde par la première. Selon cette théorie, les grandeurs physiques sont hétérogènes aux nombres; elles ne sont pas quantifiables parce qu'elles sont numériques, mais parce qu'elles « ressemblent »⁵ à des nombres, c'est-à-dire que les relations entre grandeurs physiques présentent une structure telle que les nombres sont adaptés à leur représentation.

Pendant, toute propriété n'est pas nécessairement une grandeur mesurable, et cela doit être vérifié de façon empirique. Une grandeur physique doit ainsi satisfaire un certain nombre de conditions de mesurabilité, attachées à une opération de mesure particulière⁶. Ainsi, la masse est une grandeur mesurable parce qu'elle satisfait à deux conditions. D'une part, on peut définir une opération *empirique* de comparaison des masses qui aboutit soit à une inégalité qui,

1. von Helmholtz (1887). Pour une étude détaillée de la position de Helmholtz, de ses inspirations et de sa réception, voir Darrigol (2003).

2. Nous ignorons ici la différence que l'on peut faire valoir entre « quantité » et « grandeur », qui ne sont pas distingués dans le VIM et dans le GUM, mais qui ne renvoient pas à la même idée. À ce sujet, voir de Courtenay (2015).

3. de Courtenay (2008), p.219 (de Courtenay souligne).

4. Michell a toutefois contesté ce point. Il considère que la conception de Helmholtz ne se distingue pas spécifiquement des conceptions classiques de la mesure issues d'Euclide. Il situe plutôt l'origine de la théorie représentationnelle chez Bertrand Russell : « à ma connaissance, Russell a été le premier à développer une théorie représentationnelle de la mesure », Michell (1993), p.203. L'approche représentationnelle a atteint l'une de ses formes les plus abouties chez Krantz, Luce, Suppes et Tversky, Suppes *et al.* (1971); Suppes *et al.* (1989); Suppes *et al.* (1990).

5. « Les propriétés mesurables d'un objet doivent ressembler d'une façon spécifique à la propriété qu'est le nombre, puisqu'elles peuvent être décomptées représentées par les mêmes symboles; elles doivent avoir une qualité en commun avec les nombres. », Campbell (1921), p.112.

6. Ce n'est pas pour autant un opérationnalisme, car la définition de la grandeur n'est pas attachée à une opération de mesure spécifique. Voir à ce titre Michell (1994), pp.390–391.

à l'instar des relations d'ordre mathématiques, est transitive et antisymétrique, soit à une égalité, transitive et symétrique comme l'est l'égalité algébrique. Par conséquent, la masse présente un ordre. D'autre part, on peut définir une opération *empirique* de sommation des masses qui possède les mêmes propriétés formelles que la somme arithmétique, parmi lesquelles la commutativité et l'associativité. Les propriétés empiriques de ces deux opérations doivent être vérifiées pour tous les objets particuliers ayant une masse. En vertu de cette deuxième opération, la masse est une propriété additive, et elle est donc une grandeur mesurable, que l'on appelle parfois aussi grandeur « fondamentale »⁷ Certaines propriétés, par exemple la température, présentent un ordre mais ne sont pas additives. Dans ce cas, on parle de grandeurs dérivées si l'on peut les définir à partir d'une loi physique impliquant des grandeurs mesurables⁸.

Malgré les nombreuses critiques qu'a reçue l'approche représentationnelle, on retient généralement l'idée que les grandeurs physiques sont une sous-catégorie des propriétés physiques – elles sont l'ensemble des propriétés physiques qui sont quantifiables, c'est-à-dire qui vérifient effectivement les conditions de mesurabilité⁹.

Une fois définie une grandeur physique, la *valeur d'une grandeur* peut à son tour être explicitée. Pour cela, il faut voir au préalable que le terme de « grandeur » recouvre encore deux acceptions différentes. Une grandeur physique est d'abord une catégorie générale qui désigne une propriété quantifiable. Ainsi, la longueur est une grandeur. On parle alors de *type de grandeur*¹⁰. Mais la grandeur est également une caractéristique d'un objet particulier : l'on parlera ainsi de la longueur *d'un* objet spécifique ou d'une situation matérielle spécifique (le diamètre « de la Terre », la distance « Terre-Soleil », etc.). On parle ici de *grandeur particulière*, ou grandeur individuelle. La grandeur particulière est une instance d'une quantité physique en regard d'un cas matériel particulier ; la grandeur est alors *attachée* à ce cas.

C'est lorsqu'on s'intéresse au cas d'une grandeur particulière que l'on peut alors introduire la notion de « valeur d'une grandeur ». Ainsi, la « longueur », comprise comme grandeur générale, n'a pas de « valeur ». Toutefois, c'est bien parce que la longueur est elle-même une grandeur générale qu'il y a un sens à parler de valeur de la longueur d'un objet particulier. En effet, comme l'ont souligné Giordani et Mari¹¹, la catégorie générale de grandeur établit une comparabilité : deux objets peuvent être comparés en vertu la catégorie générale de longueur. On peut donc associer à chaque objet donné sa grandeur particulière¹², et l'opération de mesure permet ensuite de faire correspondre à une grandeur particulière donnée sa valeur, en comparant la grandeur avec une grandeur de même type, prise pour unité. La valeur de la grandeur particulière sera alors le nombre de fois que l'unité est contenue dans la grandeur mesurée. C'est pourquoi, à la longueur *de tel objet* peut être associée la valeur de la longueur

7. C'est le terme qu'utilise Campbell, voir [Campbell \(1920\)](#), p.277.

8. [de Courtenay \(2008\)](#), p.221.

9. Cette appartenance se retrouve dans le VIM, où grandeur est défini comme une « propriété d'un phénomène, d'un corps ou d'une substance, que l'on peut exprimer quantitativement sous forme d'un nombre et d'une référence », [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.2.

10. [Mari \(2009\)](#)

11. [Mari et Giordani \(2012\)](#), p.757.

12. [Mari et Giordani \(2012\)](#), p.757.

de l'objet ¹³. On peut alors écrire, en reprenant la notation instituée par Maxwell :

$$Q = \{Q\}[Q] \quad (6.1)$$

où « Q » est la grandeur particulière mesurée, et « $[Q]$ » est l'unité, c'est-à-dire la grandeur de même nature avec laquelle la grandeur mesurée Q est comparée. Le symbole « $\{Q\}$ » désigne la valeur numérique de Q relativement à l'unité choisie, indiquant le nombre de fois que l'unité $[Q]$ est contenue dans Q . Le terme « $\{Q\}[Q]$ » est donc la valeur ¹⁴ de la grandeur Q . L'interprétation du signe « = » est complexe du fait que, suivant Maxwell lui-même ¹⁵, on ne doit pas le comprendre comme une égalité, mais comme signifiant « Q est exprimé par $\{Q\}[Q]$ ». Une entité concrète, la grandeur Q , est mise en relation avec une entité abstraite, la valeur de la grandeur, en vertu de quoi la grandeur Q , initialement connue par désignation, devient ensuite connue par description ; c'est la fonction épistémologique de la mesure ¹⁶.

Les développements précédents ont successivement examiné comment une propriété concrète du monde physique peut être représentée par des entités mathématique, c'est-à-dire comment cette propriété peut être considérée comme une grandeur physique ; puis ce qu'est l'élément linguistique par lequel cette représentation est amenée, à savoir la valeur de la grandeur. Cependant, cela ne dit rien de la justesse des valeurs attribuées à des grandeurs particulières en vertu d'un mesurage particulier. C'est en s'attelant à cette question précise que l'on en vient à introduire l'idée de valeur « vraie » d'une grandeur physique. La première chose à dire est qu'une valeur vraie est une valeur, c'est-à-dire qu'elle est à situer dans la catégorie de la représentation – elle sert à désigner quelque chose, comme le fait toute valeur d'une grandeur. Si l'on parle de valeur « vraie », cela signifie par ailleurs qu'il faut distinguer les valeurs qui sont justement attribuées à des grandeurs physiques des valeurs qui sont injustement attribuées aux grandeurs. Cela fait intervenir l'idée d'erreur de mesure.

Au début du XX^e siècle, Campbell a cherché à développer un raisonnement qui intègre un questionnement sur la notion d'erreur de mesure. Suivant les traces de Poincaré ou Fechner ¹⁷, Campbell reconnaît que les conditions de mesurabilité ne sont en fait jamais parfaitement vérifiées empiriquement par les relations et opérations physiques entre quantités concrètes. Si l'on compare deux à deux les masses de plusieurs corps A , B et C à l'aide d'une balance à plateaux, et que l'on constate que A et B sont de même masse, puis que B et C sont de même masse, les conditions de mesurabilité stipulent qu'on doit également constater que A et C sont de même masse. Or, fait valoir Campbell, les résultats de l'observation ne s'accordent pas toujours avec cette exigence. En fait, la transitivité de l'égalité nous « apparaît nécessaire uniquement parce que nous sommes familiers avec les balances » ¹⁸. Le désaccord entre les conditions de mesura-

13. Patrick Suppes a décrit la mesure comme une fonction – un morphisme de groupes – qui établit une correspondance entre un modèle de l'observation empirique – la grandeur physique – et un modèle numérique – la valeur de la grandeur ; Suppes (1993).

14. « Valeur numérique » de la grandeur et « valeur » de la grandeur sont deux notions différentes. La valeur numérique est un nombre pur alors que la valeur de la grandeur est dimensionnée, elle inclut l'unité.

15. Giordani et Mari (2011), section 2.2.

16. Giordani et Mari (2011)

17. de Courtenay (2008), p.232.

18. Campbell (1928), p.30.

bilité et les résultats de l'observation peut s'expliquer qualitativement, par exemple en constatant un défaut dans la conception de la balance, dont les bras pourraient ne pas être exactement de la même longueur, ou encore en invoquant des phénomènes de frottement. Cependant, ce désaccord est tout particulièrement problématique si l'on se place dans une démarche de *fondation* de la mesure, où l'on cherche justement à déterminer la mesurabilité de la grandeur « masse » en constatant qu'elle respecte les conditions de mesurabilité. Campbell en déduit que ces violations empiriques des conditions de mesurabilité justifient l'introduction du concept d'erreur de mesure. Il propose de conclure que les conditions de mesurabilité ne sont pas des lois empiriques vérifiées par l'expérience, mais des lois théoriques dont on admet la vérité et qui deviennent alors des *axiomes* de la mesure, l'erreur de mesure venant expliquer pourquoi ces lois ne sont pas vérifiées empiriquement¹⁹. De Courtenay²⁰ a insisté sur le fait que dès lors, l'erreur de mesure est pour Campbell un concept théorique²¹. En accompagnement du concept d'erreur de mesure, Campbell introduit également celui de « grandeur réelle »²², à savoir une grandeur fictive, postulée, qui respecte *par définition* les conditions de mesurabilité que l'on ne parvient pas à vérifier empiriquement. La démarche de Campbell consiste donc à postuler la notion de grandeur réelle, dont les propriétés sont celles qu'auraient les grandeurs physiques si l'on constatait expérimentalement qu'elles vérifiaient les conditions de mesurabilité, c'est-à-dire s'il n'y avait pas d'erreur de mesure. La notion de « grandeur réelle » de Campbell renvoie à ce qui est aujourd'hui appelé « valeur vraie » d'une grandeur²³, communément comprise comme la valeur d'une grandeur particulière que l'on obtiendrait par un mesurage si celui-ci était parfait, et donc, de façon équivalente, en l'absence d'erreur de mesure. Le problème de cette interprétation courante de la valeur vraie est qu'elle apparaît comme une circularité, puisque, comme nous l'avons déjà mentionné à plusieurs reprises, l'erreur de mesure est elle-même définie comme l'écart à la valeur vraie. Définir la valeur vraie comme la valeur obtenue en l'absence d'erreur est elle-même une tautologie ! La démarche de Campbell évite en fait ce cercle vicieux en trouvant une caractérisation empirique de l'erreur de mesure, qui est déduite de la violation *observable* des conditions de mesurabilité. La circularité rend cependant difficile une définition rigoureuse du concept de valeur vraie. Ainsi, le VIM1 mentionne-t-il une « valeur qui caractérise une grandeur parfaitement définie, dans les conditions qui existent lorsque cette grandeur est considérée »²⁴ ; mais que signifie alors « caractériser » ? De même, la définition du

19. « Dans certains cas, lorsque nous trouvons qu'une loi que nous croyions vraie n'est pas strictement une loi, nous devons l'abandonner entièrement et recommencer de zéro [...] [Mais] nos idées sur la mesure du poids et de la plupart des autres grandeurs fondamentales ont été fondées par ceux qui croyaient (inconsciemment, bien sûr) que les lois de l'égalité et de l'addition étaient vraies ; quand il a été découvert qu'elles n'étaient pas strictement vraies, les vieilles idées étaient encore retenues et les nouvelles idées développées sans abandonner totalement les lois à partir desquelles les vieilles idées étaient basées. La conception des "erreurs de mesure" était alors introduite. », Campbell (1920), p.439.

20. de Courtenay (2008), p.232.

21. « L'usage même du terme d'erreur implique une théorie ; il implique qu'il existe un certain système de mesure idéal, à partir duquel est constituée l'erreur. Un tel système idéal n'existe pas réellement. », Campbell (1928), p.137.

22. Campbell (1928), p.138.

23. Plus précisément, la « grandeur réelle » de Campbell est une grandeur, donc un élément du domaine concret. La « valeur vraie » est un élément du domaine symbolique. L'on peut dire que la « valeur vraie » est la valeur de ce que Campbell appelle « grandeur réelle ».

24. Organisation internationale de normalisation (ISO) (1984), p.9.

VIM3 est peut-être moins satisfaisante encore, voyant la valeur vraie comme la « valeur d'une grandeur compatible avec la définition de la grandeur »²⁵; cette fois-ci, c'est le terme « compatible » qui reste très obscur. Cette définition a été perçue par certains métrologues comme une tautologie²⁶. La compréhension de la définition est rendue difficile par l'opacité du terme « compatible ». Il semble que ces définitions tentent précisément d'éviter la tâche, certes ardue, consistant à spécifier ce que signifie le qualificatif « vrai ».

Dans tous les cas, le concept de valeur « vraie » d'une grandeur apparaît dès lors que l'on intègre à la question générale de la mesure la problématique de l'exactitude des mesurages. Dans ce contexte, la « valeur vraie » de la grandeur devient l'objectif idéal de la mesure : le mesurage d'une grandeur particulière aboutit à une indication, laquelle n'est pas exacte, mais que l'on souhaiterait aussi proche que possible de la valeur vraie de la grandeur que l'on aurait obtenue en l'absence d'erreur. L'activité de mesure est alors une activité d'évaluation, l'évaluation de la valeur vraie de la grandeur. La difficulté provient du fait que cette valeur vraie étant postulée, renvoie à une notion idéale, et, de ce fait, inconnue – c'est l'origine de l'argument d'inconnaissabilité que nous développerons dans la prochaine section.

Précisons pour conclure un point qui nous sera essentiel pour la suite. La construction d'une notion de « valeur vraie » d'une grandeur nous pousse à distinguer deux types bien différents de valeur. Si, d'un côté, la valeur vraie de la grandeur constitue la visée de la mesure, que l'on cherche à évaluer, il ne s'agit pas d'une valeur qui est *attribuée* à la grandeur. La valeur attribuée à une grandeur est le résultat effectif d'une mesure, que ce résultat soit directement l'indication d'un instrument de mesure, ou qu'il soit obtenu après un traitement complet incluant une réduction statistique des erreurs aléatoires et une correction des erreurs systématiques. Il nous faut donc distinguer la « valeur vraie » de la grandeur visée, qui constitue la cible inconnue de la mesure, et la « valeur attribuée » à la grandeur, qui correspond à ce que les scientifiques vont utiliser dans leurs modèles, et en particulier dans leurs équations, en partant du principe que cette valeur attribuée est une estimation de la valeur vraie qu'ils aimeraient connaître. Ainsi, lorsque le CODATA propose :

$$\text{Masse de l'électron : } m_e = 9,109\,383\,56(11) \cdot 10^{-31} \text{ kg} \quad (6.2)$$

l'incertitude sur le résultat (représentée par le nombre entre parenthèses, qui porte sur les deux dernières décimales du résultat) est le marqueur d'une différence de nature entre cette valeur, la valeur *attribuée*, et la « vraie » masse, inconnue, de l'électron.

6.2 La critique de la valeur vraie dans le VIM et le GUM : l'« inconnaissabilité »

L'introduction du VIM fait référence au changement conceptuel que nous avons désigné comme un « tournant épistémique » en métrologie. L'angle d'approche est toutefois un peu différent. Laissant de côté les questions liées à l'interprétation des probabilités et aux méthodes

25. [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.20.

26. [Pavese \(2009\)](#), p.1300.

statistiques, elle se focalise sur le statut du concept de valeur vraie dans la mesure, et raisonne en termes d'*objectifs* de la mesure. Ainsi peut-on lire :

Le changement dans le traitement de l'incertitude de mesure, d'une approche « erreur » (quelquefois appelée approche traditionnelle ou approche de la valeur vraie) à une approche « incertitude », a conduit à reconsidérer certains des concepts correspondants qui figuraient dans la deuxième édition du VIM. L'objectif des mesurages dans l'approche « erreur » est de déterminer une estimation de la valeur vraie qui soit aussi proche que possible de cette valeur vraie unique. [...] L'objectif des mesurages dans l'approche « incertitude » n'est pas de déterminer une valeur vraie le mieux possible. On suppose plutôt que l'information obtenue lors d'un mesurage permet seulement d'attribuer au mesurande un intervalle de valeurs raisonnables, en supposant que le mesurage a été effectué correctement.²⁷

On retrouve dans cette remarque la tension, que nous avons soulignée à la section précédente, entre *évaluation* de la valeur vraie (« détermination », dans le vocabulaire du VIM) et *attribution* de la valeur mesurée²⁸. Celle-ci prend la forme d'une opposition entre une démarche normative – une évaluation qui a pour but d'être « aussi proche que possible » d'une cible – et une démarche descriptive – attribuer des valeurs à une grandeur sans chercher à déterminer une valeur vraie.

L'introduction du VIM reste toutefois un peu évasive, et les intentions des auteurs du document semblent apparaître plus clairement à la lumière d'un article écrit conjointement par Ehrlich, Dybkaer, et Wöger, trois membres du JCGM WG2 – le groupe de travail consacré à la maintenance et la révision du VIM. L'article, intitulé “Evolution of philosophy and description of measurement”²⁹, a été publié en 2007 en amont de la troisième édition du VIM (VIM3, 2008), afin d'annoncer les contours des changements prévus par rapport à l'édition précédente (VIM2, 1993). Ehrlich *et al.* identifient deux prémisses de l'approche « erreur » :

- (1) « Pour un mesurande bien spécifié, il existe une valeur unique, appelée valeur vraie, qui est cohérente avec la définition du mesurande »³⁰
- (2) « (I)l est possible de déterminer la valeur vraie d'un mesurande par la mesure, au moins en principe, si une mesure 'parfaite' est effectuée »³¹

Étant données ces prémisses, l'objectif d'une mesure est alors de « déterminer une estimation de la valeur vraie du mesurande, d'“aussi près” que possible »³², formulation reprise dans l'introduction du VIM. Or, comme le veut le constat ordinaire, les résultats de mesure sont toujours affectés d'erreurs de mesure. Pour obtenir un résultat « parfait », il faut donc corriger ces erreurs, obtenir une erreur finale aussi faible que possible, et se rapprocher ainsi de la valeur vraie

27. Joint Committee for Guides in Metrology (JCGM) (2012), p.x. Cette remarque du VIM fait écho à une situation déjà évoquée en 1973 par les métrologues du *National Physical Laboratory* (NPL, Londres) qui mentionnent l'existence de « deux écoles de pensée », l'une défendant le concept de valeur vraie, l'autre défendant une position qu'ils désignent comme « les observables uniquement », *Campion, Burns et Williams* (1973), p.26.

28. Mari a étudié cette tension, voir *Mari* (1997).

29. *Ehrlich, Dybkaer et Wöger* (2007)

30. *Ehrlich, Dybkaer et Wöger* (2007), p.202.

31. *Ehrlich, Dybkaer et Wöger* (2007), p.202.

32. *Ehrlich, Dybkaer et Wöger* (2007), p.203.

-
- (i) L'objectif est d'estimer la valeur vraie de la grandeur visée.
 - (ii) Pour cela, on procède à une mesure de la grandeur.
 - (iii) Cependant, le résultat est affecté d'une erreur de mesure, qui est la différence entre le résultat de mesure et la valeur vraie de la grandeur.
 - (iv) On aimerait donc corriger l'erreur pour remonter à la valeur vraie elle-même.
 - (v) Mais on ne connaît pas l'erreur commise, puisqu'il faudrait pour cela connaître la valeur vraie de la grandeur – qui est précisément ce que l'on cherche à évaluer depuis le début. On est alors renvoyé à (i).
-

TABLEAU 6.1 – Circularité soulignée par Ehrlich *et al.*, caractéristique de l'argument d'« inconnaisabilité ».

de la grandeur. Le problème réside alors dans le constat selon lequel il n'est jamais possible de savoir si toutes les erreurs ont été corrigées ou éliminées, et donc de savoir si le résultat final est proche ou non de la valeur vraie de la grandeur. Comme l'expliquent Ehrlich *et al.* :

Jusqu'à présent, aucune méthode satisfaisante n'a été trouvée pour identifier, et encore moins pour corriger, toutes les erreurs de mesure. [...] Puisqu'il est virtuellement impossible de savoir avec certitude s'il y a un autre élément [d'erreur systématique] [...], l'erreur systématique "totale" est inconnue [...] [et] la valeur vraie ne peut pas être connue.³³

L'argument des auteurs revient à souligner la circularité suivante : pour connaître la valeur vraie, il faut connaître l'erreur ; et pour connaître l'erreur, il faudrait connaître préalablement la valeur vraie (tableau 6.1). Cette limite est celle de l'« inconnaisabilité » de la valeur vraie. Les auteurs font également valoir, dans un autre registre, les limites liées au traitement statistique traditionnel des données, que nous avons traitées en détail dans la partie précédente. L'argument d'« inconnaisabilité » s'en distingue cependant, en ce qu'il n'est pas un argument d'abord méthodologique ou technique, mais immédiatement épistémologique.

L'argument d'inconnaisabilité peut également être interprété comme une variation autour de la question de l'exactitude de mesure. Quel que soit le niveau de finesse d'une mesure expérimentale et de son traitement théorique, il peut toujours subsister, en bout de chaîne, une erreur systématique non identifiée – et donc inconnue – qui rend le résultat erroné. Cela est reporté dans le GUM :

Le résultat d'un mesurage (après correction) peut, sans qu'on le sache, être très proche de la valeur du mesurande (et, en conséquence, avoir une erreur négligeable) même s'il possède une incertitude élevée. C'est pourquoi l'incertitude du résultat d'un mesurage ne doit pas être confondue avec l'erreur résiduelle inconnue. [...] Naturellement, un effet systématique non mis en évidence ne peut pas être pris en compte dans l'évaluation de l'incertitude du résultat d'un mesurage mais il contribue à son erreur.³⁴

33. Ehrlich, Dybkaer et Wöger (2007), pp.205–206.

34. Joint Committee for Guides in Metrology (JCGM) (2008d), p.6.

Par conséquent, rien ne peut jamais être affirmé quant à la valeur vraie de la grandeur étudiée sans subordonner cette affirmation à une *hypothèse*, celle selon laquelle aucune erreur n'a subsisté après correction.

Les réflexions proposées dans le VIM se font l'écho de développements déjà engagés dans le GUM, où l'argument d'inconnaisabilité apparaît également, comme en témoigne l'extrait suivant :

La définition de l'incertitude de mesure [...] est une définition opérationnelle qui se focalise sur le résultat de mesure et son incertitude évaluée. Elle n'est cependant pas incompatible avec d'autres concepts d'incertitude de mesure tels que :

- mesure de l'erreur possible sur la valeur estimée du mesurande telle que fournie par le résultat d'un mesurage.
- estimation caractérisant l'étendue des valeurs dans laquelle se situe la valeur vraie d'une grandeur mesurée.

Bien que ces deux concepts traditionnels soient valables en tant qu'idéaux, ils se focalisent sur des grandeurs inconnues : respectivement l'« erreur » du résultat d'un mesurage et la « valeur vraie » du mesurande (par opposition avec sa valeur estimée).³⁵

Le constat d'inconnaisabilité est par ailleurs répété à plusieurs autres reprises dans le corps du document³⁶. L'une des conséquences notables du constat du caractère idéal et inconnaisable du concept de valeur vraie réside dans le fait que les auteurs du GUM entreprennent d'éviter l'usage du terme lui-même :

Le terme valeur vraie [...] a été traditionnellement utilisé dans les publications sur l'incertitude, mais non dans ce *Guide*³⁷

La démarche du GUM peut être résumée ainsi. (i) Il n'est pas possible de connaître la valeur vraie d'un mesurande (ou l'erreur finale, ce qui est équivalent) (ii) Par conséquent, il serait illusoire de prétendre qu'un résultat de mesure a pour objet la valeur vraie du mesurande. (iii) Il faut au contraire parler de ce que l'on peut connaître, c'est-à-dire les valeurs que l'on *attribue* au mesurande : (iv) C'est pourquoi l'incertitude de mesure ne caractérise plus « l'étendue des

35. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.3. L'« inconnaisabilité » apparaît de façon plus explicite dans la version originale du document, en anglais. Celle-ci emploie, en lieu et place de « grandeurs inconnues », l'expression “*unknowable quantities*”, [Joint Committee for Guides in Metrology \(JCGM\) \(2008a\)](#), p.3. Les deux expressions peuvent difficilement être considérées comme synonymes et le choix de traduction (de l'anglais vers le français) est plutôt surprenant. Nous considérerons par la suite que le terme « inconnu » est employé ici pour signifier « inconnaisable », ce qui est en bien meilleure adéquation avec les remarques formulées dans le reste du document.

36. « (L)e concept d'erreur est idéal et les erreurs ne peuvent pas être connues exactement », [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.5; « ni la valeur de la grandeur réalisée ni celle du mesurande ne peuvent jamais être connues exactement; tout ce qu'on peut connaître est leurs valeurs estimées. », [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.52; « (a)lors que les valeurs exactes des contributions à l'erreur d'un résultat de mesurage ne sont pas connues et ne peuvent pas l'être, les incertitudes associées aux effets aléatoires et systématiques responsables de l'erreur peuvent être évaluées. », [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.53.

37. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.51.

valeurs dans laquelle se situe la valeur vraie d'une grandeur mesurée »³⁸ mais « la dispersion des valeurs attribuées à un mesurande »³⁹.

Le point focal de ce *Guide* concerne le résultat de mesure et son incertitude évaluée, plus que les grandeurs inconnues valeur « vraie » et erreur. En adoptant le point de vue opérationnel que le résultat d'un mesurage est simplement la valeur à attribuer au mesurande et que l'incertitude de ce résultat est une mesure de la dispersion des valeurs qui pourraient être raisonnablement attribuées au mesurande, ce *Guide* rompt en fait la liaison souvent déroutante entre l'incertitude et les grandeurs inconnues valeur « vraie » et erreur.⁴⁰

(v) Dès lors, tout appel à la notion d'erreur de mesure n'a plus lieu d'être :

À vrai dire, l'approche opérationnelle du présent Guide, où l'accent est mis sur la valeur observée (ou estimée) d'une grandeur et sur la variabilité observée (ou estimée) de cette valeur rend entièrement inutile tout recours au concept d'erreur.⁴¹

Et la terminologie « valeur vraie » est vidée de son sens :

Dans ce *Guide*, on évite l'emploi des termes « valeur vraie d'un mesurande », ou « valeur vraie d'une grandeur » (souvent abrégés en « valeur vraie »), parce que le mot « vrai » est considéré comme redondant.⁴²

Les remarques du GUM ne vont pas sans poser un certain nombre de problèmes significatifs. Si l'on rend inutile tout recours au concept d'erreur, comment justifie-t-on alors de faire des corrections ? Si l'expression « valeur vraie » est tenue pour redondante, cela signifie-t-il qu'il n'y a pas de différence entre la valeur attribuée à une grandeur et la valeur vraie de la grandeur⁴³ ? Dire que « le résultat d'un mesurage est simplement la valeur à attribuer au mesurande » n'est-il pas une tautologie ? De plus, cette formulation n'occulte-t-elle pas un élément essentiel, à savoir le critère sur lequel on attribue une valeur au mesurande ? Comment définit-on alors ce que l'on cherche ? Comme nous le défendrons par la suite, la plupart des revendications que fait valoir le GUM apparaissent injustifiées, ou bancales. Cependant, l'argument d'inconnaisabilité lui-même reste ouvert.

Nous voyons ici apparaître deux questionnements assez distincts, selon la façon dont on interprète les intentions déployées dans le GUM, et que le VIM prête à l'approche dite « incertitude » de la mesure. La première question est une question technique, qui porte sur le contenu des méthodes d'analyse de données : *est-il nécessaire, pour caractériser et exploiter les résultats de mesure, de faire appel à un paramètre mathématique tel que la « valeur vraie » d'une grandeur ?* En d'autres mots, est-ce que les scientifiques utilisent effectivement dans leurs calculs la notion

38. Organisation internationale de normalisation (ISO) (1984), p.16.

39. Joint Committee for Guides in Metrology (JCGM) (2012), p.25.

40. Joint Committee for Guides in Metrology (JCGM) (2008d), p.62.

41. Joint Committee for Guides in Metrology (JCGM) (2008d), p.62.

42. Joint Committee for Guides in Metrology (JCGM) (2008d), p.52, annexe D.3.5.

43. Comme le fait valoir Pavese, « le GUM, tout en abandonnant le terme "vraie", n'ajoute pas à "valeur" la spécification élémentaire "mesurée" », Pavese (2009), p.1298. De fait, en raison de l'amendement effectué, la terminologie du GUM devient remarquablement vague.

de valeur vraie ? Cette question présente un intérêt tout particulier par la façon dont elle conditionne les discussions plus philosophiques que nous voudrions ensuite aborder. Si l'approche « incertitude » fait l'économie de la notion de valeur vraie, cela engage une transformation substantielle avec un fort impact sur la pratique des scientifiques. Mais si l'approche « incertitude » maintient l'usage de la notion de valeur vraie, c'est que l'amendement qui est apporté au concept porte sur son interprétation et sur le sens que l'on donne à l'activité de mesure, sans pour autant que cela n'ait d'influence sur la pratique des scientifiques. La distinction que nous proposons ici entre questions « techniques » et questions « philosophiques » repose donc en particulier sur le statut du *formalisme* de l'analyse d'incertitude : si les conclusions relatives à une question donnée – par exemple la légitimité du concept de valeur vraie d'une grandeur – n'ont aucune conséquence sur le formalisme lui-même, c'est qu'elles sont seulement d'ordre interprétatives ; d'où le fait de parler, un peu abusivement toutefois, de questions « philosophiques ». Dans la section suivante, nous défendons que les approches les plus courantes de la mesure continuent de faire usage d'une notion de valeur vraie.

Cela nous amène par conséquent à un second questionnement : *quelle est la signification du terme « valeur vraie », et en particulier du qualificatif « vrai » ?* Cette question renvoie à un problème éminemment philosophique, car il revient d'une part à questionner le statut des termes théoriques dans les théories et les modèles scientifiques, et d'autre part à s'interroger sur la nature de la « vérité » que l'on peut associer à ces termes théoriques. Parler de « valeur vraie » d'une grandeur peut donner l'impression que l'on prête à la mesure des vertus métaphysiques fortes, à savoir la possibilité d'accéder à une réalité indépendante de nous. De ce fait, des questionnements internes à la métrologie, d'abord introduits par les métrologues eux-mêmes, nous renvoient à des problématiques générales de la philosophie des sciences. Nous explorerons cette question dans un second temps, après avoir répondu au questionnement plus technique de l'usage de la valeur vraie dans les modèles métrologiques.

6.3 L'usage de la valeur vraie dans les méthodes statistiques appliquées à la mesure

6.3.1 Les principales conceptions de la mesure en métrologie ne font pas l'économie de la valeur vraie

Nous avons développé à la partie I les ressorts techniques de l'analyse statistique des données expérimentales selon deux modèles essentiels de la pratique métrologique, l'approche fréquentiste traditionnelle et une approche bayésienne en plein développement.

Rappelons que l'approche fréquentiste vise à partir d'un échantillon de données pour remonter à la valeur vraie de la grandeur en construisant un *estimateur* de cette valeur, typiquement la moyenne arithmétique de l'échantillon, estimateur dont on peut chercher à évaluer l'exactitude en s'appuyant sur la variabilité des données expérimentales, traitée statistique, et sur la possible amplitude des erreurs systématiques dont il faut rendre compte par ailleurs. Dans l'approche fréquentiste, l'introduction de la valeur vraie est opérée au niveau de ce que nous avons désigné par l'hypothèse (H2) (chapitre 3.2.2), qui établit que le résultat moyen obtenu en l'absence d'erreur systématique est la valeur vraie de la grandeur.

L'approche bayésienne fonctionne par révision successive de connaissances : les données expérimentales viennent enrichir une croyance *a priori*, ce qui est implémenté par le théorème de Bayes, lequel permet de calculer une probabilité *a posteriori* décrivant les connaissances de l'expérimentateur à propos de la grandeur au vu des informations dont il dispose. Le rôle de la valeur vraie n'y est pas particulièrement différent de celui qu'elle joue dans l'approche fréquentiste. La valeur vraie apparaît en particulier dans l'expression de la vraisemblance. Rappelons que le théorème de Bayes, appliqué aux données répétées de mesure (chapitre 4.5.1), indique que :

$$p(\chi | \{x_i\}) \propto L(\chi | \{x_i\}) \times p(\chi) \quad (6.3)$$

La fonction $L(\chi | \{x_i\})$ est la fonction de vraisemblance. On peut l'exprimer par exemple selon une hypothèse gaussienne (voir l'annexe A.3) :

$$p(x_i = t | \mu, V) \propto \frac{1}{\sqrt{V}} \exp \left[-\frac{1}{2V} (t - \mu)^2 \right] \quad (6.4)$$

où l'on fait apparaître l'espérance d'une distribution de probabilité. À l'instar du cas fréquentiste, cette espérance s'identifie à la valeur vraie à condition qu'il n'y ait pas d'erreur systématique. L'hypothèse (H2) du modèle fréquentiste s'exprime ici aussi, mais de façon sous-jacente. Au-delà des apparences, la conception du paramètre « valeur vraie » est très similaire dans les deux approches.

Les deux approches font donc intervenir un critère d'attribution d'une valeur à la grandeur qui fait intervenir un paramètre que l'on appelle traditionnellement « valeur vraie ». De façon remarquable, la valeur vraie joue un rôle *normatif* dans les deux approches. En principe, il est tout à fait possible d'appliquer l'une ou l'autre des méthodes sans se poser aucune question sur l'existence d'une telle valeur « vraie » – il n'y a qu'à appliquer la méthode mathématique fournie. Cependant, rien ne vient *justifier* l'application de ces méthodes s'il n'y a pas dans le modèle statistique un postulat, celui selon lequel la grandeur possède une valeur vraie que l'on cherche à évaluer. Ainsi, la valeur vraie donne une assise au caractère *inférentiel* des statistiques employées, qui ne se bornent pas à un rôle descriptif. L'objectif, dans l'approche fréquentiste, n'est pas de décrire les propriétés statistiques d'un échantillon de données (la moyenne, l'écart-type des données expérimentales), mais d'employer ces propriétés *pour inférer* quelque chose à propos de la valeur vraie de la grandeur visée. L'approche bayésienne fait valoir le rôle normatif de la valeur vraie de façon dérobée : en axant la représentation sur les états de croyance, elle peut donner l'impression que la valeur vraie n'intervient plus. Cependant, celle-ci réapparaît lorsqu'on s'intéresse par exemple à l'expression de la fonction de vraisemblance, comme nous l'avons défendu plus tôt.

6.3.2 Peut-on vraiment parler d'une approche « incertitude » ?

À la lumière des objections formulées dans cette section, nous voyons que la description donnée dans le GUM, puis environ quinze ans plus tard dans le VIM et dans certains articles de la littérature métrologique comme celui d'Ehrlich, Dybkaer et Wöger, ne semble pas vraiment

correspondre aux pratiques effectives. Du moins, il y a un décalage apparent entre l'objectif affichés du GUM, tenant à faire disparaître la valeur vraie dans l'*expression* d'un résultat de mesure, sur la base d'un constat d'inconnaisabilité (disparition que l'on retrouve dans la définition d'« incertitude de mesure » dans le VIM), et la présence de la valeur vraie comme paramètre dans les équations générales employées. Cela rejoint le constat émis par Walter Bich, le président du groupe de travail JCGM WG1, mandaté par le BIPM pour maintenir et réviser le GUM :

La discussion qui suit est basée sur l'expérience personnelle de l'auteur. Au moment où j'ai appris les concepts [...] concernant l'approche bayésienne des probabilités appliquée aux mesures, la valeur vraie était soumise à de sévères critiques, particulièrement en Italie. L'approche basée sur la valeur vraie et l'erreur était mise en doute parce qu'elle était fondée sur des quantités inconnaisables, c'est-à-dire des concepts idéalisés. Les termes mêmes étaient presque bannis de la littérature, et quiconque osait les utiliser était considéré avec suspicion comme un partisan de vieilles idées. Par la suite, les échos de cette critique ont également résonné dans le GUM [...] Cependant, lors du passage de la philosophie aux calculs, des erreurs et des valeurs (vraies) de grandeurs pouvaient toujours être facilement identifiées dans les formules, malgré le nouveau cadre destiné à éviter leur usage.⁴⁴

Bich situe la seule transformation véritablement significative de la métrologie dans l'espace des méthodes statistiques employées, et non dans des remarques conceptuelles sur le statut de la « valeur vraie » d'une grandeur.

En recensant les arguments que le GUM apporte à l'encontre de l'usage du terme « valeur vraie », nous avons vu qu'il conclut en particulier au fait que ce terme, « valeur vraie » est redondant. Comment peut-on comprendre cette remarque ? Le nom « valeur » peut être associé à deux qualificatifs différents, « vraie » et « attribuée » (ou valeur « mesurée » dans le VIM⁴⁵), représentant ainsi deux concepts différents. Le mouvement opéré dans le GUM vise à placer la valeur mesurée au centre de l'attention. Plus encore, il consiste à affirmer que le concept de valeur « mesurée » est le seul valide. Dès lors, vu que toute référence à la valeur « vraie » est supprimée, la différenciation entre « valeur vraie » et « valeur attribuée » n'est plus nécessaire, et le choix est pris de ne plus utiliser que le simple terme « valeur », sans qualificatif. Cependant, cela mène à de grandes mésententes, puisque, comme nous le prétendons – en accord avec Bich – le concept de valeur vraie n'est *pas* éliminé dans le GUM. Le choix effectué dans le GUM est à la source d'un important manque de clarté, et mène à la confusion, précisément, entre les notions de valeur « vraie » (l'idéal-cible de la mesure) et de valeur « attribuée » (le résultat effectif, reporté dans les tables de référence et employé dans les calculs numériques). Ce sont finalement Ehrlich, Dybkaer et Wöger eux-mêmes qui révèlent le mieux l'ambiguïté de la position du GUM : « le GUM décourage l'utilisation du terme (mais pas du concept) "valeur vraie" »⁴⁶. La critique de la notion de valeur vraie semble alors se réduire à une question d'interprétation et non d'usage, et, contrairement à ce qui est annoncé, l'approche GUM ne remet

44. Bich (2012a), p.2156.

45. Joint Committee for Guides in Metrology (JCGM) (2012), p.19.

46. Ehrlich, Dybkaer et Wöger (2007), p.210.

pas en question la notion de « valeur vraie » :

Il est important de reconnaître que [l'approche du GUM] ne signifie pas que le concept de valeur vraie est découragé ou ignoré dans le GUM. [...] (D)ans l'approche du GUM, le concept de valeur vraie est nécessaire pour décrire l'objectif de la mesure. Le concept de valeur vraie est aussi nécessaire pour formuler un modèle de mesure.⁴⁷

Bien que le VIM adopte dans l'ensemble les principes du GUM, le choix a été pris de maintenir les catégories traditionnelles de « valeur vraie »⁴⁸, d'une part, et de « valeur mesurée »⁴⁹, d'autre part. Ce choix semble, en regard de notre analyse, une meilleure décision. Celle-ci n'était pourtant pas immédiatement acquise, d'après les témoignage de Marc Priel :

Les premiers projets de la version 3 [du VIM], en 2004, conduisaient à la suppression du concept d'erreur et auraient à terme généré l'obligation de modifier un très grand nombre de normes. Plusieurs instituts de métrologie et différentes organisations ont réagi en mettant en évidence les risques d'une évolution aussi radicale.⁵⁰

Priel explique également que :

Pendant [des] années de réflexion, les experts se sont lancés dans ce grand débat sur l'existence ou non de la valeur vraie. En 2004, ils envisageaient même dans un premier projet, de ne présenter que la seule méthode [incertitude], relayant en annexe ce concept de valeur vraie. Plusieurs pays, dont la France, ont émis des doutes et des inquiétudes. Finalement, ils sont revenus à une position intermédiaire, en présentant ces approches en parallèle. Mais il est certain qu'à l'avenir nous parlerons de moins en moins de valeur vraie.⁵¹

À l'avenir, parlerons-nous de moins en moins de la valeur vraie, comme le pressent Priel ? L'une des réponses réside dans la façon dont le prochain GUM (le « GUM2 », en préparation) traitera du sujet. Sans pouvoir présager du contenu du futur document, les nuances que nous avons apportées jusqu'ici nous rendent enclins à tempérer cette affirmation.

L'analyse que nous proposons ici nous amène à restreindre la portée de l'évolution dont témoignent en particulier le GUM et l'introduction du VIM. Certes, l'idée que la métrologie a connu un tournant épistémique demeure tout à fait valable ; cependant, les conséquences exactes de ce tournant sur le concept de valeur vraie lui-même sont moins évidentes. En particulier, l'approche contemporaine n'induit aucune transformation notable dans le formalisme vis-à-vis de la valeur vraie d'une grandeur – tout au plus suggère-t-elle une révision de la terminologie et de l'interprétation du concept. Cela nous amène à discuter du questionnement philosophique qui en découle : comment peut-on comprendre ce concept de « valeur vraie » d'une grandeur ?

47. Ehrlich, Dybkaer et Wöger (2007), p.211.

48. Joint Committee for Guides in Metrology (JCGM) (2012), p.20.

49. Joint Committee for Guides in Metrology (JCGM) (2012), p.19.

50. Priel (2010), p.4.

51. Priel (2008), p.21.

6.4 Questionnement philosophique : valeur « vraie » et réalisme scientifique

Les métrologues s'attaquent à la valeur vraie dans une discussion qui, bien qu'elle ne soit pas explicitement formulée en ces termes, mais qui nous semble clairement engager des problématiques philosophiques. Le questionnement principale que nous souhaitons mener ici vise à expliciter la façon dont on comprend le rapport entretenu entre « valeur vraie » et vérité. En particulier, l'emploi d'un concept de « valeur vraie » nécessite-t-il de s'engager dans une forme de réalisme scientifique ? Plus encore, l'analyse de ce concept nous incite-t-elle à plaider en faveur de l'adhésion au réalisme scientifique ? L'argument d'inconnaisabilité semble effectivement marquer une distanciation avec une conception réaliste ; du moins, si elle ne l'est pas explicitement, nous défendrons que c'est ce qu'elle implique de façon logique. Nous affirmerons alors qu'il nous semble pouvoir interpréter cette prise de distance comme le reflet de la méfiance qu'entretiennent les scientifiques envers les types de questionnement qui semblent dépasser le cadre de l'activité scientifique proprement dite – en particulier, des questionnements d'ordre métaphysique – quand bien même les scientifiques auraient tendance à conserver naturellement un attachement à un réalisme scientifique modéré.

6.4.1 Valeur vraie et vérité

Dans la section précédente, nous avons défendu que les approches les plus classiques de la métrologie (l'approche fréquentiste traditionnelle, l'approche du GUM ou encore celle de son supplément 1) continuent d'utiliser dans leur machinerie technique un paramètre inconnu et idéal que l'on peut ou non être tenté d'appeler « valeur vraie ». Celui-ci joue un rôle normatif, en présidant la façon dont une valeur est attribuée à une grandeur à partir des données expérimentales et théoriques à disposition. La présence effective de ce terme théorique dans les approches métrologiques traditionnelles et contemporaines nous incite à conclure que s'il peut y avoir une critique du concept de valeur vraie en métrologie, celle-ci doit porter sur la signification du concept – et en particulier sur la légitimité de le dénommer ou non « valeur vraie ». C'est là une question pleinement philosophique qui nous invite à étudier de plus près le lien entre « valeur vraie » et « vérité ». Ne pourrait-on pas simplement voir dans le terme « valeur vraie » un abus de langage ? Pourquoi en effet parler de valeur « vraie » à propos de ce qui n'est au premier abord qu'un paramètre mathématique dans un modèle ? Le concept n'est-il pas seulement un outil mathématique fonctionnel, sans aucune relation effective avec une quelconque « vérité » ? Il s'agirait là d'une interprétation instrumentale qui semble, du moins au premier abord, tout à fait tenable et en accord avec des thèses philosophiques très classiques. L'attitude inverse consiste à interpréter le terme de « valeur vraie » de façon littérale et de considérer ainsi qu'il s'agit d'une valeur qui présente la propriété d'être vraie. Dans ce cas, le qualificatif « vraie » est particulièrement problématique par la portée philosophique, et en particulier métaphysique, qu'il engage : dans quelle mesure le terme de « valeur vraie » peut-il légitimement renvoyer à une idée de « vérité » au sens propre ?

De ce fait, nous discernons ici deux façons tout à fait distinctes de comprendre le concept, l'une selon une interprétation instrumentale, l'autre selon une interprétation littérale. Cela

revient à découper de deux façons différentes l'expression globale commune « vraie valeur d'une grandeur » :

- (1) {vraie} {valeur d'une grandeur}
- (2) {vraie valeur} {d'une grandeur}

Le premier découpage proposé, « {vraie} {valeur d'une grandeur} », suggère une interprétation littérale. Dans ce cas, la valeur vraie d'une grandeur est une valeur particulière, qui porte une valeur de vérité que n'ont pas, à l'opposé, les autres valeurs possibles de la grandeur. Ici, « vraie » est alors effectivement un qualificatif de « valeur », et « valeur vraie » tire sa signification de l'association de deux idées, celle de valeur d'une grandeur et celle de vérité. Cela revient à dire que si l'on affirme :

(P) « La masse de cet objet est 5 kg »

cette proposition est susceptible d'être vraie ou fausse, et que la « vraie masse » de l'objet est la seule valeur qui rend vraie la proposition (P) précédente⁵². Dans ce cas, il faut alors s'interroger sur ce qui est entendu, de façon plus générale, par le qualificatif « vraie » – autrement dit, cela revient à s'interroger sur la nature de la vérité, ce qui est une question que l'on retrouve à la racine des principales théories de la vérité dans la philosophie contemporaine⁵³. Ce faisant, nous en venons à spécifier deux aspects en particulier. D'une part, quel élément précis vient caractériser la valeur de vérité ? D'autre part, en vertu de quoi une valeur vraie est-elle considérée comme vraie⁵⁴ ? Nous explorerons cette question à la sous-section 6.4.3.

La seconde acception de « valeur vraie » est donnée par « {valeur vraie} {d'une grandeur} ». Dans ce second cas, « valeur vraie » n'est pas à comprendre comme l'association de deux idées mais comme un tout, c'est-à-dire comme une simple dénomination qui n'engage à rien quant à un quelconque statut de vérité – et, par extension, n'engage à rien sur le plan métaphysique. Le terme « valeur vraie » devrait alors être compris comme *un* terme singulier exprimant *une* notion. Sous cet angle, le choix du terme « vraie » lui-même devient arbitraire ; parler de « valeur vraie » n'a plus pour rôle que d'introduire une distinction syntaxique avec la valeur *attribuée* à la grandeur. Il en découle, d'un point de logique tout du moins, que n'importe quelle terminologie aurait pu satisfaire dans cette situation précise ; c'est pourquoi certains ont proposé l'usage de « valeur théorique »⁵⁵ ou encore « valeur cible »⁵⁶. Pour distinguer cette acception de la précédente, nous proposons d'amender la terminologie correspondant à cette interprétation précise, et de parler à son sujet de « valeur-vraie », le trait d'union indiquant l'absence de

52. Comme l'exprime Mari, selon cette conception, « "la valeur vraie de la quantité X est Y " devrait en fait être interprété comme "la proposition 'la valeur de la quantité X est Y ' est vraie" », Mari (2005), p.262.

53. Glanzberg (2013) : « [Les théories de la vérité] tentent toutes de répondre directement à la question de la nature : quelle est la nature de la vérité ? Elles prennent cette question selon sa valeur nominale : il y a des vérités, et la question qui se pose concerne leur nature. En répondant à cette question, chaque théorie fait de la notion de vérité une partie d'une métaphysique ou d'une épistémologie plus profonde ».

54. Dans l'épistémologie anglophone contemporaine, ces deux thématiques sont désignées comme la recherche de l'identification des "truthbearers", qui portent la valeur de vérité, et des "truthmakers", qui rendent les "truthbearers" vrais ou faux (Glanzberg, 2013).

55. « La notion de « valeur vraie » d'une grandeur n'a en général pas de sens et il est préférable de parler de valeur théorique, de référence ou admise », Robert-Schwartz et Treiner (2003), p.3.

56. Eisenhart (1963), p.171. Willink (2013), p.6.

relation qualificative entre les deux mots employés. L'interprétation en termes de valeur-vraie permet de faire l'économie d'une réflexion philosophique sur la vérité, mais ne rend pas moins nécessaire de comprendre ce que désigne l'objet « valeur-vraie ». Puisque l'on se place alors dans un cadre où la vérité des énoncés n'est pas un élément essentiel, cela signifie que l'interprétation amorcée trouve un cadre naturel d'exposition dans une philosophie anti-réaliste. Nous chercherons à voir dans quelle mesure cette posture peut être développée, ainsi que ce qu'elle engage, à la sous-section 6.4.2.

Nous développons notre argumentation en deux temps. Nous explorons d'abord ce que pourrait être une conception de « valeur-vraie », comprise dans le sens (2) développé précédemment, c'est-à-dire qui n'est pas attachée à une notion de vérité. Puis, à l'inverse, nous développons ensuite ce que peut être une « valeur vraie » comprise dans le sens (1), en distinguant selon les deux catégories principales de vérité, métaphysique et épistémique, que nous expliciterons. Cela nous permettra d'en venir au dernier point : le rôle que jouent les positions philosophiques dans la pratique scientifique.

6.4.2 Une « valeur-vraie » ? Ébauche empiriste

Comme nous l'avons déjà mentionné, l'expression « valeur-vraie » que nous proposons pour l'usage présent vise à rendre plus explicite l'absence de rapport qu'entretient le terme avec une quelconque notion de vérité. Le trait d'union a pour but, en particulier, de rappeler que l'expression elle-même ne véhicule aucun sous-entendu et qu'il pourrait lui être substitué toute autre désignation, comme celles, déjà évoquées, de paramètre, valeur théorique, valeur cible, etc. L'expression « valeur-vraie » correspond à une interprétation instrumentale du concept, qui n'est de ce point de vue qu'une fiction utile pour développer un mécanisme d'analyse de données expérimentales qui permet la formulation d'un résultat final sous la forme numérique. Cette approche est donc typique d'une position anti-réaliste, à laquelle il faut donner corps : on peut par exemple chercher à caractériser la valeur-vraie dans un cadre empiriste. Afin de tester la proposition selon laquelle la valeur vraie d'une grandeur serait à comprendre uniquement selon l'interprétation instrumentale que nous désignons par l'expression « valeur-vraie », nous proposons – à titre d'illustration, plus que d'analyse systématique – de l'intégrer dans une position telle que l'empirisme constructif de Bas van Fraassen.

L'empirisme constructif de van Fraassen⁵⁷ est vraisemblablement l'une des plus notables positions empiristes récentes et abouties. La démarche de van Fraassen procède d'une double volonté. D'une part, van Fraassen cherche à rendre les critères de choix théoriques indépendants d'exigences métaphysiques – c'est là sa motivation de départ pour fonder un empirisme. D'autre part, il tire les leçons de l'échec de l'empirisme logique, et se démarque notablement de ce dernier en s'efforçant de ne pas fonder sa propre position sur des considérations linguistiques. Ce faisant, il renouvelle un débat classique de l'opposition entre réalisme et empirisme, en réaffirmant la possibilité de tracer une frontière entre entités « observables » et « inobservables » – non sans avoir soigneusement distingué les catégories « inobservable » (qui concerne

57. Explicité en particulier dans van Fraassen (1980).

les entités) et « théorique » (qui concerne les termes et concepts)⁵⁸. Bien que van Fraassen reconnaisse le caractère arbitraire et donc flottant de la frontière tracée entre ces catégories, cette dichotomie est à la racine de ses développements et lui permet ensuite de déployer une épistémologie anti-réaliste qui fonde les critères d'acceptation des théories scientifiques sur leur capacité à être « empiriquement adéquates »⁵⁹. Dire d'une théorie qu'elle est empiriquement adéquate, c'est dire que toutes ses prédictions *observables* sont effectivement vérifiées par l'expérience⁶⁰. Dédire des expériences effectuées qu'une théorie est empiriquement adéquate reste une démarche ampliative, en ce qu'elle ne concerne pas seulement les observations disponibles à l'instant présent, mais aussi l'ensemble des observations que l'on peut envisager de faire⁶¹. L'adéquation empirique d'une théorie n'est donc jamais acquise. Cependant, van Fraassen considère que l'empirisme constructif apporte un gain substantiel par rapport au réalisme, en ce que le premier est moins exigeant que le second. L'empirisme constructif fonde l'acceptation des théories sur la croyance en l'adéquation empirique de ces dernières, et n'exige rien de plus ; le réalisme fonde quant à lui l'acceptation des théories sur la croyance que celles-ci sont vraies. Or, l'adéquation empirique ne porte *que* sur les conséquences *observables* de la théorie, donc sur une *portion* de la théorie, contrairement à un critère de vérité qui porterait non seulement sur les conséquences observables de la théorie mais aussi sur l'ensemble des constructions théoriques qui y sont développées⁶². La position de van Fraassen se démarque donc d'un réalisme scientifique en ce qu'elle n'implique aucun engagement ontologique : rien n'exige que les termes et concepts employés dans les théories ne renvoient à des entités existantes. Ce que van Fraassen défend, c'est que l'adéquation empirique permet de faire l'économie de questions métaphysiques qui n'apportent rien – en l'occurrence, ici, le lien entre termes théoriques et réalité. Ce faisant, affirme van Fraassen, l'empirisme constructif « nous délivre de la métaphysique »⁶³, suivant ainsi un objectif qui était déjà celui de l'empirisme de Carnap un demi-siècle plus tôt, ou encore de ceux de Mach, Helmholtz, Poincaré, Duhem, etc. Notons cependant que cela n'interdit pas pour autant de continuer à croire que les structures des théories scientifiquement valides sont (au moins en partie) contraintes par l'état d'une réalité extérieure à nous – mais cela suggère fortement qu'il n'y a pas de sens à envisager que les théories soient l'image ou le reflet de la réalité⁶⁴.

Au-delà de l'exposition très générale que nous venons de proposer, nous apercevons qu'un

58. van Fraassen (1980), "The Theory/Observation Dichotomy", pp.13–19.

59. van Fraassen (1980), p.18.

60. Comme l'expliquent Monton et Mohler (2014), la définition précise de l'adéquation empirique nécessite un développement plus sophistiqué, que l'on retrouve explicitement en détail dans van Fraassen (1980), et qui fait partie de ce qui est appelé conception « sémantique » des théories. « Une théorie est adéquate empiriquement, donc, si les apparences – "les structures qui peuvent être décrites dans les rapports expérimentaux et de mesure" – sont isomorphes aux sous-structures empiriques d'un modèle de la théorie », Monton et Mohler (2014), section 1.5 (dernière consultation le 25 janvier 2016). Pour une analyse de ce qu'est la conception sémantique des théories, et de la façon dont elles se distinguent des approches linguistiques de la première moitié du XX^e siècle, voir Vorms (2011).

61. Monton et Mohler (2014), section 3.4 : « Observable versus Observed » (dernière consultation le 25 janvier 2016).

62. van Fraassen (1980), pp.68-69.

63. van Fraassen (1980), p.69.

64. Soler (2008), p.143.

positionnement anti-réaliste permet de structurer l'usage et l'interprétation d'un concept instrumental de valeur-vraie. Si l'on admet l'adéquation empirique comme seul critère essentiel d'acceptation d'une théorie, alors une théorie qui fait l'usage d'un concept tel que celui de valeur-vraie sera jugée en fonction de ses seules prédictions observables, selon qu'elles soient ou non vérifiées par l'expérience. Ce qui nous est essentiel dans le cas présent, c'est que l'empirisme de van Fraassen ne prétend à aucun moment que l'on doive s'interdire d'employer un quelconque terme théorique – a fortiori celui de valeur-vraie. Il convient seulement d'accepter qu'il n'est pas nécessaire que ce concept soit plus qu'une fiction, qu'un outil de construction théorique au service de la structuration des raisonnements quantitatifs, des prédictions théoriques, et des mesures expérimentales, qui, elles, permettront de juger « l'adéquation empirique » de la théorie. L'empirisme constructif de van Fraassen permet en revanche de tempérer la portée métaphysique du terme « valeur vraie » ; il a pour vertu principale de ne pas chercher à faire correspondre la valeur-vraie à un quelconque état de la réalité.

Une telle position anti-réaliste ne rend pas nécessaire l'abandon du terme théorique lui-même ; en revanche, elle revient à en réviser sérieusement la signification. Il est difficile d'identifier si le positionnement des métrologues doit être interprété comme une adhésion à un anti-réalisme ou à un empirisme. Certains d'entre eux font parfois appel à des principes explicitement philosophiques. Ainsi Franco Pavese fait-il référence à Wittgenstein en appliquant le principe de vérification à la métrologie⁶⁵. Mais cette démarche est très isolée ; et le GUM, en particulier, ne fait preuve d'aucun positionnement philosophique explicite. Le concept instrumental de valeur-vraie semble tout à fait viable en principe – tant que l'approche philosophique générale dans laquelle elle s'insère, que celle-ci soit l'empirisme de van Fraassen ou autre chose, est elle-même valide. Cependant, si nous cherchons à mieux situer le positionnement des métrologues, il nous faut explorer d'autres options possible. Cela nous invite à envisager la valeur vraie sous la seconde acception que nous avons exposé au préalable, à savoir comme une interprétation littérale : « {vraie} {valeur d'une grandeur} ». Il nous faut donc explorer brièvement les théories philosophiques de la vérité.

6.4.3 Une valeur « vraie » ? Vérités métaphysique et épistémique

Comment comprendre « valeur vraie » selon une acception littérale, c'est-à-dire comme une « {vraie} {valeur d'une grandeur} » ? Selon une telle acception, le qualificatif « vraie » énonce une valeur de vérité effective ; pour élaborer une réponse, il nous faut donc préciser au préalable la signification que l'on accorde à la notion de vérité que l'on associe à un tel concept. On trouve dans la philosophie générale des sciences différentes théories de la vérité. Nous chercherons donc à déterminer si nous pouvons donner un cadre philosophique précis à la notion de valeur

65. « L'objectif du GUM peut être considéré comme compatible avec le “principe de vérification” de Wittgenstein : “la signification d'une question est la méthode d'y répondre”, une formulation qui dans d'autres traductions de l'allemand se lit “la signification d'un énoncé est sa méthode de vérification”. Sans entrer dans la question de la validité générale de ce principe – par lui-même non vérifiable – qui a été très contrastée dans d'autres domaines, à savoir ceux de la théologie et de la philosophie transcendantale, il y a peu de doutes quant à sa validité en science, à savoir dans la physique moderne, où la preuve expérimentale est généralement exigée pour étayer les affirmations théoriques et les modèles. Par conséquent, le concept de “valeur vraie” serait dénué de sens selon ce principe, puisqu'il est réputé invérifiable », Pavese (2009), p.1298.

vraie employée par les métrologues, et voir ainsi si les problématiques auxquels ces derniers se heurtent trouvent des éclaircissements à la lumière des développements déjà existants en philosophie.

Glanzberg a proposé une synthèse des conceptions « néo-contemporaines » de la vérité⁶⁶. Parmi celles-ci, la théorie de la « vérité-correspondance » est peut-être la forme la plus intuitive ou évidente de vérité⁶⁷ : un énoncé⁶⁸ est vrai en vertu de sa correspondance avec un élément de la réalité. Cela engage à s'attacher à un positionnement réaliste. En effet, l'adoption de cette notion implique de postuler l'existence d'une réalité objective indépendante de nous⁶⁹. De plus, elle implique de poser cette réalité comme arbitre de la valeur de vérité des énoncés. Interpréter la « valeur vraie » en termes de vérité-correspondance semble effectivement plutôt intuitif – la valeur d'une grandeur étant un élément d'une *proposition* sur l'état physique de la grandeur, elle ne s'identifie pas par essence à la réalité elle-même, mais est un moyen de description, exprimé par le langage, qui *correspond* à la réalité. L'on peut toutefois envisager d'aller plus loin, en se référant à une autre théorie de la vérité, celle de la vérité-identité : « selon la théorie de la vérité-identité, les propositions vraies ne correspondent pas aux faits, elles *sont* des faits »⁷⁰. La valeur vraie de la grandeur serait alors un élément de la réalité en propre, en considérant par exemple que le monde physique est quantitatif par essence, et que les nombres sont constitutifs non pas de la façon dont nous nous le représentons, mais de sa nature même⁷¹. L'appel à la vérité-correspondance semble moins exigeant et moins radical qu'une interprétation formulée dans les termes de la vérité-identité. Cependant, nous ne chercherons pas à rentrer ici dans des discussions sur le contenu détaillé des deux théories de la vérité mentionnées jusqu'ici, qu'il s'agisse de correspondance ou d'identité. Plutôt que de chercher à faire l'exégèse des approches, et de constater comment elles diffèrent de l'une à l'autre, nous souhaitons au contraire pointer vers une caractéristique remarquable que celles-ci ont en commun : elles sont des conceptions *métaphysiques* de la vérité. Or, les travaux contemporains sur la vérité nous rappellent que nous pouvons envisager encore d'autres façons de concevoir

66. Glanzberg (2013)

67. À tel point que son caractère évident constitue même l'un des principaux arguments en sa faveur : « (l) principal argument positif donné par les partisans de la théorie de la correspondance de la vérité est son évidence », Marian (2015), section 4.

68. Dans un premier temps, nous utiliserons de façon générique le terme « énoncé » pour désigner les objets qui portent une valeur de vérité, tout en gardant à l'esprit que nous aurions également pu parler par exemple de « croyances », et que leur nature fait débat – même si ces débats ont récemment perdu en intensité, comme le précise Glanzberg (2013), section 6.1 : « les débats contemporains sur la vérité ont été beaucoup moins préoccupés par la question des porteurs de vérité que ne le furent les débats classiques. »

69. « La théorie de correspondance de la vérité est en son cœur une thèse ontologique : une croyance est vraie s'il existe une entité appropriée », Glanzberg (2013), section 1.1.2.

70. Marian (2015), section 8.3 (l'auteur souligne).

71. Joel Michell a défendu une position réaliste de la mesure qui s'apparente à une telle position. Pour Michell, « quand l'hypothèse selon laquelle un attribut est quantitatif est acceptée, alors en accompagnement de cette hypothèse, comme part du même emballage théorique, il est admis que différentes grandeurs de l'attribut se tiennent dans des relations de rapport, ces relations étant des instances de nombres réels. Ainsi, les nombres réels sont tenus pour être situés comme des caractéristiques intrinsèques dans le contexte empirique de la mesure », Michell (2005), p.292. Il n'est pas nécessaire de chercher à faire rentrer la position de Michell dans une case bien particulière, mais il est clair que celle-ci dénote par son caractère fortement réaliste.

la vérité, sans engagement métaphysique.

On distingue ainsi deux catégories de théories de la vérité. Il existe, d'un côté, celles qui accordent à la vérité une valeur métaphysique : dans ce type de théories, c'est l'adéquation de la proposition avec la réalité qui fait foi. Dans ce cas, un agent peut croire ou non en la vérité ou la fausseté d'une proposition, celle-ci étant déterminée par un état extérieur des choses indépendant de l'état mental de l'agent ; mais l'état de vérité de la proposition n'est pas décidable, précisément parce que le critère de vérité met en jeu des éléments indépendants de l'esprit de l'agent. À cela s'oppose une seconde catégorie de théories de la vérité, pour lesquelles la vérité est décidable sur des critères qui nous sont accessibles, tels que la croyance, la connaissance, la vérification, etc.⁷². Ainsi, le vérificationnisme défend l'idée qu'un énoncé est vrai s'il est vérifiable, « c'est-à-dire qu'il existe une procédure de vérification que nous pourrions en principe mener et qui nous indiquerait que l'énoncé en question a été vérifié »⁷³. Le cohérentisme stipule qu'un énoncé est vrai s'il s'insère dans un ensemble cohérent d'énoncés⁷⁴. On retrouve aussi des formes issues du pragmatisme⁷⁵, ou encore une conception sociale fondée sur le consensus, où un énoncé est vrai *en vertu du fait* qu'il est communément accepté⁷⁶. Dans tous les cas, on est alors en présence d'une conception de la vérité fondamentalement différente de celle de vérité-correspondance ou de vérité-identité, puisque l'existence d'une réalité objective (quelle qu'elle soit) n'intervient plus dans le critère de vérité. D'ailleurs, de façon remarquable, ces théories épistémiques de la vérité rendent possible la détermination exacte de la valeur de vérité d'un énoncé : par exemple, la détermination de la cohérence d'un ensemble d'énoncés ne fait pas appel à des éléments extérieurs inaccessibles. Ainsi, « (p)ar rapport à la vérité-correspondance, la définition de la vérité comme cohérence a l'avantage d'offrir un critère empirique tout à fait opérant pour décider concrètement de la vérité ou de la fausseté des théories »⁷⁷. Dans ce cas, le statut de vérité d'un énoncé est temporaire et peut évoluer au gré des connaissances – en effet, la cohérence d'une théorie peut être remise en question à la lumière de nouveaux résultats expérimentaux ou de nouveaux développements théoriques. Pour résumer, on peut donc mettre en regard deux catégories générales de théories de la vérité, l'une métaphysique où le critère de vérité implique une relation avec un réel indépendant de nous, l'autre épistémique où la vérité d'un énoncé est décidable sur des critères qui nous sont directement accessibles. C'est cette distinction qui nous intéresse spécifiquement, au-delà de la multiplicité des théories individuelles.

Une fois encore, l'adoption de l'une ou l'autre des théories ainsi que la discussion de leurs vertus et de leurs limites est une question générale de philosophie qui ne trouve pas de réponse définitive et que le cas d'étude spécifique de la métrologie contemporaine ne permet certainement pas de résoudre. Nous préférons, de ce fait, nous orienter vers le rapport qu'en-

72. Glanzberg (2013), section 4.2.

73. Glanzberg (2013), section 4.2.

74. « Une croyance est vraie si et seulement si elle fait partie d'un système cohérent de croyances », Glanzberg (2013), section 1.2.

75. « Peirce et James sont tous deux associés avec le slogan selon lequel : la vérité est satisfaisante à croire », Glanzberg (2013), section 1.3.

76. Soler (2008), pp.58–59.

77. Soler (2008), pp.57–58.

treignent les acteurs de la métrologie avec les questions philosophiques. Nous avons identifié une catégorisation des théories de l'erreur en des conceptions métaphysiques et d'autres épistémiques ; est-il possible de capturer la façon dont les métrologues conçoivent les concepts qu'ils emploient selon l'une ou l'autre de ces catégories ? Une fois encore, l'argument d'inconnaissabilité peut nous servir de marqueur pour interpréter le positionnement philosophique des métrologues. Il nous apparaît assez clairement le concept de valeur vraie doit être interprété selon une conception métaphysique de la vérité. En effet, nous l'avons vu, les conceptions épistémiques de la vérité sont décidables selon des critères accessibles. Si le statut de la valeur vraie d'une grandeur était conçu en termes épistémiques, par exemple cohérentistes ou sociaux, il serait alors tout à fait possible de déterminer si une valeur est ou non « vraie » – et par conséquent, il n'y aurait pas de raison de tenir la valeur vraie d'une grandeur pour « inconnaissable ». L'argument d'inconnaissabilité entre donc en contradiction avec les formes épistémiques de vérité, ce qui indique qu'il faut se tourner vers une interprétation métaphysique de la vérité. L'argument d'inconnaissabilité prend alors la forme suivante : si l'on interprète la « valeur vraie » selon une interprétation métaphysique de la vérité, alors le caractère « vrai » ou « faux » d'un résultat expérimental n'est pas décidable ; puisqu'il n'est jamais possible de savoir si un résultat de mesure est vrai ou faux, alors il n'est pas possible de connaître la « valeur vraie » ; il faut donc interpréter différemment le concept.

Les catégories épistémiques de la vérité ne sont pas pour autant satisfaisantes pour réinterpréter le concept de valeur vraie. En effet, il existe déjà un concept correspondant à une valeur *consensuelle* ou à une valeur *cohérente*. En effet, l'opposition entre les deux catégories de vérité, métaphysique et épistémique, nous semble tout particulièrement affleurer lorsque l'on confronte le concept de valeur vraie d'une grandeur à un autre terme de métrologie, celui de « valeur conventionnelle », qui a connu un sort singulier lors des dernières décennies. Puisque la valeur vraie d'une grandeur visée soit réputée inconnaissable, ce n'est pas à elle que les opérateurs et les usagers vont se référer pour prendre des décisions ou encore juger de la validité d'une hypothèse théorique. Il leur faut se référer à quelque chose d'empiriquement accessible, et ils s'accordent ainsi sur une valeur consensuelle qu'ils estiment légitime d'utiliser : c'est la valeur *attribuée* à la grandeur, une estimation de la valeur vraie dont l'usage n'est que temporaire. Les scientifiques et les usagers s'accordent donc sur le choix d'une estimation considérée comme la meilleure à un instant donné, qui est *conventionnellement* adoptée – temporairement, jusqu'à ce que de nouvelles mesures suggèrent de changer de valeur conventionnelle. L'appel à une valeur conventionnelle vise donc à écarter l'écueil de l'inconnaissabilité de la valeur vraie : cette dernière étant inaccessible, on s'en *réfère* à une valeur conventionnelle considérée comme acceptable (selon le contexte). Le calcul détaillé de cette valeur conventionnelle est l'affaire à la fois des méthodes statistiques qui sont mobilisées en métrologies, décrites à la partie I, et d'un processus social de consensus, qui peut par exemple se conclure sur la publication de valeurs recommandées⁷⁸, lesquelles sont alors *cohérentes* entre elles et cohérentes avec les théories physiques acceptées.

La notion de valeur conventionnelle est définie dans le VIM depuis la première édition. La lecture des différentes éditions du VIM révèle d'ailleurs une évolution significative. Dans les

78. C'est ce que fait le CODATA pour les constantes de la physiques ; nous renvoyons pour cela à la partie III.

deux premières éditions du VIM (VIM1, 1984 et VIM2, 1993) ce concept est nommé « valeur conventionnellement vraie » (tableau 6.2). Cependant, cette terminologie est abandonnée dans la troisième édition du VIM (VIM3, 2008), où l'on retrouve la notion sous l'intitulé « valeur conventionnelle », la définition qui lui correspond marquant de plus une nette distanciation avec les éditions précédentes, comme l'atteste la note n° 3 : « le terme “valeur conventionnellement vraie” est quelquefois utilisé pour ce concept, mais son utilisation est déconseillée » (voir tableau 6.2). De façon remarquable, le VIM3 opère un deuxième mouvement dans la continuité de cette première transformation. Ainsi, le terme « erreur de mesure » ne se rapporte plus directement à la valeur vraie de la grandeur comme cela est traditionnellement envisagé. En lieu et place, l'erreur de mesure est définie comme la « différence entre la valeur mesurée d'une grandeur et une valeur de référence »⁷⁹ (tableau 6.3). Deux cas distincts sont envisagés, et à ces deux cas correspondent en fait deux concepts bien différents d'erreur de mesure. Dans le premier cas, l'erreur de mesure est une erreur d'*étalonnage* : l'indication donnée par un instrument de mesure est erronée parce qu'elle ne correspond pas à la valeur communément admise de la grandeur mesurée. Il n'est pas exclu, par ailleurs, que l'indication soit égale à la valeur vraie de la grandeur mesurée ; seulement, ce n'est pas le critère d'erreur employé ici. Dans le second cas, l'erreur de mesure est *inconnue* et est relative au rôle *prédictif* de la mesure : il s'agit de chercher à déterminer la valeur vraie de la grandeur visée, par exemple pour participer à la construction d'une nouvelle valeur conventionnelle, qui sera tenue pour meilleure que la précédente. La double signification de l'erreur de mesure a été discutée par les métrologues, qui ont perçu la différence de nature qui sépare ces deux définitions⁸⁰. Pour rendre plus explicite la dualité de la définition de l'erreur de mesure, le VIM3 introduit un concept intermédiaire, celui de « valeur de référence », lui-même défini dans la section 5 du document (tableau 6.4). Selon les cas, la valeur de référence peut être la valeur vraie de la grandeur, ou une valeur conventionnelle.

Ce qui nous semble remarquable ici, c'est que cette double définition pointe vers une dualité de l'activité de mesure elle-même – viser l'accord intersubjectif d'une part (valeur conventionnellement vraie), et accéder à une réalité objective extérieure à nous, d'autre part (valeur vraie) ; ce que l'on peut mettre en miroir des deux catégories générales de théories de la vérité que nous avons discutées plus tôt dans cette section. Ainsi, la « valeur conventionnellement vraie » est « vraie » dans le sens où elle est le produit d'un consensus – l'un des avatars épistémiques de la vérité, ce consensus provenant du fait qu'elle forme un ensemble cohérent avec les autres valeurs conventionnelles et les théories acceptées. Par contraste, la « valeur vraie » reste à interpréter selon une acception métaphysique de la vérité. On peut comprendre la confusion que pouvait engendrer la superposition, dans les VIM 1 et 2, de deux types bien différents de vérités ; et expliquer ainsi la contre-indication spécifiée dans la troisième édition du VIM : « le terme “valeur conventionnellement vraie” est quelquefois utilisé pour ce concept, mais son utilisation est déconseillée » (tableau 6.2 p.185). Cette contre-indication, associée à l'abandon du terme « valeur conventionnellement vraie » dans le VIM3, est une preuve supplémentaire que

79. Joint Committee for Guides in Metrology (JCGM) (2012), p.25.

80. La double nature de l'erreur de mesure est discutée par Ehrlich et Dybkaer (2012). Les rédacteurs du VIM semblent s'orienter vers l'introduction d'une notion d'« erreur d'indication » pour désigner les erreurs se rapportant à une valeur conventionnelle ; voir Ehrlich (2014).

| | |
|---------------------------------------|--|
| VIM1 (1984) (p.10) | Valeur conventionnellement vraie (d'une grandeur) : valeur d'une grandeur qui peut être substituée à la valeur vraie dans un but déterminé. <i>Note</i> Une valeur conventionnellement vraie est, en général, considérée comme suffisamment proche de la valeur vraie pour que la différence puisse être non significative pour le but donné. <i>Exemple</i> Au sein d'une organisation, la valeur attribuée à un étalon de référence peut être prise comme étant la valeur conventionnellement vraie de la grandeur réalisée par l'étalon. |
| <hr/> | |
| VIM2 (1993) (p.17) | Valeur conventionnellement vraie (d'une grandeur) : valeur attribuée à une grandeur particulière et reconnue, parfois par convention, comme la représentant avec une incertitude appropriée pour un usage donné [...] |
| <hr/> | |
| VIM3 (2008) (pp. 20–21) | Valeur conventionnelle : valeur attribuée à une grandeur par un accord pour un usage donné Exemple 1. Valeur conventionnelle de l'accélération due à la pesanteur ou accélération normale de la pesanteur, $g_n = 9,806\ 65\ \text{m} \cdot \text{s}^{-2}$. [...] NOTE 1 Le terme « valeur conventionnellement vraie » est quelquefois utilisé pour ce concept, mais son utilisation est déconseillée. NOTE 2 Une valeur conventionnelle est quelquefois une estimation d'une valeur vraie. NOTE 3 Une valeur conventionnelle est généralement considérée comme associée à une incertitude de mesure convenablement petite, qui peut être nulle. |

TABLEAU 6.2 – Différentes définitions de « valeur conventionnellement vraie » dans les deux premières éditions du *Vocabulaire International de Métrologie* (VIM) et de « valeur conventionnelle » dans la troisième édition du VIM. Notons que l'insistance sur le caractère conventionnel est plus marqué dans la définition du VIM3 : la valeur est attribuée « par accord ». Dans le VIM2, la valeur était attribuée « parfois par convention ». Dans le VIM1, le caractère conventionnel n'apparaît que dans l'intitulé du concept.

| | |
|----------------------------------|--|
| VIM3 (2008) (p.25) | <p>Erreur de mesure : différence entre la valeur mesurée d'une grandeur et une valeur de référence</p> <p>NOTE 1 Le concept d'erreur peut être utilisé</p> <p>a) lorsqu'il existe une valeur de référence unique à laquelle se rapporter, ce qui a lieu si on effectue un étalonnage au moyen d'un étalon dont la valeur mesurée a une incertitude de mesure négligeable ou si on prend une valeur conventionnelle, l'erreur étant alors connue,</p> <p>b) si on suppose le mesurande représenté par une valeur vraie unique ou un ensemble de valeurs vraies d'étendue négligeable, l'erreur étant alors inconnue. [...]</p> |
|----------------------------------|--|

TABLEAU 6.3 – « Erreur de mesure » dans la troisième édition du *Vocabulaire International de Métrologie* (VIM).

| | |
|----------------------------------|--|
| VIM3 (2008) (p.53) | <p>Valeur de référence : valeur d'une grandeur servant de base de comparaison pour les valeurs de grandeurs de même nature</p> <p>NOTE 1 La valeur de référence peut être une valeur vraie d'un mesurande , et est alors inconnue, ou une valeur conventionnelle , et est alors connue. [...]</p> |
|----------------------------------|--|

TABLEAU 6.4 – « Valeur de référence » dans la troisième édition du *Vocabulaire International de Métrologie* (VIM).

les métrologues, et les scientifiques en général, on tendance à réserver le terme « vrai » pour une acception métaphysique de la vérité, et que la révision de l'interprétation du concept de « valeur vraie » ne passe pas par l'adoption d'une conception épistémique de la vérité.

Il nous semble donc que la question que nous avons soulevée quant à l'interprétation de l'usage du terme « vrai » par les métrologues peut être tranchée sans ambiguïté. La notion courante de « valeur vraie » qui est conceptualisée dans la littérature métrologique est bien une notion métaphysique – et c'est d'ailleurs précisément en vertu de cela qu'elle se révèle aussi problématique. Ayant interprété le positionnement des métrologues sur ce point, nous pouvons désormais revenir sur l'argument central d'inconnaissabilité et ses conséquences quant à la validité du concept de valeur vraie d'une grandeur.

6.5 La valeur vraie à la croisée de questionnements scientifiques et philosophiques

À ce niveau, il nous semble important de distinguer deux modes d'analyse, auxquels correspondent deux questions sensiblement différentes. Le premier mode a trait à la légitimité des différents positionnements philosophiques possibles. Y a-t-il, parmi les différentes options recensées, *une* conception correcte de la valeur vraie ? Ou bien, au contraire, est-ce que l'adoption de l'une ou l'autre des doctrines est du ressort d'un choix, qu'il soit individuel ou collectif, guidé par des critères épistémologiques et méthodologiques ? Nous n'entrevoions pas d'issue claire. Nous considérons que cette question est subordonnée aux débats engagés dans la philosophie générale des sciences – et qu'elle ne trouve pas de réponse spécifique dans le cadre précis de la mesure. S'il s'agit de trancher quant à la meilleure théorie de la vérité, alors nous renvoyons ce problème à une question de philosophie générale des sciences, tant il nous semble acquis que notre cas d'étude spécifique ne nous permettra pas de le régler. S'il s'agit par ailleurs de se positionner comme arbitre des pratiques scientifiques, nous émettons également d'importantes réserves. En effet, les pratiques et concepts scientifiques sont discutés et redéfinis en permanence dans l'espace spécialisé. Dès lors, l'attitude qui consisterait à vouloir se substituer au travail réflexif permanent des scientifiques dans leur propre espace conceptuel, et à appliquer sans discernement des principes philosophiques *a priori* nous semble discutable. Cela ne signifie pas pour autant qu'il soit impossible au philosophe de prendre part au débat conceptuel ; simplement qu'il ne peut que difficilement prétendre posséder un statut privilégié par rapport aux acteurs eux-mêmes. Notre réticence à nous placer comme arbitre, ainsi que l'indécision des questions générales de philosophie à ce sujet nous incite à la plus grande prudence quant au positionnement que nous pourrions être amené à adopter. Plutôt que de chercher à *nous* positionner philosophiquement, et à identifier la meilleure doctrine que les scientifiques et métrologues devraient adopter, nous préférons au contraire partir du positionnement de ces derniers, et identifier ce que ce positionnement révèle quant aux rapports entre science et philosophie. En effet, si nous assumons une dépendance « de haut en bas » envers les débats plus théoriques de la philosophie générale des sciences, il nous semble essentiel d'isoler ce qui, dans l'étude de notre objet, est susceptible d'enrichir en retour la philosophie des sciences. Ici, notre analyse s'intéresse tout particulièrement à l'attitude qu'adoptent les scientifiques (plus spéci-

fiquement les métrologues) vis-à-vis des questionnements épistémologiques et métaphysiques qui ne manquent pas de s'offrir à eux. Nous entrevoyons donc ici un second mode d'analyse. Celui-ci concerne le positionnement philosophique des acteurs eux-mêmes. Est-il possible de faire entrer le positionnement épistémologique global d'un document comme le GUM dans une catégorie précise ?

Il nous faut d'ailleurs nuancer d'emblée la teneur des propos que nous serions tentés de formuler ici. Nous cherchons ici à identifier des positionnements philosophiques chez les acteurs. Or, nous n'avons pas l'intention de recenser les positionnements *individuels* des membres de la communauté des métrologues, et nous n'avons aucunement la prétention de pouvoir discerner les convictions philosophiques de chacun, qui peuvent différer grandement d'un acteur à l'autre, et qui, pour la plupart, ne sont pas exprimées dans les publications. Nous nous attachons plutôt à comprendre la philosophie « appliquée » qui peut être dégagée de la lecture de la littérature spécialisée dans son ensemble, sans chercher à affirmer que cette attitude philosophique est adoptée par tous sans discernement.

Ayant exprimé la prudence avec laquelle nous souhaitons nous diriger dans la voie choisie, nous développerons notre discussion en deux temps. Dans un premier temps, nous identifierons la position des métrologues comme un acte de méfiance envers des discussions d'ordre métaphysique, sans pour autant constituer un désengagement envers le réalisme scientifique. Dans un second temps, nous reviendrons sur l'argument de l'inconnaisabilité, et nous discuterons la portée de cet argument en métrologie. Cela nous permettra de conclure sur les vertus épistémiques du réalisme appliqué à la question de l'erreur en science.

6.5.1 Un désengagement anti-métaphysique

Nous avons soutenu, dans la section 6.3 l'idée que l'approche contemporaine du GUM et de ses suppléments reste fondamentalement structurée autour d'un terme théorique, utilisé dans l'appareillage technique, qui est usuellement appelé « valeur vraie ». Pourquoi retrouve-t-on, dans la littérature métrologique récente, des questionnements autour de la notion de valeur vraie ? Nous émettons l'hypothèse suivante : les métrologues cherchent à éliminer du champ scientifique les contenus qui leur semblent ne pas être du ressort de l'activité scientifique elle-même. Le questionnement serait alors celui de la pertinence d'une réflexion éminemment philosophique, et la réponse associée serait dans ce cas d'affirmer que tout principe philosophique auquel le contenu scientifique est indifférent n'a pas sa place dans le discours scientifique. C'est du moins l'une des façons dont on peut interpréter, en extrapolant, la remarque formulée par Cohen et DuMond en 1965 :

Personne ne peut garantir que l'évaluation des constantes fondamentales à une époque donnée aboutit à de « vraies » valeurs. La vérité absolue, si ces mots ont un sens, est au-delà du royaume de la physique.⁸¹

Cohen et DuMond expriment assez clairement l'idée qu'un questionnement sur la « vérité absolue », comprise comme une notion transcendante et métaphysique, dépasse le domaine de

81. Cohen et DuMond (1965), p.540. On peut aussi se référer à Eisenhart citant Cassius J. Keyser : « (l)a certitude absolue est un privilège des esprits incultes – et des fanatiques. C'est, pour le peuple scientifique, un idéal inatteignable », Eisenhart (1963), p.171.

la physique – ce faisant, ils marquent une neutralité, ou une indifférence, quant à la question de l'interprétation du concept de valeur vraie d'une grandeur. Les problèmes qui incluent une dimension métaphysique, comme peut l'être un réalisme qui accorde une valeur vraie aux grandeurs physiques, sont directement concernés par un tel amendement – ce réalisme apparaîtrait ainsi, pour les scientifiques, comme un choix auquel leurs pratiques seraient indifférentes. En particulier, une forme spécifique de réalisme qui accorderait au concept de valeur vraie d'une grandeur un statut métaphysique bien particulier n'est pas nécessaire pour développer les méthodologies ordinaires d'analyse des données expérimentales. C'est en un sens ce que Bich défend :

(L) a question de savoir si la valeur vraie et les erreurs sont ou non des concepts logiquement valides est immatérielle au développement d'une théorie de l'incertitude formellement correcte.⁸²

Dans la continuité, si l'interprétation réaliste de la valeur vraie n'est pas nécessaire, on peut également se demander si elle n'est pas alors même un obstacle à la compréhension des concepts. C'est le questionnement qu'expriment en creux Ehrlich, Dykaer et Wöger :

Si la valeur vraie [...] n'est pas connaissable en principe, alors la question se pose de savoir si le concept de valeur vraie est nécessaire, utile ou même nuisible!⁸³

En amenant une notion réputée « inconnaissable » comme la valeur vraie ou l'erreur, un réalisme jugé trop naïf pourrait être perçu négativement, puisqu'il serait ainsi à l'origine de l'introduction de difficultés conceptuelles pourtant inutiles dans les méthodes d'analyse d'incertitude. Le tournant épistémique qu'a pris la métrologie, que nous avons évoqué au chapitre 5 de la partie I, et que nous avons illustré par l'adhésion aux thèses bayésiennes, participe de ce mouvement général de purification du contenu scientifique qui vise à éviter de tirer des conclusions inappropriés de remarques qui n'ont pas initialement leur place dans le discours scientifique.

Mentionnons, sans la développer, une autre hypothèse qui consiste à suggérer que le positionnement exprimé dans les documents généraux de métrologie pourrait également être interprété comme l'affirmation d'une forme de maturité des méthodes de calcul d'incertitude, qui dès lors, s'émanciperait d'un certain nombre de questionnements conceptuels et philosophiques. La maîtrise grandissante des concepts et théories associés au calcul d'incertitude permettrait ainsi aux métrologues et aux scientifiques de se passer des questions philosophiques autrefois nécessaires pour donner un cadre intelligible à l'ensemble de la discipline, à l'instar de ce qui a eu lieu dans l'après-guerre à propos de la physique quantique, où les questionnements philosophiques des « pères fondateurs » sont apparus de moins en moins essentiels pour comprendre la théorie elle-même⁸⁴ – le cadre formel devenant de plus en plus autonome.

L'apparent désengagement anti-métaphysique des métrologues ne nous semble pas automatiquement induire un désengagement envers le réalisme, comme cela pourrait paraître au

82. Bich (2012a), p.2156.

83. Ehrlich, Dybkaer et Wöger (2007), p. 209.

84. Voir Lévy-Leblond (2006), p.920.

premier abord. Il nous semble que les métrologues sont hostiles à l'idée générale de « valeur vraie » de par sa teneur trop métaphysique et son engagement réaliste trop marqué, mais que leur pensée conserve un arrière-fond réaliste. La littérature métrologique ne contient certes pas une unique réponse à l'argument d'inconnaisabilité, mais la façon dont certains métrologues mobilisent cet argument pour chercher à rendre silencieux le terme théorique de « valeur vraie » nous semble montrer qu'ils n'envisagent pas vraiment la possibilité de conserver ce terme dans une approche pleinement anti-réaliste. Le désir d'éliminer la valeur vraie n'est qu'une très légère touche d'anti-réalisme, qui concerne seulement l'idéalité des modèles mathématiques où les grandeurs ont des valeurs bien définies, et qui constate l'inaccessibilité empirique (triviale) de ces valeurs. Cela laisse beaucoup de marge pour des formes atténuées de réalisme, comme le réalisme convergent où l'on envisage de resserrer de plus en plus la distribution des valeurs possibles d'une grandeur. Nous interprétons donc la position dominante des documents de métrologie comme un attachement *non viscéral* à un réalisme modéré qui resterait alors le mode majeur d'interprétation des contenus ; en cela, notre position est somme toute assez ordinaire⁸⁵.

Nous pouvons essayer d'aller plus loin que ce premier constat. En quoi l'attachement des scientifiques envers le réalisme scientifique présente-t-il un intérêt pour la philosophie des sciences ? La discussion que les métrologues ont engagé autour de questionnements épistémologiques, comme celui qui concerne le statut de la valeur vraie, permet d'interroger les relations science-philosophie et de croiser différentes formes ou réflexions, techniques, conceptuelles, et philosophiques. Sur la base des acquis des sections précédentes, nous voulons désormais analyser la façon dont les considérations philosophiques jouent un rôle au sein même de la pratique scientifique. En particulier, ici, notre propos visera à montrer que l'attachement au réalisme scientifique a des conséquences effectives sur le contenu méthodologique des sciences – contenu méthodologique qui est précisément l'un des objets de travail de la métrologie.

6.5.2 Inconnaisabilité, valeur vraie et localisation du réalisme scientifique dans la pratique scientifique

Nous l'avons vu, l'argument central de la critique de la valeur vraie en métrologie gravite autour de l'idée d'« inconnaisabilité ». Puisque la valeur vraie est réputée inconnaisable, elle apparaît comme une pure spéculation dont il peut être utile de faire l'économie, ou qu'il n'est pas nécessaire d'interpréter littéralement. Cependant, il nous faut à ce stade interroger l'usage du terme « inconnaisable » et sa signification. L'emploi de ce terme peut sembler justifié, en ce qu'il apparaît que, si la valeur vraie se rapporte à une réalité extérieure de nous et indépendante de notre esprit, alors elle est, par nature, inaccessible. De fait, c'est là l'application à la valeur vraie d'un problème philosophique fondamental, que Tiercelin formule ainsi :

(L)e réalisme est soumis à une tension caractéristique : si le monde est indépendant de notre connaissance, comment pouvons-nous réellement le connaître ?⁸⁶

Compris ainsi, l'argument, tel qu'il est employé par les métrologues, se réduirait à un constat d'échec, celui de l'impossibilité de formuler une quelconque connaissance à propos du « réel ».

85. Ce réalisme est à rapprocher de la description qu'en fait Wimsatt par exemple, [Wimsatt \(1987\)](#), pp.23–24.

86. [Tiercelin \(2006\)](#), p.936.

Or, il est probable que bien peu de métrologues seraient tentés d'accepter une limite aussi radicale – c'est ce que nous avons défendu à la sous-section précédente. L'argument d'inconnaissabilité ne fait qu'exprimer d'une autre façon l'idée selon laquelle, dès lors que l'on adhère à une conception métaphysique de la vérité, il n'est possible que de croire en la vérité d'un énoncé, sans que ce statut de vérité ne soit décidable de façon définitive (tout du moins dans les sciences empiriques) – puisqu'il va de soi que nous n'avons accès à l'« envers du décor » que par des moyens de connaissance imparfaits. La discussion se fait alors le miroir des débats qui ont eu lieu en philosophie, en particulier autour de la notion de « vérité-correspondance », que Marian résume ainsi :

L'objection qui pourrait bien avoir été la plus efficace pour provoquer le mécontentement avec la théorie de la correspondance est basée sur une préoccupation épistémologique. En un mot, l'objection tient à ce qu'une théorie de la vérité correspondance doit inévitablement mener au scepticisme à propos du monde extérieur puisque la correspondance exigée entre nos pensées et la réalité n'est pas déterminable. [...] Il est généralement souligné que nous ne pouvons pas sortir de notre propre esprit pour comparer nos pensées avec la réalité indépendante de l'esprit. [...] Puisque cela est impossible, puisque tout notre accès au monde passe par notre cognition, la théorie de la correspondance rend la connaissance impossible (cf Kant [...]).⁸⁷

De fait, comme le signale Marian, cette position remonte effectivement à une problématique déjà abordée par Kant, pour qui la définition de la vérité comme accord avec l'objet est un « diallèle »⁸⁸, c'est-à-dire un cercle vicieux ; en effet, dit-il, « puisque l'objet est hors de moi et que la connaissance est en moi, tout ce que je peux apprécier c'est si ma connaissance de l'objet s'accorde avec ma connaissance de l'objet »⁸⁹. En d'autres termes on ne peut pas accéder à l'objet « en soi » pour le comparer à la connaissance qu'on en a⁹⁰.

L'argument d'inconnaissabilité repose sur le fait qu'il est impossible d'attester que les énoncés que nous formulons soient vrais. Mais ce constat est uniquement un constat d'*incertitude*, et il n'implique en rien l'impossibilité que ces énoncés soient vrais⁹¹, ni qu'il faille abandonner toute prétention à tendre vers la vérité, et par association abandonner toute prétention à la connaissance. Dédire de l'argument d'inconnaissabilité qu'il faut abandonner ou réformer l'idée même d'une valeur « vraie » est une position qui en principe semble tout à fait tenable – à condition de trouver un système alternatif viable –, mais ce n'est pas pour autant une conclusion *nécessaire* et indispensable. Il est tout à fait possible de maintenir l'attachement à un réalisme modéré, et de conserver l'usage de la valeur vraie tout en constatant, dans le même temps, que nos connaissances scientifiques sont toujours marquées d'une forme d'in-

87. Marian (2015), section 9.2.

88. Kant (2007), p.54.

89. Kant (2007), p. 54.

90. Je remercie Philippe Stamenkovic de ses explications sur la conception kantienne de la vérité, et d'avoir attiré mon attention sur les extraits cités ici.

91. Mari rappelle également ce point : « En effet, la reconnaissance pragmatique de la présence inévitable d'incertitude dans la mesure n'a pas nécessairement de conséquence sur la vérité dans la mesure : être incertain à propos de *X* n'implique pas de refuser que la vérité sur *X* peut être exprimée », Mari (2005), p.265.

certitude. L'idée même d'associer à un résultat de mesure une incertitude de mesure témoigne d'une quête permanente, celle de délimiter les contours de notre (manque de) connaissance, et de réduire toujours plus l'espace dans lequel nous pourrions situer les différentes valeurs des constantes physiques. La tendance toujours plus intense à la réduction progressive des incertitudes de mesure des différentes constantes physiques trouve un cadre philosophique naturel particulièrement intuitif dans le réalisme convergent : on pense connaître de mieux en mieux la valeur vraie de la constante en question parce qu'on tend asymptotiquement, avec parfois quelques cahots, vers la valeur vraie qui constitue l'horizon inaccessible de la quête expérimentale. Il demeure bien entendu en permanence un doute irréductible. Cependant, l'on peut à l'inverse constater également qu'il existe de très bonnes raisons de *croire* en la validité des résultats de mesure, des théories scientifiques, etc. De fait, observer que certains métrologues statuent que toute valeur vraie d'une grandeur est « inconnaisable » semble un peu surprenant lorsque, dans le même temps, les scientifiques affirment connaître la valeur de l'anomalie du moment magnétique de l'électron avec une précision portant sur la onzième décimale, considérée comme l'un des tests les plus puissants de la théorie de l'électrodynamique quantique⁹². Plus encore, il est surprenant de voir les textes de métrologie affirmer que la valeur vraie d'une grandeur est inconnaisable, tout en expliquant que la formulation probabiliste d'un résultat de mesure, selon l'interprétation épistémique, est la mesure de l'état « de connaissance » de l'expérimentateur à propos de la valeur vraie de la grandeur.

On pourra rétorquer de nouveau, et à raison, que la précision annoncée ne peut pas être garantie ; mais c'est là la seule véritable conclusion que l'on peut tirer de l'argument d'inconnaisabilité, et que l'on peut résumer de la façon suivante : nos connaissances sont faillibles. Nous pensons qu'il serait peut-être opportun de préférer le terme d'« inaccessible » à celui d'« inconnaisable », pour caractériser la tension inhérente au réalisme qu'ont souligné les philosophes, parmi lesquels Kant. Le terme « inaccessible » véhicule une idée différente et nous semble plus pertinent – ce n'est pas parce que l'on ne peut pas accéder à la valeur vraie d'une grandeur que celle-ci n'est pas connaissable, parce que l'idée commune de connaissance présente un caractère inévitablement faillible.

Comment peut-on finalement comprendre le concept de « valeur vraie » d'une grandeur ? Il ne s'agit pas nécessairement de prétendre que le monde physique est quantitatif par nature, et que la mesure n'est qu'un dévoilement de la réalité numérique du monde ; plus simplement, cela revient à affirmer, contre une conception purement instrumentale, que les énoncés scientifiques quantitatifs présentent une valeur de vérité. La question ontologique – quelle est la forme des objets dans la réalité ? – et celle de la nature précise du rapport qu'entretiennent les propositions théoriques avec les entités de la réalité restent tout à fait ouvertes, sans qu'il soit nécessaire d'y répondre pour conserver un attachement à la valeur vraie.

De façon plus concrète, le concept de « valeur vraie » véhicule l'idée que si le test d'un résultat expérimental – test de la cohérence entre différents résultats, comparaison avec une prédiction théorique – est négatif, alors il doit être possible, d'une façon ou d'une autre, d'*améliorer* les mesures effectuées ou encore les théories concernées par le test. Autrement dit, le concept de valeur vraie intègre l'idée qu'il est possible de *corriger* les échecs théoriques et expérimentaux ;

92. Gabrielse et Hanneke (2006), Mohr, Taylor et Newell (2012)

cela vient, par exemple, en identifiant une « erreur systématique », identification à laquelle est associée une forte connaissance *positive* – s’il y a une erreur, il doit être possible d’en trouver la cause. La valeur vraie d’une grandeur apparaît ici nécessaire pour représenter l’idée selon laquelle la mesure est une activité guidée par une *cible*. Pour pouvoir corriger une erreur, il faut qu’il y ait au départ la *possibilité* d’une erreur (rappelons Kyburg : « voici ce qui est presque la pierre de touche de l’objectivité : la possibilité d’une erreur »⁹³). Si l’on ne reconnaît pas l’existence de l’erreur de mesure, alors on ne reconnaît pas la possibilité de faire des corrections théoriques ou des améliorations expérimentales, et donc on nie la possibilité d’un *progrès* expérimental. Mais il ne s’agit pas simplement de dire que la valeur vraie permet de concevoir une correction ; on pourrait envisager une vision purement cohérentiste, sans valeur vraie, dans laquelle l’idée de correction serait également présente. Plutôt, le concept donne également à l’expérimentateur une raison de *poursuivre* l’enquête et de ne pas se contenter d’un simple constat de cohérence : le processus de correction est alors permanent et perpétuel. Finalement, l’engagement réaliste est celui qui consiste à dire qu’on *peut* et que l’on doit traquer les erreurs – même si cette traque n’est jamais le produit d’une comparaison des résultats de mesure avec la valeur vraie elle-même. Notons qu’en aucun cas cette position de prétend justifier *a priori* l’adhésion au réalisme scientifique. Elle défend plutôt que cette forme de réalisme présente des vertus épistémiques. En ce sens, la valeur vraie d’une grandeur accompagne l’affirmation du rôle épistémologique positif du « problème de l’erreur » que Hon n’a eu de cesse de rappeler.

Si l’on adhère à l’idée que l’une des formes du progrès scientifique réside dans l’identification, la correction et l’élimination des erreurs de mesure, et que par conséquent, notre connaissance partielle des grandeurs tend à s’améliorer progressivement de façon à ce que nous pouvons raisonnablement affirmer connaître de mieux en mieux la valeur vraie d’une grandeur, il nous faut alors disposer d’une conception satisfaisante de la notion d’approximation. Cela permet d’intégrer le caractère approximatif et incertain des résultats de mesure, en particulier dans une perspective de réalisme convergent. Cette problématique apparaît de façon encore plus nette lorsqu’on est amené à considérer un second type d’arguments objectés à la valeur vraie : celui de la non-unicité. En effet, quand bien même on accepterait la validité de principe du concept de valeur vraie, il apparaît qu’il est impossible d’affirmer, pour certaines grandeurs, que celles-ci sont caractérisées par une *unique* valeur vraie. Comme nous le défendons dans le chapitre suivant, l’une des possibilités consiste à introduire le concept d’une valeur *approximativement* vraie. Un réalisme trop radical, qui ne rendrait pas compte de l’idée de proximité avec la vérité mais qui verrait toute proposition comme soit vraie, soit fausse, expose à des difficultés, car il revient à classer une grande partie des propositions approximatives comme fausses. L’approximation introduit une plasticité qui permet de mieux rendre compte de l’état de vérité des théories scientifiques.

93. Kyburg (1992b), p.147.

Chapitre 7

Valeur vraie : non unicité et incertitude définitionnelle

Avant d'aborder l'argument d'inconnaissabilité, nous avons vu en introduction de cette partie que la critique de la notion de valeur vraie qui a été récemment émise au sein de la communauté métrologique, et en particulier dans le GUM, mélange ce que nous avons identifié comme étant deux arguments distincts. L'un a trait au caractère « connaissable » de la valeur vraie : si celle-ci présente la singularité d'être à la fois la valeur que l'on recherche et une valeur à jamais inconnue, c'est qu'il est peut-être préférable de concevoir la mesure autrement qu'au travers de ce concept. Cet argument est l'objet du chapitre précédent. Le second argument découle d'un constat assez immédiat : même si l'on accepte l'idée générale de valeur vraie d'une grandeur, il se révèle toutefois impossible de caractériser certaines grandeurs physiques par une valeur vraie *unique*. Dès lors, l'idée même d'une multiplicité de valeurs « vraies » pour une même grandeur amène à interroger la légitimité du concept – ou au moins son appellation.

La littérature métrologique du XX^e siècle contient à plusieurs reprises des remarques relatives à la non-unicité de la valeur vraie. Les différentes éditions du VIM, ainsi que le GUM, intègrent cette réflexion, mais elle n'y est présente que de façon parcimonieuse. De manière générale, cette problématique semble n'avoir jamais été considérée comme centrale et n'a jamais été développée de façon extensive ; lorsqu'elle est évoquée, c'est d'abord pour se prémunir des interprétations considérées comme trop naïves de la notion de valeur vraie d'une grandeur. Les publications du GUM et des dernières éditions du VIM marquent néanmoins une étape remarquable en ce qu'elles introduisent une notion nouvelle, bien qu'encore instable : celle d'« incertitude intrinsèque », rebaptisée par la suite « incertitude définitionnelle », qui apparaît de prime abord comme une limite ultime à l'exactitude possible de la mesure d'une grandeur donnée.

Dans ce chapitre, nous souhaitons explorer la structure de cette seconde critique de la valeur vraie d'une grandeur, en la considérant comme indépendante de la précédente critique sur l'inconnaissabilité ; et nous souhaitons nous intéresser à la nature de cet objet nouveau qu'est l'incertitude définitionnelle. Notre questionnement principal est le suivant : le constat de la multiplicité des valeurs vraies pour nombre de grandeurs physiques impose-t-il d'abandonner l'idée d'une « valeur vraie » d'une grandeur physique ? Comment peut-on maintenir

une conception classique des grandeurs physiques tout en prenant acte de ce constat de non-unicité? Cette question se pose tout particulièrement si l'on souhaite adhérer à une forme, même modérée, du réalisme scientifique.

Dans un premier temps, nous exposerons les fondements de ce que nous appellerons par la suite le constat de « non-unicité » de la valeur vraie d'une grandeur; c'est l'objet de la section 7.1. Nous prendrons comme point de départ la littérature métrologique, mais nous affirmerons que celle-ci n'a pas accordé une grande importance à cette thématique. Nous en extrairons malgré tout deux idées essentielles. La première est l'émergence du concept d'« incertitude définitionnelle » à partir de la publication du GUM. La seconde est un questionnement général sur la notion de valeur vraie. Nous chercherons ainsi à prendre au sérieux le constat de non-unicité de la valeur vraie et à reprendre à notre compte les arguments des métrologues pour rendre plus explicites les problèmes que ce constat mène à soulever.

Dans un second temps, à la section 7.2 nous nous efforcerons de déterminer la nature de la notion d'incertitude définitionnelle, introduite dans le GUM. Nous verrons en particulier que celle-ci peut être comprise comme une incertitude dans le sens où elle est un concept épistémique, c'est-à-dire relatif, d'une façon ou d'une autre, à un corpus de connaissances. Nous verrons toutefois que, contrairement à l'acceptation ordinaire de l'incertitude de mesure, l'incertitude définitionnelle ne caractérise pas une *limite* de connaissances.

Dans un troisième temps, nous développerons un raisonnement qui nous amène à défendre la conception classique de la valeur vraie d'une grandeur. Cela sera développé à la section 7.3. Nous soutenons que nous pouvons réconcilier l'idée de valeur vraie avec celle d'une multiplicité de valeurs d'une grandeur, sans nécessairement renoncer à une conception réaliste de la vérité. Nous montrerons qu'il faut pour cela faire appel à une épistémologie de l'approximation qui fait émerger une notion satisfaisante de « vérité approchée ». Nous comprendrons alors que la notion d'incertitude définitionnelle présente avant tout la vertu de rappeler le caractère intentionnel de la notion de mesurande, et nous montrerons qu'elle guide l'expérimentateur en lui indiquant le niveau de précision qu'il doit apporter à un modèle pour atteindre les buts qui sont les siens.

7.1 Non-unicité de la valeur vraie d'une grandeur

7.1.1 Mesurande et grandeur réalisée

Afin de bien comprendre ce qui est entendu par « non-unicité » de la valeur vraie, il est nécessaire de revenir au préalable sur la notion de mesurande, qui joue ici un rôle essentiel. Le terme de « mesurande » a été introduit assez récemment dans la littérature scientifique. Il est très rare de le retrouver dans les publications qui précèdent la première édition du VIM, en 1984; depuis, son usage s'est progressivement généralisé. Le tableau 7.1 présente l'évolution de l'acceptation de « mesurande » au fil des éditions du VIM. Initialement, le concept visait surtout à désigner de façon simple la grandeur mesurée¹ : c'est ce qui transparait des énoncés des

1. En 1984, à l'occasion de la publication de la première édition du VIM, Clifford a apporté quelques précisions sur les motivations qui ont poussé les rédacteurs du VIM à introduire le terme « mesurande ». Il indique ainsi

| | | |
|-------------|--|--------|
| VIM1 (1984) | Mesurande : grandeur soumise à mesurage | (p.12) |
| VIM2 (1993) | Mesurande : grandeur particulière soumise à mesurage | (p.20) |
| VIM3 (2008) | Mesurande : grandeur que l'on veut mesurer | (p.17) |

TABLEAU 7.1 – Différentes définitions de « mesurande » dans les trois éditions du *Vocabulaire International de Métrologie* (VIM). Les définitions du VIM1 et du VIM2 sont identiques dans leur esprit, bien qu'une précision soit apportée à la définition du VIM2, afin de signaler que le mesurande est une grandeur particulière, associée à un objet (la longueur d'un objet spécifique), et non une grandeur générale (la notion générale de longueur). La définition du VIM3 donne une place à l'expérimentateur (« l'on ») et caractérise le mesurande de façon *intentionnelle*, ce qui marque une différence majeure avec les définitions précédentes.

VIM 1 et 2² : un mesurande est une « grandeur [particulière] soumise à mesurage »³. Le GUM s'appuie sur le vocabulaire défini dans la seconde édition du VIM (les deux documents ayant été publiés la même année, en 1993), et emploie donc cette même acception de « mesurande ». Pourtant, à la lecture de l'annexe D, nous remarquons que quelques libertés sont prises quant à la signification du terme. Ainsi, il est indiqué que « (l)orsqu'on réalise un mesurage, la première étape consiste à spécifier le mesurande, c'est-à-dire la grandeur à mesurer »⁴. Le fait que le mesurande soit la grandeur « à mesurer » et non la grandeur « mesurée » ou « soumise à la mesure » préfigure la définition qui sera ensuite adoptée quinze ans plus tard dans le VIM3, où est introduite l'idée d'*intention* : le mesurande est alors la « grandeur que l'on veut mesurer »⁵. Cette nuance est motivée par le constat selon lequel, dans de nombreux cas, la grandeur soumise à mesure n'est pas la grandeur que l'expérimentateur souhaite mesurer, comme en atteste la nécessité d'effectuer des *corrections*. C'est pourquoi il est jugé nécessaire, dans le GUM, de marquer la différence entre le mesurande, compris comme grandeur *à mesurer*, et « la grandeur réalisée »⁶. L'on pourrait penser que ce n'est qu'une question de choix de terminologie : le mesurande doit-il désigner ce que l'on mesure effectivement ou ce que l'on mesure une fois les corrections effectuées ? Dans ce cas, l'introduction du terme présente l'intérêt de rappeler

« "Quantité mesurée" s'applique à une mesure déjà effectuée : "quantité à mesurer" s'applique à une mesure future. Il est commode d'avoir un seul terme qui sera toujours approprié. Le terme "mesurande" a parfois été utilisé dans le passé et a été adopté pour ce Vocabulaire », Clifford (1985), p.73. Il s'agit, à notre connaissance, d'un terme presque entièrement nouveau en métrologie, à quelques exceptions près.

2. La différence entre les définitions du VIM1 et du VIM2 tient au fait que le mesurande n'est plus une grandeur mais une grandeur « particulière ». Cette précision semble essentiellement vouée à marquer la différence entre un concept général de grandeur, par exemple la longueur, et une grandeur particulière, comme la longueur d'un objet spécifique (voir chapitre 6.1).

3. Organisation internationale de normalisation (ISO) (1984), p.12; Organisation internationale de normalisation (ISO) (1993), p.20.

4. Joint Committee for Guides in Metrology (JCGM) (2008d), p.51.

5. Joint Committee for Guides in Metrology (JCGM) (2012), p.17.

6. « De manière idéale, la grandeur réalisée pour le mesurage devrait être totalement conforme à la définition du mesurande. Il arrive cependant souvent qu'on ne puisse réaliser une telle grandeur et le mesurage est effectué sur une grandeur qui est une approximation du mesurande. », Joint Committee for Guides in Metrology (JCGM) (2008d), section D.2, p.51.

à l'expérimentateur qu'il n'est pas toujours possible de mesurer directement une grandeur⁷. Néanmoins, la nuance que nous venons d'exhiber engage plus que cela. En effet, elle induit une différence de nature épistémologique entre « mesurande » et « grandeur mesurée » : en paraphrasant Mari, un mesurande n'est pas *donné*, mais est *construit* – il n'est pas simplement *ce que l'on mesure* mais *doit être défini*⁸. C'est en ce sens, nous semble-t-il, qu'il faut comprendre la direction donnée à la définition de la notion de « valeur vraie » qui est donnée dans le VIM3, bien que celle-ci soit par ailleurs assez obscure : « valeur d'une grandeur compatible avec la définition de la grandeur »⁹.

Suivant l'évolution de la notion de mesurande, ce dernier est déterminé par l'expérimentateur et non par le protocole de mesure que ce dernier emploie, et devient ainsi *idéal*¹⁰. En particulier, la définition d'un mesurande donnée par un expérimentateur peut être irréalisable, par exemple si cette définition inclut des conditions d'expérience idéales et atteignables uniquement à la limite – l'on peut penser ainsi à la mesure d'une grandeur dont la réalisation nécessiterait d'obtenir un vide absolu¹¹. L'acte concret de mesure d'une grandeur est une *réalisation* particulière à partir de laquelle l'expérimentateur vise à remonter au mesurande qu'il a défini préalablement, en effectuant les corrections nécessaires. Comme indiqué dans le GUM,

Le résultat du mesurage de la grandeur réalisée est corrigé de la différence entre cette grandeur et le mesurande, pour ramener le résultat de mesure à ce qu'il aurait été si la grandeur réalisée avait en fait satisfait complètement à la définition du

7. Ce constat peut sembler évident dans certaines disciplines, comme pour la mesure des constantes de physique fondamentale où l'activité de correction est omniprésente, et où la question de la cible est depuis longtemps une question essentielle de la pratique scientifique. Mais ce n'est pas automatiquement le cas pour tous les domaines, et De Bièvre a ainsi insisté sur les conséquences positives que génère une telle prise de conscience en chimie : voir De Bièvre (2007) et De Bièvre (2008).

8. Luca Mari, communication privée; voir aussi Mari (2015), p.74. Notons que le GUM et le VIM n'explicitent que très mal ce qu'est « définir » un mesurande, ce qui n'est pas sans poser certains problèmes, comme l'ont souligné Baratto (2008) puis Mari (2008).

9. Joint Committee for Guides in Metrology (JCGM) (2012), p.20. Cette définition est assez similaire à celle proposée en 1993 dans le VIM2, mais marque en revanche une différence assez nette avec celle du VIM1. Voir également les remarques que nous formulons p.6.1.

10. Il faut toutefois distinguer deux types de mesurande. Dans le premier cas, le mesurande est défini théoriquement – il comporte alors une part d'idéalisation. Dans le second cas, le mesurande peut être simplement *désigné* : il est « cette » grandeur réalisée par « cette » opération de mesure particulière. Le guide publié par l'EURACHEM à l'attention de la communauté des chimistes indique ainsi : « Dans les mesures analytiques, il est particulièrement important de distinguer entre les mesures destinées à produire des résultats qui sont indépendants de la méthode utilisée, et celles qui ne sont pas destinées à cela. Ces dernières sont souvent désignées comme des méthodes empiriques ou des méthodes définies de façon opérationnelle », Ellison et Williams (2012), p.121. Dans ce cas, il n'y a plus de différence entre mesurande et grandeur réalisée. Comme le font remarquer Giordani et Mari, la contrepartie de cette définition opérationnelle est qu'il n'est, en toute rigueur, plus possible de transférer le résultat de mesure en dehors de son contexte particulier de production : « Cette position a l'avantage opérationnel immédiat d'éviter la nécessité de mesurer une quelconque grandeur d'influence, mais elle empêche de projeter l'information acquise en dehors du contexte strict (et inconnu) dans lequel elle a été obtenue », Giordani et Mari (2012), p.2148. L'un des apports de la nouvelle conception du mesurande est le pouvoir d'abstraction que celui-ci permet. Dans la suite de ce chapitre, nous nous intéresserons, sauf exception, aux mesurandes définis théoriquement.

11. Prenons par exemple la définition de l'ampère : « (l)'ampère est l'intensité d'un courant constant qui, maintenu dans deux conducteurs parallèles, rectilignes, de longueur infinie, de section circulaire négligeable et placés à une distance de 1 mètre l'un de l'autre dans le vide, produirait entre ces conducteurs une force égale à 2×10^{-7} newton par mètre de longueur. », Bureau International des Poids et Mesures (2006), p.23.

mesurande.¹²

Les grandeurs réalisées ne sont pas nécessairement des réalisations *directes* du mesurande ; la nécessité d'effectuer des corrections en atteste, justement. La grandeur réalisée corrigée peut toutefois être encore appelée « réalisation » du mesurande. La distinction entre mesurande et grandeur réalisée est essentielle pour comprendre pourquoi l'on parle de non-unicité de la valeur vraie d'une grandeur, en raison de la multiplicité des réalisations possibles du mesurande.

7.1.2 Non-unicité de la valeur vraie

L'importance de la notion de mesurande tient en ce qu'elle fait intervenir une définition. Cela suscite le constat suivant : la définition d'un mesurande ne renvoie pas inmanquablement à *une* grandeur physique unique ; il peut lui correspondre un ensemble de grandeurs physiques proches mais distinctes. C'est le cas si la définition du mesurande est insuffisamment restrictive. La *cible* de la mesure est alors floue¹³, et l'on comprend que cette cible puisse être atteinte en réalisant des grandeurs différentes. En reprenant à notre compte une expression de Tal, le mesurande est « réalisable de façon multiple »¹⁴. Nous pouvons également présenter les choses du point de vue des corrections effectuées, où le raisonnement employé est symétrique. Puisque le mesurande est insuffisamment défini, on ne peut pas déterminer de façon univoque la correction à appliquer pour « ramener le résultat de mesure à ce qu'il aurait été si la grandeur réalisée avait en fait satisfait complètement à la définition du mesurande » (en reprenant les termes du GUM cité précédemment).

Dès lors qu'à un mesurande donné correspondent différentes réalisations possibles, la valeur de chacune de ces grandeurs peut être dite « valeur vraie » du mesurande ; et, par conséquent, le mesurande n'est plus caractérisé par une valeur vraie unique, mais par un *ensemble* de valeurs vraies également légitimes. L'annexe du GUM illustre ce constat général par un exemple qui gagne à être reproduit extensivement (et que nous citons en deux temps) :

D.3.2 À titre d'exemple, supposons que le mesurande soit l'épaisseur d'une feuille donnée d'un matériau à une température spécifiée. L'éprouvette est portée à une température proche de la température spécifiée et son épaisseur est mesurée à un endroit particulier avec un micromètre. L'épaisseur du matériau, à cet endroit et à cette température, sous la pression exercée par le micromètre, est la grandeur réalisée.

D.3.3 On détermine, au moment de la mesure, la température du matériau et la pression appliquée. On corrige alors le résultat non corrigé du mesurage de la grandeur réalisée en prenant en compte la courbe d'étalonnage du micromètre,

12. Joint Committee for Guides in Metrology (JCGM) (2008d), p.51.

13. Comme l'expriment Vosk et Emery, « (c)ette incomplétude laisse l'identité de ce qui est mesuré quelque peu vague », Vosk et Emery (2014), p.200.

14. Tal (2011), p.1087. Tal emploie cette expression à propos de la réalisation de la définition des unités de mesure. Dans la situation qu'il décrit, ces unités peuvent être réalisées selon différents principes expérimentaux, en exploitant des phénomènes physiques différents, et au moyen d'étalons matériels différents – mais, dans tous les cas, ce sont des réalisations (imparfaites) de la *même* grandeur. Le cas présent est un peu différent : il s'agit de remarquer que *différentes* grandeurs sont compatibles avec une même définition ; dès lors, il existe différentes façons de réaliser la définition.

l'écart de température de l'éprouvette par rapport à la température spécifiée et la légère compression de l'éprouvette sous la pression appliquée.

D.3.4 Le résultat corrigé peut être appelé meilleure estimation de la valeur « vraie », « vraie » dans le sens où c'est la valeur d'une grandeur censée satisfaire pleinement à la définition du mesurande; mais si le micromètre avait été appliqué à un endroit différent de la feuille de matériau, la grandeur réalisée aurait été différente, avec une valeur « vraie » différente. Cependant, cette valeur « vraie » aurait été compatible avec la définition du mesurande parce que cette dernière ne spécifie pas que l'épaisseur doit être déterminée à un endroit particulier de la feuille.¹⁵

Dans cette citation, un premier principe est énoncé : la *cause* de la multiplicité des valeurs vraies compatibles avec la définition d'un mesurande donné se trouve dans l'*incomplétude* de la définition en question. La multiplicité des valeurs vraies est associée à des lacunes définitionnelles qui ajoutent chacune un degré d'arbitraire possible dans la réalisation du mesurande. Nous renvoyons au tableau 7.2 pour une présentation de quelques exemples de mesurandes et des sources d'incomplétude de leurs définitions.

Notons que cette caractérisation en termes d'incomplétude suppose que la définition d'un mesurande soit complétable, ce qui ne semble être toujours le cas, en particulier quand la limitation est quantique, par exemple dans le cas de la fréquence des raies spectrales (voir tableau 7.2), qui ont toujours une largeur en raison de la durée de vie finie des états d'énergie excités d'un atome. Dans ce cas, la limitation semble être intrinsèque aux objets mesurés, et non une question de définition. Cependant, il faut en fait séparer ce qui est de l'ordre du phénomène physique et de la définition d'un mesurande. Dire que la définition de la fréquence d'une raie spectrale est incomplète ne revient *pas* à affirmer que si la définition venait à être complétée, la raie spectrale aurait alors une fréquence unique – cela est contraire à la théorie quantique. Cela signifie par contre qu'il n'est pas possible de décrire la raie spectrale par une seule grandeur, et qu'il faudrait plutôt la caractériser simultanément par (au moins ?) deux grandeurs, à savoir sa fréquence centrale et sa largeur, qui, elles, sont susceptibles de présenter une valeur vraie unique (à moins, bien sûr, que d'autres sources d'incomplétude ne soient identifiées par ailleurs).

Nous verrons qu'à la suite de cette remarque, le GUM tire la conséquence du constat de non-unicité de la valeur vraie en introduisant un nouveau concept, celui d'incertitude définitionnelle. Avant d'en venir à cette idée, nous souhaitons d'abord développer les conséquences du constat de non-unicité de la valeur vraie d'une grandeur sur la validité conceptuelle de la notion de « valeur vraie » elle-même.

7.1.3 Critiques de la valeur vraie

Dans le chapitre précédent, nous avons montré comment les métrologues ont récemment porté leur attention sur le caractère « inconnaissable » d'une notion de valeur « vraie » d'une grandeur. Nombre des remarques adressées envers la notion de valeur vraie sont effectivement du ressort de la problématique d'inconnaissabilité. Le constat de non-unicité de la valeur vraie

15. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#) pp.51–52.

| Mesurande générique | Facteurs influençant la complétude de la définition | Exemples de définition plus complète |
|---|---|---|
| Longueur d'une feuille de matériau [†] | Forme de la feuille Dépendance en température | Longueur moyenne Longueur maximale ou minimale Longueur selon un axe donné Longueur de l'objet à une température donnée. |
| Taille d'un individu ^{††} | Phénomène de « tassement diurne » (« Chez un adulte, cette taille varie d'environ un centimètre entre le lever et le coucher ») ^{††} | Taille au lever Taille au coucher Taille moyenne sur 24h |
| Population d'un pays ^{††} | Flux de population Décès et naissances (l'intervalle moyen de variation est d'environ 25 secondes) | Population à un instant donné (défini à quelques secondes près) |
| Température du point triple de l'eau ^{†††} | Composition isotopique de l'eau | Température du point triple pour : • la composition naturelle de l'eau • la composition isotopique de l'eau des océans ("Vienna Standard Mean Ocean Water") ^{††††} |
| Fréquence d'une raie spectrale ^{††} | Largeur de la raie spectrale, liée à la durée de vie des états excités | Fréquence moyenne de la raie spectrale Intervalle de fréquences couvertes par la raie spectrale |

TABLEAU 7.2 – Exemples de mesurandes incomplètement définis, ainsi que des facteurs possibles à l'origine de l'incomplétude de ces définitions, et des exemples de mesurandes définis de façon plus précise. † Exemple tiré du GUM; †† Exemple tiré de [Robert-Schwartz et Treiner \(2006\)](#), p.2.; ††† Exemple documenté par [Pavese \(2007\)](#), p.330.; †††† Voir la brochure éditée par le BIPM au sujet du Système International d'unités, [Bureau International des Poids et Mesures \(2006\)](#), p.82.

d'une grandeur est à la racine d'une seconde critique¹⁶ qui n'est pas tout à fait du même ordre que l'argument d'inconnaissabilité. Dans le GUM, les questions liées à l'unicité de la valeur vraie d'une grandeur, dont la citation que nous avons reproduite fait partie, sont introduites dans l'annexe D, intitulée « valeur 'vraie', erreur et d'incertitude », laquelle s'ouvre sur une remarque annonçant que seront expliquées les raisons pour lesquelles le terme « valeur vraie » n'est pas utilisé dans le document¹⁷. L'annexe présente ainsi une somme de remarques pertinentes dont certaines touchent à la question de l'unicité de la valeur vraie et à d'autres celle de sa connaissabilité. Nous avons déjà émis quelques réserves au chapitre précédent (section 6.3.2) quant à la portée des critiques émises dans le GUM, dont l'examen attentif révèle le manque de clarté. Nous avons montré en particulier qu'il semblerait que le mouvement effectué consiste simplement à masquer le terme « vrai » tout en continuant à utiliser le concept de la même façon, ce qui n'apporte aucun bénéfice tout en introduisant une ambiguïté (que les rédacteurs

16. Des remarques du même type peuvent être trouvées chez [Teller \(2013\)](#) qui n'emploie cependant pas le même vocabulaire et aborde la question sous un angle un peu différent. Les réflexions de Teller l'amènent à proposer d'abandonner ce qu'il appelle le « réalisme de l'exactitude de mesure » [*measurement accuracy realism*].

17. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.51.

du VIM3 n'ont d'ailleurs pas souhaité perpétuer). De ce fait, il serait inutile de revenir en détail sur ce point spécifique ; et l'argumentaire du GUM lui-même, présenté dans l'annexe D, ne mérite peut-être pas d'être analysé dans le détail – nous verrons à la section suivante que le véritable apport du GUM, bien que présent sous forme d'ébauche uniquement, est avant tout d'introduire une forme nouvelle d'incertitude, l'« incertitude définitionnelle ».

En étudiant la littérature scientifique, nous apercevons par ailleurs que le constat de non-unicité de la valeur vraie d'une grandeur, effectué dans le GUM, n'a en fait rien de particulièrement nouveau en soi. Eisenhart critiquait déjà la notion de valeur vraie dans son article de 1962, au terme d'une discussion assez similaire à celle présentée dans le GUM¹⁸ ; lui-même se référait également à des réflexions antérieures trouvées dans la littérature métrologique. Il apparaît ainsi que la reconnaissance du constat selon lequel certaines grandeurs sont des « concepts flous »¹⁹ remonte au moins au-delà des années 1940. Chez Eisenhart comme chez d'autres, les discussions alimentent surtout une réflexion d'ordre philosophique, ou tout du moins conceptuelle ; elles ne conduisent à aucun développement technique. Eisenhart juge que la non-unicité de la valeur vraie plaide pour un changement de terminologie, et juge qu'il serait préférable de substituer à « valeur vraie » le terme « valeur cible »²⁰, tout en reconnaissant que le poids des habitudes rend difficile un tel changement. Au-delà de cette question de terminologie, la discussion qu'il engage n'a pas pour autant de conséquence supplémentaire. Nous pourrions d'ailleurs nous demander quel est le but précis que poursuivent les remarques qui y sont émises ; il nous semble, dans la continuité de ce que nous avons suggéré dans le chapitre précédent, que ces remarques découlent en premier lieu d'un souci d'éviter les considérations trop métaphysiques que le terme de valeur *vraie* semble charrier avec lui. Notre constat bibliographique donne une autre perspective à la position développée dans le VIM3 et dans le GUM, décrite dans le chapitre précédent, selon laquelle la critique de la valeur vraie caractérise la transition d'une approche dite « erreur » à une approche « incertitude » où le caractère inconnaissable de la valeur vraie d'une grandeur est explicitement reconnu. En effet, contrairement à ce que nous proposons, l'argumentaire développé dans ces documents est un *mélange* de questions de *connaissabilité* et de questions de *non-unicité* ; or, de toute évidence, l'argument de non-unicité était déjà reconnu dans les approches traditionnelles de la mesure²¹, comme le montre l'exemple d'Eisenhart et de ses prédécesseurs.

Au-delà des limites formelles de l'argumentation donnée dans la littérature métrologique, et dans le GUM en particulier, le constat de non-unicité de la valeur vraie doit malgré tout être

18. Eisenhart (1963), pp.170–171.

19. Eisenhart (1963), p.171.

20. Terme que nous avons déjà mentionné au chapitre précédent, et que reprend en particulier Willink.

21. La note 1 de la définition de « valeur vraie » dans le VIM3 témoigne du mélange opéré entre les deux types d'arguments : « Dans l'approche 'erreur' de description des mesurages, la valeur vraie est considérée comme unique et, en pratique, impossible à connaître. L'approche 'incertitude' consiste à reconnaître que, par suite de la quantité intrinsèquement incomplète de détails dans la définition d'une grandeur, il n'y a pas une seule valeur vraie mais plutôt un ensemble de valeurs vraies compatibles avec la définition. Toutefois, cet ensemble de valeurs est, en principe et en pratique, impossible à connaître. », p.20. Notre analyse montre qu'il est inexact d'attribuer à l'« approche erreur » l'idée d'une valeur vraie unique. Ehrlich, Dybkaer et Wöger reconnaissent ce fait en indiquant : « il n'est pas évident de savoir à quelle approche de la mesure associer la prémisse d'un manque d'unicité de la valeur vraie d'un mesurande », Ehrlich, Dybkaer et Wöger (2007), pp.207–208.

pris au sérieux afin de discuter de la validité de la notion de valeur vraie elle-même. De fait, on conçoit tout à fait pourquoi un tel constat – la non-unicité – peut être interprété comme une limite de la notion de valeur vraie : si la notion de « valeur vraie », interprétée littéralement (comme nous l'avons proposé au chapitre précédent), est épistémologiquement valide, alors à toute grandeur physique ne peut correspondre qu'une seule et unique valeur vraie²². La solution adoptée dans le VIM, depuis la seconde édition, consiste pourtant à élargir la notion de valeur vraie à un *ensemble* de valeurs vraies. Ainsi la note n° 3 de la définition de « valeur vraie » dans le VIM2 indique

L'article indéfini 'une' plutôt que l'article défini 'la' est utilisé en conjonction avec 'valeur vraie' parce qu'il peut y avoir plusieurs valeurs correspondant à la définition d'une grandeur particulière donnée.²³

Cependant, la dénomination est alors clairement contre-intuitive : comment différentes valeurs peuvent-elles être simultanément « vraies » pour une même grandeur ?

Dans un contexte un peu différent, le physicien Jacques Treiner a récemment émis une critique similaire. La démarche de Treiner est avant tout à visée didactique ; elle s'inscrit dans un contexte de rédaction et d'explicitation des programmes de lycée dans les académies françaises²⁴. Selon Treiner, en introduisant la notion de valeur vraie, les scientifiques et les enseignants induisent dans l'esprit des élèves une incompréhension qui grandira au fur et à mesure que le vécu de ces élèves viendra mettre à l'épreuve la légitimité d'une telle notion. Partant d'un constat similaire à celui que l'on trouve chez Eisenhart ou dans le GUM, Treiner montre que différentes grandeurs classiques en physique expérimentale ou en sciences sociales (taille d'un individu, longueur d'une table, fréquence et largeur en fréquence d'une raie spectrale, population d'un pays, etc.) présentent une variabilité intrinsèque, ce qui démontre également que la notion de valeur vraie est inappropriée :

Allons tout de suite à l'essentiel : cette notion de « valeur vraie ». Elle est inconnue, nous dit-on, puisqu'on la cherche, mais en tout cas *elle existe*. Or, à l'évidence – une évidence que l'élève peut percevoir – cette valeur vraie *n'existe pas!*²⁵

puis,

Que conclure de ces exemples ? Qu'il n'y a pas de valeur vraie, que toute grandeur physique est distribuée suivant une loi de probabilité qui dépend du système. La notion première, c'est donc celle de *variabilité* des grandeurs physiques.²⁶

Pour Treiner, l'idée qu'une grandeur serait caractérisée par une valeur vraie est dissoute dès lors que l'on réalise que du fait de la variabilité intrinsèque des grandeurs physiques, l'opéra-

22. C'est ainsi que Ehrlich, Dybkaer et Wöger présentent également la chose : « dans l'approche classique, il ne serait plus possible de parler de "la valeur vraie" d'un mesurande, ou de l'"erreur systématique" associée à un résultat de mesure, puisque de telles valeurs uniques n'auraient plus de sens », p.208.

23. [Organisation internationale de normalisation \(ISO\) \(1993\)](#), p.16.

24. [Ministère de la Jeunesse, de l'Éducation nationale et de la Recherche \(2002\)](#). Cet extrait du « document d'accompagnement », un guide à destination des enseignants de classe terminale, a été rédigé par Jacques Treiner et Daniel Beaufils, comme Treiner l'indique lui-même dans [Treiner \(2011\)](#), p.10.

25. [Treiner \(2011\)](#), p.10. Treiner souligne.

26. [Treiner \(2011\)](#), p.12.

tion de mesure revient toujours à un tirage aléatoire²⁷. Le message central de Treiner est qu'il ne faut assimiler variabilité et erreur (ou incertitude) *qu'avec prudence*, et qu'il serait même peut-être préférable de ne pas les assimiler du tout, puisque savoir qu'un phénomène physique présente une variabilité intrinsèque « représente une connaissance positive »²⁸, non un état de doute. Par cette remarque, Treiner touche là un élément essentiel pour comprendre la notion d'incertitude définitionnelle introduite par le GUM; nous y reviendrons à la prochaine section.

À ce point de l'exposition, il nous faut mentionner un point essentiel, que ni Eisenhart, ni le GUM, ni Treiner n'explorent dans leurs différents arguments. Il s'agit de la conséquence du constat de la multiplicité des valeurs vraies d'une grandeur physique sur l'interprétation des équations de la physique. En effet, l'idée de valeurs vraies multiples d'une grandeur entre en dissonance avec l'usage traditionnel des équations de la physique, qui expriment à la fois une relation physique entre des quantités et une relation numérique entre les (vraies) valeurs des grandeurs impliquées²⁹. Considérons par exemple l'équation de la trajectoire d'un objet en chute libre dans le vide, dans un référentiel galiléen, soumis à une pesanteur uniforme :

$$z = \frac{1}{2}gt^2 \quad (7.1)$$

(où z est l'altitude de l'objet, t la durée de la chute, et g l'accélération de pesanteur) Selon l'interprétation traditionnelle, cette équation est vérifiée par les « valeurs vraies » des grandeurs impliquées. C'est parce que l'on considère que les valeurs vraies des grandeurs z , g et t vérifient l'équation (7.1) qu'une étude expérimentale de la chute libre d'un corps visera à vérifier à quel point les valeurs *attribuées* à ces grandeurs, c'est-à-dire les valeurs mesurées, que l'on espère proches de leurs valeurs vraies respectives, vérifient elles aussi l'équation. Mais si ces grandeurs ne peuvent pas être caractérisées par une valeur vraie unique, il est impossible de continuer à affirmer que leurs multiples valeurs « vraies » vérifient l'équation (7.1).

En résumé, le constat de non-unicité de la valeur vraie d'une grandeur mène à soulever deux questions que nous ne pouvons pas ignorer. (i) Quelle est la signification du terme « vraie » pour la « valeur vraie » d'une grandeur dès lors que celle-ci n'est pas unique ? (ii) Comment peut-on interpréter les équations de la physique si aux grandeurs mises en jeu ne correspondent pas une valeur vraie unique ?

Nous proposons une réponse articulée en deux temps. Dans un premier temps, nous opérons un détour par la notion d'« incertitude définitionnelle » qui a été introduite dans le GUM. Interprétée correctement, celle-ci devient un outil intéressant pour traiter les questions que nous avons soulevées. Nous verrons cependant que la réponse adoptée par le GUM n'est pas totalement satisfaisante, en particulier parce qu'elle occulte les questionnements épistémologiques que nous soulevons ici. Dans un second temps, nous revenons sur la question de la signification de la valeur vraie d'une grandeur, à la fois du point de vue d'une grandeur seule

27. « (M)on but ici est de souligner que l'on est *toujours* face à des grandeurs aléatoires. [...] En ce sens, *effectuer une mesure doit être considéré comme le tirage d'une variable aléatoire* dont on cherche à déterminer la loi de probabilité ! », Treiner (2011), p.12 (Treiner souligne).

28. Treiner (2011), p.9.

29. Voir en particulier de Courtenay (2015).

et du point de vue des équations de la physique. Nous défendons qu'il est possible de sauvegarder la notion, à condition de faire valoir un concept satisfaisant de « vérité approchée », exprimé dans un cadre réaliste modéré.

7.2 L'« incertitude définitionnelle »

7.2.1 L'incertitude définitionnelle dans les documents de métrologie

S'il est compris depuis bien avant la publication du GUM que la multiplicité des valeurs possibles d'une grandeur physique constitue une limite potentielle de la notion de « valeur vraie », le concept d'« incertitude définitionnelle » est apparu quant à lui pour la première fois dans les documents récents de métrologie. Le terme n'apparaît qu'une seule fois dans le GUM. On le retrouve dans la continuité de la citation que nous avons exposée à la section 7.1.2 :

(E)n raison d'une définition incomplète du mesurande, la valeur « vraie » présente une incertitude qui peut être évaluée à partir de mesurages réalisés en différents emplacements de la feuille. À un niveau donné, tout mesurande possède une telle incertitude « intrinsèque » qui peut être estimée en principe d'une façon ou d'une autre. C'est l'incertitude minimale avec laquelle on peut déterminer un mesurande et chaque mesurage qui atteint une telle incertitude peut être considéré comme le meilleur mesurage possible du mesurande. Pour obtenir une valeur de la grandeur en question avec une incertitude plus petite, il faut que le mesurande soit défini plus complètement.³⁰

Alors qu'Eisenhart, par exemple, ne tirait aucune conclusion pratique de la non-unicité de la valeur vraie, le GUM va plus loin et tente de rendre compte de cette unicité dans le calcul d'incertitude. La *conséquence* de la multiplicité de la valeur vraie, dans le GUM, est une « incertitude intrinsèque », qui est une limite basse, incompressible, de l'incertitude de mesure finale sur le mesurande considéré. L'« incertitude intrinsèque » étant ainsi introduite dans le GUM, l'on pourrait s'attendre à ce qu'elle soit définie dans l'édition du VIM2 de la même année, sur laquelle le GUM s'appuie. Ce n'est pourtant pas le cas, et il faut attendre l'édition de 2008 pour y retrouver le concept sous le nom d'« incertitude définitionnelle : composante de l'incertitude de mesure qui résulte de la quantité finie de détails dans la définition d'un mesurande »³¹. Dans la suite de ce chapitre, nous conserverons uniquement ce dernier terme, n'utilisant qu'« incertitude définitionnelle » seule, au détriment de l'« incertitude intrinsèque » du GUM; nous suivons en cela l'usage qui s'est assez rapidement imposé en métrologie.

L'introduction de l'incertitude définitionnelle présente l'intérêt particulier d'être un essai de *réponse* au problème posé par la non-unicité de la valeur vraie; les métrologues ne se contentent alors pas simplement d'énoncer un constat, comme cela était le cas dans les discussions antérieures, mais ambitionnent d'élaborer une approche qu'ils puissent intégrer à leur pratique. Que ce soit par le terme lui-même, ou par l'usage qui en est fait, on constate que le GUM et le VIM suggèrent déjà deux propositions. (i) L'incertitude définitionnelle *est* une

30. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#) p.52.

31. [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.25.

incertitude de mesure, de même nature que les composantes habituellement traitées dans la pratique scientifique. Aussi anodin qu'il peut avoir l'air au premier abord, nous verrons que ce constat n'a en fait rien d'évident. La seconde proposition est (ii) l'incertitude définitionnelle est *quantifiable* au même titre que toute autre composante d'incertitude de mesure. Ainsi l'incertitude définitionnelle est-elle considérée dans le VIM et le GUM comme une composante *quantitative* d'incertitude, qui peut, de ce fait, être calculée et contribuer au bilan d'incertitude global. Cependant, comme nous le verrons plus en détail, les modalités de ce calcul ne sont pas explicitées dans les documents généraux de métrologie, contrairement à ce à quoi l'on pourrait s'attendre une fois ces concepts introduits.

Si l'incertitude définitionnelle est une composante d'incertitude qui doit être déterminée au même type que les autres composantes, alors l'absence de traitement quantitatif de l'incertitude définitionnelle est une lacune notable du GUM. Cela, cependant, n'est peut-être pas le premier reproche que nous pourrions adresser au document. Il semble raisonnable qu'avant même de chercher à quantifier l'incertitude définitionnelle, la nature de ce concept soit proprement explicitée ; or, cette réflexion est également absente du GUM. De fait, dans le GUM, la nature de l'incertitude définitionnelle n'apparaît pas de façon très claire. En particulier, elle semble être attachée au mesurande ou à la valeur vraie du mesurande (nous soulignons, voir la citation complète que nous avons reproduite p.205) :

- « la valeur "vraie" *présente* une incertitude »
- « tout mesurande *possède* une telle incertitude "intrinsèque" »

tout en étant une caractéristique du résultat lui-même :

- « c'est l'incertitude minimale *avec laquelle on peut déterminer* un mesurande »

Or, c'est dire deux choses assez différentes, et cela révèle une certaine confusion conceptuelle de la part du GUM. Il est possible d'interpréter différemment l'incertitude de mesure selon l'approche adoptée, et en particulier la méthode statistique employée pour la calculer. Mais elle n'est à aucun moment l'incertitude *du mesurande*, ou *de la valeur vraie* de ce mesurande. L'incertitude de mesure dépeint un rapport entre le résultat de mesure et la valeur vraie de la grandeur, ou entre l'expérimentateur et la valeur de la grandeur – quoi qu'il en soit, elle est une caractéristique du résultat de mesure, et non de la grandeur elle-même. Bien entendu, un phénomène ou une grandeur physique particulière peuvent présenter une variabilité intrinsèque, ou un flou intrinsèque. Mais nous nous sommes précisément attachés à montrer que l'incertitude de mesure ne s'identifie pas à une variabilité physique : elle est le doute que cette variabilité induit lorsque celle-ci échappe à nos moyens de connaissance – c'est-à-dire qu'il n'est pas possible de l'expliquer. Ainsi, par exemple, les « incertitudes » quantiques ne sont pas des « incertitudes de mesure »³² ; les premières portent sur les objets eux-mêmes, alors que

32. Comme l'écrit Lévy-Leblond, en physique quantique, « l'idée d'"incertitudes" résulte une fois de plus de l'importation illicite d'une notion dans un domaine autre que celui où elle vaut ; il s'agit ici d'une assimilation injustifiée entre la marge d'indéfinition numérique d'une grandeur physique quantique et la marge d'ignorance inévitablement associée à une mesure expérimentale – ce qu'on appelle traditionnellement une incertitude », Lévy-Leblond (2000).

les secondes sont de nature épistémique et expriment un rapport à la connaissance³³. Au vu de notre analyse, seule la dernière des différentes formulations proposées dans le GUM nous semble valable (« l'incertitude minimale avec laquelle on peut déterminer un mesurande »)³⁴.

À ce stade, il nous faut préciser que notre objectif n'est pas tant de critiquer le GUM, lequel cherchait à introduire un concept nouveau et encore mal maîtrisé, mais de montrer qu'il reste nécessaire d'identifier la nature exacte du concept d'incertitude définitionnelle, et revenir sur les deux prémisses du GUM, qui ne sont pas acquises par avance : (i) l'incertitude définitionnelle est-elle une incertitude de mesure ? (ii) L'incertitude définitionnelle peut-elle être quantifiée comme n'importe quelle autre composant d'incertitude, et incorporée à un bilan classique d'incertitude ? Dans la suite de cette section, nous défendrons l'idée selon laquelle l'incertitude définitionnelle peut être effectivement *interprétée* comme une incertitude dans le sens où elle présente effectivement une dimension épistémique, mais qu'elle n'est pas pour autant une composante d'incertitude du même type que les composantes ordinaire d'un bilan d'incertitude.

7.2.2 L'incertitude définitionnelle est un concept épistémique

Délimitons en premier lieu la nature de l'incertitude définitionnelle : pourquoi peut-on l'appeler « incertitude » ? Pour répondre à cette question, il nous faut insister d'abord sur ce que nous entendons par « incertitude », et en particulier sur le fait que nous interprétons l'incertitude de mesure comme un concept *épistémique*. Nous avons plusieurs fois rappelé la dualité qui caractérise la façon dont on vient évaluer la fiabilité et la validité d'un résultat de mesure. Il y a d'un côté le domaine des concepts qui caractérisent l'objet mesuré en lui-même, le dispositif de mesure ou le résultat obtenu ; c'est le versant physique, ou objectif, de la mesure. D'un autre côté viennent les concepts qui caractérisent la connaissance que nous avons de ces objets ; c'est le versant épistémique de la mesure, ou, en suivant la direction de certains bayésiens, son versant subjectif.

La valeur « vraie » d'une grandeur, l'erreur de mesure ou l'exactitude de mesure concernent le versant physique. Mesurer une grandeur avec une faible erreur de mesure pourrait témoigner d'une bonne connaissance de la grandeur ; mais, encore une fois, l'un des problèmes principaux de l'évaluation d'un résultat de mesure vient du fait que l'on peut obtenir une erreur faible *sans le savoir* – ce qui n'est rien d'autre qu'un avatar du problème de l'« inconnaissabilité ». Par conséquent, il faut distinguer l'erreur de mesure que commet un expérimentateur de la connaissance que l'expérimentateur a de cette erreur ; la première est *indépendante* de la seconde. À l'inverse, l'incertitude de mesure est un concept épistémique, puisqu'elle vient ca-

33. Rappelons que, même dans le cas fréquentiste, les incertitudes de mesure expriment un rapport à la connaissance ; la différence avec le cas bayésien est qu'elles ne quantifient pas *directement*, dans l'expression probabiliste elle-même, un état de connaissance, mais qu'elles proposent plutôt de déduire l'état de connaissance de constats probabilistes portant sur les issues physiques du dispositif de mesure. Nous renvoyons à ce sujet au chapitre 5.

34. Notons également en passant, mais sans nous attarder sur ce point, que l'incertitude définitionnelle peut également *résulter* en une dispersion additionnelle du processus de mesure, selon la façon dont les mesures sont effectivement menées. Cela est brossé de façon assez rudimentaire chez Mari et Macii (2008). Dans cet article, les auteurs pointent de façon pertinente le fait qu'une incertitude définitionnelle peut amener à devoir réviser le mode de calcul de l'incertitude de type A dans certaines situations. Nous estimons cependant que cette question reste mineure dans le cadre de notre présente discussion.

ractériser l'état de connaissance de l'expérimentateur à propos de la grandeur mesurée ; elle ne se situe pas dans le domaine des objets étudiés (des événements ou des phénomènes physiques) comme l'exactitude de mesure ou l'erreur de mesure, mais est relative au sujet connaissant. Lorsque l'incertitude de mesure considérée est la conséquence d'une imperfection instrumentale, d'une variabilité des résultats de mesure ou encore d'un effet physique qu'il n'est possible de corriger parfaitement, ce tempérament épistémique apparaît de façon claire, même s'il n'est pas forcément simple à quantifier. C'est ce que nous avons décrit à la partie I, où nous avons déjà eu l'occasion d'expliquer la façon dont l'incertitude de mesure est interprétée selon les approches statistiques employées. Les métrologues s'accordent, quelle que soit leur préférence statistique, sur le fait que, lorsque le constat de départ est celui de la présence d'erreurs inhérentes au dispositif ou au modèle employés, ces erreurs entraînent par ricochet une limitation dans la connaissance que la mesure fournit à propos du mesurande – l'incertitude qualifie alors un doute à propos de la valeur du mesurande. L'interprétation épistémique de l'incertitude de mesure est une acception récente de la métrologie contemporaine. En particulier, elle accompagne assez naturellement le développement des méthodes bayésiennes ; pour autant, adhérer à cette interprétation n'engage pas automatiquement à embrasser la perspective bayésienne. Marquer une séparation entre ce qui vient caractériser les objets physiques et ce qui vient caractériser notre connaissance nous semble une avancée conceptuelle féconde, indépendamment des méthodes statistiques employées. Ceux qui souhaitent éviter de basculer dans le subjectivisme préconisé par certains défenseurs de l'approche bayésienne en métrologie sont tout à fait en droit de contester l'intérêt d'un concept tel que celui d'incertitude de mesure, mais il nous semble préférable qu'ils en reviennent alors aux concepts plus traditionnels comme exactitude, erreur ou même « erreur probable » plutôt que de donner à l'incertitude de mesure une interprétation objective.

Au vu de cette discussion, il nous semble important de voir si l'incertitude définitionnelle peut être interprétée comme un concept épistémique, si l'on veut comprendre en quoi elle peut effectivement être qualifiée d'« incertitude ». Nous défendons que c'est effectivement le cas, mais nous nuancions immédiatement ce constat en remarquant que le rapport entre incertitude définitionnelle et connaissance n'est pas le même que celui qu'entretiennent ordinairement incertitude de mesure et connaissance.

On perçoit intuitivement les raisons pour lesquelles il est possible de voir dans l'incertitude définitionnelle l'expression d'un doute. Néanmoins, il faut noter que ce doute ne porte pas sur l'exactitude de la valeur estimée. Plutôt, puisque le mesurande n'est pas parfaitement défini, il n'est pas possible de savoir précisément à *quoi se rapporte* la valeur tirée d'une mesure donnée : il y a un doute sur l'*identité* même de ce qui est à mesurer³⁵. De ce fait, contrairement à la forme que prend ordinairement l'incertitude de mesure, l'incertitude définitionnelle ne qualifie pas une limite de connaissance : elle émane d'un flou dans la définition d'un mesurande, et tant que celle-ci n'est pas modifiée, le flou persiste avec un accroissement des connaissances

35. « (D)es variations dans les procédures ou les conditions peuvent conduire à la mesure de quantités de valeurs différentes mais satisfaisant toutes la définition spécifiée du mesurande. [...] Correctement comprise, alors, ce qui est désigné comme l'incertitude définitionnelle n'est pas due à notre état imparfait de connaissance à propos de la valeur vraie d'un mesurande. Plutôt, elle est due à notre état imparfait de connaissance au sujet de quel mesurande sera effectivement mesuré », Vosk et Emery (2014), p.200-201.

théoriques ou de la précision expérimentale. En effet, l'incertitude définitionnelle est associée à l'imprécision d'une *définition*, et non pas aux limites et à l'imprécision d'une connaissance donnée. Cela se comprend en dehors de toute considération sur la mesure elle-même, et même en dehors du domaine scientifique, comme l'illustre un exemple tiré de la vie ordinaire. Si le client d'un restaurant commande un fruit quelconque, sans plus de précision, le serveur pourra par exemple lui apporter une orange ou une pomme, lesquelles sont toutes deux compatibles avec la requête du client. Cela ne signifie en aucun cas que le client ne sait pas distinguer entre une pomme et une orange... en revanche, ce dernier ne sait pas à l'avance quel fruit lui sera apporté : il y a un doute quant à la *réalisation* de la commande parce qu'il y a un flou dans la *définition* de la commande. Cette subtilité ne semble pas avoir été discernée par les rédacteurs du GUM, où ne perce d'ailleurs aucune remarque afférente aux rapports entretenus entre incertitude définitionnelle et connaissance. À l'inverse, Treiner, dans un contexte un peu différent, a insisté sur la nécessité de bien séparer ce qui qualifie l'objet ou le phénomène lui-même de ce qui qualifie notre connaissance. Cette séparation est essentielle, en ce qu'elle signale quelque chose qui n'a pas toujours été bien explicité en métrologie, et que nous avons déjà eu l'occasion de souligner à la partie I : la variabilité n'est pas toujours synonyme d'incertitude, parce que la première est du domaine du phénomène là où la seconde est du domaine de la connaissance. L'incertitude de mesure est parfois une conséquence de la variabilité, lorsque cette variabilité, en perturbant le processus de mesure, est un obstacle à la connaissance d'une grandeur physique. Mais la variabilité peut être une propriété intrinsèque à la cible que l'on cherche à mesurer. Selon Treiner, la confusion est souvent entretenue sur ce sujet, ce qui est à la source de certaines incompréhensions :

Le plus simple, c'est de caractériser [la distribution des valeurs de la grandeur] par une valeur moyenne et un écart-type [...]. L'écart-type est directement relié à ce qu'on a pris l'habitude d'appeler « incertitude », mais en fait, ainsi que les exemples le montrent, le terme d'incertitude est inapproprié pour décrire une propriété du système. La variabilité est intrinsèque, elle est incompressible, elle fait partie de la connaissance que l'on a du système. [...] (P)our terminer sur le sens des mots, il est utile également de distinguer deux sens (au moins) du terme « incertitude » : l'incertitude au sens de l'écart-type d'une loi de probabilité, et l'incertitude non statistique liée à une ignorance. [...] Il existe des sources d'incertitudes liées à notre ignorance de certains processus. La présence de nuages, par exemple, est difficile à prévoir, car la formation des gouttelettes est un phénomène hors équilibre qui est très sensible à la présence d'aérosols dans l'atmosphère, lesquels servent de centres de nucléation pour le changement de phase vapeur liquide. Or les nuages ont un double effet sur le climat : ils ont tendance à réchauffer les basses couches de l'atmosphère, tout en réfléchissant une partie du rayonnement solaire. C'est la raison pour laquelle les incertitudes sur le rôle des nuages dans le réchauffement climatique ne diminuent pas. Mais si l'on maîtrisait mieux la théorie, ou la façon de l'implémenter dans les codes climatiques, ces incertitudes diminueraient. *Les incertitudes liées à la variabilité d'un phénomène, au contraire, sont stables et ne diminuent pas lorsque les connaissances progressent.*³⁶

36. Treiner (2011), p.13 (nous soulignons). Notons que l'article de Treiner ne mentionne pas l'incertitude défini-

Suivant Treiner, nous constatons aisément que l'incertitude définitionnelle ne décroît pas lorsque nos connaissances sont enrichies ; c'est en fait le phénomène inverse qui a lieu. Lorsque nos connaissances théoriques et expérimentales augmentent, nous pouvons découvrir de nouvelles façons de considérer que la définition d'un mesurande est incomplète. Dans le cadre d'une mesure de longueur, si l'on ignore le phénomène de dilatation thermique, il n'y a aucune raison de penser qu'un mesurande défini comme la longueur d'un objet indépendamment de sa température est un mesurande incomplètement défini. Dès lors que l'on prend conscience que la longueur d'un corps dépend de sa température, il devient nécessaire de prendre en compte cette dépendance dans la définition du mesurande ; ou, alternativement, de l'ignorer, ce qui se fait au prix d'une incertitude définitionnelle. Dans ce cas, nous aboutissons à une conclusion en apparence paradoxale : *l'incertitude définitionnelle est la conséquence d'un accroissement de nos connaissances.*

De ce point de vue, l'incertitude définitionnelle dépend effectivement de l'état des connaissances à l'instant considéré. On aurait pu penser, au premier abord, que le domaine dans lequel s'exprime l'incertitude définitionnelle n'est pas celui qui est propre à l'incertitude de mesure : puisque la première nommée marque l'imprécision d'une *définition*, elle apparaît comme une caractéristique de l'objet défini, non d'une quelconque connaissance. Le terme d'incertitude « intrinsèque » utilisé dans le GUM et dans le VIM2 allait dans ce sens en occultant la problématique de la définition d'une grandeur et en laissant entendre que l'incertitude était une caractéristique de la grandeur en elle-même (c'est ainsi que l'on pouvait comprendre la phrase, bien que bancale, du GUM : « tout mesurande possède une telle incertitude "intrinsèque" »³⁷). Le terme « incertitude définitionnelle » nous semble plus adapté et nous permet de mieux comprendre le fait que cette dernière n'est pas une mesure de l'imprécision intrinsèque d'une grandeur, ni même d'ailleurs de l'incomplétude intrinsèque de la définition d'un mesurande, mais *décrit la façon dont l'expérimentateur croit, ou sait, que la définition de ce mesurande est incomplète.* Notre description de l'incertitude définitionnelle ne doit pas donner l'impression que celle-ci est purement rattachée à des questions de modélisation de la part de l'expérimentateur ; elle a bien entendu des causes physiques, comme la variabilité d'un phénomène physique, une limite quantique, etc. Toutes ces causes sont des causes de non-unicité de la valeur vraie, mais l'incertitude définitionnelle elle-même n'a pas trait directement à ces causes, mais à la façon dont l'expérimentateur en tient compte, lorsque ces causes sont connues. L'incertitude définitionnelle se situe donc à un croisement : entre causes physiques de non-unicité d'une valeur vraie, connaissance de ces causes, et choix de modélisation qui se traduit en particulier au travers de la définition du mesurande. L'incertitude définitionnelle ne caractérise ni un manque de connaissance ni une propriété intrinsèque des grandeurs, mais plutôt une relation entre notre connaissance et les définitions des mesurandes, c'est-à-dire une incomplétude des définitions *relativement à nos connaissances* des causes de non-unicité, et relativement aux modèles utilisés.

tionnelle, d'autant que l'auteur relate sa méfiance vis-à-vis des textes officiels de métrologie : « [Des] incohérences [...] sont, hélas, présentes dans un certain nombre de textes officiels concernant la métrologie. Ce n'est pas une raison pour les reprendre telles quelles dans notre enseignement. Il est plus... éducatif, de leur montrer que même les textes officiels, parfois... », pp.9–10. Cependant, l'article déploie des arguments qui nous semblent éclairer en retour la notion d'incertitude définitionnelle.

37. [Joint Committee for Guides in Metrology \(JCGM\) \(2008a\)](#), p.52. Voir la citation que nous reproduisons p.205.

Ainsi, plus précisément encore, seules les causes *connues* de non-unicité de la valeur vraie d'un mesurande peuvent être considérées comme source d'incertitude définitionnelle. Ces causes sont sciemment écartées de la modélisation du problème, par choix de l'expérimentateur – dès lors, il n'est plus possible d'en faire état dans le bilan d'incertitude classique.

À la lumière de ces différentes remarques, l'incertitude définitionnelle apparaît comme la conséquence d'un flou définitionnel, résultant à la fois de certains choix de la part de l'expérimentateurs, et de certaines contraintes imposées par la forme des phénomènes physiques étudiés. Certains choix de l'expérimentateur ont une finalité essentiellement pragmatique, par exemple ne pas prendre en compte des phénomènes de dilatation thermique afin de ne pas aboutir à un modèle déraisonnablement complexe. L'expérimentateur n'a cependant pas une latitude infinie dans les choix de définition, lesquels sont contraints, en particulier, par l'échelle d'étude du phénomène. Nous développerons cela à la section 7.3, où nous affirmerons que l'incertitude définitionnelle est le produit d'un double compromis. D'une part, sur le plan de la conceptualisation, l'incomplétude de la définition d'un mesurande est ce qui permet l'identification même d'une grandeur physique – il n'y a pas de sens à définir la longueur d'une table en termes atomiques car on perdrait alors de vue l'objet « table » lui-même. D'autre part, sur le plan pragmatique, il est tout à fait possible de faire l'économie d'un certain nombre de spécifications dans la définition d'un mesurande, comme par exemple négliger un phénomène dont la conséquence est marginale sur la grandeur étudiée, sans pour autant que cela rende impossible la mesure et l'exploitation de la mesure du mesurande en question.

7.2.3 L'incertitude définitionnelle comme *composante* d'incertitude

La discussion précédente répond à une question qui n'a pas été abordée dans le GUM, à savoir pourquoi il est légitime de considérer l'incertitude définitionnelle comme une notion épistémique, et, de ce fait, de l'appeler « incertitude ». La description du GUM présente un second défaut notable : l'incertitude définitionnelle y est immédiatement considérée comme *quantifiable*, au même titre que toute autre composante d'incertitude ; autrement dit, le GUM affirme sans questionnement qu'elle peut être intégrée à un bilan d'incertitude global³⁸. De fait, le GUM et le VIM présentent l'incertitude définitionnelle comme la *limite basse* et incompressible de l'incertitude de mesure pour un mesurande donné³⁹. C'est donc qu'en définitive, l'incertitude définitionnelle a été additionnée – « propagée » – d'une façon ou d'une autre avec les autres composantes d'incertitude. La figure 7.1, tirée du GUM, montre sans ambiguïté possible que le document fait de l'incertitude définitionnelle une composante d'incertitude comme les autres. Si cette incertitude peut être quantifiée, puis propagée, se pose alors la question suivante : comment cela peut-il être effectué sur le plan mathématique ? Chercher à calculer l'incertitude définitionnelle, c'est rouvrir pour ce type d'incertitude les questions déjà abordées pour les autres composantes : comment peut-on la déterminer ? Quelles informations doivent

38. Notons qu'on retrouve à peine plus d'interrogations à ce sujet dans la littérature métrologique plus récente, en particulier parce que l'incertitude définitionnelle n'y a été effleurée qu'avec un intérêt plutôt mineur jusqu'ici.

39. « C'est l'incertitude minimale avec laquelle on peut déterminer un mesurande et chaque mesurage qui atteint une telle incertitude peut être considéré comme le meilleur mesurage possible du mesurande. », [Joint Committee for Guides in Metrology \(JCGM\) \(2008a\)](#), p.52.

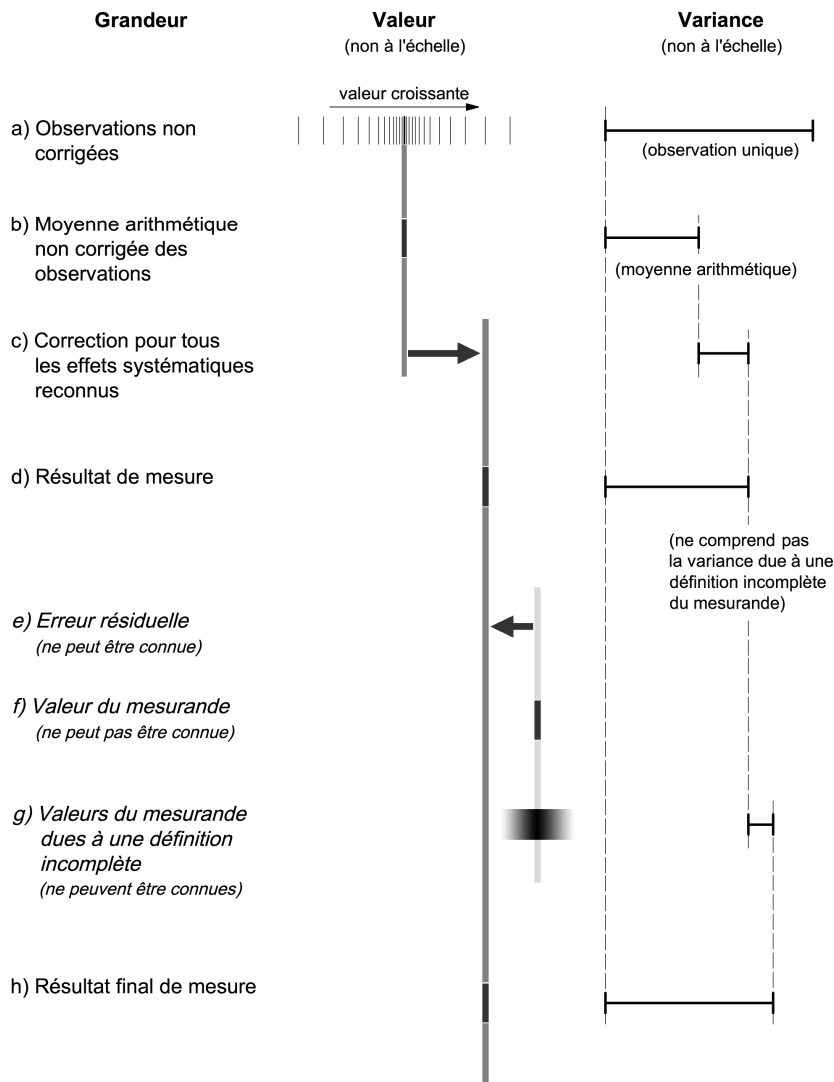


FIGURE 7.1 – Figure reproduite à partir du GUM (figure D.2, p.55), intitulée « Illustration graphique des valeurs, de l'erreur et de l'incertitude ». La figure met en regard une description des résultats de mesure en termes d'erreurs et en termes d'incertitudes. Dans la colonne du milieu, sous l'intitulé « valeur », sont représentées les données expérimentales dispersées (a), dont on prend la moyenne arithmétique (b) que l'on corrige pour les effets systématiques que l'on a identifiés (c). L'erreur de mesure (e) est alors l'écart qui subsiste après correction avec la (les) valeur(s) vraie(s) du mesurande (f, g). L'incertitude de mesure est représentée dans la colonne de droite, sous l'intitulé « variance ». La variance de la moyenne arithmétique (b) est combinée avec l'incertitude-type associé aux erreurs systématiques connues (c) pour obtenir l'incertitude composée (d) : c'est la propagation des incertitudes, représentée géométriquement par la concaténation des segments (b) et (c). Mais on constate aussi que l'incertitude composée (d) est encore combinée avec l'incertitude définitionnelle (g) pour obtenir l'incertitude finale (h). Cette figure témoigne du fait que, dans le GUM, l'incertitude définitionnelle est comprise comme une composante ordinaire d'incertitude de mesure, quantifiable et que l'on peut combiner dans un bilan d'incertitude pour obtenir une résultante globale.

être exploitées? Doit-on, et peut-on, en proposer une interprétation probabiliste? Comment, enfin, peut-on combiner la composante correspondante avec les autres composantes d'incertitude?

Il se trouve que le GUM ne donne presque aucune direction quant à la méthode d'évaluation de l'incertitude définitionnelle. Tout au plus est-il expliqué sur l'exemple déjà cité précédemment que « la valeur 'vraie' [de la longueur d'une feuille de matériau] présente une incertitude qui peut être évaluée à partir de mesurages réalisés en différents emplacements de la feuille »⁴⁰, ce qui laisse entendre que l'incertitude définitionnelle peut être déterminée expérimentalement. Par la suite, dans le document, une telle question est immédiatement évacuée : il est précisé que son spectre d'application se limite aux cas pour lesquels il existe une valeur (vraie) « essentiellement unique »⁴¹. La solution du GUM consiste donc à rappeler que la valeur « vraie » d'un mesurande ne peut que très rarement être considérée comme unique, pour mieux souligner l'idéalité du concept; puis à en conclure à l'existence d'une composante spécifique d'incertitude, l'incertitude définitionnelle, pour enfin supposer cette composante d'incertitude négligeable dans les cas couverts par le document. Cette solution présente le remarquable intérêt de ne pas rendre le problème démesurément complexe : elle rappelle à l'expérimentateur qu'il doit apporter un grand soin à la définition du mesurande, mais ne l'oblige pas à effectuer en supplément une tâche fastidieuse qui ne lui apporterait qu'un maigre bénéfice.

Tout pragmatique qu'il soit, le choix effectué dans le GUM apparaît comme une échappatoire un peu trop opportuniste : si le GUM évacue la question de la détermination quantitative de l'incertitude définitionnelle, ce n'est peut-être pas seulement dans le but de simplifier la tâche des utilisateurs, mais également parce que cette question ne trouve pas de réponse claire ! Il y a donc un décalage entre les questions soulevées dans le GUM, et la façon dont il y répond : si l'incertitude définitionnelle est une composante d'incertitude comme les autres, il devrait exister un moyen de l'évaluer quantitativement de façon rigoureuse. Mais en fait, rien jusqu'ici dans le GUM n'est venu justifier le fait que l'incertitude définitionnelle doive être comprise comme une composante ordinaire d'incertitude. Une autre solution au problème d'ensemble consiste donc à affirmer qu'au contraire, l'incertitude définitionnelle *n'est pas* une composante d'incertitude, et que dès lors, il n'est peut-être pas indispensable de chercher à la quantifier.

Cette seconde solution est la position que nous défendons. Une incertitude de mesure caractérise un doute quant à la valeur du mesurande; cela suppose que la valeur du mesurande

40. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.52; nous renvoyons à la section 7.2.1 de cette thèse pour la citation complète d'où est tirée cette remarque.

41. Nous nous référons ici à la version anglaise du document, [Joint Committee for Guides in Metrology \(JCGM\) \(2008a\)](#), p.1, plus expressive que la version française qui fait état « d'une grandeur physique bien définie – le mesurande – qui peut être caractérisée en première approximation par une valeur unique. », [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.1. Dans le VIM3, il est fait mention d'une valeur « par essence unique », [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.x, ce qui ne semble pas porter tout à fait la même signification, car les expressions employées dans le GUM font quant à elles sentir le caractère *approximatif* de l'unicité de la valeur du mesurande. Plus loin dans le VIM3, ce caractère approché refait clairement surface : « Lorsque l'incertitude définitionnelle associée au mesurande est considérée comme négligeable par rapport aux autres composantes de l'incertitude de mesure, on peut considérer que le mesurande a une valeur vraie par essence unique. », [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.20.

est donnée d'une part, et qu'elle n'est pas parfaitement connue d'autre part. Or, l'incertitude définitionnelle traduit le fait que la valeur du mesurande est mal *définie* – c'est-à-dire que la cible elle-même est floue – mais pas nécessairement mal *connue*. L'incertitude définitionnelle ne caractérise pas une limite de connaissances. Par conséquent, nous pensons qu'incertitude de mesure et incertitude définitionnelle ne représentent tout simplement pas la même chose, et qu'il n'y a pas de sens à les combiner ensemble. La suite de notre analyse nous mènera à suggérer que l'incertitude définitionnelle doit être comprise d'abord comme un outil de réflexion sur la validité et la portée des modèles employés. Nous plaiderons alors pour une appréhension semi-quantitative de l'incertitude définitionnelle : un ordre de grandeur suffit. Pour cela, nous passons d'abord par une analyse plus détaillée de ce qu'est la « non-unicité » de la valeur vraie d'une grandeur, pour revenir ensuite sur la notion d'incertitude définitionnelle.

7.3 « Non-unicité » et grandeurs physiques

Couplée avec quelques remarques éparses sur le caractère idéal de la notion de valeur vraie elle-même, la façon dont est introduite puis gérée l'incertitude définitionnelle dans le GUM permet de s'extraire sans trop de dommages de la problématique que pose inévitablement l'existence d'une multiplicité de valeurs, dites « vraies », compatibles avec la définition d'un mesurande donné. En suivant les instructions du GUM, un expérimentateur conclura que, s'il souhaite éviter toute complication, il n'aura plus qu'à rester conscient de l'importance que joue la définition du mesurande dans toute démarche expérimentale, et à s'assurer que le mesurande considéré possède une valeur vraie « essentiellement unique » – dans le cas contraire, il lui faudra évaluer l'incertitude définitionnelle correspondante de façon à l'intégrer au bilan global d'incertitude. Cependant, cette solution s'appuie sur un pré-requis implicite. En effet, elle suppose que le mode de mathématisation des grandeurs physiques n'est pas affecté par le constat de non-unicité de la valeur vraie. Ce faisant, elle fait l'économie des questions, pourtant essentielles, que nous avons évoquées à la section 7.1.3 : (i) Quelle est la signification du terme « vraie » pour la « valeur vraie » d'une grandeur dès lors que celle-ci n'est pas unique ? (ii) Comment peut-on interpréter les équations de la physique si aux grandeurs mises en jeu ne correspondent pas une valeur vraie unique ?

Considérer l'incertitude définitionnelle comme une composante d'incertitude est un mouvement un peu particulier. Rappelons que l'incertitude définitionnelle diffère des composantes habituelles de l'incertitude de mesure en ce qu'elle ne caractérise pas les limites d'une connaissance quant à la valeur du mesurande. Elle est l'expression d'un constat, selon lequel il n'est pas possible de caractériser le mesurande par *une* valeur « vraie » *unique* – elle traduit donc un doute sur l'identité même de la cible de la mesure. Avant de chercher à quantifier une incertitude dite définitionnelle, il est primordial d'interroger au préalable les conséquences du constat de non-unicité de la valeur vraie d'un mesurande sur la validité des modes classiques de mathématisation des grandeurs physiques⁴². Il s'agit bien sûr de revenir sur le concept de

42. Lors de la conférence organisée au National Physical Laboratory (NPL, Londres) en novembre 2013, à l'occasion du vingtième anniversaire de la publication du GUM, Ehrlich a témoigné du fait que les métrologues sont tout à fait conscients de cette difficulté. Il note en particulier que « si l'incertitude définitionnelle est significative en comparaison avec l'incertitude-type composée [...] alors il n'y a pas de valeur vraie essentiellement unique du

valeur vraie lui-même : si une telle valeur vraie n'est pas unique, le concept conserve-t-il un sens ? Par ricochet, cela nous invite également à revenir sur les équations qui décrivent mathématiquement les « lois de la physique ». Ces équations sont traditionnellement comprises comme des équations entre valeurs des grandeurs. Or, l'idée de valeurs vraies multiples entre en conflit avec cette conception. Ce conflit peut-il être résolu ? Ou bien faut-il envisager que les équations aux grandeurs portent sur d'autres objets ? Ces questions sont deux faces d'une même problématique plus générale quant à la notion de *grandeur physique* elle-même.

Dans cette section, nous cherchons à répondre à ces questions en montrant qu'il reste possible de conserver l'interprétation traditionnelle des grandeurs physiques et des équations aux grandeurs, à condition de se pencher sur la structure des modèles qui rendent possible leur expression. Dans ce cadre, l'incertitude définitionnelle peut être remise à contribution, mais avec un rôle différent que celui que lui accorde le GUM. Plutôt que de la considérer comme une composante à ajouter à un bilan complet d'incertitude de mesure, nous avancerons l'idée selon laquelle elle est avant tout un indicateur (parmi d'autres) de la validité de nos modèles. L'incertitude définitionnelle permet de comparer les modèles que nous utilisons avec les *objectifs* que servent ces modèles.

7.3.1 Mesurandes et modèles

Nous souhaitons d'abord montrer en quoi la notion de mesurande et l'argument de non-unicité de la valeur vraie ne peuvent être compris que si l'on invoque deux notions étroitement liées : celle de modèle et celle d'approximation.

De Courtenay a insisté sur le caractère théorique de la notion de grandeur physique. L'expérience montre que les propriétés physiques ne vérifient pas empiriquement les conditions de mesurabilité – nécessaires pour que l'on puisse les qualifier de « mesurables ». La définition des grandeurs physique nécessite un acte théorique, où l'on *postule* que les grandeurs sont mesurables et l'on *explique* les divergences entre les constats empiriques et les propriétés idéales d'une grandeur mesurable par la présence d'erreurs de mesure⁴³ – les conditions de mesurabilité deviennent alors des axiomes de la mesure. Dans le domaine de la métrologie, Giordani et Mari ont insisté à leur tour sur la place grandissante qui est accordée au rôle des modèles⁴⁴. Comme ces derniers l'ont souligné, la définition d'un mesurande par un expérimentateur nécessite que l'expérimentateur construise un *modèle* de l'objet soumis à la mesure⁴⁵, sans quoi il est impossible d'identifier la grandeur elle-même. Ainsi, parler de « longueur » pour la feuille de matériau décrit dans l'exemple du GUM⁴⁶, ou encore une simple feuille de papier, tient au fait que cette feuille est assimilée à une structure géométrique (typiquement un parallépipède rectangle) pour laquelle la grandeur mathématique qu'est la « longueur » est bien définie. Cette

mesurande, puisque toutes les valeurs dans le spectre de l'incertitude définitionnelle satisfont la définition du mesurande, et par conséquent la signification de la courbe de distribution de probabilité [décrivant le résultat de mesure] devient quelque peu obscure », Ehrlich (2014), p.S151.

43. de Courtenay (2008)

44. Giordani et Mari (2012), Mari et Giordani (2014)

45. Giordani et Mari (2012), p.2147.

46. Nous renvoyons à la section 7.1.2.

construction renvoie à ce que McMullin appelle l'« idéalisation mathématique »⁴⁷ : bien que les objets traités par les sciences physiques ne soient pas des objets mathématiques⁴⁸, il est possible de les représenter comme tels⁴⁹, mais cela nécessite une étape d'abstraction. À l'intérieur du modèle géométrique, la feuille possède effectivement une longueur bien définie. La construction du mesurande à l'intérieur d'un modèle permet, par suite, d'insérer ce modèle dans un (ou plusieurs) modèles phénoménologiques qui explicitent des relations physiques entre grandeurs physiques. Ainsi, la mesure des dimensions d'un ensemble de meubles permettent de prévoir une façon de les agencer dans une pièce. De même, mesurer la résistance d'un conducteur ohmique nécessite de construire un modèle de résistance, dans un circuit électrique, résistance qui est supposée vérifier la loi phénoménologique d'Ohm, laquelle établit une relation entre trois grandeurs (courant traversant le conducteur, tension aux bornes du conducteur, et résistance du conducteur)⁵⁰.

Nous pourrions tenter de contourner cette nécessité par une définition opérationnaliste de la longueur de la feuille : si la longueur de la feuille est définie comme ce qui est effectivement mesuré par l'instrument choisi (un micromètre dans l'exemple du GUM), alors il n'y a, en apparence, pas d'ambiguïté quant à ce qui est mesuré, et cela ne fait intervenir aucune hypothèse quant à la forme effective de la feuille. Cependant, comme nous l'avons déjà noté lorsque nous avons introduit la notion de mesurande, ce mouvement opérationnaliste est insatisfaisant sur plusieurs points. Il fait barrage à une opération essentielle de l'activité de mesure, à savoir le processus de *correction*, puisque, suivant l'acception opérationnaliste, l'expérimentateur mesure toujours exactement ce qu'il souhaite mesurer. Dans la continuité, cela fait que le résultat obtenu devient inutilisable en dehors du contexte immédiat de sa mesure – le résultat ne peut pas être utilisé par d'autres, il n'est pas communicable, et il ne participe donc pas à l'entreprise scientifique *collective*.

L'idéalisation mathématique est l'un des ressorts de l'idée même de mathématisation d'une grandeur. Ici, la question n'est pas de savoir si l'on peut parler de valeur « vraie » pour une grandeur physique, mais de comprendre s'il y a un sens à parler de grandeur tout court pour un objet particulier. Il n'est possible de parler de longueur pour une feuille de matériau que parce l'idéalisation mathématique permet de concevoir proprement cette grandeur pour cet objet particulier⁵¹. De plus, souligne alors McMullin, à l'étape d'abstraction que fait intervenir

47. McMullin (1985), pp.248–254.

48. « La géométrie est une abstraction, une idéalisation. Elle laisse de côté le détail qualitatif qui constitue le singulier physique comme physique », McMullin (1985), p.249.

49. C'est la position que Galilée exprime par la voix de Salviati, dans son *Dialogue sur les deux grands systèmes du monde*, et que McMullin résume ainsi : « la réalisation en question n'est pas une barrière à l'intelligibilité en des termes géométriques, et [les] conséquences des "obstacles" dus à la difficulté d'appliquer des concepts géométriques simples aux complexités du monde sensible peuvent elles-mêmes, en principe, être saisies », McMullin (1985), pp.251–252. McMullin insiste sur le fait que l'opération d'idéalisation mathématique n'est nullement une *réduction* de la physique aux mathématiques : « Ce qui est arrivé est que les "mathématiques" utilisées en physiques ont graduellement incorporé de plus en plus de contenu physique. Le Livre de la Nature n'est *pas* écrit dans le langage des mathématiques à strictement parler. La *syntaxe* est mathématique, mais la *sémantique* ne l'est pas. [...] Le Livre de la Nature, bien qu'il emploie une grammaire mathématique, est écrit dans le *langage* de la physique (ou de la chimie ou de la biologie...) », pp.252–253.

50. Cet exemple est repris de Giordani et Mari (2012), pp.2147–2149, où il est développé plus longuement.

51. Notons qu'il ne s'agit pas ici d'affirmer qu'un modèle est nécessaire pour définir la longueur comme grandeur *générale* (ou type de grandeur). Si la longueur est une grandeur physique, c'est parce que c'est le résultat d'une

l'idéalisation mathématique s'ajoute également une étape d'approximation : la représentation de la Terre par une sphère n'a de valeur que parce que la Terre est *approximativement* une sphère; et cette approximation n'a elle-même de valeur que parce que les conséquences des inexactitudes que cette approximation induit peuvent être maîtrisées d'une façon ou d'une autre⁵². En fin de compte, le constat de non-unicité de la valeur vraie apparaît ici tout à fait immédiat : il n'est pas possible de « forcer » la réalité dans le modèle mathématique choisi, et donc de mesurer une valeur vraie unique. L'incertitude définitionnelle est alors comprise comme la contrepartie de l'idéalité des grandeurs physiques, et des approximations du modèle employé vis-à-vis de l'objet matériel soumis à la mesure.

Comme nous l'avons répété à plusieurs reprises, la définition d'un mesurande est également suspendue aux « grandeurs d'influence », c'est-à-dire aux phénomènes physiques qui affectent le résultat de la mesure. Si l'on considère un objet ordinaire, comme un crayon, la définition d'un mesurande tel que « la longueur d'un crayon » apparaît incomplète sous plusieurs angles. D'abord, parce que le crayon n'est pas une figure géométrique régulière – cela renvoie à l'idéalisation mathématique mentionnée précédemment. Mais également parce que la longueur du crayon est affectée par différents effets physiques : la dilatation thermique, ou encore l'élongation sous l'effet de la pesanteur. Ainsi, le crayon, tenu horizontalement ou verticalement à la surface de la Terre, n'aura pas la même longueur; en raison de l'élasticité du matériau, la force de pesanteur induit une variation de longueur dont l'amplitude est de l'ordre de quelques micromètres⁵³. On peut ainsi envisager différents niveaux de précision dans la définition d'un mesurande correspondant à ce que l'on considérerait de manière générale être la « longueur du crayon » :

- (i) « Longueur du crayon » – sans spécification supplémentaire.
- (ii) « Longueur du crayon à la température T »
- (iii) « Longueur du crayon à la température T , tenu (par exemple) horizontalement »

L'incertitude définitionnelle diminue de (i) à (iii). On considère généralement que ce que l'on pourrait appeler la « longueur horizontale » et la « longueur verticale » du crayon sont deux instances d'une *même* grandeur (particulière) qui varie selon les conditions d'observation – et qu'il en est de même pour la dépendance de la longueur en température. Il est possible d'identifier la « longueur horizontale » et la « longueur verticale » du crayon car quelque chose relie ces différentes définitions, à savoir l'identification du « crayon » comme un individu matériel bien défini. Apparaît alors une idéalisation supplémentaire – et une difficulté associée – relative à l'identification des objets matériels qui sont supposés « porter » une grandeur physique bien

idéalisation des conditions de mesurabilité de l'espace. En ce sens, le concept de longueur ne précède pas celui de mesure. Cependant, pour un objet spécifique – une feuille de matériau – la possibilité de définir une longueur, comprise comme une grandeur *particulière*, est également attachée à la construction d'un modèle géométrique (puis physique) de l'objet. Pour une discussion de la différence entre type de grandeur et grandeur particulière, voir [Mari \(2009\)](#).

52. C'est une étape d'idéalisation supplémentaire que McMullin nomme « idéalisation de construction », [McMullin \(1985\)](#), pp.254–259.

53. Je remercie Marc Priel, du LNE (Laboratoire National de métrologie et d'Essais, Paris) à qui je dois cet exemple particulier.

précise. Certaines grandeurs ne sont attachées à aucun objet corporel *spécifique*. C'est le cas par exemple des constantes physiques, comme la constante de Planck, ou de constantes phénoménologiques comme la conductivité d'un métal particulier. Mais il existe une autre catégorie de grandeurs *particulières* associées à des objets eux-mêmes particuliers. Or, ces objets n'ont d'identité qu'à une certaine échelle. Ainsi, à l'échelle microscopique, il est impossible de bien définir la frontière entre le crayon et l'air environnant⁵⁴. Le constat général est le suivant : ajouter des spécifications à la définition du mesurande mène à réduire l'incertitude définitionnelle, mais résulte en un détachement progressif de l'objet originellement étudié pour en arriver à une étude de ses sous-structures, auquel il ne peut pas forcément se réduire simplement. De ce fait, l'identification et l'isolation d'un individu comme « le crayon », nécessaire à l'entreprise menée, se fait au prix d'une incertitude définitionnelle. Il serait toujours possible, par exemple, de définir la longueur d'un crayon comme étant la distance entre les centres de masses de deux noyaux identifiés comme appartenant à deux extrémités du crayon, et définis à l'avance. Dans ce cas, l'incertitude définitionnelle disparaîtrait (ou diminuerait grandement), mais la grandeur ainsi définie ne présenterait aucune signification – et sa mesure ne serait d'aucune utilité. C'est parce que l'expérimentateur veut mesurer la longueur d'un objet, le « crayon », et non pas une configuration de particules élémentaires à l'échelle microscopique, qu'il a besoin de construire un modèle de cet objet à l'échelle macroscopique. Il est naturel pour un expérimentateur d'adapter la définition du mesurande à ses besoins, et l'incertitude définitionnelle apparaît ici la conséquence d'un *choix*, le choix d'étudier des phénomènes à une certaine échelle et dans un certain but. Nous en arrivons ainsi à la conclusion suivante : l'incertitude définitionnelle est l'expression du coût de nos choix de modélisation, et se situe donc à l'interface entre le modèle et l'objet « réel », *en fonction de notre connaissance sur cette interface*.

L'incertitude définitionnelle apparaît ici liée à une problématique touchant de près à la question du réductionnisme en science. Elle caractérise le fait que, *dans nos théories et modèles*, le mesurande choisi est un individu *complexe* qui présente une sous-structure (i) à laquelle il n'est pas dit que ses propriétés se réduisent nécessairement ou (ii) que l'expérimentateur peut *choisir* de ne pas explorer. Par contraste, le *VIM3* mentionne que « (d)ans le cas particulier des constantes fondamentales, on considère la grandeur comme ayant une seule valeur vraie »⁵⁵ ; cela est valable précisément parce que ce sont des grandeurs considérées comme fondamentales. Il n'est pas nécessaire pour cela d'interpréter le terme « fondamental » dans une perspective ontologique ; l'incertitude définitionnelle est, nous l'avons défendu, épistémique et donc attachée à un état de connaissance donné. L'élément décisif est ici qu'il n'existe pas de sous-structure *connue* aux constantes fondamentales, c'est-à-dire encore que les théories n'en proposent aucune *explication*. (Notons que pour ces grandeurs considérées comme fondamentales, l'absence d'incertitude définitionnelle ne provient pas d'une profusion de détails apportés à la définition du mesurande. Au contraire, dans ce cas précis, les mesurandes ont souvent une définition remarquablement concise.) Au vu de cette discussion, il nous semble

54. Robert et Treiner font valoir un argument du même type : « si un bois possède une odeur, c'est bien parce que des molécules le quittent sans arrêt. La notion usuelle de 'largeur' perd donc son sens en deçà de l'échelle de quelques molécules, ce qui, reconnaissons-le, ne pose pas grande difficulté pour la vie quotidienne. », *Robert-Schwartz et Treiner (2006)*, p.2.

55. *Joint Committee for Guides in Metrology (JCGM) (2012)*, p.20.

| | Variables d'état | Constantes générales |
|-------------------|--|---|
| phénoménologiques | masse d'un échantillon d'uranium raideur d'un ressort | conductivité électrique du cuivre densité de l'eau pure à température et pression données |
| fondamentales | État quantique d'un électron donné | Masse de l'électron Constante de Planck |

TABLEAU 7.3 – Essai de classification des grandeurs selon un critère double. D'un côté, les grandeurs physiques peuvent être des variables d'état, qui caractérisent un objet spécifique et ont une dynamique que l'on caractérise par une équation dépendante du temps; ou des constantes physiques, « qui décrivent des propriétés universelles des interactions de la matière » (Uzan, 2004, p.14). D'un autre côté, elles peuvent être des grandeurs phénoménologiques ou fondamentales selon qu'il existe ou non une ou plusieurs théories qui viennent expliquer leur état.

légitime de lier l'identification de l'incertitude définitionnelle à une double classification des grandeurs physiques : entre grandeurs phénoménologiques et fondamentales d'un côté, et entre variables d'état et constantes d'un autre côté (voir tableau 7.3). Répétons que cette classification ne suggère pas que les statuts des grandeurs soient définitivement figés. Des grandeurs comme la masse de l'électron, ou même la constante de Planck sont uniquement fondamentales *dans l'état présent de nos principales théories*; elles pourraient ne plus l'être dans des théories en gestation comme la théorie des cordes.

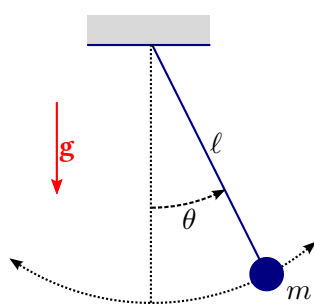
Dans la suite de cette section, nous explicitons de façon plus précise ce que nous entendons par une distinction entre le niveau fondamental et le niveau phénoménologique. Pour cela, nous nous appuyons sur un débat qui a opposé séparément Nancy Cartwright à Ernan McMullin et Ronald Laymon, et dont l'un des enjeux porte sur le niveau de réalisme que l'on peut accorder aux lois de la physique. Cartwright a énoncé ce désormais célèbre aphorisme selon lequel « les lois de la physique mentent »⁵⁶ – en fait, plus spécifiquement les lois fondamentales de la physique. L'analyse de ce débat nous permettra de mettre en évidence la façon dont les modèles sont un lieu naturel pour l'expression des approximations – et de réévaluer la notion de valeur vraie à travers ce prisme.

7.3.2 Modèles et approximation

L'épistémologie des modèles et celle de l'approximation ont donné lieu à une littérature abondante qui les a traitées sous de nombreux angles⁵⁷. Nous ne souhaitons pas revenir exhaustivement sur ces questions, mais plutôt l'aborder avec une perspective qui nous intéresse ici, c'est-à-dire en regard d'une problématique bien spécifique : que pouvons-nous apprendre à propos du statut de vérité des théories physiques, et, plus précisément, des équations qui ex-

56. Cartwright (1983)

57. Franck Varenne a proposé une synthèse de ces débats articulée autour de la question du réalisme scientifique, Varenne (2012). Voir aussi Frigg et Hartmann (2012). Olivier Darrigol fait valoir une piste quelque peu différente dans sa description de la structure modulaire des théories physiques, voir Darrigol (2007). Il critique en particulier la conception des modèles proposée par Cartwright et Morrison.



Résolution classique de manuel (mouvement plan)

- « Principe fondamental de la dynamique » : $m \mathbf{a} = m \mathbf{g} + \mathbf{T}$, où le vecteur \mathbf{a} représente l'accélération de l'objet et \mathbf{T} la force de tension exercée par le fil.

- Par calcul est déduite l'équation dite « du mouvement »

$$\ddot{\theta} + \frac{g}{\ell} \sin \theta = 0$$

- L'équation est ensuite linéarisée : $\ddot{\theta} + (g/\ell) \theta = 0$
- La résolution de l'équation différentielle permet de remonter à la trajectoire de l'objet, sa période d'oscillation, etc.

⇒ possibilité de confrontation expérimentale

FIGURE 7.2 – Modèle du pendule simple : un objet de masse m est suspendu à un fil de longueur ℓ , dont l'extrémité opposée est accrochée à un support fixe. Le champ de pesanteur est représenté par le vecteur \mathbf{g} . L'angle entre le fil et la verticale est appelé θ . Sous l'effet de la pesanteur, l'ensemble fil et masse a tendance à effectuer des oscillations, dont on peut chercher à déterminer la dynamique, et en particulier la période d'oscillation, en s'appuyant sur certaines hypothèses additionnelles.

priment mathématiquement les « lois » physiques ? Cela dessine en creux un questionnement classique de la physique : comment les lois et théories physiques sont-elles testées et en quoi ces tests expérimentaux peuvent-ils être considérés comme les garants de leur statut de vérité ?

Dans cette perspective, il importe tout d'abord de reconnaître le caractère indispensable des idéalizations et de la construction de modèles physiques dès lors que l'on cherche à confronter une théorie à l'expérience. Laymon nomme cette thèse « l'omniprésence des idéalizations »⁵⁸. Cela peut être illustré à l'aide de l'exemple du modèle du « pendule simple », un grand classique des manuels de sciences physiques (figure 7.2). Un objet est suspendu à un fil dont l'autre extrémité est accrochée à un support fixe, et est lâché d'une position initiale quelconque, avec une vitesse initiale nulle ; sous l'effet de la pesanteur, l'objet effectue des oscillations. Le modèle permet de retrouver le caractère oscillatoire de la trajectoire de l'objet, et en particulier de prédire la période de ses oscillations. Le cadre général du problème est celui de la mécanique classique : sa résolution met à contribution le contenu de cette théorie (en particulier ses lois). Mais les lois de la mécanique classique seules ne permettent ni la mise en place ni la résolution du problème, et il faut passer pour cela par une importante liste d'idéalizations et d'approximations :

- La masse est supposée ponctuelle, le fil inextensible, le support rigide et fixe.
- La masse est supposée être soumise aux seules forces de pesanteur et de tension du fil (pas de frottement, pas d'effet thermique, etc.)
- Le champ de pesanteur est supposée uniforme
- Le référentiel d'étude est supposé galiléen

Sous ses conditions, les lois de la mécanique newtonienne permettent de dériver une « équation du mouvement », c'est-à-dire une équation différentielle dont l'inconnue est un paramètre

58. Laymon (1989), p.357.

de la trajectoire de l'objet, ici sa position angulaire. L'équation du mouvement obtenue est non-linéaire ; elle peut être résolue au moyen des « intégrales elliptiques », et de la « fonction elliptique de Jacobi » introduites par Jacobi vers 1830. Cette résolution reste assez fastidieuse en regard du problème posé, et le cas d'étude classique est donc celui des « petites oscillations », ce qui permet, par une approximation supplémentaire, d'obtenir une équation différentielle linéaire, dont la solution (fonction des conditions initiales) donne le mouvement de l'objet à travers l'expression de l'angle en fonction du temps. L'on peut ensuite déterminer la période des oscillations. C'est alors, et alors seulement, que les prédictions du modèle peuvent être confrontées avec les résultats obtenus expérimentalement.

Il apparaît donc que les théories physiques ne peuvent pas être confrontées directement à l'expérience. Elles doivent être mises en situation dans un modèle qui permet leur expression : les lois de la théorie testée permettent alors de dériver le résultat prédit à partir des hypothèses du modèle⁵⁹. De plus, les hypothèses des modèles ne sont généralement pas guidées par la théorie elle-même (ainsi Humphreys a souligné qu'entrent régulièrement en jeu des critères pratiques liés à la simplicité et à la maniabilité des modèles⁶⁰, ce qui se traduit par exemple par l'importance des modèles linéaires). Dans tous les cas, les modèles ne se *déduisent pas* des théories⁶¹. Cela a conduit Morgan et Morrison à considérer les modèles comme largement indépendants des théories qu'ils mobilisent, et à souligner le rôle des modèles comme « médiateurs », comme « agents autonomes »⁶². À cette indépendance entre théorie et modèle s'ajoute un second point. En reprenant l'étude de ce qu'Achinstein a appelé « modèles théoriques »⁶³, Redhead a montré que, de façon remarquable, les prédictions de ces modèles sont en contradiction avec les éléments des théories plus fondamentales qu'ils incorporent⁶⁴. Le modélisateur du pendule simple sait que le « référentiel terrestre » n'est pas rigoureusement galiléen, que l'objet oscillant n'est pas un point matériel, etc. De même, les solutions de l'équation linéarisée ne sont pas des solutions de l'équation non-linéaire d'origine. De fait, chaque idéalisation et chaque approximation introduite dans le modèle ajoute une distance supplémentaire avec les lois initialement considérées comme vraies. Pour autant, ces modèles possèdent des vertus épistémiques, exploratoires autant que de construction, et guident l'enquête scientifique. Par-

59. Ainsi Sheldon Smith voit-il les modèles comme des « recettes » dont les lois de la théorie testée font partie des ingrédients, [Smith \(2002\)](#). La position de Smith est résumée dans [Kistler \(2011\)](#), pp.106–107.

60. [Humphreys \(2004\)](#)

61. [Imbert \(2008\)](#), pp.345–346.

62. [Morgan et Morrison \(1999\)](#). Cette conception des modèles s'écarte de la façon dont ceux-ci sont compris en logique et en mathématiques, où un modèle est une structure qui satisfait les axiomes de la théorie, et qui vient ainsi l'interpréter en la rendant vraie. Comme le résume Varenne, Patrick Suppes défend quant à lui l'idée qu'« il n'y a pas de différence fondamentale entre les divers types de modèles, et que l'on peut donc penser tout modèle scientifique à l'image du modèle tel qu'il intervient en logique et en métamathématique. », [Varenne \(2012\)](#), p.180. Mais selon Achinstein, Redhead, ou encore Morgan et Morrison, la conception logicienne des modèles ne correspond pas à la façon dont ceux-ci sont véritablement utilisés dans l'activité scientifique : « L'une des critiques les plus perspicaces de l'approche sémantique est que celle-ci se méprend quant à la localisation des modèles dans l'édifice scientifique. Les modèles sont relativement indépendants de la théorie, au lieu d'en être constitutifs ; ou, pour utiliser le slogan de Morrison (1998), ils sont des "agents autonomes". Cette indépendance a deux aspects : la construction et le fonctionnement », [Frigg et Hartmann \(2012\)](#), section 4.2.

63. [Achinstein \(1968\)](#)

64. [Redhead \(1980\)](#)

tant de ce double constat – (i) les modèles sont (au moins en partie) indépendants des théories qu'ils mobilisent, et (ii) les résultats de ces modèles sont en contradiction avec ces théories – on en arrive à la conclusion que le test expérimental des théories est toujours indirect et délicat. En effet, on teste les conséquences d'un modèle construit partiellement à partir de la théorie en question. Les idéalizations et les approximations introduites sont à l'origine d'une opacité entre théorie et expérience⁶⁵. Si un expérimentateur enregistre un désaccord entre la période d'oscillation du pendule simple prédite par le modèle et celle qu'il a mesurée expérimentalement, il ne songera vraisemblablement pas avoir réfuté la mécanique classique ; s'il doit envisager une révision à l'échelle théorique, c'est d'abord sur son modèle – et sur la validité des résultats expérimentaux eux-mêmes – qu'il va se pencher.

Partant des constats précédents, Cartwright a développé une approche anti-réaliste et anti-réductionniste des lois de la physique⁶⁶. Elle s'appuie pour cela sur une distinction de nature entre lois dites « phénoménologiques » et lois « fondamentales », qu'il faut au préalable expliciter. Cartwright fait d'abord valoir le décalage considérable entre l'usage des termes « phénoménologique » et « théorique » lorsque ceux-ci sont utilisés par les philosophes ou par des scientifiques. Chez les philosophes, la distinction est relative aux objets sur lesquelles portent les lois, observables ou inobservables, et trouve son importance dans des questionnements liés au réalisme et à l'empirisme⁶⁷. Cartwright adopte pour sa part la conception qui est généralement celle des scientifiques. Dans cette conception, les lois phénoménologiques et fondamentales se distinguent par leur fonction : les premières visent à *décrire* les phénomènes (indépendamment de la question de l'observabilité de ces phénomènes) alors que les secondes visent à *expliquer* les phénomènes. Ainsi, la « loi d'Ohm » de la conduction électrique entre dans la première catégorie, alors que les équations de Maxwell, décisives pour expliquer la loi d'Ohm, entrent dans la seconde⁶⁸.

Puisque les lois phénoménologiques décrivent ce qui a lieu, ce sont elles qui peuvent être confrontées à l'expérience. Les lois fondamentales, quant à elles, ne peuvent être testées qu'indirectement. Bertrand Russell avait déjà fait valoir une idée similaire :

Dans toute science, nous devons distinguer deux espèces de lois : en premier lieu, celles qui sont vérifiables empiriquement, mais qui ne sont probablement qu'ap-

65. Laymon emploie l'analogie suivante : « (c)'est comme si une lentille malformée était interposée entre le scientifique et la réalité », Laymon (1995), p.354.

66. Cartwright (1983)

67. La distinction entre observable et inobservable est à la racine des différentes approches empiristes. C'est pourquoi van Fraassen (1980) s'efforce de la maintenir, pp.13–19.

68. L'approche de Cartwright, que nous reprenons ici à notre compte, présente au moins deux limites notables. D'une part, il serait préférable de parler de *théories* fondamentales ou phénoménologiques, car il n'est jamais possible de prendre en compte les lois de façon individuelles et isolées. D'autre part, la vision de Cartwright est très anhistorique, et ne tient pas compte du fait que les catégories sont elles-mêmes fluides et évoluent avec le temps. Le travail de perfectionnement des théories s'accompagne en général d'une élimination des surplus de contenu qui entraîne des changements dans leurs statuts. Ainsi, les théories fondamentales d'hier sont-elles souvent les théories phénoménologiques d'aujourd'hui. Par exemple, si le modèle standard de la physique des particules a pu être autrefois considéré comme une théorie fondamentale, il est aujourd'hui considéré comme un assemblage à *expliquer* par une théorie plus fondamentale encore. Par la suite, nous considérerons comme loi fondamentale (respectivement phénoménologique) une composante d'une théorie considérée comme fondamentale (respectivement phénoménologique) à un instant donné. Je remercie Olivier Darrigol de m'avoir fait part de ces remarques.

proximatives; en second lieu, celles qui ne sont pas vérifiables, mais qui peuvent être exactes.⁶⁹

Cartwright rejoint Russell sur la possibilité ou non de vérifier empiriquement les lois de la physique, mais elle s'en démarque quant à la question de leur exactitude. En effet, selon Cartwright, ce sont les lois fondamentales qui « mentent » :

En physique moderne [...] les lois phénoménologiques sont vouées à décrire, et souvent y parviennent raisonnablement bien. Mais les équations fondamentales sont vouées à expliquer, et paradoxalement le coût du pouvoir explicatif est l'adéquation descriptive. Les lois au pouvoir explicatif réellement puissantes telles que celles trouvées dans la physique théorique ne disent pas la vérité.⁷⁰

Selon Cartwright, les lois fondamentales d'une théorie sont idéales, et la validité expérimentale des modèles dans lesquelles elles sont impliquées ne doit pas faire croire que ces lois elles-mêmes ont été vérifiées :

Les lois fondamentales de la théorie sont vraies des objets dans le modèle, et elles sont utilisées pour dériver un compte-rendu de la façon dont ces objets se comportent. Mais les objets du modèle ont seulement « la forme et l'apparence des choses » et, en un sens très fort, n'ont pas leur « substance ou leurs qualités propres ». [...] La leçon quant à la vérité des lois fondamentales est claire : *les lois fondamentales ne gouvernent pas les objets dans la réalité; elles gouvernent les objets dans les modèles.*⁷¹

L'un des principaux arguments de Cartwright repose sur son objection contre l'inférence à la meilleure explication. Puisque les lois fondamentales expliquent les lois phénoménologiques, et que ces dernières sont soumises au test empirique, les premières sont considérées comme vraies en vertu du fait qu'elles expliquent des lois considérées comme vérifiées : c'est une inférence à la meilleure explication. Selon Cartwright, cette inférence n'est pas valide⁷². Suivant son raisonnement, les lois phénoménologiques sont directement testables, mais l'idéalisation empêche de remonter jusqu'aux lois fondamentales : l'expérience ne permet pas le test de ces dernières et il n'est pas justifié de les considérer comme vraies. Certaines approximations améliorent même les lois⁷³; or, s'il faut corriger une loi pour obtenir des conclusions valides, il est raisonnable de la tenir pour fautive. La représentation des phénomènes par des modèles élémentaires hautement idéalisés est un acte forcé. Le rôle des lois fondamentales est d'expliquer mais il y a un compromis entre explication et vérité : plus une loi explique, moins elle est vraie⁷⁴.

69. Russell (1912), p.172 (la traduction, proposée par Kistler, est tirée de Kistler (2011).

70. Cartwright (1983), p.3.

71. Cartwright (1983), pp.17-18 (nous soulignons).

72. Cartwright lui oppose l'explication causale. Voir par exemple son analyse des travaux de Perrin, Cartwright (1983), pp.82-85.

73. Cartwright (1983), pp.106-118.

74. Cartwright (1983), p.58. Cartwright prolonge sa critique dans son ouvrage *The Dappled World*, en s'attaquant en particulier au modèle « déductif-nomologique », dû à Hempel et Oppenheim : « Il s'agit d'une conception qui sert notre croyance en un seul grand système, un système composé d'un ensemble restreint de premiers principes bien coordonnés, admettant une formulation simple et élégante, à partir desquels tout ce qui a lieu [...] peut être dérivé. [...] Mais le traitement des systèmes réels ne sont pas déductifs, ni approximativement déductifs, ni déductifs

En réponse à cette position anti-réaliste vis-à-vis des lois, McMullin et Laymon ont développé indépendamment et successivement deux réponses qui se rejoignent sur un point majeur : le statut de vérité des lois physiques impliquées dans des modèles idéalisés est tributaire de la possibilité de *perfectionner* lesdits modèles. Ces réponses mettent l'accent sur le rôle de l'approximation dans la problématique développée. Dans son analyse épistémologique de l'idéalisation chez Galilée, McMullin rappelle qu'une prémisse approximative, bien qu'inexacte et donc, en toute rigueur, fausse, se démarque d'une prémisse rigoureusement fausse par le fait qu'on peut en estimer le degré d'approximation, et s'assurer que cette dernière ne s'éloigne pas déraisonnablement de l'idée de départ⁷⁵. De ce fait, les modèles peuvent ou non présenter une possibilité de perfectionnement selon deux axes remarquables. (i) Il est parfois possible de les dé-idéaliser progressivement, c'est-à-dire de réduire la quantité d'idéalisation qu'il contiennent, et de constater qu'il prédisent ensuite *mieux* les phénomènes observés. (ii) Il est possible de les élargir, de les étendre à des domaines qu'ils ne couvraient pas initialement. McMullin prend l'exemple du spin de l'électron, qui n'était pas contenu dans les premiers modèles quantiques de l'électron, mais a pu y être intégré sans que cela ne rende nécessaire de changer de modèle⁷⁶. McMullin fait remarquer que les deux propositions (i) et (ii) ne sont pas possibles pour des modèles qui ne sont pas déjà approximativement valables dès le départ, et qui ne contiennent pas en eux-mêmes des germes de leur perfectionnement. Ce faisant, il s'oppose à la position anti-réaliste de Cartwright :

Prendre au sérieux le modèle, en tant qu'une description approximativement vraie, est ce qui nous mène à *nous attendre* que la correction produise une prédiction vérifiable. Le fait que l'idéalisation formelle fonctionne *effectivement* en ce sens de façon plutôt cohérente constitue un argument puissant en faveur d'une version modérée du réalisme scientifique.⁷⁷

Ronald Laymon adopte une position similaire. Il fait remarquer que si l'on reprend l'idée de Cartwright selon laquelle ce sont les lois phénoménologiques qui sont directement confrontées à l'expérience, celles-ci ne peuvent pas pour autant être dites « vraies » : elles ne sont en fait qu'*approximativement* vraies⁷⁸. Dès lors, il n'y a rien de surprenant à ce que, pour dériver des lois phénoménologiques à partir de lois fondamentales, il soit nécessaire de passer par des idéalizations et des approximations ; et cela n'est pas incompatible avec l'assertion de Russell selon laquelle les lois fondamentales peuvent être exactes. D'où la position de Laymon :

Les arguments [de Cartwright] [...] ne prennent pas en compte la pertinence de l'amélioration graduelle des idéalizations et des approximations, et donc des améliorations correspondantes des prédictions opérées.⁷⁹

Laymon reconnaît toutefois qu'il est effectivement impossible de tester directement les lois fondamentales. La validité empirique d'un modèle est dans une certaine mesure indépendante du

avec des corrections, ni approchant de façon plausible toujours plus près du caractère déductif au fur et à mesure que nos théories progressent. », Cartwright (1999), p.9, traduit par Imbert (2008), p.346.

75. McMullin (1985), p.256.

76. McMullin (1985), pp.263-264.

77. McMullin (1985), p.262 (McMullin souligne).

78. Laymon (1989), p.354.

79. Laymon (1989), p.353.

statut de vérité des lois que ce modèle met à contribution, puisque l'accord empirique peut être induit par des ajustements à différents niveaux du modèle : l'idéalisation introduit effectivement une opacité dans le test des théories. Par conséquent, il n'est pas possible de soutenir une proposition telle que « si les lois employées dans le modèle sont vraies, alors les conclusions du modèle seront empiriquement valides » ; et encore moins, réciproquement : « si les conclusions du modèle seront empiriquement valides, alors les lois employées dans le modèle sont vraies ». C'est pourquoi Laymon envisage une solution qui consiste à contourner ce problème en testant les lois fondamentales de façon indirecte :

Si un ensemble de lois fondamentales est vrai, alors nous pouvons en principe faire des corrections suffisantes afin d'obtenir de meilleures prédictions.⁸⁰

La force de cet argument réside dans le fait que Laymon parvient à réintroduire une clause relative au statut de *vérité* des lois fondamentales dans la confrontation entre modèle et expérience, en dépit de « l'omniprésence de l'idéalisation ». Laymon développe ainsi deux « principes de confirmation » :

(C1) Un ensemble de lois fondamentales reçoit confirmation si l'utilisation de spécifications plus réalistes des conditions initiales ou aux limites conduit de fait à des prédictions plus exactes. [...] (C2) Un ensemble de lois fondamentales reçoit infirmation si l'utilisation de spécifications plus réalistes des conditions initiales ou aux limites ne conduit pas à des prédictions plus exactes.⁸¹

Laymon prend pour exemple le rapport entretenu entre les « lois de Kepler » et les lois de Newton du mouvement⁸². Les lois de Kepler décrivent les propriétés principales du mouvement des planètes autour du Soleil (et par extension, du mouvement d'un corps soumis à une unique force centrale dont l'intensité décroît en raison inverse du carré de la distance séparant le corps du centre attracteur). Ces trois lois peuvent être testées à partir d'observations astronomiques des trajectoires des planètes dans le système solaire. Elles correspondent à ce que Cartwright désigne en tant que « lois phénoménologiques ». Les lois de Newton du mouvement décrivent la dynamique d'un corps soumis à des forces extérieures dans un référentiel donné, et établissent en particulier l'égalité (vectorielle) entre ces forces et le produit masse fois accélération du corps. Elles permettent donc, une fois recensées les forces auxquelles est soumis un corps matériel, de prédire la trajectoire de ce corps. Dans le cadre de la mécanique newtonienne, les lois de Newton sont des lois fondamentales. En principe, les lois de Kepler peuvent être déduites des lois de Newton ; c'est d'ailleurs l'un des résultats les plus classiques des manuels de physique d'université. Mais les lois de Newton ne sont pas mobilisées directement ; elles le sont dans un modèle qui ne met en jeu que deux corps massifs, en interaction gravitationnelle l'un avec l'autre, en l'absence de toute autre force extérieure : c'est le « problème à deux corps ». Lorsque le modèle fait intervenir plus de deux corps en interaction gravitationnelle – problème « à N corps » – les conclusions deviennent différentes, et entrent en contradiction avec les lois de Kepler. En particulier, les trajectoires ne sont plus des courbes fermées (la dynamique est

80. Laymon (1989), p.359.

81. Laymon (1989), pp.359–360.

82. Laymon (1989), p.364. Duhem avait déjà abordé cet exemple dans Duhem (1906) pour critiquer en particulier l'épistémologie inductiviste ; voir Brenner (1990), p.32. Les remarques de Duhem avaient également inspiré à Feyerabend son « principe de contre-induction » (Feyerabend, 1975).

en fait chaotique), et ne peuvent donc pas être des ellipses. Il est clair que le problème « à N corps » décrit le système solaire de façon plus fine que ne le fait le problème à deux corps : la Terre, par exemple, est soumise à l'attraction du Soleil, mais également de la Lune et des autres planètes du système solaire (principalement Jupiter). Pour déduire les lois de Kepler des lois de Newton, il faut donc passer par une idéalisation qui, de toute évidence, nous éloigne de la réalité. Or, défend Laymon, cela n'est pas un problème puisque les lois de Kepler ne sont qu'approximativement vraies ; les lois de Newton, quant à elles, continuent à être utiles pour les problèmes à plus de deux corps (bien que la complexité de ces problèmes soit notoire) et permettent justement d'indiquer comment perfectionner les lois de Kepler, ou de délimiter le domaine dans lequel ces dernières sont approximativement valides.

Les principes défendus par McMullin et Laymon vont dans le sens d'une défense d'une forme de réalisme scientifique, le *réalisme convergent*, selon lequel l'activité scientifique mène progressivement à formuler des propositions, lois et théories, qui ne sont jamais exactes, mais tendent en général à s'en rapprocher de plus en plus. Une grande part des vertus épistémiques des modèles est de donner une place à l'approximation et à l'idéalisation dans le but de permettre le progrès scientifique dans en donnant une direction qui est celle de la vérité des théories scientifiques⁸³. Box et Draper ont exprimé à leur façon cette caractéristique des modèles, par une formule devenue consacrée : « fondamentalement, tous les modèles sont faux, mais certains sont utiles ; la question pratique est de savoir à quel point ils doivent être faux pour ne pas être utiles »⁸⁴ ; cela renvoie également à un aspect pragmatique : les modèles ont un caractère local et provisoire, adapté à un *usage* précis.

Que pouvons-nous retenir de cette analyse pour notre cas d'étude spécifique ? Elle rappelle d'abord que l'idée d'une valeur vraie *unique* d'une grandeur n'a de sens qu'à l'intérieur d'un modèle. Il est illusoire de penser que des théories imparfaites, exploitées dans des modèles idéalisés de phénomènes contenant leur lot d'approximations, puissent définir des grandeurs physiques qui seraient quant à elles parfaitement exactes. Il nous semble raisonnable de penser que la méfiance des métrologues à l'encontre de la teneur trop métaphysique de certains termes classiques du vocabulaire qu'ils utilisent – dont celui de « valeur vraie » – exprime justement leur volonté de fuir cette *illusion* d'exacitude. Pour autant, la solution qui consiste à évacuer purement et simplement le concept de valeur vraie n'est peut-être pas non plus satisfaisante. De façon cruciale, les lois phénoménologiques dans lesquelles sont mises à contribution les grandeurs de la même catégorie sont des lois *approchées* – elles sont exprimées à l'intérieur de modèles de phénomènes qui, encore une fois, incorporent leur lot d'idéalisations et d'approximations. Ce sont les modèles qui mettent en jeu des relations entre valeurs numériques bien définies – mais cela ne nécessite pas que les grandeurs correspondantes soient caractérisées par des valeurs vraies uniques. Il n'y a aucun sens à chercher à caractériser une grandeur par une valeur infiniment précise alors que ces grandeurs interviennent dans des équations qui ne sont par nature *jamais* exactes. Suivant cette idée, l'incertitude définitionnelle apparaît à l'échelle phénoménologique. Dans ce cadre, l'idée d'une multiplicité de valeurs « vraies »

83. C'est ainsi ce qu'exprime Wimsatt par son expression « de faux modèles comme moyen vers des théories plus vraies », Wimsatt (1987).

84. Box et Draper (1987), p.424.

d'une grandeur physique est acceptable, à condition de reconnaître que ces valeurs ne soient qu'*approximativement* vraies. Cela nécessite d'avoir à disposition un concept satisfaisant de « vérité approchée »⁸⁵.

7.3.3 Incertitude définitionnelle et objectifs de la mesure

Si la notion de valeur vraie reste épistémologiquement valide, cela ne résout pas pour autant les problèmes qui se posent à l'expérimentateur dans sa *pratique* scientifique. Comment celui-ci doit-il prendre en compte l'incertitude définitionnelle ?

Notre discussion montre que la problématique de l'incertitude définitionnelle doit être comprise en relation avec un impératif pragmatique touchant aux *objectifs* de l'expérimentateur. C'est ce que soulignent Giordani et Mari :

Bien sûr, une telle idéalisation n'est pas imposée par la modélisatrice, qui, au moyen de ces modèles, décide en fait des concepts (ou des objets soumis à la mesure et du mesurande) qu'elle considère appropriés en fonction de ses objectifs.⁸⁶

Rappelons que la notion même de mesurande est attachée à une *intention*. Nous avons vu à différentes reprises, et de nouveau à la section 7.1.1 que la mesure d'une grandeur est généralement attachée à une entreprise de correction, au cours de laquelle l'expérimentateur relie la « grandeur réalisée », attachée aux conditions concrètes de l'expérience, et le mesurande, qui contient une part d'idéalisation. Tal a décrit ce cheminement comme une « dé-idéalisation » progressive de la grandeur définie idéalement vers la grandeur réalisée⁸⁷ ; chaque composante d'incertitude associée à une correction est alors le résultat d'une opération qui vise à « reconstruire » la grandeur réalisée. L'incertitude définitionnelle est ici encore singulière, en ce qu'elle procède du mouvement inverse. Elle correspond à la quantité d'idéalisation qui n'a *pas* été introduite dans la définition du mesurande, mais qui aurait été nécessaire, au vu des effets physiques connus affectant l'objet soumis à la mesure, si les objectifs de l'expérimentateur avaient

85. Barberousse rappelle que l'épistémologie de l'approximation et la réflexion sur une notion de vérité approchée étaient déjà très présentes, il y a un siècle, chez des philosophes comme Pierre Duhem, Émile Borel ou encore Hans Reichenbach (Barberousse, 2008, p.54). Chakravartty donne un aperçu des différentes tentatives d'explicitation formelle de la vérité approchée à partir du milieu du XX^e siècle (Chakravartty, 2015, section 3.4). Il recense ainsi l'introduction par Karl Popper de la « vérisimilitude » en 1972. Vient également l'approche des « mondes possibles » proposée par Pavel Tichý dès 1976, Graham Oddie ou encore Ilkka Niiniluoto. Chakravartty présente également l'approche « hiérarchique », avancée par Jerrold Aronson, Rom Harré et Eileen Cornell Way dans les années 1990. Ces différentes positions sont également exposées dans Oddie (2016) (dernière consultation le 8 février 2016). Richard Boyd voit également dans le réalisme structural de John Worrall une autre approche possible du problème de l'approximation (Boyd, 2010, section 3.5, dernière consultation le 8 février 2016). Boyd fait par ailleurs remarquer que les notions d'approximation et de vérité approchée sont souvent essentielles dans la défense du réalisme (Boyd, 1990). De fait, une conception trop rigide de la vérité des théories prête le flanc à certaines critiques du réalisme scientifique, comme par exemple l'argument de la « méta-induction pessimiste » qui énonce que puisque la plupart des théories passées ont été réfutées un jour, il n'y a, par induction, aucune raison de croire dans la vérité des théories présentes, même si celles-ci paraissent extrêmement solides à l'instant présent. Dans Laudan (1981), Larry Laudan a présenté une forme de cet argument très classique (que l'on trouvait déjà chez Montaigne) où il attaque le réalisme convergent et, à travers lui, l'idée de vérité approximée. En réponse, différencier les théories fausses des théories seulement approximativement vraies mène à sérieusement nuancer la méta-induction pessimiste.

86. Giordani et Mari (2012), p.2148

87. Tal (2011), pp.1088-1090.

requis une définition plus fine. Dans ce cas, l'amplitude de l'incertitude définitionnelle associée à un mesurande donnée est le produit d'un choix de modélisation effectué par l'expérimentateur. Par exemple, dans de nombreuses situations ordinaires, la structure microscopique des objets mesurés ne présente aucun intérêt pour un usager, qui souhaite se placer à l'échelle la plus appropriée pour ses besoins.

Les métrologues sont largement conscients du rapport qui est entretenu entre la précision des mesures et les *objectifs* d'une enquête expérimentale, étant donné que la métrologie est une discipline devant assurer la cohésion d'un spectre large de domaines fondamentaux (recherche fondamentale) et appliqué (industrie, ingénierie, métrologie légale, etc.) et est donc en permanence confrontée à des questionnements de type pragmatique. Ainsi est-il fait part de cet aspect chez Eisenhart⁸⁸ et, de façon notable, dans le GUM :

Dans la pratique, la spécification ou la définition exigée pour le mesurande est dictée par l'exactitude de mesure exigée pour le mesurage. Le mesurande doit être défini de façon suffisamment complète en rapport avec l'exactitude exigée de sorte que sa valeur soit unique pour tous les objectifs pratiques associés au mesurage. C'est dans ce sens qu'on utilise l'expression « valeur du mesurande » dans ce *Guide*.

EXEMPLE Si l'on doit déterminer la longueur nominale d'une barre d'acier de longueur un mètre au micromètre près, sa spécification doit comprendre la température et la pression auxquelles la longueur est définie. Le mesurande peut alors être spécifié comme, par exemple, la longueur de la barre à 25,00 °C et 101 325 Pa (avec, en plus, tout autre paramètre de définition jugé nécessaire, tel que la manière de supporter la barre). Cependant, si l'on ne doit déterminer la longueur de la barre qu'au millimètre près, sa spécification ne nécessitera pas la définition d'une température, ou d'une pression, ou de tout autre paramètre.⁸⁹

Ce que le GUM appelle « exactitude de mesure exigée pour le mesurage » peut être retrouvé sous une forme proche dans le VIM3, où est définie l'« incertitude cible » : « incertitude de mesure spécifiée comme une limite supérieure et choisie d'après les usages prévus des résultats de mesure »⁹⁰.

Nous avons défendu l'idée que l'incertitude définitionnelle ne doit pas être comprise comme une composante de l'incertitude totale ; elle mesure plutôt l'incomplétude de la définition du mesurande que nos connaissances nous permettent d'identifier (et donc de l'étendue des valeurs compatibles qui s'en suit). Dans la continuité de cette position, les remarques précédentes nous permettent de mieux comprendre le rôle de l'incertitude définitionnelle. En confrontant l'in-

88. « Dans l'analyse finale, la "valeur vraie" de la grandeur d'une quantité est définie par un accord entre experts sur une *méthode modèle* pour la mesure de son amplitude – c'est la moyenne limite d'un *processus modèle* conceptuel qui est une réalisation idéale de la méthode modèle convenue. Et le raffinement dont il faut faire preuve pour spécifier le processus modèle dépendra des raisons pour lesquelles une détermination de la grandeur de la quantité concernée est nécessaire – pas simplement les raisons immédiates pour lesquelles les mesures doivent être prises, mais aussi les autres usages que ces mesures, ou les valeurs finales ajustées qui en sont dérivées, peuvent servir », Eisenhart (1963), p.170.

89. Joint Committee for Guides in Metrology (JCGM) (2008d), p.4.

90. Joint Committee for Guides in Metrology (JCGM) (2012), p.27.

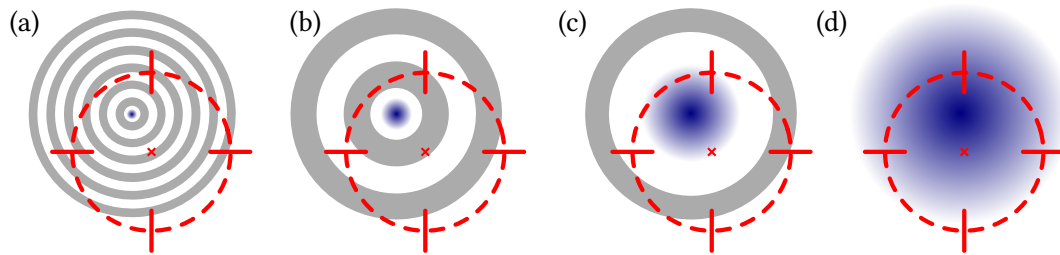


FIGURE 7.3 – Interprétation de l’incertitude définitionnelle à l’aide de la métaphore des cibles. Les quatre situations (a), (b), (c) et (d) schématisent la mesure d’une même grandeur avec une incertitude de plus en plus faible. La valeur vraie de la grandeur visée est représentée par le disque au centre de la cible (en bleu et en dégradé). Le dégradé indique qu’en raison de l’incertitude définitionnelle, la grandeur visée est floue; l’incertitude définitionnelle elle-même correspond au rayon du disque central. Le résultat de mesure est représenté par la valeur centrale (croix, en rouge) et l’incertitude de mesure associée (cercle en pointillés, en rouge). Dans le cas (a), l’incertitude de mesure est grande devant l’incertitude définitionnelle, et le disque central apparaît presque comme un point : c’est le cas de travail du GUM, où la grandeur peut être caractérisée par une « valeur vraie essentiellement unique ». La question de l’incertitude définitionnelle peut être écartée et l’on revient à une situation classique. L’incertitude diminuant de (a) à (d), on peut zoomer peu à peu sur la cible. La grandeur visée apparaît alors de plus en plus floue au fur et à mesure que l’incertitude de mesure devient comparable à l’incertitude définitionnelle. Dans le cas (d), on atteint les limites du modèle : une incertitude de mesure plus faible encore ne correspondrait à rien de tangible.

certitude définitionnelle avec l’incertitude de mesure déterminée par les méthodes statistiques usuelles, on peut vérifier si le modèle employé est adéquat ou non en regard des objectifs de l’enquête expérimentale, caractérisés en particulier par l’« incertitude cible ». Si l’incertitude définitionnelle est du même ordre de grandeur que l’incertitude cible, c’est qu’il est vraisemblablement nécessaire d’envisager d’affiner le modèle employé. En revanche, si l’incertitude définitionnelle est très petite devant l’incertitude cible, c’est que le modèle est satisfaisant du point de vue de la précision des définitions des grandeurs qu’il mobilise. L’incertitude définitionnelle n’entre donc pas dans une relation de combinaison mais dans une relation de *comparaison* avec l’incertitude de mesure. Le second cas correspond au domaine d’application du GUM, où la grandeur peut être caractérisée par une « valeur vraie essentiellement unique ». Dans ce cadre de travail, les problèmes liés à la multiplicité des valeurs du mesurande peuvent être ignorés : ce sont des problèmes essentiellement conceptuels sans conséquence pratique. Notons toutefois que lorsque les mesures effectuées gagnent en précision, l’incertitude cible diminue, en conséquence de quoi l’incertitude définitionnelle devient de moins en moins négligeable, et les problèmes liés à la non-unicité de la valeur vraie sont alors peu à peu susceptibles de refaire surface. Cette interprétation de l’incertitude définitionnelle est représentée à la figure 7.3 à l’aide de la métaphore des cibles.

Si l'on admet la position défendue ici, il faut toutefois reconnaître qu'elle laisse ouverte une question délicate. En effet, si l'incertitude définitionnelle entre dans une relation de comparaison avec l'incertitude de mesure, c'est qu'on doit la quantifier, et donc que la question de l'évaluation quantitative de l'incertitude définitionnelle persiste. La réponse que nous suggérons consiste à affirmer qu'il est possible de rendre compte de l'incertitude définitionnelle sur un mode quantitatif, mais qu'il n'est pas nécessaire de chercher à la déterminer par un nombre précis qui aurait une signification univoque. De fait, nos connaissances théoriques et expérimentales permettent de nous essayer à prédire les conséquences qu'aura l'incomplétude de la définition d'un mesurande sur l'étendue des valeurs vraies possibles de ce mesurande. Si l'on prédit par exemple qu'une barre de métal peut voir sa longueur varier d'environ un centimètre en raison de fluctuations de sa température entre certaines valeurs limites, alors l'incertitude définitionnelle sur la longueur de la barre, associée à une absence de spécification de la température à laquelle la longueur de la barre est définie, est *de l'ordre* du centimètre. Pour autant, il ne sera pas possible de résumer l'incertitude définitionnelle par une unique valeur, sous la forme d'une incertitude-type par exemple. Mais cela n'est pas nécessaire : ce qui importe, c'est de pouvoir donner une idée du domaine dans lequel se trouve l'incertitude définitionnelle, de façon à pouvoir la comparer à l'incertitude-cible, mais pas d'établir un rapport précis entre les valeurs des deux incertitudes. Tant que l'incertitude définitionnelle est négligeable par rapport à l'incertitude cible, il n'est pas nécessaire de déterminer une composante quantitative d'incertitude définitionnelle selon une procédure éprouvée et rigoureuse qui permette de l'intégrer à un bilan complet d'incertitude. Si, à l'inverse, l'incertitude définitionnelle n'est pas négligeable par rapport à l'incertitude cible, nous défendons qu'il est souhaitable d'améliorer le modèle de façon à la réduire et à la rendre de nouveau négligeable – dès lors, il n'y a pas non plus d'intérêt à chercher à en déterminer une valeur précise. Notre conception est donc résumé par l'idée qu'une incertitude définitionnelle non négligeable devrait inciter à construire de meilleurs modèles⁹¹.

91. Ayant essentiellement fondé notre réflexion sur des problématiques en lien avec la physique, nous ne pouvons écarter la possibilité que l'analyse d'autres disciplines révèle la nécessité d'adopter une position plus souple à ce sujet. L'on pourrait envisager que, dans certaines situations, l'incertitude définitionnelle soit structurellement appelée à être significative et ne puisse jamais être rendue négligeable par rapport à l'incertitude cible. Le GUM mentionne une telle possibilité : « (b)ien qu'il faille définir un mesurande suffisamment en détail pour que toute incertitude provenant des lacunes de sa définition soit négligeable par rapport à l'exactitude requise pour le mesurage, on doit reconnaître que cela ne peut pas toujours se faire. », *Joint Committee for Guides in Metrology (JCGM) (2008d)*, p.52. Depuis la fin des années 2000, le JCGM s'est activement attelé à songer à la rédaction de la prochaine édition du GUM. Les considérations relatives au cas où l'incertitude définitionnelle n'est pas négligeable font partie des discussions ouvertes. En 2012, l'organisme a mené une enquête auprès de nombreux laboratoires, à l'échelle internationale, afin de s'assurer que le nouveau GUM soit conçu en regard des besoins de ses utilisateurs. L'enquête prend la forme d'une liste de questions relatives au GUM, ses principes, son usage, sa forme, et ses objectifs. L'une des questions fait indirectement référence à la problématique de l'incertitude définitionnelle, et laisse entendre que lorsqu'on sort du domaine de la physique, et que l'on s'intéresse par exemple à des disciplines comme la chimie, la biologie ou la médecine, les problèmes liés à l'incertitude définitionnelle pourraient être susceptibles de réapparaître, car il ne serait pas toujours possible de rendre l'incertitude définitionnelle négligeable (Bich, 2012b); ce rapport, publié sur le site web du BIPM, n'a pas de numéro de page. Les réponses des laboratoires à ce questionnaire ont été compilées et rendues publiques par le JCGM à l'adresse suivante : http://www.bipm.org/wg/JCGM/JCGM-WG1/Allowed/sub-committee_5/WG1-SC5-N12-15_JCGM_GUM_Survey_Collated_responses.pdf. Les deux documents ont été consultés le 21 octobre 2015). Vosk et Emery proposent un cas d'étude issu de la métrologie légale, en

L'idée que l'incertitude définitionnelle constitue une limite inférieure de l'incertitude totale de mesure reste intuitivement acceptable, si l'on remarque que réduire l'incertitude de mesure en-deçà de cette limite n'apporte *aucune* connaissance supplémentaire du système – il serait absurde de prétendre atteindre une précision plus importante que celle que le niveau de finesse du modèle déployé permet d'envisager en premier lieu. Certaines grandeurs physiques n'ont de sens qu'à une certaine échelle, et leur valeur vraie n'a également de signification qu'à cette échelle. De ce point de vue, l'incertitude définitionnelle ne décrit pas tant une limite ultime de la précision des mesures que l'expérimentateur peut effectuer, mais plutôt la limite effective de précision de ses modèles. Nous interprétons en fin de compte l'incertitude définitionnelle comme la résolution avec laquelle les lois phénoménologiques sont supposées être approximativement vraies.

7.4 Synthèse

Ce chapitre s'est attaché à montrer les origines d'une critique spécifique de la valeur vraie, qui trouve son origine dans un constat simple : si l'on parle de valeur « vraie » d'une grandeur, celle-ci devrait être unique pour une grandeur donnée. Or, plusieurs exemples suffisent à se convaincre que, dans de très nombreux cas, cela n'est pas concevable. Un tel aperçu peut être généralisé en revenant sur le rôle fondamental de la notion de mesurande, et en particulier sur son caractère *intentionnel*. La non-unicité de la valeur vraie d'un mesurande trouve une origine physique dans les propriétés de l'objet considéré – variabilité de certaines grandeurs, dépendance envers un paramètre extérieur, flou intrinsèque de l'objet, par exemple lié à des aspects quantiques – et dépend de la façon dont le mesurande est défini, et, en particulier, du niveau de détail avec lequel sa définition est formulée. Chaque lacune définitionnelle est susceptible d'élargir le spectre de valeurs vraies compatibles avec la définition du mesurande – en fonction des propriétés physiques de l'objet étudié.

Or, si l'on cherche à préciser la cible en ajoutant des spécifications à la définition du mesurande, on se retrouve généralement confronté à l'impossibilité de franchir certaines limites conceptuelles à moins de transformer substantiellement la façon dont on conçoit les objets sur lesquels on travaille, chose qui mène à perdre de vue la cible initiale. C'est pourquoi nous aboutissons à l'idée que la présence d'un flou dans la définition d'un mesurande n'est pas simplement le signe que l'expérimentateur a péché par manque d'efforts ou de rigueur, mais plutôt que ce flou est nécessaire afin de pouvoir trouver la perspective la plus adaptée au problème que l'on se pose. De fait, la définition de la longueur d'un objet ne nécessite pas de prendre en compte les mêmes phénomènes d'influence selon que sa mesure vise à tracer les limites d'un terrain de sport (afin de respecter les dimensions réglementaires fixés par les instances offi-

étudiant un exemple de mesure du taux d'alcoolémie par une analyse de l'haleine, Vosk et Emery (2014), pp.200–203. L'absence de littérature détaillée sur la question rend très difficile de percevoir la portée que peuvent avoir ces différentes remarques. La thèse que nous défendons ne va pas dans le même sens, mais il nous semble important d'évoquer ces perspectives qui pourraient nous amener à devoir perfectionner au moins en partie notre positionnement, car nous considérons encore et toujours que la pratique des scientifiques et des métrologues ne saurait être légiférée *a priori* par des considérations d'ordre philosophique.

cielles) ou à construire une pièce essentielle d'un moteur de fusée visant à placer un satellite en orbite. Ainsi, la question de la définition d'un mesurande est relative à un but, et les problèmes liés à la non-unicité des valeurs d'une grandeur découlent des types d'objets étudiés et des choix de représentation d'un expérimentateur en fonction de ses besoins – en articulation avec les limites physiques qui sont la conséquence des propriétés de l'objet lui-même (il est illusoire d'espérer définir un mesurande de façon plus précise que ce que les limites quantiques autorisent à faire). Adopter une telle perspective pragmatique ne signifie pas que l'on doive abandonner toute prétention à décrire le « réel », au moyen de propositions qui portent une valeur de vérité. Cependant, nous défendons l'idée qu'il n'y a, dans une certaine mesure, aucune possibilité d'aborder la notion de vérité à l'échelle phénoménologique sans faire appel à celle de « vérité approchée ». La valeur « vraie » d'une grandeur phénoménologique est ainsi sa valeur « approximativement vraie », et elle intervient dans des équations physiques elles-mêmes considérées comme seulement approximativement vraies. Celle d'une grandeur fondamentale, comme la constante de Planck, peut être *considérée comme* unique car elle intervient dans des équations décrivant des lois physiques *considérées comme* exactes. La non-unicité de la valeur vraie d'une grandeur ne traduit pas l'idée qu'il n'y a plus rien de tel que la « vérité » dans les valeurs des grandeurs physiques. Seulement, l'idée est que, *même si* les résultats de mesures portent une certaine valeur de vérité, celle-ci ne *peut pas* forcément prendre la forme qu'on souhaiterait lui donner – à savoir une valeur numérique unique.

Dans le GUM, le problème de la non-unicité de la valeur vraie est traité par l'introduction d'un concept nouveau en métrologie, celui d'incertitude « définitionnelle ». Nous interprétons l'appellation d'incertitude comme le témoignage du caractère épistémique du concept, tout en nuancant immédiatement la portée de ce terme : l'incertitude définitionnelle n'est pas une composante quantitative de l'incertitude de mesure. Elle exprime une incomplétude de la définition du mesurande relativement à nos connaissances, au sein d'un modèle donné, et permet d'évaluer si un modèle expérimental est adéquat en regard des objectifs fixés – si l'incertitude définitionnelle est trop élevée, c'est qu'il faut affiner le modèle employé.

Notre appel à la notion de vérité approchée d'une part, et notre conception des équations fondamentales et phénoménologiques de la physique d'autre part, nécessitent l'adhésion à une forme de réalisme scientifique. Pour autant, nous estimons qu'il n'est aucunement question de chercher ici à *démontrer* une position philosophique comme le réalisme scientifique – ou, inversement, un anti-réalisme. En revanche, il nous semble établi que le constat de non-unicité de la valeur vraie *n'est pas* incompatible avec une conception réaliste de l'activité scientifique – tout comme il peut être également compris dans une perspective anti-réaliste. Dans tous les cas, notre constat nous amène à conclure que, pour continuer à comprendre la notion de façon cohérente, il est nécessaire de fournir une explication de ce que peut être une vérité « approchée ».

Conclusions de la seconde partie

Les deux arguments explorés dans cette partie, l'inconnaissabilité de la valeur vraie d'une part, sa non-unicité d'autre part, sont assez différents par leur nature, même s'ils ont en commun de rappeler tous deux l'idéalité du concept, et même si les réponses apportées font toutes deux appel à la notion de valeur vraie comme cible de la mesure. Nous défendons l'idée qu'aucun des ces deux arguments ne mène véritablement à abandonner le concept de valeur vraie. L'argument d'inconnaissabilité se traduit principalement par le fait qu'il n'est jamais possible de garantir l'exactitude d'un résultat de mesure. Pour autant, l'intérêt du concept de valeur vraie tient à son caractère normatif, indiquant que la mesure est tournée vers une cible, ce qui laisse ouverte la quête perpétuelle d'identification et de correction des erreurs de mesure dans le but de perfectionner notre connaissance imparfaite de la valeur vraie de la grandeur visée. L'argument de non-unicité rappelle que la valeur d'une grandeur s'exprime à l'intérieur d'un modèle, et il amène à réfléchir sur le caractère approché des connaissances expérimentales et des modèles théoriques dans lesquels les grandeurs mesurées s'insèrent. L'identification de la cible de la mesure est du ressort de l'expérimentateur et se rapporte à un objectif bien identifié qui détermine le niveau de précision nécessaire dans la définition d'un mesurande.

Les réponses apportées aux questions développées dans cette partie nous suggèrent qu'il n'est peut-être pas nécessaire d'embrasser trop vite la perspective épistémique mise en avant par les métrologues bayésiens. Toutefois, les arguments que ces derniers font valoir restent en partie valables, et l'argument d'inconnaissabilité demeure un problème dès lors que l'on cherche à donner la signification d'une incertitude de mesure. Puisqu'on ne peut pas garantir l'exactitude d'une mesure, c'est qu'un constat d'incertitude, suivant l'interprétation bayésienne, est d'abord un état de croyance de l'expérimentateur. Dans quelle mesure pourrait-on alors prétendre que cette incertitude est une évaluation de l'erreur de mesure commise? À quelles conditions et en quelles circonstances? Cela reflète une tension entre *incertitude de mesure* et *exactitude de mesure*, que nous avons déjà exhibée à la partie précédente, et sur laquelle nous reviendrons dans la prochaine partie, en étudiant les ajustements des constantes de la physique. Ceux-ci fournissent une illustration de la façon dont les physiciens conçoivent l'usage de l'incertitude de mesure et intègrent les différents questionnements que les métrologues ont soulevé durant la seconde moitié du XX^e siècle et qui ont fait l'objet des deux premières parties de cette thèse. Cela nous permettra de revenir plus en détail sur ce que nous avons évoqué ici de façon abstraite, à savoir l'identification de la traque des erreurs de mesure comme l'une des matrices du progrès scientifique permettant la production d'une connaissance *positive*. Cette

analyse fait voir la question de la valeur vraie dans une perspective différente, celle d'une activité collective qui croise les évaluations de différents scientifiques.

Troisième partie

Les ajustements de constantes de la physique : une perspective différente sur le rapport entre incertitude et exactitude

Vue d'ensemble de la partie III

Dans la partie I, nous avons vu en quoi la transition d'une approche fréquentiste traditionnelle de la mesure à une approche bayésienne participe d'un « tournant épistémique » qui caractérise à la fois un changement dans les modes de représentation et dans les objectifs de la mesure. Les probabilités bayésiennes visent à décrire non pas les phénomènes physiques eux-mêmes mais la connaissance que l'on peut en avoir. La discussion révèle ainsi qu'il est nécessaire de marquer une distinction entre « exactitude » de mesure, qui traduit une distance avec la vérité, et l'« incertitude » de mesure, qui est la largeur des valeurs attribués à une grandeur suite à une mesure donnée, et que les bayésiens interprètent de façon épistémique, c'est-à-dire comme la mesure d'un état de connaissance. En particulier, l'exactitude d'une mesure ne peut jamais être garantie, puisqu'il faudrait pour cela connaître la valeur vraie de la grandeur mesurée, que l'approche traditionnelle voit justement comme la visée inconnue de la mesure. Nombre de métrologues n'ont pas manqué de souligner cette circularité de l'approche traditionnelle et suggèrent que le problème provient du concept de « valeur vraie » d'une grandeur lui-même. C'est le fondement de l'argument d'inconnaissabilité, décrit dans la partie II, et auquel l'approche traditionnelle se heurte invariablement : puisque la valeur vraie d'une grandeur est inconnaissable, il n'y a pas sens à évaluer la mesure en regard de cette valeur vraie. Si l'on ajoute que la valeur vraie d'une grandeur ne peut pas être unique, l'on est amené à la conclusion de certains métrologues que la finalité d'une mesure n'est plus de déterminer une valeur vraie de la façon la plus exacte possible mais d'exprimer un état de connaissance de la façon la plus raisonnable possible. Un résultat de mesure est alors l'attribution à une grandeur physique de valeurs raisonnables en adéquation avec un objectif donné. Le « tournant épistémique » se caractérise alors par l'abandon de la vérité comme norme régulatrice de la mesure, au profit d'une conception qui met l'accent sur la *description* des états de connaissance⁹².

Nous avons défendu dans les deux parties précédentes que la solution épistémique au problème de l'inconnaissabilité de la valeur vraie n'est pas pour autant pleinement satisfaisante, en particulier car elle met en retrait une notion essentielle de la perception du progrès expérimental, celle d'erreur de mesure. Nous défendons ainsi que la notion de valeur vraie vise à engager les expérimentateurs dans un processus permanent de correction.

Dans cette partie, nous souhaitons ajouter une dimension supplémentaire à cette défense,

92. Cela ne signifie pas que l'approche épistémique – et *a fortiori* la conception bayésienne de la mesure – n'admet aucune norme régulatrice : la rationalité, en particulier, reste au cœur du processus d'élicitation des croyances et des connaissances.

en observant l'usage de l'incertitude de mesure dans une activité spécifique de la physique de précision, celle des « ajustements des constantes de la physique ». Nous reprenons l'histoire de ces ajustements depuis leurs débuts avec l'initiative de Raymond Birge en 1929 jusqu'aux derniers travaux du CODATA au début du XXI^e siècle, en apportant une attention à deux épisodes significatifs de cette histoire, qui montrent chacun comment l'effort de *mise en commun* de résultats expérimentaux mène à une réflexion poussée sur le rôle et la signification de l'incertitude de mesure en physique.

Nous montrons que l'incertitude de mesure intervient de façon essentielle dans deux mécanismes propres à la mise en commun des connaissances expérimentales : la comparaison et la combinaison des résultats de mesure. Ce faisant, nous verrons que les physiciens se retrouvent eux aussi confrontés aux problématiques développées dans les parties précédentes, mais qu'ils n'adhèrent ni entièrement à l'objectivisme de l'approche traditionnelle ni entièrement à la solution épistémologique promue par les métrologues ces dernières décennies. Nous affirmerons que notre défense du concept de valeur vraie d'une grandeur s'accorde avec le positionnement des physiciens, lesquels conçoivent le progrès scientifique principalement comme une traque des erreurs de mesure en vue de leur correction, et gardent ainsi un attachement à la dimension évaluative de la mesure.

Cette partie est structurée selon quatre chapitres, organisés de la façon suivante. Au préalable, dans le chapitre 8 nous introduisons le problème général de la mise en commun des résultats d'observation à partir d'un épisode de l'histoire de l'astronomie qui date du XVIII^e siècle, et oppose les attitudes épistémologiques de Leonhard Euler et Tobias Mayer à propos du traitement des données observationnelles. Cela met en relief la problématique de la « combinaison des observations » soulevée par Stigler⁹³. Cet épisode porte ainsi en germe l'idée d'un traitement statistique de l'erreur de mesure, qui prendra une forme plus aboutie dans la méthode des moindres carrés de Legendre et Gauss, et préfigure l'évaluation d'un résultat de mesure en vue de juger de la possibilité de l'intégrer à un ensemble plus large de connaissances – c'est-à-dire l'analyse d'incertitude.

Le chapitre 9 est consacré au travail de pionnier de Raymond Birge sur les « ajustements des constantes de la physique » en 1929. De façon essentielle, l'initiative de Birge vient illustrer l'usage qui est fait de l'incertitude de mesure en physique de précision, en montrant comment cette dernière sert de base pour comparer et combiner des résultats de mesure. Ainsi, l'incertitude de mesure est un outil central pour quantifier et traiter mathématiquement le désaccord ou l'accord à l'intérieur d'une collection de résultats expérimentaux caractérisant des grandeurs reliées théoriquement entre elles par un système d'équations. L'usage de l'incertitude de mesure dans l'ajustement de Birge nous fera pressentir que celle-ci n'est ni une évaluation de l'exactitude de la mesure ni une indication de la confiance que l'expérimentateur porte dans le résultat.

Le travail de Birge est une initiative scientifique d'abord isolée, qui est progressivement devenue une pratique régulière chez les physiciens, jusqu'à être institutionnalisée. La pratique des ajustements des constantes de la physique a engendré de nombreuses discussions. En 1970, à l'occasion d'une conférence internationale organisée au National Bureau of Standards, la

93. Stigler (1986)

légitimité des ajustements est remise en question. Le chapitre 10 est consacré à l'examen de ces discussion, et permet de comprendre les vertus épistémiques que les acteurs accordent aux ajustements des constantes comme de la physique. Ainsi, l'intérêt que présentent ceux-ci n'est pas tant d'obtenir de « bonnes » valeurs des constantes que d'induire un travail constant de traque et de correction des erreurs de mesure.

Nous achèverons alors cette partie par le chapitre 11 où nous conclurons sur l'interprétation du concept d'incertitude de mesure et sur les rapports qu'il entretient avec l'exactitude de mesure.

Chapitre 8

Combinaison des résultats de mesure et ajustements aux moindres carrés

Dans ce chapitre, nous revenons sur un épisode spécifique du développement de la notion d'erreur statistique au milieu du XVIII^e siècle. Nous nous appuyons ici essentiellement sur la littérature secondaire, et tout particulièrement sur le travail de Stephen Stigler. Stigler a exhibé un mécanisme, la *combinaison des observations* qui vient débloquent un problème épistémologique majeur, à savoir : peut-on mettre en relation, et en particulier mettre en commun, des résultats d'observations obtenus dans des conditions très différentes ? L'étude historique menée par Stigler montre que cela n'avait au départ rien d'évident pour les astronomes du XVIII^e siècle, avant que ceux-ci n'en viennent progressivement à adopter la combinaison des observations, qui revient à affirmer qu'il y a effectivement un bénéfice épistémologique à mélanger de tels résultats d'observation. La combinaison des observations s'appuie sur une conception de l'erreur de mesure qui prend une forme aboutie dans la méthode des moindres carrés de Legendre et Gauss, et qui préfigure le traitement probabiliste des résultats de mesure.

En discutant un épisode précis de cette histoire, nous montrerons que le geste épistémologique de combinaison des observations engage, *sans l'achever*, un processus d'élargissement de l'idée de *répétabilité* des mesures, qui est une première étape vers une conception complète de l'idée de « reproductibilité » qui est au fondement de la possibilité de faire *circuler* les résultats de mesure, et dont le ressort principal est l'analyse de l'incertitude de mesure. Les conditions de reproductibilité sont ainsi le moteur de la mise en commun des résultats expérimentaux, qui est elle-même l'élément central de la quête d'identification (puis de correction) des erreurs de mesure dites « systématiques », puisque celles-ci se manifestent de façon invisible dans des conditions de répétabilité trop restreintes, c'est-à-dire lorsqu'un même expérimentateur répète des mesurages dans des conditions à chaque fois pratiquement identiques. Nous voyons ainsi apparaître en quoi l'étude de cette période historique, et des fermentes de l'idée de reproductibilité, est le prélude à l'objet que nous étudierons ensuite, à savoir l'ajustement des constantes de la physique, qui est une mise en œuvre systématique de la comparaison et de la combinaison de résultats de mesures obtenus dans des conditions différentes, afin, en particulier, de faire apparaître des effets physiques jusqu'alors inconnus qui viennent perturber différemment chacun de ces résultats. De fait, la problématique commune à ce chapitre et aux chapitres suivants est

la façon dont sont réalisées la comparaison et la combinaison des résultats de mesure, ce qui nous amène à concevoir la mesure dans sa dimension collective.

Dans la section 8.1, nous explicitons ce qui est entendu par « combinaison des observations ». Puis, dans la section 8.2, nous introduisons un épisode historique illustrant les attitudes opposées de deux savants, Euler et Mayer, concernant l'appréhension quantitative de l'erreur de mesure et sa conséquence sur la possibilité de combiner les observations. Dans la section 8.3, nous exposons l'importance d'une caractéristique commune aux problèmes qui se posent aux deux savants, à savoir la surdétermination des systèmes d'équations auxquels ils aboutissent chacun. Dans la section 8.4, nous montrons que l'une des différences principales entre les solutions proposées par les deux savants à ce problème de surdétermination réside dans la façon dont ils conçoivent la *propagation des erreurs*. Dans la section 8.5, nous relierons cette différence d'attitude aux conditions dans lesquelles ont été effectuées les observations exploitées par chacun des savants. En conclusion, dans la section 8.6, nous ouvrons sur la méthode des moindres carrés à laquelle Legendre et Gauss donneront jour environ un demi-siècle plus tard, qui vient synthétiser l'avancée épistémologique que constitue la combinaison des observations, et qui est à la racine des méthodes d'ajustement des constantes de la physique que nous étudierons dans les chapitres suivants.

8.1 La « combinaison des observations »

Dans la partie I, nous avons explicité en quoi la notion d'*erreur de mesure*, et en particulier son acception au sens statistique du terme, est un pivot de la conception moderne de la mesure physique. La genèse de cette notion a été documenté en détail par Stephen Stigler dans son livre consacré à la mesure de l'incertitude avant 1900. Stigler affirme que l'erreur statistique de mesure prend ses racines au milieu du XVIII^e siècle dans les disciplines que sont l'astronomie et la géodésie. Il fait remonter le développement progressif de la pratique régulière de l'erreur de mesure à une période initiale comprise entre 1748, année à laquelle Tobias Mayer s'attaque au phénomène astronomique alors inexplicable de « libration » de la Lune, et 1805, année de la publication par Legendre de la méthode dite des « moindres carrés »¹, ouvrant la voie à un traitement probabiliste qui atteindra une première maturité vingt ans plus tard avec la « synthèse de Laplace-Gauss »².

Selon le compte-rendu de Stigler, l'étincelle provient d'un acte épistémologique fondateur, qu'il appelle « combinaison des observations »³ en reprenant à son compte une expression de l'époque. La *combinaison des observations* exprime l'idée selon laquelle il est possible et profitable d'exploiter ensemble les résultats de mesures effectuées par différents observateurs, en différents lieux, avec différents instruments de mesure, ou encore selon des principes et

1. « Les moindres carrés ont été le dernier maillon d'une chaîne de développement qui a commencé environ en 1748, et [...] à la fin des années 1780 des méthodes qui étaient largement connues et utilisées se révélaient, pour des raisons pratiques, adéquates pour les problèmes rencontrés », Stigler (1986), p.15. Voir plus généralement le chapitre "Least Squares and the Combination of Observations", pp.11-61.

2. Stigler (1986), "The Laplace-Gauss Synthesis", pp.139-158.

3. Stigler (1986), p.11.

méthodes de mesure variés. Selon Stigler, il faut attendre la seconde moitié du XVIII^e siècle pour que commence à émerger une conception élargie de la combinaison d'observations différentes. Jusqu'alors, l'emploi de la moyenne est restreinte à un cas très précis, à la condition que les observations soient effectuées dans des circonstances *considérées comme identiques* :

Les astronomes moyennaient les mesures qu'ils considéraient comme équivalentes, les observations qu'ils jugeaient de précision intrinsèque égale parce que les mesures avaient été faites par le même observateur, au même moment, en un même lieu, avec le même instrument, et ainsi de suite. Les exceptions, des instances dans lesquelles des mesures qui n'étaient pas considérées équivalentes étaient combinées, étaient rares avant 1750.⁴

L'analyse de Stigler montre comment l'usage alors rudimentaire de la moyenne arithmétique a été progressivement enrichi et raffiné par les astronomes dans la seconde moitié du XVIII^e siècle. Pour illustrer ce changement d'usage, Stigler a étudié les attitudes des savants renommés du XVIII^e siècle que sont Tobias Mayer et de Leonhard Euler. Dans leurs travaux respectifs d'astronomie, Mayer et Euler ont développé deux façons sensiblement différentes de combiner les collections de données expérimentales qu'ils avaient à leur disposition. Pourtant concomitantes et explorant des questions assez similaires, les approches de Mayer et Euler contrastent nettement, et leur confrontation illustre de façon frappante le rôle épistémique de l'erreur de mesure dans le rapport entre théorie et expérience. Stigler voit même dans les travaux de Mayer un « succès statistique » et dans ceux d'Euler un « échec statistique »⁵ ; en particulier, le raisonnement que Mayer tient quant à l'erreur de mesure lui a permis de débloquent un verrou conceptuel essentiel pour ouvrir la porte à la combinaison des observations. Dans ce chapitre, nous reprenons à notre compte l'analyse de Stigler, afin de comprendre le rôle que joue la notion d'erreur de mesure pour rendre conceptuellement possible la combinaison des résultats de mesure.

8.2 Mayer et Euler face à deux problèmes d'astronomie

Tobias Mayer (1723-1762), astronome allemand, a produit un travail à la fois théorique et expérimental, et a de ce fait effectué de nombreuses observations astronomiques au cours de sa vie. Entre 1748 et 1750, il s'attaque à un problème astronomique remarquable de son époque : l'explication de la « libration de la Lune ». La libration de la Lune est un phénomène complexe en vertu duquel il est possible, pour un observateur situé sur Terre, d'observer presque les deux tiers de la surface de Lune au cours d'un mois lunaire, alors que, la rotation de la Lune sur elle-même étant synchrone avec sa révolution autour de la Terre, la Lune devrait en principe toujours présenter le même hémisphère. Leonhard Euler (1707-1783), mathématicien suisse, s'est lui aussi consacré, dans un mémoire daté de 1749⁶, à une question épineuse et alors inexpliquée de l'astronomie de son époque, concernant une variation séculaire (en apparence

4. Stigler (1986), p.16.

5. « Parce qu'ils étaient largement lus, ils ont grandement influencé les travaux ultérieurs et, d'un point de vue statistique, forment un contraste unique et dramatique dans le traitement des données d'observation. Ensemble, ils racontent l'histoire d'un succès statistique (par un astronome de renom, Mayer) et d'un échec statistique (par un mathématicien de premier plan, Euler) », Stigler (1986), p.17.

6. Euler (1749)

apériodique) dans les trajectoires relatives de Jupiter et Saturne. Les problèmes engagés par les deux savants sont en grande partie similaires : il s'agit de mettre à contribution les principes de la dynamique newtonienne pour expliquer les phénomènes observés ; la crédibilité de cette dernière est alors renforcée si elle résiste aux anomalies apparentes que l'observation astronomique révèle. Cependant, ces deux phénomènes, libration de la Lune et mouvement séculaire de Saturne et Jupiter, présentent une différence notable qui se révèle essentielle lorsqu'on s'intéresse de plus près aux attitudes épistémologiques respectives de Mayer et Euler. La libration de la Lune est la manifestation d'un mouvement périodique rapide, dont la périodicité est de l'ordre du mois. Même si l'anomalie du mouvement que cherche à expliquer Euler présentait une périodicité, celle-ci serait au minimum de plusieurs siècles. Cette différence influe sur le statut des observations exploitées par chacun des deux savants. En effet, il est possible d'analyser le premier phénomène à partir d'un échantillon d'observations effectuées au cours de quelques périodes de révolution, donc quelques mois, ce qui est une échelle de temps raisonnablement courte. En particulier, il est possible, pour Mayer, d'exploiter uniquement ses propres observations. Ainsi Mayer s'appuie-t-il sur vingt-sept de ses propres observations, effectuées entre 1748 et 1749. En revanche, Euler ne peut étudier le second phénomène qu'en le suivant sur une échelle de temps nettement plus longue. De fait, Euler exploite soixante-quinze jeux de mesures de sept grandeurs, mesurés par *différentes personnes* entre 1582 et 1745 ; aucune d'entre elles n'est de sa propre main.

Les cas respectifs de Mayer et Euler présentent un intérêt notable car les deux savants font face à un problème mathématique identique. Dans les deux cas, l'analyse astronomique s'appuie sur une équation centrale, l'« équation de condition », qui prend typiquement la forme suivante :

$$f(a_1, a_2, \dots, a_p, \alpha_1, \alpha_2, \dots, \alpha_q) = 0 \quad (8.1)$$

Dans cette équation, les a_i sont des grandeurs variables accessibles à l'observation (leur variation traduit le mouvement des astres), et les α_i sont différentes constantes inconnues que l'on cherche à connaître (pour les unes comme pour les autres, il s'agit souvent d'angles, par exemple une latitude, dans le cas de problèmes d'astronomie). L'équation de condition est issue d'une construction théorique : pour les deux problèmes étudiés ici, elle met en jeu les lois de la dynamique newtonienne et est donc valable à chaque instant, si la théorie est correcte. Par conséquent, à chaque fois que l'observateur effectue une observation groupée des p grandeurs variables a_i , il peut injecter les mesures de ces dernières dans l'équation d'observation, et aboutir à une équation à q inconnues – les q constantes α_i qui ne varient pas d'une observation à l'autre. Ainsi, en mobilisant l'équation de condition de façon répétée, l'observateur dispose en fin de compte d'autant d'équations qu'il a effectué d'observations groupées. Dans le cas de Mayer, l'équation de condition relie trois variables observables et trois constantes inconnues ; pour le problème d'Euler, l'équation relie sept inconnues et huit variables observables⁷ (tableau 8.1).

7. Voir [Stigler \(1986\)](#), p.21, pour Mayer ; et p.26, pour Euler.

| Mayer | Euler |
|--|--|
| Équation de condition : | Équation de condition : |
| $\beta - (90^\circ - h) = \alpha \sin(g - k) - \alpha \sin \theta \cos(g - k)$ | $\begin{aligned} \varphi = \eta - 23525'' \sin q + 168'' \sin 2q \\ - 32'' \sin 2\omega - 257'' \sin(\omega - q) \\ - 243'' \sin(2\omega - p) + m'' - x'' \sin q \\ + y'' \sin 2q - z'' \sin(\omega - p) \\ - u(\alpha + 360\nu + p) \cos(\omega - p) \\ + Nn'' - 0,11405k'' \cos q \\ + (1/600)k'' \cos 2q \end{aligned}$ |
| Trois inconnues : α, β, θ Trois variables observables : g, h, k Vingt-sept jeux d'observation de g, h, k | Huit inconnues : $x, y, m, z, \alpha, k, n, u$. Sept variables observables : $\varphi, \eta, q, \omega, p, N, \nu$ Soixante-quinze jeux d'observation de $\varphi, \eta, q, \omega, p, N, \nu$ |

TABLEAU 8.1 – Équations de conditions rencontrées par Euler et Mayer. L'étude de Mayer prend pour objet la position du cratère lunaire « Manilius », et l'équation de condition relie sa latitude β avec les mesures des positions angulaires d'autres points de repère de la surface lunaire. Celle d'Euler se concentre sur la longitude héliocentrique η de Saturne. Les deux équations de condition sont obtenues par application de la mécanique newtonienne puis par différentes opérations de linéarisation visant à simplifier les expressions obtenues.

8.3 Surdétermination des systèmes d'équations

Le point nodal, et commun aux deux problèmes, est la *surdétermination* des systèmes d'équations mis en jeu : Mayer et Euler ont tous deux à disposition plus d'observations que nécessaire pour déterminer les constantes inconnues de leurs problèmes respectifs. Mayer compte vingt-sept jeux d'observations pour trois inconnues à déterminer ; Euler exploite soixante-quinze jeux d'observations afin d'identifier huit inconnues. Par conséquent, ils aboutissent chacun à un système d'équations dans lequel le nombre d'équations est supérieur aux nombres d'inconnues : le système est dit surdéterminé. La résolution des systèmes surdéterminés d'équations est un problème mathématique classique, et le système n'admet une solution, en toute rigueur, que si les équations sont toutes compatibles entre elles. Si le système admet une solution, c'est que les observations concordent entre elles et sont cohérentes avec les prédictions de la théorie. Or, que ce soit dans les travaux de Mayer ou dans ceux d'Euler, les observations ne concordent pas et les systèmes d'équations n'admettent pas de solution mathématique. De fait, la surdé-

termination des systèmes d'équations mis en jeu, qui est pourtant la conséquence d'un état de faits tout à fait *positif* – lié à la profusion des observations disponibles – est la cause d'un problème sérieux pour les deux savants : que conclure s'il n'y a pas de solution mathématique au problème ?

C'est ici que réapparaît, en filigrane, la notion d'erreur de mesure : si les équations ne présentent pas de solution, cela peut être parce que les observations ne sont pas totalement exactes, c'est-à-dire qu'elles sont affectées d'une erreur, selon le raisonnement essentiel qui structure invariablement les théories de l'erreur et que nous avons exposé dans les différentes parties de ce travail, et qui est précisément une instance de ce que Hon appelle le « problème de l'erreur »⁸. Si les observations effectuées sont imprécises, il n'est pas surprenant d'obtenir des équations incompatibles entre elles. Certes, l'absence de solution mathématique pourrait être imputée à la validité de la théorie newtonienne, que Mayer et Euler ont tous deux pour ambition de tester. Mais elle pourrait également être reportée sur la qualité du contenu empirique – les observations astronomiques elles-mêmes. Toute la difficulté est alors de trouver le bon moyen de trancher entre l'hypothèse selon laquelle les observations effectuées sont imprécises et celle selon laquelle la théorie est en cause.

Mayer et Euler savent pertinemment qu'ils ne font pas simplement face à un problème purement mathématique – résoudre des systèmes d'équations – et tous deux comprennent qu'il faut y adjoindre un raisonnement *physique* sur la précision des observations effectuées. Ainsi Stigler rapporte-t-il de Mayer la remarque suivante :

[Mayer] notait que même dans des conditions favorables, une observation individuelle d'un arc pouvait seulement être comptée comme exacte dans une limite de 10 à 15 minutes, et il affirmait que l'effet d'une erreur de cette amplitude [...] sur les déterminations finales pouvait être retracée au travers des formules qu'il avait données.⁹

En soi, l'idée même d'erreur expérimentale n'est pas nouvelle pour les savants du XVIII^e siècle¹⁰. C'est dans la façon de *traiter* l'erreur de mesure chez Mayer qu'émerge quelque chose de nouveau. L'approche de Mayer contient un embryon de traitement de type statistique qui, selon Stigler, fait écho à l'idée de *combinaison des observations*.

8.4 Propagation des erreurs

Doit-on reporter l'incompatibilité entre les observations et la théorie sur la qualité des observations ou sur la validité de la théorie ? Mayer et Euler sont tous deux confrontés à cette question mais élaborent deux façons différentes d'y répondre. La résolution du problème posé amène les deux savants à réfléchir explicitement sur la manière dont interagissent les différentes erreurs qui affectent chaque observation – ce qu'en termes contemporains on appellerait

8. Hon (1998)

9. Stigler (1986), p.23.

10. Nous pouvons, à titre d'exemple, renvoyer vers les travaux de Hon sur le problème de l'erreur dans l'astronomie grecque, chez Kepler et Galilée. Voir Hon (1987), Hon (1989a), Hon (2004).

la « propagation » des erreurs. Euler choisit de regrouper les observations par petits groupes, chaque groupe formant un système d'équations non surdéterminé. Il les résout séparément et compare alors les solutions obtenues en jugeant si leurs écarts sont ou non satisfaisants. Euler considère en particulier que, lorsque des résultats sont combinés entre eux, les erreurs s'additionnent¹¹. Sans préjuger d'une quelconque filiation, nous voyons assez clairement la similarité que ce raisonnement entretient avec les méthodes de propagation linéaire des « limites d'erreur » que nous avons décrites au chapitre 3 de la partie I. Stigler attribue en partie cette attitude à une forme de conservatisme lié à son statut de mathématicien :

Il a longtemps été dans la pratique des mathématiciens de penser en termes d'erreur *maximale* qui pourrait se produire dans un calcul complexe, plutôt qu'en termes d'erreur vraisemblable, et de penser en termes de limites absolues d'erreur (qui augmenteraient typiquement avec l'agrégation) plutôt qu'en termes de tailles vraisemblables d'erreur (qui ne le feraient pas).¹²

Pour Euler, l'élément décisif réside dans le fait que des observations de moins bonne qualité risquent de polluer¹³ le résultat final si elles sont mélangées avec les observations identifiées comme plus fiables. Dès lors, il est préférable de ne pas combiner les différentes observations.

À l'inverse, Mayer adopte la solution du « mélange ». Celle-ci consiste à considérer que l'excès d'information que traduit la surdétermination des systèmes d'équation peut être exploité avantageusement, de façon à obtenir un résultat d'autant plus affiné que le nombre d'observations menées est important. Mayer se dirige dans cette direction en utilisant une méthode *ad hoc* qui revient, à grands traits, à réduire la taille du système d'équations en combinant les équations par groupe selon une opération de moyennage¹⁴. L'avancée qu'apporte le travail de Mayer réside dans l'idée suivante : *en combinant des résultats d'observation, même pris dans des conditions différentes, il est possible d'aboutir à un résultat plus fiable encore*. Comme l'explique Stigler,

L'important n'est pas qu'il ait trouvé une méthode particulièrement astucieuse pour combiner les vingt-sept équations mais qu'il simplement qu'il a trouvé utile de combiner les équations plutôt que de, disons, se contenter de sélectionner trois « bonnes » équations et de résoudre les inconnues à partir de celles-ci, comme il le faisant à titre d'illustration.¹⁵

La méthode *ad hoc* de Mayer introduit ainsi un mouvement intellectuel consistant à penser que la multiplication des observations conduit effectivement à une réduction de l'erreur finale – c'est une intuition de ce que nous pouvons interpréter comme une forme de *réduction* statistique des erreurs. Mayer évalue que l'erreur est réduite proportionnellement au nombre total d'observations mobilisées :

Parce que ces dernières valeurs [basées sur l'ensemble des vingt-sept équations] ont été tirées de neuf fois plus d'observations, on peut donc conclure qu'elles sont

11. « Par la combinaison de deux équations ou plus, les erreurs des observations et des calculs se multiplient » (Euler, 1749, p.135, cité p.29 dans Stigler).

12. Stigler (1986), pp.28–29.

13. Stigler emploie le terme « contaminer », Stigler (1986), p.28.

14. Stigler (1986), pp.21–23.

15. Stigler (1986), pp.25–26.

neuf fois plus correctes (*neunmal richtiger*); par conséquent l'erreur (*Fehler*) dans chacune de ces constantes est en raison inverse du nombre de leurs observations. (Mayer, 1750, p.155).¹⁶

Bien qu'il n'y ait à ce moment aucune considération de type probabiliste chez Mayer, le contraste entre la façon dont Euler et Mayer conçoivent la combinaison des erreurs de mesure fait écho aux deux approches, probabiliste et non probabiliste, qui ont coexisté dès les débuts de la théorie des erreurs et perduré au XX^e siècle dans les pratiques des scientifiques¹⁷. L'exemple d'Euler et Mayer ne nous montre pas l'émergence d'une idée d'*erreur de mesure* en tant que telle, mais la progression de l'idée selon laquelle combiner les résultats mène à réduire l'erreur affectant chaque observation individuelle. C'est en ce sens que l'approche de Mayer est révélatrice d'une attitude qui contient, en germe, la possibilité d'un traitement statistique des données expérimentales. Si Mayer ne formalisera pas son approche selon une méthode systématique, celle-ci se révélera tout de même influente¹⁸. Quelques décennies plus tard, Legendre publiera en 1805 la méthode dite aux « moindres carrés ». Laplace et Gauss viendront alors en donner une interprétation probabiliste, aboutissant à une théorie probabiliste des erreurs.

8.5 Conditions d'observation et reproductibilité des mesures

Au vu de l'analyse de Stigler, nous serions tentés de conclure que l'histoire a « donné raison » à Mayer sur Euler¹⁹. Cependant, il nous faut revenir ici sur une différence essentielle qui sépare les deux problèmes et qui semble influencer sensiblement les attitudes épistémiques qu'adoptent les deux savants – au-delà des traditions et intuitions séparées des mathématiciens et des astronomes. Euler exploite quant à lui des observations qu'il n'a pas mené lui-même. De ce fait, il ne peut pas juger leur fiabilité relative, et il peut effectivement craindre que les

16. Stigler (1986) citant Mayer, p.23. Les termes d'origine (en italique) sont reproduits par Stigler dans son texte original.

17. Nous renvoyons au chapitre 3.5 pour le détail de la discussion correspondante. Rappelons en particulier qu'Armatte a montré que, dès ses débuts, la théorie des erreurs a été marquée par une dualité dans les conceptions de l'erreur, associée à une dualité de la façon de les combiner, linéairement ou non (Armatte, 2010).

18. Stigler (1986), pp.30–31. Elle est connue sous le nom de « méthode des moyennes » de Mayer, et restera utilisée bien après le développement de la méthode des moindres carrés, à en croire Birge (1932a), p.213, note n° 12.

19. Jan Lacki (communication privée) suggère que la différence d'attitude entre Mayer et Euler s'explique en partie par la divergence de leurs objectifs. Euler se place dans une perspective de test de la théorie newtonienne. Il inscrit son travail dans une réflexion générale sur la validité de la loi de gravitation qu'avait formulée Newton, et en particulier sur la validité de sa décroissance en raison inverse des carrés des distances. De ce fait, il accorde une grande importance à la recherche d'alternative théoriques qui s'accorderaient mieux avec les observations. Mayer n'a pas cette intention, en raison de quoi son travail se rapproche plus d'une simple recherche du meilleur compromis entre théorie et observation. Dans son analyse, Stigler n'a pas vraiment ouvert la question épistémologique de l'accord entre théorie et expérience car son propre objectif est de baliser la route qui mène de la *combinaison des observations* à la méthode des moindres carrés et à leur explication probabiliste par Laplace et Gauss. Je remercie Jan Lacki d'avoir attiré mon attention sur cet aspect. Sur un autre plan, Maarten Bullynck nuance également le portrait bien dressé par Stigler de cette période. Stigler place la question statistique au centre de son étude, ce qui, selon Bullynck, a pour conséquence d'occulter le rôle d'autres acteurs concernés plutôt par les questions instrumentales et techniques (Bullynck [2014]). Le travail de Stigler n'en demeure pas moins remarquablement précieux, car il met en avant certains mécanismes épistémiques que l'on retrouve par la suite dans les travaux des physiciens et des métrologues du XX^e siècle jusqu'à nos jours.

meilleures observations soient « contaminées » par d'autres, moins bonnes. Cela explique en partie pourquoi il est réticent à les combiner entre elles. Sur ce plan, sa situation est à l'opposé de celle de Mayer, qui exploite des observations qu'il a effectuées lui-même, en raison de quoi il peut légitimement juger qu'elles sont également fiables. De ce point de vue, le problème de Mayer est moins complexe que celui d'Euler sur le plan épistémologique. Mayer effectue bel et bien une extension des conditions de répétabilité strictes sous lesquelles les astronomes s'autorisaient à considérer la moyenne de leurs observations, mais les conditions d'observations restent proches. En particulier, les instruments utilisés, ainsi que l'observateur, sont les mêmes pour chaque observation. Par conséquent, l'extension qu'opère Mayer n'est pas aussi significative que celle qui aurait été nécessaire à Euler pour appliquer le même raisonnement. Mayer propose un premier raisonnement quant à la façon dont les erreurs se propagent, mais il ne dit rien de ce qu'il est envisageable de faire lorsque les erreurs diffèrent à chaque observation. Or, cette étape nécessite une conceptualisation plus poussée du traitement mathématique des erreurs – qu'apportera la méthode des moindres carrés. Il manque en fait à Euler, comme à Mayer, une notion de *pondération* des résultats de mesure qui deviendra par la suite un ingrédient fondamental de la « combinaison des observations », ce que nous verrons lorsque nous nous intéresserons à l'ajustement des constantes de la physique. La quantification de l'erreur de mesure, puis le développement de méthodes de pondération, sont précisément ce qui fonde aujourd'hui la notion de *reproductibilité* d'une mesure.

Notre lecture de l'exemple de Mayer et Euler gravite autour d'une idée essentielle, formulée en deux temps : (i) il est possible combiner les résultats de mesures effectuées dans des conditions différentes, et ce même s'ils ne sont pas exactement en accord ; (ii) il est généralement profitable d'effectuer une telle combinaison, en ce que l'erreur sur le résultat moyenné est plus faible que celle affectant chaque donnée individuelle. Ce mouvement épistémologique nous apparaît comme crucial en ce qu'il nous semble être l'une des raisons mêmes d'être de la notion contemporaine d'« incertitude de mesure ». Le point (i) fonde une conception élargie de l'idée de *reproductibilité* d'une mesure : l'on peut extraire un résultat de son contexte immédiat de production pour le communiquer à d'autres utilisateurs, et ce *malgré* les contingences du contexte de production ; et c'est précisément parce que l'incertitude de mesure vise à rendre compte des contingences en questions qu'elle permet la circulation des résultats. Le point (ii) rejoint une autre idée, tout particulièrement soulignée dans les approches modernes de la mesure, à savoir que l'amplitude d'une incertitude de mesure est avant toute chose la conséquence de la *quantité d'information* disponible à propos la grandeur mesurée²⁰. Ajouter à un corpus de résultats expérimentaux une observation même imprécise peut, en principe, apporter une information supplémentaire et donc perfectionner la connaissance sur le résultat final. Cependant, comme la discussion sur Euler et Mayer le montre, cela passe par une réflexion sur les degrés de confiance que l'on accorde aux résultats que l'on utilise, c'est-à-dire sur la fiabilité que l'on pense que présentent les résultats d'observation ainsi que les méthodes utilisées pour y aboutir. Autrement dit, il est nécessaire de faire intervenir un raisonnement quantitatif sur la fiabilité des résultats de mesure – c'est précisément ici qu'intervient l'incertitude de mesure.

20. Cette formulation ne doit pas donner l'impression qu'elle rend immédiatement nécessaire l'adhésion à une interprétation bayésienne de l'incertitude de mesure, qui n'est pas absolument nécessaire ici.

Comme l'exprime Stigler :

(L)a mesure seule n'est pas suffisante. Pour servir les fins de la science les mesures doivent être susceptibles de comparaison. Et la comparabilité des mesures exige une certaine compréhension commune de leur exactitude.²¹

L'exemple des ajustements des constantes de la physique, développé sur les deux prochains chapitres, nous permettra de montrer l'une des façons dont les physiciens ont mis en application cette idée dans le courant du XX^e siècle.

8.6 Conclusion : la méthode des moindres carrés

Concluons pour terminer sur la méthode des moindres carrés. L'un des apports de la méthode d'ajustement aux moindres carrés, publiée par Legendre en 1805, et que l'on peut voir comme la continuation des efforts de Mayer, est de fournir une méthode qui vient généraliser l'usage de la moyenne arithmétique chez les astronomes du milieu du XVIII^e siècle. La méthode s'appuie sur l'hypothèse de l'erreur de mesure, mais son constat essentiel n'est pas cette hypothèse en elle-même. En effet, l'hypothèse de l'erreur est déjà sous-jacente dans la moyenne arithmétique simple : s'il est nécessaire d'effectuer une moyenne, c'est que les résultats répétés sont différents entre eux, et que l'opération de moyennage vient réduire cette différence, c'est-à-dire permet de réduire l'erreur de mesure. L'apport de la méthode des moindres carrés est d'exploiter l'extension des conditions de reproductibilité réalisée par Mayer. Legendre lui-même la décrit comme une sorte de moyenne généralisée :

Ces formules sont les mêmes par lesquelles on trouveroit le centre de gravité commun de plusieurs masses égales, situées dans les points donnés ; d'où l'on voit que le centre de gravité d'un corps quelconque jouit de cette propriété générale. [...] On voit donc que la méthode des moindres carrés fait connoître, en quelque sorte, le centre autour duquel viennent se ranger tous les résultats fournis par l'expérience, de manière à s'en écarter le moins qu'il est possible.²²

La méthode des moindres carrés concerne typiquement le cas où plusieurs grandeurs sont interdépendantes, c'est-à-dire reliées par une équation qui exprime une loi physique donnée. Prenons pour simplifier le cas d'une loi affine reliant deux grandeurs physiques x et y en fonction de deux paramètres constants a et b (dans ce cas la méthode se réduit à ce que l'on appelle couramment une « régression linéaire ») :

$$y = f(x) = ax + b \quad (8.2)$$

Le principe fondamental de la méthode consiste à considérer que les paramètres a et b sont des paramètres « ajustables », c'est-à-dire que l'on peut (et on doit) déterminer le couple (a, b) auquel correspond le meilleur accord entre le modèle et l'expérience. Pour déterminer cet accord, il faut comparer les données expérimentales avec le modèle. Chaque mesure est un enregistrement d'un couple de valeurs (x_i, y_i) . Si le modèle est correct et que les mesures sont exemptes

21. Stigler (1986), p.1.

22. Legendre (1805), p.75.

d'erreur, chaque couple de valeurs doit être tel que $y_i = f(x_i)$, c'est-à-dire que la représentation graphique des données doit former une droite. Or, on ne constate (presque) jamais un accord parfait entre théorie et expérience. En général, le nuage de points formé par les couples (x_i, y_i) de données expérimentales n'est pas parfaitement aligné. Par conséquent, quelles que soient les valeurs de a et b , la droite $y = ax + b$ ne recoupera pas tous les points expérimentaux – voire n'en recoupera aucun. Pour déterminer la meilleure valeur de a et de b , il faut donc trouver le meilleur compromis entre les données et l'équation (8.2). L'écart entre un couple de données expérimentales (x_i, y_i) et la droite $y = f(x)$ représentant le modèle est caractérisé par le terme $y_i - f(x_i)$, appelé « résidu ». La méthode des moindres carrés consiste à choisir – « ajuster » – les valeurs de a et de b de telle façon que la somme des carrés des résidus :

$$\sum_{i=1}^n (y_i - f(x_i))^2 \quad (8.3)$$

soit minimale. Cette opération de minimisation n'admet qu'un seul couple (a, b) de solutions²³. Un exemple est donné à la figure 8.1. Legendre reconnaît que la méthode contient une part d'« arbitraire »²⁴ : minimiser la somme des carrés des résidus est un choix, et un autre choix aurait pu consister à minimiser la somme des valeurs absolues, ou la somme d'une autre puissance arbitraire, de ces résidus. Ce choix était d'abord motivé par l'objectif de « faire en sorte que les erreurs extrêmes, sans avoir égard à leur signes, soient renfermées dans les limites les plus étroites qu'il est possible »²⁵. C'est d'ailleurs cet arbitraire que Laplace et Gauss chercheront par la suite à justifier, non sans difficulté, par un modèle probabiliste.

23. James Woodward a proposé une analyse de quelques problématiques philosophiques liées à la régression statistique, dans son article [Woodward \(1988\)](#). Le problème qu'il étudie couvre un champ plus large que celui qui est concerné par l'exemple de Mayer et Euler présenté ici, ou que l'ajustement des constantes de la physique qui sera étudié dans les prochains chapitres. En effet, les cas que nous étudions ici concernent une situation dans laquelle une théorie physique permet de dériver une relation mathématique entre grandeurs. La relation s'inscrit donc dès le départ dans le contexte d'une loi physique, dont l'ajustement permet ensuite de tester la cohérence avec les données expérimentales. Woodward envisage une situation plus large dans laquelle on cherche à comprendre la relation entre deux variables, sans que celles-ci soient liées *a priori* par une loi. Dans ce cas, la régression linéaire est d'abord une étude de corrélation entre les données. Vient alors la question du rapport de *causalité* d'une variable à l'autre que l'on pourrait éventuellement en déduire. Dans les cas que nous traitons ici, le rapport de causalité est déjà supposé dans les théories physiques qui sont mobilisées (par exemple la mécanique newtonienne). L'article de Woodward explore ainsi deux thèmes qui ne sont pas centraux dans notre analyse, celui de la causalité et celui de l'explication dans les sciences. Aris Spanos a également porté son attention sur les méthodes d'ajustement, du point de vue de l'inférence inductive ([Spanos, 2007](#)). Spanos critique une perspective centrée sur l'idée d'*approximation*, pour laquelle le seul critère d'ajustement réside dans la proximité entre les données et le modèle – c'est le cas, par exemple, de la méthode de Legendre. Il affirme que le critère d'approximation, utilisé seul, peut produire de mauvais ajustements, par exemple si les résidus varient eux-même de façon déterministe avec la variable d'ajustement, alors qu'on serait en droit de s'attendre à ce qu'ils prennent la forme d'un « bruit blanc » (p.1050). Spanos y oppose une seconde perspective, qui puise dans l'« approche erreur-statistique » héritée de [Mayo \(1996\)](#), et insiste sur une réflexion préalable sur la validité des modèles statistiques employés. La position de Spanos n'est pas en contradiction avec la pratique ordinaire des physiciens ; il est rare que ceux-ci se contentent d'un ajustement pour tirer des conclusions théoriques définitives – comme nous le verrons dans les chapitres suivants, les physiciens insistent au contraire sur le caractère provisoire des valeurs tirées des ajustements, l'ajustement venant plutôt donner une idée de la cohérence des idées testées.

24. [Legendre \(1805\)](#), p.72.

25. [Legendre \(1805\)](#), p.72.

Depuis sa conception par Legendre et Gauss, la méthode des moindres carrés a été enrichie d'un grand nombre d'outils statistiques visant à évaluer la justesse et la cohérence de l'ajustement, et à raisonner sur les paramètres d'incertitude. Elle a connu un très long développement technique qu'il ne serait ni utile, ni possible, de développer ici. Nous souhaitons plutôt évoquer sa mise en application spécifique dans le cadre des « ajustements des constantes de la physique », qui consiste à mettre en commun toutes les mesures des constantes les plus fondamentales de la physique. Cette activité nécessite de raisonner dans un cadre de reproductibilité élargi, bien plus vaste encore qui avait permis à Mayer de concevoir la combinaison des observations. En élargissant le cadre de reproductibilité, les ajustements des constantes nécessitent une réflexion beaucoup plus poussée encore sur l'analyse d'erreur et le calcul de l'incertitude de mesure.

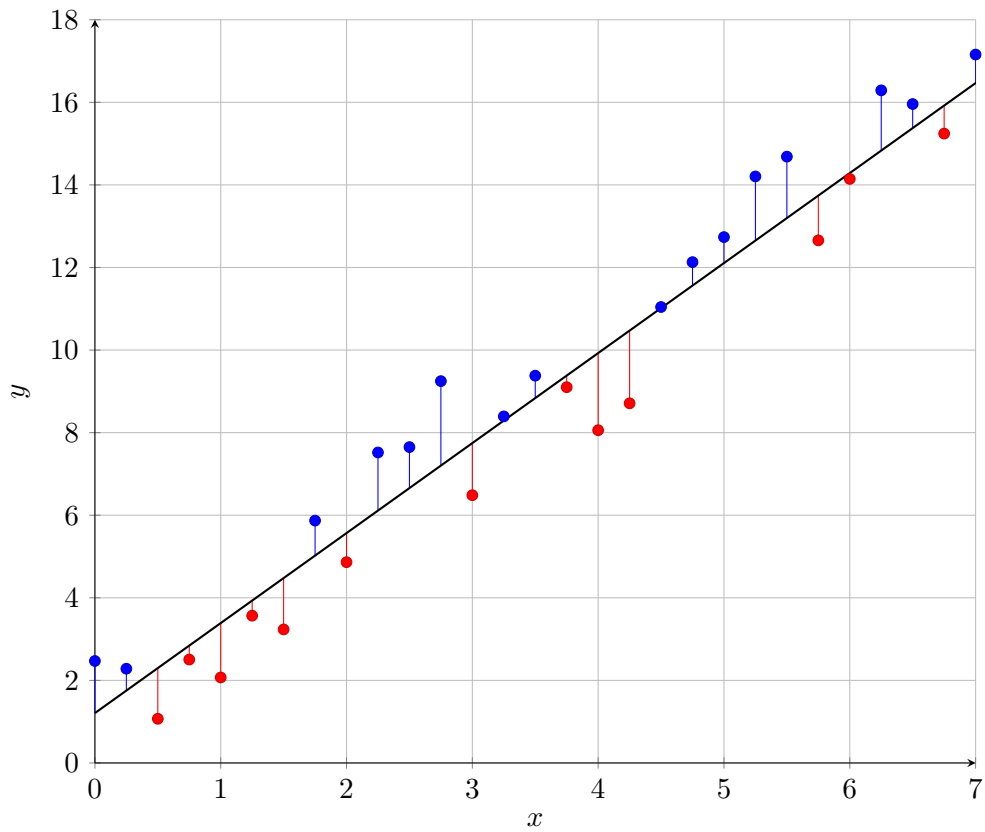


FIGURE 8.1 – Illustration de la méthode des moindres carrés sur un exemple fictif arbitraire, pour un modèle affine $y = ax + b$ (« régression linéaire »). Les données expérimentales sont représentées par les points (en bleu et en rouge), et le modèle ajusté est représenté par la droite (en noir). Cette droite est l'unique droite qui minimise la somme des carrés des résidus. Les résidus, qui quantifient l'écart entre le modèle et chaque donnée expérimentale, sont représentés par des segments qui joignent la droite aux points (résidus positifs en bleu, au-dessus de la droite; résidus négatifs en rouge, en-dessous de la droite). La répartition égale des résidus au-dessus et en-dessous de la droite ajustée fait assez nettement apparaître en quoi cette opération est similaire au calcul d'une moyenne.

Chapitre 9

L'ajustement des constantes de la physique : l'initiative de Birge

Le développement du chapitre précédent souligne un élément essentiel du développement du concept d'incertitude de mesure : pour pouvoir opérer une *combinaison des observations*, il faut avoir une idée de la qualité des mesures qui sont combinées entre elles, car cela conditionne la façon dont les erreurs de mesure se répercutent sur le produit final. Parce que Mayer était en droit de considérer que chacune des mesures qu'il avait effectuées présentait sensiblement la même précision, il pouvait envisager la possibilité que moyenner ces mesures revienne à réduire l'erreur finale. Euler n'avait pas cette possibilité, parce qu'il n'avait pas une idée précise de la qualité des observations qu'il exploitait. Mayer avait donc la possibilité d'élargir les conditions de reproductibilité dans lesquelles la combinaison des observations était possible, *parce qu'il avait mis à contribution* un raisonnement qui ressemble en tout point, bien que de manière plus élémentaire, à une analyse d'incertitude. Cela vient rappeler avec force un constat classique de la métrologie : l'incertitude de mesure est l'outil qui permet la communication des résultats de mesure et leur exploitation dans d'autres conditions, par d'autres personnes, en d'autres lieux et d'autres temps. L'élargissement des conditions de reproductibilité et l'affinage des méthodes de calcul d'incertitude permet aux physiciens du début du XX^e siècle de venir régulièrement comparer leurs mesures des constantes de la physique. Raymond Birge va venir enrichir encore le dialogue entre les physiciens en proposant de chercher à réunir l'ensemble des mesures existantes d'un petit groupe de constantes, de façon à en tirer des valeurs plus fiables encore, les valeurs « (les plus) probables » de ces constantes.

C'est cet épisode de l'histoire des sciences, le travail fondateur de Birge, que nous souhaitons étudier dans le présent chapitre. Cette étude nous permettra en particulier de mettre en évidence l'usage qui est fait de l'incertitude de mesure dans le domaine de la physique de précision. Nous verrons alors que la pratique des ajustements révèle une limite à la combinaison des observations, car il est nécessaire de faire intervenir préalablement une étape de sélection des données expérimentales. Or, cette limite nous amène à réexaminer une interprétation classique et très courante de l'incertitude de mesure, à savoir qu'elle quantifie la « confiance » que l'on peut avoir en ce résultat, ou encore sa « qualité » ou « fiabilité », autant de termes qui se révèlent difficiles à cerner pleinement. De fait, nous affirmons que cette interprétation est in-

valide si l'on s'en tient à l'usage qui en fait dans les ajustements des constantes de la physique.

Dans la section 9.1, nous débutons par une présentation générale de l'initiative de Birge, en montrant que Birge a d'abord été guidé par des motivations méthodologiques, qui l'ont mené à réfléchir aux différentes façons dont l'erreur probable était évaluée chez ses contemporains.

Nous exposerons ensuite, à la section 9.2 les principes généraux de la procédure d'ajustement, qui consiste à *modifier* un jeu de valeurs expérimentales des constantes de la physique de façon à ce que les nouvelles valeurs ainsi obtenues soient « ajustées » aux équations aux grandeurs qu'impliquent les théories physiques acceptées – d'où le nom de valeurs « ajustées ». Ce travail fait intervenir la méthode d'ajustement « aux moindres carrées » héritée de Legendre et de Gauss.

Nous rentrerons alors plus en détail dans le raisonnement mis en place par Birge, en nous appuyant sur l'exemple de deux constantes particulières, la célérité de la lumière dans le vide c , et la constante de Planck h . Le premier exemple révèle la nécessité de procéder parfois à une étape de *sélection*, lorsque les mesures recensées ne présentent pas un accord suffisant pour qu'elles soient mélangées entre elles. Le cas de la constante de Planck montre comment le mélange est effectué lorsque les différentes mesures recensées sont en bon accord. Cela permet de mettre en évidence que le mélange est effectué au travers d'une moyenne pondérée pour laquelle l'incertitude de mesure permet de déterminer les coefficients de pondération. Ce développement est l'objet de la section 9.3.

Ayant décrit les différentes étapes de la démarche de Birge, nous pourrions alors revenir sur l'interprétation générale de l'incertitude de mesure. Nous contesterons l'interprétation courante selon laquelle l'incertitude de mesure vient caractériser la qualité d'un résultat, ou encore la confiance que l'on accorde en un résultat. En effet, dans les cas où Birge doit sélectionner les mesures en amont de l'ajustement, l'incertitude de mesure elle-même n'entre pas dans le critère de sélection. En effet, il superpose un raisonnement supplémentaire sur la qualité supposée du résultat ; c'est donc que l'incertitude elle-même ne mesure pas cette qualité. Cette conclusion, portée dans la section 9.4, annonce un développement plus détaillé au chapitre 11.

9.1 L'idée initiale de Birge

L'ajustement des constantes de la physique est une activité qui remonte à l'année 1929, avec les travaux du physicien Raymond T. Birge (13 mars 1887–22 mars 1980). Birge entre en 1918 à la faculté de physique de l'Université de Berkeley, puis, en 1938, devient le directeur du département de physique, jusqu'à sa retraite en 1955¹. Il y consacre une large partie de ses travaux à la mesure de précision, se spécialisant d'abord en spectroscopie, où son activité consiste en partie à déterminer avec la meilleure précision possible certaines constantes physico-chimiques. Il s'intéressera tout particulièrement à la mesure de la constante de Planck². Élargissant son champ de recherche, Birge va être amené à s'intéresser à l'état général des connaissances sur les constantes physiques : avec quelle précision les connaît-on et quelles en sont les « meilleures

1. La vie professionnelle de Birge a été brièvement résumée par A. Carl Helmholz dans une notice nécrologique puis dans un mémoire biographique. Voir [Helmholz \(1980\)](#), [Helmholz \(1990\)](#)

2. [Birge \(1919\)](#)

valeurs » ? Ces constantes étant reliées les unes aux autres au travers d'un réseau d'équations, Birge réalise qu'il ne faut pas les considérer indépendamment les unes des autres, mais mettre en commun l'ensemble des connaissances disponibles à leur sujet. C'est ainsi qu'en 1929, il publie un article intitulé "Probable Values of the General Physical Constants"³, long de soixante-dix pages, dans lequel il cherche à effectuer une synthèse critique des mesures de précision des constantes physiques effectuées par ses contemporains, afin d'en dériver un ensemble unifié et cohérent de « valeurs probables » de trente-et-une constantes de la physique⁴. Cet article est publié dans *The Physical Review Supplement*, supplément qui changera de nom peu de temps après : l'article devient ainsi la première publication du premier numéro de la revue *Reviews of Modern Physics*.

Un tel travail de synthèse poursuit plusieurs buts. À court terme, il s'accompagne de la production d'un ensemble de valeurs recommandées (les « valeurs probables » ou valeurs « les plus probables » des constantes étudiées). C'est là l'essence même du travail d'ajustement : collecter de multiples mesures de chaque constante physique considérée, et les synthétiser en un jeu de valeurs ajustées, considérées comme les « meilleures » à l'instant donné au vu des connaissances disponibles. Birge insiste cependant sur le caractère perpétuellement provisoire de telles valeurs, puisque de nouveaux résultats expérimentaux sont en permanence susceptibles de venir nourrir la connaissance scientifique. À plus long terme, l'initiative de Birge stimule surtout une harmonisation des pratiques employées, couplée à une réflexion méthodologique et épistémologique, afin que les résultats expérimentaux soient communiqués par chacun d'une façon qui en permette la même utilisation par tous. Passant en revue la littérature de physique de la précision de son époque, Birge s'aperçoit qu'y prospèrent de nombreuses lacunes méthodologiques, et constate qu'aucune étude systématique de l'ensemble des constantes « générales » de la physique n'a été effectuée, à l'exception de certaines initiatives plus localisées disciplinairement :

(U)ne investigation des valeurs des constantes générales utilisées actuellement dans la littérature révèle un surprenant manque d'uniformité, aussi bien en ce qui concerne les valeurs effectivement adoptés et l'origine de ces valeurs. Ceci est probablement dû au fait qu'il est presque impossible de trouver une étude critique de ces meilleures valeurs, suffisamment à jour pour être vraiment fiable, et suffisamment détaillée pour expliquer les incohérences entre les anciennes tables. La situation est bien meilleure dans le cas de certains groupes de constantes. Ainsi, la meilleure valeur de la masse atomique de chaque élément est déterminée annuellement par certains comités de masse atomique, et la nécessité d'une telle liste de masses atomiques est évidente pour tout chimiste. Il y a certainement un besoin similaire dans le cas de constantes plus importantes encore, telles que la vitesse de la lumière, la charge de l'électron, la constante de Planck, etc.⁵

Dans son article, Birge vise ainsi à combler cette lacune, en entreprenant un travail critique de

3. Birge (1929b).

4. Notons que dans ce premier article, Birge admet lui-même s'être limité pour l'essentiel à des mesures effectuées dans des laboratoires situés aux États-Unis – car la communication avec ces différents laboratoires s'en trouvait facilitée; Birge (1929b), p.2.

5. Birge (1929b), p.2.

synthèse sur la mesure des constantes de la physique. Il conçoit tout à fait que ce travail ne soit qu'un prototype pour des développements futurs⁶. Il ne s'agit que d'une première tentative, d'autant que cette résolution réclame un travail collaboratif sérieux à l'échelle internationale⁷ et n'est pas vouée à rester le travail d'une seule personne. Notons que Birge n'est pas le premier à effectuer un travail de cette nature ; il précise lui-même qu'il s'aide substantiellement de deux sources principales d'inspiration⁸, mais il s'en distingue en poussant la démarche bien plus avant. Il systématise l'idée qu'il est important de réunir les informations disponibles, et que cette réunion est l'occasion d'une réflexion critique fructueuse quant aux méthodes employées pour obtenir ces informations.

Quel est le principe général du travail de Birge ? Il nous faut d'abord expliciter ce qui est entendu par « constantes de la physique » (ou « constantes physiques »). Celles-ci sont une catégorie particulière de grandeurs physiques, qui ne sont pas attachées à un objet en particulier mais décrivent des propriétés communes à une classe d'objets ou à une classe de phénomènes. Différents termes sont couramment utilisés ; ainsi, Jean-Philippe Uzan les nomme « constante de la nature » et les caractérise de la façon suivante :

La description de la nature à l'aide de lois mathématiques fait intervenir deux types de quantités dont le statut semble différent. Les premières décrivent l'état des systèmes étudiés (position, température, vitesse...) : elles sont sujettes à des équations d'évolution et peuvent donc varier au cours du temps. Les secondes sont des quantités supposées invariantes dans le temps et dans l'espace, les *constantes de la nature*. Ces constantes semblent décrire des propriétés universelles des interactions de la matière. Pas le moindre calcul qui ne dépende d'une d'elles et elles interviennent dans la formulation des lois physiques quel que soit le système particulier considéré.⁹

Birge s'intéresse pour sa part à ce qu'il appelle les « constantes générales de la physique »¹⁰ car il restreint son propos à un nombre réduit de constantes, mais mentionne ensuite le terme de « constantes universelles »¹¹ que reprend Gilles Cohen-Tannoudji selon une acception très spécifique et plus restrictive¹². Certaines constantes sont qualifiées de « constantes fondamentales »¹³, lorsqu'il est considéré que leur valeur ne peut pas être expliquée à partir de celles d'autres grandeurs physiques. Le simple terme de « constante physique », comme l'emploie

6. (« L)es résultats donnés ici devraient être considérés plus comme une présentation de la situation que comme une solution définitive du problème », Birge (1929b), p.2.

7. « (C)ela nécessite la coopération impartiales de nombreuses personnes situées dans des laboratoires scientifiques partout dans le monde », Birge (1929b), p.2.

8. Ces deux sources sont The International Critical Tables (McGraw-Hill Co., 1926) et le "Handbuch der Physik" de Hans Geiger et Karl Scheel (1926) ; voir Birge (1929b), p.7.

9. Uzan (2004), p.14. Voir aussi Uzan et Lehoucq (2005) et Cohen-Tannoudji (1998) pour des réflexions sur le statut des constantes physiques.

10. Birge (1929b), p.1.

11. Birge (1929b), p.1.

12. Cohen-Tannoudji (1998). Pour Cohen-Tannoudji, les constantes universelles sont une sous-catégorie particulière des constantes de la physique, et sont la constante de Planck, la célérité de la lumière dans le vide, la constante gravitationnelle et la constante de Boltzmann.

13. Cohen et DuMond (1965)

Lévy-Leblond, dans son essai de classification¹⁴, semble néanmoins suffire pour décrire le travail d'ajustement proprement dit.

Hacking a suggéré qu'il était plausible de faire remonter la préhistoire du concept de constante physique à 1832, lorsque Charles Babbage (1772–1871) réclame la publication d'une synthèse de toutes les constantes connues « dans les sciences et dans les arts »¹⁵ :

Churchill Eisenhart, du US Bureau of Standards, m'a suggéré un jour que le pamphlet de Babbage marque le début de l'idée moderne de "constantes de la nature". [...] Le point est que Babbage résume ce travail en déclarant officiellement ce qui était dans l'esprit de nombreux de ses contemporains, à savoir que le monde pourrait être défini par un ensemble de nombres, qui seraient appelés constantes.¹⁶

Le point essentiel derrière la démarche de Babbage n'est pas tant la mise en commun de savoirs, que l'idée que le monde peut être défini par un ensemble de grandeurs qu'on appelle « constantes ». George Smith, quant à lui, considère que c'est à l'année 1897, avec les travaux de Thomson sur l'électron, que l'on peut retracer les débuts d'une véritable attention portée à la mesure *précise* des constantes de la physique, et à l'intérêt qu'une telle précision représente¹⁷. Il recense ainsi six constantes prééminentes dans l'histoire de la mesure de précision ; durant la période couvrant les années 1897 à 1913, une importance grandissante est accordée à leur mesure. Il s'agit de la vitesse de la lumière dans le vide, la charge élémentaire, la constante de Planck, la constante d'Avogadro, la constante de Faraday, et la constante de Rydberg. Voyant dans les travaux de Birge la continuité de ce premier effort, Smith note que le nombre de constantes concernées n'a ensuite cessé de croître, dépassant cent-trente en 2006.

Pourquoi Birge ressent-il le besoin d'effectuer un tel travail d'ajustement des constantes ? L'article de Birge n'est nullement consacré à la question théorique ou même ontologique que pourrait suggérer une réflexion sur le caractère « fondamental » ou non des constantes de la physique ; son article ne présente aucune réflexion de nature philosophique ou classificatoire quant à la notion de « constante (fondamentale) de la physique »¹⁸. L'objet de l'article est dual ;

14. Lévy-Leblond (1977)

15. Hacking (1983), p.235.

16. Hacking (1983), p.235.

17. « Une appréciation complète, néanmoins, des éléments de preuve qui peuvent être obtenus d'une mesure précise des constantes fondamentales semble avoir émergé seulement durant la période de 1897 à 1913, en commençant par la mesure de J.J. Thomson du rapport masse à charge m/e de ce que nous pouvons maintenant appeler l'électron », Smith (2010), p.566. Voir pp.566–569 pour ses développements sur le sujet.

18. Birge laisse d'ailleurs de côté un élément essentiel que ses successeurs ne manqueront pas de remarquer : le statut et la détermination des constantes de la physique entretiennent un lien très fort avec la définition des unités de mesure. Ainsi, la précision des mesures des constantes physiques est décisive pour juger des opportunités de redéfinir ou non le Système International d'unités. Le travail d'ajustement effectué au sein du CODATA à la fin des années 2000 et au début des années 2010 est en permanence tourné vers la perspective d'une réforme majeure, engageant en particulier la redéfinition du kilogramme et désormais prévue pour 2018. Les questions relatives à la définition des unités sont entièrement absentes des articles de Birge et ne semblent pas avoir motivé son travail du tout. Elles seront rendues plus explicites quelques décennies plus tard ; voir par exemple l'article de Branscomb à ce sujet en 1971, Branscomb (1971). Pour une discussion sur le rôle des constantes physiques dans la définition des unités, voir par exemple Petley (1992). Mirowski a décrit l'histoire des étalons, des dimensions physiques, et des constantes adimensionnées comme le produit d'une volonté initiale de débarrasser le système métrique de tout anthropomorphisme. Il défend cependant l'idée d'un retour de l'anthropomorphisme à la fin du XX^e siècle

il procède à la fois d'un questionnement sur le rôle des constantes dans l'architecture théorique de la physique d'une part, et d'une démarche méthodologique et épistémologique d'autre part. Cette dualité est sous-jacente dans la première phrase de l'article :

Certains des résultats les plus importants de la science physique sont incorporés, directement ou indirectement, dans les amplitudes numériques des diverses constantes universelles; et la détermination exacte de ces constantes a engagé le temps et le travail d'un grand nombre de scientifiques les plus éminents de ce monde.¹⁹

Un premier élément réside dans le fait que l'étude des constantes de la physique présente un intérêt particulier, du fait du rapport qu'elles entretiennent avec les fondements des théories physiques – la mesure précise des constantes de la physique est un outil puissant de test de ces théories. Dans sa deuxième partie, la phrase citée embrasse sur le versant épistémologique et méthodologique du problème : comment réunir et organiser l'ensemble des connaissances expérimentales relatives aux constantes physiques ?

La lecture de l'article révèle que c'est le second point qui y est le plus mis en avant. Birge souhaite mener une réflexion critique sur la connaissance que la communauté scientifique possède à propos de ces constantes, et sur la façon dont cette connaissance a été acquise, qui implique de nombreux savants, instruments, méthodes, connaissances théoriques et expérimentales, etc. Le travail engagé est donc une analyse critique des méthodes employées par les différents laboratoires impliqués sur les questions de mesure des constantes de la physique, qu'elles soient des techniques instrumentales ou méthodes d'analyse statistique des données expérimentales. Il amène Birge à relever une grande inhomogénéité dans les pratiques de ses contemporains, et constate que de nombreuses informations sont occultées lorsque les scientifiques rendent compte de leur travail. Il déplore que certains travaux voient leur qualité détériorée du fait de choix malheureux dans le traitement des données, et ce au détriment de la qualité des procédures expérimentales elles-mêmes.

Le travail de Birge est remarquable du point de vue de l'importance qui y est accordé à une notion comme celle de l'« erreur probable », que l'on peut en un certain sens considérer comme l'ancêtre de l'incertitude de mesure contemporaine²⁰ :

Le terme « erreur probable » est un terme technique, appliqué correctement uniquement quand les erreurs sont distribuées selon une courbe gaussienne d'erreurs.

(Mirowski, 1992).

19. Birge (1929b), p.1.

20. Dans la suite de ce chapitre et dans les chapitres suivants, nous assimilerons le plus souvent *erreur probable* et *incertitude de mesure*. Bien que les termes ne soient pas rigoureusement équivalents sur le plan technique, ni même quant à leur interprétation, ils n'en sont pas moins très proches et découlent tous deux de la même idée générale. Leur distinction est utile pour discuter certains aspects spécifiques, relatifs à l'analyse détaillée de leur histoire et de leur conceptualisation, certes indiscutables mais néanmoins indispensables lorsqu'il s'agit de comprendre l'emploi de ces concepts dans une perspective plus générique, comme tel est le cas pour l'étude des ajustements. Nous nous attacherons toutefois à rester fidèle au vocabulaire utilisé par les acteurs et maintiendrons dès lors l'usage du terme « erreur probable » lorsque celui-ci est présent dans la littérature que nous étudions ; cependant, il convient de noter que cette attention ne vise pas tant à marquer une différence entre les concepts que pour éviter les anachronismes linguistiques.

Elle est définie comme l'amplitude numérique que la vraie erreur est susceptible d'excéder ou non. Ce concept est bien défini et en tant que tel fournit un critère *objectif* d'erreur.²¹

L'erreur probable est donc définie de telle façon qu'il y ait une probabilité 1/2 que l'erreur finale soit inférieure à l'erreur probable. Birge réaffirme l'importance qu'il y a à inclure un constat d'incertitude – une mention de l'erreur probable – à tout résultat expérimental²² :

Dans l'investigation présente, je me suis efforcé d'obtenir non seulement la valeur la plus probable d'une constante donnée, mais aussi son "erreur probable." Une estimation de l'erreur probable est souvent aussi importante que la constante elle-même.²³

L'analyse de l'erreur probable est articulée avec la méthode des moindres carrés héritée des travaux de Legendre et Gauss.

Dans les années qui suivront, Birge publiera à plusieurs reprises des synthèses méthodologiques qui résument la démarche qu'il a employée²⁴. C'est en tant que pionnier²⁵ de la pratique des ajustements des constantes de la physique (et non, bien sûr, de la méthode d'ajustement aux moindres carrés elle-même), que Birge est tout particulièrement reconnu dans le milieu de la métrologie et de la physique de précision. Rappelons à ce propos que Birge n'invente ici ni l'idée de départ, ni le détail technique de la méthode des moindres carrés, dont les bases ont été posées dès 1805 par Legendre et Gauss. Son apport réside dans le fait d'effectuer un ajustement sur l'ensemble des constantes principales de la physique fondamentale, et sur le travail d'harmonisation, d'aplanissement méthodologique et de prise de recul qu'il génère en conséquence.

9.2 L'ajustement : principes généraux

Birge se concentre sur trente-et-une constantes physiques « principales »²⁶, reliées entre elles par un réseau d'équations qui traduisent les lois physiques auxquelles sont soumises les grandeurs : c'est ce que Cohen et DuMond appelleront par la suite la « nature enchevêtrée [*entangled*] des données »²⁷. Parmi les équations fondamentales exposées par Birge, considérons,

21. Birge (1929b), p.4.

22. Comme le fait remarquer Terry Quinn, l'importance de l'erreur probable avait déjà été reconnue de façon formelle par le Comité International des Poids et Mesures dès la première décennie de son existence, au début des années 1880 (Quinn, 2012, pp.112–113)

23. Birge (1929b), p.4.

24. Notons par exemple Birge (1932a), Deming et Birge (1934), Birge et Weinberg (1947). La liste complète est beaucoup plus longue.

25. C'est ainsi que le qualifie A. Carl Helmholtz : « Son article de 1929, "Probables Values of the General Physical Constants," [...] est un travail de pionnier classique dans le domaine de la détermination des meilleures valeurs des constantes », Helmholtz (1980), p.63. Il considère que cet article est « probablement le plus important que Birge ait jamais écrit », Helmholtz (1990), p.77.

26. Exposées dans Birge (1929b), tableau a, p.59.

27. DuMond et Cohen (1948), p.85.

à titre d'exemple l'expression suivante (exprimée dans le système d'unités « CGS »)²⁸ :

$$R_{\infty} = \frac{2\pi^2 e^5}{h^3 c^2 e/m} \quad (9.1)$$

L'équation (9.1) établit une relation entre cinq grandeurs : la constante de Rydberg²⁹ R_{∞} , la charge élémentaire e , la constante de Planck h , la célérité de la lumière dans le vide c , et le rapport e/m entre la charge élémentaire et la masse de l'électron. Elle n'est qu'une branche particulière d'un réseau de p équations reliant entre elles un total de q constantes (toutes n'apparaissent pas nécessairement dans chacune des équations). On dispose par ailleurs de n valeurs expérimentales correspondant aux mesures (indépendantes) de constantes différentes. À l'instar des cas étudiés par Euler et Mayer en leur temps, le système est ici surdéterminé : il existe plus d'équations que d'inconnues, c'est-à-dire que $n + p > q$ ³⁰.

Dans un système où le nombre d'équations égale le nombre d'inconnues, les équations sont utilisées selon une visée *prédictive*, à savoir la détermination de la valeur des grandeurs jusque là inconnues à partir de la valeur des grandeurs qui ont été mesurées³¹. En regard de cette situation la présence d'un système surdéterminé aboutit, sur le plan épistémologique, à une réflexion d'un autre type, portant sur la *cohérence* des données expérimentales³². Pour Birge, la question est donc d'abord la suivante : les résultats des mesures des différentes constantes sont-ils cohérents entre eux ? À cette question, la réponse est négative, ce qui se manifeste de deux façons différentes. D'une part, les multiples résultats de mesure d'une même grandeur ne concordent pas, en particulier lorsqu'elles diffèrent par la date et le lieu où elles ont été menées, ainsi que par le groupe de scientifiques qui a mené la mesure, ou encore par la méthode employée³³. D'autre part, les valeurs collectées ne respectent *pas* les équations aux grandeurs auxquelles elles sont censées se soumettre dès lors que la théorie est correcte. Pour la seule équation (9.1), on ne trouve aucun jeu de valeurs collectées « d'origine »³⁴, que nous notons

28. Birge (1929b), p.49. « CGS » signifie « centimètre-gramme-seconde ».

29. Il s'agit plus précisément de la constante « infinie » de Rydberg. Le produit $R_{\infty} h c$ mesure l'énergie d'ionisation d'un atome hydrogène dans le cas fictif où le noyau aurait une masse infinie.

30. Selon cette description, le résultat de la mesure d'une constante est lui-même compris comme une équation du système : valeur de la grandeur = valeur mesurée.

31. Lorsqu'une équipe de recherche calcule la vitesse (moyenne) v de neutrinos par mesure d'un temps de vol T sur une distance parcourue D , elle s'appuie sur une équation physique $D = v T$ et deux mesures, celles de D et T . Cette seule équation permet la prédiction de v , la seule véritable inconnue du problème, à partir de la connaissance de D et T .

32. Birge exprimera cette idée plus explicitement dans son ajustement de 1932 : « Un énoncé général de la procédure correcte est le suivant. S'il existe une équation théorique contenant, disons, e , h , et e/m , cette équation ne peut pas être réputée fournir une valeur d'aucune de ces constantes. Si, toutefois, il existe trois différentes relations théoriques impliquant ces trois constantes, on peut utiliser ces trois relations simultanément pour l'évaluation des trois constantes. Si plus que ces trois relations existent, les valeurs *les plus probables* de ces constantes sont déterminées par les moindres carrés », Birge (1932b), p.231. Voir aussi DuMond et Cohen (1948), p.82, pour une discussion à ce sujet.

33. « Certaines de ces constantes peuvent être évaluées par divers procédés. Chacun a été étudié par diverses personnes, à diverses reprises, et chaque enquête produit normalement un résultat numérique plus ou moins différent de celui de toute autre enquête », Birge (1929b), p.1.

34. Nous reprenons ici le vocabulaire de DuMond et Cohen (1948), p.82.

ici $R_{\infty(\text{or})}$, $e_{(\text{or})}$, $h_{(\text{or})}$, $c_{(\text{or})}$, et $(e/m)_{(\text{or})}$, qui soit tel que :

$$R_{\infty(\text{or})} = \frac{2\pi^2 e_{(\text{or})}^5}{h_{(\text{or})}^3 c_{(\text{or})}^2 (e/m)_{(\text{or})}} \quad (9.2)$$

Il en résulte deux questionnements : (i) peut-on dépasser le simple constat d'incohérence et évaluer (quantitativement) le *niveau* d'incohérence d'un jeu de données, et (ii) en fonction de la réponse apportée à la question (i), que peut-on alors faire d'un jeu de données incohérent ? Les réponses aux deux questions (i) et (ii) font appel à l'erreur probable (puis l'incertitude de mesure à partir des années 1970), qui joue un rôle fondamental : permettre la comparaison et la combinaison de résultats de mesure. La similarité avec la problématique rencontrée simultanément par Mayer et Euler est immédiate. Le questionnement prend sa racine dans un même constat, celui de surdétermination des systèmes d'équation, bien que la nature de cette surdétermination soit différente d'un cas à l'autre (une même « équation de condition » appliquée sur plusieurs séries d'observations dans un cas, un ensemble d'équations exprimant différentes lois physiques dans l'autre cas). La question (i) ne fait que répéter le motif qui traversait les travaux de Mayer et Euler : doit-on attribuer le désaccord constaté à une erreur dans la théorie ou à une erreur dans les mesures ? Il s'agit donc de s'interroger sur le caractère *significatif* ou non d'un désaccord entre deux résultats expérimentaux. Durant les deux siècles environ qui séparent les travaux des deux savants à ceux de Birge, cette question est devenue très classique et a déjà trouvé une réponse assez largement explorée, à savoir que l'accord doit être jugé dans les limites de l'erreur de mesure évaluée : deux résultats sont *significativement* différents si l'écart qui les sépare est sensiblement supérieur à l'erreur probable évaluée.

L'« ajustement des constantes » proprement dit est une opération qui vient répondre à la question (ii), à savoir : que peut-on alors faire d'un jeu de données incohérent ? L'idée centrale est la suivante : *si le désaccord entre résultats n'est pas significatif*, il est possible, et même souhaitable, d'*ajuster* ces données, c'est-à-dire de les *modifier*, en les mélangeant, de façon à aboutir à un unique jeu cohérent de données, lui-même en accord avec les équations du système. Il y a « ajustement » car les données expérimentales sont ajustées sur les équations du système ; ainsi, les valeurs ajustées $R_{\infty(\text{aj})}$, $e_{(\text{aj})}$, $h_{(\text{aj})}$, $c_{(\text{aj})}$, et $(e/m)_{(\text{aj})}$ des grandeurs impliquées dans l'équation (9.1) doivent être telles que :

$$R_{\infty(\text{aj})} = \frac{2\pi^2 e_{(\text{aj})}^5}{h_{(\text{aj})}^3 c_{(\text{aj})}^2 (e/m)_{(\text{aj})}} \quad (9.3)$$

L'activité d'« ajustement » des constantes implique la transformation du jeu originel de valeurs en un jeu de valeurs ajustées. Les résultats de mesure sont altérés *sans effectuer d'expérience ou de mesure supplémentaire*, par la seule réunion des différents résultats expérimentaux connus. C'est une opération de *combinaison des observations* ; il s'agit en effet de voir si, par la réunion d'informations provenant de différentes expérimentations, sur différents objets, par différentes méthodes, etc. il devient possible de *perfectionner* le jeu de données disponibles. On retrouve les considérations développées à propos de l'épisode historique décrit au chapitre 8 concernant Mayer et Euler : la réunion d'informations expérimentales quantitatives en apparence discor-

dantes peut aboutir à des résultats finaux de meilleure qualité.

Le lien avec le cas de Mayer et Euler se poursuit dans la façon dont sont implémentées les méthodes statistiques permettant de réaliser concrètement l'ajustement. Stigler a insisté sur la filiation entre l'approche de Mayer et la méthode des moindres carrés de Legendre et Gauss. C'est cette même méthode des moindres carrés qui est également mobilisée dans la démarche de Birge. Il y a une part d'arbitraire dans le choix de la méthode des moindres carrés – arbitraire que soulignait Legendre lui-même, comme mentionné au chapitre 8 – mais celui-ci est fondé sur une idée plus élémentaire encore, à savoir que l'ajustement vise à calculer quel est le jeu de valeurs *cohérentes* qui *s'éloigne le moins* des valeurs originelles. C'est ce que Cohen et DuMond appellent la « moindre violence »³⁵ faite aux données originelles. Le principe technique de minimisation des « moindres carrés » est justement une réalisation mathématique de ce principe de *moindre violence* – c'est éventuellement dans le choix de cette méthode spécifique que l'on peut trouver une forme d'arbitraire, comme ne manqueront pas de le faire remarquer certains commentateurs ultérieurs. Le travail d'ajustement apparaît donc comme la recherche d'un meilleur *compromis* entre une information empirique – les données expérimentales – et une information théorique – le réseau d'équations aux grandeurs qui structure les théories physiques dans lesquelles les constantes interviennent. Notons que le résultat de l'ajustement n'est valable qu'un temps, jusqu'à ce que de nouvelles mesures soient produites. Il n'est composé que des valeurs alors appelées « les plus probables »³⁶ eu égard aux résultats expérimentaux obtenus jusqu'alors. Il ne s'agit donc en aucun cas d'aboutir à un jeu de valeur *correct* ou *exact* – ni de croire par exemple que les valeurs ajustées s'identifieraient aux valeurs « vraies » des grandeurs étudiées – mais d'abord de répondre à une exigence de *cohérence*.

Nous détaillerons dans la section suivante le mécanisme que Birge met en place pour procéder à l'ajustement de 1929. Notons au préalable que l'étude de la façon dont Birge met en place ce mécanisme nous permet de mieux repérer le rôle qu'y joue l'incertitude de mesure. Comme nous l'avons vu, seule l'analyse d'erreur et d'incertitude permet de juger de la cohérence effective d'un jeu de valeurs. De plus, la « proximité » entre le jeu de valeurs ajustées et le jeu de valeurs originelles ne peut être mesurée que si l'on prend en compte les incertitudes associées à chacune des valeurs des constantes mises en jeu. On retrouve ici l'idée développée à la section précédente : *l'incertitude de mesure permet la comparaison entre résultats de mesure*, et ce sur un plan quantitatif. L'erreur et l'incertitude de mesure peuvent toutefois être évaluées de différentes façons. En particulier, l'analyse d'erreur se décline, une fois encore, sous une forme probabiliste et une forme déterministe. Ainsi, l'évaluation de l'erreur peut amener à une « erreur probable » ou encore à des « limites d'erreur ». À l'époque de Birge, les pratiques et les interprétations sont multiples, ce qui rend difficile la mise en commun des résultats expérimentaux, d'autant plus que les scientifiques eux-mêmes ne sont alors pas toujours conscients que manque d'uniformité méthodologique génère un grand risque d'ambiguïté. C'est ce que constate Birge,

35. Cohen et DuMond (1965), p.540.

36. Plus tard, l'expression « valeurs recommandées » lui sera préférée : « Je suggère que le CODATA ne devrait pas essayer de publier un ensemble de “meilleures” valeurs [...] Il devrait publier [...] un ensemble de valeurs “recommandées” qui seraient utilisées par les personnes dont les besoins sont satisfaits par moins que l'exactitude totale qui est potentiellement disponible », Branscomb (1971), p.7.

qui regrette un manque de rigueur dans l'exposition des résultats, dont la conséquence est l'impossibilité, pour un lecteur, de reconstituer le résultat de façon à le rendre utilisable dans un autre contexte; en particulier, « dans certains cas, on ne peut pas dire si l'erreur publiée est destinée à représenter une erreur probable ou une limite d'erreur »³⁷. Les faiblesses constatées ne se résument toutefois pas seulement à cette question certes essentielle mais toutefois *conventionnelle* de mise en forme des résultats par tous sur un même mode; sur un deuxième plan, c'est également la qualité des pratiques employées qui est remise en question :

(D)ans de nombreux cas, il est tout à fait impossible de déterminer, à partir d'un article imprimé, quelles valeurs des constantes auxiliaires ont été effectivement adoptées. Souvent, il n'est même pas possible de déterminer la formule utilisée pour calculer le résultat final. [...] L'auteur a trouvé dans la littérature des cas où une part appréciable de l'exactitude, obtenue durant des années par le plus habile des travaux expérimentaux, a été sacrifiée, bien sûr sans le vouloir, par des méthodes de calcul approximatives ou mal avisées. [...] Dans de nombreux cas, aucun réexamen n'était possible, puisque les données originales nécessaires pour un tel calcul n'avaient pas été publiées.³⁸

Le travail de Birge s'appuie donc sur une revue critique poussée des travaux de ses contemporains³⁹, ce qui met en évidence le rôle d'analyse critique des ajustements, au-delà même de leur objectif initial, à savoir produire un ensemble cohérent de valeurs ajustées. Dans la section suivante, nous laisserons de côté ce travail critique et verrons plus en détail le mécanisme qu'emploie Birge pour réaliser l'ajustement. Cela nous permettra de décrire le rôle qu'y joue l'incertitude de mesure.

9.3 Procédure d'ajustement : l'approche de Birge

La problématique méthodologique et épistémologique qu'aborde Birge est la conséquence de deux facettes d'un même problème : d'un côté, il faut faire la conjonction de mesures différentes d'une même constante; d'un autre côté, il faut ajuster entre elles les valeurs des différentes constantes au réseau d'équations qui les relie. Birge choisit de réévaluer les constantes une par une, en commençant par la vitesse de la lumière dans le vide, qu'il est possible de déterminer indépendamment des autres. Il procède ensuite par itération, en répétant ce procédé pour chaque constante, et en utilisant si nécessaire les valeurs ajustées des constantes précédentes qu'il vient de déterminer⁴⁰. Birge désignera plus tard cela comme une méthode « par approximations successives »⁴¹, par opposition à un ajustement simultané des constantes (qu'il cherchera à développer dans un article ultérieur, en 1932). Il est intéressant de constater

37. Birge (1929b), p.3.

38. Birge (1929b), pp.3-4.

39. « L'article était ouvertement critique envers le travail des autres, et Birge dit dans des interviews qu'il aurait bien pu ne plus avoir d'ami une fois que celui-ci avait été publié », Helmholtz (1990), p.78.

40. Birge (1929b), p.3. Les constantes calculées précédemment jouent alors le rôle de constante « auxiliaire » dans le calcul en cours.

41. Birge (1932b), p.228.

qu'en suivant ce protocole, la méthode des moindres carrés elle-même n'intervient que de façon tangentielle. Birge en propose un bref résumé⁴², dont il profite pour expliciter en passant les raisons qui motivent l'adoption de cette méthode précise plutôt que les alternatives parfois plébiscitées par certains de ses contemporains⁴³.

Pour illustrer la façon dont Birge mène à bien la procédure d'ajustements, prenons pour commencer l'exemple de la célérité c de la lumière dans le vide avec lequel Birge débute son étude. Ce dernier en recense les mesures les plus remarquables ; ce sont les travaux de Michelson qui font alors référence :

La dernière et la plus précise des déterminations directes de la vitesse de la lumière est celle de Michelson, en 1921-26. De fait, cette investigation surpasse tant en exactitude ce qui précède que seule celle-ci doit être prise en compte.⁴⁴

Birge repère également d'autres travaux plus récents, qui déterminent c de façon indirecte. D'une part :

La vitesse des ondes électriques stationnaires courtes a également été mesurée et montre un accord approximatif avec la vitesse de la lumière, dans d'assez grandes limites d'erreur. Le meilleur travail à ce sujet est probablement celui de Mercier.⁴⁵

D'autre part :

La vitesse des ondes électromagnétiques a été obtenue indirectement par le rapport du système électrique d'unités électrostatique (es) avec le système électromagnétique (em), selon la théorie électromagnétique de la lumière généralement acceptée. La meilleure valeur de ce rapport, qui est ici désignée par c' , est indubitablement celle trouvée par Rosa et Dorsey.⁴⁶

Les trois résultats sont représentés sur le graphe de la figure 9.1. Sans surprise, ces trois déterminations n'aboutissent pas à des valeurs rigoureusement égales. C'est ici que l'erreur probable joue son premier rôle d'outil de comparaison des résultats de mesure. Le résultat de Rosa et Dorsey est en « bel accord »⁴⁷ avec celui de Michelson, ce qui peut être constaté de façon simple – les « barres d'erreurs » des deux résultats se recouvrent – ou pourrait être explicité de façon plus poussée, par un test statistique. À l'inverse, celui de Mercier diffère *significativement* des deux résultats précédents : l'écart avec le résultat de Michelson, par exemple, est de vingt-quatre fois l'erreur probable sur ce dernier. De ce fait, il n'est pas possible d'immédiatement invoquer la *combinaison des observations* et d'adopter une valeur finale calculée par un mélange des trois résultats présents. Il faut d'abord, au préalable, opérer un *arbitrage* entre les

42. Birge (1929b), pp.4–6.

43. Notons cependant que, alors que la méthode générale des moindres carrés sert usuellement à ajuster *plusieurs* grandeurs à des données expérimentales selon une loi qui relie ces différentes grandeurs (comme dans le cas de la régression linéaire que nous avons décrite à la section 8.6), Birge ne l'exploite, dans son article de 1929, que dans le cas restreint de l'étude *d'une seule* grandeur. La méthode des moindres carrés n'intervient en effet que pour justifier l'adoption d'une moyenne arithmétique pondérée et dériver la valeur de l'erreur probable associée à cette moyenne.

44. Birge (1929b), p.9.

45. Birge (1929b), p.9.

46. Birge (1929b), p.10

47. Birge (1929b), p.10.

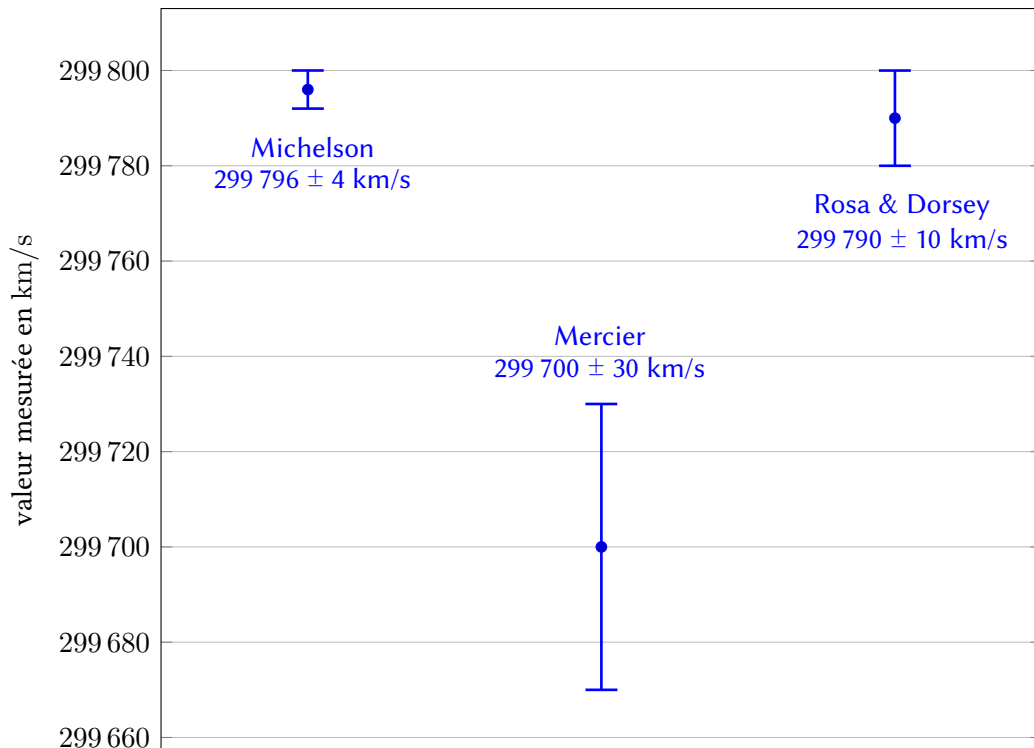


FIGURE 9.1 – Mesures de la vitesse de la lumière dans le vide c recensées par Birge en 1929. Le résultat de Mercier est sensiblement différent des deux autres, ce qui fait dire à Birge que l'erreur probable a été sous-estimée.

deux jeux de valeurs. Plusieurs arguments entrent alors en compte : un résultat est d'autant plus crédible que son erreur probable est faible et qu'il est en accord avec d'autres résultats obtenus indépendamment par d'autres. Il s'agit là de paramètres qui sont contenus dans l'expression numérique des résultats. Mais Birge ne s'en tient pas à ces seuls arguments ; il note également que le résultat de Mercier est fragilisé par « des incertitudes dans la théorie »⁴⁸, alors qu'à l'inverse, l'erreur probable évaluée par Rosa et Dorsey lui semble très prudente – trop prudente même, et, par conséquent, vraisemblablement surestimée. Sur ces bases, Birge écarte le résultat de Mercier. L'étude de l'ajustement de la vitesse de la lumière dévoile ainsi une opération qui constitue l'une des deux opérations essentielles de l'activité d'ajustement : il s'agit de la *sélection* des données d'entrée. Si l'ajustement des constantes peut être vu comme une opération de mise en commun des résultats expérimentaux (qui prend la forme d'une version évoluée de moyennage), il n'est pas pour autant la mise en commun de *tous* les résultats dans une même moyenne. Il est nécessaire, au préalable, de s'interroger individuellement sur la validité des résultats recensés. Birge le fait remarquer dès son introduction :

Si une observation particulière dévie trop grandement d'une courbe lisse, elle de-

48. Birge (1929b), p.9.

vrait être rejetée avant de tenter de traiter les données par les moindres carrés.⁴⁹

Toutefois, Birge ne pousse pas la réflexion plus loin. Il n'explicite pas ce que signifie, pour deux résultats, d'être en désaccord significatif. Nous verrons cependant au chapitre 10, en nous intéressant à des épisodes ultérieurs de l'histoire des ajustements des constantes de la physique, que l'opération de sélection – ou d'arbitrage entre les données – est particulièrement sujette à discussion, d'autant qu'elle n'est pas décidable sur des critères simplement mathématiques, et fait donc appel à un élément de jugement qui peut passer pour une forme d'arbitraire.

En fin de compte, Birge choisit d'adopter directement pour valeur finale celle obtenue par Michelson; celui de Rosa et Dorsey n'est pas écarté, mais est néanmoins négligé en raison de sa faible précision⁵⁰. Dans le cas de la vitesse de la lumière dans le vide, il n'y a pas d'ajustement « aux moindres carrés » à proprement parler, simplement un arbitrage entre résultats.

Un exemple plus complexe est donné par le cas de la constante de Planck h , qui vient en treizième position dans l'ordre dans lequel Birge ajuste les constantes. Nous représentons sur la figure 9.2 les résultats recensés pour la constante de Planck. Ces résultats ont été préalablement retravaillés par Birge, qui ne se contente pas des valeurs brutes qui lui ont été communiquées. En particulier, il se trouve que les mesures de h sont indirectes; à chaque fois, la détermination de h fait intervenir la connaissance de la charge élémentaire e . Ce n'est pas directement h qui est mesuré, mais un rapport h/e^α , où α est un coefficient qui varie selon le principe de mesure. À titre d'exemple⁵¹, l'effet photoélectrique permet de mesurer le rapport h/e ; la spectroscopie à rayons X permet quant à elle de mesurer le rapport $h/e^{4/3}$. Mais si les laboratoires fournissant les mesures de h emploient chacun une valeur différente de la charge élémentaire e , cela a des conséquences sur l'ajustement. Par conséquent, Birge recalcule⁵² les différentes valeurs d'entrée recensées pour h à partir de la valeur ajustée qu'il a lui-même obtenue pour e .

Dans le cas de h , les résultats sont tous en bon accord dans les limites des erreurs probables évaluées. Par conséquent, il n'est pas nécessaire, dans ce cas précis, de passer par une étape d'arbitrage; au contraire, tous les résultats mentionnés peuvent être intégrés au calcul de la valeur ajustée. Contrairement au cas de la vitesse de la lumière, on a bien ici une procédure de mélange, qui fait intervenir la méthode des moindres carrés (sur un cas élémentaire, l'ajustement à une loi constante $h = a$). La valeur ajustée est le produit d'une *combinaison des observations*; cependant, elle n'est pas simplement la moyenne arithmétique des différents résultats recensés. Pour la calculer, Birge fait appel à un principe devenu classique, qui découle de la modélisation probabiliste de la mesure: la combinaison des différents résultats doit être effectuée selon une moyenne *pondérée*, le poids accordé à chaque valeur étant inversement proportionnel au carré de son erreur probable. Ainsi, si nous appelons h_i les six différentes valeurs

49. Birge (1929b), p.6.

50. Birge (1929b), p.10. Comme nous le verrons par la suite, ce choix est justifié par la méthode de pondération des résultats dans l'ajustement. Lorsque l'incertitude de mesure associée à une valeur est élevée devant celles des autres valeurs exploitées pour l'ajustement, la contribution de cette valeur au résultat ajusté est très faible. Au-delà d'un certain seuil, cette contribution est donc négligée. Cette opération est cependant à distinguer nettement de l'opération de *sélection* qui consiste à écarter *préalablement* certains résultats de l'ajustement.

51. Birge (1929b), pp.51–54.

52. Voir par exemple le début de la page 54 dans Birge (1929b), qui montre également que les manipulations de Birge ne se limitent pas à remplacer e par sa valeur ajustée.

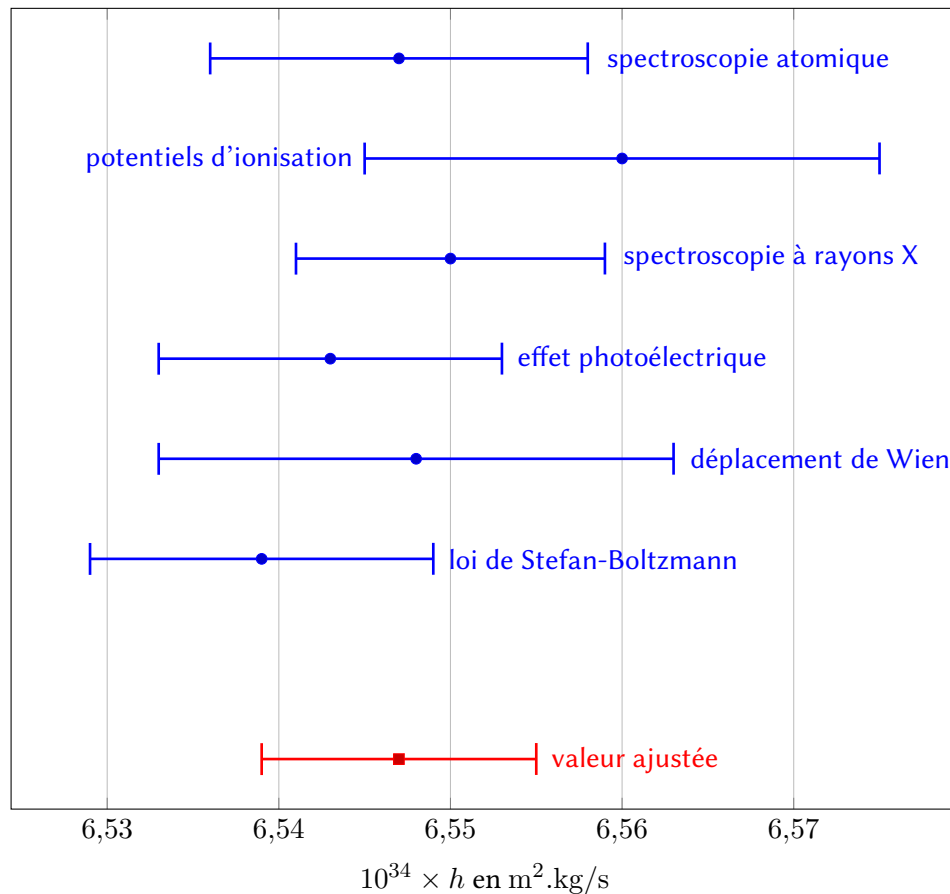


FIGURE 9.2 – Mesures de la constante de Planck h recensées par Birge en 1929 (en bleu, les valeurs centrales sont représentées par des ronds) et valeur ajustée (en rouge, représentée par un carré). Les résultats sont tous en bon accord, et la valeur ajustée est la moyenne pondérée des valeurs individuelles. L'erreur probable sur la valeur ajustée dépend à la fois de la dispersion des différentes valeurs et de l'erreur probable sur la charge élémentaire e , dont h dépend directement.

recensées par Birge, et u_i les erreurs probables respectives, nous avons :

$$h_{\text{ajustée}} \propto \sum_{i=1}^6 \frac{h_i}{u_i^2} \quad (9.4)$$

La somme des coefficients de pondération devant être égale à 1, nous avons alors l'expression complète suivante :

$$h_{\text{ajustée}} = \frac{\sum_{i=1}^6 \frac{h_i}{u_i^2}}{\sum_{i=1}^6 \frac{1}{u_i^2}} \quad (9.5)$$

L'équation (9.5) peut être interprétée comme une conséquence de la méthode des moindres carrés, ce que Birge s'attachera à démontrer dans un article méthodologique publié trois ans plus tard⁵³. Elle demeure aujourd'hui une formule classique du calcul d'incertitude. En particulier, elle est toujours à la base des processus d'ajustement du début du XXI^e siècle. La justification des coefficients de pondération tire sa justification de la modélisation probabiliste de la mesure ; nous la brosons très succinctement dans l'annexe A.5. Ce choix de pondération apparaît comme la meilleure façon de réunir les informations contenues dans des résultats de mesure dont la précision est jugée comme différente. De ce fait, l'équation (9.5) vient *étendre* le principe de « combinaison des observations » en y adjoignant une idée essentielle. En effet, la combinaison des observations effectuée par Mayer en 1760 est une combinaison simple, sans pondération ; celle-ci n'est possible que parce que Mayer a effectué lui-même les observations et est en mesure de juger qu'elles sont toutes également précises. Birge exploite quant à lui des mesures qui ne sont pas toutes également précises, ce qui se traduit par des erreurs probables variées. En utilisant les erreurs probables pour déterminer les poids de chaque valeur dans l'ajustement, l'équation (9.5) exprime une combinaison des observations bien plus élargie encore, puisqu'elle permet d'intégrer des résultats obtenus dans des conditions entièrement différentes, tout en prenant acte du fait que ceux-ci ne sont pas nécessairement tous effectués avec une même précision. Mayer avait proposé ce qui apparaît rétrospectivement comme une extension de la combinaison des observations depuis des « conditions de répétabilité » strictes à des « conditions de reproductibilité » restreintes ; l'équation (9.5) exprime comment élargir encore cette combinaison des observations à des conditions de reproductibilité larges⁵⁴, à la condition expresse d'introduire un raisonnement suffisamment élaboré, et quantitatif, sur l'erreur probable (aujourd'hui l'incertitude de mesure) relative aux résultats considérés. Nous voyons une fois encore en quoi l'erreur probable est un ingrédient indispensable à la formulation d'un résultat de mesure si l'on souhaite que celui-ci puisse être communiqué et utilisé par d'autres.

Précisons que Birge calcule également l'erreur probable associée à la valeur ajustée. Celle-ci provient de la combinaison de deux termes. En effet, comme nous l'avons rappelé un peu plus tôt, les différentes mesures de h sont obtenues indirectement, à partir de la mesure d'un rapport h/e^α où le coefficient α dépend du principe de mesure. Une fois mesuré le rapport

53. Birge (1932a)

54. « (C)ette façon de calculer [...] donne la possibilité [...] de réunir les connaissances obtenues dans des expériences parfois très différentes », Protassov (2002), p.101.

h/e^α , la connaissance de e (par une mesure séparée) permet de calculer h . Ainsi, les différentes mesures de h sont dites « corrélées » (elles dépendent toutes d'une même détermination de e) et le calcul de l'erreur probable sur la valeur ajustée de h doit intégrer une composante liée à la mesure de e . Birge décide de simplifier le problème en utilisant un coefficient α moyen sur l'ensemble des méthodes utilisées⁵⁵, et aboutit à l'écriture de h comme le produit de deux termes :

$$h = \left(\frac{h}{e^{1,26}} \right) \times e^{1,26} \quad (9.6)$$

Ce sont les mesures de $h/e^{1,26}$ qui peuvent être considérées comme indépendantes entre elles. On trouve alors, selon un calcul classique de propagation des erreurs⁵⁶, que :

$$\frac{u(h)}{h} = \sqrt{\left(\frac{u(h/e^{1,26})}{h/e^{1,26}} \right)^2 + \left(1,26 \frac{u(e)}{e} \right)^2} \quad (9.7)$$

Où les $u(h)$, $u(h/e)$ et $u(e)$ sont les erreurs probables respectives associées à h , h/e et e . L'erreur probable associée à h est donc la composition de deux termes, le premier étant déterminé à partir de la dispersion des mesures de $h/e^{1,26}$, que l'on peut retrouver à partir de la dispersion des valeurs de h représentées figure 9.2; et le second étant déterminé lors de l'ajustement de la constante e que Birge a mené plus tôt dans son article. C'est ainsi qu'est calculée l'erreur probable associée la valeur ajustée de h , représentée figure 9.2.

9.4 Incertitude de mesure et évaluation d'un résultat

Les travaux de Birge nous donnent à réfléchir, en retour, sur ce que représente l'incertitude de mesure, et, en particulier, sur ce qu'elle *évalue* quant à un résultat de mesure. La méthode de pondération présentée à la section précédente illustre l'idée courante selon laquelle l'incertitude de mesure représente la *confiance* que l'on accorde à un résultat de mesure. En effet, plus un résultat présente une incertitude faible, plus son poids dans l'ajustement sera élevé. Birge lui-même voit l'erreur probable comme une mesure de la « fiabilité » du résultat⁵⁷. Nous pourrions dire, en suivant ce constat, que l'incertitude de mesure permet de rendre *également utilisables* deux résultats dont on a initialement une confiance *inégalement*. Ainsi, deux résultats compatibles dont la précision diffère peu peuvent être tous deux exploités dans l'ajustement, tant que leurs incertitudes de mesure associées ont été déterminées de façon rigoureuse et réaliste, et à condition d'adopter la méthode de pondération adaptée; et il n'y a aucune raison de devoir *choisir* à quel résultat s'en tenir. Les deux résultats sont rendus *également fiables*, au sens où il est aussi sage de s'appuyer sur l'un que sur l'autre, pour peu qu'on aie conscience des incertitudes associées, et ce même si le résultat le plus précis se révélera le plus efficace (ou, pour

55. L'exposant α prend alternativement les valeurs 1 (pour trois résultats), 4/3 (pour deux résultats) et 5/3 (une seule fois). L'exposant moyen vaut donc 1,26. Voir Birge (1929b), p.57.

56. Il s'agit ici d'appliquer la formule classique de « propagation des erreurs » (devenue par la suite « propagation des incertitudes » selon la dénomination employée dans le GUM) qui s'appuie sur les propriétés de l'algèbre des variables aléatoires. Nous renvoyons à la partie I pour le détail de ces aspects techniques.

57. « Maintenant, l'erreur probable une quantité mesurée peut être appelée une mesure absolue de sa fiabilité », Birge (1932a), p.216.

reprendre une métaphore de Barry Taylor, le plus « aiguisé »⁵⁸) dès lors qu'il s'agit tester une théorie ou de prendre une décision.

Cependant, le terme de *fiabilité* employé notamment par Birge porte à considération. En effet, si l'incertitude de mesure est un indicateur de la fiabilité d'un résultat, cette seule donnée devrait suffire pour juger s'il est opportun ou non d'accepter la validité du résultat. Certes, les choses sont relativement simples dans l'exemple précédent de la constante de Planck, car les résultats sont tous en bon accord, et l'on peut alors appliquer directement, et sans questionnement, le critère mathématique de pondération pour obtenir une valeur ajustée qui n'est pas *a priori* problématique. Que faire, cependant, lorsque certains résultats présentent un désaccord significatif, au-delà des marges d'erreur ? Nous avons vu que le problème s'est présenté à Birge dès la première constante étudiée – la vitesse de la lumière. Birge a choisi de passer par une opération de *sélection*, dont nous avons brièvement discuté des tenants à la section 9.3 ; dans ce cas précis, celle-ci s'est révélée raisonnablement simple en ce que l'un des résultats semblait nettement moins crédible. Cependant, l'information que Birge a mise à contribution pour juger de la crédibilité du résultat de Mercier n'était *pas* contenue dans le résultat numérique lui-même, et n'était pas liée à l'amplitude de l'erreur probable correspondante. Ainsi, l'arbitrage des données adéquates à l'ajustement passe par une évaluation *supplémentaire*, indépendante de l'amplitude de l'incertitude de mesure elle-même, et qui repose sur un jugement *externe* du résultat de mesure dans son ensemble. Or, s'il est nécessaire d'ajouter à l'analyse d'incertitude une évaluation externe supplémentaire de la qualité ou de la fiabilité du résultat, c'est que l'incertitude de mesure *échoue* à représenter elle-même ces caractères essentiels. La remarque que nous soulevons ici est au fondement d'une difficulté assez sérieuse qui va poursuivre les physiciens durant toute l'histoire des ajustements. Elle n'est toujours pas résolue au début du XXI^e siècle, et a atteint un sommet en 1970 lorsque, dans une conférence organisée au NIST, les métrologues et physiciens ont discuté de la légitimité des ajustements aux moindres carrés. Dans le chapitre suivant, nous développons plus en détail cet épisode et nous nous attachons à montrer qu'il a nourri des discussions techniques et épistémologiques très riches ; sa résolution a demandé une maturation plutôt longue et n'a pas nécessairement emporté l'unanimité.

Notre analyse nous montre en fin de compte qu'il existe, dans le champ disciplinaire des ajustements des constantes de la physique, *deux* évaluations de la fiabilité d'un résultat : une évaluation interne, quantifiée par l'incertitude de mesure, et qui indique comment exploiter le résultat *une fois que celui-ci a été accepté* ; et une évaluation externe, qui indique s'il faut ou non accepter le résultat. En ce sens, l'incertitude de mesure ne vient pas qualifier un résultat de mesure ; plutôt, elle *fait partie intégrante* du résultat. De fait, nous avons vu, dans la partie I consacrée aux approches probabilistes de la mesure, que l'évolution des méthodes statistiques en métrologie nous mène désormais à considérer le résultat comme l'ensemble constitué *à la fois* de la valeur centrale et de l'incertitude de mesure, et qu'un tel résultat peut être formulé de diverses façons, par exemple directement par un intervalle d'incertitude, ou encore par une distribution de probabilité ; autant d'approches qui n'accordent aucun statut particulièrement privilégié à la valeur centrale. Si l'on considère qu'un résultat de mesure complet n'est pas la

58. Taylor (1971)

donnée d'une unique valeur, mais est la donnée d'un ensemble de valeurs raisonnables, caractérisées par une valeur centrale et une incertitude, ou par un intervalle, ou par une distribution de probabilité, alors l'incertitude de mesure ne peut plus être interprétée comme une évaluation de la fiabilité ou de la qualité du résultat. L'incertitude de mesure ne peut pas être à la fois un constituant interne du résultat lui-même et un jugement externe sur le résultat.

Doit-on considérer qu'un résultat expérimental est de bonne qualité dès lors que l'incertitude de mesure associée est faible ? Ce n'est certainement pas de cette façon que le concept est interprété dans la pratique des ajustements. Ce n'est pas l'amplitude de l'incertitude de mesure elle-même qui permet de déterminer avec quelle forme un agent doit avoir ou non confiance en un résultat de mesure ; c'est plutôt l'assurance que cette incertitude de mesure a été *correctement* évaluée. Dès lors, le jugement ne porte plus sur le résultat lui-même mais sur l'évaluateur. Cela nous renvoie de fait au désaccord interprétatif – décrit à la partie I qui oppose les tenants d'une approche fréquentiste objective de la mesure aux défenseurs d'une approche bayésienne, qui pour de nombreux métrologues prend la forme d'une épistémologie subjectiviste. En associant un intervalle d'incertitude à un taux de succès effectif, les fréquentistes souhaitent rendre solidaires les évaluations interne et externe de la fiabilité d'un résultat de mesure ; comme semble nous indiquer le présent chapitre, et comme le confirmera notre prochain chapitre, cette exigence ne correspond pas à la conception de l'incertitude de mesure qui est adoptée dans la pratique des ajustements. À l'inverse, les bayésiens omettent d'expliquer comment rendre compte de l'évaluation externe de la fiabilité d'un résultat : l'incertitude de mesure représente peut-être le degré de croyance de l'expérimentateur quant à la valeur obtenue, mais avec quel degré de croyance un *autre* expérimentateur qui souhaiterait faire usage dudit résultat doit-il considérer *à la fois* la valeur obtenue *et* l'incertitude associée au résultat ?

L'analyse que nous menons ici nous fait pressentir que les physiciens impliqués dans les ajustements ne prennent pas tout à fait la même direction que les métrologues en charge de la rédaction des guides tels le GUM. Nous voyons que la signification de l'incertitude de mesure n'est pas figée, mais dépend des *usages* de l'incertitude en fonction des disciplines scientifiques, ainsi que de l'interprétation des probabilités qui accompagne ces usages. Nous reprendrons ce questionnement plus en détail dans le chapitre 11 conclusif, où nous reviendrons sur les rapports entre incertitude de mesure et confiance, et où nous serons renvoyés à un point essentiel qui traverse notre questionnement de bout en bout : la relation entretenue entre *incertitude* et *exactitude* de mesure. Auparavant, nous souhaitons développer un deuxième épisode de l'histoire des ajustements des constantes de la physique, en remontant jusqu'au début des années 1970, où ont été discutées des problématiques liées au traitement des données discordantes ; ces problématiques font remonter une fois encore les questions de sélection et de mélange, et le rôle et la signification de l'incertitude de mesure s'y retrouvent explicitement réexaminés.

Chapitre 10

Le traitement des données discordantes dans les ajustements des constantes de la physique

Dans ce chapitre, nous poursuivons l'analyse des ajustements des constantes de la physique en constatant la façon dont l'initiative de Birge est progressivement devenue une tradition d'ajustement, puis a été centralisée en étant intégrée à une structure institutionnelle internationale. Nous décrivons donc l'évolution des travaux d'ajustement pendant les quarante années qui succèdent le travail initial de Birge, et nous nous attardons en particulier sur l'état des lieux de la pratique en 1970. L'analyse de l'article de Birge de 1929 nous a permis de dévoiler l'existence d'une tension entre sélection et mélange, tension dont Birge ne prend pas véritablement acte. Ce chapitre montre que cette tension deviendra de plus en plus problématique au fur et à mesure que les ajustements prendront de l'ampleur et deviendront un enjeu scientifique et institutionnel. En décrivant les discussions épistémologiques et méthodologiques qui ont accompagné l'évolution des ajustements des constantes, nous souhaitons montrer que l'opposition entre sélection et mélange est liée à l'usage qui est fait de l'incertitude de mesure dans les ajustements, et à l'interprétation qui lui est accordée. Ainsi, nous en viendrons à conforter l'idée, ébauchée dans le chapitre précédent, selon lequel l'incertitude de mesure n'est pas une évaluation de la qualité d'un résultat ou de la confiance que l'on porte en un résultat.

La section 10.1 décrit comment les physiciens ont peu à peu repris l'idée initiale de Birge, d'abord sur la base d'initiatives personnelles, puis en donnant à ce travail une dimension institutionnelle avec son intégration au CODATA à la fin des années 1960.

Cependant, la pratique des ajustements a également été critiquée à différentes reprises durant cette période. C'est l'objet de la section 10.2. En particulier, l'ajustement présente une étape *sélection* des données d'entrée qui est perçue négativement par certains physiciens. D'une part, elle revient à trancher entre plusieurs résultats en apparence également fiables. D'autre part, elle provoque une sous-estimation des incertitudes sur les valeurs ajustées qui s'observe dans la fluctuation de ces valeurs d'un ajustement à l'autre. Cela amène certains physiciens à avancer l'idée que les valeurs obtenues elles-mêmes ne sont pas l'objectif le plus important des

ajustements.

La section 10.3 revient plus en détail sur cet aspect en décrivant en particulier deux positionnements qui ont été confrontés lors d'une conférence menée en 1970 à propos des ajustements. P.L. Bender préconise que les résultats des ajustements doivent être autant que possible exacts, en conséquence de quoi la présence de résultats discordants devraient amener les physiciens à élargir leurs incertitudes. B.N. Taylor affirme quant à lui que cela reviendrait à amoindrir la fécondité des ajustements, dont les vertus épistémiques principales ne résident pas tant dans les valeurs produites mais dans le travail d'analyse et de recherche des erreurs qui accompagne la procédure.

Nous montrons alors, dans la section 10.4, que cette problématique reste aujourd'hui ouverte au sein du CODATA, comme le montre l'exemple très récent de l'« énigme du rayon du proton ».

10.1 Une tradition d'ajustements

En 1932, trois ans après son premier article, Birge publie une nouvelle tentative d'ajustement de quatre constantes physiques : la charge élémentaire e , la constante de Planck h , le rapport e/m entre la charge élémentaire et la masse de l'électron, et enfin la constante de structure fine α . Cet article intervient alors qu'a lieu une intense discussion quant à la valeur de la constante α , en particulier autour des considérations théoriques développées par Arthur Eddington. Eddington affirmait alors depuis quelques années avoir obtenu, par une dérivation entièrement théorique, la valeur *exacte* de cette constante (plus précisément de son inverse α^{-1} qu'il tenait pour un nombre entier¹) – et ce à l'aide de la récente mécanique quantique. Ce ne sont pas ces aspects historiques² qui nous intéresseront ici ; notons simplement que les travaux d'Eddington avaient été accueillis avec scepticisme par la communauté scientifique, dont Birge, lequel n'avait pas manqué de pointer l'écart significatif de cette prédiction théorique avec les meilleurs résultats expérimentaux³. Suite à ces réactions, Wilfrid Bond, un physicien spécialiste d'hydrodynamique, avait publié plusieurs articles tendant à montrer une correspondance entre les prédictions d'Eddington et certaines déterminations expérimentales⁴. C'est dans ce contexte que Birge publie en 1932 son second article d'ajustements. Il y reconnaît la fécondité du principe général qui sous-tend les calculs de Bond, mais n'adhère pas au détail du raisonnement, et rejette les conclusion que Bond en tire⁵. Malgré les réserves émises par Birge, la

1. Eddington avait d'abord prédit $\alpha^{-1} = 136$, puis, reconnaissant le désaccord expérimental pointé par la communauté des physiciens, avait ensuite conclu $\alpha^{-1} = 137$.

2. Helge Kragh a documenté l'histoire de la constante de structure fine, et en particulier les débats concernant Eddington et la communauté scientifique ; voir Kragh (2003).

3. Birge (1929b), p.72. Anticipant la publication de son article d'ajustements, Birge avait également publié dans *Nature* une très brève critique du résultat d'Eddington, voir Birge (1929a).

4. Bond (1930), Bond (1931).

5. « Le principe majeur qui sous-tend la nouvelle méthode de calcul a été suggéré par Bond, mais dans ses deux articles, Bond ne parvient pas à mener la théorie à sa conclusion logique. L'auteur n'est pas d'accord du tout avec le choix de données de Bond, ni avec sa pondération des données, et c'est pour cette raison qu'un réexamen du matériau disponible a été effectué. Tous les calculs de Bond sont menés pour des méthodes de moindres carrés, mais dans certains cas il semble avoir utilisé des formules qui ne sont pas applicables à la situation », Birge (1932b), p.231.

méthodologie développée dans les articles de Bond lui semble suffisamment féconde pour qu'il décide d'en reprendre l'idée centrale, dans le but de réviser la façon dont il s'était lui-même attaqué aux ajustements dans son article de 1929. Le principal objectif est alors pour Birge de trouver un substitut à la très imparfaite méthode « par approximations successives », dont un défaut majeur est de donner un ordre de priorité aux constantes en les évaluant tour à tour. En lieu et place, la méthode de Bond permet une évaluation *simultanée* des différentes constantes, que Birge reprend à son compte. Si l'article de 1929 peut être vu comme l'initialisation d'une *idée* générale, celle des ajustements, l'article de 1932 peut quant à lui être considéré comme la mise en place d'une *technique* spécifique qui sera reprise par d'autres scientifiques après lui. Du fait du changement de méthode employée, l'ajustement de 1932 marque une différence importante avec celui de 1929 en ce qu'il emploie la méthode des moindres carrés de façon plus substantielle. Dans l'article de 1929, celle-ci ne servait qu'à justifier l'expression de l'erreur probable associée à la mesure répétée d'une même grandeur dans des conditions différentes. C'est seulement avec la méthode d'évaluation simultanée de l'article de 1932 que Birge fait appel à une méthode de régression linéaire qui exploite pleinement la méthode des moindres carrés, et qui deviendra par la suite le standard pour les pratiques d'ajustements des constantes de la physique. Les travaux de Birge aboutissent en particulier au développement d'une méthode graphique s'appuyant sur des diagrammes qui seront par la suite appelés diagrammes de « Birge-Bond »⁶ et qui seront, durant un temps, l'un des outils centraux des procédures ajustements des constantes.

Dans son article de 1929, Birge décrivait explicitement son travail comme une simple tentative, un premier essai dont l'objectif était avant tout de pointer certaines lacunes méthodologiques dans les travaux de ses contemporains, et de stimuler l'harmonisation des pratiques d'analyse d'erreur. Cette initiative va connaître un grand succès. Si Birge lui-même publie plusieurs autres ajustements jusqu'en 1945, ses travaux sont également réappropriés et réinvestis par différents scientifiques, et diffusent assez rapidement dans la communauté scientifique. L'on voit ainsi peu à peu apparaître une *tradition* d'ajustements, d'abord présente sous la forme d'une somme de productions à l'échelle individuelle. Mentionnons, dans une liste non exhaustive, le travail de Dunnington en 1939⁷, ceux de Cohen et DuMond en 1948, 1952, 1955 (avec Layton et Rollett), 1958 et 1965⁸, ainsi que ceux de Bearden, avec Watts en 1951, et avec Thomsen en 1957⁹, et enfin celui de Taylor, Parker et Langenberg en 1969¹⁰. Mais ce qui est encore plus ou moins de l'ordre de l'initiative personnelle¹¹ au milieu du XX^e siècle prend un tour

6. Voir par exemple [Dunnington \(1939\)](#) ou encore [DuMond et Cohen \(1948\)](#).

7. [Dunnington \(1939\)](#)

8. [DuMond et Cohen \(1948\)](#), [DuMond et Cohen \(1953\)](#), [Cohen et al. \(1955\)](#), [Cohen et DuMond \(1965\)](#)

9. [Bearden et Watts \(1951\)](#), [Bearden et Thomsen \(1957\)](#)

10. [Taylor, Parker et Langenberg \(1969\)](#)

11. L'ajustement mené par Cohen et DuMond en 1965 avait été quant à lui commandé par un comité du Conseil national de la recherche des États-Unis, le "joint committee of the Divisions of Chemistry and Physics of the U.S. National Research Council", comme indiqué dans le compte-rendu, [Cohen et DuMond \(1965\)](#), p.538. Celui publié par Taylor, Parker et Langenberg était une initiative personnelle des auteurs, et, de l'aveu du premier nommé, se situait « totalement en dehors de la structure normale [*totally outside of the normal structure*] » des ajustements; Taylor indique d'ailleurs sur le ton de l'humour que les auteurs pouvaient être considérés, en un certain sens, comme des « intrus [*interlopers*] ». Taylor et Cohen furent ensuite amenés à collaborer au sein du CODATA pour les ajustements

nouveau à la fin des années 1960 avec la mise en place d'une collaboration internationale, le CODATA, qui prend désormais en charge la pratique d'ajustement. Le CODATA, pour "Committee on Data for Science and Technology" est un organisme créé en 1966 sous l'égide du Conseil International des Unions Scientifiques (ICSU, lui-même fondé en 1931, désormais appelé « Conseil International pour la Science ») pour répondre à l'augmentation massive des flux de données quantitatives constatée depuis un siècle¹². Le cas des constantes de la physique y fait l'objet d'un groupe de travail spécifique, le « groupe de travail sur les constantes fondamentales »¹³, dont la création se situe entre 1968 et 1970¹⁴. Sous l'impulsion du CODATA, le groupe de travail publie un nouvel ajustement en 1973¹⁵; le rythme des ajustements ralentit alors, les deux publications suivantes étant en 1986¹⁶ puis 1998¹⁷. À l'occasion de l'ajustement de 1998, il est décidé d'adopter un rythme régulier et soutenu d'ajustements, à raison d'un ajustement tous les quatre ans¹⁸, ce qui est rendu possible, en particulier, par les progrès de l'informatique et des méthodes de calcul numérique¹⁹.

Nous voyons ainsi en quoi l'initiative personnelle de Birge, reprise par différents scientifiques, a mené à une tradition d'ajustements qui s'est progressivement structurée et standardisée jusqu'à devenir institutionnalisée, et, de ce fait, *centralisée*. L'héritage de Birge se fait en particulier sentir selon trois apports significatifs issus des deux articles de 1929 et 1932 :

- (1) L'idée même de mettre en commun un maximum de connaissances expérimentales relatives aux constantes physiques. Ce faisant, Birge met en avant le bénéfice que la communauté scientifique trouverait à adopter une même méthode pour calculer les erreurs probables et exprimer les résultats de mesure. Ce plaidoyer sera répété à plusieurs reprises, par exemple chez Cohen et DuMond²⁰ ou encore chez Cohen et Taylor²¹;
- (2) L'insistance portée sur l'importance d'un concept comme l'« erreur probable », auquel est attaché une signification *statistique*, en faveur d'un calcul sur les « limites d'erreur ». Ce constat est formulé un demi-siècle avant que la tension entre les deux pratiques ne

suivants, après avoir été mis en contact par l'entremise de Lewis Branscomb, alors directeur du NIST et éditeur de la revue *Reviews of Modern Physics*. Ces informations ont été collectées lors d'un entretien (non publié) avec Taylor [Taylor (2013)].

12. Lide et Wood (2012), p.1.

13. Ce groupe de travail a été nommé "CODATA Task Group on Fundamental Constants" jusqu'en 2014, année à laquelle son statut administratif a évolué; il est désormais devenu un "CODATA Committee on Fundamental Constants".

14. Lide et Wood (2012), p.7.

15. Cohen et Taylor (1973)

16. Cohen et Taylor (1987)

17. Mohr et Taylor (1999)

18. Cohen et Taylor (1987), p.1718. Quatre ajustements ont été produits depuis; voir Mohr et Taylor (2005), Mohr, Taylor et Newell (2008), Mohr, Taylor et Newell (2012). Le compte rendu de l'ajustement de 2014 n'a pas encore été publié au 23 novembre 2015.

19. Birge faisait remarquer en 1932 que « la principale objection [à l'évaluation simultanée des constantes] est que les calculs nécessaires pourraient aisément prendre des siècles », Birge (1932b), p.231. L'exécution de l'algorithme actuel d'ajustement, développé et maintenu par les services du NIST, prend désormais quelques secondes tout au plus (François Nez, communication privée).

20. Cohen et DuMond (1965), p.539.

21. Cohen et Taylor (1987), p.1124.

devienne le moteur de la conception du GUM, et ce n'est pas sans raison que le travail d'ajustement des constantes de la physique influencera par la suite assez profondément la conception du GUM, en particulier du fait de la proximité entre les communautés impliquée – à titre d'exemple, l'un des principaux rédacteurs du GUM, Barry Taylor, physicien au NIST, est également un acteur majeur des ajustements des constantes de la physique, comme ce chapitre le montrera ;

- (3) L'application de la méthode des moindres carrés pour une évaluation simultanée de l'ensemble des constantes considérées dans l'ajustement. Ce travail sera continuellement perfectionné par ses successeurs. Nous pourrions dire qu'il s'agit presque de l'aspect le moins important du travail de Birge, car il ne forme qu'une solution technique à la problématique scientifique et épistémologique très profonde que Birge a eu le mérite de mettre en lumière.

Saisi par l'inhomogénéité des pratiques de son époque et par le manque de littérature précise sur la question de l'erreur probable, Birge met d'abord en avant la vertu méthodologique de son travail. Cohen et DuMond, en 1948, valoriseront avant tout le travail de synthèse que les ajustements apportent, avant de souligner avec plus d'insistance en 1965 leur intérêt quant au *test des théories*. Ce n'est qu'ensuite que les ajustements seront de plus en plus perçus comme un outil fondamental dans le travail de redéfinition des unités de mesure qui court encore jusqu'au début du XXI^e siècle.

Si l'initiative de Birge a connu un franc succès et a laissé une trace marquante dans la communauté de la physique de précision, la pratique des ajustements a également rencontré des critiques. Dans la prochaine section, nous évoquons l'une d'entre elles, qui a trait à la façon dont les résultats recensés sont sélectionnés en amont d'un ajustement.

10.2 La critique des ajustements

Avec l'émergence d'une tradition d'ajustements dans le milieu du XX^e siècle vient un lot d'objections qui s'élèvent contre le principe même des ajustements. DuMond et Cohen notent, en 1948, que :

La procédure pour obtenir des valeurs ajustées par des moyennes pondérées de nombreuses expériences différentes (dont le désaccord est probablement en partie le résultat d'erreurs systématiques) a été sévèrement critiquée dans de nombreux milieux. L'école de pensée qui s'oppose à une telle procédure n'a aucune meilleure alternative à offrir, cependant, que d'attendre avec espoir de meilleurs résultats et, en attendant, de ne prendre aucun engagement quand à l'adoption de valeurs. Le résultat pratique de cette attitude, néanmoins, est un manque lamentable d'uniformité dans la littérature puisque les auteurs sont continuellement obligés de faire des calculs impliquant les constantes.²²

Cohen et DuMond n'explicitent pas la nature de ces critiques. Il nous semble légitime, toutefois, de mettre en premier ligne la problématique de la *sélection* des données d'ajustement. Cette pro-

22. DuMond et Cohen (1948), pp.83–84

blématique traverse l'histoire des ajustements de bout en bout, dès le premier article de Birge jusqu'au début du XXI^e siècle. Pour la comprendre, il nous faut d'abord brièvement synthétiser le fonctionnement général d'un ajustement des constantes, sur la base de notre présentation du travail de Birge au chapitre 9. Celui-ci peut être schématisé selon trois étapes essentielles.

La première étape est l'étape de *recensement* des résultats expérimentaux. Elle ne se limite pas à la collecte des seules données numériques mais répond également à la nécessité de récupérer de nombreuses informations supplémentaires sur les résultats recensés, afin de pouvoir évaluer la façon dont ces derniers entreront ensuite dans le mélange.

La seconde étape est celle de *l'analyse critique* des résultats collectés, qui se fait elle-même en deux phases. Les calculs sont vérifiés, de façon à s'assurer de leur validité ; si nécessaire, certains d'entre eux sont repris, par exemple lorsque les auteurs ont utilisé un raisonnement correct et un protocole expérimental valide, mais se sont appuyés sur des tables de valeurs obsolètes ou incorrectes. Par ailleurs, il est nécessaire de contrôler que le résultat final est formulé de façon adéquate – puisqu'il est essentiel que tous le soient de la même façon. Ces opérations peuvent se révéler triviales si tous les résultats sont obtenus et formulés selon un protocole commun standardisé (par exemple, celui qui a été promu dans le GUM), mais elles étaient tout particulièrement indispensables lors des premiers ajustements de Birge. Nous constatons donc que cette première facette de l'étape d'analyse critique engage une *intervention* de la part des responsables de l'ajustement : les résultats ne sont pas collectés *tels quels* mais sont retravaillés lorsque cela est jugé nécessaire ; de même cela peut venir avec une réévaluation de l'incertitude de mesure associée. Une fois que les résultats collectés ont été retravaillés de façon à ce qu'ils puissent être comparés sans ambiguïté, la seconde phase de l'étape d'analyse critique consiste à *sélectionner* ces derniers en fonction de la façon dont leur apport potentiel à l'ajustement a été évalué. Cela signifie qu'un résultat n'est pas automatiquement considéré comme acceptable ; il est nécessaire de juger au préalable si son intégration à l'ajustement peut être considéré ou non comme bénéfique.

Vient alors la troisième étape : celle du *mélange*. C'est l'étape de synthèse ; l'ensemble des résultats est restitué sous la forme d'un jeu de valeurs recommandées, accompagné des incertitudes associées. C'est ici qu'intervient l'ajustement aux moindres carrés, et l'étape se résume essentiellement à cette opération mathématique. En fin de compte, le mélange est peut-être l'étape la moins problématique de toutes, car une fois acceptés les différents principes statistiques qui la sous-tendent, elle devient une routine.

Revenons plus en détail sur l'opération de sélection. Dans le chapitre précédent, nous avons présenté le travail de Birge comme une extension du mouvement épistémologique si remarquable que Stigler appelle *combinaison des observations*, en ce qu'il s'agit de réunir des résultats obtenus dans différentes conditions, par différentes personnes, au moyen de différentes méthodes de mesures. Bearden et Thomsen n'ont pas manqué de pointer en quoi cela contraste avec certaines pratiques antérieures :

Dans les premiers travaux de physique atomique, il y avait une tendance à sélectionner la meilleure expérience pour fixer une constante particulière ou un rapport. Par exemple, l'expérience de la goutte liquide de Millikan a longtemps été

considérée comme donnant la valeur la plus fiable de e . Cette approche a évolué graduellement vers une revue plus critique de toutes les données pertinentes.²³

Le geste épistémologique que constitue la *combinaison des observations* est basé sur la démonstration que le mélange de résultats expérimentaux différents peut être fécond, plutôt que de ne conserver que celui qui apparaît comme le meilleur d'entre eux. Pour autant, il semble exister des *bornes* à la combinaison des observations. Rappelons par exemple que Birge, dans son traitement de la vitesse de la lumière, avait écarté l'une des mesures recensées. Birge n'avait pas les moyens de *garantir* que le choix effectué était le bon, mais il ne pouvait pas se contenter de mélanger sans discernement l'ensemble des trois résultats. L'étape de sélection était rendue impérative par le fait que le résultat de Mercier était en désaccord significatif avec ceux obtenus séparément par Michelson d'un côté, et Rosa et Dorsey de l'autre. En effet :

Avant de clore la discussion sur les méthodes de calcul, il semble nécessaire d'attirer l'attention sur le fait que l'on doit utiliser un certain jugement dans l'application des méthodes des moindres carrés. Autrement, les résultats pourraient bien être absurdes.²⁴

Au cours des deux décennies d'ajustements durant lesquelles ils ont publié plusieurs ajustements, Cohen et DuMond n'ont eu de cesse de répondre à ces critiques et de justifier le principe même de sélection des données. Leur article de 1965 réserve un assez long développement consacré à l'explication des ressorts de ce qu'ils appellent le « rejet des données »²⁵. Leur défense met en évidence le rôle essentiel des erreurs systématiques dans la procédure. Un désaccord significatif entre deux résultats témoigne que l'un des deux résultats (au moins) est probablement affecté d'une *erreur systématique* :

Il devrait être parfaitement clair à tout physicien raisonnable que les données d'entrée, auxquelles doivent être attribuées des poids *a priori*, ne peuvent pas être pesées en toute sécurité par le rapport inverse des carrés des estimations des erreurs d'entrée si ces données sont mutuellement incompatible par trois, quatre fois ou peut-être même plus, que l'écart-type des différences entre les différentes données. Dans de telles circonstances, on soupçonne avec une presque certitude que le désaccord n'est pas une fluctuation statistique mais une erreur systématique dans l'une ou plusieurs des données. Dans un tel cas la pondération appropriée est probablement tout à fait sans rapport avec les écarts-types supposés. Nous sommes donc obligés de faire un choix de rejet de certaines données.²⁶

Si l'un des résultats est affecté d'une erreur systématique, alors il n'est plus possible de procéder à une combinaison des observations car cela aurait pour effet de reporter l'erreur systématique en question sur l'ensemble des autres résultats obtenus, alors que ceux-ci sont peut-être plus exacts. Cela justifie qu'il faille faire appel à une opération de sélection. Mais il reste encore à justifier au cas par cas chaque choix spécifique de sélection.

23. Bearden et Thomsen (1957), p.270.

24. Birge (1929b), p.6.

25. Cohen et DuMond (1965), pp.577-584.

26. Cohen et DuMond (1965), pp.539-540.

L'opération de sélection a donc pu être perçue comme problématique à plusieurs égards. Elle va d'abord, en apparence, à l'encontre de l'objectif initial de l'ajustement, à savoir la réunion des connaissances disponibles sur les constantes physiques. Apparaît une tension qui résulte de l'antinomie qu'il y a entre l'idée fondatrice de mélange et celle de sélection, qui font pourtant toutes deux partie intégrante de la procédure. Pourquoi sélectionner les données alors même que la vertu première des ajustements des constantes de la physique est de faire la réunion des différents résultats expérimentaux disponibles ? Si l'on effectue un mélange, pourquoi sélectionner ? Ou, à l'inverse, si l'on sélectionne, pourquoi s'embarrasser à chercher à mettre en place une quelconque combinaison des observations ? En se plaçant à mi-chemin entre mélange et sélection, la procédure d'ajustement n'apparaît pas à tous comme pleinement cohérente.

De plus, l'opération de sélection ne se fait pas sur un critère systématique et automatique. La comparaison de deux résultats fait appel à un test statistique qui permet d'évaluer si un désaccord est ou non significatif ; la combinaison de résultats qui ne sont pas en désaccord significatif se fait au moyen de la méthode de pondération décrite au chapitre précédent. Dans les deux cas, les seules informations numériques données dans les résultats, valeur finale et incertitude de mesure, suffisent à opérer. Concernant la sélection, on a affaire à un travail qualitatif, qui, comme l'énoncent Birge, puis Cohen et DuMond, fait appel au jugement, et donc à la subjectivité :

(C)eux qui effectuent une revue critique des constantes sont obligés d'utiliser le jugement et de prendre des décisions qui doivent inévitablement impliquer, avec un certain degré, l'élément subjectif.²⁷

Mais si cette dose de subjectivité est effectivement défendable, elle n'en est pas moins problématique car elle laisse l'impression que la sélection des données peut se faire sur la base d'un choix arbitraire. Cela vient nourrir la tension entre sélection et mélange, d'autant qu'à la question épistémologique de la justesse du choix de sélection s'ajoute d'ailleurs un enjeu institutionnel : pour un laboratoire, voir son résultat incorporé à l'ajustement est un facteur de reconnaissance, qui devient encore plus saillant à partir du moment où le CODATA pilote les ajustements. La censure d'une donnée peut donc aussi être perçue, en négatif, comme un désaveu et une critique des efforts et de la légitimité d'un groupe de travail.

Un troisième point réside dans la signification que l'on peut accorder aux résultats ajustés. Sélectionner les données d'entrée a pour conséquence de réduire les incertitudes finales, puisque cette « censure », pour reprendre le vocabulaire de Taylor *et al.*, revient à atténuer les écarts statistiques entre les valeurs exploitées dans l'ajustement :

(O)n doit réaliser que les valeurs finales des constantes résultant d'un LSA [*least-squares adjustment*] sont généralement obtenues d'un groupe de données d'entrée hautement filtré. La plupart des données discordantes ou « mauvaises » sont rejetées [...] Cette censure résulte dans le fait que les données de sortie sont tellement cohérentes qu'elles pourraient apparaître comme inviolées. Une telle confiance n'est pas justifiée!²⁸

27. Cohen et DuMond (1965), p.539.

28. Taylor, Parker et Langenberg (1969), p.379.

Taylor, Parker et Langenberg prennent pour exemple l'évolution de la valeur ajustée de la constante de structure fine sur les vingt années qui précèdent leur article. Leur représentation graphique de cette évolution est reproduite figure 10.1. L'étude de l'évolution de la valeur ajustée de la constante de structure fine montre qu'à plusieurs occasions, la valeur nouvellement ajustée diffère significativement de celle qui avait été obtenue lors de l'ajustement précédent. Les auteurs étendent leur constat à plusieurs autres constantes, également représentées figure 10.1 ; de plus, un examen attentif, au-delà des exemples pointés par les auteurs, montrerait que l'on retrouve le même constat, de façon régulière, tout au long de l'histoire des ajustements. Un tel constat peut surprendre si l'on pense que les ajustements ont pour but de produire les valeurs les plus exactes possibles des constantes, et si l'on s'attend, dans la continuité, à observer un progrès continu. Comme le discutent Taylor, Parker et Langenberg, ce constat peut également donner l'impression que les incertitudes sont généralement sous-estimées, c'est-à-dire que les métrologues en charge des ajustements sont trop optimistes quant aux résultats formulés, et qu'il faudrait donc envisager des projections moins ambitieuses afin de garantir l'exactitude des résultats obtenus. Mais le progrès des connaissances vient avec l'identification d'erreurs systématiques ou la découverte de nouvelles méthodes de mesure, qui amènent à réviser substantiellement les connaissances que l'on a des constantes. De fait, la solution qui consisterait à surévaluer les incertitudes de mesure ne satisfait pas les auteurs, notamment car cela revient à « jeter de l'information »²⁹. Cohen et DuMond avaient déjà formulé une réponse à ce type de critiques, et mis en avant le caractère hautement provisoire des résultats ajustés :

Il n'y a vraiment pas de danger dans l'utilisation de la théorie statistique pour le calcul des valeurs ajustées des constantes de la physique avec leurs erreurs probables résultantes, à condition que nous n'oublions par le caractère provisionnel des résultats. Si des erreurs systématiques sont présentes dans les données d'origine, nous pouvons seulement espérer qu'avec un nombre suffisant d'expériences indépendantes, même les erreurs systématiques tendront à avoir une distribution aléatoire. Puisque nous devons déterminer les valeurs provisionnelles à partir d'un réservoir de données partiellement incohérentes, l'outil le plus pratique que nous avons est l'ajustement aux moindres carrés. Cela fournit une méthode d'analyse impartiale pour déterminer un ensemble de valeurs qui constitue un compromis qui fait le moins possible de violence à toutes nos sources d'information.³⁰

La réponse adoptée par Cohen et DuMond amène toutefois à s'interroger de nouveau sur la finalité d'un ajustement : quelle importance accorder aux valeurs finales si celles-ci sont amenées à changer significativement de façon régulière ? Nous verrons à la section suivante que Taylor, quelques années plus tard, remettra explicitement en question l'utilité réelle des valeurs ajustées – tout en défendant l'intérêt de procéder à des ajustements. Notons pour l'instant que Cohen et DuMond remarquaient déjà en 1953 que les valeurs finales ne constituaient pas le seul intérêt de l'ajustement :

Deux utilisations peuvent être faites d'une solution aux moindres carrés de ce type : (1) elle peut être utilisée sur un ensemble assez cohérent d'équations surdéterminées pour déterminer les meilleures valeurs des constantes. (2) Elle peut être

29. Taylor, Parker et Langenberg (1969), p.379.

30. DuMond et Cohen (1948), p.84 (les auteurs soulignent).

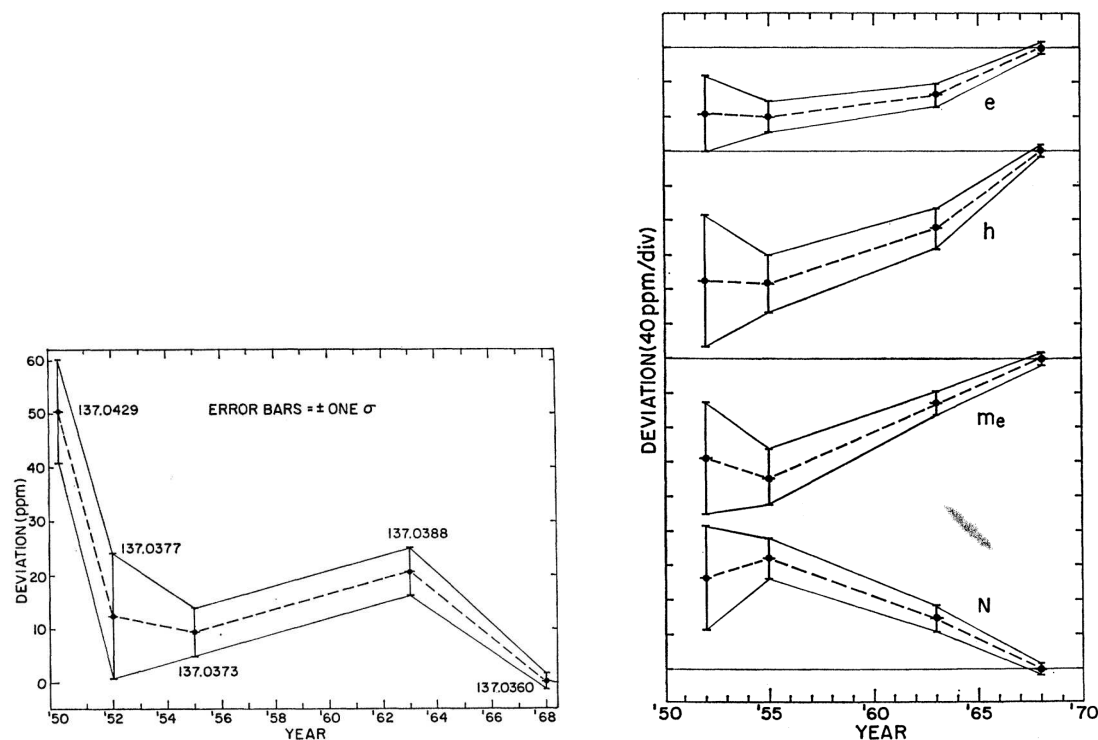


FIGURE 10.1 – Reproduction de deux graphes tirés de Taylor, Parker et Langenberg (1969) (à gauche, p.379; à droite, p.481) illustrant l'évolution de différentes constantes physiques depuis 1950. Le graphe de gauche représente l'inverse de la constante de structure fine; sur le graphe de droite sont représentées la charge élémentaire e , la constante de Planck h , la masse de l'électron m_e et la constante d'Avogadro N . En ordonnée est indiquée non pas la valeur ajustée absolue mais sa déviation par rapport au résultat de l'ajustement de 1969, en ppm (partie par millions). Chaque valeur ajustée, marquée par un point, est accompagnée de l'incertitude correspondante, figurée par un segment centré sur la valeur. En abscisse est reportée l'année à laquelle chaque valeur ajustée a été obtenue. Concernant la constante de structure fine, on constate à deux reprises, en 1952 et 1969, que la nouvelle valeur ajustée diffère significativement de celle qui avait été obtenue lors de l'ajustement précédent. Les autres constantes présentent également des « sauts » du même type. Ces constats sont assez communs et révèlent le caractère hautement provisoire des résultats ajustés : il n'est jamais possible de garantir leur exactitude.

utilisées sur un ensemble incohérent d'équations comme un outil pour déterminer, si possible, laquelle des équations contient l'erreur systématique cachée.³¹

Nous voyons donc émerger une tension importante entre deux objectifs. Le premier objectif que l'on peut concevoir est l'objectif *d'exactitude*, qui nous ramène à la problématique de la valeur vraie que nous avons abordée dans la partie II. Intuitivement, il est raisonnable de souhaiter que les valeurs fournies par les ajustements des constantes soient les plus exactes possibles, et ce d'autant plus qu'il s'agit de valeurs recommandées qui seront utilisées par tous, des chercheurs aux industriels. Les différents commentaires que l'on retrouve à cette période dans les comptes-rendus d'ajustement mettent en évidence la difficulté de cet objectif. Les ajustements ne peuvent informer que sur la cohérence interne d'un jeu de données expérimentales à un instant donné, mais ils ne garantissent en aucun cas l'*exactitude* des résultats ajustés. Nous voyons ici affleurer de nouveau la problématique de l'« inconnaisabilité » de la valeur vraie que nous avons décrite au chapitre 6, et dont nous avons vu une illustration avec la remarque de Cohen et DuMond, en 1965, exprimant l'idée que la « vérité absolue » n'est pas l'objet de la physique³². Taylor, Parker et Langenberg tiennent un discours semblable, où l'idée de vérité semble presque tendre à la foi religieuse :

[La représentation graphique de l'évolution de notre connaissance des constantes] illustre notre position selon laquelle aucun ensemble de constantes fondamentales ne devrait être considéré comme la vérité révélée [*Gospel truth*]. Alors que nous pouvons espérer que l'ajustement présent nous rapproche de la vérité, le réalisme nous oblige à reconnaître que d'autres changements significatifs dans notre connaissance des constantes sont susceptibles d'apparaître. Goethe aurait tout à fait pu parler de cela lorsqu'il a écrit "Es irrt der Mensch, solang' er strebt." [*l'homme se trompe tant qu'il cherche*]³³

En lieu et place de l'objectif d'exactitude, les auteurs mettent en avant une autre vertu du processus d'ajustement, à savoir permettre l'identification d'erreurs de mesure que l'on peut alors *corriger*. De ce fait, l'opération de *comparaison*, qui précède l'étape de sélection et qui en décide la nécessité, est au moins aussi importante que celle de combinaison qui régit le mélange final des données expérimentales.

Les tensions que nous décrivons ici se sont peu à peu cristallisées, de telle sorte qu'elles ont été explicitement débattues par les physiciens lors d'un colloque mené en 1970. Dans la prochaine section, nous nous intéresserons à deux interventions données lors de ce colloque. Nous nous attacherons en particulier à montrer la solution qu'a défendue Barry Taylor, et qui demeure encore aujourd'hui à la racine du mode de fonctionnement du groupe de travail du CODATA. Cette solution se situe dans la continuité des considérations de Cohen et DuMond, et présente l'intérêt de relier les deux tensions que nous avons présentées ici, en montrant que trouver un bon compromis entre sélection et combinaison permet d'envisager le progrès scientifique comme une quête perpétuelle, celle de la recherche et de la correction des erreurs

31. DuMond et Cohen (1953), p.693.

32. Voir p.188.

33. Taylor, Parker et Langenberg (1969), p.478.

– à condition d’accepter d’adoucir l’objectif d’exactitude à *l’instant présent* et de le considérer dans une perspective *dynamique*.

10.3 Deux attitudes épistémiques : précision contre exactitude

En 1970 est tenu un colloque international dans les locaux du National Bureau of Standards, à Gaithersburg, dans le Maryland aux États-Unis, réunissant des physiciens et métrologues spécialisés dans différents domaines de la physique de précision. Le compte-rendu de ce colloque, publié en 1971³⁴, nous donne accès à une très intéressante réflexion sur la façon dont sont interprétés les différents concepts et méthodes de l’ajustement des constantes de la physique. Il s’agit de la première réunion de cette ampleur sur le sujet³⁵; celle-ci survient peu après la création du groupe de travail du CODATA consacré aux constantes fondamentales. À cette occasion, les différentes communications balaient un large spectre de discussions, essentiellement tournées vers la physique fondamentale, classées par disciplines (étalons de temps et de fréquences, longueurs et vitesse de la lumière, rayons X, etc.). Le dernier thème abordé présente un intérêt tout particulier quant à la question des incertitudes de mesure. Il concerne la méthodologie du traitement de données expérimentales, et est présenté sous forme d’une question : « faudrait-il abolir les ajustements des constantes fondamentales ? »³⁶ L’un des thèmes centraux du débat est le traitement des données discordantes, qui fait écho aux différentes critiques évoquées dans les précédents comptes-rendus d’ajustements, critiques que nous avons décrites à la section précédente. Le débat est l’occasion d’une synthèse sur la thématique des ajustements – et pour certains, d’en défendre la légitimité.

La discussion s’ouvre avec un article critique de l’astrophysicien Peter Bender, qui se place dans la continuité des critiques que nous avons décrites à la section précédente. Bender commence par rappeler que le travail d’ajustement devient problématique lorsque les mesures des constantes physiques ne sont pas toutes en accord. Il ne milite pas pour autant pour un abandon des ajustements :

Quand les incertitudes attribuées par les expérimentateurs semblent raisonnables sur des bases physiques mais que les résultats sont en désaccord flagrant, un point de vue est de dire qu’il n’y a pas de façon sensée par laquelle nous pouvons attribuer une valeur la plus probable et une incertitude. Il y a une quantité considérable de vérité dans cette affirmation, mais je ne pense pas que cela devrait nous conduire à essayer d’arrêter la publication des évaluations.³⁷

Bender, cependant, est en désaccord avec la solution adoptée habituellement dans les différents comptes-rendus d’ajustement publiés jusqu’alors. Son désaccord porte principalement sur l’étape de sélection des données. Celle-ci présente le défaut d’être trop radicale, en particulier car elle implique des décisions binaires :

34. Langenberg et Taylor (1971b)

35. « Nous pensons que c’est la première fois que cela a été fait d’une telle façon délibérée et globale », avant-propos par E. Ambler, in Langenberg et Taylor (1971b), p.iii.

36. Langenberg et Taylor (1971b), pp.493–525.

37. Bender (1971), p.493.

Jusqu'à maintenant, les gens qui font des évaluations se sont sentis obligés de prendre des décisions binaires crues pour traiter les désaccords, et ils ne tiennent pas compte, dans les incertitudes finales, du fait qu'un mauvais choix a pu être fait.³⁸

Si une telle méthode de sélection est adoptée, il faut alors justifier les différents choix de sélection effectués. Sur ce plan, Bender pointe en particulier l'illusion d'objectivité que donnent les ajustements des constantes, alors même que l'opération de sélection manifeste tout le contraire :

Je pense que le moment est venu pour nous de reconnaître le degré de subjectivité et de flou qui existe réellement³⁹

La critique de Bender s'articule autour d'un idéal bien précis, caractérisé par la remarque suivante :

[La plupart des utilisateurs] préféreraient utiliser des valeurs qui sont au moins cohérentes entre elles et que quelqu'un qui a étudié le problème avec soin considère comme aussi proches que possibles que les vraies valeurs, au moment de leur adoption. [...] La question majeure est de savoir comment faire de la valeur et de l'incertitude qui sont choisies [pour les utilisateurs] des valeurs aussi peu biaisées que possible.⁴⁰

Bender émet ainsi le souhait que les données transmises aux utilisateurs soient exemptes de biais. Ce faisant, il met en avant une vertu bien particulière des résultats de mesure, leur *exactitude*. Il marque effectivement une différence avec les positions exprimées par Cohen et DuMond durant les deux décennies précédentes, dans lesquelles ces deux auteurs avaient identifié un tel objectif comme inatteignable. Bender n'accepte pas de se résoudre aux propositions des auteurs des ajustements précédents, décrivant les résultats de l'ajustement comme hautement provisoires. Nous avons vu que Taylor, Parker et Langenberg n'avaient pas d'objection à voir les valeurs ajustées des constantes effectuer régulièrement des « sauts » au fur et à mesure des ajustements. Cela ne satisfait pas Bender :

[Une] estimation juste de l'incertitude réelle dans le résultat [est] l'objectif le plus important. Pour résumer, le désir fondamental est de trouver un moyen d'éviter que l'incertitude citée dans les résultats soit systématiquement trop petite en raison du rejet de certaines données.⁴¹

Bender estime que ces remarques sont à même de justifier un aménagement des méthodes d'ajustement, de façon à faire disparaître la phase de sélection. Selon lui, il n'est pas légitime que la méthodologie de l'ajustement rende impérative une opération de sélection des données d'entrée, dès lors que rien ne permet d'affirmer sans équivoque que l'une est préférable aux autres. Bender suggère dans son article une méthode alternative, qu'il développe à partir de l'exemple suivant :

38. Bender (1971), p.494.

39. Bender (1971), p.494.

40. Bender (1971), pp.493-494.

41. Bender (1971), pp.494.

Considérons des résultats qui tombent en deux groupes, avec peut-être quatre valeurs dans l'ensemble supérieur et deux valeurs dans l'ensemble inférieur. [...] [Supposons qu'] il n'y ait aucun autre motif pour se méfier de l'un des résultats.⁴²

Bender envisage deux conclusions logiquement possibles. Il se peut, d'un côté, que l'un ou l'autre groupe de valeurs soit affecté d'une erreur systématique qui n'a pas été identifiée. Puisque l'erreur n'a pas été identifiée, il est impossible de la corriger, et il est également impossible de savoir quel groupe de valeurs elle affecte (le problème posé consiste justement à chercher à naviguer en eaux inconnues, et la seule chose qui peut être dite est que si l'un des groupes de résultats est correct, et que les incertitudes ont été correctement déterminées, l'autre est alors, quant à lui, automatiquement incorrect). Pour autant, le groupe (inconnu) de valeurs affecté par l'erreur systématique devrait en principe être éliminé de l'ajustement. Il se peut, d'un autre côté, que ce ne soient pas les valeurs mais les incertitudes qui soient affectées d'un défaut d'évaluation. Si les incertitudes associées aux différents résultats ont été sous-estimées, c'est que, par miroir, les différences entre les résultats en question ont été surévaluées et ne devraient peut-être pas être considérées comme significatives. Les différences entre les différents résultats deviennent mécaniquement significatives, plus que de raison. Dans ce second cas, la solution au problème consiste à *étendre* les incertitudes.

Trois solutions distinctes peuvent donc être envisagées : (1) abandonner le premier groupe de résultats ; (2) abandonner le second groupe de résultats ; (3) étendre les incertitudes. Mais si, comme l'a supposé Bender, « il n'y a pas d'autre raison pour se méfier de l'un des résultats », il est mécaniquement impossible de trancher entre les trois options précédentes. C'est ici que l'étape de sélection achoppe car elle nécessite alors l'appel à un choix arbitraire. Faire le choix exclusif de l'un des groupes de résultats sur le second, c'est prendre le *risque* de se tromper et de privilégier un ensemble de valeurs biaisées. S'il y a une symétrie dans les données, celle-ci doit se retrouver dans le résultat de l'ajustement ; or, l'étape de sélection brise cette symétrie. Pour éliminer ce type de biais, Bender propose plutôt d'associer une probabilité à chacune des trois options possibles (1), (2), et (3), leur « probabilité d'être correcte »⁴³ ; puis d'utiliser ces probabilités pour pondérer les résultats que produiraient chaque méthode prise séparément. Cela permettrait d'obtenir un résultat « moyen » qui tiendrait compte de façon équitable des différents résultats recensés (le calcul viendrait alors se *superposer* à la pondération classique des résultats selon leurs incertitudes respectives). Bender reconnaît que la méthode d'attribution des probabilités aux différentes hypothèses laisse une part importante à la subjectivité. Cependant, affirme-t-il, le choix de sélection promu dans les comptes-rendus d'ajustements donne encore plus de poids à la subjectivité, en ce qu'il contraint à des décisions radicales qui ne laissent pas de place à la nuance. Bender estime que sa solution présente la vertu de laisser les choix venir par degrés, et, de ce fait, de faciliter l'accord entre différents ajustements menés par différentes personnes. Suivant sa position, il vaut mieux un compromis qui incorpore un peu de tous les résultats disponibles qu'un choix sélectif qui peut se révéler erroné.

À notre connaissance, la suggestion de Bender n'a pas donné de suite dans la communauté scientifique. Cependant, il est aisé de constater une proximité frappante avec les techniques

42. Bender (1971), p.494.

43. « Je demanderais [à l'évaluateur] d'estimer les probabilités que chaque hypothèse soit correcte », Bender (1971), p.494.

bayésiennes qui commenceront à s'imposer deux décennies plus tard. Bien qu'il n'y soit fait nullement mention dans l'article de Bender, les probabilités qu'il propose d'assigner aux différentes propositions pourraient tout à fait être interprétées comme des probabilités *a priori* – le simple fait que ces probabilités portent sur des propositions et non des événements montrant déjà qu'il s'agit de probabilités épistémiques⁴⁴. Certains arguments sont communs avec ceux que développerons plus tard les bayésiens, en particulier celui selon lequel il est préférable d'assumer la subjectivité des évaluations plutôt que de la dissimuler. Il est intéressant, toutefois, de noter que Bender fait appel à de telles méthodes afin de sauvegarder un objectif d'exactitude, ce qui va dans le sens inverse de ce que plaideront ensuite les bayésiens.

L'article de Bender introduit un débat entièrement consacré aux ajustements des constantes, et intitulé « les ajustements aux moindres carrés des constantes fondamentales de la physique devraient-ils être abolis ? » L'ensemble des réponses qui y sont apportées sont remarquables pour ce qu'elles révèlent sur les principes et conceptions sous-jacentes à la notion de mesure en métrologie et aussi quant à l'évolution de l'interprétation des probabilités utilisées dans la métrologie. Dans le cas présent, c'est d'abord la réponse apportée par Taylor qui mérite l'attention.

Pour répondre à Bender, Taylor s'appuie sur l'ajustement qu'il a mené avec Parker et Langenberg peu avant cette conférence, en 1969. Cet ajustement incorpore de nouvelles connaissances théoriques et expérimentales issues de la découverte récente de l'« effet Josephson », et avec lui de nouvelles mesures de la grandeur $2e/h$. Le désaccord entre Taylor et Bender apparaît nettement comme la conséquence d'une différence dans la représentation des objectifs d'un ajustement. Bender, nous l'avons vu, voit l'ajustement des constantes comme un moyen de production des meilleures valeurs possibles des constantes – c'est-à-dire ici des valeurs les plus exactes – afin que les utilisateurs de ces constantes disposent du meilleur « produit » possible. Mais Taylor rappelle quant à lui que les ajustements possèdent d'autres vertus que de simplement fournir des jeux de valeurs ajustées à des utilisateurs divers. Il considère même que les valeurs finales n'ont qu'un intérêt mineur en regard des autres bénéfices des ajustements :

Certes, il est utile d'avoir à une époque donnée un ensemble cohérent de constantes qui peuvent être utilisées par tous les chercheurs qui en ont besoin. Mais ceci est le résultat le moins important d'un ajustement aux constantes – la contribution la plus précieuse de telles études à la connaissance humaine est l'information recueillie au cours de l'examen critique qui accompagne l'ajustement.⁴⁵

Ainsi, les ajustements sont utiles, mais « peut-être pas pour les raisons les plus évidentes »⁴⁶. Taylor insiste sur les vertus épistémiques du *processus* d'ajustement et de l'ensemble du travail

44. Cohen et DuMond adoptent eux-mêmes une conception qui s'approche remarquablement des modèles bayésiens : « il est clair que pour répondre utilement à la question de savoir s'il est justifié de rejeter une donnée, nous devons avoir une information *a priori* quant à l'estimation de la précision de toutes les données du groupe. Nous ne faisons presque jamais de mesure sans avoir une préconception (qu'elle soit vague ou précise) de la précision à attendre de nos méthodes, de notre appareil, de nos compétences en tant qu'observateurs. D'un certain point de vue, la vie peut être considérée comme rien d'autre qu'une succession de ces jugements *a priori*, suivis et vérifiés par des conclusions *a posteriori* », Cohen et DuMond (1965), p.579.

45. Taylor (1971), p.495.

46. Taylor (1971), p.495.

que celui-ci engage. La production d'un jeu de valeurs ajustées ne constitue que l'un des cinq apports principaux que Taylor recense. Les quatre autres apports concernent (i) l'examen critique des données d'entrée et des modèles théoriques. C'est là un point sur lequel Birge avait déjà beaucoup insisté : certains résultats nécessitent d'être retravaillés, telle la valeur théorique de la constante de Rydberg dans l'ajustement de 1969. Ils concernent également (ii) l'harmonisation des pratiques scientifiques concernant l'analyse d'incertitude et la formulation des résultats de mesures – encore un aspect longuement souligné par Birge, et qui a coïncidé avec, si ce n'est guidé, l'adoption généralisée de l'erreur probable puis de l'écart-type expérimental dans la pratique des métrologues. Ils concernent enfin (iii et iv) l'identification de désaccords expérimentaux et le test général de la cohérence des théories physiques. Cela aboutit à une traque des erreurs systématiques et au développement de nouveaux éléments de théorie. Selon Taylor, tous ces points peuvent être observés dans l'ajustement de 1969.

Taylor insiste donc sur le fait qu'en définitive, l'ensemble de valeurs obtenues en sortie du processus d'ajustement n'est pas d'une grande importance⁴⁷. D'une part, la majorité des utilisateurs n'a pas besoin de valeurs aussi précises : « la plupart des travailleurs n'ont pas besoin des dernières décimales dans les valeurs des constantes »⁴⁸. Nombre d'entre eux auraient même pu se contenter de valeurs non ajustées⁴⁹. D'autre part, les utilisateurs qui ont besoin des valeurs les plus précises ne peuvent pas se contenter du résultat de l'ajustement et vont avoir besoin de fouiller plus profondément dans les travaux de recherche qui ont été menés à ce sujet :

(C)es scientifiques qui ont réellement besoin d'utiliser les dernières décimales ne se contenteront pas seulement de prendre les nombres dans une table mais remonteront à l'article originel.⁵⁰

Dès lors, les ajustements ne sont pas tournés vers une large population d'utilisateurs, mais sont un travail par et pour les spécialistes de la physique de précision :

(P)uisque la majorité ne se soucie pas particulièrement des nombres qu'elle utilise, les ajustements devraient être adaptés à la minorité, petite mais plus importante, qui le fait et qui a besoin des nombres les plus utiles et les plus stimulants sur lesquels ils peuvent mettre leur main.⁵¹

Dans un article issu de la même conférence, Thomsen (lui-même coauteur de plusieurs ajustements dans les années 1950) appuie cette position. Selon lui, la revue critique constitue l'essentiel du travail ; l'ajustement lui-même n'est qu'une « routine »⁵² et les données ajustées deviennent trop rapidement obsolètes pour qu'elles soient réellement utiles pour les études les plus pointues. Taylor illustre cette démarche par une métaphore :

47. « On peut dire que la valeur numérique particulière des ensembles finaux de “meilleures” constantes ou de constantes “recommandées” est seulement d'une importance secondaire. L'attitude de la plupart des scientifiques à propos des constantes soutient également cette thèse », Taylor (1971), p.496.

48. Taylor (1971), p.496.

49. « Ces travailleurs qui se contentent d'utiliser les nombres tels qu'ils leurs sont donnés sans se soucier de leur provenance pourraient utiliser presque n'importe quel nombre », Taylor (1971), p.496.

50. Taylor (1971), p.496.

51. Taylor (1971), p.497.

52. Thomsen (1971), p.503.

Notre philosophie, dans l'ajustement de 1969 est de fournir le meilleur objet tranchant pour les travailleurs qui en ont besoin et qui sont prêts à l'utiliser tout en sachant le soin qu'il faut apporter à ne pas se couper eux-mêmes.⁵³

C'est en rapport avec cette conception de l'apport des ajustements que l'étape controversée de sélection des données lui apparaît non seulement légitime mais même indispensable. Taylor confronte les positions consistant d'une part à sélectionner les données d'entrée, et d'autre part à prendre en compte l'intégralité des données en étendant les incertitudes si nécessaire, de façon à faire disparaître les désaccords. Il ne voit dans la seconde solution que la recherche d'un compromis entre des données pourtant incompatibles entre elles, que l'on réconcilierait artificiellement en prétendant a posteriori qu'elles seraient moins précises que ce que le travail d'analyse porte à affirmer. Cela n'est pas satisfaisant :

(C)'est un aveu de défaite [...] la seule solution serait de supposer que toutes les données sont mauvaises. Par conséquent, cela revient à jeter de l'information en rendant les quantités plus inconnues qu'elles ne le sont vraiment.⁵⁴

Selon Taylor, cela revient à *masquer* les erreurs, un constat déjà effectué par Cohen et DuMond lors de l'ajustement de 1965 :

L'idée qu'une surestimation de l'erreur "pour la sécurité" est en quelque sorte plus louable ou vertueuse que l'effort d'être aussi exact que possible avec le biais le plus faible possible que ce soit dans l'amplification ou la minimisation de l'estimation de l'écart-type, est déplorablement répandue. Nous demandons : pour qui une telle surestimation est-elle "sûre"? Certainement pas pour la communauté scientifique générale qui souhaite utiliser le résultat. Pour elle c'est une dissimulation de la réalité des faits en ce qui concerne les résultats de mesure. Pour l'auteur de ce résultat cela peut sembler "sûr" en ce qu'à une date ultérieure sa mesure est moins susceptible d'être contredite par un travail futur; mais même pour lui cette timidité inutile peut être une sécurité illusoire car, en raison de l'exagération injustifiée de son estimation d'erreur, un désaccord crucial, susceptible de révéler un fait nouveau fondamental et important, pourrait avoir été enterré et perdu pour toujours, d'aussi loin que sa réputation est concernée. Pour les hommes courageux il n'y a qu'un seul refuge "sûr", la vérité simple et sans fard quant à leurs méthodes et leurs résultats.⁵⁵

53. Taylor (1971), p.497.

54. Taylor (1971), p.496. Thomsen appuie également cette position : « Une alternative est de ne rien rejeter et de simplement augmenter toutes les erreurs de façon suffisante pour obtenir une image raisonnablement cohérente. Cependant, cela [...] empêche tout jugement critique de la part de l'évaluateur », Thomsen (1971), p.504. Pour autant, Thomsen ne partage pas totalement la position défendue par Taylor. Il estime qu'il reste profitable de chercher à tenir compte des résultats discordants dans l'ajustement, et propose une brève analyse de la façon dont cela pourrait être fait. Il affirme ainsi « il existe un juste milieu évident entre l'augmentation sans discernement des erreurs et le rejet d'une donnée dans son entièreté », p.505. Nous ne décrivons par en détail la position de Thomsen car il ne serait pas utile de résumer l'ensemble des positions défendues par chaque participant au débat. Notons simplement que cette position rejoint en partie celle de Bender quant à l'idée de départ, à savoir la légitimité d'une sélection : Bender n'est donc pas isolé sur ce plan.

55. Cohen et DuMond (1965), p.541.

Inversement, la solution de sélection est la solution la plus utile, car elle stimule le travail expérimental et théorique, et valorise l'ensemble des vertus épistémiques de l'ajustement, plutôt que de se focaliser sur la seule production de jeux de valeurs finales. Taylor l'exprime ainsi :

Le défi de prouver que les ajustements sont incorrects, forçant ainsi un nouvel ajustement, est bien plus stimulant que le défi consistant à obtenir un nombre qui sera simplement ajouté au “pot”⁵⁶

Pour Taylor, effectuer une sélection dans les données d'entrée est la meilleure méthode à condition de prendre pour perspective *les investigations à venir*. La projection dans le futur apparaît de façon très nette dans la citation suivante, extraite de l'ajustement que Taylor a mené en 1969 avec Parker et Langenberg :

(L)es mesures des constantes physiques fondamentales à des niveaux toujours plus élevés d'exactitude sont importantes, pas seulement parce qu'elles ajoutent “une nouvelle décimale” et nous fournissent un ensemble de constantes plus cohérent avec lequel travailler, mais parce qu'elle *peuvent mener* à la découverte d'incohérences autrefois inconnues, ou à l'élimination d'une incohérence dans notre description physique de la nature. Nous avons pris ce point de vue comme principe directeur dans l'ensemble de la présente étude.⁵⁷

Même lorsque le choix de sélection est erroné, les conséquences peuvent être bénéfiques. Il prend pour exemple les mesures de la constante de structure fine, dont deux valeurs incompatibles avaient été recensées en 1965 (voir figure 10.2). Cohen et DuMond avaient privilégié l'un des deux résultats, motivés en cela par des considérations théoriques – un choix contredit ensuite par de nouvelles mesures exploitant l'effet Josephson. Selon Taylor, « il est désormais clair que la décision incorrecte de Cohen et DuMond s'est révélée très utile »⁵⁸.

La confrontation des articles de Bender et de Taylor met en évidence deux façons tout à fait différentes de concevoir l'exigence d'exactitude associée à un ajustement des constantes de la physique. En privilégiant l'exactitude, et ce pour que les utilisateurs soient le moins enclins à commettre d'erreurs, Bender se focalise sur le résultat immédiat de l'ajustement, dans une conception des ajustements que l'on pourrait décrire comme *statique*. En soulignant les différentes vertus épistémiques de l'ajustement, Taylor déploie une conception *dynamique* des ajustements dans laquelle les scientifiques se projettent en permanence vers le futur avec pour objectif majeur, non l'exactitude immédiate des résultats ajustés, mais la quête d'un *progrès* scientifique, expérimental (identifier des erreurs systématiques) et théorique (expliquer les désaccords par un perfectionnement des modèles et des théories employées) – quitte, parfois, à devoir faire marche arrière lorsque certains choix se révèlent erronés⁵⁹. La réponse de Taylor montre à quel point la recherche d'une exactitude *immédiate* est secondaire dans la démarche d'ajustement.

56. Taylor (1971), p.496.

57. Taylor, Parker et Langenberg (1969), p.377 (nous soulignons).

58. Taylor (1971), p.497.

59. « Il est bien sûr évident que le prix à payer pour la solution que je préconise est que s'il s'avère plus tard que les mauvaises données ont été utilisées, alors les valeurs ajustées de certaines constantes peuvent changer d'un ou de deux écarts-types, voire plus. ([...] ceci a été la règle plutôt que l'exception.) », Taylor (1971), p.497. Nous renvoyons à la section précédente pour une discussion plus complète de cette question précise.

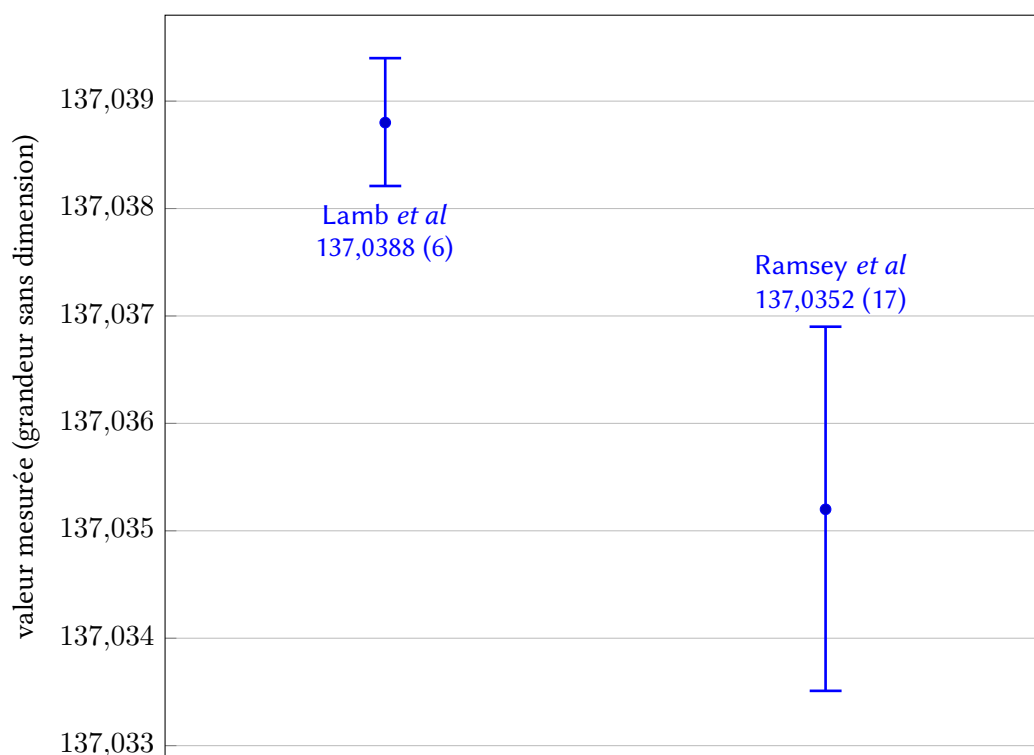


FIGURE 10.2 – Mesures de la constante de structure fine en 1965. Les nombres indiqués entre parenthèses représentent les incertitudes respectives sur les deux dernières décimales du résultat, conformément à la notation usuelle adoptée par les métrologues depuis le milieu du XX^e siècle. Les deux résultats ont été obtenus selon deux principes de mesure différents et présentent un désaccord significatif. Dans leur ajustement de 1965, Cohen et DuMond ont choisi de sélectionner le résultat obtenu par Lamb *et al.* à partir d'un raisonnement qui ne mobilise pas les amplitudes des incertitudes elles-mêmes.

Bien sûr, l'on peut toujours arguer que le progrès scientifique qui découle de l'identification et la maîtrise des sources d'erreurs systématiques, ainsi que du perfectionnement des théories, permet en retour une amélioration de l'exactitude des valeurs des constantes de la physique. Mais cela n'est en quelque sorte qu'un « effet secondaire », à plus long terme, d'un ensemble de vertus épistémiques qui ne se limite pas aux seules valeurs obtenues. Le mélange des résultats expérimentaux ne produit pas en lui-même de nouvelle connaissance sur les constantes de la physique ; il permet, en revanche, d'enclencher un processus qui quant à lui est à l'origine de nouvelles connaissances de par l'analyse critique qu'il permet de mettre en place.

10.4 Conclusion sur les ajustements

L'histoire des ajustements ne s'arrête pas en 1971, mais son intégration à la structure institutionnelle du CODATA en constitue un tournant. Le rythme des ajustements est un temps ralenti, pour être ensuite accéléré et rendu régulier. Les publications, réunies sous la bannière du CODATA, ne laissent plus apparaître les disparités interprétatives comme le permettait le fourmillement des années 1930 à 1960. Ces disparités n'ont pas pour autant disparu ; seulement, en centralisant le travail d'ajustement, l'action du CODATA contribue à les rendre plus silencieuses. À partir de 1970, les comptes-rendus d'ajustement possèdent un statut de référence, et de ce fait, sont produits à travers un long et indispensable travail de polissage, qui tend à (au moins partiellement) masquer les idiosyncrasies mises en jeu dans la conception de leur contenu : difficultés rencontrées, désaccords entre spécialistes (réglés ou non), ainsi que les diverses aspérités du travail ordinaire de recherche.

Le compte-rendu que proposent Cohen et Taylor de l'ajustement mené en 1973⁶⁰ contient essentiellement des remarques beaucoup plus techniques et ne revient pas sur les questionnements épistémologiques précédents. Celui de 1986, toujours par Cohen et Taylor⁶¹, incorpore un nouvel élément de standardisation en se conformant aux directives de la recommandation INC-1 émise par le BIPM (voir partie I ainsi que l'épilogue historique proposé au chapitre 12 de la présente partie). Il intègre par ailleurs de nombreuses révisions techniques dont certaines sont rendues possibles par l'avènement du calcul numérique. Le constat est similaire pour le compte-rendu de 1998, par Mohr et Taylor⁶², où il est fait explicitement référence au GUM, et qui revient par ailleurs en détail sur la méthodologie employée. Les ajustements suivants prennent pour référence le texte de 1998, et se contentent essentiellement d'y renvoyer le lecteur pour les questionnements méthodologiques généraux.

Nous pensons qu'il serait inutile de décrire ici en détail les évolutions techniques qui caractérisent les ajustements de 1986, de 1998, et les ajustements ultérieurs. Celles-ci sont loin d'être anecdotiques ; leur étendue montre que les questions posées dans les publications précédant la création du CODATA continuent à être discutées, et que les réponses apportées s'appuient sur le progrès régulier des méthodes statistiques⁶³. Ainsi, l'algorithme d'ajustement a été sensible-

60. [Cohen et Taylor \(1973\)](#)

61. [Cohen et Taylor \(1987\)](#)

62. [Mohr et Taylor \(1999\)](#)

63. Suite aux travaux de Cohen dans les années 1950, les modèles ont intégré l'analyse des corrélations entre les données d'entrée. Cette évolution a joué un rôle important dans l'adoption d'une formalisation qui fait appel au

ment raffiné et complexifié ; il en est de même pour les méthodes de comparaison des données et d'analyse de leur cohérence d'ensemble, ou encore pour les procédés de sélection des données d'entrée. Cependant, les questionnements épistémologiques principaux demeurent essentiellement inchangés. Les questions posées en 1970 lors de la conférence organisée par le NIST sont toujours valides aujourd'hui, même si elles ne sont pas nécessairement énoncées dans les mêmes termes. Un exemple récent, parmi d'autres, vient illustrer en quoi le traitement des données discordantes peut être considéré – jusqu'ici – comme un relatif invariant : il s'agit de la mesure du « rayon du proton ».

Le « rayon du proton » est une expression qui désigne la dispersion spatiale de la charge électrique d'un proton – laquelle, en physique quantique, n'a pas de frontière précise. Plus précisément, il s'agit du rayon quadratique moyen de la distribution de charges, défini par :

$$r_p = \sqrt{\langle r_p^2 \rangle} = \sqrt{\frac{1}{e} \int_{r=0}^{+\infty} r^2 \rho(r) dr} \quad (10.1)$$

où $\rho(r)$ est la densité de charge sur une surface de rayon r (la formule donnée est valable pour une symétrie de type sphérique)⁶⁴. Pendant un temps, le rayon du proton a été mesuré au moyen d'expériences sur des atomes d'hydrogène, selon deux principes différents : par diffusion d'électrons à travers un gaz d'hydrogène, et par spectroscopie⁶⁵. Cela a mené à des résultats compatibles, qui, moyennés, suggèrent un rayon $r_p = 0,8770 \pm 0,0045$ fm. Plus récemment, l'équipe de Randolph Pohl, à l'institut Paul Scherrer, en Suisse, a adopté une nouvelle méthode basée sur la spectroscopie d'un atome d'« hydrogène muonique » formé non pas d'un proton et d'un électron, mais d'un proton et d'un muon. Cela a permis à l'équipe d'améliorer d'un facteur dix la précision des mesures, et d'aboutir à une prédiction donnant $r_p = 0,8409 \pm 0,0004$ fm. Les deux résultats sont représentés à la figure 10.3 qui montre sans ambiguïté qu'ils sont en net désaccord. Les physiciens ont constaté qu'il était pour l'instant impossible de résoudre l'incohérence entre les données, que ce soit par une explication théorique ou par la découverte d'un biais expérimental. De fait, ils ont appelé ce problème « l'énigme du rayon du proton [*the proton radius puzzle*] »⁶⁶. Lors d'une réunion organisée dans les locaux du BIPM, à Sèvres, les 3 et 4 novembre 2014, l'énigme du rayon du proton a occupé l'une des demi-journées de travail⁶⁷, durant laquelle ont été discutées de nombreuses pistes théoriques et expérimentales. Les membres du CODATA ont également dû prendre une décision quant aux données à incorporer à l'ajustement de 2014. Lors des discussions a émergé une problématique en de nombreux points semblable à celle discutée par les scientifiques tels Bender ou Taylor en 1970 : peut-on trouver un moyen de rendre compte équitablement des deux résultats, ou bien est-il nécessaire

calcul matriciel, bien plus poussé que les premiers calculs menés par Birge.

64. Pohl *et al.* (2010)

65. La récente histoire de la mesure du rayon du proton peut être retrouvée sous une forme très accessible dans le journal *Pour la science*, Bernauer et Pohl (2014)

66. Pohl *et al.* (2013)

67. Nadine de Courtenay et moi-même avons eu le privilège d'être invités par les membres du CODATA à venir assister aux deux jours de réunion. J'en profite pour remercier ici chaleureusement Claudine Thomas, David Newell, François Nez, et les autres membres concernés du BIPM ou du CODATA de nous avoir aimablement accueilli en leur sein durant ces deux jours.

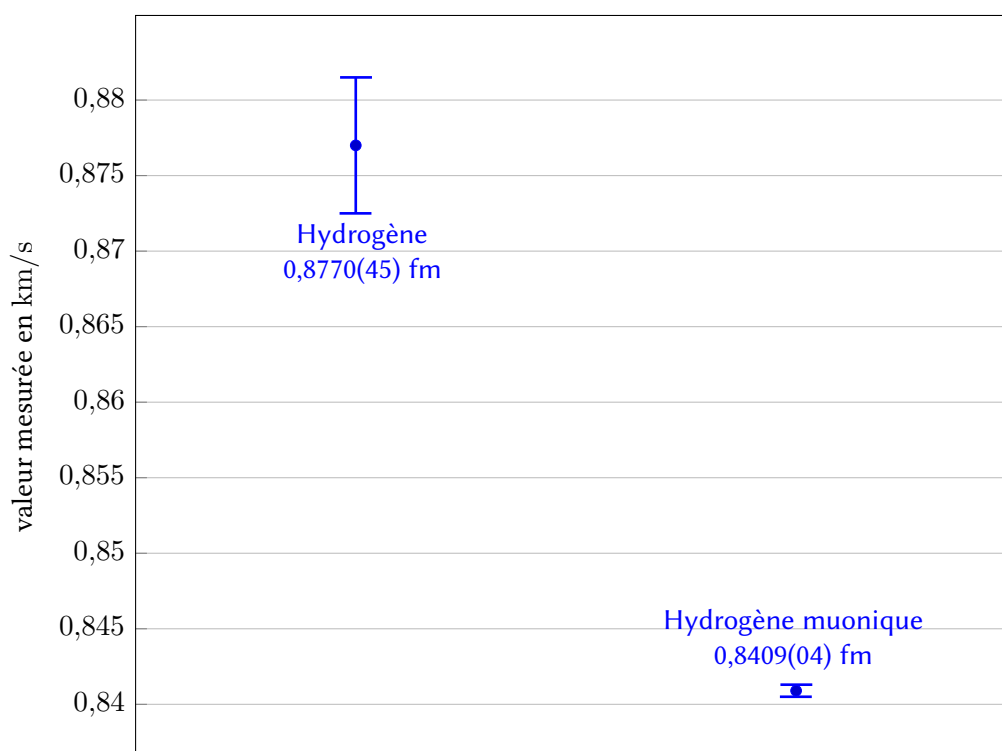


FIGURE 10.3 – Mesures du rayon du proton à partir de principes de mesure orientés autour de l'hydrogène d'une part et de l'« hydrogène muonique » d'autre part, recensées lors de l'ajustement du CODATA de 2014. Les nombres entre parenthèse représentent les incertitudes sur les deux derniers chiffres significatifs. Les résultats sont clairement en désaccord, en raison de quoi le CODATA a dû arbitrer l'intégration de ces résultats à l'ajustement final, en attendant une résolution théorique ou expérimentale du problème.

de (temporairement) trancher entre les deux ? Le fait que ce questionnement perdure rappelle que sa résolution n'est pas une question de connaissance scientifique mais d'un choix de méthodologie, qui contient une part de convention, voire d'arbitraire, et qui n'est pas vouée à être partagé par tous à tout moment. Lors de la réunion du CODATA, certains ont ainsi rappelé que rien ne justifie de rejeter un résultat s'il n'est en contradiction avec rien d'autre qu'un second résultat. D'autres, à l'instar de Bender en 1970, ont suggéré qu'il fallait agir en relation avec les attentes des utilisateurs. D'autres encore ont estimé que l'objectif d'un ajustement n'est pas d'obtenir les « bonnes » valeurs (ce que nous appelons un objectif d'*exactitude*) mais de formuler un état de connaissance (ce que représente l'*incertitude* dans son acception contemporaine). En fin de compte, la décision a été prise d'écarter les résultats obtenus avec l'hydrogène muonique, par prudence d'une part (s'agissant du résultat le plus récent), et d'autre part pour « maintenir l'énigme comme une énigme », et, en laissant active cette tension, rendre plus évident encore le besoin de lui trouver une véritable réponse par la théorie ou par l'expérience. L'ensemble des

membres présents s'est pourtant accordé sur le fait que la mesure par l'hydrogène muonique constitue un résultat *valide*. L'un des membres du CODATA a d'ailleurs pointé ce paradoxe, en estimant que le rejet du résultat suggère en fait que les physiciens ne lui accordent pas autant de confiance qu'ils le prétendent. En fin de compte, certains ont insisté sur le fait que l'apparent problème que pose le désaccord entre les deux résultats est avant tout quelque chose de *fécond*, puisqu'il stimule la recherche théorique et expérimentale.

Que révèle l'analyse de l'histoire des ajustements depuis l'initiative de Birge jusqu'aux travaux récents du CODATA ? Nous avons focalisé notre attention sur le traitement des données discordantes, et la façon dont les physiciens choisissent de s'attaquer à ce problème par un processus de *sélection* qui va en apparence à l'encontre du mouvement de *combinaison des observations* lui-même au fondement de l'idée d'incertitude de mesure et du travail d'ajustement⁶⁸. Sur ce plan, quatre points essentiels ressortent. D'abord, ces épisodes confirment la conclusion que nous avons tirée de l'analyse du travail de Birge, à savoir que l'incertitude de mesure ne peut pas être trop rapidement comprise comme l'estimation de la fiabilité d'un résultat – l'exemple du rayon du proton montre que c'est parfois le résultat dont l'incertitude est la plus élevée qui est conservée par les scientifiques. Ensuite, les réponses apportées au problème s'articulent autour de l'idée d'erreur systématique. Ces erreurs systématiques constituent une difficulté apparente de l'analyse d'incertitude car elles ne peuvent pas être incorporées à un bilan d'incertitude tant qu'elles n'ont pas été identifiées. C'est la raison pour laquelle l'incertitude de mesure ne peut pas être comprise comme une évaluation de l'exactitude de mesure, c'est-à-dire de la proximité entre le résultat obtenu et la « valeur vraie » de la constante mesurée. Apparaît donc une tension entre exactitude et incertitude que l'on retrouve à différentes reprises dans les discussions des physiciens, et que nous retrouvons par la suite dans le débat opposant les interprétations fréquentiste et bayésienne de la mesure. Cette tension prend également la forme d'une tension entre objectivité et subjectivité dans la mesure. En effet, le traitement de l'erreur systématique fait appel au jugement de l'expérimentateur, et cela nous renvoie au fait que ce traitement est aussi la raison première pour laquelle les méthodes bayésiennes ont pris leur essor. De fait, la question du traitement des erreurs systématiques, la tension entre incertitude et exactitude, et celle entre subjectivité et objectivité sont des ingrédients remarquables du débat entre fréquentistes et bayésiens que nous avons décrit dans la partie I.

Dans le prochain chapitre, nous souhaitons conclure sur les questionnement relatifs aux rapports entretenus entre incertitude et exactitude en réinvestissant le débat opposant fréquentistes et bayésiens, à la lumière des processus épistémiques que l'étude des ajustements a permis de mettre en évidence.

68. Notons que la question de la sélection des données expérimentales ne s'est pas posée qu'aux physiciens effectuant des ajustements. On la retrouve de façon régulière en métrologie, comme le montrent les discussions évoquées dans Demeyer (2011), p.27.

Chapitre 11

Incertitude et exactitude

Le cas des ajustements des constantes nous donne un aperçu assez complet de l'une des façons dont les scientifiques peuvent se mettre d'accord sur un résultat. L'ajustement des constantes est un processus de construction d'un consensus scientifique temporaire qui s'exprime par le jeu de valeurs recommandées mais aussi par les choix de sélection des données d'entrée et par le poids qui est accordé à chacune de ces données. Aucun de ces aspects n'est effectué de façon automatique par un algorithme mathématique ; ils sont le produit de négociations impliquant des acteurs épistémiques dont le caractère social ne peut pas être mis en retrait. Ainsi, les ajustements apparaissent comme une plate-forme épistémique et sociale de mise en commun des résultats de mesure.

Aucun consensus, bien entendu, ne garantit l'exactitude des valeurs recommandées des constantes physiques. De fait, les ajustements des constantes ne répondent pas vraiment, ou du moins seulement partiellement, à une seconde question : quelle crédibilité accorde-t-on à un résultat de mesure, et, par extension, pourquoi en fait-on usage ?

Rappelons, à ce stade, les définitions respectives d'« incertitude de mesure » et d'« exactitude de mesure », dans le VIM3. En premier lieu, l'incertitude est définie comme telle :

Incertitude de mesure : paramètre non négatif qui caractérise la dispersion des valeurs attribuées à un mesurande, à partir des informations utilisées.¹

La définition du VIM3 propose en quelque sorte une acception générique de l'incertitude : elle indique qu'il s'agit de la largeur des valeurs attribuées à un mesurande, sans pour autant expliquer comment cette largeur doit être interprétée. Doit-on considérer que l'incertitude de mesure indique la largeur de valeurs dans laquelle la valeur vraie du mesurande se situe vraisemblablement ? Cette question est celle du lien entre incertitude de mesure et exactitude de mesure, cette dernière étant définie dans le VIM comme :

Exactitude de mesure : étroitesse de l'accord entre une valeur mesurée et une valeur vraie d'un mesurande²

1. [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.25.

2. [Joint Committee for Guides in Metrology \(JCGM\) \(2012\)](#), p.21.

Nous considérerons, dans la suite de ce chapitre, l'incertitude de mesure à travers la définition générique du VIM3, et nous nous interrogerons sur la façon dont on peut l'interpréter, selon qu'elle prenne la forme d'une exactitude de mesure, ou d'une incertitude au sens que lui donnent les bayésiens.

Dans la section 11.1, nous rappelons les deux horizons limites que notre analyse des conceptions de la mesure en métrologie a fait apparaître : un horizon fréquentiste objectiviste, et un horizon bayésien épistémique. Nous expliquons en quoi ces deux horizons ne nous apparaissent pas pleinement satisfaisants.

Les sections 11.2 et 11.3 analysent alors successivement l'éclairage que les ajustements des constantes de la physique apportent quant à ces deux conceptions limites de la mesure. Nous proposons une définition de l'incertitude de mesure relative à chacune de ces conceptions, tout en indiquant les limites relatives à chacune des approches.

La section 11.4 aborde brièvement un autre épisode de l'histoire de la physique - celui de l'annonce de la détection de neutrinos « supraluminiques », c'est-à-dire se déplaçant à une vitesse supérieure à celle de la lumière dans le vide. Ce nouveau cas d'étude, qui n'est pas directement relié aux ajustements des constantes de la physique étudiés jusqu'ici, nous permet d'observer en contexte la façon dont un résultat de mesure est reçu et accepté ou non par la communauté scientifique. Nous constatons ainsi que l'incertitude de mesure n'est pas directement reliée au rapport de confiance et de crédibilité qu'entretiennent les différents acteurs quant au résultat de mesure, mais que ce rapport est alimenté par des informations externes au résultat lui-même, que l'incertitude de mesure ne peut pas contenir.

Dans la section 11.5, nous notons pour finir qu'il est important de distinguer deux perspectives différentes, qui engagent des enjeux distincts. On compte d'une part une perspective de recherche, orientée vers l'amélioration de la connaissance, par exemple des grandeurs physiques comme dans le cas des ajustements. Dans cette perspective, l'objectif est de sécuriser la découverte à *terme* des erreurs de mesure de façon à stimuler le progrès scientifique. Dans ce cas, à un instant donné, l'exigence d'exactitude n'est pas prioritaire. Cependant, on compte d'autre part une perspective appliquée, qui est par exemple celle des ingénieurs ou des industriels : l'objectif sera ainsi d'assurer la qualité d'une chaîne de production, de garantir la fiabilité d'un système ou d'un produit de consommation, etc. Dans ce cas, l'exigence d'exactitude est prioritaire, et la perspective de ne corriger les erreurs qu'à moyen ou long terme n'est pas acceptable. On se retrouve alors non plus dans une dynamique de progrès tournée vers le futur, mais dans une logique de *risque* tournée vers le présent.

11.1 Deux horizons insatisfaisants

Notre étude des ajustements vient compléter de façon très instructive le questionnement que nous avons développé tout au long des différentes parties de cette thèse. Ces parties nous ont fait identifier deux horizons limites insatisfaisants : un horizon objectiviste fréquentiste, et un horizon subjectiviste bayésien.

Selon l'approche objectiviste, la mesure a pour but premier de permettre l'évaluation de

l'état d'une réalité physique ; un résultat de mesure est alors d'autant meilleur que la valeur obtenue est proche de la valeur vraie de la grandeur mesurée. L'erreur de mesure et l'exactitude de mesure sont alors les outils privilégiés pour qualifier un résultat de mesure.

L'évaluation d'un résultat de mesure selon l'approche objectiviste est problématique en ce qu'il est *impossible* de connaître l'erreur de mesure commise – l'expérimentateur n'a pas accès à l'envers du décor, il ne peut pas s'appuyer sur la donnée de la valeur vraie de la grandeur pour évaluer l'erreur de mesure, puisqu'il cherche précisément à évaluer la valeur vraie à partir de la donnée de l'erreur de mesure ; c'est le problème de l'inconnaissabilité de la valeur vraie. C'est pourquoi certains fréquentistes, comme Willink, insistent sur l'importance d'identifier un critère objectif de performance qui permette de justifier les raisons pour lesquelles on accorde du crédit à un résultat de mesure. Le critère de performance qu'ils proposent est le taux de succès de la procédure de mesure, c'est-à-dire sa propension à produire des intervalles qui contiennent effectivement la valeur vraie de la grandeur avec une fréquence limite satisfaisante (voir le chapitre 5). Willink reconnaît bien sûr qu'il n'est pas possible de mesurer le taux de succès d'une procédure, puisque, encore une fois, la valeur vraie est inconnue ; mais il affirme que l'on peut prédire ce taux de succès et, sur la base de cette prédiction, alors avoir de bonnes raisons de penser que la mesure donne accès à la connaissance de la valeur vraie avec une assurance que l'on peut décrire par une probabilité.

Nous avons insisté sur les vertus d'une telle conception : celle-ci met en avant un caractère normatif de la mesure en rappelant la possibilité d'être en erreur ; et c'est cette possibilité qui permet l'initialisation d'un processus de *correction*. Dès lors, identifier la *cible* de la mesure, représentée par la valeur vraie de la grandeur visée, fournit aux scientifiques un critère d'action pour engager des processus de correction permettant la continuation du progrès expérimental. Cependant, l'horizon objectiviste n'est pas totalement satisfaisant car le problème de l'inconnaissabilité reste en partie valable, du fait qu'il est illusoire de prétendre connaître l'exactitude d'un résultat. Si l'argument d'inconnaissabilité ne mène pas nécessairement à remettre en question le concept de valeur vraie d'une grandeur, il rappelle toutefois que l'on ne peut jamais garantir l'erreur de mesure qui affecte le résultat obtenu, c'est-à-dire, de façon équivalente, l'exactitude de mesure qui constitue pourtant l'objectif de l'évaluation d'un résultat de mesure selon l'approche fréquentiste.

C'est pourquoi les bayésiens voient dans la formulation fréquentiste d'un résultat de mesure un objectif inatteignable. Les fréquentistes ne peuvent à aucun moment garantir que le résultat qu'ils formulent est correct, et, de ce fait, leurs exigences d'exactitude n'ont pas non plus de sens. Cela signifie qu'il faut se résoudre à accepter certaines objections émises par les bayésiens, en particulier l'importance accordée à la notion d'incertitude de mesure comprise comme un concept épistémique qui ne qualifie pas le résultat lui-même mais la connaissance que l'expérimentateur pense avoir de la grandeur visée.

Pour autant, l'horizon subjectiviste que font percevoir les bayésiens est également insatisfaisant car, en formulant un résultat de mesure uniquement en terme de croyance, il interdit que celui-ci dise quelque chose sur le monde. En particulier, le résultat de mesure ne peut plus être évalué et testé. De ce fait, on abandonne toute prétention à produire des propositions vraies sur le monde. En un sens, la formulation subjectiviste des résultats de mesure apparaît à certains

métrologues fréquentistes comme une tautologie : un expérimentateur exprime par un résultat de mesure ce qu'il croit qu'il connaît, mais la connaissance qu'il exprime prend seulement la forme d'un degré de croyance.

La perspective bayésienne présente un autre écueil. En adoptant une position qui est celle d'une description des degrés de croyance, les bayésiens centrent l'étude sur une analyse des connaissances *dans le présent* et éliminent un aspect normatif essentiel de la mesure – la valeur vraie comme cible – masquant ainsi les raisons qui les enjoignent de continuer le processus de correction *dans le futur*. Il leur faut alors mobiliser de l'extérieur un critère d'action, c'est-à-dire ajouter à la structure bayésienne qui permet la description des connaissances des agents une structure supplémentaire qui explique ce que cette connaissance permet de conclure ou de faire.

N'aurait-on donc le choix qu'entre ces deux horizons insatisfaisants, entre un objectif inatteignable et un objectif atteignable mais qui ne dit rien sur le monde? La question n'est donc pas tant de savoir si les statistiques fréquentiste et bayésienne sont chacune adaptées ou non à l'analyse d'incertitude, mais de savoir si la conceptualisation de la mesure qui l'accompagne est féconde. C'est là que les ajustements des constantes de la physique se révèlent particulièrement intéressants sur le plan épistémologique, car ils permettent de contraster les conceptualisations parfois très théoriques que nous avons exposées dans les premières parties, avec un *usage*, celui que font les acteurs de la physique de précision, en regard de leurs traditions d'une part, et des enjeux de leur recherche d'autre part. Cette usage nous permet de confronter la façon dont l'incertitude de mesure est mise à contribution dans les raisonnements des physiciens avec la façon dont celle-ci est définie dans les différents modèles statistiques de la mesure auxquels nous nous sommes intéressés jusqu'ici. Or, cet usage, dont témoigne par exemple la position tenue par Taylor au début des années 1970, montre que les physiciens se placent dans un entre-deux qui nous semble fécond, sans être une solution universelle aux problèmes posés. Les physiciens reconnaissent les limites de la notion d'exactitude de mesure, mais n'abandonnent cependant pas la dimension normative véhiculée par le concept d'erreur de mesure. Ils ne résolvent pas pour autant les questions philosophiques que nous soulevons, mais ils permettent de nuancer les velléités exprimées dans les discussions animées par les métrologues fréquentistes et bayésiens ces dernières décennies. L'exemple des ajustements des constantes illustre de façon très parlante en quoi les horizons décrits ici sont tous deux insatisfaisants. Il est un bon exemple de la façon dont les scientifiques se placent non pas à l'extrême de l'un ou l'autre des horizons, mais quelque part au milieu.

11.2 Ajustements et perspective objectiviste

Quelle est la conception de l'incertitude de mesure qui ressort de l'usage qui est fait de ce concept dans les ajustements de constantes de la physique? Nous pouvons d'abord raisonner par la négative, en cherchant à identifier ce que l'incertitude de mesure n'est pas. Une première proposition est la suivante :

Proposition 1.1. Plus l'incertitude de mesure associée à un résultat est faible, plus le résultat est considéré comme exact.

L'on peut encore reformuler la proposition, de façon presque équivalente, par :

Proposition 1.2. L'incertitude de mesure est une estimation de l'amplitude de l'erreur finale.

Quelle que soit la position tenue, ces propositions ne fonctionnent pas. Nous avons vu à de multiples reprises que l'incertitude de mesure ne peut pas être considérée comme la mesure de l'exactitude. En effet, il n'est jamais possible de garantir que toutes les erreurs de mesure ont été correctement identifiées, et certaines erreurs inconnues peuvent engendrer un biais bien supérieur à l'amplitude de l'incertitude de mesure. Si nous reconnaissons qu'il n'est jamais possible d'être certain qu'aucune erreur de mesure inconnue ne subsiste à la fin de l'analyse d'un résultat, alors nous devons admettre à leur suite que l'incertitude de mesure ne peut pas être une évaluation de l'erreur possible, à la suite du GUM par exemple :

[M]ême si les incertitudes évaluées sont faibles, il n'y a, pour autant, aucune garantie que l'erreur sur le résultat de mesure soit faible, parce que, dans la détermination d'une correction ou dans l'évaluation des lacunes d'une connaissance, on peut aussi négliger un effet systématique qui n'aurait pas été reconnu. Ainsi, l'incertitude du résultat d'un mesurage n'indique pas nécessairement la proximité vraisemblable du résultat de mesure et de la valeur du mesurande³

Ce constat constitue la seule véritable conséquence tangible de l'argument d'inconnaissabilité : il n'est pas possible de *garantir* que l'on connaît parfaitement l'erreur finale sur un résultat de mesure. C'est effectivement un constat indiscutable, sur lequel s'accordent tous les acteurs de la métrologie. Cependant, un désaccord émerge lorsqu'on s'intéresse à l'ambition que l'on peut nourrir dans l'évaluation d'un résultat de mesure, et à l'objectif que poursuit cette évaluation. Ainsi, les métrologues attachés au caractère objectif de la mesure, en particulier un défenseur de l'approche fréquentiste comme Willink, auront tendance à soutenir la proposition suivante :

Proposition 2.1. L'objectif de l'évaluation d'un résultat de mesure est de déterminer une incertitude de mesure qui soit la meilleure estimation possible de l'exactitude du résultat.

Cela peut être encore exprimé par l'idée suivante :

Proposition 2.2. Il est souhaitable que l'erreur finale de mesure soit approximativement contenue dans les limites de l'incertitude de mesure.

L'introduction du critère de « taux de succès » par les fréquentistes mène également à réécrire la proposition de façon différente :

Proposition 2.3. Si l'évaluation du résultat de mesure est effectuée correctement, alors l'incertitude de mesure doit estimer l'erreur finale avec un taux de succès supérieur au niveau de confiance annoncé.

La conception de l'incertitude de mesure dans les ajustements ne correspond pas non plus à ces propositions. L'opposition entre l'approche sécuritaire que l'on devine chez Bender et l'approche de Taylor, décrite au chapitre précédent, illustre ce point. La position sécuritaire a été

3. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.53.

critiquée à maintes reprises dans les comptes-rendus d'ajustement, mais aussi dans le GUM⁴. La position défendue par Taylor, en réponse à Bender, est explicite : dans un ajustement des constantes de la physique, les valeurs produites elles-mêmes n'ont pas une grande importance. Taylor insiste sur le fait que l'exactitude des valeurs recommandées est un souci mineur, le point le plus important étant que *celles-ci puissent être corrigées dans le futur* si jamais elles se révèlent erronées. Plus encore, fait remarquer Taylor, si l'on surestime une incertitude de mesure, afin de s'assurer que celle-ci soit supérieure à l'erreur finale commise, on affaiblit le pouvoir conclusif du résultat de mesure, et on peut de ce fait masquer certains erreurs que l'analyse aurait amené à dévoiler. À l'inverse, relâcher ponctuellement, dans l'immédiat, l'exigence d'exactitude rend plus féconde la recherche des erreurs de mesure. De fait, l'incertitude de mesure représente la résolution avec laquelle le résultat de mesure semble pouvoir permettre de sonder les théories physiques ; il est donc parfaitement contre-productif de la surestimer. L'objectif n'est pas de valoriser le fait de faire des erreurs mais d'accepter de *s'exposer* à l'erreur.

En ce sens, il semble y avoir des similitudes avec l'approche objectiviste prônée par les métrologues fréquentistes, en ce que les deux conceptions mettent en avant l'importance de raisonner sur les succès *possibles*, dans le futur, de l'enquête expérimentale. Cependant, l'exemple des ajustements, et, plus généralement, les exemples issus de la mesure des constantes de la physique⁵, montrent que le critère de taux de succès défendu par les fréquentistes ne fonctionne pas. Ainsi que l'ont fait remarquer par exemple Taylor, Parker et Langenberg, les valeurs ajustées de nombreuses constantes physiques ont connu des sauts significatifs (voir figure 10.1 de la section précédente). Ce constat est incompatible avec l'exigence, formulée par Willink, de voir les intervalles d'incertitude englober la valeur de la constante mesurée avec un taux de succès satisfaisant. La différence entre l'approche d'un Willink et celle adoptée dans les ajustements des constantes tient à ce qu'à chaque fois qu'une erreur significative a été identifiée, c'est la preuve que les résultats obtenus jusqu'alors étaient tout simplement erronés, et qu'il n'y a plus aucune raison de considérer que le résultat de mesure formulé alors contient la valeur vraie de la grandeur. Or, revendiquer que le caractère erroné ou non d'une valeur ajustée n'est pas problématique *dans l'instant*, comme le font Taylor et d'autres physiciens, revient à affirmer qu'il n'y a aucune raison que les résultats de mesure soient des succès dans l'immédiat, et, par suite, qu'il n'y a aucune raison à ce que le taux de succès de l'entreprise de mesure soit égale à la probabilité associée aux intervalles d'incertitude.

Doit-on en conclure qu'il n'y a aucun lien entre incertitude de mesure et erreur de mesure ? La réponse ne doit pas forcément être aussi radicale. De fait, l'incertitude de mesure est fondée sur une analyse de l'erreur. Elle s'appuie sur une estimation de l'amplitude possible des erreurs

4. Voir l'annexe E du GUM, en particulier pp.56–57. Il est ainsi précisé que « la méthode [du GUM] se trouve en désaccord avec certaines méthodes plus anciennes qui ont en commun les deux idées suivantes. [...] La première idée est que l'incertitude déclarée devait être "sûre" ou "conservatoire", ce qui signifie qu'elle ne devait jamais risquer d'être sous-estimée. En fait, parce que l'évaluation de l'incertitude d'un résultat de mesure est problématique, cette idée conduisait souvent à l'élargir délibérément. », *Joint Committee for Guides in Metrology (JCGM) (2008d)*, p.56.

5. Un exemple tout à fait similaire est présenté par *Lautrup et Zinkernagel (1999)* dans leur histoire de la mesure de l'anomalie de moment magnétique de l'électron. Cet exemple puise lui aussi dans la mesure des constantes de la physique, mais n'est pas étudié à travers le prisme des ajustements.

aléatoires à partir d'une mesure de la variabilité des résultats, et sur une estimation des limites possibles des corrections apportées aux erreurs systématiques qui ont été identifiées. Les propositions 1 et 2 ne fonctionnent *qu'à la condition supplémentaire* que toutes les sources d'erreur de mesure ont été identifiées; leur validité est donc suspendue à une hypothèse. C'est ce que conclut le GUM :

En général, l'erreur exacte du résultat d'un mesurage n'est pas et ne peut pas être connue [...] C'est seulement lorsqu'il y a une base solide pour croire que tout cela a été fait correctement, sans oublier aucun effet systématique significatif, que l'on peut supposer que le résultat de mesure est une estimation fiable de la valeur du mesurande et que son incertitude-type composée est une mesure fiable de son erreur possible.⁶

Par conséquent, on peut défendre la proposition suivante, moins forte que les précédentes :

Proposition 3. L'incertitude de mesure est une évaluation approximative de la composante de l'erreur finale due aux sources d'erreur qui ont été identifiées (erreurs aléatoires incluses).

La proposition 3 nous semble s'accorder correctement à la fois avec l'usage qui est fait de l'incertitude dans les ajustements, et avec la façon dont celle-ci est déterminée lorsqu'on suit la méthodologie du GUM.

On comprend alors le positionnement de Taylor sous un autre angle. Dans l'instant, il n'y a aucune raison de penser que toutes les sources d'erreur ont été identifiées, et, par conséquent, d'exiger que les résultats de mesure soient exacts; mais dans le futur, il est souhaitable que les sources d'erreur soient peu à peu dévoilées. L'approche de Taylor, et celle qui a été généralement adoptée dans les ajustements de la constante de la physique jusqu'ici, consiste donc à adopter une approche *dynamique* focalisée sur la possibilité de corriger les erreurs plutôt qu'une approche statique centrée sur l'exactitude *présente* des résultats de mesure.

De fait, l'abandon de l'exigence immédiate d'exactitude n'est certainement pas le signe que les physiciens abandonnent une perspective évaluative, tant leur conception reste ancrée autour du concept central d'erreur de mesure. L'attachement des physiciens à l'erreur de mesure témoigne que ces derniers n'embrassent pas pleinement la perspective descriptive véhiculée par l'approche épistémique de la mesure – et en particulier par certains métrologues bayésiens. Le constat effectué se maintient jusqu'à aujourd'hui avec l'exemple de la mesure du rayon du proton : la décision des membres du CODATA de « maintenir l'énigme comme une énigme », témoigne de leur volonté de ne pas se satisfaire d'une simple description d'un état de connaissance qui ressemblerait alors trop à un *statu quo*, mais au contraire de maintenir une certaine tension qui stimule les recherches futures.

La position adoptée au long des comptes-rendus d'ajustement, défendue en particulier par Taylor, consiste à dire que la confrontation des résultats de mesure fait émerger le progrès scientifique par la mise en évidence puis l'explication, et enfin l'élimination d'erreurs de mesure⁷. La clé du progrès scientifique est le processus de correction, qui est alors compris comme

6. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.53.

7. Le mode de fonctionnement prôné par Taylor s'accorde tout à fait avec la façon dont Lautrup et Zinkernagel

une activité sans fin. Les physiciens ne se contentent pas d'identifier la trace quantitative de la présence d'une erreur, pour en rendre compte par une théorie mathématique qui permettrait la réduction ou l'élimination numérique de cette erreur. Leur objectif est de remonter à la cause de l'erreur de mesure, d'identifier les phénomènes physiques, les erreurs théoriques, les artefacts expérimentaux, etc. qui en sont à l'origine, de façon à avoir une explication des raisons pour lesquelles leurs différentes mesures sont en désaccord. L'énigme du rayon du proton ne sera résolue que lorsque les résultats pourront être réconciliés, ou lorsqu'il sera prouvé que l'un d'entre eux est manifestement faux. De ce fait, l'approche des physiciens est en accord avec ce que préconise Hon :

D'un point de vue épistémologique, il est utile d'enquêter sur la source de l'erreur et non pas tant d'examiner les caractéristiques mathématiques de l'erreur et les moyens de l'éliminer par calcul – la fonction causale étant d'un intérêt plus élevé ici que la fonction pragmatique.⁸

L'attention des physiciens porte donc à la fois sur l'aspect causal (identifier les sources d'erreur) et l'aspect pragmatique (aboutir à un consensus sur les valeurs recommandées).

11.3 Ajustements et perspective épistémique

Nous avons vu que la position objectiviste défendue par les fréquentistes ne s'accorde pas bien avec l'usage qui est fait de l'incertitude de mesure dans les ajustements des constantes de la physique. Est-il possible d'accorder cet usage avec une interprétation de l'incertitude de mesure en termes épistémiques? Armatte a suggéré, à partir de son étude de la mesure de la longueur du méridien terrestre au XVIII^e siècle, que « (l)a précision d'une mesure est un indicateur de sa valeur sur le marché de la preuve »⁹. Cela peut être interprété au travers de l'idée courante selon laquelle l'incertitude de mesure indique la *confiance* (ou plutôt le manque de confiance) que l'on peut avoir en un résultat. On en vient alors à la proposition suivante :

Proposition 4. Plus l'incertitude de mesure associée à un résultat de mesure est faible, plus on accorde de confiance à ce résultat, dans le sens où l'on sera davantage tenté d'accepter la validité du résultat et d'en faire effectivement usage à des fins théoriques ou pratiques.

Il est intéressant de remarquer que lors de la conférence sur les ajustements menée dans les locaux du NIST en 1970, plusieurs intervenants avaient déjà développé une position préfigurant l'emploi des probabilités épistémiques – bien que cela prenait alors une forme encore assez classique quant à la conception de la mesure. Ainsi, Thomsen explique-t-il, à propos de l'estimation d'une erreur systématique :

Comment alors pouvons-nous trouver un fondement raisonnable pour attribuer une valeur numérique à une erreur systématique? Dans la plupart des cas, la réponse la plus réaliste semble être que l'expérimentateur doit reconnaître qu'il cite

décrivent l'histoire de la mesure de l'anomalie du moment magnétique de l'électron : « (v)u dans son ensemble, le développement historique consiste en un dépouillement progressif de l'environnement de l'électron, avec une élimination correspondante des erreurs systématiques », [Lautrup et Zinkernagel \(1999\)](#), p.107.

8. [Hon \(2003\)](#), p.183.

9. [Armatte \(2011\)](#), p.23.

des cotes de paris. Ainsi, s'il déclare un résultat comme (100 ± 1) cm (erreur probable) il doit affirmer qu'il y a une probabilité de 50 pourcent que la valeur vraie se situe entre 99 cm et 101 cm. S'il a constitué son estimation de l'erreur honnêtement, évitant à la fois l'excès d'optimisme et le conservatisme excessif, *il devrait être prêt à choisir n'importe laquelle des deux options du pari*. C'est là l'essence d'une estimation honnête de l'erreur.¹⁰

Son collègue Franken exprime une idée similaire :

Imaginez que vous deviez miser 500 \$ [...] sur la proposition que c'est un intervalle d'erreur correct, et que vous me laissiez choisir de quel côté du pari je me place.¹¹

Pour autant, cette proposition ne correspond toujours pas très bien à l'usage de l'incertitude de mesure dans les ajustements. Certes, elle fonctionne plutôt bien lorsqu'on s'intéresse à la façon dont les mesures sont combinées entre elles. De fait, le chapitre 9 a montré que, lorsque Birge (puis, à leur tour, les physiciens qui lui succèdent) combine différentes mesures d'une même constante dans une moyenne pondérée, le poids le plus élevé est accordé aux résultats dont l'incertitude de mesure est la plus faible. Cela semble venir nourrir une interprétation de l'incertitude de mesure en termes de confiance. Cependant, lorsque certains résultats se trouvent être en désaccord significatif, il faut procéder à une sélection ; or, dans ce cas précis, l'incertitude de mesure n'est *pas* le critère principal de choix, voire n'entre pas du tout dans la procédure de sélection. Dans les discussions sur le rayon du proton, décrites au chapitre précédent, les physiciens ont choisi de conserver (temporairement) la valeur obtenue par les expériences les plus anciennes, où la mesure est effectuée sur un hydrogène ordinaire. C'est pourtant le résultat qui présente l'incertitude de mesure la plus élevée. Mais, de fait, l'objectif n'était pas ici de juger laquelle des deux valeurs proposées était la plus vraisemblable sur la base de la valeur de l'incertitude de mesure associée à chaque résultat ; il s'agissait plutôt de juger lequel des deux résultats, valeur *et* incertitude considérés ensemble, semblait le plus solide. Ici, l'interprétation de l'incertitude de mesure comme mesure de la confiance est inadéquate. C'est pourquoi nous avons conclu les chapitres 9 et 10 par l'idée que l'incertitude de mesure ne peut pas à elle seule évaluer un résultat de mesure, et qu'il faut lui superposer un jugement extérieur qui évalue la qualité de la mesure elle-même (incluant une évaluation de la façon dont l'expérimentateur a déterminé l'incertitude de mesure).

Si l'incertitude de mesure ne décrit pas la confiance que l'on accorde à un résultat, nous pouvons nous tourner vers la formulation préconisée depuis le tournant épistémique en métrologie :

Proposition 5.1. L'incertitude de mesure est la réciproque logique d'un état de connaissance.

La difficulté de cette proposition tient au fait qu'il faut préciser ce qui est entendu par « connaissance ». La définition philosophique classique de la connaissance comme « croyance vraie justifiée »¹² est de nouveau problématique, puisque rien ne garantit l'exactitude du résultat, et,

10. Thomsen (1971), p.503 (Thomsen souligne).

11. Franken (1971), p.508.

12. Ichikawa et Steup (2014)

par suite, la validation de la condition de « vérité ». C'est pourquoi l'on sera tenté d'interpréter l'incertitude exclusivement en termes de croyances, comme le font les bayésiens subjectivistes :

Proposition 5.2. L'incertitude de mesure traduit la croyance de l'expérimentateur (ou d'un groupe d'expérimentateurs) quant aux valeurs possibles de la grandeur visée.

Cela nous renvoie à la description donnée dans le GUM :

L'incertitude de mesure peut ainsi être décrite comme la mesure de la façon dont on croit que l'on connaît la valeur vraie essentiellement unique du mesurande.¹³

Les propositions 5.1 et 5.2 semblent mieux résister à la critique que la proposition 4. En effet, en intégrant la notion de croyance à la caractérisation même de l'incertitude, on se protège contre le problème de l'inconnaissabilité qui mine chacune des propositions que nous avons tenté de formuler jusqu'ici. Mais il subsiste une difficulté, car la notion de croyance reste ici un peu ambiguë. L'incertitude de mesure a été déterminée par l'expérimentateur en prenant en compte toutes les sources de doute qu'il a pu recenser, en cherchant à n'en oublier aucune, mais sans pour autant introduire artificiellement des éléments de doute qui n'ont pas de fondement rationnel. De ce fait, l'incertitude de mesure ne représente pas tant ce que croit l'expérimentateur, mais correspond plutôt à ce qu'il est *le plus raisonnable* pour lui de croire dans l'instant, *s'il s'en tient* aux informations disponibles, et aux sources d'erreur qu'il a pu identifier. On en revient de fait à la proposition 3 : une fois encore, la croyance est assujettie à l'absence d'erreur de mesure inconnue.

De fait, que l'on cherche à formuler l'incertitude de mesure selon une acception objective ou subjective, on se retrouve confronté invariablement au même problème : la difficulté réside dans les erreurs systématiques inconnues¹⁴. Comme l'a fait remarquer Stefaan Pommé, la difficulté de l'enquête expérimentale et de l'interprétation de l'incertitude de mesure ne provient pas des « inconnues connues » mais des « inconnues inconnues »¹⁵. Ainsi, l'incertitude de mesure ne caractérise pas tant ce que l'expérimentateur ne sait pas, ni ce que l'expérimentateur sait, mais uniquement ce qu'il *sait qu'il ne sait pas*. On en vient donc à la proposition 6 :

Proposition 6. L'incertitude de mesure représente ce que l'expérimentateur sait qu'il ne sait pas, sans aucune présupposition sur ce qu'il ne sait pas qu'il ne sait pas.

13. Joint Committee for Guides in Metrology (JCGM) (2008d), p.3.

14. Eisenhart le faisait déjà remarquer : « (l)'erreur systématique globale d'un processus de mesure consiste ordinairement d'"erreurs systématiques" élémentaires dues à la fois à des causes assignables et non assignables. Celles d'origine inconnue (non envisagées, pas encore identifiées, pas encore découvertes) sont toujours à craindre », Eisenhart (1963), p.183.

15. Je reprends cette expression de l'intervention de Stefaan Pommé lors de l'atelier du BIPM consacré aux incertitudes de mesure, organisé dans les locaux du BIPM, à Sèvres, les 15 et 16 juin 2015 (Pommé [2015]), qui s'inspirait déjà en cela d'une citation devenue célèbre de Donald Rumsfeld, en février 2002, lequel était alors Secrétaire de la Défense des États-Unis et cherchait à défendre la position du gouvernement face au manque de preuves de la présence d'armes de destruction massive en Irak (nous laissons ici la formulation originale) : "reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones". Cette remarque s'adapte parfaitement à la démarche scientifique.

Cette proposition n'est rien d'autre que la proposition 3 formulée en des termes épistémiques.

La dialectique des croyances et des rapports de confiance nous renvoie à l'état de « dépendance épistémique », pour reprendre le terme introduit par John Hardwig¹⁶, dans lequel sont les différents acteurs du monde scientifique les uns envers les autres. Si un expérimentateur X obtient un résultat donné, pourquoi un autre expérimentateur Y devrait-il croire ce résultat, et en faire usage ? Y ne prend pas automatiquement pour argent comptant ce que communique X ; il intègre une réflexion quant au résultat lui-même mais également quant à X . Cette évaluation pourrait être exprimée à travers la confiance que Y a envers X . Ainsi, la confiance de Y dans le résultat est le produit de la confiance de Y envers X et de celle de X envers le résultat.

Si l'on s'en tient à une approche bayésienne subjectiviste, cette description s'intègre assez naturellement dans une conception où la probabilité est individuelle, chaque acteur ayant à propos d'une proposition donnée sa *propre* probabilité. Dans ce cas, l'incertitude de mesure est relative à un agent connaissant ; lorsque X transmet le résultat, l'incertitude qu'il communique est *son* incertitude, à laquelle Y est tenu d'adhérer ou non. Cette proposition revient alors à s'interroger sur les degrés de croyance des acteurs à l'échelle individuelle, et renvoie à la question de la dépendance épistémique à celle des rapports de dépendance envers les experts. Mais c'est également là le défaut de cette conception : elle met en avant une dimension profondément individualiste du savoir – en aplatissant la question de la valeur d'un résultat à la seule croyance subjective. À l'inverse, les remarques d'un métrologue fréquentiste comme Willink conduisent à envisager une autre perspective, ou prendre acte de la dépendance épistémique incite à garder conscience du caractère social de l'activité scientifique, et à replacer la question de l'erreur de mesure dans une conception élargie de l'expérience qui intègre les considérations sociales liées à la division du travail¹⁷.

11.4 Les neutrinos « supraluminiques »

Un exemple extérieur aux ajustements vient illustrer en quoi les physiciens n'assimilent pas incertitude de mesure et confiance, ni même incertitude de mesure et croyance. Le 23 septembre 2011, les collaborateurs de l'expérience internationale « Opera », « dédiée à l'observation d'un faisceau de neutrinos produit par les accélérateurs du CERN à Genève et détecté 730 km plus loin depuis le laboratoire souterrain de Gran Sasso en Italie »¹⁸, annoncent avoir observé des neutrinos plus rapides que la vitesse de la lumière dans le vide¹⁹, un résultat en complet désaccord avec la théorie de la relativité. L'annonce connaît un très grand retentissement, et est immédiatement reprise dans les médias généralistes. L'expérience attire d'autant plus l'intérêt des médias et du public qu'elle est liée à deux icônes de la physique moderne, la théorie de la relativité et Albert Einstein. Ainsi le quotidien en ligne *lefigaro.fr* titre-t-il « Relativité : Einstein contredit par des chercheurs du CNRS »²⁰.

16. Hardwig (1985)

17. Grégis et de Courtenay (2016), p.50.

18. CNRS (2011)

19. Adam, T. *et al.* (2011)

20. Vanlerberghe (2011)

L'écart relatif mesuré entre la vitesse moyenne des neutrinos et la vitesse théorique de la lumière dans le vide est d'environ 25 parts par million, c'est-à-dire que la différence de vitesse est de l'ordre de 25 millièmes de la vitesse de la lumière. De plus, cet écart est six fois supérieur à l'incertitude associée au résultat : c'est donc que le résultat est très significatif, ce qui pourrait être interprété en termes probabilistes en exprimant qu'il y aurait environ une chance sur un milliard que l'écart soit dû à une mauvaise estimation des sources d'erreurs connues, *s'il ne subsistait par ailleurs aucune erreur de mesure*. C'est cette dernière condition qui est la plus délicate. La plupart des réactions modérées réagissent justement en rappelant que la validité du résultat est suspendue à une identification complète des sources d'erreur. Comme le précise Jean-Marc Lévy-Leblond :

Ces neutrinos traversent la Terre sur 730 km en quelques millisecondes avant de pouvoir être détectés (pour une toute petite partie d'entre eux). Il faut mesurer les temps d'arrivée et de départ avec une précision de quelques nanosecondes et la longueur du trajet parcouru avec une précision de 2 cm (la mesure se fait par GPS). On conçoit la difficulté de telles mesures. Le point le plus faible de l'expérience est le repérage de l'instant de départ des neutrinos, qui n'est défini que statistiquement.²¹

L'hebdomadaire *Le Point* publie un entretien avec le physicien Thibault Damour, où celui-ci exprime des réserves similaires :

- Le Point.fr : Vous n'y croyez donc pas ?
- Thibault Damour : A priori, je suis très prudent. [...] il y a eu un nombre massif d'expériences concernant la relativité restreinte et la vitesse des neutrinos qui n'ont jamais trouvé de violations.
- Le Point.fr : Alors ces physiciens ont triché ou commis une erreur grave ?
- Thibault Damour : Triché ? Absolument pas. C'est une équipe très sérieuse et hors de tout soupçon. Il faut dire que j'ai déjà vu dans ma carrière, à plusieurs reprises, des annonces aussi fracassantes qui se sont révélées fausses par la suite, comme la découverte d'ondes gravitationnelles. Je veux simplement dire que la science est basée sur l'expérience et que des erreurs se produisent parfois. Du reste, l'équipe en question en a conscience. Elle livre son expérience à la communauté scientifique pour que celle-ci passe tout en revue. Et c'est sain qu'il y ait parfois des erreurs. Si la science n'en faisait jamais, cela signifierait qu'elle tricherait.²²

Les réactions des scientifiques montrent ainsi que la communauté scientifique, dans l'ensemble, accueille avec scepticisme le résultat obtenu. Ce n'est pas l'amplitude de l'incertitude de mesure évaluée par les scientifiques de la collaboration OPERA qui guide le scepticisme des commentateurs, mais une réflexion externe sur la validité du résultat dans son ensemble. L'incertitude de mesure qu'ont évaluée les collaborateurs de l'expérience ne sert à déterminer la significativité du résultat *qu'à la condition que celui-ci soit correct*, c'est-à-dire qu'aucune source d'erreur n'ait été oubliée. En revanche, pour juger de la validité du résultat, la communauté scientifique réfléchit en termes d'*erreurs de mesure* et non d'incertitude de mesure.

21. Petit et Lévy-Leblond (2011)

22. Lewino (2011)

La communauté accueille le résultat avec scepticisme. Mais, de façon plus remarquable encore, c'est également le cas des membres de la collaboration OPERA. Eux-mêmes émettent des réserves significatives :

Malgré la grande significativité de la mesure rapportée ici, et la stabilité de l'analyse, l'impact potentiellement grand du résultat motive la poursuite de nos études de façon à investiguer des effets systématiques possibles et toujours inconnus qui pourraient expliquer l'anomalie observée. Nous ne tentons délibérément pas aucune interprétation théorique ou phénoménologique des résultats.²³

Le CNRS avait même anticipé le retentissement du résultat en publiant dès la veille de l'annonce un communiqué de presse très prudent, expliquant en particulier que :

Jusqu'ici, la vitesse de la lumière a toujours été considérée comme une limite infranchissable. Si ce n'était pas le cas, cela pourrait ouvrir des perspectives théoriques complètement nouvelles. Compte tenu de l'énorme impact qu'un tel résultat pourrait donc avoir pour la physique, des mesures indépendantes s'avèrent nécessaires afin que l'effet observé puisse être réfuté ou bien formellement établi. C'est pourquoi les chercheurs de la collaboration OPERA ont souhaité ouvrir ce résultat à un examen plus large de la part de la communauté des physiciens.²⁴

De fait, la communauté scientifique ne peut pas se permettre d'abandonner brutalement une théorie aussi solide que la théorie de la relativité, sur la seule base d'une expérience menée par une unique laboratoire, quelle que soit l'apparente significativité de leurs résultats.

Durant les mois qui suivent la première annonce, la collaboration s'affaire à perfectionner la mesure et à identifier d'éventuelles erreurs. Dans un premier temps, elle publie de nouveaux résultats qui viennent confirmer l'observation initiale²⁵. Cependant, le 23 février 2012, le CERN annonce que les chercheurs ont identifié deux causes possibles de désaccord avec la théorie²⁶ ; puis, le 16 mars 2012, confirme que les pistes suivies semblent indiquer avec force que le désaccord observé n'est en fait qu'un artefact expérimental dû à une erreur de mesure. Ce constat sera entériné le 8 juin 2012²⁷ à l'issue de la 25^e Conférence internationale sur la physique des neutrinos et l'astrophysique, tenue à Kyoto. Deux erreurs ont été identifiées : l'une concerne une dérive dans l'horloge utilisée pour mesurer le temps de vol des neutrinos ; l'autre provient d'un défaut de branchement d'une fibre optique responsable de la synchronisation des horloges employées pour la mesure. En témoignage du fait que personne n'a peut-être jamais vraiment cru au résultat, le directeur de la recherche au CERN, Sergio Bertolucci, émet la remarque suivante : « (m)ême si ce résultat n'est pas aussi sensationnel que certains l'auraient souhaité [...] il correspond à ce que nous attendions tous au fond de nous-mêmes »²⁸.

23. Adam, T. *et al.* (2011), p.22.

24. CNRS (2011).

25. Adam, T. *et al.* (2013) ; l'article a initialement été publié sur la plate-forme en ligne arXiv.org le 17 novembre 2011.

26. CERN (2011)

27. CERN (2011)

28. CERN (2011)

À l'annonce de l'identification de l'erreur, les réactions sont partagées. Dans le grand public, la déception est parfois le miroir des attentes qu'avait suscité l'enthousiasme initial. En particulier, le caractère apparemment grossier de l'erreur commise contraste avec les conséquences théoriques très profondes que laissait entrevoir le résultat. Ainsi, le quotidien *Corriere della Sera* qualifie l'ensemble de l'opération de « flop des physiciens »²⁹ Dans la sphère plus spécialisée, on insiste sur le fait que tout s'est fait selon les canons de la bonne science, et on insiste sur les vertus épistémiques de la recherche de l'erreur³⁰. Certains membres de la collaboration font remarquer qu'il aurait vraisemblablement été impossible d'identifier les erreurs aussi rapidement si le résultat n'avait pas été rendu public, et concluent que la publication d'un résultat erroné a résulté dans une accélération des recherches. Pourtant, l'expérience reste toutefois vécue comme un échec et laisse des traces parmi les collaborateurs³¹. Le 30 mars, le directeur de la collaboration, Antonio Ereditato, démissionne de son poste³², au lendemain d'un dépôt, par plusieurs groupes de la collaboration, d'une motion de défiance – bien que celle-ci ait été rejetée. Au-delà des considérations liées aux relations humaines entre chercheurs, certains parmi eux reprochaient à la direction d'avoir agi trop précipitamment et de ne pas avoir attendu des confirmations supplémentaires avant de publier le résultat³³.

L'exemple des neutrinos supraluminiques vient illustrer en quoi chacune des propositions 1 et 2, ainsi que 4 et 5, ne fait jamais que recouvrir partiellement l'appréhension que l'on a du concept d'incertitude de mesure. Dans ce cas, on peut effectivement interpréter l'incertitude de mesure comme une tentative d'évaluation de l'erreur finale, mais les physiciens se doutent largement que cette évaluation est susceptible d'être sous-estimée. Elle ne correspond pas à la confiance que les physiciens accordent au résultat, puisque ni la communauté ni les membres de la collaboration eux-mêmes ne semblent réellement y croire. Ce sont les propositions 3 et 6, qui assujettissent l'interprétation de l'incertitude de mesure à une hypothèse – l'absence d'erreur de mesure inconnue – qui donnent un aperçu de ce que signifie réellement l'incertitude de mesure.

Dans cet exemple, la caisse de résonance des médias généralistes a amplifié certaines visions mythifiées et idéalisées de la science et a peut-être contribué à une déformation du débat scientifique, ce qui a amené la communauté comme le grand public à percevoir la publication de ce résultat comme un véritable échec scientifique. Pour autant, si l'on suit la position de Taylor sur les ajustements, la publication par OPERA d'un résultat erroné n'est pas en soi une erreur. L'erreur, s'il peut y en avoir une, ne porte pas sur la publication du résultat en elle-même mais dépend des conséquences que l'on en tire. Cela montre en quoi la perspective dynamique tournée vers le progrès futur est adaptée pour les perspectives de recherche fondamentale, où l'objectif principal est le perfectionnement des connaissances. Cependant, il faut contraster

29. Voir par exemple [AFP \(2012\)](#)

30. Ainsi Sergio Bertolucci, le directeur de la recherche au CERN, explique-t-il que « La nouvelle [...] a été pour le public l'occasion de voir en action ce qu'est la méthode scientifique : un résultat inattendu a été soumis à l'examen des scientifiques, a été étudié en détail, et la solution a été trouvée en partie grâce à la collaboration entre des expériences qui sont normalement concurrentes. C'est ainsi que la science avance ! », [CERN \(2011\)](#).

31. [Larousserie \(2012b\)](#)

32. [AFP \(2012\)](#)

33. [Larousserie \(2012a\)](#)

cette conception avec une conception tournée vers d'autres types d'objectifs, en particulier dès lors que l'on s'intéresse à des domaines appliqués.

11.5 Perspectives de recherche et perspectives appliquées

Toute raisonnable qu'elle soit, l'approche adoptée par les physiciens dans les ajustements des constantes n'est pas pour autant universelle. Elle répond à des problématiques spécifiques, qui sont celles de la physique de précision, dans un domaine de recherche fondamentale très pointu. Cette approche s'accorde donc avec des objectifs précis. En revanche, appliquée à d'autres types de grandeurs physiques, chimiques, biologiques, médicales, et autres, la « philosophie » des ajustements peut se révéler très insatisfaisante. Lorsqu'un laboratoire teste la toxicité d'un produit destiné à l'alimentation, en regard de limites de tolérances fixées par la loi à des fins de sécurité sanitaire, il peut sembler dangereux de se contenter de reconnaître que la mesure effectuée est probablement inexacte et qu'elle présente peut-être une erreur de mesure significative, tout en rappelant qu'une telle erreur pourra un jour être corrigée, dans le futur, si l'on continue à tester les instruments et les méthodes de mesure utilisés. Les acteurs du monde industriel souhaitent que leurs produits fonctionnent, souhaitent pouvoir le vérifier et le faire savoir. Les considérations philosophiques sur l'inconnaissabilité ne les concernent que de très loin et la perspective d'une correction à long terme ne correspond certainement pas à leurs perspectives propres : considérer qu'il est impossible d'affirmer si un produit est conforme ou non, mais affirmer que ces connaissances pourront être améliorées dans le futur ne leur est d'aucun secours. De même, l'opposition entre le choix de statistiques bayésiennes et fréquentiste leur apparaît comme une querelle de spécialiste bien loin de leurs préoccupations.

De fait, plus on s'intéresse à des utilisateurs en bout de chaîne, plus l'exigence d'exactitude de mesure devient importante. Terry Quinn le fait d'ailleurs remarquer :

Des données qui sont reproductibles sans être exactes peuvent conduire à déformer la représentation de la nature d'une multitude de façons. Une chose est sûre ; il serait imprudent de construire des avions sur la base de données reproductibles mais qui ne sont pas exactes car l'on ne pourrait garantir leur sécurité. [...] Si leur exactitude n'est pas fermement établie, il n'y a pas moyen de savoir si des résultats apparemment reproductibles sont constants dans le temps.³⁴

Lors du colloque "GUM : past, present, future" organisé à Londres en novembre 2013 à l'occasion du vingtième anniversaire du GUM, c'est un employé du fabricant de moteurs d'avions « Rolls-Royce PLC », Pete Loftus, qui a fait remarquer le décalage que l'on observe entre les préoccupations de chercheurs et ceux des industriels³⁵. Le discours de Loftus tranchait assez nettement avec celui tenu par les intervenants précédents, qui étaient principalement des

34. [Quinn \(2002\)](#), p.13. Notons que Quinn se démarque de la position d'une partie de la communauté métrologique par son attachement à la notion d'exactitude, en témoigne sa remarque formulée à l'occasion du colloque "GUM : past, present, future" organisé au NPL, à Londres, en novembre 2013 à l'occasion du vingtième anniversaire du GUM : « je pense toujours qu'"exactitude" est un meilleur mot [qu'incertitude] », [Quinn \[2013\]](#).

35. [Loftus et Guidice \[2013\]](#)

statisticiens, physiciens ou chimistes. Que cette remarque émane d'un industriel est assez frappante. Dans son intervention, Loftus s'intéressait aux enjeux liés à l'incertitude de mesure dans le cas de la production et de la commercialisation de produits industriels de grande valeur. Il faisait essentiellement valoir l'argument suivant : dans le domaine industriel, l'incertitude ne suffit pas. Une faible incertitude ne suffit pas à assurer une mesure de qualité, et, par suite, à garantir la qualité du produit. De ce fait, expliquait-il, l'incertitude ne permet pas de « capturer la confiance que nous avons dans nos données »³⁶. L'objectif de l'industriel, selon lui, n'est pas de qualifier un état de connaissance mais de contrôler les défauts de la chaîne de production et de s'assurer que celle-ci soit un succès – afin que les produits soient de qualité. Loftus faisait d'ailleurs remarquer que lorsque l'incertitude de mesure est évoquée dans l'industrie, cela est fait « de façon défensive » : l'incertitude de mesure est ce dont on parle « quand les choses ont mal tourné »³⁷. En réponse à ces problèmes, Loftus évoquait alors assez rapidement la direction que prennent les industriels : celle de la gestion du *risque*.

La gestion du risque est l'un des éléments de différenciation des approches des chercheurs et celles des industriels. On peut se demander quels sont les risques concrets que prennent les physiciens lorsqu'ils s'engagent sur les deux dernières décimales d'une constante physique. Cela dépend de l'usage qui sera fait des valeurs de ces constantes, et de la conséquence qu'aura la connaissance plus ou moins précise de ces constantes. Les remarques de Taylor font sentir que l'inexactitude des résultats de mesure des constantes fondamentales de la physique n'a pas une grande conséquence dans les domaines appliqués qui feraient usage plus ou moins directement des valeurs recommandées, où la précision nécessaire est bien inférieure à celle qu'offrent les mesures les plus précises. Ce faisant, il exprime l'idée que les physiciens ne prennent pas de gros risque en indiquant des résultats de mesure qui pourraient être erronés. Cependant, lorsqu'il s'agit par exemple de redéfinir le Système International d'unités sur la base des mesures les plus précises des constantes physiques, ce même risque n'est plus alors permis, et il se révèle nécessaire de prolonger les recherches et d'agir avec beaucoup plus de prudence – ce qui explique que la redéfinition du kilogramme, évoquée dès avant 2005³⁸, ait été reportée à plusieurs reprises avant d'être programmée pour 2018.

En conclusion, la logique engagée dans la position de Taylor est une logique de recherche orientée vers le progrès futur. Mais les industriels auront tendance à être dans une logique de risque orientée vers la qualité *présente* de leurs produits. On voit ainsi que la dialectique du progrès futur ne couvre pas l'ensemble des problématiques auxquelles font face les différents acteurs du monde scientifique et économique. À une dynamique de progrès tournée vers le futur, on opposera une dynamique de qualité et de contrôle tournée vers le présent.

36. Loftus et Guidice [2013]

37. Loftus et Guidice [2013]

38. Quinn et Burnett (2005)

Conclusions de la troisième partie

Chacun des chapitres met en évidence une avancée épistémique sur le plan de l'analyse de l'incertitude et de la mesure des grandeurs physiques. Le chapitre 8 se concentre sur les vertus épistémiques de la combinaison des résultats de mesure. Il montre comment Mayer a proposé une première extension des conditions dans lesquelles des résultats de mesure peuvent être combinés entre eux de façon à procéder à une réduction de l'erreur de mesure. Cette extension passe par un raisonnement, certes encore élémentaire chez Mayer, mais néanmoins essentiel, sur la confiance que l'on peut accorder à un résultat de mesure – et elle aboutit, sur un plus long terme, après maturation par les savants de la fin du XVII^e siècle et du début du XVIII^e siècle, parmi eux Legendre, Gauss et Laplace, à une véritable théorie probabiliste de l'erreur qui permet le calcul quantitatif de ce qui est aujourd'hui désigné par l'incertitude de mesure. C'est l'incertitude de mesure qui devient le pivot de la circulation des résultats de mesure et de leur utilisation dans des contextes différents, en dehors des conditions immédiates de leur production. Les ajustements des constantes de la physique constituent un accomplissement parmi d'autres de l'élargissement des conditions de reproductibilité initié au XVII^e siècle. L'incertitude de mesure est effectivement un concept pivot dans la méthodologie visant à combiner entre elles les différentes mesures des constantes physiques, en articulation avec un réseau d'équations, issu de la théorie, reliant ces constantes entre elles. L'étude de la mise en place du procédé d'ajustement des constantes par Birge, au chapitre 9, révèle cependant que la combinaison des résultats de mesure est également dépendante d'une opération préalable, cette de *comparaison* des résultats, qui décide de la légitimité d'effectuer la combinaison. Une fois encore, l'incertitude de mesure est le pivot de l'opération de comparaison. En fonction de ce qu'indique la comparaison, on pourra alors envisager de combiner les résultats entre eux, ou d'en sélectionner une sous-partie cohérente. La façon dont Birge procède à cette opération montre que l'incertitude de mesure ne peut pas véritablement être comprise comme une évaluation de la fiabilité d'un résultat de mesure, ou de la confiance entretenue à propos du résultat. Les successeurs de Birge ont bien compris cet aspect et l'ont même revendiqué. Dans le travail qu'ils fournissent jusqu'aux années 1970, qui constitue l'objet du chapitre 10, ils insistent sur l'importance qu'il y a à distinguer entre obtenir un résultat exact – ce qui ne constitue pas à leurs yeux l'objectif principal des ajustements – et obtenir une faible incertitude. Dans cette perspective, l'incertitude de mesure n'est pas une évaluation de la fiabilité des résultats mais représente la finesse avec laquelle les scientifiques peuvent venir sonder la validité des structures théoriques auxquelles sont articulées les expériences effectuées, et tester la présence d'erreurs de mesure qu'ils pourraient alors chercher à expliquer et à corriger. Ce faisant, les physiciens prennent

acte de la différence entre *incertitude de mesure* et *exactitude de mesure*, qui nous semble être la seule véritable conséquence tangible de l'argument d'« inconnaisabilité » qui a été plus récemment employé par les métrologues pour remettre en question les objectifs et les modes de représentation associés à la mesure physique. La relation entre incertitude et exactitude dépend en fait du contexte dans lequel le résultat est employé. On distingue une logique de progrès, tournée vers l'amélioration des connaissances futures, et plutôt adaptée à une perspective de recherche, d'une logique de risque, tournée vers la qualité et le contrôle dans le présent d'un ensemble de produits, et qui correspond à une perspective appliquée.

Quatrième partie

Épilogue : la genèse du GUM

Chapitre 12

Épilogue : la genèse du GUM

Dans les parties I à III, nous nous sommes focalisés avant tout sur les aspects conceptuels liés à l'analyse d'incertitude. Cela nous a amenés à de multiples reprises à croiser le chemin de différentes institutions, parmi lesquelles le BIPM, et, à l'intérieur du BIPM, le JCGM (Joint Committee for Guides in Metrology), responsable de la maintenance du GUM et du VIM. L'étude des ajustements des constantes de la physique nous a également fait découvrir le rôle du CODATA (Committee on Data for Science and Technology). Les très nombreuses institutions scientifiques, métrologiques, gouvernementales, et autres laissent deviner une architecture sociale extrêmement sophistiquée dont l'étude pourrait apporter un regard très différent de celui que nous avons porté jusqu'ici. L'on pourrait s'intéresser en particulier aux dynamiques de consensus, aux rapports de pouvoir, aux enjeux économiques de la métrologie et de la mesure physique en général, et analyser ainsi la façon dont les sphères sociale et scientifique s'interpénètrent.

Sans nous engager véritablement dans cette direction, nous proposons dans cette partie, en guise d'épilogue, de brosse un aperçu de la façon dont s'est déroulée la rédaction du GUM. Ce chapitre laisse donc de côté les aspects conceptuels qui ont développés en longueur dans les précédentes parties, pour proposer un exposé du rôle des individus et des institutions dans le travail de standardisation qui accompagne la rédaction du GUM. Le travail effectué s'appuie sur des sources écrites et orales, recoupant en particulier des comptes-rendus de réunion avec les témoignages de différents acteurs en lien avec cet épisode¹. Ce travail nous a permis de compléter les informations parcellaires que l'on peut assez facilement retrouver dans la littérature métrologique, mais qui ne recouvrent pas toutes les questions que l'on peut se poser à propos de la conception du GUM. Ces informations, en particulier, se contentent majoritairement d'un exposé très bref, centré sur les décisions des institutions, sur leurs déclarations d'intention et

1. Deux entretiens oraux ont été menés pour ce travail. Le premier s'est déroulé avec Barry Taylor, à Paris, en juin 2013 ; le second, avec Wolfgang Wöger, à Paris, en juin 2014. À cela s'ajoutent des discussions plus informelles avec Charles Ehrlich et Walter Bich à l'occasion d'un colloque consacré au vingt ans du GUM, organisé à Londres en novembre 2013 : "GUM : past, present, future", 7 et 8 novembre 2013, National Physical Laboratory, Teddington, Royaume-Uni. Je remercie Barry Taylor de sa disponibilité et de sa gentillesse, d'avoir accepté de mener cet entretien, et de m'avoir donné l'accès à de nombreuses sources écrites très précieuses quant à l'histoire du GUM. Je remercie également Wolfgang Wöger, Walter Bich et Charles Ehrlich de m'avoir accordé de leur temps pour répondre à mes questions.

sur les résultats effectifs, sans s'attarder sur le contenu de l'étape de développement elle-même. Nous sommes convaincus que ce travail gagne à être poursuivi dans de plus amples détails.

Nous suivons une trame chronologique, en exposant d'abord dans la section 12.1 l'origine des motivations qui, à la fin des années 1970, ont poussé le Bureau International des Poids et Mesures (BIPM) à piloter l'harmonisation des pratiques d'analyse d'incertitude, puis à suggérer la conception d'un guide – le futur GUM. Dans la section 12.2, nous évoquons les difficultés rencontrées au cours des années 1980, et comment la responsabilité du document a échoué à l'Organisation Internationale de Standardisation (ISO), pour finalement aboutir à une publication en 1993. La section 12.3 est consacrée à la période suivant la publication du GUM, lors de laquelle le BIPM a repris les commandes de cette activité et a mis en place le JCGM, un comité dédié à la maintenance et à la révision du GUM, toujours actif aujourd'hui. Cela nous permettra, dans la section 12.4 d'ouvrir brièvement quelques questionnements et perspectives situés à l'interface des champs épistémologiques et sociaux.

12.1 De la lettre d'Amblor à la « recommandation INC-1 »

Les métrologues s'accordent² sur le fait que l'impulsion initiale de la création du GUM remonte à une lettre³ adressée en 1977 par Ernest Ambler, métrologue et physicien spécialiste des radiations ionisantes, à Jean Terrien, directeur du Bureau International des Poids et Mesures. Ambler est alors directeur depuis 1976 – et jusqu'en 1989 – du National Bureau of Standards (NBS), l'institut national de métrologie des États-Unis (qui deviendra en 1988 le National Institute of Standards and Technology, NIST). Ce n'est pas là sa seule responsabilité : il est également membre du Comité International des Poids et Mesures (CIPM), dont la mission est d'organiser la structuration des unités de mesure selon le Système International d'unités ; enfin, Ambler préside par ailleurs le Consultative Committee for Measurement Standards of Ionizing Radiation (CCEMRI) au Bureau International des Poids et Mesures (BIPM), un comité chargé de mettre en place des comparaisons internationales d'étalons de référence (appelées « intercomparaisons » dans le jargon métrologique) dans le domaine des radiations ionisantes. L'implication d'Amblor dans ces différents comités internationaux dont le rôle est en grande partie axé sur la régulation et la coordination des recherches scientifiques, le place ainsi dans une position d'observation privilégiée. Or, Ambler devient de plus en plus préoccupé par la difficulté que rencontrent alors les métrologues à comparer leurs résultats de façon claire et univoque⁴. En effet, il constate que les différents laboratoires impliqués dans ce travail de regroupement des résultats de mesure n'utilisent pas les mêmes méthodes d'évaluation d'incertitude de mesure, et ne formulent pas leur résultats finaux de la même façon⁵. De ce fait, la comparaison des résultats demande une traduction permanente, au cas par cas, des informa-

2. Cox et Harris (2014), p.S143.

3. Ambler (1977). Je remercie une fois encore Barry Taylor de m'avoir transmis ce document.

4. Taylor [2013]

5. Taylor [2013]. Ce constat sera répété la même année dans un rapport du CIPM consacré à la discussion de la lettre d'Amblor et de ses conséquences : « Les usages diffèrent d'un laboratoire à l'autre, entraînant une fâcheuse incohérence. », Comité International des Poids et Mesures (1977), p.16.

tions provenant de chaque laboratoire en fonction des méthodes utilisées, ce qui rend en fin de compte difficile, voire impossible, la mise en place d'une méthode systématique. C'est à la suite d'une réunion du CCEMRI, les 18 et 19 juin 1977, qu'Ambler écrit à Terrien pour y rapporter les préoccupations que les différents membres du comité ont émises. La lettre soulève plusieurs points principaux. Elle émet le souhait de voir les résultats d'étalonnage exprimés avec une valeur unique, et souligne l'importance que revêt alors la question de la combinaison des erreurs aléatoires et systématiques. Elle relève que d'autres organisations expriment un constat similaire⁶ et suggère qu'il est préférable qu'un organisme régulateur comme le CIPM s'occupe de centraliser ses questions plutôt que de les laisser proliférer de façon anarchique dans chacune des institutions⁷. De plus, la lettre d'Ambler propose la mise en place d'un groupe dont la mission serait de fournir aux différents laboratoires des lignes directrices communes pour l'évaluation de l'incertitude de mesure. Ces lignes directrices devraient être pensées de façon à favoriser le consensus international, et « éviter autant que possible les discussions philosophiques sur la théorie statistique »⁸. Il n'est pas alors spécifiquement question de rédiger un guide; cependant, cette lettre pose les bases de ce qui prendra peu à peu la forme d'un programme dirigé vers la constitution d'un document international de référence, qui, plus tard, sera baptisé *Guide pour l'expression de l'incertitude de mesure*.

Le CIPM étudie la proposition d'Ambler lors de sa 66^e session, un mois plus tard. Le rapport de la réunion rappelle les enjeux et les difficultés d'une telle démarche, en particulier sur le plan institutionnel :

Groupe de travail sur les incertitudes. — Mr Ambler avait écrit au directeur du BIPM pour attirer son attention sur un problème complexe, celui des incertitudes et de la façon dont on devrait exprimer les résultats de mesure. Le BIPM est bien placé pour s'occuper de cette question qui n'est résolue dans aucun pays. [...] Un groupe de travail sera donc créé sous la présidence de Mr Giacomo; il sera constitué de représentants des laboratoires nationaux et des organismes intéressés (OIML, ISO, CEI, etc.). Mr Terrien souligne que le problème des erreurs est très vaste. La raison essentielle pour laquelle le CIPM doit s'y intéresser est qu'à l'heure actuelle beaucoup d'organismes sont en train d'étudier cette question. On risque de les voir émettre des recommandations contradictoires et de se trouver alors devant une situation inextricable. Il ne s'agit pas pour le CIPM de fixer des règles arbitraires, mais il serait bon que le CIPM donne des lignes directrices; il pourrait par exemple suggérer la méthode la plus convenable dans des cas particuliers. [...] Mr Perlstain indique que l'ISO et la CEI, par exemple, se sont intéressés depuis longtemps à ce problème. Il s'est avéré impossible de concilier les points de vue très divergents

6. « L'urgence de l'action du CIPM sur le problème vient également du fait que d'autres organisations internationales telles l'IEC et l'ISO étudient des problèmes similaires », Ambler (1977).

7. « Une direction de la question par le BIPM pourrait empêcher un méli-mélo de différentes approches du problème par diverses organisations internationales. Il semble aussi plus efficace, du point de vue du temps que cela prendrait aux scientifiques, que le problème soit étudié une fois pour toutes par le CIPM, plutôt que séparément par différents Comités Consultatifs et par des organisations nationales et internationales », Ambler (1977).

8. Ambler (1977)

des différents groupes de spécialistes concernés.⁹

La compte-rendu du CIPM rappelle également que l'une des origines majeures des problèmes constatés dans l'évaluation des incertitudes de mesure réside dans le traitement séparé des erreurs systématiques et des erreurs aléatoires :

Le problème est celui du traitement des incertitudes dans la pratique. Deux sections du CCEMRI ont signalé l'importance de ce problème. La théorie des erreurs aléatoires est bien connue, mais pour les erreurs systématiques on ne fait en réalité que des estimations; celles-ci dépendent de l'expérimentateur et il n'existe pas de méthode unique, généralement acceptée, pour combiner les erreurs systématiques aux erreurs aléatoires.¹⁰

Parmi toutes les questions qui ont été soulevées lors de la conception du GUM, celle de la dichotomie entre erreurs aléatoires et systématiques semble avoir été la plus préminente et la plus décisive. Nous avons déjà eu l'occasion à plusieurs reprises de décrire en détail la nature de la distinction entre ces deux types d'erreurs, et la difficulté que celle-ci pose pour le traitement probabiliste de l'incertitude de mesure. Nous avons également mentionné à diverses occasions que la métrologie, de par sa fonction, recouvre à la fois des problématiques de l'ordre de la recherche fondamentale et d'autres qui concernent des domaines beaucoup plus appliqués, jusqu'aux aspects industriels. De ce fait, la métrologie est sans cesse pensée en rapport avec ses *utilisateurs*, et la question de l'incertitude de mesure ne déroge pas à cette règle. Comment un résultat doit-il être formulé de façon à ce qu'il soit utilisable par les utilisateurs? Rappelons que nous avons déjà établi au chapitre 3.6 que la dichotomie entre deux types d'erreurs sous-tend deux positions opposées quant à la façon dont doit être exprimé un résultat de mesure ainsi que son incertitude. Au point de vue du « fournisseur » (ou de l'expérimentateur), selon lequel il est souhaitable de ne pas trop comprimer les informations qui viennent avec un résultat numérique – adopté par exemple dans les ajustements des constantes de la physique – s'ajoute un point de vue « utilisateur » (ou client) où l'objectif est de rendre l'utilisation du résultat de mesure le plus simple possible, par exemple dans une pratique routinière de tests statistiques standardisés ou même automatisés. Si les deux points de vue ne sont pas forcément inconciliables, chacun vient avec une façon différente d'exprimer l'incertitude d'une mesure. Le point de vue utilisateur réclame que l'incertitude de mesure prenne la forme d'une valeur unique¹¹, comme en témoigne le compte-rendu de la réunion du CIPM :

Pour les laboratoires d'étalonnage ou de vérification, [...] les clients de ces laboratoires veulent un nombre unique traduisant l'exactitude des mesures ou des instruments soumis à l'étalonnage et permettant par exemple de juger de leur conformité à un cahier des charges.¹²

L'obtention d'une incertitude unique nécessite de fonder un tel calcul d'incertitude sur des bases statistiques solides. Or, c'est précisément cet aspect qui se révèle le plus difficile à obtenir,

9. [Comité International des Poids et Mesures \(1977\)](#), p.16. Le préfixe « Mr » est d'origine dans l'ensemble des textes du CIPM cités dans ce chapitre, de même que les noms soulignés.

10. [Comité International des Poids et Mesures \(1977\)](#), pp.15–16.

11. Comme l'explique Quinn, pour un laboratoire fondamental, un résultat de mesure est un « constat complexe », [Quinn \[2013\]](#). Pour l'utilisateur final, ce doit préférentiellement être une donnée simple à exploiter.

12. [Comité International des Poids et Mesures \(1977\)](#), p.16.

car il est le plus inhomogène dans les pratiques des scientifiques. Un an plus, tard, en 1978, le CIPM émettra d'ailleurs la remarque suivante :

Si tout le monde est d'accord sur la nécessité de fournir le plus possible de détails sur les incertitudes pour les usages scientifiques, les difficultés surgissent dès qu'on veut soumettre ces incertitudes à un traitement mathématique. Les opinions deviennent très divergentes lorsqu'on cherche à caractériser l'incertitude globale par un nombre unique. [...] C'est un aspect que l'on ne peut négliger, même si l'on reconnaît que pour des buts purement scientifiques il convient de donner un maximum d'informations sur les différentes composantes de l'incertitude.¹³

L'on retrouve ainsi que la volonté d'exprimer une incertitude en une composante unique, et à travers cela la question de la combinaison des différentes composantes d'incertitudes, est à la racine des problématiques de l'époque.

Dès la première réunion consacrée à ce sujet, le CIPM semble accéder à la requête d'Ambler, bien que la démarche complète reste encore à définir. Il est prévu de préparer le terrain en menant une enquête préliminaire au sein des laboratoires :

Mr *Giacomo* est conscient du fait qu'il sera difficile de réunir un groupe efficace, donc restreint, de personnes compétentes. Il a l'intention de prendre contact avec les laboratoires en leur adressant un questionnaire et d'éviter d'attaquer de front l'ensemble du problème car il est trop vaste; pour commencer, il faut choisir un ou deux aspects sur lesquels un progrès sensible peut être réalisé.¹⁴

Un questionnaire est effectivement envoyé en février 1978¹⁵ à trente-deux laboratoires nationaux de métrologie « connus pour avoir un intérêt particulier sur le sujet », ¹⁶ ainsi qu'à cinq organisations internationales¹⁷. Le CIPM se conforme à la suggestion d'Ambler et son objectif désormais explicite est mettre en place, à terme, un groupe de travail dont la mission serait d'aboutir à une recommandation internationale¹⁸. Les réponses, données par vingt-et-un laboratoires de pays différents (dont l'Union Soviétique et certains pays du bloc de l'Est), sont collectées jusqu'à la fin de l'année 1978 et synthétisées dans un rapport du BIPM en mars 1980¹⁹. Au fur et à mesure de la réception des réponses apportées au questionnaire, il apparaît qu'aucun consensus marqué ne pourra être dégagé de façon simple²⁰. En octobre 1980, le BIPM réunit comme prévu un groupe de travail sur l'expression des incertitudes dont la base de réflexion s'appuie sur les réponses fournies au questionnaire. À partir de l'examen des réponses, le groupe de travail soumet au CIPM une recommandation internationale baptisée

13. Comité International des Poids et Mesures (1978), p.10.

14. Comité International des Poids et Mesures (1977), p.16.

15. Comité International des Poids et Mesures (1978), p.10.

16. Kaarls (1980), p.3.

17. Le questionnaire est retranscrit dans Bureau International des Poids et Mesures (1980). Ambler fut lui-même l'un des répondants au questionnaire, au nom du NBS, « non pas en tant que statisticien mais en tant que physicien », Taylor (2013).

18. Bureau International des Poids et Mesures (1980), appendix I "Questionnaire on uncertainties", p.1-2.

19. Bureau International des Poids et Mesures (1980), pp.2-3.

20. Quinn [2013].

« INC-1 »²¹, et adoptée ensuite par le CIPM qui, dans la recommandation « CI-1981 »²², répète l'importance de la problématique soulevée, reconnaît les progrès que le questionnaire a permis d'opérer, et préconise l'expérimentation de méthodes standardisées fondées sur la recommandation INC-1 au sein des laboratoires de métrologie, afin que « dans un délai de deux ou trois ans le BIPM fasse le point sur la mise en œuvre de ces propositions »²³. À ce moment, l'objectif est uniquement de formuler un consensus satisfaisant, afin de regrouper les différentes méthodes d'évaluation et d'expression de l'incertitude de mesure ; ce n'est que plus tard que cette recommandation deviendra l'un des soubassements essentiels à la première édition du GUM²⁴.

La recommandation INC-1 tranche notamment certains débats, et tout particulièrement celui concernant les types d'erreurs : « la distinction habituelle entre incertitudes aléatoires et systématiques n'est pas essentielle »²⁵. La recommandation propose de regrouper les deux types d'incertitude sous une même bannière statistique, tout en distinguant deux façons de les calculer : statistiquement (méthodes de « type A ») et par des moyens autres que statistiques (méthodes de « type B »)²⁶. Il est à noter que la recommandation INC-1 est très courte – elle tient sur une page – et de ce fait laisse ouverte la voie à une multitude d'interprétations. En particulier, elle ne dit pas dans le détail comment sont menées les méthodes de type A et de type B.

12.2 L'ISO prend la main

Si les métrologues se sont assez rapidement mis d'accord sur la recommandation INC-1, les années qui suivent ne sont pas vraiment fructueuses et la situation se tasse²⁷ ; par conséquent, en 1984, il est proposé que soit rédigé un document plus détaillé :

Mr Quinn souligne la nécessité de donner une nouvelle impulsion au Groupe de travail sur l'expression des incertitudes.

Mr Blevin rappelle qu'il est très important pour le CIPM de savoir où en sont les travaux sur les incertitudes. Le CIPM devrait soit confirmer les recommandations provisoires faites précédemment, soit leur apporter les modifications qu'il jugerait nécessaires. On ne peut guère laisser le problème en attente pendant encore deux

21. [Kaarls \(1980\)](#), version anglaise p.13 et version française p.14. Voir aussi [Giacomo \(1981\)](#), pp.73–74, ainsi que [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.30.

22. [Comité International des Poids et Mesures \(1981\)](#), p.26. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.31. Voir aussi [Giacomo \(1982\)](#), pp.43–44.

23. [Comité International des Poids et Mesures \(1981\)](#), p.26.

24. En 1986, Ambler relèvera que l'application de la recommandation INC-1 est lacunaire dans le cadre des inter-comparaisons : « (D)ans la récente comparaison interlaboratoire du CCPR, certains laboratoires semblaient encore ignorer la recommandation du Groupe de travail sur l'expression des incertitudes qui demande d'utiliser l'écart-type. », [Comité International des Poids et Mesures \(1986\)](#), p.14. Le CIPM adoptera par conséquent une nouvelle recommandation, la recommandation CI-1986 ([Comité International des Poids et Mesures, 1986](#), p.14 ; [Joint Committee for Guides in Metrology \(JCGM\), 2008d](#), p.30) afin de systématiser l'utilisation des principes scientifiques développés dans la recommandation INC-1, pour les travaux effectués dans le cadre du CIPM.

25. [Comité International des Poids et Mesures \(1981\)](#), p.8.

26. Nous renvoyons à la partie I pour une description plus détaillée de la mise en application de ces méthodes.

27. [Quinn \[2013\]](#)

ans. Pour Mr Ambler, le plus important est d'étudier la mise en application de la recommandation.

Mr Quinn suggère de rédiger un document plus détaillé que l'on pourrait soumettre à l'approbation de ce même Groupe de travail.²⁸

Dans l'ensemble, la recommandation INC-1 ne fait pas l'objet d'objection notables, mais son manque de détails est toutefois critiqué. En particulier, la communauté estime qu'il serait bon d'enrichir la recommandation d'exemples de précisions méthodologiques qui permettraient l'application concrète de la recommandation aux cas que rencontrent quotidiennement les métrologues²⁹. Le groupe de travail du CIPM conclut qu'il devient nécessaire d'impliquer d'autres organisations, en particulier l'ISO (Organisation Internationale de Normalisation), dans la rédaction du document³⁰. Cette suggestion est répétée en 1985³¹, puis le CIPM finit par estimer que la problématique dépasse le cadre d'action strict du BIPM, et propose que l'ISO devienne responsable de la production du document, ce qui alors entériné³².

Du 1 au 3 octobre 1986, dans les locaux du Bureau International de Métrologie Légale (le secrétariat permanent de l'OIML, Organisation Internationale de Métrologie Légale, à Paris), l'ISO réunit pour la première fois l'ISO-TAG4-WG3 (ISO Technical Advisory Group 4, Working Group 3), un groupe de travail sur les incertitudes, présidé lors de cette séance par Ronald Collé du NBS. Le groupe est constitué de dix membres provenant de diverses organisations : BIPM, ISO, OIML, AFNOR (Association française de normalisation), IEC (International Electrotechnical Commission), et de laboratoires nationaux de métrologie : NPL (National Physical Laboratory, Grande-Bretagne), PTB (Physikalisch-Technische Bundesanstalt, Allemagne), VSL (Van Swinden Laboratorium, Pays-Bas), et GOST (URSS). Robert Kaarls, du VSL, est le seul membre du groupe à avoir également participé au groupe de travail du CIPM lors de la rédaction de la recommandation INC-1 en 1980. Cette fois-ci, l'objectif explicite du groupe de travail est de composer un document qui détaille des règles communes à propos de l'expression des incertitudes de mesure, tout en continuant de s'appuyer sur les préceptes de la recommandation INC-1 :

(L)a tâche du groupe de travail est d'élaborer un document sur la base de la recommandation du Groupe de travail du BIPM sur l'incertitude, laquelle fournit des conseils sur l'expression de l'incertitude de mesure, pour une utilisation dans les services de normalisation, d'étalonnage, d'accréditation de laboratoire et des services de métrologie. L'objectif de ces directives est de promouvoir une informa-

28. Comité International des Poids et Mesures (1984), pp.24–25.

29. Comité International des Poids et Mesures (1984), p.25.

30. Comité International des Poids et Mesures (1984), p.25.

31. Comité International des Poids et Mesures (1985), p.37.

32. La façon dont l'ISO a en fin de compte été mandatée pour ce travail n'est pas très claire. Le rapport de la réunion de l'ISO-TAG4-WG3 d'octobre 1986 indique seulement : « ces dernières années, le CIPM a renvoyé cette question à l'Organisation Internationale de Normalisation (ISO) car il a estimé qu'il s'agit là d'un corps international plus logique pour essayer de parvenir à un accord et à une uniformité sur l'énoncé des incertitudes dans les organisations internationales de normalisation et de métrologie », Karp (1986), p.2. Barry Taylor avance l'hypothèse suivante : à ce moment, le secrétaire général de l'ISO, Lawrence D. Eicher, provenait du NBS, et il est possible qu'Amblér lui ait proposé oralement la constitution de ce groupe, Taylor [2013].

tion complète sur la façon dont on aboutit à un énoncé d'incertitude et de fournir une base pour la comparaison internationale des résultats de mesure.

Le GT a conclu que sa tâche est de produire un document qui sera solidement basé sur la recommandation INC-1 du BIPM (1980), mais qui sera bien plus spécifique et utilisable que la très générale Recommandation.³³

Le groupe de travail clôt son rapport sur un projet de rédaction de différents chapitres, assignés chacun à un ou plusieurs membres du groupe³⁴. Cependant, dans les années qui suivent, les progrès sont peu marqués et la situation atteint un point mort³⁵. En 1989, pour donner une nouvelle impulsion au groupe de travail, Barry N. Taylor, physicien et métrologue au NBS, devenu depuis peu le NIST, suggère de faire appel à son collègue E. Richard Cohen, avec qui il a collaboré depuis une vingtaine d'années dans le pilotage des ajustements des constantes de la physique (comme nous l'avons vu plus en détail dans les sections précédentes de ce chapitre), pour diriger le groupe de travail de l'ISO. La première réunion présidée par Cohen a lieu au cours de cette même année³⁶. Mais ce changement de direction n'apporte pas de progrès notable, d'autant que Cohen semble faire preuve d'un manque d'écoute et d'ouverture envers les propositions du reste du groupe de travail³⁷.

C'est donc un peu plus tard, constatant le manque d'avancement du groupe de travail, que Taylor s'essaie à son tour à la rédaction d'un brouillon de document. Les circonstances du moment font qu'il se retrouve à collaborer avec son collègue Chris Kuyatt, alors directeur de la section des radiations ionisantes au NIST³⁸. Début 1991, ils soumettent un document au TAG 4 de l'ISO, la sous-structure de l'ISO dont dépend le groupe de travail sur les incertitudes. Ce document est mis en concurrence avec un autre texte proposé au TAG 4 au même moment. La question est alors soulevée quant au document à choisir comme point de départ : celui de Taylor et Kuyatt est adopté et c'est ainsi qu'ils deviennent les rédacteurs principaux du brouillon du futur guide international sur les incertitudes³⁹. Est alors assignée à Taylor et Kuyatt la tâche d'intégrer les commentaires, remarques, et éléments supplémentaires qui leur sont régulièrement envoyés en réaction au premier brouillon qu'ils ont fourni, afin d'aboutir à un document acceptable pour tous : celui-ci devient une collaboration dirigée. Le document est finalement publié dans le courant de l'année 1993 en anglais et en français, avec comme titre *Guide to the expression of uncertainty in measurement*. Il est écrit au nom de sept organisations internationales : BIPM, IEC, IFCC (International Federation of Clinical Chemistry), ISO, IUPAC (International Union of Pure and Applied Chemistry), IUPAP (International Union of Pure and Applied Physics), et OIML. Il est ensuite très légèrement remanié en 1995.

Entretemps, Taylor et Kuyatt rédigent pour le NIST une note technique, d'abord publiée en janvier 1993 puis rééditée en septembre 1994⁴⁰, qui synthétise les principes du GUM en

33. Karp (1986), p.3.

34. Karp (1986), pp.11–12.

35. Taylor [2013]

36. Taylor [2013]

37. Taylor [2013]

38. Taylor [2013]

39. Taylor [2013]

40. Taylor et Kuyatt (1994)

les formulant selon les règles de pratique ayant cours au NIST⁴¹. Cette publication connaît un grand succès aux États-Unis où, à en croire Barry Taylor, plusieurs dizaines de milliers de copies sont distribuées⁴². La note technique du NIST se distingue du GUM par son caractère plus local, attaché aux pratiques d'un institut national en interaction avec des acteurs d'un espace bien défini, l'espace nord-américain. Le GUM, par son ambition plus large d'harmonisation des pratiques à l'échelle internationale, se heurte à la fois à la nécessité de formuler un consensus général tout en satisfaisant l'ensemble des acteurs visés dont les pratiques sont multiples. L'on peut alors comprendre que, bien que publiés au même moment, ce soit la note technique du NIST qui connaisse la première un grand succès aux États-Unis. Il est à noter de plus que le GUM lui-même, en tant que publication de l'ISO, était alors de ce fait un document payant. Sa réédition en 2008 par le JCGM (détaillée dans la section suivante) en fera un document librement accessible.

12.3 La création du JCGM

L'ISO publie le GUM une première fois en 1993, puis celui-ci est très légèrement remanié en 1995. Reste alors ouverte la question de sa maintenance. Quel doit être l'avenir du document ? L'un des enjeux porte sur la responsabilité de l'ISO, qui a jusqu'alors été responsable du contenu du GUM et de sa distribution. L'ISO doit-il se charger de continuer à travailler sur le document, ou doit-il laisser une autre organisation prendre le relais ? Cette question s'articule avec le statut du GUM au sein de l'ISO : il ne s'agit que d'un guide, et non pas d'une norme. Or, des désaccords émergent quant au rapport de l'ISO avec les guides⁴³. C'est l'une des raisons pour laquelle il semble que l'ISO n'apparaisse plus comme l'organisation adéquate pour effectuer la maintenance du GUM.

En 1997, Terry Quinn, en concertation avec Taylor, donne l'impulsion pour la fondation d'un comité, le JCGM (Joint Committee for Guides in Metrology)⁴⁴, qui reprenne en charge le traitement de la question des incertitudes de mesure. La première réunion officielle a lieu aux alentours de l'année 2000, et il y est déjà envisagé d'engager une révision substantielle du document, en vue d'une seconde édition. Cependant, Taylor, qui dirige alors le groupe de travail⁴⁵, suggère qu'il n'est pas opportun de modifier le GUM dans l'immédiat, alors qu'il commence à peine à être diffusé, et qu'il est déjà le fruit d'un consensus difficile à atteindre. C'est à l'occasion des débuts du JCGM que Walter Bich, métrologue à l'INRIM (Istituto Nazionale Di Ricerca Metrologica, Turin), commence à travailler sur le GUM. Il avait été convié par Terry Quinn, dès 1999, à participer aux réunions du JCGM⁴⁶. En réactions aux réticences de Taylor à proposer un nouveau document, Bich suggère que soient rédigés des textes complémentaires⁴⁷ ; les premiers de ces « suppléments » verront le jour en 2008. À l'issue de cette même réunion, Barry

41. Taylor et Kuyatt (1994), p.1.

42. Taylor [2013].

43. Taylor [2013]; Wöger [2014]

44. Bich (2013). Bich fait remarquer que les deux personnes qui avaient alors le plus d'influence dans ce domaine étaient Terry Quinn et Barry Taylor.

45. Wöger (2014).

46. Bich (2013).

47. Bich (2013).

Taylor émet le souhait de ne plus participer au JCGM : il désire avant tout se consacrer à la réforme future du Système International d'Unités.⁴⁸ C'est à cette occasion que Bich devient directeur du groupe de travail chargé de la maintenance du GUM⁴⁹.

Après sa première réunion, le JCGM n'a pas encore d'existence formelle claire : il n'est alors avant tout qu'un groupe de travail⁵⁰. L'ISO souhaite de plus réguler le nombre de membres des groupes de travail et leur appartenance : un maximum de trois membres de chaque organisation représentée doit être respecté, et aucun expert en dehors de ce cadre ne peut y assister. Il apparaît qu'il est nécessaire de trouver un cadre pour opérer ces régulations. Le rôle de l'ISO dans cette opération étant de plus en plus questionné, il est fait appel au BIPM pour héberger le JCGM et lui donner une structure formelle⁵¹.

La création du JCGM est l'occasion d'une jonction des efforts de maintenance du GUM et du VIM. La première publication du GUM avait en effet coïncidé avec celle de la deuxième édition du *Vocabulaire International de Métrologie* (VIM2), les deux documents étant produits et distribués par l'ISO, et publiés la même année, en 1993. Cela avait marqué le rapprochement des trajectoires des deux documents. La première édition du *Vocabulaire International de Métrologie* (« VIM1 »), en 1984, a probablement été constituée de façon assez indépendante et sans aucun lien avec le futur GUM. Cependant, il serait juste de dire, à en croire Taylor, que l'ISO-TAG4-WG3 a aussi été le lieu de discussion de la seconde édition du VIM, publiée en 1993⁵², parallèlement au GUM qui s'appuie (encore aujourd'hui) sur sa terminologie. Les deux documents deviennent finalement la responsabilité conjointe du JCGM à partir de la fin des années 1990.

Le rôle du JCGM au sein du BIPM est scindé en deux groupes de travail : un premier groupe de travail, le « JCGM WG1 » est dédié aux réflexions portant sur le GUM ; un second groupe de travail, le « JCGM WG2 » est dédié aux réflexions portant sur le VIM. En 2008, les groupes de travail respectifs du JCGM aboutissent d'une part à la publication d'un premier supplément au GUM⁵³, et d'autre part d'une troisième édition du VIM⁵⁴. À cette occasion, le GUM est réédité sous une forme presque identique à sa version de 1995, et est distribué gratuitement par le BIPM. D'autres suppléments sont publiés lors des années suivantes (2009, 2011, 2012) et d'autres sont prévus dans un futur proche, en attendant une seconde version du GUM entièrement remaniée.

12.4 Assise institutionnelle de l'incertitude de mesure

Nous souhaitons pour conclure évoquer, sans les développer, quelques pistes embryonnaires de réflexion que nous inspire la narration développée dans ce chapitre. Ces balises vi-

48. Taylor [2013]

49. Bich (2013).

50. Bich (2013).

51. Bich (2013).

52. Organisation internationale de normalisation (ISO) (1993)

53. Joint Committee for Guides in Metrology (JCGM) (2008b)

54. Joint Committee for Guides in Metrology (JCGM) (2008c), qui sera publié de nouveau en 2012 pour corriger quelques erreurs, Joint Committee for Guides in Metrology (JCGM) (2012).

seront surtout à repérer la façon dont peuvent être articulées les problématiques épistémologiques qui constituent la grande majorité des parties I à III, avec les problématiques institutionnelles et sociales qui apparaissent nettement dans l'histoire du GUM, ainsi d'ailleurs que dans celle des ajustements des constantes.

En un sens, il est possible de situer la rédaction du GUM dans la continuité d'un travail global de standardisation et de centralisation qui traverse la métrologie depuis plus d'un siècle. Quinn fait ainsi le lien entre la lettre d'Ambler et les motivations qui ont amené à la fondation du BIPM et du CIPM⁵⁵. La proximité entre le GUM et les travaux d'ajustements est également assez nette. Elle apparaît bien sûr lorsqu'on s'intéresse à l'identité de certains des protagonistes : Richard Cohen, Barry Taylor, et Terry Quinn sont tous les trois passés par le poste de directeur du groupe de travail du CODATA consacré aux constantes fondamentales⁵⁶ ; Ernest Ambler s'est également impliqué dans ces travaux. Le proximité ne se limite pas à cela ; les problèmes soulevés par Birge et par Ambler résultent essentiellement d'une même difficulté, à savoir la façon dont on traite de façon commune les erreurs aléatoires et systématiques de mesure pour en rendre compte par une valeur numérique présentant la même signification pour tous. Birge avait souligné la grande inhomogénéité des pratiques de son époque ; Ambler mentionne la difficulté que cette diversité entraîne pour la pratique des intercomparaisons. Nous remarquons par ailleurs que les préoccupations qui ont mené à la rédaction du GUM ont été émises depuis des domaines assez fondamentaux de la recherche, en physique et en chimie, pour remonter vers des domaines plus appliqués au fur et à mesure que le document a pris corps. Au départ, l'implication des physiciens au sein de l'ISO-TAG4-WG3 était très forte. Elle a diminué depuis, et semble moins importante au sein du JCGM, dont le spectre s'est élargi jusqu'à des domaines appliqués à forts enjeux économiques, comme l'industrie. Au bénéfice scientifique qui peut être tiré d'un tel travail de standardisation se superpose des questionnements d'ordre économique, qui se feront de plus en plus explicites au fur et à mesure que le GUM quitte le domaine de la physique fondamentale pour investir celui de l'ingénierie, de l'industrie et du commerce⁵⁷.

S'il y a travail de standardisation, il convient alors de soulever trois questions : qui est impliqué dans ce travail, comment celui-ci est mené à bien, et que produit-il effectivement ? Les deux premières questions sont fortement intriquées. La narration historique que nous proposons montre que cette histoire fait intervenir trois pôles, celui des individus, celui des institutions, et celui des concepts ; de fait, il est important de considérer séparément l'aspect individuel et l'aspect institutionnel.

D'un côté, certains acteurs incarnent avant tout l'institution qu'ils représentent – la lettre d'Ambler à Terrien est avant tout une lettre du CCERMI au BIPM. Ces institutions émettent des avis et prennent des décisions dont il est difficile de rendre compte dans le détail car celles-ci

55. Quinn (2013). Quinn revient en détail sur l'histoire de ces institutions dans Quinn (2012).

56. Lide et Wood (2012), p.14.

57. Nous pouvons citer à titre d'exemple la façon dont la note technique 1297 du NIST reconnaît explicitement cet aspect économique dans son avant-propos : « (l)le mouvement vers l'adoption internationale de l'approche du CIPM pour exprimer l'incertitude est entraînée dans une large mesure par l'économie globale et le marché ; son utilisation dans le monde entier permettra que des mesures effectuées dans différents pays et dans des secteurs aussi divers que la science, l'ingénierie, le commerce, l'industrie, et la régulation soient plus facilement comprises, interprétées et comparées », Taylor et Kuyatt (1994), p.iv.

sont la synthèse de discussions qu'il n'est pas forcément possible de retracer, d'autant que le fonctionnement des institutions les mènent précisément à effacer le rôle des individus pour promouvoir un travail commun. Une partie du travail de métrologie vise à produire des concepts portés par des institutions (crédibilité, autorité) et non par des personnes, d'où un effacement des individus, non pas par contingence mais plutôt en raison de la nature de l'activité métrologique elle-même. Par ailleurs, à l'échelle des institutions interviennent des tendances générales guidées par des traditions de recherche et de pratiques. Les grands laboratoires de métrologies, tels le Bureau National de Métrologie (France), le National Physical Laboratory (Royaume-Uni), le Physikalisch-Technische Bundesanstalt (Allemagne) ou le National Institute of Standard and Technology (États-Unis) ont chacun développé nombre de techniques, et font valoir un point de vue qu'ils ne sont pas nécessairement prêts à abandonner facilement.

Mais d'un autre côté, on constate qu'il est difficile, dans cette histoire tout du moins, de s'en tenir aux institutions et de négliger l'importance des individus qui les composent. Les négociations menées, les idées partagées ne sont pas simplement l'émanation de structures très générales et abstraites mais sont d'abord le produit de personnalités qui ne s'effacent pas toujours derrière les structures. Nous constatons ainsi l'importance des initiatives personnelles. C'est le cas pour Birge en ce qui concerne les ajustements des constantes de la physique; la structure institutionnelle, le CODATA, ne prend les choses en main qu'une fois atteinte une certaine maturité. Quant au GUM, il apparaît que le travail mené au sein du BIPM et de l'ISO s'est progressivement ralenti, voire bouché, avant que le déblocage ne provienne en partie de travaux encore une fois individuels, que ce soit le document proposé par Taylor et Kuyatt pour une première version du GUM, ou l'impulsion donnée par Taylor et Quinn pour fonder le JCGM. Dans ces différents exemples, les institutions que sont le CODATA, l'ISO et le BIPM capitalisent alors sur ces déblocages pour ensuite relancer un travail pleinement collaboratif. La dynamique individuelle fait apparaître des contingences dans ce qui pourrait autrement sembler un cheminement linéaire d'un projet bien préparé vers une réalisation de ce projet en tout point conforme avec les intentions initiales.

Par conséquent, à la question du « comment », celle du « qui » répond en partie. À en croire certaines publications de l'époque, la recherche de consensus a été biaisée par une éviction des avis contradictoires. C'est ce qu'exprime Rolf Schumacher dans un article extrêmement incisif, presque un pamphlet, commentant le déroulement des réunions du groupe de travail du BIPM sur les incertitudes. Schumacher décrit ce qu'il perçoit comme une absence de dialogue et un fort dogmatisme de la part des défenseurs de l'approche préconisée dans la recommandation INC-1⁵⁸; dont il met d'ailleurs en question la crédibilité⁵⁹; et il insiste sur l'absence d'un réel consensus, contrairement à ce qui est affirmé dans les comptes-rendus officiels de

58. « J'ai assisté à la réunion en croyant que le "Groupe de conseil" serait invité à offrir des avis et des recommandations, peut-être même de développer un consensus [...] Cette attente était partagée par d'autres participants et fut exprimée durant la réunion par Carroll Croarkin. Mais nous devions vite apprendre que cela n'était pas le cas [...] Le président de la réunion [...] a indiqué sa résistance envers toute tentative de changement [des recommandations] », [Schumacher \(1987\)](#), p.56.

59. « Dr. Eisenhart du NBS [...] dit de l'un des partisans éminents des recommandations du BIPM (avec qui il est en désaccord) : "je doute qu'il ait jamais vu un laboratoire de normalisation de l'intérieur" », [Schumacher \(1987\)](#), p.56.

réunion⁶⁰. Schumacher répète ses critiques dans un second article publié un an plus tard dont le titre est très clair : “The european unity behind the BIPM recommendations may be not so unified”⁶¹. Les deux articles de Schumacher dessinent un débat polarisé autour de deux camps entre lesquels il n’y a que très peu de dialogue constructif. Mais surtout, bien que Schumacher souligne l’existence de dissensions d’ordre nationales ou supranationales (l’opposition entre les traditions nord-américaines et européennes est rappelée à plusieurs occasions), il décrit un rapport de force où ce sont les individus, et non les institutions, qui dessinent l’orientation générale. Il semble que l’élément décisif réside alors dans le fait qu’en plus de ne pas partager les mêmes idées, les deux camps, surtout, ne partagent pas le même pouvoir. Schumacher décrit le consensus obtenu comme “un vote majoritaire des membres d’un petit club”⁶². L’impression rétrospective de Taylor sur le sujet n’est d’ailleurs pas incompatible avec la description de Schumacher. Taylor rapporte ainsi qu’Ambler, dans les années 1980, avait conscience du fait que certains membres du NBS, tels Eisenhart ou la statisticienne Carroll Croarkin, étaient assez conservateurs, et en particulier étaient réticents, voire opposés, à l’idée de combiner les incertitudes. Il fit en sorte de limiter l’influence de ces derniers dans les groupes de travail du CIPM et de l’ISO sur l’incertitude de mesure⁶³. Taylor fait d’ailleurs remarquer que les désaccords n’ont parfois pas tant été résolus par des débats rationnels que par le renouvellement des générations : certains des plus conservateurs ont emporté avec eux leurs objections⁶⁴.

La question du « qui » et du « comment » est bien entendu une question complexe qui mériterait un travail historique bien plus poussé que celui que nous avons pu mener dans le cadre du présent travail. Nous retenons principalement que notre analyse montre, du moins, qu’il est nécessaire de raisonner sur une double échelle, celle des institutions et celle des individus, la seconde ayant tendance à s’effacer dans les textes au profit de la première.

Reste la question de ce qui est produit, la question du « quoi ». Bien entendu, la réponse est en apparence évidente : le produit du travail est la publication du GUM, du VIM (la troisième édition en particulier) et des différents suppléments du GUM. L’on peut chercher à aller un peu plus loin. Alexandre Mallard a décrit différents attributs que possède l’écriture des normes⁶⁵ qu’il décrit en particulier comme la mise en convergence de réseaux sociotechniques. C’est bien ce que l’on constate dans le cas du GUM, qui a mis en relation physiciens, statisticiens, industriels, instituts de normalisation, etc. Le GUM n’est certes pas initialement une norme – il est un guide, même s’il a par la suite diffusé dans certaines normes⁶⁶ – mais il s’en approche en partie, d’autant qu’il a été conçu comme un outil de standardisation. Mallard souligne qu’une

60. « Le procès-verbal de la réunion contient l’affirmation : “la majorité des participants de la réunion se sont mis d’accord” – c’est la position prétendue qu’adopte le président de la réunion. On ne sait pas ce à quoi l’“accord” de cette phrase se réfère, mais il semble n’y avoir eu aucun signe visible d’une quelconque majorité d’opinion lors de la réunion ou lors des votes et sondages effectués », Schumacher (1987), p.58.

61. Schumacher (1988)

62. Schumacher (1987), p.58.

63. Taylor (2013)

64. Taylor (2013)

65. Mallard (1998), Mallard (2000)

66. En France, le GUM est un temps devenu la norme NF ENV 13005 d’août 1999 (Priel, 2008) ; cette norme a été annulée en août 2014.

norme présente un intérêt à la fois pour ce qu'elle contient – le texte⁶⁷ – et ce qu'elle ne contient pas – le hors-texte⁶⁸, ce sur quoi les acteurs ne sont pas d'accord. Les dynamiques de consensus arbitrent simultanément ces deux points. Au-delà du document produit – le GUM et sa famille de compléments – il serait donc fécond de s'interroger sur l'ensemble des éléments sur lequel le GUM n'a pas engagé son autorité.

Mallard a également rappelé que le travail de normalisation ne consiste pas simplement à obtenir un consensus quant à des connaissances et méthodes *déjà* existantes mais également à produire de nouvelles connaissances⁶⁹. De fait, le travail de mise en forme et de recherche de consensus qui a accompagné la rédaction du GUM a fait émerger une façon inédite de présenter l'analyse d'incertitude, centrée autour des méthodes de type A et B, sur la notion d'incertitude-type, puis celle d'incertitude « élargie », et autour de la façon de propager ces différentes incertitudes. Ce contenu n'a rien de révolutionnaire : il se nourrit de façon essentielle de pratiques déjà éprouvées, mais il reformule, restructure et synthétise ces acquis d'une façon qui constitue en soi une véritable avancée conceptuelle. Cependant, dans le même temps, ces avancées conceptuelles n'ont pas été poussées assez loin. En effet, dans la partie I, nous montrons que le GUM intègre de façon implicite un certain nombre de conceptualisations qui sont peu, voire pas du tout explicitées, en particulier concernant l'interprétation des probabilités. De ce fait, dans une histoire plus large, le GUM n'apparaît que comme un élément de transition que nombre de métrologues ont tenté de dépasser dès sa publication. Il est d'ailleurs intéressant d'observer dans les travaux préparatifs à la rédaction du GUM qu'une consigne bien spécifique revient de façon répétée : celle d'éviter au maximum les considérations philosophiques pour se concentrer sur la recherche d'un consensus technique⁷⁰. Or, l'absence de travail conceptuel, et la focalisation sur un apparent aplanissement des difficultés *techniques* liées aux traitements conjoints des erreurs aléatoires et systématiques de mesure, sera précisément l'un des foyers de critique du GUM, en particulier dans le débat qui oppose aujourd'hui partisans de conceptualisations fréquentiste et bayésienne de la mesure, que nous décrivons dans la partie I. En quelque sorte, la recherche d'un consensus à peu de frais a eu pour conséquence une réception assez mitigée. Bien qu'il soit le résultat d'un effort réussi de standardisation, le GUM ne peut certainement pas être considéré comme l'achèvement d'un travail de réflexion sur la thématique des incertitudes des mesure. En témoigne le fait qu'alors même que ce dernier commence à diffuser dans

67. « (L)a résolution de nombreux conflits qui interviennent dans la rédaction d'une norme se cristallise très précisément autour de la recherche de la bonne formulation, du mot exact permettant d'exprimer une spécification technique sous une forme acceptable par tous les participants », [Mallard \(2000\)](#), p.42.

68. « (L)e contenu d'une norme s'analyse tout autant par l'ensemble des points sur lesquels elle ne s'engage pas que par ceux sur lesquels elle spécifie quelque chose. », [Mallard \(2000\)](#), p.56.

69. « Cette analyse nous aura appris que la normalisation est indissociable de la constitution d'une série de connaissances : la rédaction d'une norme ne se réduit pas à la recherche d'une position de compromis entre différents possibles qui seraient déjà donnés [...] Le processus de normalisation engendre donc des connaissances et des savoir-faire qui ne laissent aucune trace dans le document final, mais qui ne disparaissent pas forcément. », [Mallard \(2000\)](#), p.57.

70. Ainsi, dès la lettre d'Ambler, on peut lire : « il est important que le rapport contienne des lignes directrices pour l'énoncé des incertitudes, mais évite autant que possible des discussions philosophiques insolubles sur la théorie statistique », [Ambler \(1977\)](#), p.2. Le questionnaire de 1978 contient également la remarque suivante : « il serait souhaitable d'éviter les discussions purement philosophiques et mathématiques qui ont peu d'influence sur les questions appliquées », [Bureau International des Poids et Mesures \(1980\)](#).

des milieux de plus en plus larges, dans l'enseignement supérieur en particulier, l'objectif du JCGM est désormais de le remplacer par une seconde édition dans laquelle sont prévus des remaniements significatifs⁷¹.

71. L'on peut se référer en particulier à l'enquête menée par le JCGM auprès des laboratoires de métrologie : Bich (2012b), et aux réponses à l'enquête, rendues publiques par le JCGM à l'adresse suivante : http://www.bipm.org/wg/JCGM/JCGM-WG1/Allowed/sub-committee_5/WG1-SC5-N12-15_JCGM_GUM_Survey_Collated_responses.pdf. Les deux documents ont été consultés le 21 octobre 2015.

Conclusion générale

Le travail mené dans cette thèse soulève la question suivante : quelle est la « valeur de l'incertitude » ? Cette tournure, volontairement vague, laisse apparaître un questionnement dual. En effet, interroger la valeur de l'incertitude, c'est à la fois se demander comment celle-ci peut être calculée et résumée par une expression numérique – la valeur numérique de l'incertitude –, et chercher à comprendre ce qu'elle signifie et ce que la notion elle-même induit quant à l'interprétation d'un résultat de mesure – la valeur épistémique de l'incertitude. Notre analyse montre que ces deux aspects sont intimement intriqués : d'une part, s'interroger sur les fondements techniques de l'analyse d'incertitude mène invariablement à une réflexion d'ordre épistémologique qui engage la relation entre mesure et connaissance ; d'autre part, questionner la portée épistémologique de l'activité de mesure nécessite de comprendre la façon dont les scientifiques rendent compte des limites de leurs enquêtes expérimentales au moyen du concept d'incertitude de mesure, concept qu'ils manipulent dans un formalisme qui peut prendre une forme extrêmement sophistiquée.

Il nous faut ici rester humble quant à la portée que nous prêtons à notre travail : bien que nous fassions l'étude de la métrologie, une discipline scientifique que l'on peut qualifier de généraliste du fait de sa fonction assez singulière visant à établir des fondements communs pour la méthodologie scientifique, notre analyse reste focalisée autour d'un domaine bien spécifique et délimité, en particulier, par des exemples issus pour la plupart de la physique. Cela introduit un biais qui ne nous semble aucunement rédhibitoire mais qui laisse entrevoir la possibilité d'un élargissement des cas d'étude à la chimie, la biologie, la psychologie, les sciences humaines et d'autres disciplines, pour autant de perspectives de recherche que nous espérons prometteuses.

Dans la première partie de ce travail, notre étude des modèles statistiques fréquentiste et bayésien de l'analyse d'incertitude révèle comment les questionnements techniques auxquels sont confrontés les scientifiques amènent ces derniers à développer des problématiques épistémologiques. Nous avons présenté l'ensemble de ces développements scientifiques et techniques à partir d'un point de départ qui est le « problème de l'erreur » : les mesures physiques sont affectées d'erreurs de mesure, et l'exploitation adéquate des résultats de mesure rend nécessaire une réflexion sur leur fiabilité. C'est là qu'entre en jeu l'incertitude de mesure, comme corrélat de la neutralisation des erreurs de mesure – leur élimination physique, leur correction mathématique, leur réduction statistique. En métrologie, et dans les différentes disciplines scientifiques spécialisées, il est employé pour cela un formalisme qui fait appel au langage des probabilités.

On peut s'attendre à ce que les probabilités, qui sont introduites comme un outil technique pour manipuler mathématiquement les concepts d'erreur et d'incertitude de mesure, ne soient qu'un mode de représentation pouvant indifféremment prendre plusieurs formes sans que cela n'ait réellement d'impact sur la pratique elle-même. De ce point de vue, il n'y a rien de surprenant à ce que plusieurs modèles probabilistes soient utilisés, comme tel est le cas en métrologie. Nous avons vu que coexistent un modèle fréquentiste et un modèle bayésien, qui diffèrent d'abord par l'objet que décrivent les probabilités. Ainsi, le modèle fréquentiste caractérise par des probabilités des événements susceptibles d'avoir lieu avec une certaine fréquence limite lorsque des mesures sont répétées à l'infini. Les probabilités représentent alors une instance du monde physique : le comportement physique du processus de mesure. C'est pourquoi on parle d'une approche objective de la mesure. Le modèle bayésien emploie quant à lui les probabilités pour décrire un niveau de crédibilité accordé par un agent à une proposition donnée. Ainsi, la probabilité prend pour objet une relation entre le sujet connaissant et l'objet étudié, soumis à la mesure; cette relation est un état de croyance ou de connaissance, c'est-à-dire qu'en définitive la probabilité est un concept épistémique. Les interprétations du bayésianisme (et plus précisément ici de celui qui est appliqué à la mesure) sont multiples, en particulier selon qu'on considère ou non que l'état de croyance décrit par les probabilités dépend de la subjectivité de l'individu qui porte le jugement – en ce cas, on se trouve alors dans une approche subjectiviste de la mesure. L'examen de l'état des lieux de la métrologie contemporaine montre de façon très nette qu'il n'y a pas une simple coexistence des approches fréquentiste et bayésienne dans cette sphère spécialisée : ces deux approches sont en forte concurrence, comme le montre le contenu de très nombreux articles destinés à pointer les inconvénients et les limites d'une approche pour mieux défendre l'autre.

Cela nous suggère de soulever en passant un questionnement que nous n'avons pas évoqué dans ce travail, et qui concerne la possibilité – ou non – d'un pluralisme quant aux méthodes statistiques et aux interprétations probabilistes associées. Gillies a ébauché une approche pluraliste dans un cadre général⁷². Notre étude, dans le cadre de la métrologie, suggère que peuvent être identifiées plusieurs zones de contact entre les deux modèles que nous avons décrits – lorsque, par exemple, une vraisemblance est déterminée à partir d'une distribution de fréquence. On peut également rappeler un état de fait connu depuis longtemps, à savoir que si le fréquentisme est compatible avec un déterminisme (le fréquentisme appliqué à un tirage de type pile ou face est un exemple de fréquentisme raisonnablement déterministe), c'est qu'il n'est pas la description d'un pur hasard mais bien la représentation d'une situation donnée en rapport avec un choix de modélisation – et donc, d'une certaine façon, d'un état de connaissance. Il nous faut alors souligner que l'état actuel de la philosophie générale des probabilités développe des pistes variées quant à la nature possible du concept de probabilité, ces pistes ne se limitant pas à un simple antagonisme entre probabilités fréquentistes et épistémiques. Un point de contact entre les positionnements fréquentiste et bayésien pourrait – et nous insistons ici sur le caractère conditionnel de cette suggestion – trouver une résonance dans la conception propensionniste des probabilités. Celle-ci a été introduite d'abord par Popper pour répondre au problème des cas singuliers, qui par nature ne peuvent avoir lieu qu'une seule fois – par exemple le résultat d'une élection politique bien précise – et qui semblent donc échapper à

72. Gillies (2000), pp.187–205.

une description fréquentiste⁷³. Popper et d'autres philosophes à sa suite ont cherché à développer une interprétation probabiliste objective qui puisse couvrir ce type de cas. Or, il semble que l'interprétation propensionniste puisse amener à faire dialoguer les conceptions objective et épistémique de la probabilité, plutôt que de simplement les opposer. Nous laissons ici cette réflexion ouverte.

Que les approches fréquentiste et bayésienne soient en concurrence semble nous indiquer que la question de la nature des probabilités employées dans les modèles statistiques n'est pas simplement une question de mode de représentation et de choix de langage mathématique. Si le débat tire son origine d'une question technique, à savoir « quel formalisme est le plus adapté à l'analyse d'incertitude selon les objectifs que l'on s'impose ? », il ne se limite pas seulement au choix d'un outil parmi une palette de méthodes disponibles qui pourraient se révéler également valables. De fait, nous avons vu que la confrontation de ces méthodes statistiques engagent des débats nourris qui orientent la discussion vers un certain nombre de problématiques épistémologiques. La mesure est-elle une activité objective, ou bien présente-t-elle une part incompressible de subjectivité ? Peut-on évaluer l'exactitude d'un résultat de mesure, comprise comme une mesure de l'erreur possible affectant le résultat en question, ou doit-on se contenter de décrire un état de connaissance relatif à un agent (ou à un groupe d'agents) sans prétendre dire quoi que soit à propos de la grandeur mesurée elle-même ? En quoi, par suite, peut-on estimer que la mesure physique donne accès, d'une façon ou d'une autre, à une réalité physique traduite en termes de valeurs numériques « vraies » – ce d'autant plus si, à l'issue de la première question, on a reconnu le caractère subjectif de l'activité de mesure ? Ce sont autant de questions philosophiques que les métrologues et scientifiques ne résolvent pas mais à propos desquelles ils ont tendance à se positionner lorsqu'ils discutent des fondements et des ramifications des modèles techniques qu'ils emploient. L'enjeu que nous faisons ressortir est le suivant : la mesure est-elle une pratique *évaluative*, comme le prétendent les fréquentistes, ou bien n'est-elle qu'une pratique *descriptive*, comme semble l'induire la façon dont un certain nombre de métrologues proposent d'appliquer aujourd'hui l'approche bayésienne à la mesure ? Les deux parties suivantes de cette thèse reprennent ce questionnement en toile de fond.

Les réponses à ces questions ne sauraient pas légitimer de façon absolue l'une ou l'autre des méthodes, mais sont des conséquences de discussions techniques lors desquelles les spécialistes ne peuvent plus simplement se contenter de développer un arsenal formel, mais doivent envisager les conséquences conceptuelles et philosophiques de leurs méthodologies. Nous avons défendu qu'il n'y aurait pas de sens à trancher entre les différentes méthodes statistiques, ni entre les épistémologies qui les accompagnent, en particulier parce que celles-ci ne sont pas justifiables a priori mais sont attachées à des objectifs particuliers eux-mêmes liés à un contexte qui est celui de la société contemporaine, avec ses enjeux techniques, économiques et scientifiques.

La seconde partie de ce travail pousse plus loin notre analyse de l'épistémologie appliquée des métrologues et interroge le statut d'un concept traditionnel, plutôt intuitif, mais néanmoins difficile à définir et à caractériser de façon satisfaisante : la « valeur vraie » d'une grandeur. Certains textes de la métrologie contemporaine, qui font la synthèse des évolutions observées

73. Voir par exemple Gillies (2000), pp.113–136.

depuis le milieu du XX^e siècle et les années 1970 en particulier, tentent de dissimuler ou de faire disparaître le concept du formalisme : il est affirmé qu'il est possible de décrire un résultat de mesure uniquement en des termes réputés « connaissables ». Les différents développements qui sont présentés dans ces textes peuvent être regroupés sous la même caractérisation commune d'une approche « épistémique » de la mesure.

La question des méthodes statistiques employées et de l'interprétation probabiliste associée n'est plus directement soulevée ici : la critique de la valeur vraie d'une grandeur n'est pas en soi une caractéristique spécifique de l'opposition entre approches fréquentiste et bayésienne de la mesure. D'ailleurs, l'idée que ce concept ne renvoie qu'à une vision idéalisée des grandeurs physiques, ou encore qu'il constitue un objectif inatteignable, a été évoquée à plusieurs reprises par les scientifiques, métrologues et statisticiens bien avant la critique du fréquentisme en métrologie. Cependant, la notion de valeur vraie est devenue, au fil du débat contemporain, un marqueur de l'approche fréquentiste traditionnelle, car elle renvoie à l'idéal d'objectivité cher aux fréquentistes. À l'inverse, le mouvement de réforme que constitue l'application du bayésianisme à la mesure accompagne un tournant épistémique qui engage une remise en question de la notion de valeur vraie, ce qui amène les mouvements à se recouvrir partiellement.

Nous avons vu que deux arguments spécifiques ont tout particulièrement été opposés à l'usage et à la légitimité du concept de valeur vraie : l'argument d'inconnaissabilité d'une part, et l'argument de non-unicité d'autre part. Notre position, décrite dans la seconde partie de cette thèse, consiste à concéder qu'il est impossible de jamais connaître *exactement* la valeur vraie d'une grandeur, mais qu'il demeure possible de prétendre la connaître *approximativement*. Dès lors, que la valeur vraie d'une grandeur ne soit pas unique n'est pas rédhibitoire sur le plan conceptuel, mais nous amène à réexaminer la structure des théories et des modèles dans lesquels s'insèrent les grandeurs physiques. Pour que notre réponse soit satisfaisante, il faut toutefois accepter que « connaître approximativement » la valeur vraie d'une grandeur ne signifie pas pouvoir la localiser avec certitude dans un intervalle fini de valeurs. De fait, il est impossible de garantir l'exactitude d'un résultat de mesure. Or, l'idée fondamentale qui accompagne la notion de valeur vraie d'une grandeur n'est pas qu'il est possible de donner l'exactitude d'un résultat de mesure à un instant donné, mais qu'il est possible d'améliorer cette exactitude au moyen d'un processus de correction des erreurs de mesure. De ce point de vue, s'il y a un progrès possible des connaissances, c'est qu'il y a une connaissance possible quant à la valeur vraie. S'il est impossible de justifier totalement ce point de vue – car il demande un attachement au réalisme scientifique qui ne peut pas trouver de justification logique ou empirique –, il nous apparaît, comme l'illustre d'ailleurs la troisième partie de cette thèse, que celui-ci correspond en grande partie à la conception générale qu'ont les scientifiques des grandeurs physiques et de l'activité de mesure, contrairement à ce que défendent les documents de métrologie discutés jusqu'ici.

De fait, cette position n'est pas simplement une posture épistémologique abstraite : elle peut être mise en rapport avec ce que les physiciens eux-mêmes ont perçu dans la pratique des ajustements des constantes de la physiques, dont l'étude fait l'objet de la troisième partie de cette thèse. En cherchant à déterminer les « meilleures » valeurs des constantes de la phy-

sique à un instant donné, en faisant la synthèse des connaissances disponibles à propos de ces constantes, les physiciens ont tout à fait conscience du fait qu'ils n'aboutissent pas à des valeurs vraies ou exactes, mais simplement à des valeurs cohérentes avec leurs modèles et leurs mesures, qui forment un consensus légitime. Cependant, ils ne conçoivent pas l'impossibilité de garantir l'exactitude d'une mesure comme l'impossibilité de connaître la valeur des constantes de la physique. Au contraire, ils déploient une épistémologie tournée vers un progrès futur par la mise en place d'un processus permanent et sans fin de correction des erreurs expérimentales ; ce faisant, et sans pour autant qu'ils adhèrent spécifiquement à une conception fréquentiste de la mesure, ils montrent qu'ils conservent un attachement au caractère évaluatif de la mesure que la conception épistémique contemporaine de la mesure a tendance à minorer.

En interrogeant la valeur de l'incertitude, nous attaquons la thématique de la précision des mesures physiques et des limites de la connaissance expérimentale. Le terme de précision lui-même est plutôt intuitif, mais ambigu, et il renvoie à quantité de concepts plus complexes et parfois mutuellement incompatibles. En particulier, une dualité traverse cette thèse : la dualité entre incertitude et exactitude. L'incertitude de mesure, dans son acception épistémique, porte sur un état de croyance et, par extension, de connaissance. L'exactitude de mesure renvoie au versant métaphysique de la mesure en exprimant un rapport de proximité d'un résultat avec la vérité. La question des limites de la connaissance expérimentale qui sous-tend cette thèse revient, dans le cadre de la métrologie contemporaine, à s'interroger sur le rapport qu'il est possible de faire émerger entre ces deux concepts, compris en fin de compte comme deux versants de la notion de connaissance. Il n'est bien sûr pas possible de répondre simplement et directement à cette question, puisque la connaissance est par nature toujours incertaine. Notre travail nous mène plutôt à la conclusion que « connaître » n'est pas tant dire le « vrai » que de chercher à l'atteindre.

Annexes

Annexe A

Développements techniques sur les méthodes statistiques

Notations : dans cette annexe, on désignera par $E(\hat{a})$ l'espérance d'une variable aléatoire \hat{a} et par $V(\hat{a})$ sa variance.

A.1 Estimateurs statistiques

Nous souhaitons expliciter en détail les raisons qui poussent à considérer la moyenne arithmétique \bar{x} comme un bon estimateur de $\mu_{\hat{x}}$ et nous montrerons que l'on aboutit en particulier au résultat contre-intuitif selon lequel \bar{x} peut être un bon estimateur de $\mu_{\hat{x}}$ tout en étant fortement inexact. En effet, ce n'est pas *la* grandeur statistique *particulière* \bar{x} qui est un bon estimateur de $\mu_{\hat{x}}$, mais le choix *a priori* de \bar{x} comme estimateur de $\mu_{\hat{x}}$ qui est un bon choix, en ce qu'il a *généralement* tendance à produire des estimateurs dont l'exactitude est satisfaisante. Pour cela, il faut raisonner non pas sur les valeurs obtenues, mais sur le *processus* dans son ensemble – ici, un tirage aléatoire parmi une population statistique parente \hat{x} . La nature fréquentiste des probabilités employées influence le mode de raisonnement que l'on peut employer.

Le modèle fréquentiste réduit la mesure répétée d'une grandeur au tirage aléatoire d'une valeur x_i parmi une population décrite par une variable aléatoire \hat{x} . On obtient ainsi un échantillon de n données expérimentales $\{x_i\}$. L'obtention de cet échantillon peut lui-même être considéré comme un tirage aléatoire parmi une population d'échantillons possible. Il en est de même pour l'obtention de la moyenne arithmétique \bar{x} .

De même qu'une variable aléatoire \hat{x} vient représenter la procédure de tirage aléatoire d'une donnée expérimentale x_i parmi la population de valeurs possibles, on peut également définir une variable aléatoire $\hat{\bar{x}}$ qui représente le tirage aléatoire d'un estimateur lors d'une

expérience donnée ¹

$$\hat{\bar{x}} = \frac{1}{n} \sum_{i=1}^n \hat{x}_i \quad (\text{A.1})$$

(où les \hat{x}_i sont la même variable aléatoire \hat{x} considérées comme *indépendantes* les unes des autres).

Insistons sur le fait qu'au tirage aléatoire d'un échantillon de n données expérimentales correspond celui d'un *seul* estimateur \bar{x} . En pratique, l'expérimentateur utilise la moyenne arithmétique des données dont il dispose. L'approche fréquentiste consiste à raisonner de façon hypothétique sur l'ensemble des estimateurs \bar{x} que l'expérimentateur *aurait pu* obtenir s'il avait répété l'expérience consistant à mesurer un échantillon de n mesures un grand nombre de fois. Si l'expérimentateur effectuait cette opération, il aboutirait à un échantillon d'estimateurs lui aussi dispersé, mais cette dispersion serait moins importante que celle qui affecte les échantillons de données expérimentales à partir desquels les estimateurs sont calculés.

Dans l'algèbre des variables aléatoires, l'espérance et la variance sont des grandeurs additives, c'est-à-dire que si A et B sont deux variables aléatoires indépendantes, leurs espérances et variances sont telles que :

$$\begin{cases} E(\hat{a} + \hat{b}) = E(\hat{a}) + E(\hat{b}) \\ V(\hat{a} + \hat{b}) = V(\hat{a}) + V(\hat{b}) \end{cases} \quad (\text{A.2})$$

Par conséquent, on démontre facilement que $E(\hat{\bar{x}}) = E(\hat{x})$ (ce qui était déjà contenu dans la loi des grands nombres mentionnée au chapitre 3), mais surtout que :

$$V(\hat{\bar{x}}) = \frac{V(\hat{x})}{n} \quad (\text{A.3})$$

On retrouve ici le résultat énoncé au chapitre 3, que l'on peut désormais formuler dans les termes utilisés ici : la variance de $\hat{\bar{x}}$ est d'autant plus faible que :

- (i) La dispersion des données expérimentales est faible, c'est-à-dire que l'écart-type σ de la variable aléatoire \hat{x} – ou, alternativement, sa variance $V = \sigma^2$ est faible.
- (ii) La taille de l'échantillon, c'est-à-dire le nombre n de mesures effectuées, est grande, ce qui montre que la qualité du résultat final est d'autant meilleure que la mesure est répétée, c'est-à-dire que la quantité d'informations prises à propos de la cible est grande.

La variance de $\hat{\bar{x}}$ peut être considérée comme une mesure de la fiabilité de \bar{x} comme estimateur, puisqu'elle caractérise la tendance qu'a un estimateur \bar{x} à différer de la cible $\mu_{\hat{x}}$. Cependant, cette variance est inconnue au même titre que $\mu_{\hat{x}}$ et σ .

1. On peut en fait, pour chaque grandeur relative au problème obtenue après mesure, associer une variable aléatoire qui décrira la procédure d'obtention de la grandeur. Les procédures sont décrites par des variables aléatoires (symbolisées avec un chapeau) et les résultats effectifs sont décrits par des nombres (symbolisés par des lettres minuscules).

Il est donc nécessaire de construire un estimateur de $V(\hat{x}) = V(\hat{x})/n$, c'est-à-dire de construire un estimateur de V . On définit la variance empirique s^2 de l'échantillon $\{x_i\}$ ainsi :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (\text{A.4})$$

De même que \bar{x} , s^2 est une grandeur fixe, déterminée par l'échantillon effectivement obtenu après répétition de la mesure. Mais on peut également caractériser le processus de production de l'estimateur par une variable aléatoire \hat{s}^2 définie ainsi :

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{x}_i - \hat{x})^2 \quad (\text{A.5})$$

s^2 est un estimateur « sans biais » de V car $E(s^2) = V$. Par conséquent s^2/n est un estimateur sans biais de $V(\hat{x})$. C'est la raison pour laquelle on définit l'« incertitude-type » $u(x)$ sur x par :

$$u(x) = \sqrt{V(s^2)} = \sqrt{\frac{s^2}{n}} \quad (\text{A.6})$$

On parle d'« incertitude-type » car le raisonnement employé jusqu'ici ne s'appuie que sur des calculs d'espérance et de variance, c'est-à-dire sur des calculs en termes de *limites* uniquement. Or, le modèle fréquentiste vise à aller plus loin en formulant des intervalles de confiance permettant d'associer un niveau de confiance, c'est-à-dire une probabilité, à un intervalle de valeurs choisi pour représenter la valeur de la grandeur mesurée. Pour cela, il est nécessaire d'introduire une hypothèse supplémentaire, et de mener des calculs probabilistes plus poussés. C'est ce que nous développons dans la section A.2 de cette annexe.

A.2 Intervalles de confiance

L'hypothèse (H3) stipule la forme de la distribution de probabilité de la variable aléatoire \hat{x} . Dans cette annexe, nous menons les calculs sur la base de l'hypothèse gaussienne – ce qui est le cadre du GUM.

(H3.1) \hat{x} est distribuée selon une loi normale :

$$f_{\hat{x}}(t) \propto \exp\left(-\frac{1}{2} \left(\frac{t - \mu_{\hat{x}}}{\sigma}\right)^2\right)$$

Une loi normale est entièrement définie par ses deux paramètres $\mu_{\hat{x}}$ et σ . On utilise encore la notation $\mathcal{N}(\mu_{\hat{x}}, \sigma)$ pour désigner une telle loi de probabilité.

L'hypothèse (H3) ouvre la voie à un calcul probabiliste dit « déductif ». Si l'on connaît la distribution de \hat{x} , on connaît alors celle de $\hat{\bar{x}}$, de \hat{s}^2 , ou de toute autre fonction dont la définition fait appel uniquement à \hat{x} . On peut montrer que, si $\hat{x} = \mathcal{N}(\mu_{\hat{x}}, \sigma)$, alors $\hat{\bar{x}} = \mathcal{N}(\mu_{\hat{x}}, \sigma/\sqrt{n})$ en raison de l'additivité de la loi normale².

2. Voir Saporta (2006), p.45.

La cible de l'inférence statistique est $\mu_{\hat{x}}$. Cependant, \hat{x} dépend de deux paramètres inconnus, $\mu_{\hat{x}}$ et σ . Pour pouvoir formuler une proposition à propos du paramètre $\mu_{\hat{x}}$ à partir des données, il faut réussir à l'isoler, c'est-à-dire à faire disparaître σ des calculs. Pour cela, on introduit la loi de probabilité suivante :

$$\hat{T}_{n-1} = \frac{\hat{x} - \mu}{\sqrt{\hat{s}^2/n}} \quad (\text{A.7})$$

En développant le calcul, on montre que :

$$\hat{T}_{n-1} = \frac{\frac{\hat{x} - \mu}{\sigma/\sqrt{n}}}{\frac{\sqrt{\hat{s}^2}}{\sigma}} = \sqrt{n-1} \frac{\frac{\hat{x} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\sum_{i=1}^n \left(\frac{\hat{x}_i - \hat{x}}{\sigma}\right)^2}} \quad (\text{A.8})$$

On peut étudier séparément le numérateur et le dénominateur de l'expression précédente. Au numérateur, la fonction suivante :

$$\frac{\hat{x} - \mu}{\sigma/\sqrt{n}}$$

est une « loi normale centrée réduite » $\mathcal{N}(0, 1)$, c'est-à-dire une loi normale d'espérance nulle et d'écart-type unitaire³. Au dénominateur, la fonction suivante :

$$\sum_{i=1}^n \left(\frac{\hat{x}_i - \hat{x}}{\sigma}\right)^2$$

est une loi du « khi-deux » χ_{n-1}^2 à $n - 1$ degrés de liberté, une distribution usuelle des statistiques⁴. Par conséquent, l'expression de \hat{T}_{n-1} correspond exactement à une « distribution de Student » à $n - 1$ degrés de liberté. Cette distribution a été introduite en 1908 par William Gosset, sous le pseudonyme de Student⁵ – il travaillait alors pour les brasseries Guinness et n'avait pas l'autorisation de publier en son nom⁶ – alors qu'il s'attachait justement à résoudre un problème de ce type.

L'intérêt principal de la construction de \hat{T}_{n-1} est que l'expression de cette loi de probabilité ne dépend ici que du seul paramètre inconnu $\mu_{\hat{x}}$; comme souhaité, l'écart-type inconnu σ de \hat{x} n'apparaît plus. Les distributions de Student sont bien documentées et il est possible de se référer à un tableau de valeurs des « quantiles » pour un nombre donné de degrés de liberté

3. Voir Saporta (2006), p.43.

4. Par définition, une loi du khi-deux à $n - 1$ degrés de liberté est définie comme la somme quadratique de $n - 1$ lois normales centrées réduites indépendantes; voir par exemple Saporta (2006), p.93. Le résultat, dans le cas présent, n'est pas immédiat, du fait de la présence de la moyenne arithmétique \hat{x} qui n'est pas indépendante des variables \hat{x}_i puisqu'elle est précisément définie à partir de ces dernières. Il serait inutile de développer la démonstration; nous pouvons nous contenter de mentionner que la présence de \hat{x} contribue à réduire le nombre de degrés de liberté de n à $n - 1$.

5. Student (1908)

6. Zabell (2008), p.1.

(voir tableau A.1). On peut ainsi dire que le tirage aléatoire d'une valeur parmi la population d'une loi de Student à ν degrés de liberté a une probabilité p d'aboutir à une valeur comprise dans l'intervalle $[-t_p(\nu); t_p(\nu)]$. Transcrit dans le cas de la mesure répétée d'une grandeur, un échantillon de n valeurs expérimentales sera tel que :

$$\frac{\bar{x} - \mu_{\hat{x}}}{u(x)} \in [-t_p(\nu); t_p(\nu)] \quad \text{avec une fréquence limite } p \quad (\text{A.9})$$

On peut également écrire, par une transformation simple :

$$\mu_{\hat{x}} \in I_p = [\bar{x} - t_p(\nu) u(x); \bar{x} + t_p(\nu) u(x)] \quad \text{avec une fréquence limite } p \quad (\text{A.10})$$

L'intervalle I_p est appelé « intervalle de confiance » associé au niveau de confiance p . Le quantile $t_p(\nu)$ est appelé dans le GUM « facteur d'élargissement » et y est désigné par le signe k ou k_p . Il convient encore une fois de remarquer que $\mu_{\hat{x}}$ et I_p étant fixes, les équations (A.9) et (A.10) ne valent que pour la *procédure*, c'est-à-dire pour les valeurs *potentielles* que la procédure *aurait* pu produire, si elle avait effectivement été répétée une infinité de fois.

A.3 Traitement bayésien des méthodes de type A

Dans cette annexe, nous souhaitons détailler les aspects techniques du calcul d'incertitude dans le cas d'une mesure répétée. Nous nous inspirons très fortement de la description donnée par Ignacio Lira dans son livre sur l'évaluation de l'incertitude de mesure⁷.

On considère une grandeur X accessible par la mesure et dont on a effectuée n mesures dans des conditions considérées comme identiques (conditions de répétabilité). On obtient ainsi un échantillon de n données expérimentales x_i . La méthode bayésienne se déploie en trois étapes :

- (1) Déterminer la distribution *a priori* pour $\hat{\chi}$
- (2) Déterminer la fonction de vraisemblance
- (3) Appliquer le théorème de Bayes pour obtenir la distribution *a posteriori*.

Comme nous le verrons, l'étape (3) se réduit ici à une simple multiplication.

Étape (1) : *prior* La distribution *a priori* de $\hat{\chi}$ est construite à partir des informations disponibles. Si aucune information spécifique n'est disponible, il est raisonnable d'adopter une distribution *a priori non informative*, c'est-à-dire uniforme sur l'espace des réels⁸.

7. Lira (2002)

8. Ces *prior* ne sont pas exempts de limites ; en particulier, ces distributions sont singulières du fait qu'elles ne sont pas normalisables. Toutefois, ils permettent souvent les dérivations mathématiques les plus simples et demeurent couramment utilisés. Lira adresse et tempère certaines critiques, voir Lira (2002), pp.173–175.

| Nombre de degrés de liberté $\nu = n - 1$ | probabilité p | | | | | |
|--|-----------------|-------|-------|-------|-------|-------|
| | 68,27 | 90 | 95 | 95,45 | 99 | 99,73 |
| 1 | 1,84 | 6,31 | 12,71 | 13,97 | 63,66 | 235,8 |
| 2 | 1,32 | 2,92 | 4,3 | 4,53 | 9,92 | 19,21 |
| 3 | 1,2 | 2,35 | 3,18 | 3,31 | 5,84 | 9,22 |
| 4 | 1,14 | 2,13 | 2,78 | 2,87 | 4,6 | 6,62 |
| 5 | 1,11 | 2,02 | 2,57 | 2,65 | 4,03 | 5,51 |
| 6 | 1,09 | 1,94 | 2,45 | 2,52 | 3,71 | 4,9 |
| 7 | 1,08 | 1,89 | 2,36 | 2,43 | 3,5 | 4,53 |
| 8 | 1,07 | 1,86 | 2,31 | 2,37 | 3,36 | 4,28 |
| 9 | 1,06 | 1,83 | 2,26 | 2,32 | 3,25 | 4,09 |
| 10 | 1,05 | 1,81 | 2,23 | 2,28 | 3,17 | 3,96 |
| 11 | 1,05 | 1,8 | 2,2 | 2,25 | 3,11 | 3,85 |
| 12 | 1,04 | 1,78 | 2,18 | 2,23 | 3,05 | 3,76 |
| 13 | 1,04 | 1,77 | 2,16 | 2,21 | 3,01 | 3,69 |
| 14 | 1,04 | 1,76 | 2,14 | 2,2 | 2,98 | 3,64 |
| 15 | 1,03 | 1,75 | 2,13 | 2,18 | 2,95 | 3,59 |
| 16 | 1,03 | 1,75 | 2,12 | 2,17 | 2,92 | 3,54 |
| 17 | 1,03 | 1,74 | 2,11 | 2,16 | 2,9 | 3,51 |
| 18 | 1,03 | 1,73 | 2,1 | 2,15 | 2,88 | 3,48 |
| 19 | 1,03 | 1,73 | 2,09 | 2,14 | 2,86 | 3,45 |
| 20 | 1,03 | 1,72 | 2,09 | 2,13 | 2,85 | 3,42 |
| 25 | 1,02 | 1,71 | 2,06 | 2,11 | 2,79 | 3,33 |
| 30 | 1,02 | 1,7 | 2,04 | 2,09 | 2,75 | 3,27 |
| 35 | 1,01 | 1,7 | 2,03 | 2,07 | 2,72 | 3,23 |
| 40 | 1,01 | 1,68 | 2,02 | 2,06 | 2,7 | 3,2 |
| 45 | 1,01 | 1,68 | 2,01 | 2,06 | 2,69 | 3,18 |
| 50 | 1,01 | 1,68 | 2,01 | 2,05 | 2,68 | 3,16 |
| 100 | 1,005 | 1,66 | 1,984 | 2,025 | 2,626 | 3,077 |
| ∞ | 1 | 1,645 | 1,96 | 2 | 2,576 | 3 |

TABLEAU A.1 – Tableau de valeur des quantiles $t_p(\nu)$ de la distribution de Student pour différents degrés de libertés et différentes probabilités. Pour un nombre ν de degrés de liberté donné, la probabilité indiquée en haut de colonne est la probabilité qu'il y a d'obtenir une valeur comprise dans l'intervalle $[-t_p(\nu); t_p(\nu)]$ lors d'un tirage aléatoire selon cette loi. Reproduit à partir du tableau G.2 du GUM, [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.81.

Étape (2) : vraisemblance La détermination de la vraisemblance est l'étape la plus complexe. Lira situe l'origine de la fonction de vraisemblance dans une modélisation probabiliste *fréquentiste* :

À chaque fois qu'un ensemble d'observations [...] est donné, il faut essayer d'insister sur les raisons physiques qui mènent à l'apparition de différentes valeurs. Si l'enquête expérimentale nous amène à croire que les écarts [...] à la (vraie) valeur de la quantité sont causés par plusieurs effets aléatoires et que ces déviations sont indépendantes les unes des autres, il est raisonnable de penser que les données sont tirées indépendamment d'une population infinie de valeurs dont la distribution de fréquence est normale⁹

De ce point de vue, Lira s'appuie sur la variable aléatoire fréquentiste \hat{x} décrite au chapitre 3, et le choix d'une distribution comme la loi normale relève de l'hypothèse (H3) mentionnée au 3.4. Cette probabilité est interprétée comme une probabilité *conditionnelle* : la probabilité d'obtenir une donnée $x_i = t$ est *conditionnée* aux valeurs μ et V de l'espérance et de la variance de la distribution. La densité de probabilité est une fonction qui prend pour variable la valeur *possible* t de x_i et dans laquelle μ (assimilée à la valeur vraie χ de X) et V sont traitées comme des *paramètres inconnus*.

$$p(x_i = t | \mu, V) \propto \frac{1}{\sqrt{V}} \exp \left[-\frac{1}{2V} (t - \mu)^2 \right] \quad (\text{A.11})$$

On en déduit la fonction de vraisemblance pour une mesure individuelle. Les arguments sont inversés par rapport à la fonction précédente : la vraisemblance est une fonction qui prend pour argument les variables μ et V et dans laquelle x_i est un paramètre connu et fixe.

$$L_i(\mu, V | x_i) \propto \frac{1}{\sqrt{V}} \exp \left[-\frac{1}{2V} (x_i - \mu)^2 \right] \quad (\text{A.12})$$

Comme les mesures sont considérées comme indépendantes, la fonction de vraisemblance globale pour un échantillon de données $\{x_i\}$ est obtenue par le produit des vraisemblances individuelles¹⁰ :

$$L(\mu, V | \{x_i\}) \propto \frac{1}{V^{n/2+1}} \exp \left[-\frac{1}{2V} \sum_{i=1}^n (x_i - \mu)^2 \right] \quad (\text{A.13})$$

On peut réécrire l'expression A.13 à l'aide de la moyenne arithmétique \bar{x} et de la variance empirique s^2 :

$$L(\mu, V | \{x_i\}) \propto \frac{1}{V^{n/2+1}} \exp \left[-\frac{1}{2V} ((n-1)s^2 + n(\mu - \bar{x})^2) \right] \quad (\text{A.14})$$

9. Lira (2002), p.176. Lira avait déjà énoncé cette dépendance quelques pages plus tôt : « dans la plupart des cas, la vraisemblance est construite selon un modèle probabiliste qui est supposé fonctionner pour le problème de mesure considéré. Le modèle de probabilité est essentiellement une distribution de fréquence », p.175.

10. Cette étape contient une opération plus subtile impliquant un *prior* particulier, le « prior de Jeffreys », voir Lira (2002), p.176. Cela revient à multiplier la vraisemblance obtenue par un facteur $1/V$. La dénomination « prior de Jeffreys » fait référence au mathématicien anglais Harold Jeffreys.

L'une des difficultés de l'opération repose sur le fait que l'objectif est d'obtenir une probabilité sur μ , assimilée à la valeur vraie de X , et ce *indépendamment* de la valeur de V . Or, l'expression A.14 dépend de V . Le paramètre V est donc considéré ici comme un « paramètre de nuisance »¹¹ qu'il faut faire disparaître en intégrant sur les valeurs possibles de V .

Cette dernière manipulation, nommée marginalisation¹², s'appuie sur une propriété classique des probabilités conditionnelles. Si l'univers est partitionné en événements A_i , alors la probabilité d'une proposition B se décompose comme suit¹³ :

$$P(B) = \sum_i P(B|A_i) P(A_i) \quad (\text{A.15})$$

Pour les probabilités discrètes, la relation est également valable. Appliquée au cas présent, on a alors :

$$p(\{x_i\}|\mu) \propto \int_0^{+\infty} p(\{x_i\} = \{t_i\}|\mu, V) p(V|\mu) \quad (\text{A.16})$$

Si l'on considère que toutes les valeurs positives de V sont *a priori* équiprobables, avec la seule contrainte que $V \geq 0$, alors on a :

$$p(\{x_i\}|\mu) \propto \int_0^{+\infty} \frac{1}{V^{n/2+1}} \exp\left[-\frac{1}{2V} ((n-1)s^2 + n(\mu - \bar{x})^2)\right] dV \quad (\text{A.17})$$

On peut reprendre cette expression pour une construire la nouvelle fonction de vraisemblance, qui, par conséquent, ne prend plus qu'une unique variable μ :

$$L(\mu|\{x_i\}) \propto \int_0^{+\infty} \frac{1}{V^{n/2+1}} \exp\left[-\frac{1}{2V} ((n-1)s^2 + n(\mu - \bar{x})^2)\right] dV \quad (\text{A.18})$$

Après intégration, on aboutit à :

$$L(\mu|\{x_i\}) \propto [(n-1)s^2 + n(\mu - \bar{x})^2]^{-n/2} \quad (\text{A.19})$$

Qui est l'expression recherchée de la fonction de vraisemblance. Comme souhaité, elle ne dépend que de la seule variable inconnue, μ . Or, comme nous l'avons déjà mentionné, μ est assimilé à la valeur vraie χ de X ; cela correspond à l'hypothèse (H2) du modèle fréquentiste, que nous avons exprimé à la section 3.2.2 du chapitre dévolu aux approches fréquentistes de la mesure. Nous pouvons donc écrire :

$$L(\chi|\{x_i\}) \propto [(n-1)s^2 + n(\chi - \bar{x})^2]^{-n/2} \quad (\text{A.20})$$

11. Lira (2002), p.176.

12. Lira (2002), p.177.

13. Drouet (Drouet, 2016) rappelle qu'appliquée aux croyances, cette règle est connue sous le nom de « règle de Jeffrey »; voir en particulier Jeffrey (1983). Le nom de « règle de Jeffrey » fait référence au statisticien bayésien Richard C. Jeffrey, qu'il ne faut pas confondre avec le mathématicien Harold Jeffreys mentionné précédemment.

Étape 3 : distribution a posteriori. Le théorème de Bayes stipule que :

$$p(\hat{\chi} | \{x_i\}) \propto L(\hat{\chi} | \{x_i\}) \times p(\hat{\chi}) \quad (\text{A.21})$$

Le prior choisi étant non informatif, on a donc simplement :

$$f_{\hat{\chi}}(t) \propto [(n-1)s^2 + n(t - \bar{x})^2]^{-n/2} \quad (\text{A.22})$$

que l'on peut réécrire :

$$f_{\hat{\chi}}(t) \propto \left[1 + \frac{1}{n-1} \left(\frac{\sqrt{n}}{s} (t - \bar{x}) \right)^2 \right]^{-n/2} \quad (\text{A.23})$$

Par un changement de variable $T = \sqrt{n} (t - \bar{x}) / s$, on obtient finalement :

$$f_{\hat{\chi}}(T) \propto \left[1 + \frac{T^2}{n-1} \right]^{-n/2} \quad (\text{A.24})$$

C'est l'expression de la densité de probabilité d'une loi de Student à $(n-1)$ degrés de liberté¹⁴. Par conséquent, on retrouve un résultat similaire à celui qui est obtenu, dans les mêmes conditions, par le modèle fréquentiste. Comme l'écrit Lira,

(L)'intervalle élargi symétrique [...] est [...] le même que celui qui est obtenu par une analyse conventionnelle [de type A]¹⁵

L'expression de l'incertitude-type est cependant différente, car elle est calculée directement comme l'écart-type de la distribution obtenue. Par conséquent, on a :

$$u(x) = \frac{n-1}{n-3} \sqrt{\frac{s^2}{n}} \quad (\text{A.25})$$

A.4 Randomisation des erreurs systématiques : un exemple

Dans cette section, nous décrivons sur un exemple une façon dont on peut modifier un processus de mesure de façon à transformer une erreur systématique en erreur aléatoire, ce que l'on désigne par l'expression, tirée de l'anglais, « randomisation » d'une erreur systématique. Nous prenons pour cela l'exemple de la mesure de la longueur d'un objet quelconque à l'aide d'une règle graduée au millimètre.

Cas ordinaire. Le processus de mesure ordinaire est très simple : l'une des extrémités de l'objet à mesurer est placée au niveau du 0 de la règle ; on reporte alors les graduations entres lesquelles se situe l'autre extrémité de l'objet (figure A.1).

14. Saporta (2006), pp.98–99.

15. Lira (2002), p.178.

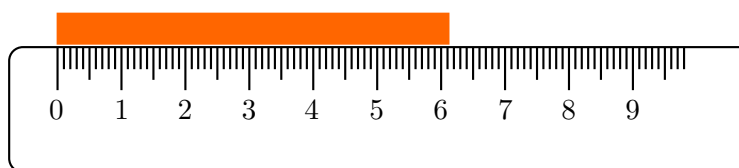


FIGURE A.1 – Mesure ordinaire de la longueur d'un objet à l'aide d'une règle graduée. L'objet est placé parallèlement à la règle de façon à ce que l'une de ses extrémités coïncide avec la graduation 0 de la règle. On peut alors lire immédiatement l'indication donnée par la graduation à laquelle se situe l'extrémité opposée de l'objet. En général, cette extrémité ne coïncide pas avec une graduation de la règle, et on en tire un encadrement de la longueur de l'objet. Il subsiste une erreur de mesure, que l'on appelle dans ce cas « erreur de résolution ». Si l'on répète la mesure, on obtient toujours le même résultat, à moins que les conditions de mesure varient de telle façon qu'on ne puisse plus considérer qu'il s'agit de conditions de répétabilité. Par conséquent, l'erreur de mesure sur la longueur de l'objet est une erreur systématique. On peut l'évaluer par des limites d'erreur, ou déterminer une incertitude-type correspondante en exploitant par exemple les méthodes de « type B » présentées dans le GUM.

Si l'extrémité opposée de l'objet se situe entre les graduations 6,1 et 6,2 d'une règle graduée au millimètre, comme sur la figure A.1, alors on peut conclure, en appelant ℓ la longueur de l'objet :

$$6,1 \text{ cm} < \ell < 6,2 \text{ cm} \quad (\text{A.26})$$

Dans des conditions de répétabilité, la répétition de la mesure amènera toujours au même résultat, car les fluctuations intrinsèques du dispositif ne dépassent pas la limite de résolution de l'appareil de mesure. Par conséquent, l'erreur commise est une erreur systématique. Il s'agit d'une erreur de résolution (ce que Burns, Campion et Williams appellent également une « limite d'estimation »¹⁶) : l'instrument n'offre pas de meilleure précision.

Dans ce cas, l'on dira par exemple que la longueur de l'objet est $\ell = 6,15 \text{ cm}$ avec une erreur de mesure comprise entre les limites $-0,05 \text{ cm}$ et $+0,05 \text{ cm}$. Il s'agit d'un cas référencé dans le GUM si l'on cherche à évaluer une incertitude par une méthode « de type B », et l'incertitude-type correspondante est :

$$u = \frac{1}{\sqrt{12}} \text{ mm} \simeq 0,3 \text{ mm} \quad (\text{A.27})$$

auquel cas on peut écrire :

$$\ell = 6,15(3) \text{ cm} \quad (\text{A.28})$$

Cas « randomisé ». Le code de pratique du NPL propose une méthode pour faire apparaître l'erreur de résolution comme une erreur aléatoire¹⁷. Nous nous inspirons pour la description

16. Campion, Burns et Williams (1973), p.3.

17. Campion, Burns et Williams (1973), pp.3-4.

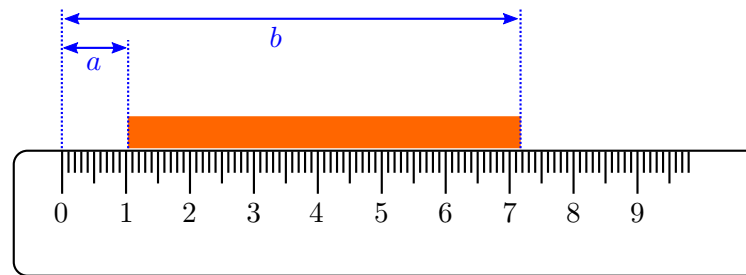


FIGURE A.2 – Méthode « randomisée » d’une mesure de longueur à l’aide d’une règle graduée. Le protocole stipule que la position initiale de l’objet à mesurer doit être tirée aléatoirement. Par conséquent, elle ne coïncide pas toujours avec une graduation de la règle. La position initiale de l’objet a une conséquence sur le nombre de graduations qui séparent les deux extrémités de l’objet à mesurer ; contrairement au cas ordinaire, ce nombre de graduations peut varier d’un mesurage à l’autre. L’erreur systématique du cas ordinaire devient une erreur aléatoire, et peut être traitée statistiquement en répétant les mesurages.

effectuée dans cette section. Pour cela, il faut placer l’objet non pas au 0 de la règle mais à une position initiale aléatoire (figure A.2).

Pour exploiter un tel protocole de mesure, on propose le modèle suivant. On note a_i la longueur entre le 0 de la règle et la position initiale de l’objet sur la règle, et b_i la longueur séparant le 0 de la règle et la position de l’extrémité opposée de l’objet sur la règle (voir figure A.2). L’indice i rappelle que ces longueurs dépendent du mesurage effectué, puisque le positionnement relatif de l’objet et de la règle est modifié à chaque instance. On note λ la « valeur vraie » de la longueur de l’objet, et on a :

$$\lambda = b_i - a_i \quad (\text{A.29})$$

La longueur λ est un nombre réel que l’on peut décomposer comme un entier (sa partie entière $\lfloor \lambda \rfloor$) et un résidu contenu entre 0 et 1 (sa partie fractionnaire $\{\lambda\}$) :

$$\lambda = \lfloor \lambda \rfloor + \{\lambda\} \quad \text{avec} \quad \lfloor \lambda \rfloor \in \mathbb{N} \quad \text{et} \quad 0 < \{\lambda\} < 1 \quad (\text{A.30})$$

Il en est de même pour la position a_i :

$$a_i = \lfloor a_i \rfloor + \{a_i\} \quad \text{avec} \quad \lfloor a_i \rfloor \in \mathbb{N} \quad \text{et} \quad 0 < \{a_i\} < 1 \quad (\text{A.31})$$

On repère $\{a_i\}$ par l’indice i car cette valeur varie à chaque mesurage, puisque la position relative de l’objet et de la règle est tirée aléatoirement.

Une fois ces éléments de modélisation construits, on peut alors conclure sur la façon de mesurer et calculer ℓ . Quelle que soit la position relative de l’objet et de la règle, on compte toujours *au moins* $\lfloor \lambda \rfloor$ graduations séparant la position a de la position b . En fonction des valeurs de $\{a_i\}$ et de $\{\lambda\}$, on comptera parfois, *selon les cas*, une graduation supplémentaire. On distingue donc deux cas :

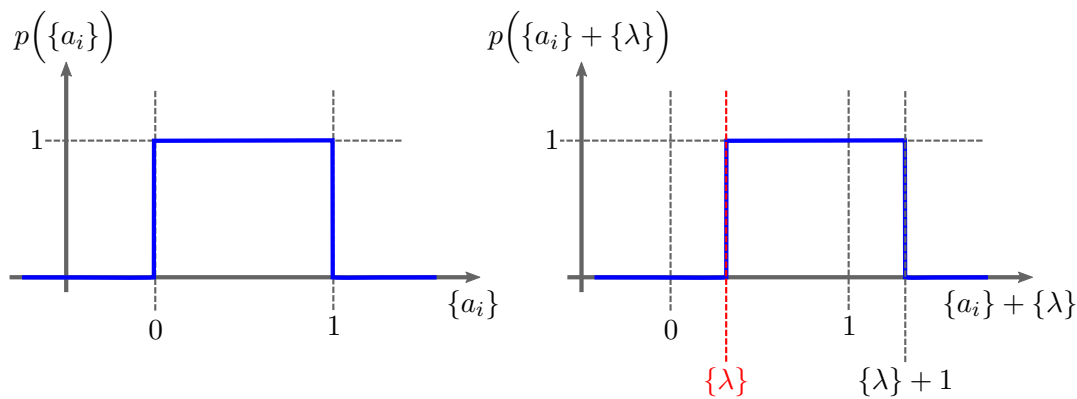


FIGURE A.3 – Distribution de probabilité de la partie fractionnaire de a_i (à gauche) et de la somme des parties fractionnaires respectives de a_i et λ (à droite). La distribution de $\{a_i\}$ est uniforme car le processus de tirage aléatoire de la position respective de l'objet et de la règle ne favorise aucune valeur. De plus $\{a_i\}$ est par définition contenue entre 0 et 1. On a par conséquent une distribution rectangulaire. La valeur $\{\lambda\}$ est fixe; la distribution de probabilité de $\{a_i\} + \{\lambda\}$ s'obtient donc par une simple translation de la distribution précédente. On peut alors calculer les probabilités respectives des cas (1) et (2) en considérant les aires contenues sous la courbe. On est dans le cas (1) si $\{a_i\} + \{\lambda\} < 1$; la probabilité correspondante est l'aire contenue sous la courbe, à gauche de la droite verticale d'abscisse 1, et vaut donc $p_1 = 1 - \{\lambda\}$. On est dans le cas (2) si $\{a_i\} + \{\lambda\} > 1$; la probabilité correspondante est l'aire contenue sous la courbe, à droite de la droite verticale d'abscisse 1, et vaut donc $p_2 = \{\lambda\}$.

- (1) Si $\{a_i\} + \{\lambda\} < 1$, on ne compte pas d'intervalle supplémentaire. Dans ce cas, on conclut, pour ce mesurage précis, que $\ell_i = \lfloor \lambda \rfloor$.
- (2) Si $\{a_i\} + \{\lambda\} > 1$, on compte un intervalle supplémentaire. Dans ce cas, on conclut, pour ce mesurage précis, que $\ell_i = \lfloor \lambda \rfloor + 1$.

Puisque la position relative de l'objet par rapport à la règle a été tirée aléatoirement, cela signifie que a , ou encore $\{a_i\}$, est le résultat d'un tirage aléatoire. Pour $\{a_i\}$, il est raisonnable de considérer que la distribution parente est uniformément distribuée pour toutes les valeurs possibles, de 0 à 1 (figure A.3, graphe de gauche). Notons qu'il s'agit là d'une distribution *fréquentiste* de probabilité. Dans ce cas, on peut également attribuer une probabilité aux deux cas (1) et (2). Pour le cas (1), la probabilité est de $p_1 = 1 - \{\lambda\}$; pour le cas (2), elle est de $p_2 = \{\lambda\}$ (voir figure A.3, graphe de droite). Si l'on prend en fin de compte pour valeur de ℓ la valeur moyenne des résultats des mesurages répétés :

$$\ell = \frac{1}{n} \sum_{i=1}^n \ell_i \quad (\text{A.32})$$

On peut alors calculer l'espérance statistique de ℓ :

$$E = p_1 \lfloor \lambda \rfloor + p_2 (\lfloor \lambda \rfloor + 1) = \lfloor \lambda \rfloor + \{\lambda\} = \lambda \quad (\text{A.33})$$

On constate ainsi que si l'on répète la mesure un grand nombre de fois, la moyenne arithmétique des résultats des mesurage, pourtant limités individuellement par une erreur de résolution, tendra vers une valeur exactement égale à la « valeur vraie » de la longueur de l'objet mesurée – si l'on suppose par ailleurs que la règle est parfaite et qu'il n'y a pas d'autre source d'erreur de mesure, ce qui est bien entendu une situation hautement idéale.

On est donc ramené à une méthode de « type A » de l'erreur aléatoire, et l'incertitude-type peut être calculée à partir de la variance empirique de la collection de données, comme décrit au chapitre 3. Notons ici que, de façon remarquable, la distribution des erreurs aléatoires n'est pas gaussienne.

Conclusions. La méthode de randomisation des erreurs systématiques illustre le fait qu'il est parfois possible de traiter statistiquement des erreurs qui semblent échapper, par leur nature, à l'analyse probabiliste. Cependant, cette méthode peut se révéler très coûteuse. Dans l'exemple cité ici, on ne voit guère le bénéfice que l'on peut retrouver à effectuer une procédure qui se révèle très laborieuse. Il est plus rapide, et vraisemblablement aussi fiable, de perfectionner la méthode classique en interpolant la position de l'extrémité de la règle à l'intérieur des graduations, ce qui permet de gagner un facteur 5 voire 10 en précision. C'est pourquoi le GUM rappelle que :

On doit avoir en mémoire qu'une évaluation de Type B d'incertitude-type peut être aussi fiable qu'une évaluation de Type A, notamment dans une situation de mesure où une évaluation de Type A est fondée sur un nombre relativement faible d'observations statistiquement indépendantes.¹⁸

L'idée générale de « randomisation », cependant, reste valable, et rappelle que la conception du dispositif de mesure, située en amont de la mesure, est au moins aussi importante que l'analyse statistique, située en aval, pour la bonne estimation des incertitudes de mesure. Dans tous les cas, l'incertitude ne peut être réduite en-deçà d'un certain seuil, en raison de la présence d'autres erreurs de mesure, par exemple les erreurs d'étalonnage de la règle.

18. [Joint Committee for Guides in Metrology \(JCGM\) \(2008d\)](#), p.12.

A.5 Combinaison de plusieurs résultats

Soit un mesurande X dont la mesure a été effectuée selon deux méthodes différentes (ou par deux laboratoires différents), et qui a abouti à deux estimations accompagnées de deux « incertitudes-types »¹⁹ associées respectives :

$$\begin{cases} x_1 = x_{10} \pm u_1 \\ x_2 = x_{20} \pm u_2 \end{cases} \quad (\text{A.34})$$

Nous avons vu à la partie I que, dans l'approche fréquentiste comme dans l'approche bayésienne de la mesure, sont associées à x_1 et x_2 des variables aléatoires centrées sur la valeur estimée, et dont l'incertitude-type constitue l'écart-type. Notons \hat{x}_1 et \hat{x}_2 ces variables aléatoires. Combiner les deux résultats, c'est construire une estimation x du mesurande X par une combinaison linéaire de x_1 et x_2 :

$$x = \lambda x_1 + (1 - \lambda) x_2 \quad (\text{A.35})$$

Les coefficients λ et $1 - \lambda$ sont deux termes de pondération, ainsi construits car le poids total attribué aux différents résultats combinés vaut nécessairement 1.

Si l'on « mélange » ainsi deux résultats, que vaut alors l'incertitude associée à la résultante obtenue ? Pour la déterminer, il faut évaluer l'écart-type de la variable aléatoire \hat{x} associée à x . Cette variable aléatoire est définie de la même façon que x :

$$\hat{x} = \lambda \hat{x}_1 + (1 - \lambda) \hat{x}_2 \quad (\text{A.36})$$

Si nous faisons l'hypothèse que les deux mesures sont totalement indépendantes l'une de l'autre, c'est-à-dire, en termes statistiques, qu'il n'y a pas de terme de covariance entre les deux variables aléatoires, alors l'incertitude-type résultante u est déterminée par la formule classique de propagation :

$$u(\lambda) = \lambda^2 u_1^2 + (1 - \lambda)^2 u_2^2 \quad (\text{A.37})$$

L'opération de combinaison des deux résultats est bénéfique si l'incertitude finale u est plus faible que les deux incertitudes u_1 et u_2 . Elle est optimale si le choix du paramètre de pondération λ la rend minimale. On peut étudier la dépendance de u en fonction de λ en calculant sa dérivée :

$$\frac{du}{d\lambda} = 2\lambda u_1^2 - 2(1 - \lambda) u_2^2 \quad (\text{A.38})$$

Cette dérivée s'annule lorsque

$$\lambda = \frac{u_2^2}{u_1^2 + u_2^2} = \frac{1/u_1^2}{1/u_1^2 + 1/u_2^2} \quad (\text{A.39})$$

19. Nous renvoyons à la partie I pour une explicitation des termes techniques des modèles probabilistes de la mesure.

Ce dernier résultat montre que la combinaison linéaire qui permet de minimiser la variance de \hat{x} est celle pour laquelle les pondérations respectives des différentes contributions sont identifiées aux inverses des carrés des incertitudes de chaque contribution :

$$\lambda \propto \frac{1}{u_1^2} \quad \text{et} \quad (1 - \lambda) \propto \frac{1}{u_2^2} \quad (\text{A.40})$$

Dans ce cas, l'incertitude-type finale est :

$$u_{\min} = \left(\frac{1}{1/u_1^2 + 1/u_2^2} \right)^2 \left(\frac{u_1^2}{u_1^4} + \frac{u_2^2}{u_2^4} \right) = \frac{1}{1/u_1^2 + 1/u_2^2} \quad (\text{A.41})$$

C'est-à-dire encore que :

$$\frac{1}{u^2} = \frac{1}{u_1^2} + \frac{1}{u_2^2} \quad (\text{A.42})$$

Les résultats précédents sont immédiatement extensibles à la combinaison d'un nombre quelconque n de valeurs expérimentales :

$$x = \sum_{i=1}^n \frac{x_i}{u_i^2} \quad \text{et} \quad \frac{1}{u^2} = \sum_{i=1}^n \frac{1}{u_i^2} \quad (\text{A.43})$$

On voit immédiatement, à la lumière de ce dernier résultat, que $(1/u^2)$ est nécessairement supérieur à chaque $(1/u_i^2)$, c'est-à-dire encore que la variance combinée u est toujours inférieure à chacune des u_i . La méthode de combinaison linéaire décrite ci-dessus garantit donc (à l'intérieur du modèle mathématique) que l'incertitude-type résultante est toujours inférieure aux incertitudes-types associées aux résultats d'origine. Cela revient à dire que si les poids attribués aux différentes résultats de mesure sont correctement choisis, la combinaison de ces résultats est toujours bénéfique : il est possible de combiner les résultats pour obtenir une valeur finale avec une incertitude plus faible. Autrement dit, les informations contenues par chacun des résultats d'origine peuvent être mélangées : il y a *sommation* des informations, au lieu de les utiliser indépendamment les unes des autres. Cette méthode de pondération consiste en la meilleure façon de réunir les informations contenues dans des résultats de mesure dont la précision est différente. En particulier, ce constat mathématique soutient une idée remarquable : même lorsque l'un des résultats est significativement plus précis qu'un autre, il peut en principe être enrichi par un autre résultat de moindre précision – si l'on s'en tient au seul raisonnement mathématique sur l'incertitude de mesure que nous explicitons ici. Ce que l'on pourrait ainsi dénommer principe de « non contamination », en reprenant à notre compte l'expression de Stigler, n'est garanti qu'à la condition d'utiliser les pondérations adéquates.

L'on comprend, à la lumière de ce principe, pourquoi Mayer et Euler n'étaient pas confrontés à la même situation (voir chapitre 8). Mayer avait une bonne idée de la précision de ses résultats, et il pouvait donc les combiner de façon bénéfique (l'on dirait aujourd'hui qu'il avait une bonne connaissance des coefficients de pondération à attribuer à chacune de ses mesures). Mais Euler, qui n'avait pas effectué les mesures lui-même, ne pouvait pas savoir quelle était la précision présumée des observations qu'il mettait à contribution. De ce fait, il pouvait légitimement pressentir qu'en attribuant à chaque observation un même coefficient, il risquait de donner trop de poids aux moins bonnes observations. En termes actuels, il lui manquait une évaluation de l'incertitude sur ses observations.

Bibliographie

- ACHINSTEIN, P. (1968), *Concepts of Science : A Philosophical Analysis*, Concepts of Science, Johns Hopkins University Press.
- ACKERMANN, R. (1989), The New Experimentalism : the Neglect of Experiment by Allan Franklin, *The British Society for the Philosophy of Science*, 40(2), pp.185–190.
- ADAM, T. ET AL. (2011), Measurement of the neutrino velocity with the OPERA detector in the CNGS beam, *ArXiv*, pp.1–24.
- ADAM, T. ET AL. (2013), Measurement of the neutrino velocity with the OPERA detector in the CNGS beam using the 2012 dedicated data, *Journal of High Energy Physics*, 2013(1), pp.1–14.
- AFP (2012), Démission à la tête de l'expérience « Opéra », sur la vitesse des neutrinos, *lemonde.fr*. 30.03.2012, mis à jour le 01.10.2012. http://www.lemonde.fr/sciences/article/2012/03/30/vitesse-des-neutrinos-demission-a-la-tete-de-l-experience_1678329_1650684.html
- AMBLER, E. (1977), Lettre de E. Ambler à J. Terrien, 25 août 1977, non publiée. Archives du National Institute of Standard and Technology, United States Department of Commerce.
- ARMATTE, M. (2004), La théorie des erreurs (1750-1820), enjeux, problématiques, résultats, in Évelyne BARBIN et LAMARCHE, J.-P. (dir.), *Histoires de probabilités et de statistiques*, pp.141–160, Paris : Ellipses.
- ARMATTE, M. (2010), Statut de la Dispersion : de l'erreur à la variabilité, *Journal Électronique d'Histoire des Probabilités et de la Statistique*, 6(1), pp.1–20.
- ARMATTE, M. (2011), L'observation scientifique est-elle un témoignage ?, *Journal Électronique d'Histoire des Probabilités et de la Statistique*, 7(1), pp.1–38.
- BARATTO, A. C. (2008), Measurand : a cornerstone concept in metrology, *Metrologia*, 45(3), pp.299–307.
- BARBEROUSSE, A. (2008), La valeur de la connaissance approchée. L'épistémologie de l'approximation d'Émile Borel, *Revue d'histoire des mathématiques*, 14(1), pp.53–55.
- BARBEROUSSE, A., KISTLER, M. et LUDWIG, P. (2000), *La philosophie des sciences au XXe siècle*, Champs Flammarion.

- BARSALOU, M. A. (2014), *Root Cause Analysis : A Step-By-Step Guide to Using the Right Tool at the Right Time*, Taylor & Francis.
- BAYES, T. (1763), An essay towards solving a problem in the doctrine of chances, *Phil. Trans. of the Royal Soc. of London*, 53, pp.370–418.
- BEARDEN, J. A. et THOMSEN, J. S. (1957), A Survey of Atomic Constants, *Nuovo Cimento Supplemento*, 5, pp.267–360.
- BEARDEN, J. A. et WATTS, H. M. (1951), A Re-Evaluation of the Fundamental Atomic Constants, *Physical Review*, 81(1), pp.73–81.
- BENDER, P. L. (1971), Handling of Discrepant Data in Evaluations of the Fundamental Constants, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.493–494, National Bureau of Standards. (Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3–7, 1970).
- BERNAUER, J. et POHL, R. (2014), Le proton, un problème de taille, *Pour la science*, 439, pp.28–35.
- BICH, W. (2012a), From Errors to Probability Density Functions. Evolution of the Concept of Measurement Uncertainty, *IEEE Transactions on Instrumentation and Measurement*, 61(8), pp.2153–2159.
- BICH, W. (2012b), Report on the GUM Online Survey (conducted 2 February – 15 June, 2012), Rapport technique, Joint Committee for Guides in Metrology (JCGM). Document disponible en ligne sur le site du BIPM : http://www.bipm.org/wg/JCGM/JCGM-WG1/Allowed/sub-committee_5/WG1-SC5-N12-14b_GUM_survey_report.pdf.
- BICH, W. (2013), entretien avec l’auteur (non publié), 8 novembre 2013.
- BICH, W. et KOOL, W. (2012), Evolution of the concept of measurement uncertainty, *OIML Bulletin*, LIII(2), pp.6–11.
- BIENVENU, A. (2007), *Un empirisme risqué : la philosophie des probabilités de Hans Reichenbach*, thèse de doctorat, Université Paris 1 Panthéon-Sorbonne.
- BIRGE, R. T. (1919), The Most Probable Value of the Planck Constant h , *Physical Review*, 14, pp.361–368.
- BIRGE, R. T. (1929a), The electronic charge e , *Nature*, 123(3096), p.318.
- BIRGE, R. T. (1929b), Probable Values of the General Physical Constants, *Reviews of Modern Physics*, 1(1), pp.1–73.
- BIRGE, R. T. (1932a), The Calculation Of Errors By The Method Of Least Squares, *Physical Review*, 40(2), pp.207–227.
- BIRGE, R. T. (1932b), Probable Values of e , h , e/m and α , *Phys. Rev.*, 40(2), pp.207–228.

- BIRGE, R. T. et WEINBERG, J. W. (1947), Least-Squares' Fitting of Data by Means of Polynomials, *Reviews of Modern Physics*, 19, pp.298–360.
- BOGEN, J. et WOODWARD, J. (1988), Saving the Phenomena, *The Philosophical Review*, XCVII(3), pp.303–352.
- BOND, W. N. (1930), The values and inter-relationships of c , e , h , M_p , m_o , G , and R , *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(67), pp.994–1003.
- BOND, W. N. (1931), The electronic charge, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 12(78), pp.632–640.
- BOUMANS, M., HON, G. et PETERSEN, A. C. (dir.) (2014), *Error and Uncertainty in Scientific Practice*, History and Philosophy of Technoscience Series, Pickering & Chatto.
- BOURGUET, M., LICOPPE, C. et SIBUM, H. (dir.) (2002), *Instruments, Travel and Science : Itineraries of Precision from the Seventeenth to the Twentieth Century*, Routledge studies in the history of science, technology and medicine, Routledge.
- BOX, G. E. et DRAPER, N. R. (1987), *Empirical model-building and response surfaces*, Wiley series in probability and mathematical statistics : Applied probability and statistics, Wiley.
- BOYD, R. (1990), Realism, Approximate Truth, and Philosophical Method, in SAVAGE, C. W. (dir.), *Scientific Theories*, pp.355–391, University of Minnesota Press.
- BOYD, R. (2010), Scientific realism, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, summer 2010.
- BRANSCOMB, L. M. (1971), Measurement Standards, Language of Discovery, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.3–8, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- BRENNER, A. (1990), *Duhem : science, réalité et apparence : la relation entre philosophie et histoire dans l'œuvre de Pierre Duhem*, Mathesis (Librairie philosophique J. Vrin), Librairie philosophique J. Vrin.
- BULLYNCK, M. (2014), À la recherche des moindres carrés, ou les distinctions de l'erreur, exposé lors de l'assemblée générale du laboratoire SPHERE, 5 décembre 2014, Université Paris Diderot.
- BUREAU INTERNATIONAL DES POIDS ET MESURES (1980), *Report on the BIPM enquiry on error statements*, Sèvres : BIPM.
- BUREAU INTERNATIONAL DES POIDS ET MESURES (2006), *The International System of Units (SI)*. Paris, 8e édition.
- BURNS, J. E., CAMPION, P. J. et WILLIAMS, A. (1973), Error and Uncertainty, *Metrologia*, 6(2), pp.101–102.

- CAMPBELL, N. R. (1920), *Physics, the Elements*, Cambridge : Cambridge University Press.
- CAMPBELL, N. R. (1921), *What is science ?*, New York : Dover.
- CAMPBELL, N. R. (1928), *An account of the principles of measurement and calculation*, New York : Longmans, Green and Company.
- CAMPION, P. J., BURNS, J. E. et WILLIAMS, A. (1973), *A code of practice for the detailed statement of accuracy*, National Physical Laboratory.
- CARNAP, R. (1936), Testability and Meaning, *Philosophy of Science*, 3(4), pp.419–471.
- CARTWRIGHT, N. (1983), *How The Laws of Physics Lie*, Oxford : Oxford University Press.
- CARTWRIGHT, N. (1999), *The Dappled World : A Study of the Boundaries of Science*, Cambridge : Cambridge University Press.
- CERN (2011), L'expérience OPERA annonce une anomalie dans le temps de vol des neutrinos allant du CERN au Gran Sasso, *Communiqué de presse du CERN*. 23 septembre 2011. <http://press.web.cern.ch/fr/press-releases/2011/09/lexperience-opera-annonce-une-anomalie-dans-le-temps-de-vol-des-neutrinos>
- CHAKRAVARTTY, A. (2015), Scientific realism, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, fall 2015. Entrée consultée pour la dernière fois le 8 février 2016.
- CHANG, H. (2004), *Inventing Temperature : Measurement and Scientific Progress*, Oxford : Oxford University Press.
- CHANG, H. et CARTWRIGHT, N. (2008), Measurement, in CURD, M. et PSILLOS, S. (dir.), *The Routledge Companion to Philosophy of Science*, pp.367–375, New York : Routledge.
- CLIFFORD, P. M. (1985), The international vocabulary of basic and general terms in Metrology, *Measurement*, 3(2), pp.72–76.
- CNRS (2011), Plus vite que la lumière ?, *Communiqué de presse du CNRS*. 22 SEPTEMBRE 2011. http://www.dr7.cnrs.fr/IMG/pdf/CP_neutrino.pdf
- COHEN, E. R. et DUMOND, J. W. M. (1965), Our Knowledge of the Fundamental Constants of Physics and Chemistry in 1965, *Reviews of Modern Physics*, 37(4), pp.537–594.
- COHEN, E. R., DUMOND, J. W. M., LAYTON, T. W. et ROLLETT, J. S. (1955), Analysis of Variance of the 1952 Data on the Atomic Constants and a New Adjustment, 1955, *Reviews of Modern Physics*, 27(4), pp.363–380.
- COHEN, E. R. et TAYLOR, B. N. (1973), The 1973 Least-Squares Adjustment of the Fundamental Constants, *J. Phys. Chem. Ref. Data*, 2(4), pp.634–734.
- COHEN, E. R. et TAYLOR, B. N. (1987), The 1986 adjustment of the fundamental physical constants, *Rev. Mod. Phys.*, 59(4), pp.1121–1148.

- COHEN-TANNOUJJI, G. (1998), *Les constantes universelles*, Hachette.
- COLCLOUGH, A. (1987), Two Theories of Experimental Error, *Journal of Research of the National Bureau of Standards*, 92(3), pp.167–185.
- COMITÉ INTERNATIONAL DES POIDS ET MESURES (dir.) (1977), *Procès-verbaux des séances, 66^e session*, volume45, Pavillon de Breteuil, F-92310, Sèvres, France. Comité International des Poids et Mesures.
- COMITÉ INTERNATIONAL DES POIDS ET MESURES (dir.) (1978), *Procès-verbaux des séances, 67^e session*, volume46, Pavillon de Breteuil, F-92310, Sèvres, France. Comité International des Poids et Mesures.
- COMITÉ INTERNATIONAL DES POIDS ET MESURES (dir.) (1981), *Procès-verbaux des séances, 70^e session*, volume49, Pavillon de Breteuil, F-92310, Sèvres, France. Comité International des Poids et Mesures.
- COMITÉ INTERNATIONAL DES POIDS ET MESURES (dir.) (1984), *Procès-verbaux de la 73^e session*, volume52, Pavillon de Breteuil, F-92310, Sèvres, France. Comité International des Poids et Mesures.
- COMITÉ INTERNATIONAL DES POIDS ET MESURES (dir.) (1985), *Procès-verbaux de la 74^e session*, volume53, Pavillon de Breteuil, F-92310, Sèvres, France. Comité International des Poids et Mesures.
- COMITÉ INTERNATIONAL DES POIDS ET MESURES (dir.) (1986), *Procès-verbaux de la 75^e session*, volume54, Pavillon de Breteuil, F-92310, Sèvres, France. Comité International des Poids et Mesures.
- de COURTENAY, N. (2008), Mesure et formation des concepts physiques, Rudolph Carnap et Norman Campbell, in BOUVERESSE, J. et WAGNER, P. (dir.), *Mathématiques et expérience. L'empirisme logique à l'épreuve (1918-1940)*, pp.213–251, Paris : Odile Jacob.
- de COURTENAY, N. (2015), The Double Interpretation of the Equations of Physics and the Quest for Common Meanings, in SCHLAUDT, O. et HUBER, L. (dir.), *Standardization In Measurement : Philosophical, Historical And Sociological Issues*, pp.53–68, Pickering & Chatto.
- COX, M. et HARRIS, P. (2014), GUM anniversary issue, *Metrologia*, 51(4), p.S141.
- D'AGOSTINI, G. (1996), A Theory of Measurement Uncertainty Based on Conditional Probability, *ArXiv Physics e-prints*.
- D'AGOSTINI, G. (2003), *Bayesian Reasoning in Data Analysis : A Critical Introduction*, World Scientific.
- DARRIGOL, O. (2003), Number and measure : Hermann von Helmholtz at the crossroads of mathematics, physics, and psychology, *Studies in History and Philosophy of Science*, 34, pp.515–573.

- DARRIGOL, O. (2007), The modular structure of physical theories, *Synthese*, 162(2), pp.195–223.
- DASTON, L. (1989), L'interprétation classique du calcul des probabilités, *Annales. Économies, Sociétés, Civilisations*, 44(3), pp.715–731.
- DASTON, L. (1995), *Classical probability in the enlightenment*, Princeton paperbacks, Princeton University Press.
- DE BIÈVRE, P. (2007), The identification of the measurand can have an effect on the magnitude of the measurement uncertainty, *Accreditation and Quality Assurance*, 12(12), pp.613–614.
- DE BIÈVRE, P. (2008), Essential for metrology in chemistry, but not yet achieved : truly internationally understood concepts and associated terms, *Metrologia*, 45(3), pp.335–341.
- DEMEYER, S. (2011), *Approche bayésienne de l'évaluation de l'incertitude de mesure : application aux comparaisons interlaboratoires*, thèse de doctorat, Conservatoire National des Arts et Métiers.
- DEMING, W. E. et BIRGE, R. T. (1934), On the Statistical Theory of Errors, *Reviews of Modern Physics*, 6(3), pp.119–130.
- DÉSENFANT, M. et PRIEL, M. (2006), Road map for measurement uncertainty evaluation, *Measurement*, 39, pp.841–848.
- DROUET, I. (2016), Le bayésianisme : éléments de définition et mutations récentes, in DROUET, I. (dir.), *Le bayésianisme aujourd'hui. Fondements et pratiques*, pp.3–25, Éditions Matériologiques.
- DUHEM, P. (1906), *La théorie physique, son objet, sa structure*, Paris : Chevalier et Rivière.
- DUMOND, J. W. M. et COHEN, E. R. (1948), Our Knowledge of the Atomic Constants F , N , m and h in 1947, and of Other Constants Derivable Therefrom, *Reviews of Modern Physics*, 20(1), pp.82–108.
- DUMOND, J. W. M. et COHEN, E. R. (1953), Least-Squares Adjustment of the Atomic Constants, 1952, *Reviews of Modern Physics*, 25(3), pp.691–708.
- DUNNINGTON, F. G. (1939), The Atomic Constants. A Revaluation and an Analysis of the Discrepancy, *Reviews of Modern Physics*, 11(2), pp.65–83.
- EHRlich, C. (2014), Terminological aspects of the *Guide to the Expression of Uncertainty in Measurement* (GUM), *Metrologia*, 51(4), pp.S145–S154.
- EHRlich, C. et DYBKAER, R. (2012), Uncertainty of error : the error dilemma, *OIML Bulletin*, LIII(2), pp.12–17.
- EHRlich, C., DYBKAER, R. et WÖGER, W. (2007), Evolution of philosophy and description of measurement (preliminary rationale for VIM3), *Accreditation and Quality Assurance*, 12, pp.201–218.

- EISENHART, C. (1963), Realistic Evaluation of the Precision and Accuracy of Instruments Calibration Systems, *Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation*, 67C(2), pp.161–187.
- EISENHART, C. (1971), Contribution to Panel Discussion on Adjustments of the Fundamental Constants, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.509–518, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- ELLISON, S. et WILLIAMS, A. (dir.) (2012), *EURACHEM / CITAC Guide CG 4 - Quantifying Uncertainty in Analytical Measurement - Third Edition*, EURACHEM/CITAC.
- ELSTER, C. (2014), Bayesian uncertainty analysis compared with the application of the GUM and its supplements, *Metrologia*, 51(4), pp.S159–S166.
- ELSTER, C., WÖGER, W. et COX, M. G. (2007), Draft GUM Supplement 1 and Bayesian analysis, *Metrologia*, 44(3), pp.L31–L32.
- ESTLER, W. T. (1999), Measurement as Inference : Fundamental Ideas, *Annals of the CIRP*, 48(2), pp.611–631.
- EULER, L. (1749), *Recherches sur la question des inégalités du mouvement de Saturne et de Jupiter, sujet proposé pour le prix de l'année 1748, par l'Académie royale des sciences de Paris*.
- FEYERABEND, P. (1975), *Against Method : Outline of an Anarchist Theory of Knowledge*, New Left Books.
- FRANKEN, P. (1971), Comments on the Assignment of Experimental Uncertainties, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.507–508, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- FRANKLIN, A. (1989), *The Neglect of Experiment*, Cambridge University Press.
- FRANKLIN, A. (1990), *Experiment, Right or Wrong*, Cambridge University Press.
- FRANKLIN, A. et PEROVIC, S. (2015), Experiment in physics, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, summer 2015. Dernière consultation le 1er février 2016.
- FRIGG, R. et HARTMANN, S. (2012), Models in science, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, spring 2012 edition. Entrée consultée le 27 octobre 2015.
- GABRIELSE, G. et HANNEKE, D. (2006), Precision pins down the electron's magnetism, *CERN Courier*.
- GALISON, P. (1987), *How Experiments End*, University of Chicago Press.

- GALISON, P. (1997), *Image and Logic*, University of Chicago Press.
- GALISON, P. (2002), *Ainsi s'achèvent les expériences*, La Découverte. Traduit de l'anglais par Bertrand Nicquevert (*How Experiments End*, 1987).
- GAUSS, C. F. (1823), *Theoria combinationis observationum erroribus minimis obnoxiae*, Henricum Dieterich.
- GIACOMO, P. (1981), News from the BIPM, *Metrologia*, 17(2), pp.69–74.
- GIACOMO, P. (1982), News from the BIPM, *Metrologia*, 18(1), pp.41–44.
- GILLIES, D. (2000), *Philosophical Theories of Probability*, Routledge.
- GIORDANI, A. et MARI, L. (2011), Quantity and Quantity Value, in *Proceedings of the TC1-TC7-TC13 14th IMEKO Joint Symposium*. IMEKO.
- GIORDANI, A. et MARI, L. (2012), Measurement, Models, and Uncertainty, *IEEE transactions on instrumentation and measurement*, 61(8), pp.2144–2152.
- GIÉ, H. et MOREAU, R. (1987), Le calcul des incertitudes, *Bulletin d'Union des Physiciens*, 691(1), pp.159–209.
- GLANZBERG, M. (2013), Truth, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, fall 2014 edition. Entrée consultée le 24 février 2015.
- GRABE, M. (1987), Principles of “metrological statistics”, *Metrologia*, 23(4), pp.213–219.
- GRABE, M. (2007), Ten Theses for a New GUM (Poster Paper), in *proceedings of PTB-BIPM Workshop on the Impact of Information Technology in Metrology, 5-7 June 2007, Berlin*.
- GRÉGIS, F. et de COURTENAY, N. (2016), Incertitude de mesure et probabilités : la confrontation des approches fréquentiste et bayésienne en métrologie, in DROUET, I. (dir.), *Le bayésianisme aujourd'hui. Fondements et pratiques*, Éditions Matériologiques. À paraître.
- HACKING, I. (1975), *The Emergence of Probability : a Philosophical Study of Early Ideas About Probability Induction and Statistical Inference*, Cambridge University Press.
- HACKING, I. (1983), *Representing and Intervening*, Cambridge University Press.
- HACKING, I. (1992), Statistical Language, Statistical Truth and Statistical Reason : The Self-Authentication of a Style of Scientific Reasoning, in McMULLIN, E. (dir.), *The Social Dimension of Science*, pp.130–157, Notre Dame : University of Notre Dame Press.
- HÁJEK, A. (2012), Interpretations of probability, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, winter 2012 edition.
- HALL, B. D. (2008), Evaluating methods of calculating measurement uncertainty, *Metrologia*, 45(2), pp.L5–L8.

- HANSON, N. (1961), *Patterns of Discovery*, Cambridge University Press.
- HARDWIG, J. (1985), Epistemic Dependence, *The Journal of Philosophy*, 82(7), pp.335–349.
- HARRÉ, R. (2003), The Materiality of Instruments in a Metaphysics for Experiments, in RADDER, H. (dir.), *The Philosophy of Scientific Experimentation*, pp.19–38, University of Pittsburgh Press.
- HEIDELBERGER, M. (2003), Theory-Ladenness and Scientific Instruments in Experimentation, in RADDER, H. (dir.), *The Philosophy of Scientific Experimentation*, pp.138–151, University of Pittsburgh Press.
- HELMHOLZ, A. C. (1980), Obituary : Raymond Thayer Birge, *Physics Today*, 33(8), pp.68–69.
- HELMHOLZ, A. C. (1990), *Raymond Thayer Birge 1887-1980, A Biographical Memoir by A. Carl Helmholtz*, Washington D.C., National Academy of Sciences.
- HIMBERT, M. (2001), Unités, références, incertitudes pour les mesures, in *La pluridisciplinarité dans les enseignements scientifiques - Tome 2 : La place de l'expérience*. Ministère de la Jeunesse, de l'Éducation nationale et de la Recherche. Actes de l'université d'été, du 9 au 13 juillet 2001, Cachan.
- HON, G. (1987), On Kepler's Awareness of the Problem of Experimental Error, *Annals of Science*, 44, p.1987.
- HON, G. (1989a), Is there a Concept of Experimental Error in Greek Astronomy?, *The British Journal for the History of Science*, 22(2), pp.129–150.
- HON, G. (1989b), Towards a typology of experimental errors : an epistemological view, *Studies in History and Philosophy of Science*, 20(4), pp.469–504.
- HON, G. (1998), Exploiting Errors, *Studies in History and Philosophy of Science*, 29(3), pp.465–479.
- HON, G. (2003), The Idols of Experiment. Transcending the “etc. List”, in RADDER, H. (dir.), *The Philosophy of Scientific Experimentation*, pp.174–197, University of Pittsburgh Press.
- HON, G. (2004), Putting Error to (Historical) Work : Error as a Tell-tale in the Studies of Kepler and Galileo, *Centaurus*, 46, pp.58–81.
- HON, G., SCHICKORE, J. et STEINLE, F. (dir.) (2009), *Going Amiss in Experimental Research*, Springer Science & Business Media.
- HOWSON, C. et URBACH, P. (1993), *Scientific reasoning : The Bayesian approach*, Open Court, Chicago.
- HUMPHREYS, P. (2004), *Extending Ourselves : Computational Science, Empiricism, and Scientific Method*, Oxford University Press.
- ICHIKAWA, J. J. et STEUP, M. (2014), The analysis of knowledge, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, spring 2014.

- IMBERT, C. (2008), *L'opacité intrinsèque de la Nature. Théories connues, phénomènes difficiles à expliquer et limites de la science.*, thèse de doctorat, Université de Paris 1 Panthéon-Sorbonne.
- ISHIKAWA, K. (1991), *Guide to Quality Control*, Asian Productivity Organization. Traduit par l'Organisation Asiatique de Productivité (Asian Productivity Organization). 2^e édition.
- JANICH, P. (1978), Physics – Natural Science or Technology ?, in KROHN, W., LAYTON, E. et WEINGART, P. (dir.), *The Dynamics of Science and Technology : Social Values, Technical Norms and Scientific Criteria in the Development of Knowledge*, pp.3–27, Springer Netherlands.
- JAYNES, E. T. (1957), Information Theory and Statistical Mechanics, *Physical Review*, 106, pp.620–630.
- JAYNES, E. T. (1968), Prior Probabilities, *IEEE Transactions on Systems and Cybernetics*, (3), pp.227–241.
- JAYNES, E. T. et BRETTHORST, G. (2003), *Probability Theory : The Logic of Science*, Cambridge University Press.
- JEFFREY, R. (1983), *The Logic of Decision*, University of Chicago Press.
- JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2008a), *Evaluation of measurement data - Guide to the expression of uncertainty in measurement*, Sèvres : JCGM.
- JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2008b), *Evaluation of measurement data - Supplement 1 to the "Guide to the expression of uncertainty in measurement" - Propagation of distributions using a Monte Carlo method*, Sèvres : JCGM.
- JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2008c), *Vocabulaire international de métrologie - Concepts fondamentaux et généraux et termes associés (VIM)*, Sèvres : JCGM.
- JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2008d), *Évaluation des données de mesure - Guide pour l'expression de l'incertitude de mesure*, Sèvres : JCGM.
- JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2009), *Evaluation of measurement data - An introduction to the "Guide to the expression of uncertainty in measurement" and related documents*, Sèvres : JCGM.
- JOINT COMMITTEE FOR GUIDES IN METROLOGY (JCGM) (2012), *Vocabulaire international de métrologie - Concepts fondamentaux et généraux et termes associés (VIM)*, Sèvres : JCGM. 3^e édition, version 2008 avec corrections mineures.
- KAARLS, R. (1980), *Report of the BIPM Working Group on the statement of measurement uncertainties (1st meeting - 21 to 23 October 1980) to the Comité International des Poids et Mesures*, Bureau International des Poids et Mesures.
- KACKER, R. et JONES, A. (2003), On use of Bayesian statistics to make the *Guide to the Expression of Uncertainty Measurement* consistent, *Metrologia*, 40(5), pp.235–248.

- KANT, E. (2007), *Logique*, Vrin. Traduit de l'allemand par Louis Guillermit (édition originale 1800).
- KARP, P. (1986), *Report of the first meeting of the ISO TAG 4 Working Group 'Uncertainties'. Convened by R. Collé, 1-3 October 1986 at the Bureau International de Métrologie Légale, 11 rue Turgot, 75009 Paris*, Organisation internationale de normalisation (ISO).
- KIRKUP, L. et FRENKEL, B. (2006), *An Introduction to Uncertainty in Measurement*, Cambridge University Press.
- KISTLER, M. (2011), Causalité, in BARBEROUSSE, A., BONNAY, D. et COZIC, M. (dir.), *Précis de philosophie des sciences*, pp.100–139, Vuibert.
- KOLMOGOROV, A. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung (Ergebnisse der Mathematik und Ihrer Grenzgebiete)*, Berlin : Springer.
- KRAGH, H. (2003), Magic Number : A Partial History of the Fine-Structure Constant, *Archive for History of Exact Sciences*, 57(5), pp.395–431.
- KUHN, T. S. (1983), *La structure des révolutions scientifiques*, Champs Flammarion. Traduit de l'anglais (1962) par Laure Meyer.
- KYBURG, H. (1992a), Measuring Errors of Measurement, in *Philosophical and functional issues in measurement theory*, pp.75–91, Hillsdale : Lawrence Elbaum Associates.
- KYBURG, H. (1992b), The Scope of Bayesian Reasoning, *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1992, pp.139–152.
- LANGENBERG, D. N. et TAYLOR, B. N. (dir.) (1971a), Panel Discussion, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.518–525, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- LANGENBERG, D. N. et TAYLOR, B. N. (dir.) (1971b), *Precision Measurement and Fundamental Constants*, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- LAPLACE, P.-S. (1812), *Théorie analytique des probabilités*, Mme Ve Courcier.
- LAROUSSE, D. (2012a), Neutrinos : retour sur une annonce trop rapide, *lemonde.fr*. 13.04.2012 à 21h04 • Mis à jour le 01.10.2012 à 10h00. http://www.lemonde.fr/sciences/article/2012/04/13/neutrinos-retour-sur-une-annonce-trop-rapide_1684719_1650684.html
- LAROUSSE, D. (2012b), Neutrinos supraluminiques : chercher l'erreur, *lemonde.fr*. 27.09.2012 à 16h02 • Mis à jour le 01.10.2012 à 10h01. http://www.lemonde.fr/sciences/article/2012/09/27/supraluminique-chercher-l-erreur_1766930_1650684.html
- LAUDAN, L. (1981), A Confutation of Convergent Realism, *Philosophy of Science*, 48(1), pp.19–49.

- LAUTRUP, B. et ZINKERNAGEL, H. (1999), $g - 2$ and the Trust in Experimental Results, *Studies in History and Philosophy of Modern Physics*, 30(1), pp.85–110.
- LAYMON, R. (1989), Cartwright and the Lying Laws of Physics, *Journal of Philosophy*, 86(7), pp.353–372.
- LAYMON, R. (1995), Experimentation and the Legitimacy of Idealization, *Philosophical Studies*, 77, pp.353–375.
- LECOUTRE, B. (2005), Et si vous étiez un bayésien qui s’ignore?, *MODULAD*, 32, pp.91–105.
- LEGENBRE, A.-M. (1805), *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris, Firmin Didot.
- LÉVY-LEBLOND, J.-M. (1977), On the Conceptual Nature of Physical Constants, *Rivista del Nuovo Cimento*, 7(2), pp.187–214.
- LÉVY-LEBLOND, J.-M. (2000), Mots & maux de la physique quantique : critique épistémologique et problèmes terminologiques, *Revue Internationale de Philosophie*, 54(212 (2)), pp.243–265.
- LÉVY-LEBLOND, J.-M. (2006), Quantique, in LECOURT, D. (dir.), *Dictionnaire d’histoire et philosophie des sciences*, pp.917–921, Paris : PUF.
- LEWINO, F. (2011), Ne détrônons pas trop vite Einstein, *lepoint*. Publié le 23/09/2011 à 13 :31 – Modifié le 24/09/2011 à 11 :34. http://www.lepoint.fr/actu-science/ne-detronons-pas-trop-vite-einstein-23-09-2011-1376688_59.php
- LIDE, D. R. et WOOD, G. H. (2012), *CODATA @ 45 Years : The Story of the ICSU Committee on Data for Science and Technology (CODATA) from 1966 to 2010*, CODATA.
- LIRA, I. (2002), *Evaluating the Measurement Uncertainty : Fundamental and Practical Guidance*, Bristol and Philadelphia, Institute of Physics Publishing.
- LIRA, I. (2008), On the long-run success rate of coverage intervals, *Metrologia*, 45(4), p.L21.
- LIRA, I. (2009), On the meaning of coverage probabilities, *Metrologia*, 46(6), pp.616–618.
- LIRA, I. et WÖGER, W. (2001), Bayesian evaluation of the standard uncertainty and coverage probability in a simple measurement model, *Measurement Science and Technology*, 12(8), pp.1172–1179.
- LIRA, I. et WÖGER, W. (2006), Comparison between the conventional and Bayesian approaches to evaluate measurement data, *Metrologia*, 43(4), pp.S249–S259.
- LOFTUS, P. et GUIDICE, S. (2013), “The assessment of measurement system performance for high-value products – the contribution and limitations of the GUM”, allocution lors de la conférence “Guide to the Expression of Uncertainty in Measurement : Past, Present, and Future”, 7 & 8 novembre 2013, National Physical Laboratory, Teddington, UK.

- MALLARD, A. (1998), Compare, Standardize and Settle Agreement : On some Usual Metrological Problems, *Social Studies of Science*, 28(4), pp.571–601.
- MALLARD, A. (2000), L'écriture des normes, *Réseaux*, 18(102), pp.37–61.
- MARI, L. (1997), The role of determination and assignment in measurement, *Measurement*, 21(3), pp.79–90.
- MARI, L. (2003), Epistemology of measurement, *Measurement*, 34, pp.17–30.
- MARI, L. (2005), The problem of foundations of measurement, *Measurement*, 38, pp.259–266.
- MARI, L. (2008), On the measurand definition, in Éric BENOIT (dir.) : *Proceedings of the 18th IMEKO World Congress on "Metrology for a Sustainable Development", September, 17-22, 2006, Rio de Janeiro, Brésil*, pp.225–230, LISTIC (Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance) - Université de Savoie, Annecy, France. IMEKO.
- MARI, L. (2009), On (kinds of) quantities, *Metrologia*, 46, pp.L11–L15.
- MARI, L. (2015), An Overview of the Current Status of Measurement Science : From the Standpoint of the *International Vocabulary of Metrology (VIM)*, in SCHLAUDT, O. et HUBER, L. (dir.), *Standardization In Measurement : Philosophical, Historical And Sociological Issues*, pp.69–80, Pickering & Chatto.
- MARI, L. et GIORDANI, A. (2012), Quantity and quantity value, *Metrologia*, 49(6), pp.756–764.
- MARI, L. et GIORDANI, A. (2014), Modelling Measurement : Error and Uncertainty, in BOUMANS, M., HON, G. et PETERSEN, A. C. (dir.), *Error and Uncertainty in Scientific Practice*, pp.79–96, Pickering & Chatto.
- MARI, L. et MACII, D. (2008), Alternative Methods To Estimate Measurand Values : Models And Operative Implications, in Éric BENOIT (dir.) : *Proceedings of the 12th IMEKO TC1 & TC7 Joint Symposium on "Man Science & Measurement", September, 3-5, 2008, Annecy, France*, pp.225–230, LISTIC (Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance) - Université de Savoie, Annecy, France. IMEKO.
- MARIAN, D. (2015), The correspondence theory of truth, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, summer 2015 edition. Dernière consultation le 25 janvier 2016.
- MAYO, D. (1996), *Error and the Growth of Experimental Knowledge*, Chicago : University of Chicago Press.
- MAZLIAK, L. (2015), Poincaré's Odds, in DUPLANTIER, B. et RIVASSEAU, V. (dir.), *Henri Poincaré 1912-2012 Progress in Mathematical Physics*, pp.151–192, Birkhäuser.
- McMULLIN, E. (1985), Galilean idealization, *Studies in History and Philosophy of Science*, 16(3), pp.247–273.

- MELLOR, D. H. (1967), Imprecision and Explanation, *Philosophy of Science*, 34(1), pp.1–9.
- MICHELL, J. (1993), The Origins of the Representational Theory of Measurement, *Studies in History and Philosophy of Science*, 24(2), pp.185–206.
- MICHELL, J. (1994), Numbers as Quantitative Relations and the Traditional Theory of Measurement, *The British Journal for the Philosophy of Science*, 45(2), pp.389–406.
- MICHELL, J. (2005), The logic of measurement : a realist overview, *Measurement*, 38, pp.285–294.
- MINISTÈRE DE LA JEUNESSE, DE L'ÉDUCATION NATIONALE ET DE LA RECHERCHE (2002), *Document d'accompagnement, physique, classe terminale scientifique*.
- MIROWSKI, P. (1992), Looking for Those Natural Numbers : Dimensionless Constants and the Idea of Natural Measurement, *Science in Context*, 5(1), pp.165–188.
- MOFFAT, R. J. (1988), Describing the Uncertainties in Experimental Results, *Experimental Thermal and Fluid Science*, 1(1), pp.3–17.
- MOHR, P. J. et TAYLOR, B. N. (1999), CODATA Recommended Values of the Fundamental Physical Constants : 1998, *J. Phys. Chem. Ref. Data*, 28(6), pp.1713–1852.
- MOHR, P. J. et TAYLOR, B. N. (2005), CODATA Recommended Values of the Fundamental Physical Constants : 2002, *Reviews of Modern Physics*, 77, pp.1–107.
- MOHR, P. J., TAYLOR, B. N. et NEWELL, D. B. (2008), CODATA Recommended Values of the Fundamental Physical Constants : 2006, *Reviews of Modern Physics*, 80, pp.633–730.
- MOHR, P. J., TAYLOR, B. N. et NEWELL, D. B. (2012), CODATA Recommended Values of the Fundamental Physical Constants : 2010, *Reviews of Modern Physics*, 84, pp.1527–1579.
- MONTON, B. et MOHLER, C. (2014), Constructive empiricism, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, spring 2014.
- MORGAN, M. S. et MORRISON, M. (dir.) (1999), *Models as Mediators : Perspectives on Natural and Social Science*, Ideas in Context, Cambridge University Press.
- MÜLLER, J. W. (1979), Some second thoughts on error statements, *Nuclear Instruments and Methods*, 163(1), pp.241–251.
- ODDIE, G. (2016), Truthlikeness, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, spring 2016.
- O'HAGAN, A., BUCK, C. E., DANESHKHAH, A., EISER, J. R., GARTHWAITE, P. H., JENKINSON, D. J., OAKLEY, J. E. et RAKOW, T. (2006), *Uncertain Judgements : Eliciting Experts' Probabilities*, John Wiley & Sons, Ltd.
- ORESQUES, N. et CONWAY, E. M. (2012), *Les Marchands de doute*, Le Pommier. Traduit de l'anglais par Jacques Treiner. Dans son édition originale "Merchants of Doubt", Bloomsbury Press, 2010.

- ORGANISATION INTERNATIONALE DE NORMALISATION (ISO) (1984), *Vocabulaire international des termes fondamentaux et généraux de métrologie*.
- ORGANISATION INTERNATIONALE DE NORMALISATION (ISO) (1993), *Vocabulaire international des termes fondamentaux et généraux de métrologie, seconde édition*.
- PAVESE, F. (2007), The definition of the measurand in key comparisons : lessons learnt with thermal standards, *Metrologia*, 47(5), pp.327–339.
- PAVESE, F. (2009), On the degree of objectivity of uncertainty evaluation in metrology and testing, *Measurement*, 42(9), pp.1297–1303.
- PEARSON, K. (1920), Notes on the History of Correlation, *Biometrika*, 13(1), pp.25–45. Paper read to the Society of Biometricians and Mathematical Statisticians, June 14, 1920.
- PETIT, P. et LÉVY-LEBLOND, J.-M. (2011), Einstein s’est-il trompé ?, *marianne.fr*. 25 septembre 2011. http://www.marianne.net/philippepetit/Einstein-s-est-il-trompe_a212.html
- PETLEY, B. W. (1988), *The fundamental physical constants and the frontier of measurement*, A. Hilger.
- PETLEY, B. W. (1992), The Role of the Fundamental Constants of Physics in Metrology, *Metrologia*, 29(2), p.95.
- POHL, R., ANTOGNINI, A., NEZ, F., AMARO, F. D., BIRABEN, F., CARDOSO, J. M. R., COVITA, D. S., DAX, A., DHAWAN, S., FERNANDES, L. M. P., GIESEN, A., GRAF, T., HÄNSCH, T. W. W., INDELICATO, P., JULIEN, L., KAO, C.-Y., KNOWLES, P., LE BIGOT, E.-O., LIU, Y.-W., LOPES, J. A. M., LUDHOVA, L., MONTEIRO, C. M. B., MULHAUSER, F., NEBEL, T., RABINOWITZ, P., dos SANTOS, J. M. F., SCHALLER, L. A., SCHUHMAN, K., SCHWOB, C., TAQQU, D., VELOSO, J. F. C. A. et KOTTMANN, F. (2010), The size of the proton, *Nature*, 466(7303), pp.213–216.
- POHL, R., GILMAN, R., MILLER, G. A. et PACHUCKI, K. (2013), Muonic Hydrogen and the Proton Radius Puzzle, *Annual Review of Nuclear and Particle Science*, 63, p.175–204.
- POINCARÉ, H. (1896), *Le calcul des Probabilités*, Georges Carré, Paris.
- POMMÉ, S. (2015), “When the model doesn’t cover reality”, allocution lors de la conférence “BIPM Workshop on Measurement Uncertainty”, 15 & 16 juin 2015, Bureau International des Poids et Mesures, Sèvres, France.
- POPPER, K. R. (1935), *Logik der forschung : zur erkenntnistheorie der modernen naturwissenschaft*, Schriften zur wissenschaftlichen Weltauffassung, Vienne : J. Springer.
- POSSOLO, A., TOMAN, B. et ESTLER, T. (2009), Contribution to a conversation about the Supplement 1 to the GUM, *Metrologia*, 46(1), pp.L1–L7.
- PRIEL, M. (2008), Guide du vocabulaire de la métrologie : le concept de la valeur vraie fait débat, *Mesures*, 804, pp.20–23.

- PRIEL, M. (2010), Vocabulaire de la métrologie, in *Mesures-Analyses*, pp.1–9, Paris : techniques de l'ingénieur.
- PROTASSOV, K. (2002), *Analyse statistique des données expérimentales*, Grenoble : EDP sciences.
- QUINN, T. et BURNETT, K. (2005), Introduction : The fundamental constants of physics, precision measurements and the base units of the SI, *Phil. Trans. R. Soc. A*, 363(1834), pp.2101–2104.
- QUINN, T. J. (2002), Metrology, its role in today's world, in LIRA, I. (dir.), *Evaluating the Measurement Uncertainty : Fundamental and Practical Guidance*, pp.1–23, Bristol and Philadelphia, Institute of Physics Publishing.
- QUINN, T. J. (2012), *From Artefacts to Atoms : the BIPM and the Search for Ultimate Measurement Standards*, Oxford University Press.
- QUINN, T. J. (2013), "Background to the development of the GUM", allocution lors de la conférence "Guide to the Expression of Uncertainty in Measurement : Past, Present, and Future", 7 & 8 novembre 2013, National Physical Laboratory, Teddington, UK.
- RADDER, H. (dir.) (2003), *The Philosophy of Scientific Experimentation*, University of Pittsburgh Press.
- RAMSEY, J. L. (1992), Towards an Expanded Epistemology for Approximations, in *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp.154–164, The University of Chicago Press on behalf of the Philosophy of Science Association.
- REDHEAD, M. (1980), Models in Physics, *The British Journal for the Philosophy of Science*, 31(2), pp.145–163.
- ROBERT-SCHWARTZ, C. et TREINER, J. (2003), Incertitudes des mesures de grandeurs, in KAHANE, J.-P. (dir.), *Commission de réflexion sur l'enseignement des mathématiques, annexe sur la statistique*, pp.6–17.
- ROBERT-SCHWARTZ, C. et TREINER, J. (2006), Incertitudes des mesures de grandeurs, *Statistix*, pp.1–12. URL= <http://www.statistix.fr/spip.php?article17>; dernière consultation le 27 octobre 2015. ce texte a initialement été publié en 2003, en annexe de la « Commission de réflexion sur l'enseignement des mathématiques », dirigée par Jean-Pierre Kahane.
- RUSSELL, B. (1912), On the Notion of Cause, in *Proceedings of the Aristotelian Society, New Series*, pp.1–26, The Aristotelian Society.
- SAPORTA, G. (2006), *Probabilités, analyse des données et statistique*, Technip.
- SCHAFFER, S. (1992), Late Victorian Metrology and Its Instrumentation : A Manufactory of Ohms, in BUD, R., COZZENS, S. et POTTER, R. (dir.), *Invisible connections : instruments, institutions, and science*, SPIE Optical Engineering Press.

- SCHLAUDT, O. (2014), La réception de Hugo Dingler par l'École d'Erlangen, *Philosophia Scientiæ*, 18(2), pp.141–159. Mis en ligne le 26 septembre 2014, consulté le 10 février 2016. URL=<http://philosophiascientiae.revues.org/954>.
- SCHUMACHER, R. B. F. (1987), A dissenting position on uncertainties, *NCSL Newsletter*, 27(4), pp.55–59.
- SCHUMACHER, R. B. F. (1988), The european unity behind the BIPM recommendations may be not so unified, *NCSL Newsletter*, 28(1), p.58.
- SHAPER, D. (1982), The Concept of Observation in Science and Philosophy, *Philosophy of Science*, 49(4), pp.485–525.
- SHEWHART, W. A. (1922), On the measurement of a physical quantity whose magnitude is influenced by primary causes beyond the control of the observer and on the method of determining the relation between two such quantities, *Proceedings of the National Academy of Sciences of the United States of America*, 8(8), pp.248–251.
- SHEWHART, W. A. (1939), *Statistical method from the viewpoint of quality control*, Courier Corporation.
- SMITH, G. E. (2010), Revisiting Accepted Science : The Indispensability of the History of Science, *The Monist*, 93(4), pp.545–579.
- SMITH, S. (2002), Violated Laws, Ceteris Paribus Clauses, and Capacities, *Synthese*, 130(2), pp.235–264.
- SOLER, L. (2008), *Introduction à l'épistémologie*, Paris : Ellipses, 2^e édition.
- SPANOS, A. (2007), Curve fitting, the reliability of inductive inference, and the error-statistical approach, *Philosophy of Science*, pp.1046–1066.
- STIGLER, S. (1986), *The History of Statistics : The Measurement of Uncertainty before 1900*, Cambridge, MA : Belknap Press of Harvard University Press.
- STIGLER, S. (1999), *Statistics On The Table. The History of Statistical Concepts and Methods*, Harvard University Press.
- STUDENT (1908), The Probable Error of a Mean, *Biometrika*, 6(1), pp.1–25.
- SUCHAUT, B. (2008), La loterie des notes au bac : un réexamen de l'arbitraire de la notation des élèves. Document de travail de l'IREDU 2008-03.
- SUPPES, P. (1993), Representation Theory and the Analysis of Structure, in *Models and Methods in the Philosophy of Science : Selected Essays*, pp.67–82, Dordrecht : Kluwer Academic Publishers. (Initialement publié dans *Philosophia Naturalis*, 25 (1988), pp.254–268.).
- SUPPES, P., KRANTZ, D. H., LUCE, R. D. et TVERSKY, A. (1971), *Foundations of Measurement*, volume 1, New York : Academic Press.

- SUPPES, P., KRANTZ, D. H., LUCE, R. D. et TVERSKY, A. (1989), *Foundations of Measurement*, volume 2, New York : Academic Press.
- SUPPES, P., KRANTZ, D. H., LUCE, R. D. et TVERSKY, A. (1990), *Foundations of Measurement*, volume 3, New York : Academic Press.
- TAL, E. (2011), How Accurate is the Standard Second?, *Philosophy of Science*, 78(5), pp.1082–1096.
- TAL, E. (2012), *The Epistemology of Measurement : A Model-Based Account*, thèse de doctorat, University of Toronto.
- TAL, E. (2015), Measurement in science, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, summer 2015.
- TAYLOR, B. N. (1971), Comments on Least-Squares Adjustments of the Constants, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.495–498, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- TAYLOR, B. N. (2013), entretien avec l’auteur (non publié), 9 juin 2013.
- TAYLOR, B. N. et KUYATT, C. E. (1994), *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST.
- TAYLOR, B. N., PARKER, W. H. et LANGENBERG, D. N. (1969), Determination of e/h , Using Macroscopic Quantum Phase Coherence in Superconductors : Implications for Quantum Electrodynamics and the Fundamental Physical Constants, *J. Phys. Chem. Ref. Data*, 41(3), pp.375–496.
- TAYLOR, J. R. (1997), *An Introduction to Error Analysis*, University Science Books. Seconde édition (première édition : 1982).
- TAYLOR, J. R. (2000), *Incertitudes et analyse des erreurs dans les mesures physiques*, Masson Sciences. Traduction de la seconde édition (1997), traduit de l’américain par Lionel et Patrick Reynaud.
- TELLER, P. (2013), Measurement accuracy realism, in *Foundations of Physics 2013*. Article présenté au colloque “Foundations of Physics 2013 : The 17th UK and European Meeting on the Foundations of Physics”. Disponible en ligne : <http://philsci-archive.pitt.edu/9740/>.
- THOMSEN, J. S. (1971), Some Aspects of Least-Squares Adjustments of Constants, in LANGENBERG, D. N. et TAYLOR, B. N. (dir.), *Precision Measurement and Fundamental Constants*, pp.503–505, National Bureau of Standards. Proceedings of the International Conference held at the National Bureau of Standards, Gaithersburg, Maryland, August 3-7, 1970.
- TIERCELIN, C. (2006), Réalisme, in LECOURT, D. (dir.), *Dictionnaire d’histoire et philosophie des sciences*, pp.936–941, Presses Universitaires de France. 4^e édition (première édition : 1999).

- TREINER, J. (2011), Variabilité, incertitude, erreur, *Bulletin d'Union des Physiciens*, 105(930), pp.9–14.
- URBANSKI, M. K. et WASOWSKI, J. (2003), Fuzzy approach to the theory of measurement inexactness, *Measurement*, 34(1), pp.67–74.
- UZAN, J.-P. (2004), Les constantes fondamentales, *Sciences et Avenir*, Hors-Série, pp.14–19. Hors série.
- UZAN, J.-P. et LEHOUCQ, R. (2005), *Les constantes fondamentales*, Belin.
- VAN FRAASSEN, B. (1980), *The Scientific Image*, Oxford : Oxford University Press.
- VAN FRAASSEN, B. (2008), *Scientific Representation : Paradoxes of Perspectives*, Oxford : Oxford University Press.
- VANLERBERGHE, C. (2011), Relativité : Einstein contredit par des chercheurs du CNRS, *lefigaro.fr*. 23/09/2011. <http://www.lefigaro.fr/sciences/2011/09/22/01008-20110922ARTFIG00686-relativite-einstein-contredit-par-des-chercheurs-francais.php>
- VARENNE, F. (2012), *Théories, réalité, modèles*, Éditions Matériologiques.
- VON HELMHOLTZ, H. (1887), Zählen und messen, erkenntnistheoretisch betrachtet, in *Philosophische Aufsätze, Eduard Zeller zu seinem fünfzigjährigen Doctorjubiläum gewidmet*. Leipzig : Fues.
- VORMS, M. (2011), *Qu'est-ce qu'une théorie scientifique ?*, Vuibert.
- VOSK, T. et EMERY, A. F. (2014), *Forensic Metrology : Scientific Measurement and Inference for Lawyers, Judges, and Criminalists*, International Forensic Science and Investigation, CRC Press.
- WANG, C. J. et IYER, H. K. (2009), Fiducial intervals for the magnitude of a complex-valued quantity, *Metrologia*, 46(1), p.81.
- WEISE, K. et WÖGER, W. (1993), A Bayesian theory of measurement uncertainty, *Measurement Science and Technology*, 4(1), pp.1–11.
- WELCH, B. L. (1947), The Generalization of 'Student's' Problem when Several Different Population Variances are Involved, *Biometrika*, 34(1/2), pp.28–35.
- WILLINK, R. (2010a), Difficulties arising from the representation of the measurand by a probability distribution, *Measurement Science and Technology*, 21, pp.1–11.
- WILLINK, R. (2010b), On the validity of methods of uncertainty evaluation, *Metrologia*, 47(1), pp.80–89.
- WILLINK, R. (2010c), Probability, belief and success rate : comments on 'On the meaning of coverage probabilities', *Metrologia*, 47(3), pp.343–346.

- WILLINK, R. (2013), *Measurement Uncertainty and Probability*, Cambridge : Cambridge University Press.
- WILLINK, R. et HALL, B. D. (2001), Does “Welch-Satterthwaite” make a good uncertainty estimate ?, *Metrologia*, 38(9), pp.9–15.
- WILLINK, R. et WHITE, R. (2012), Disentangling Classical and Bayesian Approaches to Uncertainty Analysis. Technical Report No. CCT/12-08, Rapport technique, BIPM, Sèvres.
- WIMSATT, W. C. (1987), False Models as Means to Truer Theories, in NITECKI, M. et HOFFMAN, A. (dir.), *Neutral Models in Biology*, pp.23–55, London : Oxford University Press. Reproduit dans Wimsatt, *Re-Engineering Philosophy for Limited Beings : Piecewise Approximations to Reality*, Harvard University Press, 2007.
- WINSBERG, E. (2015), Computer simulations in science, in ZALTA, E. N. (dir.) : *The Stanford Encyclopedia of Philosophy*. Stanford University, summer 2015.
- WISE, M. N. (dir.) (1997), *The Values of Precision*, Princeton University Press.
- WÖGER, W. (2014), entretien avec l’auteur (non publié), 3 juin 2014.
- WOODWARD, J. (1988), Understanding Regression, *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1988, pp.255–269.
- ZABELL, S. L. (2008), On Student’s 1908 Article “The Probable Error of a Mean”, *Journal of the American Statistical Association*, 103(481), pp.1–7.
- ZADEH, L. A. (2005), Toward a generalized theory of uncertainty (GTU)-An outline, *Information Sciences*, 172(1-2), pp.1–40.

Index des noms propres

A

Achinstein, Peter 221
Ackermann, Robert 21
Adam, Thomas 103, 309, 311
Ambler, Ernest ... 286, 320, 321, 323–325, 329,
331, 332
Armatte, Michel .. 22, 35, 55, 58, 67, 80, 90, 96,
248, 306
Aronson, Jerrold 227

B

Babbage, Charles 259
Baratto, Antonio 198
Barberousse, Anouk 19, 83, 227
Barnett, S.J. 34
Barsalou, Matthew 62
Bayes, Thomas 106
Bearden, Joyce 96, 277, 280, 281
Beaufils, Daniel 203
Bender, Peter ... 276, 286–289, 291, 292, 295,
296, 303, 304
Bernauer, Jan 295
Bernoulli, Jacques 55
Bernoulli, Jean 58
Bertolucci, Sergio 311, 312
Bich, Walter .. 43, 110, 120, 132, 139, 142, 144,
174, 189, 230, 319, 327, 328, 333
Bienvenu, Alexis 71, 72
Birge, Raymond . 28, 75, 86, 89, 90, 96, 97, 101,
122, 238, 255–272, 275–282, 290, 297,
307, 315, 330

Blevin, William 324
Bogen, James 22, 59
Bond, Wilfrid 276, 277
Bouguer, Pierre 96
Boumans, Marcel 19
Bourguet, Marie-Noëlle 23
Box, George 226
Boyd, Richard 227
Branscomb, Lewis 259, 264, 278
Brenner, Anastasios 225
Bullyncck, Maarten 248
Burnett, Keith 314
Burns, John 84, 100, 103, 139, 168, 352

C

Campbell, Norman 36, 74, 163–166
Campion, Peter ... 84, 100, 103, 139, 168, 352
Carnap, Rudolf 20, 71, 179
Cartwright, Nancy 58, 83, 219, 222–225
Chakravartty, Anjan 227
Chang, Hasok 23, 58
Clifford, P.M. 196, 197
Cohen, E. Richard ... 73, 86, 90, 96, 102, 188,
258, 261, 262, 264, 277–279, 281–283,
285, 287, 289, 291–294, 326, 329
Cohen-Tannoudji, Gilles 258
Colclough, A. R. 60
Conway, Erik 41
Cornell Way, Eileen 227
de Courtenay, Nadine ... 36, 42, 62, 119, 148,
163–166, 204, 215, 295, 309
Cox, Maurice 122, 320

Croarkin, Carroll 330, 331

D

D'Agostini, Giulio 66, 122, 144, 145

Désenfant, Michèle 44

Damour, Thibault 310

Darrigol, Olivier 163, 219, 222

Daston, Lorraine 68, 114

De Finetti, Bruno 71

de Haas, Wander Johannes 34

De Bièvre, Paul 198

Demeyer, Séverine 122, 134, 297

Deming, W. Edward 75, 89, 90, 122, 261

Dorsey, Noah 266–268, 281

Draper, Norman 226

Drouet, Isabelle 109, 145, 149, 150, 350

Duhem, Pierre 20, 179, 225, 227

DuMond, Jesse 73, 86, 90, 96, 102, 188, 258,
261, 262, 264, 277–279, 281–283, 285,
287, 289, 291–293

Dunnington, Franck 277

Dybkaer, René 77, 79, 111, 120, 134, 153, 168,
169, 173–175, 184, 189, 202, 203

E

Eddington, Arthur 134, 276

Edgeworth, Francis Ysidro 67

Ehrlich, Charles 77, 79, 111, 120, 134, 139,
153, 168, 169, 173–175, 184, 189, 202,
203, 214, 215, 319

Einstein, Albert 34, 309

Eisenhart, Churchill 43, 57, 73–76,
83, 95, 97–101, 103, 118, 119, 138, 140,
143, 177, 188, 202–205, 228, 259, 308,
330, 331

Ellison, Stephen 198

Elster, Clemens 122, 153

Emery, Ashley 199, 208, 230, 231

Ereditato, Antonio 312

Estler, W. Tyler 113, 145, 149

Euclide 163

Euler, Leonhard 238, 242–249, 251, 255,
262–264, 357

F

Fechner, Gustav 165

Feyerabend, Paul 20, 225

Franken, Peter 307

Franklin, Allan 20, 22

Frenkel, Bob 63, 119

Frigg, Roman 219, 221

G

Gabrielse, Gerald 192

Galilée 137, 216, 224, 246

Galison, Peter 20–22, 34

Galton, Francis 67

Gauss, Carl Friedrich 44, 52,
58, 88, 89, 238, 241, 242, 248, 251, 252,
256, 261, 264, 315

Geiger, Hans 258

Giacomo, Pierre 65, 102, 321, 323, 324

Gié, Hubert 86

Gillies, Donald 72, 110, 142, 336, 337

Giordani, Alessandro 137, 164, 165, 198, 215,
216, 227

Glanzberg, Michael 177, 181, 182

Goethe, Johann Wolfgang 285

Good, Irving 145

Grabe, Michael 119, 137

Guidice, Sebastiano 313, 314

H

Hacking, Ian 20, 21, 45, 68, 110, 114, 115, 259

- Hájek, Alan 71, 72, 109, 110, 115, 142, 144
 Hall, Blair D. 119, 148
 Hanneke, David 192
 Hanson, Norwood 20
 Hardwig, John 309
 Harré, Rom 23, 227
 Harris, Peter 320
 Hartmann, Stephan 219, 221
 Heidelberger, Michael 21, 23
 Helmholtz, A. Carl 256, 261, 265
 Himbert, Marc 19, 23
 Hon, Giora . . . 19, 24, 33, 37, 41, 52, 60–62, 80,
 81, 193, 246, 306
 Howson, Colin 144
 Humphreys, Paul 154, 221

I

- Ichikawa, Jonathan 307
 Imbert, Cyrille 221, 224
 Ishikawa, Kaoru 62
 Iyer, Hari 148

J

- James, William 182
 Janich, Peter 20
 Jaynes, Edwin Thompson 71, 110, 144
 Jeffrey, Richard 350
 Jeffreys, Harold 349, 350
 Jones, Albert 121, 125, 132, 134

K

- Kaarls, Robert . . 52, 65, 101–103, 115, 323–325
 Kacker, Raghu 121, 125, 132, 134
 Kant, Emmanuel 191, 192
 Karp, Pekka 325, 326
 Kepler, Johannes 137, 246

- Keynes, John Maynard 110, 142, 144
 Keyser, Cassius 188
 Kirkup, Les 63, 119
 Kistler, Max 19, 221, 223
 Kolmogorov, Andreï 109, 114
 Kool, Willem 43, 120, 142
 Kragh, Helge 276
 Krantz, David 163
 Kuhn, Thomas 20
 Kuyatt, Chris 103, 326, 327, 329, 330
 Kyburg, Henri 35, 36, 59, 146, 193

L

- Lacki, Jan 248
 Lamb, Willis 293
 Lambert, Johann Einrich 22
 Langenberg, Donald . . . 63, 277, 282–287, 289,
 292, 304
 Laplace, Pierre-Simon de . . . 88, 89, 106, 109,
 142, 242, 248, 251, 315
 Larousserie, David 312
 Laudan, Larry 227
 Lautrup, Benny 43, 304–306
 Laymon, Ronald . . . 83, 219, 220, 222, 224–226
 Layton, Thomas 277
 Lecoutre, Bruno 91
 Legendre, Adrien Marie . . . 238, 241, 242, 248,
 250–252, 256, 261, 264, 315
 Leibniz, Gottfried Wilhelm 55
 Lévy-Leblond, Jean-Marc . . . 189, 206, 259, 310
 Lewino, Frédéric 310
 Licoppe, Christian 23
 Lide, David 278, 329
 Lindley, Denis 146
 Lippman, Gabriel 89
 Lira, Ignacio . . 112, 120, 144, 149, 347, 349–351
 Locke, John 55
 Loftus, Pete 313, 314
 Luce, R. Duncan 163
 Ludwig, Pascal 19

M

Mach, Ernst 179
 Macii, David 207
 Mallard, Alexandre 331, 332
 Mari, Luca . . . 36, 137, 141, 164, 165, 168, 177,
 191, 198, 207, 215–217, 227
 Marian, David 181, 191
 Mayer, Tobias . . . 238, 242–252, 255, 262–264,
 270, 315, 357
 Mayo, Deborah 71, 89, 143, 145, 146, 251
 Mazliak, Laurent 89
 McMullin, Ernan 216, 217, 219, 224, 226
 Mellor, David 61
 Mercier, Jean 266, 267, 272, 281
 Michell, Joel 163, 181
 Michelson, Abraham 266–268, 281
 Mirowski, Philip 259, 260
 Moffat, Robert 63
 Mohler, Chad 179
 Mohr, Peter 41, 192, 278
 Monton, Bradley 179
 Moreau, René 86
 Morgan, Mary 221
 Morrison, Margaret 219, 221
 Müller, Jörg 43, 122

N

Newell, David 41, 192, 278, 295
 Newton, Isaac 248
 Nez, François 278, 295
 Niiniluoto, Ilkka 227

O

Oddie, Graham 227
 O'Hagan, Anthony 142
 Oreskes, Naomi 41

P

Parker, William H. 277, 282–285, 287, 289, 292,
 304
 Pavese, Franco 167, 171, 180, 201
 Pearson, Karl 67, 89
 Peirce, Charles Sanders 182
 Perlstain, André 321
 Perovic, Slobodan 22
 Perrin, Jean 223
 Petersen, Arthur 19
 Petley, Brian 57, 259
 Pohl, Randolph 295
 Poincaré, Henri 89, 165, 179
 Polanyi, Michel 41
 Pommé, Stefaan 308
 Popper, Karl 20, 227, 336, 337
 Possolo, Antonio 145, 149
 Priel, Marc 44, 175, 217, 331
 Protassov, Konstantin 103, 270

Q

Quetelet, Adolphe 88
 Quinn, Terry . . 45, 261, 313, 314, 322–325, 327,
 329, 330

R

Radder, Hans 23
 Ramsey, Frank 71
 Ramsey, Jeffrey 83
 Ramsey, Norman 293
 Redhead, Michael 221
 Reichenbach, Hans 71, 227
 Robert, Claudine 91, 177, 201, 218
 Rollett, John 277
 Rosa, Edward 266–268, 281
 Rumsfeld, Donald 308
 Russell, Bertrand 163, 222, 223

S

Saporta, Gilbert .. 82, 84, 88, 89, 123, 124, 345, 346, 351
 Schaffer, Simon 23
 Scheel, Karl 258
 Schickore, Jutta 24
 Schlaudt, Oliver 20
 Schumacher, Rolf 330, 331
 Shapere, Dudley 21
 Shewhart, Walter 72, 73
 Sibum, H. Otto 23
 Simpson, Thomas 80
 Smith, George 259
 Smith, Sheldon 221
 Soler, Léna 179, 182
 Spanos, Aris 251
 Stamenkovic, Philippe 191
 Steinle, Friedrich 24
 Steup, Matthias 307
 Stigler, Stephen 67, 71, 88–90, 94, 106, 124, 238, 241–244, 246–248, 250, 264, 280, 357
 Student (William Gosset) 74, 89, 346
 Suchaut, Bruno 31, 32
 Suppes, Patrick 163, 165, 221

T

Tal, Eran 19, 23, 39, 199, 227
 Taylor, Barry .. 41, 63, 103, 192, 272, 276–279, 282–287, 289–292, 294, 295, 302–305, 312, 314, 319, 320, 323, 325–331
 Taylor, John 37, 38, 81
 Teller, Paul 201
 Terrien, Jean 320, 321, 329
 Thomas, Claudine 295
 Thomsen, John ... 96, 277, 280, 281, 290, 291, 306, 307
 Thomson, Joseph J. 259
 Tichý, Pavel 227

Tiercelin, Claudine 190
 Toman, Blaza 145, 149
 Treiner, Jacques .. 177, 201, 203, 204, 209, 210, 218
 Tversky, Amos 163

U

Urbach, Peter 144
 Urbanski, Michal 153
 Uzan, Jean-Philippe 219, 258

V

van Fraassen, Bas 23, 178–180, 222
 Vanlerberghe, Cyrille 309
 Varenne, Franck 219, 221
 Vigoureux, Paul 63
 von Helmholtz, Hermann 163, 179, 261
 Vorms, Marion 179
 Vosk, Ted 199, 208, 230, 231

W

Wang, C.M. Jack 148
 Wasowski, Janusz 153
 Watts, H.M. 277
 Weinberg, Joseph 261
 Weise, Klaus 66, 113, 122, 142, 144
 Welch, Bernard 119
 White, Rod 68, 69
 Williams, Alexander ... 84, 100, 103, 139, 168, 198, 352
 Willink, Robin . 58, 65, 68, 69, 75, 91, 119, 135, 136, 139, 143, 145, 146, 148–150, 177, 202, 301, 303, 304, 309
 Wimsatt, William 190, 226
 Winsberg, Eric 153
 Wise, M. Norton 23

- Wöger, Wolfgang . . . 77, 79, 111–113, 120, 122,
134, 142–144, 153, 168, 169, 173–175,
189, 202, 203, 327
- Wood, Gordon 278, 329
- Woodward, James 22, 59, 251
- Worrall, John 227

Y

- Yule, George 67

Z

- Zabell, Sandy .L. 346
- Zadeh, Lofti 153
- Zinkernagel, Henrik 43, 304–306

La valeur de l'incertitude : l'évaluation de la précision des mesures physiques et les limites de la connaissance expérimentale

Résumé. Un résultat de mesure n'est jamais exact : il est affecté d'une « erreur de mesure », inconnue, qui caractérise l'écart entre la valeur obtenue et la « valeur vraie » de la grandeur visée. Par conséquent, un résultat de mesure acceptable ne peut pas se présenter sous la forme d'une unique valeur numérique, mais doit être accompagné d'une indication de l'« incertitude de mesure » qui lui est attachée, laquelle énonce un doute. Mais quelle est la valeur de l'incertitude de mesure ? Quelle est sa valeur numérique : comment la calcule-t-on ? Quelle est sa valeur épistémique : comment peut-on interpréter un résultat de mesure ? Dans un premier temps, nous décrivons les modèles statistiques auxquels les scientifiques font appel dans la métrologie contemporaine pour effectuer l'analyse d'incertitude, et nous montrons que la question de l'interprétation des probabilités y fait l'objet d'un débat très vif. Ce débat fait émerger des questions épistémologiques sur la nature et la fonction de la mesure physique, les métrologues insistant de plus en plus sur le caractère subjectif de cette dernière. Dans un second temps, nous examinons l'élaboration philosophique des métrologues dans leurs ouvrages techniques, où ceux-ci critiquent l'usage de la notion de « valeur vraie » d'une grandeur, et nous remettons en question cette élaboration en défendant à notre tour une telle notion. La troisième partie se tourne vers un usage spécifique de l'incertitude de mesure pour aborder la thématique sous l'angle de la physique de précision, au travers de l'activité des ajustements des constantes de la physique. Au cours de celle-ci, les physiciens développent une conception dynamique de l'exactitude des résultats de mesure, orientée vers le progrès futur en soulignant les vertus épistémiques d'un processus sans fin d'identification et de correction des erreurs de mesure.

The value of uncertainty: the evaluation of the precision of physical measurements and the limits of experimental knowledge

Abstract. A measurement result is never absolutely accurate: it is affected by an unknown “measurement error” which characterizes the discrepancy between the obtained value and the “true value” of the quantity intended to be measured. As a consequence, to be acceptable a measurement result cannot take the form of a unique numerical value, but has to be accompanied by an indication of its “measurement uncertainty”, which enunciates a state of doubt. What, though, is the value of measurement uncertainty? What is its numerical value: how does one calculate it? What is its epistemic value: how one should interpret a measurement result? Firstly, we describe the statistical models that scientists make use of in contemporary metrology to perform an uncertainty analysis, and we show that the issue of the interpretation of probabilities is vigorously debated. This debate brings out epistemological issues about the nature and function of physical measurements, metrologists insisting in particular on the subjective aspect of measurement. Secondly, we examine the philosophical elaboration of metrologists in their technical works, where they criticize the use of the notion of “true value” of a physical quantity. We then challenge this elaboration and defend such a notion. The third part turns to a specific use of measurement uncertainty in order to address our thematic from the perspective of precision physics, considering the activity of the adjustments of physical constants. In the course of this activity, physicists have developed a dynamic conception of the accuracy of their measurement results, oriented towards a future progress of knowledge, and underlining the epistemic virtues of a never-ending process of identification and correction of measurement errors.

Mots-clés. Histoire et philosophie des sciences, épistémologie; mesure, métrologie; erreur, erreur de mesure, précision, exactitude, incertitude; modèles statistiques, probabilités, fréquentisme, bayésianisme; physique, constantes physiques, ajustements

Discipline: histoire et philosophie des sciences

Université Paris Diderot–Paris 7 (5 rue Thomas-Mann 75013 Paris)

École doctorale: ED 400 – « Savoirs scientifiques, épistémologie, histoire des sciences, didactique des disciplines »

Équipe d'accueil: laboratoire SPHERE, UMR 7219