



HAL
open science

Le hasard et les sources de la variation biologique : analyse critique d'une notion multiple

Francesca Merlin

► **To cite this version:**

Francesca Merlin. Le hasard et les sources de la variation biologique : analyse critique d'une notion multiple. Philosophie. Université Panthéon-Sorbonne - Paris I, 2009. Français. NNT : . tel-00397641v1

HAL Id: tel-00397641

<https://theses.hal.science/tel-00397641v1>

Submitted on 22 Jun 2009 (v1), last revised 24 Jun 2009 (v2)

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.

Université Paris 1 Panthéon - Sorbonne, Ecole Doctorale de Philosophie

THÈSE

Pour l'obtention du grade de
docteur en Philosophie de l'Université Paris 1 Panthéon - Sorbonne

Présentée et soutenue publiquement par

Francesca MERLIN

le 08 juin 2009

Le hasard et les sources de la variation biologique : analyse critique d'une notion multiple

Sous la direction de

Jean GAYON, Professeur à l'Université Paris 1 Panthéon - Sorbonne
et de

Michel MORANGE, Professeur à l'Ecole Normale Supérieure de Paris

Composition du jury

Frédéric BOUCHARD	Professeur de Philosophie à l'Université de Montréal
Jean GAYON	Professeur de Philosophie à l'Université Paris 1 Panthéon - Sorbonne
Thierry MARTIN	Professeur de Philosophie à l'Université de Franche-Comté
Michel MORANGE	Professeur de Biologie à l'Ecole Normale Supérieure de Paris
Miroslav RADMAN	Professeur de Biologie à l'Université Paris 5

Université Paris 1 Panthéon - Sorbonne, Ecole Doctorale de Philosophie

THÈSE

Pour l'obtention du grade de
docteur en Philosophie de l'Université Paris 1 Panthéon - Sorbonne

Présentée et soutenue publiquement par

Francesca MERLIN

le 08 juin 2009

Le hasard et les sources de la variation biologique : analyse critique d'une notion multiple

Sous la direction de

Jean GAYON, Professeur à l'Université Paris 1 Panthéon - Sorbonne
et de

Michel MORANGE, Professeur à l'Ecole Normale Supérieure de Paris

Composition du jury

Frédéric BOUCHARD	Professeur de Philosophie à l'Université de Montréal
Jean GAYON	Professeur de Philosophie à l'Université Paris 1 Panthéon - Sorbonne
Thierry MARTIN	Professeur de Philosophie à l'Université de Franche-Comté
Michel MORANGE	Professeur de Biologie à l'Ecole Normale Supérieure de Paris
Miroslav RADMAN	Professeur de Biologie à l'Université Paris 5

Remerciements

Tout d'abord, je tiens à remercier mes deux directeurs de thèse, Jean Gayon et Michel Morange. Grâce à leurs encouragements et à leur confiance en moi, j'ai pu faire face aux difficultés et aux doutes qui sont apparus tout au long de cette recherche et la mener à terme dans les meilleures conditions. Les commentaires, suggestions et critiques qu'ils m'ont dispensés ont été très précieux pour ma réflexion, pour l'écriture de cette thèse, et font partie d'un bagage philosophique et scientifique essentiel pour ma carrière future. Je suis particulièrement reconnaissante à Jean Gayon pour mon initiation à la philosophie de la biologie, qui a véritablement eu lieu sous sa supervision. Il m'a transmis la rigueur et la clarté qui sont propres à sa pensée. J'exprime aussi ma gratitude à Michel Morange, qui a partagé avec moi ses réflexions, toujours d'une grande subtilité.

Je remercie Frédéric Bouchard, Thierry Martin et Miroslav Radman pour avoir gentiment accepté de faire partie de mon jury de thèse.

Je ne peux pas manquer d'exprimer ma reconnaissance à tous les membres de l'Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST), qui a représenté pour moi un environnement idéal, à la fois stimulant et chaleureux, pendant ces années. Je tiens à remercier, d'abord, le directeur Jacques Dubucs qui m'a soutenue sans aucune réticence. Entre autres, il m'a permis d'organiser le séminaire de Philosophie de la Biologie de l'IHPST au cours des trois dernières années, ce qui m'a donné l'occasion de connaître de nombreuses personnes travaillant dans le domaine et de partager mes réflexions. Merci à tous les participants fidèles de ce séminaire et aux invités qui ont accepté d'y intervenir.

Un remerciement tout spécial va à Anouk Barberousse pour son aide constante, nos collaborations, ses encouragements et son amitié : sans elle, mon travail de recherche n'aurait pu trouver un tel aboutissement. De même, je remercie Philippe Huneman et Françoise Longy pour toutes les discussions que nous avons eues au cours de ces années et pour avoir été auprès de moi dans des moments très difficiles qui, d'ailleurs, les ont touchés aussi profondément. J'aimerais remercier Peggy Cardon, dont la présence à l'Institut ne peut que rendre l'ambiance et le travail encore plus agréables.

Je remercie tous les doctorants de l'Institut, parmi lesquels figurent de grands amis. En particulier, j'exprime ma gratitude à Isabelle, Cédric, Steeves et Pierre-Alain non seulement pour le temps qu'ils ont tous consacré à la relecture de ma thèse, mais surtout pour les bons moments qu'on a vécus ensemble pendant ces années. De même, je remercie Carlo, Matteo, Francesca, Brian, Christophe, Alexandra, Hugo, Cyrille, Thomas, Neil, Arancha, Jindrich, Frédérique, Johannes, Antonine. Merci à Etienne et Valérie. Merci aussi à Eric pour son aide et sa patience.

Ma recherche a bénéficié du dynamisme du laboratoire de Miroslav Radman, en particulier des discussions avec François Taddei, Ariel Lindner et Ivan Matic, que je remercie. Je suis aussi reconnaissante à tous les étudiants du master Approches Interdisciplinaires du Vivant qui ont de nombreuses fois partagé avec moi leurs compétences scientifiques. Je remercie également tous les chercheurs, biologistes et physiciens, qui m'ont consacré du temps, entre autres, François Amblard, Alain Arneodo, Annick Lesne, Olivier Tenailon, Yan Boucher, Jérôme Wong, Etienne Couturier, Rémy Peyronnet.

J'ai pu consacrer ces années à la recherche en philosophie grâce à une allocation de l'Université Paris 1, que je remercie. Je remercie également l'U.F.R. de Philosophie, l'Ecole Doctorale et l'IHPST. Ma recherche a bénéficié d'une impulsion majeure lors du séjour de trois mois que j'ai effectué au département de Philosophie de l'Université de Davis

Californie. Cela a été possible grâce à une bourse de la région Ile-de-France, que je remercie. Je tiens à remercier très sincèrement Roberta Millstein pour m'avoir accueillie à Davis et m'avoir consacré du temps précieux : j'ai bénéficié de ses conseils, de ses critiques et de sa passion pour la philosophie. Merci également au directeur du département de Philosophie de l'Université de Davis, Jim Griesemer, et à tous ses membres, en particulier à mes amis Vadim, Mike, Ayalet et Jan.

Je suis reconnaissante à Evelyn Fox Keller pour les échanges que nous avons eus lors de ses séjours à Paris et qui m'ont permis de mieux cerner mon sujet de recherche. Ma gratitude va aussi à Joseph Berkovitz, non seulement pour ses commentaires et critiques, mais aussi pour sa générosité et son amitié.

Merci à Roselyne Richter pour les innombrables articles scientifiques et les informations concernant des séminaires qu'elle a eu la gentillesse de me donner.

Les amis qui sont loin, en Italie ou ailleurs, ont été très patients avec moi. En particulier, je remercie Priscilla, Claudio et Gloria. J'aimerais aussi exprimer ma profonde gratitude à mes amis Olivier et Félix, pour leur soutien durant ces années. Merci aussi à Jean-Sébastien.

Je suis infiniment reconnaissante à mon père, qui m'a donné la chance de venir à Paris et de faire de la recherche. Merci aussi à ma sœur pour son affection, ses encouragements et sa bonne humeur.

Un merci très spécial à Fabienne qui pendant ces années m'a beaucoup aidée et a toujours trouvé du temps pour relire mes manuscrits. Elle m'a écoutée à tout moment et m'a soutenue dans toute circonstance avec patience, générosité et beaucoup d'humour.

Enfin, ma pensée va à Marie-Claude, avec laquelle j'aurais aimé partager ce moment.

À Marie-Claude

Sommaire

INTRODUCTION	13
--------------------	----

PARTIE I – LE HASARD DANS LA PENSEE PHILOSOPHIQUE ET SCIENTIFIQUE.....	37
---	-----------

CHAPITRE 1 – HASARD SUBJECTIF ET OBJECTIF, DETERMINISME ET INDETERMINISME	43
--	-----------

1.1. LA NOTION DE HASARD : UN SUJET DE REFLEXION ANCIEN	43
1.2. LA DOUBLE NATURE DU HASARD	45
1.3. DETERMINISME ET INDETERMINISME	46
1.3.1. <i>Le déterminisme</i>	48
1.3.2. <i>L'indéterminisme</i>	52
1.4. LE HASARD SUBJECTIF	54
1.4.1. <i>Le hasard subjectif en tant qu'« ignorance des causes »</i>	54
1.5. LE HASARD OBJECTIF	57
1.5.1. <i>Le hasard objectif en tant qu'« absence d'un dessein »</i>	58
1.5.2. <i>Le hasard objectif en tant que « rencontre de séries causales indépendantes »</i>	60
1.5.3. <i>Le hasard objectif en tant que « sensibilité aux conditions initiales »</i>	65
1.5.4. <i>Le « hasard pur » du niveau quantique</i>	71
1.6. NOTRE AGNOSTICISME METAPHYSIQUE.....	75

CHAPITRE 2 – LA MATHEMATISATION DU HASARD.....	79
---	-----------

2.1. LA NOTION DE HASARD EN TANT QU'« EQUIPROBABILITE »	81
2.1.1. <i>L'équiprobabilité en termes de processus : la notion d'« échantillonnage indiscriminé »</i>	85
2.1.2. <i>Déterminisme, indéterminisme et équiprobabilité</i>	86
2.2. QU'EST-CE QU'UNE SUITE ALEATOIRE ?.....	88
2.3. LA NOTION DE HASARD EN TANT QUE « STABILITE DES FREQUENCES A LA LIMITE ».....	90
2.3.1. <i>Déterminisme, indéterminisme et stabilité des fréquences à la limite</i>	92
2.4. LA NOTION DE HASARD EN TANT QU'« INCOMPRESSIBILITE » D'UNE SUITE	95
2.4.1. <i>Déterminisme, indéterminisme et incompressibilité</i>	98
2.5. LA NOTION DE HASARD EN TANT QUE « SATISFACTION DE TOUS LES TESTS STATISTIQUES EFFECTIFS ».....	100
2.5.1. <i>Le consensus autour de la bonne définition mathématique de hasard</i>	102
2.5.2. <i>Les notions mathématiques de hasard en biologie</i>	103
2.6. LES THEORIES PHILOSOPHIQUES DES PROBABILITES.....	106
2.6.1. <i>Les interprétations épistémiques des probabilités</i>	109
2.6.2. <i>Les interprétations objectives des probabilités</i>	111
2.6.3. <i>La relation entre hasard et probabilité</i>	114

PARTIE II – HASARD ET MUTATIONS GENETIQUES 119

CHAPITRE 3 – LES MUTATIONS GENETIQUES : UNE INTRODUCTION 127

3.1. LES CONCEPTIONS DE L'ORIGINE DE LA VARIATION BIOLOGIQUE (XIX ^E -XXI ^E SIECLES)	127
3.1.1. Brève histoire de la notion de « mutation »	127
3.1.2. Lamarck, Darwin et les premiers lamarckiens	132
3.1.3. Les lamarckiens et les « néo-lamarckiens » (fin XIX ^e - début XX ^e)	134
3.1.4. De la Synthèse Moderne (1930-1950) à aujourd'hui	137
3.2. DES ALTERATIONS DIVERSES DE LA SEQUENCE D'ADN	139
3.2.1. La stabilité dynamique de l'ADN : les mécanismes de réparation	143
3.3. LE TAUX DE MUTATION	146
3.3.1. Les méthodes d'évaluation du taux de mutation	147
3.4. NOTRE PROJET : CLARIFIER LA NOTION DE HASARD DANS LA CARACTERISATION DES MUTATIONS GENETIQUES	151

CHAPITRE 4 – LA NOTION DE « HASARD EVOLUTIONNAIRE » DE DARWIN A LA SYNTHESE MODERNE 155

4.1. DARWIN	155
4.2. LA SYNTHESE MODERNE	160
4.2.1. Les « biais de mutation » à l'époque de la Synthèse Moderne	164
4.2.2. La vision consensuelle de la Synthèse Moderne quant au hasard des mutations	169
4.3. LA NOTION DE « HASARD EVOLUTIONNAIRE »	174

CHAPITRE 5 – AVANCEES ET DEBATS SUR L'ORIGINE ET LE CARACTERE DES MUTATIONS GENETIQUES, DE LA SYNTHESE MODERNE JUSQU'A NOS JOURS 179

5.1. DU TEST DE FLUCTUATION DE LURIA ET DELBRÜCK (1943) A L'HYPOTHESE DES « BACTERIES LAMARCKIENNES » (CAIRNS ET AL, 1988)	179
5.1.1. La preuve fournie par le test de fluctuation	179
5.1.2. Les limites des preuves en faveur de la vision consensuelle de la Synthèse Moderne	182
5.2. LE DEFI LANCE AU PARADIGME DE LA SYNTHESE MODERNE : DES MUTATIONS « DIRIGÉES » (1988) AUX MUTATIONS « ADAPTATIVES » (1990-2008)	186
5.2.1. Le débat controversé sur le caractère « dirigé », puis « adaptatif », des mutations génétiques	189
5.2.2. La notion de « mutation adaptative »	193
5.3. LES MECANISMES « MUTATEURS »	196
5.3.1. Les mécanismes mutateurs constitutifs	198
5.3.2. Les mécanismes mutateurs induits	202

CHAPITRE 6 – PEUT-ON ENCORE AFFIRMER QUE *TOUTE* MUTATION GENETIQUE EST LE FRUIT DU « HASARD EVOLUTIONNAIRE » ?211

6.1. LES MECANISMES MUTATEURS : UN VERITABLE DEFI A LA SYNTHÈSE MODERNE ?211	211
6.2. JABLONKA & LAMB (2005) : ANALYSE EMPIRIQUE ET CONCEPTUELLE DES MUTATIONS DUES AUX MECANISMES MUTATEURS212	212
6.2.1. <i>Le paradigme de la Synthèse Moderne selon Jablonka & Lamb (2005)</i>214	214
6.2.2. <i>Les mutations dues aux mécanismes mutateurs selon Jablonka & Lamb (1995, 2005)</i>218	218
6.2.3. <i>Jablonka et Lamb (1995, 2005) : des mutations « totalement aveugles » aux mutations « totalement dirigées »</i>223	223
6.3. POUR LE « HASARD EVOLUTIONNAIRE » DE <i>TOUTE</i> MUTATION GENETIQUE228	228
6.3.1. <i>Trois distinctions en faveur du « hasard évolutionnaire » de toute mutation génétique</i>228	228
6.3.2. <i>La notion de « hasard évolutionnaire » dans le cas des mutations les plus controversées (Wright et al, 1999)</i>230	230
6.3.3. <i>La valeur conceptuelle et empirique de la notion de « hasard évolutionnaire »</i>234	234

CHAPITRE 7 – LE HASARD DES MUTATIONS GENETIQUES : DU NIVEAU EVOLUTIONNAIRE AU NIVEAU MOLECULAIRE.....241

7.1. LES NOTIONS D'« ALEATOIRE (AU SENS) FORT » ET D'« ALEATOIRE (AU SENS) FAIBLE »243	243
7.1.1. <i>Aucune mutation génétique n'est « aléatoire (au sens) fort »</i>248	248
7.1.2. <i>De nouveaux biais de mutation</i>251	251
7.1.3. <i>Toute mutation génétique est un événement « aléatoire (au sens) faible »</i>256	256
7.2. LA DERIVE GENETIQUE « ALEATOIRE (AU SENS) FAIBLE »259	259
7.2.1. <i>Qu'est-ce que la dérive génétique aléatoire ?</i>260	260
7.2.2. <i>La controverse autour de la dérive génétique aléatoire</i>262	262
7.2.3. <i>La dérive génétique aléatoire : un processus d'échantillonnage indiscriminé sans remplacement</i>265	265
7.3. LA SIGNIFICATION ET LE ROLE DE LA NOTION D'« ALEATOIRE (AU SENS) FAIBLE »267	267
7.4. LA SIGNIFICATION ET LE ROLE DE LA NOTION DE « HASARD EVOLUTIONNAIRE »270	270

PARTIE III – HASARD ET BRUIT DANS L'EXPRESSION DES GENES277

CHAPITRE 8 – LE BRUIT DANS L'EXPRESSION DES GENES : UNE INTRODUCTION281

8.1. L'ETUDE DU BRUIT EN SCIENCE281	281
8.2. L'ETUDE DU BRUIT EN BIOLOGIE.....282	282
8.2.1. <i>« La face noire de la variation phénotypique » : le bruit développemental</i>283	283
8.2.2. <i>L'étude du bruit au cœur de la cellule</i>288	288
8.3. UNE NOUVELLE APPROCHE POUR L'ETUDE DES PROCESSUS INTRACELLULAIRES290	290
8.4. L'ORIGINE ET LES CONSEQUENCES DU BRUIT DANS L'EXPRESSION GENETIQUE293	293
8.5. LES ETUDES EXPERIMENTALES DU BRUIT DANS L'EXPRESSION DES GENES.....297	297
8.6. LA MODELISATION STOCHASTIQUE DE L'EXPRESSION DES GENES.....302	302

CHAPITRE 9 – LE HASARD ET LE BRUIT DANS L’EXPRESSION GENETIQUE	309
9.1. LE BRUIT ET LA NOTION DE « HASARD EVOLUTIONNAIRE »	309
9.1.1. <i>Structures et mécanismes de contrôle et d’exploitation du bruit</i>	311
9.1.2. <i>La régulation adaptative du bruit</i>	315
9.1.3. <i>Le bruit dans l’expression des gènes est le fruit du « hasard évolutionnaire »</i>	319
9.2. LE BRUIT ET LA NOTION « ALEATOIRE (AU SENS) FAIBLE »	324
9.2.1. <i>Les limites des modèles de l’expression des gènes</i>	325
9.2.2. <i>Le bruit « aléatoire (au sens) faible » dans l’expression des gènes</i>	330
CHAPITRE 10 – POUR UNE INTERPRETATION OBJECTIVE DES PROBABILITES DANS LES MODELES STOCHASTIQUES DE L’EXPRESSION GENETIQUE	337
10.1. UNE ANALYSE PHILOSOPHIQUE DES PROBABILITES DANS UN CADRE OBJECTIF	337
10.2. L’ANALYSE DE ROBUSTESSE	340
10.2.1. <i>Trois analyses de robustesse</i>	342
10.2.2. <i>L’analyse de robustesse et les modèles stochastiques de l’expression des gènes</i>	344
10.2.3. <i>La condition de réalisabilité de l’analyse de robustesse</i>	345
10.3. L’OBJECTIVITE DES PROBABILITES ROBUSTES	350
10.3.1. <i>Vers une notion d’objectivité pour la science</i>	352
10.4. REFLEXIONS CONCLUSIVES SUR L’ANALYSE DE ROBUSTESSE	354
CONCLUSION	359
GLOSSAIRE	365
BIBLIOGRAPHIE	377
TABLE DES FIGURES	409
INDEX DES NOMS	411

Introduction

Qu'est-ce que le hasard ? S'agit-il d'une propriété objective des événements se produisant dans le monde ou simplement d'une manière de rendre compte de l'ignorance humaine quant aux causes des événements réels ? Quelle relation la notion de hasard entretient-elle avec les notions de déterminisme, d'indéterminisme et de probabilité ? Ces questions philosophiques anciennes, d'ordre aussi bien métaphysique qu'épistémologique, constituent le cadre général de notre réflexion sur la notion de hasard dans une discipline scientifique particulière – la biologie – et, plus spécifiquement, dans la caractérisation des phénomènes à l'origine causale de la variation au sein des populations d'organismes vivants.

Dans cette introduction, nous présentons d'abord la place que la notion de hasard occupe en biologie et les problèmes qu'elle soulève dans ce cadre scientifique. Dans un deuxième temps, nous concentrons notre attention sur l'objet de notre recherche, qui est la notion de hasard dans la caractérisation des phénomènes à l'origine de la variation biologique, plus exactement, les mutations génétiques et le bruit dans l'expression des gènes. Nous soulignons l'intérêt de notre enquête, d'une part, en montrant que la question du caractère « au hasard » des sources de la variation est une problématique ancienne, en biologie et dans la réflexion philosophique, et pose encore problème aujourd'hui ; d'autre part, en mettant en évidence les enjeux épistémologiques que la notion de hasard soulève dans l'étude de ces phénomènes biologiques.

1. La notion de hasard en biologie

La notion de hasard est largement utilisée en biologie, discipline dans laquelle elle révèle une richesse conceptuelle faisant appel à une analyse épistémologique. Cette notion fait partie intégrante de la biologie dans toutes ses branches, en particulier en biologie évolutionnaire : au sein de son hypothèse de l'évolution par sélection naturelle, Darwin fut le premier à donner une place centrale à la notion de hasard dans la caractérisation de la variation des organismes vivants.

L'importance de la notion de hasard en biologie réside dans l'objet de recherche propre à cette science. En effet, la biologie étudie un ensemble d'êtres complexes – les êtres vivants – et plus précisément leur structure, leurs fonctions, leurs niveaux d'organisation, leurs relations réciproques, leurs différences, leurs nombreuses interactions avec le milieu physique et, plus généralement, leur évolution. On peut caractériser les êtres vivants comme des systèmes « à histoire », à la fois ontogénétique et phylogénétique, qui diffèrent en fonction du cours contingent et singulier de leur évolution biologique. Plus spécifiquement, c'est la dimension historique propre à l'objet de la biologie qui implique un rôle central de la notion de hasard dans les différentes branches de cette science et qui, par conséquent, attire l'attention des épistémologues sur la question de sa signification et de son rôle dans ce cadre.

La place du hasard dans le monde biologique¹ et l'importance de la notion de hasard au sein des modèles et des théories en biologie font l'objet de réflexions et de controverses scientifiques et philosophiques depuis la deuxième moitié du XIX^e siècle, notamment depuis la publication de la première édition de *L'Origine des Espèces* (Darwin, 1859). Dans l'état actuel de la recherche, les biologistes sont convaincus que le hasard joue un rôle au sein des phénomènes biologiques, du niveau moléculaire à l'échelle macroévolutive ; par conséquent, ils assignent une place centrale à la notion de hasard dans les modèles et les théories qu'ils construisent. Le problème qui les occupe ne consiste pas à se demander si le hasard intervient dans le monde biologique et, si c'est bien le cas, quels phénomènes doivent être décrits et expliqués au moyen de la notion de hasard. Ils s'interrogent plutôt sur l'importance relative du hasard par rapport à d'autres facteurs biologiques et, par conséquent, sur le degré d'importance de la notion de hasard dans les modèles et les théories en biologie.

Le problème qui nous occupe en tant qu'épistémologue porte sur un point que, à notre avis, les biologistes ne considèrent pas suffisamment et que les philosophes ont déjà traité mais jamais de manière systématique et exhaustive : c'est la question de savoir quelle notion de hasard les biologistes utilisent au sein de leurs modèles et théories. Une fois cette notion identifiée, il est encore plus intéressant de s'interroger sur sa pertinence, en particulier par rapport aux phénomènes qu'elle caractérise, et d'analyser la pertinence d'autres notions possibles de hasard en biologie. Tous ces éléments invitent à une étude, à la fois descriptive et normative, de la manière de concevoir le hasard dans les diverses branches de la biologie. Notre travail de recherche représente une tentative d'analyse épistémologique de ces questions.

¹ Nous utilisons l'expression « monde biologique » pour désigner l'ensemble des phénomènes se produisant à l'échelle du vivant.

Une telle clarification conceptuelle de la notion de hasard vise d'abord à apporter un éclairage sur la manière dont les biologistes la conçoivent et sur les raisons qui les amènent à la mobiliser. En outre, cette analyse a pour but de comprendre comment les biologistes doivent concevoir la notion de hasard pour caractériser les phénomènes biologiques concernés. L'objectif principal d'un tel travail, qui vise à dissiper l'ambiguïté liée à l'emploi de la notion de hasard en biologie, est donc de fournir un cadre conceptuel à l'utilisation de la notion de hasard dans cette discipline scientifique. Son aboutissement ouvrira la voie à une réflexion ultérieure sur le hasard, non pas au sein des modèles et des théories que les biologistes construisent, mais dans les phénomènes biologiques mêmes.

Des réflexions philosophiques anciennes

Les diverses formes conceptuelles que la notion de hasard a connues au cours de l'histoire de la pensée philosophique et scientifique, de l'Antiquité jusqu'à aujourd'hui, constituent un point de repère fondamental de notre analyse. Nous les introduirons de manière critique dans la première partie de cette thèse. Les questions centrales de ces réflexions anciennes portent sur les points problématiques suivants. Se pose d'abord la question de savoir si la notion de hasard est définie en termes objectifs ou subjectifs : renvoie-t-elle à des propriétés objectives du monde ou à certaines caractéristiques de notre connaissance des événements réels ? Ensuite, se pose le problème de définir sa signification précise, c'est-à-dire d'identifier ce à quoi elle fait référence. En effet, nous verrons qu'au sein de toute conception du hasard, qu'il soit objectif ou subjectif, cette notion peut avoir des sens différents. Reste la question du rôle descriptif ou explicatif de la notion de hasard, dont la réponse est étroitement liée à sa signification et qui est relative au domaine et au contexte dans lesquels on l'emploie.

Une autre question centrale dans le cadre des réflexions philosophiques sur le hasard porte sur les liens que cette notion entretient avec d'autres notions semblables ou qui lui sont strictement liées. Ainsi, il serait sûrement intéressant d'éclaircir les points communs et les différences entre la notion de hasard et celles de contingence et de possibilité telles qu'elles ont été définies dans la logique modale à partir de *Les premiers analytiques* d'Aristote. Cependant, nous ne nous engagerons pas dans cette analyse car la détermination univoque de ces deux notions modales comporte de telles difficultés qu'elle devrait faire l'objet d'une réflexion systématique à part. De plus, à cause de cette ambiguïté, leur utilisation n'est jamais

un élément constitutif pour la biologie² et plus généralement pour aucune discipline scientifique.

Nous souhaitons plutôt nous occuper des liens entre la conception du hasard et d'autres notions. Dans la première partie de ce travail, nous examinerons tout d'abord la relation de compatibilité ou d'incompatibilité entre les diverses notions de hasard et les différentes manières de concevoir le déterminisme et l'indéterminisme (métaphysique, scientifique, méthodologique)³. Nous nous interrogerons aussi sur la notion de probabilité, en particulier sur la relation que la notion de hasard entretient avec les diverses interprétations (épistémiques et objectives) des probabilités⁴. Ces dernières semblent entretenir une relation privilégiée, par rapport à d'autres notions, avec la notion de hasard : cela constitue l'une des hypothèses directrices de la présente recherche.⁵

Notre travail de clarification conceptuelle de la notion de hasard trouve plus spécifiquement sa place en tant que prolongement d'une longue tradition de réflexion consacrée au hasard dans les phénomènes biologiques et à la notion de hasard dans les descriptions de ces phénomènes. En effet, cette réflexion a vu le jour lors de la publication de *L'Origine des Espèces* (1859) de Darwin. Elle a ensuite trouvé une place au sein des recherches des généticiens et des naturalistes au début du XX^e siècle, puis dans celles des

² De nombreux biologistes et philosophes de la biologie ont réfléchi sur la contingence propre aux phénomènes biologiques, en particulier du processus d'évolution. Nous renvoyons à Gould, 1989[1991], *La vie est belle. Les surprises de l'évolution* ; Beatty, 2008, « Chance Variation and Evolutionary Contingency: Darwin, Simpson (*The Simpsons*) and Gould » ; Gayon, 1994, « Le hasard dans la théorie évolutionniste moderne : une analyse philosophique » ; 2005, « Chance, Explanation, and Causation in Evolutionary Theory » ; Szathmari, 2005, « Path dependence and historical contingency in biology » ; Oyama *et al* (eds.), 2001, *Cycles of Contingency. Developmental Systems and Evolution*. Dans le cadre du débat sur la possibilité de lois en biologie, le caractère contingent des généralisations produites dans cette discipline scientifique a aussi fait l'objet de nombreuses réflexions philosophiques. Nous renvoyons à Smart, 1959, « Can biology be an exact science? » ; Hull, 1974, *The Philosophy of Biological Science* ; Beatty, 1995, « The Evolutionary Contingency Thesis » ; Brandon, 1997a, « Does Biology has Laws? The experimental evidence » ; Rosenberg, 1994, *Instrumental Biology or the Disunity of Science* ; Carrier, 1995, « Evolutionary Change and Lawlikeness: Beatty on Biological Generalizations » ; Gayon, 1993, « La biologie entre loi et histoire » ; 2004, « Realism and Biological Science ».

³ Nous renvoyons aux chapitres 1 et 2.

⁴ Nous renvoyons au chapitre 2.

⁵ L'analyse des notions de hasard, de déterminisme, d'indéterminisme et de probabilité et de leurs relations ont fait l'objet de plusieurs débats en philosophie des sciences, en particulier relativement aux disciplines physiques et biologiques. En ce qui concerne la relation de compatibilité et d'incompatibilité entre déterminisme, indéterminisme, et interprétations des probabilités, nous renvoyons à Von Plato, 1982, « Probability and Determinism » ; Lewis, 1986, « A Subjectivist's Guide to Objective Chance » ; Popper, 1992, *Quantum Theory and the Schism in Physics* ; Loewer, 2001, « Determinism and Chance » ; Hoefer, 2007, « The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance » ; Schaffer, 2007, « Deterministic Chance? ». Quant à la caractérisation des notions de déterminisme, indéterminisme, hasard et probabilités et les relations entre elles, en particulier en physique, nous renvoyons à Earman, 1986, *A Primer on Determinism* ; Bishop, 2003, « On Separating Prediction from Determinism » ; 2006, « Determinism and Indeterminism » ; Krüger *et al* (eds.), 1987, *The Probabilistic Revolution* ; Gigerenzer *et al* (eds.), 1989, *The Empire of Chance: How Probability Changed Science and Everyday Life* ; Pomian (dir.), 1990, *La querelle du déterminisme. Philosophie de la science d'aujourd'hui*.

biologistes de la Synthèse Moderne⁶ (1930-1950), des biologistes moléculaires (à partir des années 1950), et aussi au sein des réflexions de plusieurs philosophes qui se sont intéressés à la biologie.⁷

Malgré cette richesse bibliographique en biologie et en philosophie, nous trouvons qu'il manque actuellement une analyse qui prenne en considération les résultats de toutes ces réflexions et les synthétise afin d'aboutir à des conclusions précises et solides, en particulier quant à la conceptualisation du hasard en biologie. Nous avons voulu par ce travail fournir une analyse de ce genre, même si nous ne nous occuperons ni du rôle du hasard en tant que facteur intervenant dans le monde biologique, ni de la notion de hasard dans toute discipline et quant à tout phénomène biologique. Nous nous proposons un objectif plus limité mais non moins ambitieux : nous limiterons notre analyse épistémologique aux phénomènes à l'origine causale de la variation individuelle, génétique et phénotypique, au sein des populations d'organismes vivants. Notre but est d'apporter une clarification conceptuelle de la notion de hasard dans ce cadre et, plus précisément, de fournir une réponse critique et actuelle à la question de sa signification et de son rôle dans la caractérisation de ces phénomènes biologiques.

⁶ Huxley introduisit l'expression « Synthèse Moderne » en 1942.

⁷ Du côté des biologistes, nous renvoyons, entre autres, à Darwin 1859, *On The Origin of Species By Means of Natural Selection or The Preservation of Favoured Races in The Struggle For Life* ; 1868, *The Variation of Animals and Plants Under Domestication* ; Weismann 1893[1892], *The Germ Plasm. A Theory of Heredity* ; Mendel, 1866[1865], « Experiments in Plant Hybridization » ; Simpson, 1953, *The Major Features of Evolution* ; 1984[1944], *Tempo and Mode in Evolution* ; Dobzhansky, 1951[1937], *Genetics and The Origin of Species* ; 1970, *Genetics of the Evolutionary Process* ; 1974, « Chance and Creativity in Evolution » ; Mayr, 1974[1970], *Populations, espèces et évolution* ; Monod, 1970, *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne* ; 1974, « On Chance and Necessity » ; Jacob, 1981, *Le jeu des possibles* ; Eble, 1999, « On the dual nature of chance in evolutionary biology and palaeology » ; Gould & Lewontin, 1979, « The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme » ; Lewontin, 2000, *The Triple Helix: Gene, Organism, and Environment*. Du côté des philosophes, nous renvoyons, entre autres, à Wimsatt, 1980, « Randomness and perceived-randomness in evolutionary biology » ; Sober, 1984, *The Nature of Selection. Evolutionary Theory in Philosophical Focus* ; Beatty, 1984, « Chance and Natural Selection » ; 1987, « The probabilistic revolution in evolutionary biology. An overview » ; 2006, « Chance Variation: Darwin on Orchids » ; 2008, « Chance Variation and Evolutionary Contingency: Darwin, Simpson (*The Simpsons*) and Gould » ; Delsol, 1989, *Cause, Loi, Hasard en Biologie* ; Lenay, 1989, *Enquête sur le hasard dans les grandes théories biologiques de la deuxième moitié du XIX^e siècle* ; Gayon, 1994, « Le hasard dans la théorie évolutionniste moderne : une analyse philosophique » ; Millstein, 1997, *The Chances of Evolution: An Analysis of the Roles of Chance in Microevolution and Macroevolution* ; 2000, « Chance and Macroevolution » ; 2006, « Discussion of 'Four Case Studies on Chance in Evolution: Philosophical Themes and Questions' » ; Weber, 2001, « Determinism, Realism, and Probability in Evolutionary Theory » ; Shanahan, 1991, « Chance as an Explanatory Factor in Evolutionary Biology ».

Esquisse topographique de la notion de hasard en biologie

L'une des principales raisons motivant l'entrée de la notion de hasard en biologie est le phénomène de mutation génétique⁸. En effet, cette source causale de variation biologique – que Darwin ne put pas connaître mais qu'il chercha à comprendre et dont il observa les effets en termes de variation phénotypique individuelle au sein des populations naturelles – est habituellement caractérisée comme un événement ou un processus « au hasard » ou « aléatoire ». Au cours du temps, grâce aux développements des recherches en génétique (à partir du début du XX^e siècle), à la naissance de la biologie moléculaire (habituellement fixée à l'année 1953, date de la découverte de la structure de l'ADN en tant que support de l'information génétique) et aux avancées dans cette dernière discipline, la notion de hasard a commencé à être utilisée pour caractériser d'autres sources de la variation biologique, par exemple la recombinaison⁹ et la transposition¹⁰. Plus récemment, durant les vingt dernières années, les biologistes ont commencé à employer la notion de hasard, souvent de manière implicite, par rapport à certains processus moléculaires impliqués dans le développement. Plus précisément, ils mobilisent cette notion afin de rendre compte des fluctuations stochastiques (ou bruit) qui affectent les réactions biochimiques se produisant au cours de l'expression des gènes¹¹. Les biologistes ont montré que ce bruit, en perturbant le déroulement du développement des organismes vivants, représente une autre source de variation au sein des populations naturelles.

La notion de hasard est aussi utilisée, entre autres, dans la caractérisation des phénomènes biologiques suivants. i) L'hérédité génétique chez les organismes eucaryotes¹² à reproduction sexuée, en particulier l'échange de fragments de matériel génétique (le phénomène de *crossing over*)¹³ entre chromosomes homologues (appariés) pendant le processus de formation des gamètes (méiose)¹⁴; la ségrégation indépendante des

⁸ Une mutation génétique est une altération de la séquence nucléotidique de l'ADN qui n'est pas due au processus de recombinaison.

⁹ La recombinaison est l'échange de fragments de matériel génétique entre deux brins d'ADN appariés.

¹⁰ La transposition est le processus par lequel des séquences d'ADN plus ou moins longues changent de position sur le chromosome.

¹¹ L'expression génétique est le processus biochimique par lequel des protéines diverses sont produites à partir de la séquence d'ADN présente à l'intérieur de la cellule.

¹² Les organismes eucaryotes sont constitués de cellules avec un nucleus (cellules eucaryotes). La plupart des organismes pluricellulaires sont eucaryotes.

¹³ En 1916, Morgan avança l'hypothèse du *crossing over* entre chromosomes homologues sur la base de ses études expérimentales des mutations héréditaires chez la *Drosophila*. En 1931, Creighton et McClintock montrèrent le mécanisme physique à la base de ce phénomène chromosomique chez le maïs.

¹⁴ La méiose consiste en une suite de deux divisions cellulaires donnant des cellules aploïdes (avec n chromosomes) à partir de cellules diploïdes (avec $2n$ chromosomes). Chez les organismes diploïdes (constitués de cellules avec $2n$ chromosomes) à reproduction sexuée, la méiose se déroule pendant la gamétogenèse,

chromosomes (des caractères, selon l'expression de Mendel)¹⁵ durant la méiose ; la rencontre des gamètes mâle et femelle qui en fusionnant (la fécondation) forment une nouvelle cellule diploïde (le zygote)¹⁶. ii) La dérive génétique aléatoire, expression par laquelle les biologistes désignent des phénomènes divers, à des échelles différentes (microscopiques et macroscopiques), mais qui correspondent tous à des fluctuations d'échantillonnage et sont à l'origine du changement de fréquences géniques au sein des populations naturelles.¹⁷ iii) Les événements divers (et imprédictibles) se produisant au niveau des écosystèmes, et même au niveau géologique, et étant à l'origine d'événements de spéciation¹⁸ et d'extinctions¹⁹.

Ce travail de recherche ne pourra certes nous permettre d'identifier la notion de hasard invoquée dans la caractérisation de chacun de ces phénomènes biologiques ; il pourra plutôt contribuer, dans la mesure de nos compétences et relativement à nos objets d'analyse, à faire progresser la réflexion philosophique sur la place occupée par cette notion en biologie.

Une pluralité de termes et de notions de hasard

Malgré l'utilisation assez répandue de la notion de hasard dans toute discipline biologique, en particulier quant aux phénomènes présentés ci-dessus, le terme « hasard » n'apparaît pas très souvent en biologie dans les articles techniques, théoriques et expérimentaux, probablement à cause du caractère flou et imprécis de ses significations possibles²⁰. Dans le cadre des descriptions verbales des phénomènes biologiques considérés comme des événements régis, au moins en partie, par le hasard, les biologistes semblent préférer d'autres termes, soit aussi vagues et ambigus que le terme « hasard » (« aveugle », « accidentel », « fortuit »), soit un peu plus précis mais toujours susceptibles d'interprétations diverses (« imprédictible », « incertain », « indéterministe »), soit des termes techniques et, pour cela, apparemment plus pertinents (« statistique », « probabiliste », « aléatoire », « stochastique »). Quel que soit le terme qu'ils emploient, les biologistes ne spécifient que

immédiatement avant la fécondation, alors que chez les organismes aploïdes (constitués de cellules avec n chromosomes), elle se produit après la fécondation.

¹⁵ En 1900, de Vries, Correns et Tschermak redécouvrirent de façon indépendante les lois de l'hérédité que Mendel avait déjà formulées auparavant.

¹⁶ La fécondation fut observée et décrite pour la première fois en 1854 par Thuret chez l'algue brune *Fucus*.

¹⁷ Pour une introduction à la dérive génétique aléatoire, voir Gillespie, 2004, *Population Genetics: A Concise Guide* ; Beatty, 1992, « Random Drift ». Nous renvoyons aussi au chapitre 7 du présent travail.

¹⁸ Voir, par exemple, Lamotte, 1988, « Phénomènes fortuits et évolution ».

¹⁹ Gould & Eldredge, 1972, « Punctuated Equilibria: an Alternative to Phyletic Gradualism » ; Gould, 1989[1991], *La vie est belle. Les surprises de l'évolution*.

²⁰ En revanche, le terme « hasard » est employé massivement dans les ouvrages et les articles de vulgarisation des sciences biologiques.

rarement sa signification exacte. Cependant, dans les publications de leurs études, ils donnent l'impression de considérer tous les termes ci-dessus comme des synonymes renvoyant à une même notion, celle de hasard.

De plus, ces termes ne semblent pas désigner une notion univoque en biologie, mais plutôt renvoyer à des notions avec des significations et des rôles différents selon la discipline, selon le phénomène biologique considéré, et même selon la perspective sous laquelle ce dernier est décrit. Par exemple, dans la deuxième partie de ce travail, nous verrons que deux notions de hasard, qui n'ont pas la même signification, peuvent être utilisées en même temps pour caractériser les mutations génétiques. D'une part, depuis Darwin, le terme « hasard » fait référence à une notion désignant le caractère non dirigé des mutations (en termes darwiniens, de la variation) par rapport au besoin adaptatif des organismes porteurs. D'autre part, ce même terme peut renvoyer à une notion de hasard qui caractérise les mutations génétiques au niveau moléculaire, relativement aux processus physico-chimiques qui les causent et sans aucune référence à leur utilité du point de vue de l'évolution. Cet exemple nous montre que le terme « hasard », et les autres termes que les biologistes utilisent dans la caractérisation de certains phénomènes biologiques, ne renvoient pas forcément à une même notion de hasard et, par conséquent, ne sont pas nécessairement synonymes. Une telle pluralité terminologique semble donc être la manifestation d'une variété conceptuelle. C'est là l'objet de notre analyse épistémologique : nous souhaitons poser la question de savoir quels sont la signification et le rôle des diverses notions de hasard en biologie, en particulier lorsqu'elles sont utilisées pour caractériser les phénomènes à l'origine causale de la variation.

Les notions de hasard et les probabilités en biologie

En outre, il est intéressant de remarquer que les biologistes utilisent des modèles probabilistes pour décrire, prédire et expliquer les phénomènes qu'ils désignent en tant que phénomènes « au hasard » : le caractère probabiliste de ces formulations mathématiques nous amène encore une fois à avancer l'idée d'une relation privilégiée entre les probabilités et la notion – ou les notions – de hasard.²¹

Plus précisément, la large utilisation de la modélisation probabiliste en biologie peut soulever deux hypothèses quant à la relation entre les notions de hasard et de probabilité et

²¹ Nous rappelons que la nature de ce rapport fera l'objet d'une réflexion critique dans le cadre de cette recherche.

quant à leurs significations et à leurs rôles respectifs. D'une part, ces deux notions pourraient ne pas avoir de relation d'identité (c'est-à-dire que ces deux notions n'auraient pas la même signification), ni de relation privilégiée entre elles. Dans ce cas, les probabilités ne pourraient être considérées ni comme l'expression mathématico-formelle de la notion de hasard, quelle que soit la manière dont cette dernière est conçue, ni comme un indice fort de son utilisation dans la caractérisation du phénomène biologique considéré. On devrait donc rendre compte de l'utilisation de modèles probabilistes pour représenter des phénomènes biologiques indépendamment de la notion de hasard.

D'autre part, les notions de hasard et de probabilité pourraient entretenir une relation privilégiée et même une relation d'identité. Par conséquent, elles pourraient désigner, bien que différemment, les mêmes caractéristiques d'un phénomène biologique. Dans ce cas, les probabilités pourraient être considérées comme l'expression mathématico-formelle de la notion de hasard ou comme un indice fiable de son utilisation dans la caractérisation du phénomène considéré.

Si la première hypothèse se révélait être la bonne, la question se poserait de savoir si les probabilités en biologie sont utilisées de manière complètement indépendante par rapport à la mobilisation de la notion de hasard ou si, au contraire, elles entretiennent une relation avec au moins une des notions possibles de hasard. Au contraire, si la deuxième hypothèse se révélait être la bonne, il serait intéressant d'identifier les critères pour déterminer, à partir de l'utilisation de modèles probabilistes, la notion de hasard qui est en jeu dans la caractérisation des phénomènes biologiques considérés et qui, donc, entretiendrait une relation d'identité ou, du moins, privilégiée avec les probabilités figurant dans les modèles de ces phénomènes²².

Notre enquête épistémologique sur la conception du hasard en biologie comprendra donc une analyse de la relation entre les notions de hasard et de probabilité : elle nous amènera à nous demander quelle est l'interprétation appropriée des probabilités dans les descriptions formelles de certains phénomènes biologiques (notamment, les sources de la variation)²³. Nous consacrerons le dernier chapitre de notre thèse à l'analyse de cette question.

²² Pour une analyse de la relation entre hasard et probabilités en biologie, nous renvoyons à Sober, 1984, *The Nature of Selection. Evolutionary Theory in Philosophical Focus* (en particulier, le chapitre 4).

²³ Pour une introduction aux interprétations des probabilités, nous renvoyons au chapitre 2 de ce travail.

2. La notion de hasard et les sources de la variation biologique

L'étude des phénomènes à l'origine causale de la variation biologique représente une voie privilégiée de notre enquête sur la notion de hasard en biologie. En effet, comme nous l'avons déjà remarqué plus haut, le terme « hasard » et tous les termes utilisés par les biologistes apparemment en tant que synonymes entrent en jeu avant tout dans la caractérisation des causes de la variation qu'on peut observer au niveau populationnel. Cette richesse terminologique semble être la manifestation d'une pluralité conceptuelle sous-jacente. C'est là la deuxième hypothèse directrice de notre enquête.

Une deuxième raison faisant des sources de la variation un objet privilégié d'analyse réside dans le fait que le caractère « au hasard » des ces phénomènes biologiques est au centre d'un débat controversé, qui a commencé au XIX^e siècle et qui est encore ouvert aujourd'hui. Notamment, le sens précis de la notion de hasard, que les biologistes évoquent largement et souvent de manière non critique dans la caractérisation des sources de la variation, a eu beaucoup d'échos pendant les trente dernières années, à la fois dans la réflexion philosophique et en biologie.

Nous avons choisi de concentrer notre analyse sur la signification et le rôle de la notion de hasard dans la caractérisation des mutations génétiques (sources de variation génétique et phénotypique) et du bruit dans l'expression des gènes (source de variation phénotypique). Avant de montrer les enjeux épistémologiques que ces deux déterminants causaux de la variation soulèvent quant à la notion de hasard, nous souhaitons présenter ici quelques éléments historiques au sujet des sources de la variation biologique, du XIX^e siècle jusqu'à aujourd'hui.²⁴ Par ce récit, nous visons à mettre en évidence que l'utilisation de la notion de hasard dans la caractérisation des sources de la variation est une problématique ancienne et centrale en biologie.

La variation biologique

Dans la nature il n'existe pas deux organismes individuels parfaitement identiques du point de vue à la fois génétique et phénotypique : la variation est une caractéristique de tout groupe d'organismes vivants et sa présence est l'une des trois conditions pour qu'il puisse y

²⁴ Pour plus de détails, nous renvoyons à la deuxième partie de ce travail.

avoir évolution par sélection naturelle dans le monde biologique²⁵. C'est pourquoi il s'agit d'un sujet central en biologie, en particulier en biologie évolutionnaire, à la fois du point de vue conceptuel et historique. Il est important de souligner que la variation est une notion qui s'applique non pas au niveau de chaque organisme individuel mais à l'échelle d'un groupe d'individus (par exemple, une population) ou d'un ensemble de populations²⁶ : la variation dite « individuelle » concerne « les différences des individus d'une population particulière » ; la variation dite « de groupe » « se rapporte aux différences constatées entre les populations ».²⁷

En biologie, le terme « variation » renvoie conventionnellement, à l'époque moderne, au fait de la diversité des caractères phénotypiques et génétiques présents au sein d'un groupe d'organismes vivants (par exemple, au sein d'une population ou d'un ensemble de populations). L'évolution peut être représentée très schématiquement comme un processus en deux étapes : d'abord, la production de la variation dans un groupe d'organismes individuels ; ensuite, le changement dans le temps, d'une génération à la suivante, des proportions des différents variants au sein du groupe. Par conséquent, la variation pouvant jouer un rôle dans l'évolution est la variation héréditaire qui, selon le paradigme de la Théorie Synthétique de l'Evolution, n'est que génétique²⁸.

On peut distinguer deux sortes de variation relativement aux caractéristiques des organismes vivants concernés : la variation génétique et la variation phénotypique. Le génotype d'un organisme individuel (ou d'un groupe d'individus isogéniques) consiste en sa constitution génétique, autrement dit en la séquence nucléotidique de son ADN (ou en sa

²⁵ Lewontin, 1970, « The Units of Selection ». Nous renvoyons au chapitre 3.

²⁶ Nous utilisons le terme « population » pour désigner un ensemble d'individus interféconds vivant dans un endroit donné. Voir Mayr, 1974[1970], *Populations, espèces et évolution*.

²⁷ *Ibid.*, p. 96. Cependant, nous verrons qu'on peut parler de variation relativement à un organisme individuel si on le considère tout au long de son développement. Par exemple, nous verrons que le bruit intervenant dans l'expression des gènes peut être une source de variation phénotypique chez un même individu au cours du temps.

²⁸ Cependant, des recherches relativement récentes dans diverses disciplines biologiques (entre autres, la biologie moléculaire, la biologie du développement, l'écologie, la biologie du comportement) ont montré que certaines variations dites « épigénétiques » au sens large, car elles ne concernent pas le matériel génétique, sont héréditaires et par conséquent jouent un rôle central dans l'évolution (Jablonka & Lamb, 1995, *Epigenetic Inheritance and Evolution* ; 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life* ; Laland *et al*, 2003, *Niche Construction: The Neglected Process in Evolution* ; West-Eberhard, 2003, *Developmental Plasticity and Evolution*). L'importance de ces variations du point de vue évolutif fait l'objet d'un débat entre biologistes et entre philosophes. Parmi ces derniers, les partisans d'une nouvelle perspective philosophique quant au développement et à l'évolution biologique, la Théorie des Systèmes en Développement (DST), ont élargi le concept d'hérédité jusqu'à y englober les caractéristiques du milieu où les organismes habitent, et ont insisté sur l'importance évolutive de tout type de variation héréditaire – selon leur définition – dans les organismes vivants. (Griffiths & Gray, 1994, « Developmental Systems and Evolutionary Explanation » ; 2004, « The Developmental Systems Perspective. Organism-environment systems as units of development and évolution » ; Oyama *et al* (eds.), 2001, *Cycles of Contingency. Developmental Systems and Evolution*). Pour une analyse critique de la DST, en particulier du concept d'hérédité étendue qu'il propose, voir Merlin, à paraître, « On Griffiths and Gray's concept of expanded and diffused inheritance ».

combinaison de gènes)²⁹. Le phénotype d'un organisme individuel (ou d'un groupe d'organismes semblables et se différenciant d'autres groupes) peut être défini comme la totalité de ses caractéristiques biochimiques, physiologiques, morphologiques, et comportementales. Deux individus avec deux génotypes différents diffèrent souvent phénotypiquement ; inversement, deux individus avec deux phénotypes différents diffèrent souvent génotypiquement. Cependant, un même génotype peut donner lieu à des phénotypes divers et un même phénotype peut être dû à des génotypes différents. Les variations génétique et phénotypique sont donc, respectivement, la diversité des séquences nucléotidiques d'ADN (ou des combinaisons de gènes) et la diversité des caractéristiques biochimiques, physiologiques, morphologiques et comportementales au sein d'un groupe d'organismes individuels (par exemple, une population)³⁰ ou au sein d'un ensemble de populations.

À la recherche des sources de la variation

La question des sources de la variation date de la fin du XVIII^e siècle - début du XIX^e siècle, lorsque l'importance de la variation de n'importe quelle caractéristique d'un organisme³¹ commença à être reconnue³². Cependant, les hypothèses avancées à cette époque pour rendre compte de la variation biologique correspondaient toujours à des idées assez vagues, bien loin de descriptions et d'explications précises des processus causaux à l'origine de la variation phénotypique, et encore moins génétique, au sein des populations naturelles.

Darwin³³ fit des causes de la variation – qu'il observa entre espèces, entre populations d'une même espèce, et entre individus d'une même population – un problème central en biologie. Bien qu'il ait proposé quelques hypothèses quant aux sources de la diversité des caractères parmi les organismes vivants (par exemple, les influences environnementales sur le système reproducteur des parents ; les effets directs de l'environnement sur les caractères des organismes individuels; les effets de l'usage et du non-usage des organes), il n'est jamais

²⁹ Les biologistes utilisent l'expression « pool génétique » (par exemple, en référence à une population) pour désigner la totalité des gènes d'un groupe donné existant à un moment donné (Mayr, 1974[1970], *op. cit.*).

³⁰ Futuyma, 1998[1979], *Evolutionary Biology*.

³¹ À cette époque, la distinction entre phénotype et génotype n'avait pas encore été proposée. Elle fut avancée bien plus tard, au début du XX^e siècle, par le généticien danois Johannsen. Par conséquent, de la fin du XVII^e jusqu'au début du XX^e, les biologistes ne purent observer et essayer d'expliquer que la variation qu'on appelle aujourd'hui phénotypique (la variation dans les apparences des organismes vivants), par exemple, la diversité des caractères physiologiques, morphologiques et comportementaux au sein d'une population naturelle.

³² Mayr, 1989[1982], *Histoire de la biologie. De Darwin à nos jours*, p. 879.

³³ Darwin, 1859, *On The Origin of Species By Means of Natural Selection or The Preservation of Favoured Races in The Struggle For Life* (du chapitre I au chapitre V).

parvenu à en fournir une description et une explication précises : les causes de la variation biologique restèrent une « boîte noire » dans la plus grande partie de son œuvre.

Il faut cependant souligner que Darwin, en niant le caractère téléologique des causes de la variation, émit une idée très influente au cours des années : la variation est le résultat du hasard au sens où elle n'est pas produite en réponse aux besoins adaptatifs des organismes concernés. De cette manière, il s'opposa aux visions qui l'avaient précédé à partir de la fin du XVIII^e siècle. En effet, Darwin prit le contre-pied de la pensée typologique et fixiste, selon laquelle la variation individuelle n'avait pas d'importance, même si des déviations de la norme propre à chaque espèce étaient admises et pouvaient se produire à cause des changements de l'environnement. En outre, il s'opposa aux transformistes du début du XIX^e siècle, selon lesquels la variation était caractérisée soit par une direction adaptative due aux besoins des organismes par rapport à leur environnement (Lamarck), soit par une direction linéaire, régulière mais non adaptative, due à une prédisposition inhérente au développement des organismes (les théories orthogénétiques).

De la fin du XIX^e siècle jusqu'à la Synthèse Moderne (1930-1950), les idées téléologiques lamarckiennes et orthogénétiques trouvèrent une suite chez les partisans de diverses théories néo-lamarckiennes de l'évolution.³⁴ Pour les néo-lamarckiens les plus fidèles à la pensée de Lamarck, la variation étaient produite en réponse aux changements de l'environnement et en vue de l'adaptation des organismes concernés ; pour Darwin, au contraire, la variation n'était produite en vue d'aucune fin.

La théorie darwinienne de l'évolution et, en particulier, la caractérisation non téléologique des causes de la variation trouvèrent beaucoup de partisans dans les années 1880 (Weismann, les naturalistes). Après une période de crise dans les années 1890, les idées de Darwin devinrent l'objet d'un consensus et d'une reformulation importante par les biologistes de la Synthèse Moderne. Cette dernière, qui surgit de l'articulation entre la génétique classique mendélienne et l'évolutionnisme darwinien, se différença des théories néo-lamarckiennes en ce qui concerne le caractère téléologique des sources de la variation. De plus, contrairement aux idées de Darwin, elle nia tout phénomène d'hérédité flexible³⁵, cette dernière ayant été réfutée pour la première fois par Weismann dans les années 1880³⁶.

³⁴ Bowler, 1983, *The Eclipse of Darwinism. Anti-Darwinian Evolution Theories in the Decades around 1900*.

³⁵ L'hérédité flexible correspond à l'idée selon laquelle des modifications que les organismes individuels ont acquises au cours de leur développement, sous l'influence de l'environnement, puissent être transmises à leurs descendants.

³⁶ Weismann, 1892[1883], « De l'Hérédité ». Lenay (2003, « Les limites naturelles de la durée de vie et la question de l'hérédité de l'acquis ») montre que l'argument théorique de Weismann contre l'hérédité des caractères acquis émergea de ses réflexions sur les causes de la mort par vieillissement.

La Synthèse Moderne fournit une explication encore incomplète mais assez satisfaisante de la variation biologique³⁷. Cependant, seules les découvertes de la chimie du matériel génétique (1944) et de la structure de l'ADN en tant que support matériel de l'information génétique (1953) ainsi que la naissance de la biologie moléculaire, permirent d'identifier les processus moléculaires à l'origine causale de la variation, en premier lieu génétique, et donc de donner une explication causale à la variation individuelle au sein des populations.

Au cours des années, quelques aspects du paradigme de la Synthèse Moderne (notamment, le principe que toutes les mutations génétiques se produisent « au hasard ») ont été mis en cause par les avancées de la recherche en biologie et sont encore au centre de controverses parmi les biologistes et les philosophes. Nous analyserons ces failles dans le consensus au sujet de la Théorie Synthétique de l'Evolution. Nous consacrerons en particulier une grande partie de ce travail (la deuxième) à l'analyse du débat actuel sur la caractérisation des sources de la variation par la notion de hasard : nous souhaitons montrer que le vrai objet de consensus au sein de la communauté des biologistes n'est pas la Synthèse Moderne dans sa totalité, mais ses aspects les plus fidèles à la pensée de Darwin.

Les sources de la variation

Quelles sont donc les sources de la variation individuelle, génétique et phénotypique, qu'on peut observer au sein des populations naturelles ? Dans l'état actuel des recherches, les biologistes détiennent des descriptions et des explications assez précises des phénomènes à l'origine de la variation biologique. Les causes possibles de la variation phénotypique sont les suivantes : les différences génotypiques – et donc les phénomènes à l'origine causale de la variation génétique ; les différences non génétiques ou épigénétiques au sens large³⁸.

Les causes principales de la variation génétique sont la mutation et la recombinaison. Au sein de ces deux phénomènes à l'origine de la variation génétique – et par conséquent,

³⁷ Nous renvoyons à Fisher, 1999[1930], *The Genetical Theory of Natural Selection: a complete variorum edition* ; Haldane, 1966[1932], *The Causes of Evolution* ; Dobzhansky, 1951[1937], *Genetics and The Origin of Species* ; Mayr, 1942, *Systematics and the Origin of Species* ; Huxley, 1948[1942], *Evolution. The Modern Synthesis* ; Simpson, 1953, *The Major Features of Evolution*.

³⁸ Par exemple, les différences dans les *patterns* de méthylation de l'ADN provoquant de la variation dans l'expression des gènes ; les différences au cours du développement des organismes individuels comme les fluctuations stochastiques (ou bruit) dans l'expression des gènes ; les différences dans les influences maternelles et paternelles au cours du développement de la progéniture ; les différences dans les conditions environnementales. Pour une présentation détaillée de sources possibles de la variation phénotypique d'origine génétique ou non génétique voir Futuyma, 1998[1979], *Evolutionary Biology*; Ridley, 2004, *Evolution*.

même si ce n'est pas de façon nécessaire, sources de variation phénotypique³⁹ –, on peut distinguer différents types de mutations et de recombinaisons. Dans la deuxième partie, au chapitre 3, nous présenterons la manière dont les biologistes distinguent et définissent tous les types de mutations génétiques.

Il est intéressant de souligner les différences entre les phénomènes de mutation et de recombinaison afin de comprendre leur potentiel respectif en tant que sources causales de la variation génétique et du point de vue de l'évolution. En effet, alors qu'une modification du matériel génétique produite par tout type de mutation consiste en l'introduction d'une nouveauté – favorable, défavorable ou neutre – dans la séquence nucléotidique de l'ADN concernée, la recombinaison remanie le matériel génétique préexistant et produit des combinaisons génétiques – pouvant aussi être favorables, défavorables ou neutres – nouvelles. Bien que dans ce travail nous ayons décidé d'analyser le phénomène de mutation, et non pas la recombinaison, nous souhaitons souligner que cette dernière joue aussi un rôle fondamental en tant que source de variation. En effet, en brassant le matériel génétique, elle peut arriver à séparer les changements favorables des changements défavorables, tous deux produits par des mutations, de manière à ce qu'ils puissent être soumis séparément au crible de la sélection naturelle. Cela permet d'éliminer les allèles défavorables du pool génétique d'une population sans perdre l'avantage apporté par d'autres allèles⁴⁰.

Lorsqu'on parcourt la littérature sur les causes de la variation au sein des populations naturelles, des travaux de Darwin jusqu'aux articles les plus récents sur le sujet, on constate de façon évidente l'emploi du terme « hasard » et l'utilisation encore plus fréquente des termes « aveugle », « accidentel », « fortuit », « imprédictible », « incertain », « indéterministe », « statistique », « probabiliste », « aléatoire », « stochastique ». Ces choix terminologiques sont particulièrement frappants dans les études sur les mutations génétiques et, depuis quelques décennies, dans les publications sur le bruit dans l'expression des gènes. Ainsi, ces deux sources de variation représentent un objet de réflexion particulièrement riche et intéressant dans le cadre d'une enquête sur la notion de hasard en biologie.

³⁹ En effet, contrairement à la célèbre hypothèse « un gène – une enzyme » tirée des premiers résultats en biologie moléculaire, la relation entre le génotype et le phénotype est très complexe, influencée par beaucoup de facteurs extra-génotypiques et extra-phénotypiques (environnementaux au sens large) du niveau microscopique au niveau macroscopique, et ne se réduit pas à une relation de « un-à-un », mais plutôt de « plusieurs-à-plusieurs ». Voir Hallgrímsson & Hall (eds.), 2005, *Variation. A Central Concept in Biology*.

⁴⁰ Nous renvoyons à l'analyse à la fois biologique et philosophique de la recombinaison par Jacob, 1981, *Le jeu des possibles*.

Hasard, mutations génétiques et bruit dans l'expression des gènes

Pour mener à bien notre enquête, nous nous proposons de poursuivre deux sous-projets, l'un d'ordre descriptif, l'autre d'ordre normatif, qui sont strictement liés. C'est pourquoi nous ne les traiterons pas séparément. Plutôt, tout au long de notre analyse, nous les poursuivrons parallèlement car ils portent sur le même problème – celui de la notion de hasard – en suivant deux approches différentes mais strictement entremêlées et se recoupant souvent.

Le premier sous-projet d'ordre descriptif consiste en l'observation et la description de la conceptualisation du hasard dans l'étude théorique et expérimentale des mutations génétiques et du bruit dans l'expression des gènes. Autrement dit, nous voulons comprendre la façon dont les biologistes conçoivent la notion de « hasard » à laquelle la multiplicité de termes qu'ils emploient semble renvoyer. Il s'agit donc de poser des questions à réponse descriptive même s'il s'agit d'une réponse nécessairement basée sur une analyse conceptuelle préalable.

Parallèlement à ce sous-projet épistémologique, nous nous engagerons dans un deuxième sous-projet d'ordre normatif qui, en réalité, représente le but principal de notre travail. Il consiste à identifier, analyser et définir une ou plusieurs notions de hasard appropriées en biologie pour caractériser les mutations et le bruit. Nous aborderons donc des questions à réponse normative, c'est-à-dire dont la réponse est basée sur l'état actuel de la recherche au sujet de ces deux sources de la variation et sur une analyse critique de la pertinence à la fois conceptuelle et empirique de la notion de hasard mobilisée dans ce cadre. Plus explicitement, nous souhaitons poser la question de savoir quelle est la signification et le rôle de la notion de hasard afin de rendre compte de manière pertinente de l'origine causale de la variation biologique.

Hasard et mutations

Nous nous occuperons d'abord des mutations génétiques, auxquelles nous consacrerons la partie la plus importante de notre thèse (la deuxième). Rappelons qu'une mutation génétique est habituellement définie comme une altération de la séquence d'ADN (ou d'ARN, par exemple dans le cas de certains virus), exception faite pour les changements dus au processus de recombinaison, pouvant être à l'origine de variation génétique et

phénotypique au sein d'une population naturelle⁴¹. La Théorie Synthétique de l'Evolution (1930-1950) attribua une place centrale à la notion de « hasard » dans la caractérisation des mutations génétiques : les biologistes considérés comme les pères fondateurs de la Synthèse Moderne affirmèrent que les mutations sont « aléatoires »⁴² ou « statistiquement aléatoires »⁴³, se produisent « au hasard » ou sont contrôlées par l'« accident »⁴⁴, toujours par rapport à la direction adaptative de l'évolution. Nous analyserons toutes ces formulations, et bien d'autres, dans le chapitre 4.

Malgré les avancées que la recherche sur les mutations génétiques a connues depuis l'époque des premières publications de la Synthèse Moderne, cet ensemble de formulations, et même la manière actuelle de caractériser les mutations génétiques par des termes renvoyant à la notion de hasard, comportent deux ambiguïtés : l'une concerne la définition de mutation génétique, l'autre la notion de hasard. D'une part, il y a une ambiguïté quant à ce que les biologistes entendaient et entendent encore aujourd'hui par « mutation génétique ». En effet, comme tout phénomène biologique, la mutation peut être définie en termes de processus ou en termes de résultat. Soit les biologistes parlent de mutation génétique pour désigner le processus causal produisant une altération de la séquence d'ADN, soit ils désignent le résultat de ce processus, c'est-à-dire la modification même de la séquence d'ADN⁴⁵.

D'autre part, l'ambiguïté porte sur la notion à laquelle les biologistes renvoyaient et renvoient encore aujourd'hui dans la caractérisation des mutations génétiques en termes de « hasard », « aléatoire », « accidentel », « statistique », etc. C'est là l'objet central de la deuxième partie de notre recherche, dans laquelle nous voulons répondre aux questions suivantes. S'agit-t-il d'une notion de hasard subjectif ou objectif ? Autrement dit, désigne-t-elle les bornes de la connaissance humaine quant au phénomène de mutation ou renvoie-t-elle à une caractéristique propre à cette source de variation ?⁴⁶ Quelle est la signification précise

⁴¹ La notion de mutation est le résultat d'un chemin historico-conceptuel au cours duquel elle n'a pas toujours désigné une modification du matériel génétique et n'a pas toujours joué le rôle que la Théorie Synthétique lui attribua. Nous renvoyons au chapitre 3 où nous tracerons l'histoire de cette notion, de la redécouverte de la génétique mendélienne au début du XX^e siècle jusqu'à aujourd'hui.

⁴² Dobzhansky, 1970, *Genetics of the Evolutionary Process*, p. 65.

⁴³ Stebbins, 1971, *Processes of Organic Evolution*, p. 35 ; Simpson, 1953, *The Major Features of Evolution*, p. 86.

⁴⁴ Mayr, 1974[1970], *Populations, espèces et évolution*, p. 117 ; 1976, *Evolution and the Diversity of Life. Selected Essays*, p. 33.

⁴⁵ Au cours de la présente recherche, nous soutiendrons qu'une notion de hasard susceptible de fournir, non simplement une description, mais aussi bien une explication du phénomène biologique qu'elle caractérise, doit désigner en premier lieu les caractéristiques du processus, et non seulement celles du résultat, d'un phénomène donné.

⁴⁶ La caractérisation des mutations génétiques par la notion de hasard et sa double ambiguïté remontent à bien avant la Synthèse Moderne et le développement de la génétique classique mendélienne. Darwin parlait déjà, non pas de mutations génétiques, mais de l'occurrence de variations « spontanées », « accidentelles » ou « au

de la notion de hasard et quel rôle joue-t-elle lorsque les biologistes l'utilisent pour caractériser les mutations génétiques ? S'agit-il d'une notion appropriée, à la fois du point de vue conceptuel et empirique, relativement à ces sources de la variation ? Nous analyserons toutes ces questions en considérant les mutations selon une perspective évolutionnaire et d'un point de vue strictement moléculaire.

Malgré l'ambiguïté qui caractérise encore aujourd'hui la définition des mutations génétiques et la notion de hasard, dix ans avant la découverte de la molécule d'ADN, Luria et Delbrück⁴⁷ affirmaient déjà avoir apporté une preuve expérimentale, à leur avis décisive, en faveur du postulat de la Synthèse Moderne quant au caractère « au hasard » des mutations. Cependant, comme nous le verrons plus en détail au chapitre 5, la validité de leur expérience sur les bactéries en tant que démonstration du caractère « au hasard » des mutations génétiques a été et est encore sujette à controverse. En effet, cette question dépend directement de la définition de mutation génétique en termes de processus ou de résultat et de la notion de hasard invoquée.

En revanche, à la fin des années 1980, la conception du caractère « au hasard » des mutations génétiques propre à la Synthèse Moderne et datant de l'hypothèse de l'évolution par sélection naturelle de Darwin, fut mise en discussion par des expériences avec les bactéries : ce fut la découverte de ce que les biologistes appelèrent les « mécanismes mutateurs », qui régulent le taux de mutation en réponse à des changements de l'environnement. La possibilité de l'existence de mutations lamarckiennes ou « dirigées », c'est-à-dire de changements produits en vue de l'adaptation, resurgit⁴⁸.

Un débat controversé éclata et se prolongea jusqu'à la fin des années 1990 : il vit s'opposer les partisans de la Synthèse Moderne, c'est-à-dire des mutations « au hasard », aux partisans des mutations « dirigées ». Dans ce cadre, le terme « hasard » et tous les autres termes, apparemment utilisés en tant que synonymes de « hasard », furent employés de manière massive en référence aux mutations génétiques – de la part des partisans de la Synthèse Moderne pour les caractériser, de la part de ses opposants pour nier la validité de leur application dans le cas de certaines mutations génétiques. Cependant, les biologistes, les historiens et les philosophes de la biologie impliqués dans ce débat proposèrent rarement des

hasard » chez les organismes vivants, même s'il n'avait aucune connaissance quant aux causes de ces variations. En premier lieu, il n'est pas toujours possible de savoir clairement si Darwin faisait référence à la variabilité présente au sein d'un groupe d'organismes individuels (le résultat) ou à la manière dont cette variation s'était produite (le processus). En second lieu, Darwin caractérisait la variation en utilisant non pas une mais plusieurs notions de hasard que nous analyserons dans la deuxième partie de ce travail (voir le chapitre 4).

⁴⁷ Luria & Delbrück, 1943, « Mutations of Bacteria From Virus Sensitivity to Virus Resistance ».

⁴⁸ Cairns *et al*, 1988, « The Origin of Mutants ».

définitions précises de la notion de hasard. Enfin, même si cette controverse se conclut en faveur de la Synthèse Moderne, le débat sur le caractère des mutations s'ouvrit à nouveau et est encore vif aujourd'hui⁴⁹. Il fera l'objet d'une analyse approfondie dans les chapitres 5 et 6 de notre thèse, où nous proposerons un argument en faveur du caractère « au hasard » des mutations génétiques.

Hasard et bruit

La troisième partie de ce travail est consacrée à l'analyse de la notion de hasard dans la caractérisation du bruit dans l'expression des gènes. Rappelons qu'on parle de bruit pour désigner les perturbations qui affectent les réactions biochimiques se produisant au cours de l'expression des gènes et, de cette manière, engendrent de la variation phénotypique au sein de populations d'organismes vivants.

Les études quantitatives du bruit intervenant dans l'expression génétique datent d'une époque assez récente : l'intérêt pour ce phénomène intracellulaire, source de variation phénotypique, a connu une phase d'ascension importante au cours des années 1990.⁵⁰ C'est pourquoi nous avons décidé de le traiter dans ce travail : c'est un sujet de recherche relativement nouveau en biologie et qui n'a pas encore bénéficié d'une enquête approfondie. La conviction qui a guidé notre travail de clarification de la notion de hasard dans la caractérisation de cette source de variation est qu'elle pourrait apporter beaucoup non seulement à la réflexion philosophique sur ce sujet, mais aussi et surtout aux biologistes dans leurs études théoriques et expérimentales du bruit.

Au chapitre 8, où nous présentons les études récentes sur ce phénomène intracellulaire, nous verrons que le bruit intervenant dans l'expression des gènes est défini comme la variation « aléatoire » autour de la valeur moyenne de ce processus, c'est-à-dire autour du résultat phénotypique moyen de l'expression des gènes. Beaucoup d'études sur le bruit portent sur l'expression d'un même gène dans une population d'individus isogéniques ou chez un même individu au cours du temps en gardant des conditions environnementales

⁴⁹ Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life* ; Shapiro, 2005, « A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering » ; Sternberg, 2002, « On the Roles of DNA Repetitive Elements in the Context of a Unified Genomic-Epigenetic System » ; Wright, 2000, « A Biochemical Mechanism for Nonrandom Mutations and Evolution » ; Fox Keller, 2003[2000], *Le siècle du gène*.

⁵⁰ Cependant, il est possible de repérer quelques articles à ce sujet à partir de la fin des années 1950. Voir les références dans Paulsson, 2005, « Models of stochastic gene expression ».

stables et homogènes. Dans ce cas, le résultat de l'expression du gène considéré est souvent décrit en termes d'abondance de protéines produites à partir de ce gène ; de même, le bruit est évalué et défini sur la base de la distribution des fluctuations, considérées comme « stochastiques », autour de la quantité moyenne de protéines.

Il faut remarquer que, dans les études sur le bruit dans l'expression génétique, nous retrouvons la double ambiguïté que nous avons évoquée à propos de la définition des mutations génétiques et de leur caractère « au hasard ». En premier lieu, nous venons de préciser que le bruit dans l'expression des gènes est habituellement défini en termes de résultat. Cependant, il pourrait aussi être conçu et défini en termes de processus, plus précisément comme une suite aléatoire d'états physico-chimiques se produisant à l'intérieur de la cellule au cours du temps. Cette deuxième manière de définir le bruit dans l'expression des gènes est rarement prise en considération par les biologistes car elle ne permet pas de le représenter et de le quantifier de manière précise.

En second lieu, l'ambiguïté porte sur la notion de hasard que les biologistes semblent invoquer dans les descriptions de ce phénomène intracellulaire. Il est intéressant de remarquer que, dans les études sur le bruit dans l'expression génétique, le terme « hasard » est presque absent. En revanche, les biologistes utilisent une terminologie plus technique se limitant souvent aux termes « aléatoire », « stochastique » et « probabiliste », qu'ils semblent employer de manière interchangeable. Par exemple, il est courant de lire que « la variabilité est inévitable dans les systèmes biologiques à cause de la nature aléatoire [ou stochastique] des réactions chimiques dans la cellule »⁵¹. Même si dans le cadre des modèles de l'expression des gènes et du bruit qui l'affecte, le caractère « aléatoire », « stochastique » ou « probabiliste » de ce dernier est traduit par des formalisations mathématiques précises⁵², ce genre d'affirmation invite à poser les mêmes questions que celles soulevées plus haut quant aux mutations génétiques. Nous nous demandons si ces termes apparemment synonymes renvoient à une même notion de hasard. Une fois cela compris, suit la question de savoir quels sont la signification et le rôle de cette notion et si elle est pertinente pour caractériser le bruit dans l'expression des gènes.

Nous pouvons d'ores et déjà indiquer que notre analyse de la notion de hasard dans la caractérisation des mutations et du bruit nous amènera, conformément à l'une des hypothèses directrices de notre recherche, à identifier non pas une mais plusieurs notions de hasard définies par des significations et/ou des rôles différents. Nous montrerons que cette pluralité

⁵¹ Raser & O'Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control », p. 2010.

⁵² Nous traiterons cet aspect dans la troisième partie, au chapitre 8, section 8.6.

conceptuelle n'est aucunement incohérente ni contradictoire. Au contraire, elle témoigne de la présence de points de vue et de projets différents, propres à des disciplines différentes, et complémentaires sur un même phénomène biologique. Nous montrerons donc que la mobilisation de plusieurs notions de hasard relativement à un même phénomène biologique, par exemple les mutations génétiques ou le bruit dans l'expression des gènes, peut contribuer à lui fournir une description plus riche et exhaustive et une explication plus diversifiée et solide.

Les probabilités, les mutations et le bruit

Nous n'avons jusqu'ici considéré que les descriptions verbales des mutations génétiques et du bruit dans l'expression des gènes. Dans le cadre de notre projet figure aussi l'analyse de l'aspect mathématico-formel des études consacrées à ces phénomènes biologiques, plus précisément le caractère probabiliste des modèles que les biologistes construisent pour décrire, prédire et expliquer les phénomènes à l'origine de la variation au sein des populations naturelles. Ainsi, l'emploi constitutif des probabilités et de l'analyse statistique dans l'étude des mutations et du bruit dans l'expression génétique soulève la question du lien entre la notion de hasard invoquée dans la caractérisation de ces sources de la variation et l'utilisation des probabilités dans leurs formalisations mathématiques.⁵³

Conformément à l'une des hypothèses directrices de notre recherche, nous soutenons que l'utilisation de modèles mathématiques probabilistes pour décrire, prédire et expliquer un phénomène biologique est un indice du fait que le phénomène modélisé est considéré comme se produisant « au hasard ». Sur cette base, nous dirigeons notre analyse vers l'identification de l'interprétation appropriée des probabilités que les biologistes utilisent dans les descriptions formelles des phénomènes à l'origine de la variation génétique et phénotypique. Cela nous permettra de mieux élucider la nature de la relation entre les notions de hasard et de probabilité et, nous le verrons, la notion même de hasard.

Nous traitons la question de l'interprétation des probabilités dans la troisième partie de ce travail, qui est consacrée à l'étude du bruit dans l'expression des gènes. Nous poserons donc les questions suivantes : doit-on concevoir les probabilités dans les modèles stochastiques de l'expression génétique selon une interprétation épistémique en tant que mesure du degré d'ignorance ou de croyance des biologistes quant au bruit à l'origine de la

⁵³ Pour les notions de calcul des probabilités et de statistique, nous renvoyons au Glossaire.

variation phénotypique? Ou doit-on les interpréter de manière objective en tant que reflet de certaines propriétés de ce phénomène intracellulaire ? Les questions que nous venons de mentionner sont importantes non seulement d'un point de vue philosophique, mais aussi pour la biologie : leur réponse pourrait apporter des éléments intéressants à la compréhension du bruit dans l'expression des gènes ou, du moins, à l'orientation de la recherche à ce sujet.

Nous pouvons d'ores et déjà indiquer que, dans le dernier chapitre, nous proposerons une méthode – l'analyse de robustesse – pour argumenter en faveur d'une interprétation objective des probabilités dans les modèles mathématiques du bruit dans l'expression génétique. Ensuite, nous soutiendrons que cette méthode peut aussi être appliquée dans le cas des probabilités relatives à toute source causale de la variation et, plus largement, à tout phénomène biologique et physique. Enfin, nous préciserons le sens dans lequel les probabilités sont objectives : nous montrerons que la notion d'objectivité à laquelle nous faisons référence est, selon notre analyse, pertinente en science et n'implique aucun engagement métaphysique pour une vision déterministe ou indéterministe des phénomènes biologiques concernés. Notre attitude agnostique à ce sujet est motivée par le caractère indécidable de ce problème d'ordre métaphysique à partir des moyens d'accès que nous avons aux phénomènes biologiques, notamment les modèles toujours partiels et simplifiés qu'on construit pour les décrire.

Dans le cadre de notre recherche, cependant, nous évoquerons à plusieurs reprises la relation de compatibilité ou d'incompatibilité entre, d'une part, les notions de hasard et les interprétations des probabilités, et d'autre part, les diverses doctrines déterministes et indéterministes. Cela nous fournira les bases pour comprendre si la notion de hasard et l'interprétation des probabilités que nous suggérerons ont des implications quant à la nature fondamentalement déterministe ou indéterministe des phénomènes biologiques considérés. Notre but consistera à proposer une notion de hasard et une interprétation des probabilités qui soient appropriées dans la caractérisation des sources de la variation biologique et qui soient indépendantes de toute prise de position quant au déterminisme et à l'indéterminisme de ces phénomènes et, plus généralement, du monde biologique.

Afin de répondre aux questions soulevées dans cette introduction, nous avons organisé ce travail en trois parties. La première (incluant les chapitres 1 et 2) est une présentation critique de quelques-unes des formes conceptuelles que la notion de hasard a connues au cours de l'histoire de la pensée philosophique et scientifique, de l'Antiquité jusqu'à

aujourd'hui. Dans ce cadre, nous nous interrogerons sur le rapport entre les diverses manières de concevoir le hasard, le déterminisme, l'indéterminisme et les probabilités.

La deuxième partie (incluant les chapitres 3, 4, 5, 6 et 7) est la plus importante et la plus développée de notre thèse et porte sur la notion de hasard dans la caractérisation des mutations génétiques. Elle a pour but de montrer que ce que nous appellerons le « hasard évolutionnaire » et l'« aléatoire (au sens) faible » correspondent à deux manières pertinentes de concevoir le hasard par rapport aux mutations génétiques, selon une perspective évolutionnaire et au niveau moléculaire.

La troisième partie (incluant les chapitres 8, 9 et 10) est focalisée sur le bruit dans l'expression des gènes. Nous défendrons ici la pertinence des notions de « hasard évolutionnaire » et d'« aléatoire (au sens) faible » dans la caractérisation de cette source de variation phénotypique. C'est aussi dans ce cadre que nous proposerons l'analyse de robustesse pour argumenter en faveur d'une interprétation objective des probabilités dans les modèles de l'expression génétique et du bruit qui l'affecte.

**Partie I – Le hasard dans la pensée
philosophique et scientifique**

La notion de hasard est un objet de réflexion au moins depuis les premiers rayonnements de la pensée philosophique en Grèce antique. Au cours des siècles, elle a été déclinée, en philosophie et dans les sciences, sous des formes conceptuelles différentes, qui peuvent pourtant être regroupées en deux conceptions majeures : le hasard subjectif et le hasard objectif. Parmi les différentes définitions du hasard, les plus techniques ont vu le jour au sein de la théorie des probabilités et de la théorie algorithmique de l'information. Chaque manière de définir ce qu'est le hasard, subjectif ou objectif, entretient une relation de compatibilité ou d'incompatibilité avec d'autres notions, en particulier celles de déterminisme, d'indéterminisme et de probabilité épistémique ou objective. La nature de ces rapports est aussi un objet central de la réflexion en philosophie.

La première partie de notre thèse est, en premier lieu, une introduction critique à quelques-unes des notions de hasard (subjectif et objectif) qui ont vu le jour au cours de l'histoire de la pensée philosophique et scientifique, de l'Antiquité jusqu'à nos jours. En deuxième lieu, elle comporte une analyse du rapport entre ces notions et les doctrines du déterminisme et de l'indéterminisme sous leurs différentes formes (métaphysique, scientifique, méthodologique). Enfin, nous fournissons dans ce cadre une première compréhension de la relation apparemment privilégiée que la notion de hasard entretient avec la notion de probabilité.

De nombreuses raisons nous ont motivée à concevoir et à écrire cette première partie. D'abord, comme dans tout travail de recherche, nous estimons qu'il est très important de tracer les grandes lignes du cadre historique et philosophique où s'insère notre réflexion sur la notion de hasard. En effet, bien que notre travail soit focalisé sur la signification et le rôle de cette notion au sein des disciplines biologiques – en particulier, lorsqu'elle est utilisée pour caractériser les phénomènes à l'origine causale de la variation au sein des populations naturelles –, il soulève et traite un certain nombre de questions philosophiques qui ont été au centre d'un débat très dense au cours des siècles.

Deuxièmement, la clarification des conceptions différentes du hasard et de leurs implications épistémologiques et métaphysiques représentent un enjeu majeur, non seulement en philosophie, mais aussi bien en science et, en particulier, pour la biologie théorique et expérimentale. En effet, une telle clarification conceptuelle peut permettre de mieux comprendre la place de la notion de hasard au sein des théories biologiques et ce que son utilisation implique quant aux présupposés théoriques et aux affirmations empiriques en biologie. Elle peut aussi permettre d'évaluer l'impact que la notion de hasard, dans le cadre des

descriptions verbales et formelles (mathématiques) des phénomènes biologiques, peut avoir sur le développement à venir des recherches en biologie.

Les questions philosophiques au sujet de la notion de hasard et de la relation qu'elle entretient avec les notions de déterminisme, d'indéterminisme et de probabilité ont suscité, au cours du siècle dernier, beaucoup d'intérêt au sein des disciplines biologiques à la fois de la part des biologistes et des philosophes des sciences dont la réflexion porte sur la biologie. Un certain nombre de notions de hasard subjectif et objectif, que nous allons introduire dans les deux chapitres de cette première partie de notre thèse, ont été évoquées à plusieurs reprises non seulement par les philosophes penchés sur ces questions, mais aussi bien par des biologistes dès l'époque de Darwin. Ces derniers ont ainsi essayé de donner un contenu à la notion de hasard lorsqu'ils l'utilisent pour décrire un certain nombre de phénomènes biologiques, du niveau moléculaire au niveau macroévolutif. En outre, durant les vingt dernières années, un débat controversé en philosophie des sciences a eu comme objet la nature fondamentalement déterministe ou indéterministe du processus évolutif et l'origine du caractère probabiliste de la théorie de l'évolution. Il a vu s'opposer, d'une part, les partisans de l'idée selon laquelle l'évolution est un processus fondamentalement indéterministe et d'une interprétation objective des probabilités au sein de la théorie de l'évolution¹ et d'autre part, les partisans de la nature déterministe du processus évolutif et donc du caractère épistémique des probabilités figurant dans sa modélisation.² Notre travail de recherche s'insère dans le cadre de ces réflexions philosophiques qui sont plus spécifiques, car elles sont focalisées sur des phénomènes et des théories biologiques.

Troisièmement, cette première partie nous permettra de justifier certains choix caractérisant notre réflexion sur la notion de hasard en biologie. Par exemple, nous pourrions justifier le fait que, tout au long de notre travail de recherche, nous évacuerons l'analyse de certaines notions de hasard (par exemple, celles qui ont émergé de la théorie algorithmique de l'information), qui ne nous semblent pas appropriées pour caractériser les phénomènes biologiques au centre de notre enquête, c'est-à-dire les sources causales de la variation. De plus, nous pourrions motiver notre choix de ne pas aborder certaines questions philosophiques que notre travail de recherche soulève, notamment le problème de la nature déterministe ou

¹ Brandon & Carson, 1996, « The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables' Proof But No Room for Determinism Either ».

² Graves *et al*, 1999, « Is Indeterminism the Source of the Statistical Character of Evolutionary Theory? ». Pour une introduction à cette controverse sur l'origine du caractère probabiliste de la théorie de l'évolution, nous renvoyons à Malaterre & Merlin, à paraître, « L'(in)déterminisme de l'évolution naturelle : quelles origines pour le caractère stochastique de la théorie de l'évolution ? ».

indéterministe, du point de vue métaphysique, des phénomènes biologiques et, plus particulièrement, des sources causales de la variation biologique.

Enfin, la quatrième raison justifiant la conception de cette première partie, et en particulier son caractère général et l'ampleur des questions qui y sont traitées, réside dans la nécessité d'une opération préalable de clarification de certaines notions qui nous serviront par la suite en tant qu'outils conceptuels : ces notions sont celles de hasard, de probabilité, de déterminisme et d'indéterminisme, mais aussi celles de prédictibilité et d'objectivité. De manière plus précise, dans cette première partie, nous visons à introduire des distinctions et des définitions sur lesquelles nous nous appuierons afin de clarifier et de mieux expliciter les résultats de notre travail. Ces outils conceptuels serviront aussi à attribuer une place à notre analyse et aux conclusions auxquelles elle aura abouti parmi les autres réflexions philosophiques, anciennes et récentes, au sujet de la notion de hasard.

Dans le premier chapitre, d'abord, nous distinguerons les deux conceptions majeures du hasard, le hasard subjectif et le hasard objectif. Ensuite, après avoir fourni une brève analyse critique des doctrines du déterminisme et de l'indéterminisme sous leurs différentes formes (métaphysique, scientifique, méthodologique), nous présenterons un certain nombre de notions de hasard, subjectif et objectif, dont nous examinerons la signification. Plus précisément, nous procéderons à une introduction de leur définition respective et de leurs implications relativement aux doctrines du déterminisme et de l'indéterminisme.

Dans le deuxième chapitre, nous présenterons d'abord quelques-unes des notions de hasard qui ont été développées dans le cadre de la théorie des probabilités et de la théorie algorithmique de l'information et qui ont reçu des définitions formelles et rigoureuses. De même que dans le premier chapitre, au cours de notre analyse, nous examinerons la relation de compatibilité ou d'incompatibilité entre ces notions mathématiques de hasard et, en particulier, les doctrines du déterminisme et de l'indéterminisme (métaphysique, scientifique, méthodologique). Enfin, nous nous intéresserons à la relation entre la notion de hasard et la notion de probabilité qui, comme nous le verrons, se révèle centrale relativement à quelques-unes des notions de hasard définies mathématiquement. Nous présenterons les diverses théories philosophiques des probabilités et, sur cette base, nous suggérerons une réponse possible à la question de savoir quelle est la relation entre les notions de hasard et de probabilité. Nous focaliserons notre attention sur le rapport entre leur signification respective – plus spécifiquement, entre les notions de hasard subjectif et objectif et les interprétations

épistémique et objective des probabilités – et sur le rapport entre les rôles que ces deux notions jouent dans un même cadre théorique.

Chapitre 1 – Hasard subjectif et objectif, déterminisme et indéterminisme

1.1. La notion de hasard : un sujet de réflexion ancien

La notion de hasard¹ est depuis longtemps utilisée pour caractériser des événements qui, dans la plupart des cas, sont imprédictibles. Elle a connu des déclinaisons conceptuelles diverses au cours de l'histoire de la pensée philosophique et scientifique, de l'Antiquité jusqu'à nos jours. Pendant la période antique, les philosophes ont consacré une part importante de leur réflexion à la recherche du principe métaphysique de la totalité des choses : l'alternative entre le hasard et une Fin, une Cause Ultime ou une Intelligence Divine, gouvernant la totalité du monde est ancienne. Plus précisément, les philosophes de la Grèce antique parlaient de *τό αὐτόματον* et de *ἡ τύχη* – qu'on traduit respectivement par « spontanéité » et « fortune » et, plus généralement, par le terme « hasard ». Ces notions ne visaient pas à mettre en discussion le principe de causalité selon lequel rien ne naît sans cause², c'est-à-dire l'idée selon laquelle tout événement est le produit d'un autre événement (ou cause) qui le précède. En effet, ils considéraient ces notions comme compatibles avec la vision déterministe selon laquelle tous les phénomènes ont lieu selon l'ordre imprimé par la raison divine (*λογος*) sur le monde dans sa totalité (*κόσμος*).³ De manière générale, les

¹ Le mot français « hasard » trouve son origine au Moyen-Orient, apparemment autour du XII^e siècle, et il apparaît en référence aux jeux de dés. Son étymologie nous révèle qu'il dérive de l'expression arabe « *az zahr* » qui signifie « jeu de dés ». Ce sens propre, dominant en ancien français, a cédé la place au sens plus général et populaire « risque », « chance », « fortune », déjà courant au Moyen Age, et qui est passé ensuite aux langues voisines (Bloch & von Wartburg, 1996, *Dictionnaire étymologique de la langue française*). Jusqu'au XIX^e siècle, on emploie le mot « hasard » en rapport avec des activités ludiques diverses, pour désigner l'incertitude des témoignages dans le cadre de la justice et, ensuite, dans les études de statistiques utilisées surtout par les sociétés d'assurance.

² D'Aristote à Leibniz, le lien entre la cause et l'effet est conçu comme nécessaire. Hume est le premier à mettre en discussion l'existence de ce genre de connexions nécessaires dans la nature. Au XX^e siècle, un certain nombre de théories de la causalité, déterministe et probabiliste, se développent. Nous précisons que, dans le cadre de ce travail, nous utilisons le terme de « cause » au sens large et commun d'un événement qui en produit un autre, sans adhérer à aucune conception particulière de la causalité. Pour une introduction aux théories de la causalité dans la philosophie contemporaine, nous renvoyons à Salmon, 1998, *Causality and Explanation* ; Psillos, 2003, *Causation and Explanation* ; Kistler, 2004, « La causalité dans la philosophie contemporaine » ; à paraître, « La causalité ».

³ La pensée de Lucrèce, qui témoigne de la philosophie épicurienne dont on a des sources lacunaires, s'éloigne de la vision déterministe du monde soutenue par la plupart des philosophes grecs anciens. Plus spécifiquement, Lucrèce s'oppose à la doctrine fataliste des Stoïciens selon laquelle le destin impose au monde et à chaque être le

philosophes grecs anciens tendaient plutôt à considérer la « spontanéité » et la « fortune » comme appartenant à l'ordre des causes, non pas nécessaires mais « par accident » (*συμβεβηκότως*), et introduisant donc un élément de contingence dans l'ordre propre à la nécessité cosmique.⁴

La conception du hasard comme absence radicale de cause déterminante – c'est-à-dire, comme négation de ce que Leibniz appela le « principe de raison suffisante »⁵, ou « principe de la raison déterminante » – est rejetée durant toute la période qui va de l'Antiquité (sauf chez les épicuriens⁶) au début du XX^e siècle. En effet, les Grecs anciens aussi bien que de nombreux philosophes et scientifiques à partir du XVII^e siècle jusqu'au début du XX^e ne pouvaient l'admettre au nom du principe de la raison humaine qui ne doit accepter un événement que s'il y a un autre événement qui le précède et le détermine nécessairement, et donc une explication causale qui le justifie. Un tel rejet de la conception du hasard en tant qu'absence de cause fut toujours accompagné et soutenu par une vision déterministe du monde dont la version minimale affirme que tout événement suit de manière nécessaire d'un événement qui le précède et qui en est la raison d'être (ou cause).

Dans les années 1920-1940, la naissance de la physique quantique provoqua un bouleversement de cette longue tradition déterministe dont la mécanique classique de Newton était considérée comme la pierre angulaire.⁷ En effet, nous verrons que, selon le « principe d'indétermination » d'Heisenberg, il est impossible de connaître simultanément les valeurs exactes de la position et de la quantité de mouvement (la masse multipliée par la vitesse) d'un système quantique, car ces quantités physiques, ainsi que d'autres qui définissent l'état d'un

composant un parcours prédéterminé, unique et nécessaire, donc inchangeable et prévisible. Dans *De rerum natura*, Lucrèce affirme que l'ordre du monde est à la fois le résultat de la nécessité, déterminant la chute parallèle des atomes assujettis à leur propre pesanteur, et du *clinamen*, c'est-à-dire de très petits écartements des atomes de leur trajectoire verticale se produisant à des moments et à des endroits indéterminés (II, 218-219). En provoquant la rencontre imprédictible des atomes, le *clinamen* est à l'origine causale de la formation des choses du monde. Cependant, le *clinamen* lui-même n'a pas de cause et l'on ne peut que dire qu'il se produit de manière totalement indéterminée. Ainsi, chez Lucrèce le hasard est conçu comme l'absence de toute cause à l'origine des déviations des atomes ; pour autant, une telle conception n'implique pas la négation du déterminisme dû à la nécessité physique.

⁴ Voir Conche, 2005, « La métaphysique du hasard » ; Lestienne, 1993, *Le hasard créateur*.

⁵ « Jamais rien n'arrive, sans qu'il y ait une cause ou du moins une raison déterminante, c'est-à-dire quelque chose qui puisse rendre raison *a priori*, pourquoi cela est existant plutôt que non existant [...] Ce grand principe a lieu dans tous les événements [...] Il ne souffre aucune exception. », Leibniz, 1969[1710], *Essai de théodicée : sur la bonté de Dieu, la liberté de l'homme et l'origine du mal*, 1, 44.

⁶ Voir la note 3 ci-dessus.

⁷ Cependant, Earman (1986, *A Primer on Determinism*) montre que la mécanique de Newton, aussi bien que la théorie de la relativité, ne sont pas complètement compatibles avec une vision déterministe du monde. De plus, il souligne que la mécanique quantique pose un défi majeur à l'ontologie de la physique newtonienne et de la théorie de la relativité plutôt qu'à l'hypothèse du caractère déterministe du monde. Voir aussi Salmon, 1998, *op. cit.*, p. 34-36.

système, sont dans une relation d'indétermination réciproque.⁸ La naissance de la physique quantique ouvrit donc la voie à une conception du hasard qui avait été longuement refoulée auparavant : il s'agit de la conception selon laquelle le hasard est objectif, ou pur, et désigne une propriété ou une entité métaphysique intrinsèquement probabiliste et existant dans le monde (du moins, au niveau microscopique ou quantique).⁹

1.2. La double nature du hasard

L'histoire de la conceptualisation du hasard, en philosophie et en science, nous révèle que cette notion est double car elle englobe deux idées distinctes¹⁰. D'une part, le hasard concerne la connaissance que nous avons des événements réels, autrement dit la relation d'ordre épistémique que nous entretenons avec les phénomènes objectifs du monde : c'est la notion de hasard subjectif. D'autre part, le hasard désigne, au contraire, une propriété inhérente à la structure du monde réel et qui est indépendante de la connaissance que nous en avons : c'est la notion de hasard objectif.¹¹ Nous verrons que ces deux manières de concevoir le hasard émergent aussi des théories biologiques, cadre dans lequel elles sont utilisées pour décrire des phénomènes divers, du niveau moléculaire à l'échelle macroévolutive.

⁸ Nous verrons qu'il existe plusieurs interprétations de la théorie quantique dont certaines n'ont rien d'indéterministe (par exemple, la théorie à variables cachées de Bohm (1952, « A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables »), élaborée à partir des travaux initiaux que de Broglie avait réalisés dans les années 1920). À son tour, toute théorie quantique peut être interprétée de manière réaliste ou non réaliste, c'est-à-dire qu'elle peut être considérée comme une description vraie de la structure du monde microscopique ou simplement comme une représentation empiriquement adéquate et appropriée du point de vue explicatif tout en étant une description fautive de la réalité quantique.

⁹ La théorie métaphysique de la réalité du philosophe américain Peirce (le Tychisme) est inspirée et se rapproche de l'indéterminisme du monde microscopique suggéré par la mécanique quantique (Peirce, 2001[1923], *Chance, love, and logic: philosophical essays*). Cependant, dans le cadre de notre travail, nous ne la prendrons pas en considération. En effet, bien que le Tychisme de Peirce trouve des fondements scientifiques pertinents dans la mécanique quantique, cette théorie de la présence irréductible du hasard dans la nature est en partie basée sur des spéculations métaphysiques qui n'ont pas beaucoup d'intérêt pour notre recherche épistémologique sur la notion de hasard en biologie. Pour une présentation critique de la « métaphysique scientifique réaliste » de Peirce, voir Tiercelin, 2003, « Le projet peircien d'une métaphysique scientifique réaliste » ; Hacking, 1995[1990], *The Taming of Chance* (chapitre 23).

¹⁰ Hacking, 2002[1975], *L'Émergence de la probabilité*.

¹¹ La notion de hasard objectif présuppose l'indépendance totale entre les propriétés des phénomènes réels et la connaissance que nous en avons. Cependant, dans le chapitre 10, nous proposerons de parler d'objectivité – de manière générale et, en particulier, en ce qui concerne les probabilités mobilisées dans les formalisations mathématiques des phénomènes biologiques – dans un sens qui nous paraît indispensable dans tout cadre scientifique. Selon cette notion – qui ne correspond pas à la notion métaphysique d'objectivité –, une propriété peut être considérée comme objective si elle est invariante (ou robuste) par rapport à un ensemble divers de perspectives.

Dans ce chapitre, nous allons procéder à l'examen de quelques notions de hasard en les regroupant selon qu'elles correspondent à l'idée du hasard subjectif ou objectif. En particulier, nous allons concentrer notre analyse sur les implications métaphysiques, épistémologiques et méthodologiques de chaque manière de concevoir le hasard, sur la relation entre les diverses notions de hasard et l'imprédictibilité des phénomènes qu'elles décrivent, et sur la relation de compatibilité ou d'incompatibilité entre chaque notion de hasard et les notions de déterminisme et d'indéterminisme.¹² C'est pourquoi, avant de nous plonger dans la présentation et l'analyse de diverses notions de hasard subjectif et de hasard objectif, nous consacrons la prochaine section à l'élucidation des notions de déterminisme et d'indéterminisme.

1.3. Déterminisme et indéterminisme

Les premières lignes du livre du philosophe américain Earman intitulé *A Primer on Determinism* (1986) expriment de manière très puissante le caractère problématique de la notion de déterminisme et, par conséquent, d'indéterminisme. Earman écrit :

« Le déterminisme est un sujet pérenne de discussion philosophique. Pour apercevoir la Tour de Babel que représente le débat sur ce sujet, il ne faut connaître que peu de la littérature philosophique : [...] Voilà, dira le cynique, un sujet philosophique *par excellence* ! »¹³.

L'évocation de cette notion, en effet, a toujours été accompagnée par une ambiguïté qui est liée, en particulier, aux discours causaux mobilisés pour en fournir une définition. De plus, le caractère équivoque de la notion de déterminisme s'estompe, mais demeure même lorsqu'elle est définie rigoureusement en termes mathématiques¹⁴.

¹² Des analyses philosophiques des différentes manières de concevoir la notion de hasard figurent dans Nagel, 1961, *The Structure of Science. Problems in the Logic of Scientific Explanation* ; Gayon, 1994, « Le hasard dans la théorie évolutionniste moderne : une analyse philosophique » ; 2005, « Chance, Explanation, and Causation in Evolutionary Theory » ; Millstein, 2006, « Discussion of 'Four Case Studies on Chance in Evolution: Philosophical Themes and Questions » ; Saint-Sernin, 2008, « Hasard » ; Hajek, 2008, « Interpretations of Probability ».

¹³ Italiques dans l'original ; « *Determinism is a perennial topic of philosophical discussion. Very little acquaintance with the philosophical literature is needed to reveal the Tower of Babel character of the discussion: [...] Here we have, the cynic will say, a philosophical topic par excellence!* », Earman, 1986, *A Primer On Determinism*, p. 1, (notre traduction).

¹⁴ Nous faisons référence à la notion de solution unique d'un système d'équations différentielles. En effet, contrairement à ce qu'on pensait jusqu'à la fin du XIX^e siècle – c'est-à-dire, avant les travaux pionniers de Poincaré sur les systèmes physiques (dits « chaotiques ») qui sont hautement sensibles à de petits changements

Bien que la plupart des réflexions sur le déterminisme et l'indéterminisme en philosophie des sciences soient focalisées sur les questions qu'ils soulèvent en physique, un certain intérêt pour ces notions et des difficultés pour en déceler la signification précise se sont manifestés, notamment au cours des vingt dernières années, au sein des réflexions sur les phénomènes et les théories biologiques. Nous avons déjà mentionné plus haut le débat philosophique sur le caractère fondamentalement déterministe ou indéterministe du processus évolutif et sur l'origine du caractère probabiliste de la théorie de l'évolution. Les arguments que certains philosophes ont avancés en faveur de l'indéterminisme de l'évolution s'appuient sur des considérations concernant le caractère fondamentalement indéterministe de quelques-uns des phénomènes impliqués dans le processus évolutif (entre autres, les mutations génétiques). De plus, ce débat philosophique a été caractérisé par de nombreux glissements entre des questions d'ordre métaphysique, épistémologique et méthodologique, dont les propos respectifs ont été confondus. Une clarification des notions de déterminisme et d'indéterminisme – et, plus particulièrement, des différentes manières de les concevoir – se révèle donc nécessaire pour mieux comprendre ce genre de controverses et pour essayer d'en dissiper toute ambiguïté. Plus généralement, seul un éclaircissement préalable de ces notions pourra nous permettre de poursuivre sans ambiguïté notre projet d'analyse des notions de hasard subjectif et objectif, de manière générale et plus spécifiquement en biologie. De plus, au cours de notre travail de recherche, une telle opération de clarification conceptuelle nous permettra d'évaluer ce qu'implique l'utilisation de certaines notions de hasard, subjectives ou objectives, pour décrire des phénomènes biologiques relativement au caractère déterministe ou indéterministe de ces derniers.

Dans cette section, nous ne prétendons pas apporter une présentation exhaustive et une clarification des notions de déterminisme et d'indéterminisme.¹⁵ Nous nous limitons à

des conditions initiales –, la notion de solution unique d'un système d'équations différentielles n'est pas toujours associée à la notion de connaissance certaine (voir Suppes, 1993, « The Transcendental Character of Determinism »). Plus précisément, Poincaré montra que certains systèmes peuvent être décrits par des équations différentielles, tout en étant complètement imprédictibles. Nous renvoyons à Barberousse, 2000, *La physique face à la probabilité*, p. 84. Au sujet des systèmes chaotiques, voir la section 1.5.3 de ce chapitre.

¹⁵ Pour une analyse approfondie des notions de déterminisme et d'indéterminisme, en particulier en physique, nous renvoyons Earman, 1986, *op. cit.* Voir aussi Montague, 1974, « Deterministic Theories » ; Popper, 1984[1982], *L'univers irrésolu. Plaidoyer pour l'indéterminisme* ; Pomian (dir.), 1990, *La querelle du déterminisme* ; Gayon, 1998, « Le déterminisme : origines d'un mot, évaluation d'une idée » ; Salmon, 1998, *Causality and Explanation* ; Atmanspacher & Bishop (eds.), 2002, *Between Chance and Choice: Interdisciplinary Perspectives on Determinism*. Pour une introduction générale aux notions de déterminisme et d'indéterminisme, et aux problématiques les concernant, voir Suppes, 1993, « The Transcendental Character of Determinism » ; Roberts, 2006, « Determinism » ; Hofer, 2008, « Causal Determinism » ; Balibar & Macherey, 2008, « Déterminisme ».

distinguer et à définir quelques déclinaisons de ces notions qui émergent de notre reconstruction des positions philosophiques disponibles.

1.3.1. Le déterminisme

Considérons d'abord la notion de déterminisme.¹⁶ Mise à part la doctrine du fatalisme que nous présentons d'abord, nous proposons de distinguer trois manières de concevoir le déterminisme selon trois perspectives différentes : le déterminisme métaphysique, le déterminisme scientifique et le déterminisme méthodologique.

Au XVIII^e siècle, en particulier dans le cadre de la théologie et de la philosophie morale, le mot « *Determinismus* » désignait la doctrine du fatalisme (du latin *Fatum*) selon laquelle tout phénomène et tout ce qui advient dans l'univers dans sa totalité sont complètement fixés à l'avance (ou prédéterminés) car dus au déploiement nécessaire et inéluctable du destin, de la volonté divine, de la nécessité naturelle ou des lois gouvernant l'histoire. Dans la mythologie antique et dans d'autres religions plus tardives, c'était le destin qui, sous la forme d'un agent surnaturel (par exemple, la mort), était censé déterminer à l'avance et de manière irrémédiable toute chose dans l'univers. Dans la philosophie des Grecs anciens, il s'agissait plutôt de la raison divine (ou *λόγος*) qui, par exemple chez les Stoïciens, était à l'origine de la nécessité naturelle et prédéterminait donc l'ordre nécessaire et rationnel au niveau cosmologique. Dans les religions chrétienne et musulmane, c'est le Dieu tout puissant et omniscient qui incarne le destin et détermine de manière inconditionnelle le devenir de toute chose : Lui seul n'est pas prédéterminé et possède donc la liberté de décision qui est au contraire niée à tout être vivant.

Le fatalisme a depuis longtemps été rejeté en tant que fruit des superstitions primitives et, en particulier par les croyants, au nom de l'idée selon laquelle Dieu ne prive aucune de ses créatures d'un certain degré de liberté. De plus, il a été rejeté par de nombreux philosophes (par exemple, par Leibniz, bien que certains, notamment Bernard¹⁷, aient attribué à celui-ci la

¹⁶ Le mot « déterminisme » est équivoque. Aujourd'hui il renvoie le plus souvent à la vision laplacienne du monde physique (Laplace, 1986[1814], *Essai philosophique sur les probabilités*), bien que Laplace n'ait jamais utilisé ce mot. Le terme « déterminisme » fit son entrée dans la langue française en 1836 (Bloch & Wartburg, 1996, *Dictionnaire étymologique de la langue française*) et fut popularisé un demi-siècle plus tard par Bernard mais dans un sens différent du déterminisme à la fois métaphysique et scientifique de Laplace. Pour une analyse critique de l'histoire du mot et de la notion de déterminisme et des implications philosophiques de cette dernière, voir Gayon, 1998, *op. cit.*

¹⁷ Bernard, 1966[1878], *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux*, p. 55-56.

doctrine du fatalisme à laquelle, au contraire, il s'opposa) en tant que forme de déterminisme trop radicale car niant la possibilité du libre-arbitre.

Passons donc à l'analyse de la notion de déterminisme sous ses formes conceptuelles diverses, qui sont souvent difficiles à distinguer. Le déterminisme peut être défini, soit selon une perspective métaphysique en tant que propriété de la structure du monde réel, soit en termes épistémologiques comme une caractéristique d'une théorie, soit du point de vue méthodologique en tant que maxime pour la science¹⁸. Selon leur degré respectif d'exigence en ce qui concerne les hypothèses qui les fondent, nous présentons dans l'ordre les doctrines déterministes suivantes.

En premier lieu, nous utilisons l'expression « déterminisme métaphysique » pour désigner la doctrine selon laquelle l'état de l'univers dans sa totalité à un instant donné et les lois régissant son évolution déterminent de manière univoque et nécessaire l'état de l'univers entier à tout instant futur et passé. Cette formulation exprime le déterminisme métaphysique universel ou global qu'on identifie souvent à la vision de Laplace¹⁹, bien que ce dernier n'ait pas utilisé cette expression. Le déterminisme métaphysique peut aussi bien recevoir une définition plus restreinte, c'est-à-dire qu'il ne concerne pas l'univers entier mais un ensemble particulier de phénomènes (un système particulier). Cela revient à affirmer que l'état complet d'un système donné à l'instant présent et les lois régissant son évolution déterminent de manière univoque et nécessaire son état à tout moment futur et passé : c'est le déterminisme métaphysique local. En outre, le déterminisme métaphysique peut concerner, non pas toute l'histoire future et passée de l'univers entier (ou d'un système physique particulier), mais simplement un segment fini de son évolution²⁰.

Afin de distinguer de manière claire la vision du monde propre respectivement au déterminisme métaphysique universel ou global et au déterminisme métaphysique local, nous soulignons que, selon le premier, tous les composants de l'univers forment une totalité solidaire qui est régie par des lois universelles et immuables et qui ne laisse aucune place à la contingence, c'est-à-dire à des événements pouvant être ou ne pas être. Au contraire, selon la version locale du déterminisme métaphysique, seuls les composants d'un système particulier sont reliés entre eux par une nécessité nomique²¹. La relation qu'ils entretiennent avec les

¹⁸ Gayon, 1998, « Le déterminisme : origines d'un mot, évaluation d'une idée ».

¹⁹ Laplace, 1986[1814], *Essai philosophique sur les probabilités*. Nous renvoyons à la section 1.4.1 de ce chapitre.

²⁰ Earman, 1986, *A Primer on Determinism*.

²¹ En suivant Kistler (2006, « La causalité comme transfert et dépendance nomique ») nous utilisons l'adjectif « nomique » pour désigner ce qui est en rapport aux lois de la nature et l'adjectif « nomologique » pour les énoncés exprimant les lois de la nature.

composants d'autres systèmes particuliers dans le monde n'est pas du même ordre et n'est donc pas conforme à des lois immuables et universelles.²² Dans ce cas, le déterminisme devient compatible avec l'existence d'événements contingents qui sont dus à la rencontre de nécessités nomiques indépendantes, c'est-à-dire l'occurrence d'événements qui auraient aussi bien pu ne pas avoir lieu. La doctrine du déterminisme métaphysique (global et local) est donc exprimée par un énoncé d'ordre métaphysique et n'implique nullement la prévisibilité scientifique, en principe, de l'évolution de l'univers dans sa totalité ou d'un système particulier.

En deuxième lieu, nous parlons de « déterminisme scientifique » pour désigner la doctrine selon laquelle « *tout événement peut être rationnellement prédit, au degré de précision voulu, à condition qu'une description suffisamment précise des événements passés, ainsi que toutes les lois de la nature, nous soit donnée* »²³. Autrement dit, à partir de la connaissance de l'état complet de l'univers (ou d'un système particulier) à un instant donné et sur la base des lois régissant son évolution, il est en principe possible de prédire, avec n'importe quel degré de précision souhaité, l'état de l'univers dans sa totalité (ou du système considéré) à tout instant futur et passé. Cette doctrine porte donc sur la prédictibilité scientifique, en principe toujours existante et infiniment précise, de l'univers dans sa totalité (dans sa formulation cosmologique) ou d'un système particulier (dans sa formulation locale). Elle se situe au niveau des modèles et des théories scientifiques construits et utilisés pour décrire, expliquer et prévoir les événements se produisant dans le monde réel, et ne concerne donc pas directement la structure de ce dernier (ou d'un système particulier). Cependant, le déterminisme scientifique connaît habituellement une déclinaison plus stricte qui ne se place pas simplement au niveau des théories, mais concerne aussi bien la réalité même. Cette déclinaison est plus exigeante non seulement par rapport au déterminisme scientifique conçu

²² Cournot (1984[1843], *Exposition de la théorie des chances et des probabilités*) soutient un déterminisme métaphysique local et l'exprime en s'opposant, avec ironie, à la conception du monde en tant que totalité solidaire. Il écrit : « [...] c'est un principe de sens commun qu'il y a des séries de phénomènes *solidaires*, ou dépendant les unes des autres, et d'autres séries qui se développent parallèlement ou successivement, sans qu'il y ait entre elles aucune dépendance, aucun lien de solidarité. À la vérité, certains philosophes se sont imaginé que tout dans le monde se tenait, et ils l'ont prouvé à leur manière, ou par des arguments subtils, ou par des plaisanteries ingénieuses ; mais, ni leurs subtilités, ni leurs plaisanteries ne sauraient prévaloir contre les croyances du sens commun. Personne ne pensera sérieusement qu'en frappant la terre du pied il dérange le navigateur qui voyage aux antipodes, ou qu'il ébranle le système des satellites de Jupiter. », p. 55.

²³ Popper, 1984[1982], *L'univers irrésolu. Plaidoyer pour l'indéterminisme*, p. 1, italiques dans l'original ; voir aussi la définition que Popper en donne à la page 5 du même ouvrage. Il est important de souligner que Popper est critique envers le déterminisme scientifique et fournit un argument philosophique élaboré contre sa validité. En résumé, il affirme que toute connaissance des phénomènes réels n'est qu'approximative et toute théorie scientifique n'est qu'une conjecture faillible au sujet de la réalité. Par conséquent, le déterminisme scientifique, selon Popper, n'est qu'une illusion due au fait que nous prenons les simplifications de nos modèles et théories pour des propriétés objectives du monde physique.

en termes de prédictibilité théorique, mais aussi par rapport au déterminisme métaphysique qu'elle suppose.²⁴ Elle requiert, premièrement, la possibilité de connaître l'état complet de l'univers (ou d'un système particulier) ; deuxièmement, des capacités de calcul infiniment puissantes ; troisièmement, l'uniformité des lois de la nature dans le temps et dans l'espace. Dans sa déclinaison plus stricte, la doctrine du déterminisme scientifique (global et local) est donc exprimée par un énoncé épistémologique – dont l'objet est la prédictibilité, en principe, des phénomènes dans le monde – mais qui présuppose et trouve sa légitimité dans l'énoncé métaphysique du déterminisme métaphysique (global et local).

Une troisième déclinaison du déterminisme est habituellement appelée le « déterminisme méthodologique » : celui-ci ne porte, ni sur la structure du monde, ni sur la prédictibilité en principe des phénomènes, mais consiste en une maxime méthodologique qui oriente la pratique scientifique. Le « déterminisme expérimental » de Bernard²⁵ en fournit un exemple. En tant qu'« axiome » méthodologique, le déterminisme correspond au principe d'uniformité des lois de la nature – aussi dit « principe de légalité » à la fin du XIX^e siècle – selon lequel tous les êtres du monde physique, inorganique et organique, sont régis par des lois fixes, universelles et immuables²⁶. Bernard pose ce principe général en tant que condition de possibilité de toute science expérimentale.²⁷ Le déterminisme méthodologique peut aussi consister à favoriser dans la pratique scientifique des hypothèses d'allure déterministe, c'est-à-dire non probabilistes. Une telle maxime repose sur le fait que les hypothèses déterministes contiennent *a priori* des informations plus précises que les hypothèses probabilistes et sont donc plus faciles à confirmer ou à falsifier à travers l'expérience. Popper²⁸, tout en se déclarant indéterministe, soutient cette forme de déterminisme méthodologique, en particulier en ce qui concerne les théories scientifiques particulières : c'est à son avis le seul déterminisme qui mérite d'être sauvegardé pour le progrès de la recherche scientifique.

²⁴ Nous verrons que la pensée de Laplace représente un cas emblématique de ce lien d'implication entre le déterminisme métaphysique et la prédictibilité, en principe, au moyen des théories scientifiques.

²⁵ Bernard, 1865, *Introduction à l'étude de la médecine expérimentale*.

²⁶ « Dans les corps vivants comme dans les corps bruts les lois sont immuables, et les phénomènes que ces lois régissent sont liées à leurs conditions d'existence par un déterminisme nécessaire et absolu. », *ibid.*, p. 118.

²⁷ Bernard (*ibid.*, p. 150-151) parle aussi du déterminisme d'un phénomène particulier pour désigner, non pas le principe méthodologique de légalité de la nature, mais la « condition d'existence » d'un phénomène, sa cause « prochaine », « efficiente », « déterminante », sa condition « nécessaire », dont la connaissance peut permettre à l'expérimentateur d'agir et de maîtriser le phénomène considéré. Il est aussi important de rappeler que Bernard (1966[1878], *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux*, p. 55-56) met en évidence le fait que son déterminisme méthodologique (qu'il appelle aussi « déterminisme physiologique ») ne doit pas être confondu avec le « déterminisme philosophique » qu'il associe à la pensée de Leibniz en accusant ce dernier de fatalisme. Pour une analyse critique du déterminisme méthodologique chez Bernard, voir Gayon 1998, « Le déterminisme : origines d'un mot, évaluation d'une idée » ; 2006a, « Les réflexions méthodologiques de Claude Bernard : structure, contexte, origines ».

²⁸ Popper, 1984[1982], *L'univers irrésolu. Plaidoyer pour l'indéterminisme*.

1.3.2. L'indéterminisme

Sur la base de la distinction des trois notions de déterminisme présentées ci-dessus, considérons l'indéterminisme et procédons à la clarification de cette notion.

Dans sa version globale, l'indéterminisme métaphysique est la doctrine selon laquelle l'état de l'univers dans sa totalité à un instant donné et les lois de la nature ne déterminent pas de manière univoque et nécessaire l'état de l'univers entier à tout autre instant futur et passé. Dans sa version locale, cette doctrine consiste à affirmer que l'état complet d'un système particulier à l'instant présent, avec les lois régissant son évolution, ne détermine pas de manière univoque et nécessaire tout état futur et passé de ce système. L'indéterminisme métaphysique (global ou local) est donc une thèse au sujet de la structure de l'univers ou d'un système particulier et n'implique ni qu'il est possible, ni qu'il est impossible de prédire, au moyen d'une théorie scientifique, leur évolution au cours du temps.

Selon une perspective épistémologique, on peut utiliser l'expression d'« indéterminisme scientifique » pour désigner la négation du déterminisme scientifique. Il correspond donc à la vision selon laquelle, à partir de la connaissance de l'état de l'univers entier (ou d'un système particulier) à l'instant présent et des lois régissant son évolution, il n'est pas toujours possible de prédire, avec n'importe quel degré de précision souhaité, l'état de l'univers entier (ou du système considéré) à tout moment donné futur ou passé. D'une part, cette thèse peut n'être qu'épistémologique et ne porter que sur les modèles et les théories scientifiques construits et utilisés en science, sans qu'aucune vision métaphysique sur la structure du monde soit supposée. D'autre part, l'indéterminisme scientifique peut aussi consister en une position plus exigeante, affirmant que l'imprédictibilité en science trouve son origine dans l'indéterminisme constitutif de l'univers entier (dans sa version globale) ou d'un système particulier (dans sa version locale). Il est donc exprimé par un énoncé épistémologique présupposant une thèse beaucoup plus forte, d'ordre métaphysique.²⁹ Les deux doctrines de l'indéterminisme métaphysique et scientifique trouvent leur place dans le cadre des réflexions sur les phénomènes quantiques de la part des physiciens et des philosophes des sciences.³⁰

En ce qui concerne la possibilité que l'indéterminisme joue le rôle d'une maxime méthodologique pour la pratique scientifique, il est ardu d'identifier ce qu'elle pourrait

²⁹ Aussi bien que le déterminisme métaphysique et scientifique, l'indéterminisme métaphysique et l'indéterminisme scientifique peuvent ne concerner qu'une portion limitée de l'histoire de l'univers entier ou d'un système particulier.

³⁰ Entre autres, voir Popper 1984[1982], *op. cit.*

apporter pour l'avancement de la science. En reformulant en termes indéterministes les idées respectives de Bernard et de Popper que nous avons introduites plus haut, l'« indéterminisme méthodologique » consisterait à poser en tant qu'axiome méthodologique la négation de l'uniformité des lois de la nature dans l'espace et dans le temps, ou encore le fait de favoriser des hypothèses d'allure indéterministe, c'est-à-dire probabilistes, en science. Cependant, des hypothèses de ce genre ne nous semblent pas appropriées en tant que maximes pratiques pour faire avancer la recherche scientifique. En ce qui concerne la première, la négation de l'uniformité des lois de la nature entrerait en conflit avec l'une des formes de raisonnement les plus utilisées dans les sciences expérimentales – l'induction. En effet, cette dernière se fonde sur l'idée de la constance des phénomènes, dans l'espace et dans le temps, et consiste à inférer des généralisations à partir de l'observation de faits particuliers.³¹ Quant à la deuxième, bien qu'une hypothèse probabiliste puisse aussi être confirmée ou falsifiée par l'utilisation de méthodes statistiques, de manière générale, elle n'est pas à préférer à une hypothèse déterministe lorsque cette dernière est disponible. Si un phénomène semble pouvoir être expliqué à la fois sans et avec le recours aux probabilités, il est préférable de tester, d'abord, l'hypothèse déterministe car elle contient des informations précises au sujet du phénomène étudié et est susceptible d'être réfutée de manière rigoureuse et non simplement statistique.

Nous venons de clarifier les différentes manières dont on peut parler de déterminisme et d'indéterminisme. Quel type de relations entretiennent-elles avec les notions de hasard subjectif et de hasard objectif ? Afin de répondre à cette question, nous allons introduire de manière critique quelques-unes des notions relatives aux deux conceptions du hasard, subjectif et objectif. Notre but est de mettre en évidence les différences entre ces notions, premièrement, en ce qui concerne leur signification précise ; deuxièmement, quant au rapport de compatibilité ou d'incompatibilité avec les doctrines du déterminisme et de l'indéterminisme. Ces deux éléments sont fondamentaux pour comprendre quelle est place de la notion de hasard lorsqu'elle est utilisée au sein d'une théorie scientifique et ce qu'elle implique aux points de vue méthodologique, théorique et philosophique.

³¹ Il ne faut pas oublier que la connexion entre les prémisses et la conclusion d'un argument inductif est probabiliste, c'est-à-dire que les prémisses n'impliquent pas logiquement la conclusion mais ne fournissent qu'un certain degré de soutien en faveur de cette dernière.

1.4. Le hasard subjectif

Nous avons déjà indiqué plus haut que la notion de hasard englobe deux idées distinctes : le hasard subjectif et le hasard objectif. Dans cette section, nous allons nous occuper du hasard subjectif, notion qui ne renvoie pas aux propriétés des événements se produisant dans le monde, mais concerne les caractéristiques de la connaissance que nous en avons, autrement dit la relation d'ordre épistémique que nous entretenons avec les phénomènes objectifs du monde. Il n'y a donc de hasard que relativement à un sujet connaissant.

En science, bien que le terme « hasard » soit utilisé avec prudence et que son sens soit rarement explicité, la notion de hasard subjectif est souvent évoquée par les scientifiques, de manière plus ou moins implicite, afin de rendre compte du caractère partiel de la connaissance des phénomènes (physiques, chimiques, biologiques, etc.) qu'ils étudient. Dans le domaine scientifique qui nous intéresse, la biologie, il suffit de remonter aux affirmations du père inspirateur de la théorie de l'évolution par sélection naturelle, Darwin, pour montrer que la notion de hasard subjectif fait partie des outils conceptuels des biologistes. Darwin attribuait le caractère spontané des variations qu'il observait au sein des populations naturelles, entre autres, à notre ignorance des causes de chaque variation particulière³². Il invoquait donc la notion de hasard subjectif comme ignorance des causes que nous allons présenter et analyser ci-dessous.

1.4.1. Le hasard subjectif en tant qu'« ignorance des causes »

La notion de hasard subjectif en tant qu'ignorance des causes consiste à considérer le hasard comme une simple manière d'exprimer notre ignorance des causes sous-jacentes aux événements se produisant dans le monde. Cette notion est souvent associée à la vision du déterminisme universel ou global, à la fois métaphysique et épistémologique, qu'est la doctrine du déterminisme scientifique dans sa déclinaison stricte. Cependant, elle n'est pas

³² Darwin, 1859, *On The Origin of Species By Means of Natural Selection or The Preservation of Favoured Races in The Struggle For Life*, p. 131. Dans la deuxième partie de ce travail, nous introduirons deux autres manières dont Darwin rendait compte du caractère « spontané », « accidentel », « au hasard » de la variation biologique.

incompatible avec une vision indéterministe de la structure du monde qui, au contraire, est le plus souvent associée à une conception du hasard en termes objectifs, c'est-à-dire qui renvoie aux caractéristiques des phénomènes dans le monde, indépendamment de la connaissance que nous en avons.

La pensée de Laplace³³ illustre bien l'idée selon laquelle le hasard est ignorance des causes, plus précisément, en tant qu'ignorance de l'événement qui précède et qui produit de manière nécessaire et univoque un autre événement donné. Cette notion émerge du cadre de la vision épistémologique de Laplace selon laquelle l'état de l'univers dans sa totalité est, en principe, prédictible de manière infiniment précise à tout instant dans le temps. Au sein de la réflexion laplacienne, une telle vision – correspondant à ce qu'on appelle aujourd'hui doctrine du déterminisme scientifique global – se fonde sur la conception métaphysique de l'univers en tant que totalité solidaire et régie par des lois nécessaires – qui correspond à ce qu'on appelle aujourd'hui doctrine du déterminisme métaphysique universel³⁴. En effet, Laplace ne se contente pas d'affirmer que l'univers entier et chaque particule qu'il comprend ont été, sont et seront toujours gouvernés par les mêmes lois, ce qui correspond au principe d'uniformité des lois de la nature avancé par Laplace, sous la forme d'une thèse métaphysique sur la structure de l'univers dans sa totalité, de la manière suivante :

« Tous les événements, ceux même qui par leur petitesse semblent ne pas tenir aux grandes lois de la nature, en sont une suite aussi nécessaire que les révolutions du Soleil. [...] Les événements actuels ont avec les précédents une liaison fondée sur le principe évident, qu'une chose ne peut pas commencer d'être sans une cause qui la produise. [...] Nous devons donc envisager l'état présent de l'univers comme l'effet de son état antérieur et comme la cause de celui qui va suivre. »³⁵

³³ Laplace, 1986[1814], *Essai philosophique sur les probabilités*.

³⁴ Voir Earman, 1986, *A Primer on Determinism*, p. 10. Nous avons déjà précisé plus haut que le mot « déterminisme » n'a jamais été utilisé par Laplace pour qualifier sa vision du monde physique. Cependant, nous avons décidé de l'utiliser en référence à la pensée laplacienne simplement afin de rendre notre analyse plus claire et homogène. Le mot « déterminisme » a été appliqué pour la première fois et à raison par un élève de Leibniz (Wolff) pour désigner la notion de « détermination » chez son maître, même si ce dernier n'avait jamais utilisé le terme « déterminisme ». Plus précisément, Leibniz (1969[1710], *Essai de théodicée : sur la bonté de Dieu, la liberté de l'homme et l'origine du mal*), distinguait de manière stricte la « nécessité » – propre aux vérités logiques et où opère le principe de contradiction – de la « détermination » – propre aux vérités des faits (physiques et moraux) et où opère le principe de « raison suffisante » (ou « déterminante »). Le mot « déterminisme » n'est entré dans le langage philosophique et scientifique qu'en 1865, dans l'*Introduction à l'étude de la médecine expérimentale* de Bernard, pour désigner à la fois la condition nécessaire sans laquelle un événement ne peut se produire et un principe méthodologique pour toute science expérimentale. Voir Gayon, 1998, « Le déterminisme : origines d'un mot, évaluation d'une idée ».

³⁵ Laplace, 1986[1814], *op. cit.*, p. 32.

Laplace soutient aussi que, par conséquent³⁶, dans l'univers – considéré comme un tout et, distributivement, comme l'ensemble exhaustif de ces composants – tous les événements sont en principe prédictibles de façon infiniment précise, univoque et avec une certitude absolue.

« Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'Analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux. »³⁷

La fiction laplacienne d'une Intelligence supérieure capable de prévoir de manière exacte les événements de l'univers traité comme un tout a été l'objet de plusieurs analyses critiques³⁸. L'interprétation la plus répandue suggère que cette Intelligence est propre, non pas à une entité divine, mais à un physicien idéalisé : elle n'est caractérisée que par une différence quantitative, et non pas qualitative, par rapport à l'intelligence humaine. Laplace estime qu'il n'y a de hasard que pour un sujet connaissant qui ne connaît pas toutes les causes des phénomènes réels et, plus généralement, la nécessité nomique régissant la totalité solidaire qu'est l'univers. Il exprime cette idée de la manière suivante :

« Dans l'ignorance des liens qui les unissent au système entier de l'univers, on les a fait dépendre des causes finales ou du hasard, suivant qu'ils arrivaient et se succédaient avec régularité ou sans ordre apparent : mais ces causes imaginaires ont été successivement reculées avec les bornes de nos connaissances, et disparaissent entièrement devant la saine philosophie, qui ne voit en elles que l'expression de l'ignorance où nous sommes des véritables causes. »³⁹

En résumé, dans le cadre de son déterminisme global qui porte à la fois sur la structure du monde et sur sa prédictibilité en principe par la science, Laplace estime que la notion de

³⁶ Laplace semble affirmer que le déterminisme de la structure de l'univers entier implique le déterminisme scientifique global, c'est-à-dire la prédictibilité, à partir de l'état complet de l'univers à un instant donné et avec les lois de la nature, de tout son état futur et passé. Cependant, l'analyse de certains contextes physiques (par exemple, les phénomènes chaotiques), montre que le déterminisme métaphysique n'implique nullement la prédictibilité de l'état de ce dernier à tout instant dans le temps. De plus, l'examen d'autres contextes physiques (par exemple, les phénomènes quantiques), révèle aussi que l'indéterminisme inhérent à un système physique n'implique pas sa totale imprédictibilité : la possibilité d'évaluer à l'avance, même si c'est en termes probabilistes, l'évolution d'un système physique (quantique) n'est pas mise en cause.

³⁷ Laplace, 1986[1814], *op. cit.*, p. 32.

³⁸ Earman, 1986, *op. cit.* ; Gayon, 1998, *op. cit.* ; Pomian (dir.), 1990, *La querelle du déterminisme* ; Kojève, 1990[1932], *L'idée du déterminisme dans la physique classique et dans la physique moderne* ; Popper, 1984[1982], *L'univers irrésolu. Plaidoyer pour l'indéterminisme* ; Bishop, 2006, « Determinism and Indeterminism ».

³⁹ Laplace, 1986[1814], *op. cit.*, p. 33.

hasard ne peut que désigner l'ignorance propre aux sujets connaissant des causes qui déterminent nécessairement et univoquement l'univers dans sa totalité. C'est en ce sens qu'on lui attribue la notion de hasard subjectif en tant qu'ignorance des causes.

1.5. Le hasard objectif

Nous avons déjà annoncé plus haut que, de manière générale, la notion de hasard objectif désigne une propriété inhérente à la structure du monde réel, indépendamment de la connaissance que nous en avons. Dans cette section, nous allons introduire quatre manières dont on peut parler de hasard objectif : « absence d'un dessein », « rencontre de séries causales indépendantes », « sensibilité aux conditions initiales » et « hasard pur ou indéterministe ».

Avant de procéder à l'analyse de ces notions de hasard, il est nécessaire de souligner que, contrairement à la vision du sens commun à ce sujet, le hasard conçu en tant qu'indépendant de notre connaissance des phénomènes réels peut être incompatible, mais ne l'est pas nécessairement, avec les doctrines du déterminisme métaphysique et scientifique. À la lumière des réflexions d'Aristote, de Cournot et de Poincaré sur la notion de hasard, nous soulignerons que la relation de compatibilité ou d'incompatibilité du hasard objectif avec le déterminisme, en particulier avec sa déclinaison métaphysique, dépend directement de l'extension de ce dernier (autrement dit, du fait qu'il est global ou local, qu'il concerne l'histoire complète de l'univers entier ou d'un système particulier ou un segment de leur évolution, etc.).

En second lieu, nous verrons que certaines des notions de hasard objectif que nous allons présenter ci-dessous ont émergé de réflexions et de théories scientifiques, relevant en particulier de la physique. De plus, nous mettrons en évidence que ces notions de hasard sont toutes évoquées, bien que parfois ce ne soit que de manière implicite ou métaphorique, au sein des disciplines biologiques pour décrire des phénomènes divers, des niveaux moléculaire et cellulaire jusqu'à l'échelle macroévolutive. Dans le cadre de notre travail sur la notion de hasard en biologie, il est donc central de clarifier les significations qui peuvent être attribuées à la notion de hasard objectif, d'en évaluer les implications quant à la question du déterminisme et de l'indéterminisme des phénomènes biologiques. Seule une telle démarche pourra nous permettre d'évaluer ensuite le caractère approprié de ces notions dans le cas qui

nous intéresse plus spécifiquement, c'est-à-dire lorsqu'elles sont utilisées pour caractériser les sources causales de la variation biologique.

1.5.1. Le hasard objectif en tant qu'« absence d'un dessein »

Aristote est un défenseur éminent de la notion de hasard objectif en tant qu'absence d'un dessein (ou « défaut de finalité »), bien que ce ne soit pas la seule manière dont il ait conçu le hasard. Cette notion a ensuite été évoquée, au cours des siècles, en philosophie et en science.⁴⁰ En particulier, nous verrons que, en biologie, elle est sous-jacente à la caractérisation de la variation biologique, plus précisément à la manière dont les biologistes conçoivent les phénomènes à son origine par rapport au cours adaptatif du processus d'évolution. En effet, depuis la formulation de l'hypothèse de l'évolution par sélection naturelle par Darwin jusqu'à aujourd'hui, on considère que la variation au sein des populations naturelles n'a pas lieu en vue de l'adaptation des organismes concernés et, dans ce sens, est le fruit du hasard.⁴¹

Afin de définir cette notion de hasard objectif, analysons la pensée d'Aristote à ce sujet. Dans son traité *De l'interprétation*, il distingue les événements terrestres et sublunaires des événements des sphères célestes externes et, dans ce cadre, il envisage longuement la notion de contingence (*τό ἐνδεχόμενον*). Pour Aristote, les événements des sphères célestes sont nécessaires et les événements terrestres et sublunaires sont contingents, c'est-à-dire qu'ils sont conçus comme pouvant être ou ne pas être. Cette distinction s'insère dans une vision déterministe, au sens métaphysique, du monde qui est moins exigeante que le déterminisme scientifique et universel laplacien car elle n'englobe pas les événements de l'univers entier dans une totalité solidaire régie, dans son ensemble, par une même nécessité nomique. En effet, selon Aristote tout événement est dû à une cause qui, à l'échelle des phénomènes terrestres et sublunaires, peut être par soi (déterminée) ou par accident (indéterminée) : elle peut découler de la substance d'une chose ou bien d'un attribut qui n'appartient pas à la

⁴⁰ Entre autres, cette manière de concevoir le hasard objectif, en particulier en tant que négation de tout dessein divin, est mobilisée par les scientifiques défendant la théorie de l'évolution afin de contrer les arguments des partisans du « Dessein Intelligent ». Ce dernier correspond à l'idée spéculative et nullement scientifique selon laquelle tout être dans la nature est tellement complexe et bien adapté à son propre environnement qu'il ne peut pas être le résultat de processus aléatoires, mais seulement le produit planifié de manière rationnelle par une sorte d'Architecte Divin.

⁴¹ Dans la deuxième partie de notre travail, nous clarifierons la signification précise de la notion de hasard lorsqu'elle est utilisée en référence à l'origine de la variation biologique dans le cadre de la théorie de l'évolution.

définition substantielle d'une chose. Elle est à l'origine, dans le premier cas, de ce qui se produit toujours ou fréquemment et, dans le deuxième cas, de ce qui ne se produit que très rarement⁴². En revanche, une fois qu'une cause par accident est présente, elle produit de manière nécessaire son effet aussi bien qu'une cause par soi.

C'est dans le deuxième livre de *La physique*⁴³ qu'Aristote expose ce qu'on pourrait appeler sa théorie du hasard. Il distingue, d'une part, les phénomènes nécessaires et constants et les phénomènes qui se produisent la plupart du temps et, d'autre part, les phénomènes qui se produisent par exception à ceux-là. Il affirme que ces derniers existent et sont engendrés par la « spontanéité » (*τό αὐτόματον*) et par la « fortune » (*ἡ τύχη*)⁴⁴ qui sont des causes « par accident » (*συμβεβηκότως*).

« [...] la fortune et le hasard [spontanéité] sont des causes par accident, pour des choses susceptibles de ne se produire ni absolument, ni fréquemment, et en outre susceptibles d'être produites en vue d'une fin. »⁴⁵

Aristote explique qu'on attribue un effet à la spontanéité ou à la fortune si et seulement si cet effet est obtenu de manière accidentelle dans le domaine des événements qui sont produits en vue d'une fin. Autrement dit, le « hasard » (la spontanéité et la fortune) intervient dans le monde en tant que cause (par accident) lorsqu'une rupture se produit entre l'ordre des causes efficientes et l'ordre des causes finales qui, au contraire, sont associées la plupart du temps, en particulier dans les actions humaines car un individu agit en vue d'une fin. Aristote remarque que ce type de ruptures se produit souvent dans le monde humain comme dans le monde de la nature et, entre autres, il en donne l'exemple suivant :

« [...] le fait pour un homme de venir sur la place par fortune, et d'y rencontrer celui qu'il voulait, mais sans qu'il y eût pensé, a pour cause le fait d'avoir voulu se rendre sur la place pour affaires [...] »⁴⁶.

Selon Aristote, il y a spontanéité (*τό αὐτόματον*) – terme désignant l'idée d'un processus automatique qui se meut de soi-même – lorsqu'un fait se révèle avantageux pour un être sans aucune intentionnalité et dont la cause ne peut que lui être extérieure. Quant à la

⁴² Aristote, *La métaphysique*, D, 30.

⁴³ Aristote, *La physique*, II, 5, 195b 31-198a 53.

⁴⁴ Les termes grecs « *τό αὐτόματον* » et « *ἡ τύχη* » sont habituellement traduits en français par le mot « hasard ».

⁴⁵ *Ibid.*, II, 5, 197a 32-36.

⁴⁶ *Ibid.*, II, 4, 36.

fortune (*ἡ τύχη*), elle concerne un être doué de la capacité de décision et qui aurait donc pu imaginer le but de l'événement qui s'est produit, donc un être humain ; en d'autres termes, même si cet événement est un effet purement accidentel et n'a rien d'intentionnel, l'être concerné aurait pu le souhaiter, le craindre, le vouloir ou vouloir l'empêcher.⁴⁷ Dans les deux cas, Aristote insiste sur le fait que l'interférence entre les deux ordres de causes, efficaces et finales, doit avoir un sens et importer à l'individu concerné par l'événement produit par accident⁴⁸.

Pour conclure en ce qui concerne la notion de hasard objectif en tant qu'absence d'un dessein, nous soulignons qu'elle se distingue de l'idée bergsonienne selon laquelle le hasard réside dans le fait que des événements qui ne sont pas intentionnels ni le fruit d'un dessein nous paraissent comme le produit d'une causalité intentionnelle⁴⁹. En effet, la notion de hasard en tant qu'absence d'un dessein est définie par les propriétés de la structure causale sous-jacente à un événement dans le monde et n'est nullement relative au sentiment humain de surprise face aux caractéristiques de ce dernier.

1.5.2. Le hasard objectif en tant que « rencontre de séries causales indépendantes »

La philosophie de Cournot représente une tentative de concevoir le hasard comme objectif sans pour autant nier le déterminisme inhérent à la structure du monde. Sa réflexion sur la notion de hasard rappelle la pensée d'Aristote, en particulier la conception de la notion d'accident, que nous allons présenter plus en détail dans cette sous-section. De plus, nous verrons que la conception cournotienne du hasard figure de manière plus ou moins explicite dans la réflexion au sein de plusieurs disciplines scientifiques, en particulier en biologie de l'évolution. Dans *Le hasard et la nécessité* (1970), le biologiste moléculaire Monod fournit un exemple paradigmatique de l'utilisation de cette notion, qu'il utilise afin de mettre en évidence

⁴⁷ D'ailleurs, selon Aristote la spontanéité est une notion plus ample dont la fortune est un cas particulier car les faits dus à la première forme de hasard peuvent avoir pour cause déterminée une nature quelconque, alors que les faits dus à la fortune ne peuvent avoir qu'une pensée comme cause.

⁴⁸ Nous examinerons davantage la notion d'accident chez Aristote dans le cadre de la section suivante. En effet, bien qu'elle soit au fondement de la notion de hasard objectif en tant qu'absence d'un dessein chez Aristote, elle est au cœur d'un deuxième volet de cette notion aristotélicienne qui, nous le verrons, correspond plutôt à l'idée du hasard objectif en tant que « rencontre de séries causales indépendantes ».

⁴⁹ Bergson, 1948[1932], *Les deux sources de la morale et de la religion*.

l'opposition entre le caractère fortuit des mutations génétiques et la nécessité propre au fonctionnement téléonomique⁵⁰ des organismes vivants.

Afin de définir le hasard objectif comme rencontre de séries causales indépendantes, analysons la pensée de Cournot à ce sujet. Dans son livre intitulé *Exposition de la théorie des chances et des probabilités* (1984[1843]), il formule d'abord ce qu'il appelle « le principe souverain et régulateur de la raison humaine » selon lequel « aucun phénomène ou événement n'est produit sans cause » : c'est le « principe de causalité » que Cournot conçoit comme « une règle absolue et nécessaire »⁵¹. Ensuite, il formule sa définition générale du hasard :

« Les événements amenés par la combinaison ou la rencontre de phénomènes qui appartiennent à des séries indépendantes, dans l'ordre de la causalité, sont ce qu'on nomme des événements *fortuits* ou des résultats du *hasard*. »⁵²

On peut immédiatement remarquer deux aspects qui, d'une part, rapprochent la notion de hasard de Cournot de la réflexion aristotélicienne au sujet de la spontanéité, de la fortune et de la structure déterministe du monde⁵³ et, d'autre part, l'opposent à la conception laplacienne du hasard subjectif en tant qu'ignorance des causes.

Une première différence avec la pensée de Laplace est évidemment la nature objective du hasard chez Cournot : ce dernier, de la même manière qu'Aristote, utilise cette notion pour désigner une propriété inhérente aux choses et non pas liée à la connaissance que nous en avons. Plus précisément, la notion de hasard chez Cournot désigne une certaine configuration des liens de causalité entre événements dans le monde.

En second lieu, la conception du monde de Cournot est déterministe selon une perspective métaphysique, mais se distingue du déterminisme global – scientifique et métaphysique – de Laplace. En effet, dans un univers déterministe au sens laplacien, la notion

⁵⁰ Selon Monod, la téléonomie est une propriété des organismes vivants et correspond au fait d'être doué d'un projet. Il définit le projet téléonomique essentiel de tout organisme vivant comme la transmission, d'une génération à la suivante, du contenu d'invariance caractéristique de l'espèce à laquelle il appartient (Monod 1970, *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, p. 30).

⁵¹ Cournot, 1984[1843], *Exposition de la théorie des chances et des probabilités*, p. 53.

⁵² *Ibid.*, p. 55. Cournot même, dans l'*Essai sur les fondements de nos connaissances et sur le caractère de la critique philosophique* (1975[1851]), signale que Laplace avait déjà formulé une définition du hasard semblable dans le *Traité des jeux de hasard* (1714, La Haye, cité par Cournot 1975[1851], *op. cit.*). Voir Lalande, 2002[1926], *Vocabulaire technique et critique de la philosophie*, p. 402. En outre, Mill (1866[1843], *Système de logique déductive et inductive, exposé des principes de la preuve et des méthodes de recherche scientifique*) avait aussi proposé une définition du hasard en termes d'indépendance des causes.

⁵³ D'ailleurs, Cournot fournit beaucoup d'exemples qui reviennent à ceux d'Aristote, comme celui d'un voyageur qui prend un train et est victime d'un déraillement. Cependant, il ne faut pas oublier qu'entre les deux philosophes, il y a eu le calcul des probabilités : Cournot peut donc exprimer aussi bien la notion de hasard dans le langage des jeux (Cournot, 1979[1875], *Matérialisme, vitalisme, rationalisme*, p. 183-183).

de hasard en tant que rencontre de séries causales indépendantes n'a pas de sens car la totalité des événements dans l'univers entier est conçue comme formant un tout solidaire régi par une même nécessité nomique ; selon Laplace, tout événement est en principe prévisible d'une façon infiniment précise et univoque. Au contraire, chez Cournot, on retrouve plutôt un déterminisme métaphysique local concernant des segments du monde. Il conçoit ce dernier comme relativement cloisonné et permet donc que des événements puissent appartenir à des îlots causaux distincts qui ne sont pas solidaires mais indépendants entre eux du point de vue nomique.⁵⁴

Afin de mieux comprendre la vision déterministe du monde de Cournot et la place qu'il donne au hasard dans ce cadre, il est intéressant de rappeler la distinction qu'il fait entre « raison » et « cause ». Les raisons des phénomènes sont les lois constantes qui sont valables, en tout temps et en tout lieu, par une nécessité qui tient à l'essence permanente des choses. Les événements dépendants, qui font partie de la même série causale, sont liés par une raison commune. Ce n'est pas le cas pour les événements appartenant à des séries causales indépendantes qui, au contraire, n'ont pas de raison commune, mais appartiennent à l'ordre des causes, des faits, c'est-à-dire des conditions initiales sur lesquelles portent les lois. C'est à ce niveau que le hasard peut apparaître, selon Cournot, sans entrer en contradiction avec la nécessité des lois universelles et immuables de la nature.

Dans *Probabilités et critique philosophique selon Cournot* (1996), le philosophe français Martin propose l'argument suivant afin de rendre compte de la compatibilité, chez Cournot, de la notion de hasard objectif en tant que rencontre de séries causales indépendantes et de sa vision déterministe du monde. Étant donné deux séries causales X et Y, avec leur trajectoire bien déterminée – la série X amenant l'individu A au point M à l'instant t et la série Y amenant l'individu B au point M au même instant t –, le fait qu'elles vont se rencontrer semble bien déterminé et même nécessaire. Mais la rencontre de ces deux séries est fortuite dans la mesure où elles sont indépendantes, c'est-à-dire qu'il n'y a, dans la série X amenant A en M à l'instant t , aucune raison que la série Y amène B au même point M et au même instant t . Ce qui est nécessaire n'est pas la rencontre des deux séries en elle-même, mais la rencontre *sachant que* A et B seront au même lieu et en même temps.

⁵⁴ Voir Saint-Sernin, 1998, *Cournot. Le réalisme* ; voir Gayon, 2005, « Chance, Explanation, and Causation in Evolutionary Theory ». La vision du monde de Cournot n'implique donc pas que, à partir de l'état complet de l'univers entier et des lois régissant son évolution, tout son état futur et passé est déterminé de manière univoque et nécessaire. Elle n'implique pas non plus que, à partir de la connaissance infiniment précise de l'état actuel de l'univers entier et des lois régissant son évolution, on puisse calculer rigoureusement aussi bien l'ensemble de ses états futurs que ses états antécédents.

« On peut dire alors que la rencontre est objectivement fortuite, mais s'offre subjectivement comme nécessaire, pour celui qui considère les séries déjà réalisées et figées dans leur invariabilité de fait. Ou encore, de même qu'on distingue un hasard objectif dont l'existence réelle est affirmée et un hasard subjectif résultant de notre ignorance, on peut, pour comprendre la pensée de Cournot, distinguer une *nécessité objective* d'une *nécessité subjective*, mais en un sens inverse ; la subjectivité de la nécessité supposant non pas l'ignorance, mais au contraire la pleine connaissance de l'intégralité du phénomène. »⁵⁵

Si l'on niait l'objectivité de ce que la notion de hasard désigne, on refuserait par avance aux séries causales leur possibilité de varier, et ainsi on établirait une solidarité entre elles qui n'existe pas, quand en fait elles sont indépendantes et ne sont invariables qu'après coup seulement. On ne peut déterminer parfaitement la rencontre de deux séries causales indépendantes qu'en les reliant dans un seul et même système, ce qui n'est pas légitime. Mais si on le fait, en ce cas-là c'est *pour nous* que la rencontre est invariablement déterminée, et non par l'une ou l'autre des deux séries.

Analysons brièvement la relation entre les deux notions de hasard objectif que nous avons introduites en nous appuyant respectivement sur la pensée d'Aristote et de Cournot. Nous avons déjà indiqué que la conception du hasard objectif chez Aristote contient en son sein à la fois la notion d'absence d'un dessein et celle de rencontre de séries causales indépendantes.⁵⁶ C'est plus précisément dans le cadre de la définition de l'accident qu'Aristote anticipe la pensée de Cournot à ce sujet. Nous rappelons que la spontanéité et la fortune (le hasard) sont pour Aristote des causes par accident, c'est-à-dire qu'elles ne découlent pas de la substance d'une chose et sont à l'origine de ce qui advient par exception aux événements nécessaires et fréquents. Dans *La métaphysique*, Aristote fournit deux définitions de l'accident :

« Accident se dit de ce qui appartient à un être et peut en être affirmé avec vérité, mais n'est pourtant ni nécessaire ni constant [...] tout attribut qui appartient à un sujet, mais non parce que le

⁵⁵ Martin 1996, *Probabilités et critique philosophique selon Cournot*, p. 128-129.

⁵⁶ Delsol (1989, *Cause, Loi, Hasard en Biologie*, p. 200) suggère que l'idée de Cournot de la rencontre de séries causales indépendantes contient en son sein la notion aristotélicienne de hasard au sens d'absence d'un dessein (ou défaut de finalité). Pour montrer cela, il apporte l'exemple d'un événement qui pourrait être fortuit au sens de Cournot mais qui, au contraire, se produit en vue d'une fin. La situation que Delsol décrit est la suivante : un conducteur de train décide d'accélérer pour écraser une voiture qu'il voit passer sur les voies car il reconnaît dans la personne au volant son ennemi mortel ; de cette manière, le conducteur rend la série causale qui dirige le train dépendante de la série causale qui dirige la voiture de l'ennemi vers les voies. En absence de la décision du conducteur du train ces deux séries causales auraient été indépendantes et leur rencontre aurait pu se produire en absence de toute finalité (par hasard).

sujet était précisément ce sujet, ou le temps, ce temps, ou le lieu, ce lieu, cet attribut sera un accident [...] » ;

« Accident s'entend encore d'une autre façon : c'est ce qui, fondé en essence dans un objet, n'entre cependant pas dans sa substance [...] »⁵⁷

Nous proposons de réinterpréter et de redéfinir l'accident chez Aristote en termes de hasard objectif en tant que rencontre de séries causales indépendantes. L'accident peut être compris comme la rencontre, dans une même chose, de ce qui appartient à sa définition substantielle propre – sa quiddité, ce qu'elle est par soi – et d'un attribut qui ne fait pas nécessairement partie de sa substance et qui est donc, dans ce sens, accidentel. Par exemple, en utilisant une situation imaginée par Aristote même⁵⁸, le fait qu'un architecte soit médecin est un accident au sens où cela est dû à la rencontre, dans le même individu, de sa capacité à construire des maisons – ce qui fait partie de sa définition substantielle en tant qu'architecte – et de sa capacité à produire de la santé – qui n'est qu'un attribut externe à la substance de cet individu en tant qu'architecte.

En conclusion de cette section consacrée à la notion de hasard objectif en tant que rencontre de séries causales indépendantes, il est intéressant de préciser la manière dont Monod utilisa cette notion pour caractériser les mutations génétiques. Dans *Le hasard et la nécessité* (1970), Monod utilise le terme « hasard » en référence aux mutations génétiques pour désigner un fait insolite, exceptionnel, inattendu, qui s'oppose à tout ce qui, au contraire, est constant, régulier, prédictible. Le hasard, c'est-à-dire l'inconstant, l'irrégulier, l'imprédictible, essentiellement opposé à l'ordre et à la régularité, s'introduit à travers les mutations dans un système soumis à la nécessité, l'organisme vivant caractérisé par l'invariance et la téléonomie. Par cette inscription, le hasard devient nécessaire. En effet, la mutation fortuite va s'intégrer au tout de la nécessité à travers l'action médiatrice de la sélection naturelle, qui l'accepte ou la rejette si elle conserve, renforce et enrichit la compatibilité de l'organisme vivant avec le tout. La notion de hasard en tant que rencontre de séries causales indépendantes émerge lorsqu'il discute la relation entre « une erreur dans la réplication du message génétique et ses conséquences fonctionnelles »⁵⁹. Le hasard se situe à la rencontre de deux événements qui n'appartiennent pas à un même enchaînement causal : il

⁵⁷ Aristote, *La métaphysique*, D, 30.

⁵⁸ *Ibid.*, E, 2.

⁵⁹ Monod, 1970, *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, p. 149.

s'agit, d'une part, de la série causale se déroulant au niveau moléculaire et aboutissant à une altération de la séquence d'ADN d'un organisme individuel donné ; d'autre part, de la série causale à l'origine des conditions environnementales au sens large (internes et externes) où l'organisme concerné se trouve et, par conséquent, à l'origine de la détermination des besoins adaptatifs de ce dernier.

« Les événements élémentaires initiaux qui ouvrent la voie de l'évolution à ces systèmes intensément conservateurs que sont les êtres vivants sont microscopiques, fortuits et sans relation aucune avec les effets qu'ils peuvent entraîner dans le fonctionnement téléonomique. »⁶⁰

Ce passage nous révèle que Monod n'évoque pas simplement la formulation de Cournot, mais aussi bien la notion aristotélicienne de hasard en tant qu'absence d'un dessein (ou défaut de finalité), pour caractériser les mutations génétiques. En effet, il souligne que les mutations ne se produisent pas en vue de l'adaptation des organismes individuels concernés et qu'il n'y a donc aucune finalité à l'origine causale de tels phénomènes biologiques.

1.5.3. Le hasard objectif en tant que « sensibilité aux conditions initiales »

Une troisième notion appartenant au visage objectif du hasard émerge du cadre de la théorie du chaos. Cette dernière décrit le comportement de systèmes dynamiques instables car hautement sensibles aux conditions initiales. Dans ce cadre, la notion de hasard correspond à l'idée même que des différences minimales dans l'état d'un système chaotique à un instant donné peuvent engendrer, au bout d'un certain temps, des différences importantes dans son état aux instants futurs. C'est la notion de hasard en tant que sensibilité aux conditions initiales.

Cette notion de hasard est invoquée non seulement en physique, mais aussi dans le cadre de diverses disciplines biologiques. Les biologistes en font surtout une utilisation métaphorique pour exprimer le caractère imprédictible de certains effets à partir de certaines conditions initiales. Par exemple, elle est utilisée en biologie moléculaire pour exprimer le fait qu'une altération chimique très petite affectant la séquence nucléotidique de l'ADN (par exemple, la substitution d'une seule base nucléotidique) peut avoir une incidence très

⁶⁰ Monod, 1970, *op. cit.*, p. 155.

importante au niveau de la protéine synthétisée à partir de la séquence modifiée et peut même avoir des conséquences disproportionnées et radicalement différentes au niveau des traits phénotypiques de l'organisme concerné. Cette même idée est exprimée, entre autres, par le philosophe américain Beatty⁶¹ lorsqu'il présente les observations et les études de Darwin sur les variétés d'orchidées. Plus précisément, Beatty souligne que la grande diversité d'espèces existantes de cette plante est très probablement le résultat de petits événements biochimiques, que sont les mutations génétiques, se produisant au sein des populations ancestrales et très semblables d'orchidées. Enfin, le paléontologue américain Gould⁶² évoque systématiquement la notion de sensibilité aux conditions initiales pour expliquer l'idée selon laquelle, si l'on pouvait rebobiner et dérouler à nouveau le film de la vie, l'évolution suivrait chaque fois une voie différente de celle que nous connaissons.

Examinons cette notion de hasard au sein du cadre théorique d'où elle a émergé. La théorie du chaos⁶³ n'a été développée qu'à partir des années 1960-1970.⁶⁴ Cependant, à la fin du XIX^e siècle, Poincaré et Hadamard avaient déjà étudié des systèmes très simples caractérisés par un petit nombre de degrés de liberté, mais dont le comportement était très instable (respectivement, un système à trois corps célestes en interaction gravitationnelle et un modèle mathématique abstrait décrivant le mouvement d'une particule glissant sans friction sur une surface à courbure négative). Ils avaient déjà eu l'intuition qu'une telle instabilité dynamique était due au fait que ces systèmes étaient hautement sensibles aux changements des conditions initiales.

L'idée du hasard en tant que sensibilité aux conditions initiales a été discutée de manière approfondie par Poincaré. Dans *Science et méthode* (1920[1908]), Poincaré dédie un chapitre entier (le quatrième) à la notion de hasard. Il commence son analyse en posant la question suivante :

« Comment oser parler des lois du hasard ? Le hasard n'est-il pas l'antithèse de toute loi ? ». La probabilité est opposée à la certitude ; c'est donc ce qu'on ignore et par conséquent semble-t-il ce

⁶¹ Beatty, 2006, « Chance Variation: Darwin on Orchids ».

⁶² Gould, 1989[1991], *La vie est belle. Les surprises de l'évolution*.

⁶³ Le terme « chaos » n'a été introduit qu'en 1975 par les mathématiciens Li et Yorke.

⁶⁴ Pour une introduction à la fois historique et critique à la théorie du chaos, nous renvoyons à Gleick, 1989, *La théorie du chaos* ; Dalmedico et al (dir.), 1992, *Chaos & déterminisme* ; Ruelle, 1991, *Hasard & chaos* ; Ekeland, 1995, *Le chaos* ; Charpentier et al (dir.), 2006, *L'héritage scientifique de Poincaré* ; Franceschelli et al (dir.) 2007, *Chaos et systèmes dynamiques. Éléments pour une épistémologie*.

qu'on ne saurait calculer. Il y a là une contradiction au moins apparente et sur laquelle on a déjà beaucoup écrit. »⁶⁵

Ensuite, afin de répondre à la question « qu'est-ce que le hasard ? », il critique la tradition selon laquelle le hasard n'est que la mesure de l'ignorance humaine des causes sous-jacentes à un événement⁶⁶ et, en revendiquant l'appartenance de la notion qu'il propose au visage objectif du hasard, il caractérise les faits qu'habituellement on regarde comme le fruit du hasard de la manière suivante :

« Une cause très petite, qui nous échappe, détermine un effet considérable que nous ne pouvons pas ne pas voir, et alors nous disons que cet effet est dû au hasard. Si nous connaissions exactement les lois de la nature et la situation de l'univers à l'instant initial, nous pourrions prédire exactement la situation de ce même univers à un instant ultérieur. Mais, lors même que les lois naturelles n'auraient plus de secrets pour nous, nous ne pourrions connaître la situation initiale qu'*approximativement*. Si cela nous permet de prévoir la situation ultérieure *avec la même approximation*, c'est tout ce qu'il nous faut, nous disons que le phénomène a été prévu, qu'il est régi par des lois ; mais il n'en est pas toujours ainsi, il peut arriver que des petites différences dans les conditions initiales en engendrent de très grandes dans les phénomènes finaux ; une petite erreur sur les premières produirait une erreur énorme sur les derniers. La prédiction devient impossible et nous avons le phénomène fortuit. »⁶⁷

Afin d'explicitier une telle notion de hasard objectif, Poincaré utilise de nombreux exemples parmi lesquels figurent l'instabilité d'un cône qui repose sur sa pointe, l'instabilité et l'imprédictibilité des phénomènes météorologiques et le caractère imprédictible de l'issue du jeu de la roulette.⁶⁸ À la lumière de chacun de ces exemples, son analyse aboutit toujours à la même conclusion :

« Voici donc une fois de plus une petite cause et un grand effet ; ou mieux des petites différences dans la cause et de grandes différences dans l'effet. »⁶⁹

Selon Poincaré, la notion de hasard désigne donc une certaine disproportion entre la cause d'un phénomène et son effet : cette notion doit donc être invoquée lorsque des différences très petites dans les conditions initiales d'un système donné (l'ordre des causes) engendrent des conséquences importantes dans son état final (l'ordre des effets). C'est à cause

⁶⁵ Poincaré, 1920[1908], *Science et méthode*, p. 64.

⁶⁶ *Ibid.*, p. 66.

⁶⁷ *Ibid.*, p. 68-69, italiques dans l'original.

⁶⁸ *Ibid.*, p. 68, p. 29, p. 71.

⁶⁹ *Ibid.*, p. 70.

d'une telle disproportion entre causes minimales et effets considérables, et non pas simplement à cause du fait qu'on ne peut connaître qu'approximativement les conditions initiales, que la prédiction de l'évolution de ce genre de systèmes physiques⁷⁰ devient impossible et qu'on est obligé d'utiliser le calcul des probabilités.⁷¹

Si l'on parle de Poincaré comme d'un philosophe appartenant à la tradition du hasard objectif et, plus généralement, si l'on caractérise la notion de sensibilité aux conditions initiales comme faisant partie du visage objectif du hasard, il est nécessaire de s'interroger sur la signification du terme « objectif » dans ce cas. En effet, même si la notion de hasard en tant que sensibilité aux conditions initiales semble désigner une caractéristique de certains systèmes physiques, la manière dont elle est définie ne nous empêche pas de l'envisager comme appartenant à la tradition du hasard subjectif. La caractérisation des causes en tant que très petites et des effets comme importants par rapport à ces dernières semble supposer l'intervention d'un critère d'évaluation qui repose sur les sens et les jugements humains, c'est-à-dire sur quelque chose de subjectif.

En ce qui concerne spécifiquement la pensée de Poincaré, il refuse la notion de hasard subjectif et soutient que l'objectivité est ce qui est valable pour tous les hommes, et non pas pour tel ou tel autre⁷². De manière générale, il est important de souligner que la disproportion entre l'ordre des causes et l'ordre des effets en ce qui concerne un système donné peut être interprétée soit comme une propriété relative à un sujet connaissant et donc due à la perception humaine, soit comme une propriété du système réel qui découle des caractéristiques propres à la structure de ce dernier, soit comme une propriété du modèle mathématique utilisé pour décrire le système réel et pour en prédire l'évolution.

Dans le premier cas, on devrait parler de hasard subjectif et attribuer donc l'imprédictibilité de l'évolution du système considéré à l'ignorance humaine des causes qui lui sont sous-jacentes. Mais on pourrait aussi comprendre la sensibilité aux conditions

⁷⁰ On parle d'« amplification exponentielle » ou de « progression géométrique » pour désigner le fait que l'instabilité des systèmes chaotiques (par exemple, des phénomènes météorologiques, du phénomène de turbulence hydrodynamique, de certaines réactions chimiques, etc.) rend leur comportement de plus en plus imprédictible au cours du temps.

⁷¹ Poincaré, 1920[1908], p. 68-71. Dans *Science et méthode*, Poincaré introduit deux autres notions de hasard dont il revendique aussi la nature objective. D'une part, il définit le hasard comme la complexité et la multiplicité des causes d'un phénomène : dans ce cas, la notion de hasard s'applique à des événements que l'on n'arrive pas à prévoir car on aurait besoin d'un ensemble de connaissances trop difficiles, complexes, nombreuses et précises pour le faire (p. 73-76). D'autre part, Poincaré reprend la notion de hasard de Cournot en tant que rencontre de séries causales indépendantes mais il ne l'analyse pas en détail, car il la considère explicitement comme moins importante que les notions de hasard en tant que complexité et multiplicité des causes et en tant que sensibilité aux conditions initiales. À son avis, ces dernières sont toujours au fondement du hasard en tant que rencontre de séries causales indépendantes (p. 76-78).

⁷² *Ibid.*, p. 88.

initiales d'un système de la même manière que Poincaré, c'est-à-dire en utilisant le terme « objectif » pour désigner le fait que la petitesse des causes, l'importance des effets et leur conséquente disproportion ne sont pas dues à la perception d'un seul individu mais sont l'objet d'un consensus intersubjectif et, dans ce sens, sont objectives.

Dans le deuxième cas, où la disproportion entre l'ordre des causes et l'ordre des effets est considérée comme une propriété du système réel, il est évident qu'on peut parler de hasard objectif pour désigner le fait que des petites différences au niveau des causes provoquent des différences importantes au niveau des effets. Dans ce cas, l'imprédictibilité de l'évolution du système considéré ne peut pas être attribuée simplement à l'ignorance humaine des causes, bien que les limites de nos connaissances puissent aussi y jouer un rôle, mais aussi et en premier lieu au caractère intrinsèquement instable du système même.

Dans la troisième situation, où la disproportion entre causes et effets est une propriété du modèle mathématique utilisé pour décrire un système réel et pour en prédire l'évolution, la question de la nature objective de la notion de hasard en tant que sensibilité aux conditions initiales devient plus délicate. D'une part, il faut considérer qu'un modèle ne fournit jamais une description complète, mais toujours partielle et approximative, d'un système réel : les propriétés de ce dernier ne peuvent être décrites que sous certains égards et selon des perspectives particulières, non pas de manière complète et absolue. Par conséquent, si la disproportion entre les événements appartenant à l'ordre des causes et les événements de l'ordre des effets au sein d'un système donné était une caractéristique du modèle décrivant ce dernier, elle pourrait être interprétée comme une propriété subjective car propre à une description particulière du système considéré. Dans ce cas, l'imprédictibilité de l'évolution du système ne serait due qu'aux bornes des connaissances humaines au sujet de sa structure causale. D'autre part, la même propriété, qu'est la disproportion entre les différences au niveau des causes et les différences qu'elles produisent en tant qu'effets, pourrait être considérée comme objective si elle se révélait être invariante (ou robuste) sous un ensemble divers de modèles décrivant tous le même système réel mais de manière différente (autrement dit, sous un ensemble de transformations du modèle décrivant le système réel considéré)⁷³. Dans ce cas, l'imprédictibilité de l'évolution dans le temps du système considéré ne serait pas simplement due aux limites des connaissances humaines à son sujet, mais dépendrait des

⁷³ Dans ce cas, nous ne faisons pas référence à la notion d'objectivité métaphysique, mais à l'objectivité en tant qu'invariance (ou robustesse) par rapport à un ensemble divers de perspectives. Pour une analyse approfondie de cette notion d'objectivité, nous renvoyons au chapitre 10.

caractéristiques communes à l'ensemble de modèles utilisés pour décrire ce système et pour en prédire l'évolution.

Avant d'en venir à la présentation et à l'analyse d'une dernière notion de hasard objectif qu'est le « hasard pur » caractérisant plus particulièrement les phénomènes quantiques, analysons la relation entre la notion de hasard en tant que sensibilité aux conditions initiales et les doctrines du déterminisme métaphysique, scientifique et méthodologique.⁷⁴

D'abord, lorsqu'on le conçoit selon une perspective métaphysique, le déterminisme est compatible avec toutes les interprétations présentées ci-dessus de la propriété qu'est la sensibilité aux conditions initiales. En effet, quelle que soit la manière dont la disproportion entre les différences dans les causes et les différences dans les effets est interprétée, elle ne met pas en question le déterminisme du système réel considéré, c'est-à-dire le fait que ses conditions initiales à un instant donné, avec les lois régissant son évolution, déterminent de manière univoque et nécessaire tous ses états futurs.

Ensuite, en ce qui concerne la thèse épistémologique du déterminisme scientifique, nous pouvons tout de suite remarquer qu'elle n'est compatible avec la sensibilité aux conditions initiales que si celle-ci est interprétée de manière subjective en tant que relative aux connaissances et aux jugements humains. Au contraire, si la disproportion entre différences dans l'ordre des causes et différences dans l'ordre des effets est considérée comme une propriété du système réel ou comme une caractéristique objective de tout modèle pouvant le représenter, au sens expliqué plus haut, la thèse épistémologique du déterminisme scientifique est ébranlée.

Enfin, le déterminisme méthodologique ne nous semble être incompatible avec aucune interprétation de la sensibilité aux conditions initiales vu que cette dernière ne peut se révéler comme un problème que pour la prédictibilité scientifique, et non par rapport à une vision déterministe de la structure du monde.

⁷⁴ Nous ne discuterons pas la relation que la notion de hasard en tant que sensibilité aux conditions initiales entretient avec la notion d'indéterminisme. En effet, dans cette section nous n'avons évoqué que les systèmes chaotiques déterministes alors qu'il y a aussi une théorie du chaos quantique. En ce qui concerne cette dernière et son interprétation indéterministe, nous renvoyons à la section suivante où nous introduirons deux notions de « hasard pur » émergeant du cadre de la théorie quantique.

1.5.4. Le « hasard pur » du niveau quantique

Une dernière facette du hasard objectif est représentée par la notion que nous appelons « hasard pur » (ou « hasard indéterministe »). Elle émerge du cadre d'une interprétation indéterministe et réaliste de la théorie physique des phénomènes quantiques (atomiques et subatomiques). Selon un certain nombre de scientifiques et de philosophes des sciences en faveur de cette interprétation, l'indéterminisme est limité au niveau microscopique (quantique). Par exemple, en biologie, « l'idée semble être que les organismes et les populations sont composés de trop nombreuses particules élémentaires pour que la chance du niveau microscopique 'percole' au niveau macroscopique que la biologie évolutionniste décrit »⁷⁵. Cependant, certains biologistes et philosophes de la biologie⁷⁶ ont essayé de montrer que le hasard du niveau quantique peut influencer les phénomènes biologiques du niveau moléculaire (par exemple, les mutations génétiques) et, de cette manière, « percoler » jusqu'à l'échelle macroscopique que la théorie de l'évolution décrit. En dépit du débat suscité par cette hypothèse, au cours des dix dernières années, au sein de la communauté des philosophes de la biologie, nous avons décidé de ne pas nous prononcer sur ce point. Nous justifierons ce choix, qui découle du fait que nous adhérons à une position agnostique du point de vue métaphysique, dans la section suivante.

Dans le cadre de cette section, nous ne prétendons pas fournir une discussion détaillée, à la fois du point de vue scientifique, historique et philosophique, des idées fondamentales de la mécanique quantique. Nous souhaitons simplement donner une idée de la raison pour laquelle la notion de hasard pur est particulièrement liée à cette théorie scientifique qu'est la mécanique quantique.⁷⁷ De plus, nous n'entendons pas nous engager pour une interprétation particulière de la théorie quantique et encore moins pour une position métaphysique quant à la nature fondamentalement déterministe ou indéterministe des phénomènes du monde microscopique (quantique). À présent, nous souhaitons simplement mettre en évidence les implications qu'une interprétation indéterministe et réaliste de la mécanique quantique aurait

⁷⁵ Sober, 1984, *The Nature of Selection. Evolutionary Theory in Philosophical Focus*, p. 121.

⁷⁶ Entre autres, Brandon & Carson, 1996, « The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables' Proof But No Room for Determinism Either » ; Stamos, 2001, « Quantum Indeterminism and Evolutionary Biology ».

⁷⁷ Pour une introduction technique mais accessible à la mécanique quantique, nous renvoyons à Ismael, 2008, « Quantum Mechanics ».

en ce qui concerne les doctrines du déterminisme et de l'indéterminisme et la notion de hasard.

Selon l'un des principes centraux de la mécanique quantique qu'est le « principe d'indétermination » énoncé par Heisenberg en 1927, les propriétés d'un système quantique, par exemple la position et la quantité de mouvement d'un électron, ne sont pas indépendantes mais liées par des relations d'indétermination : dans le cas d'un électron, la détermination de sa position avec une précision croissante rend sa quantité de mouvement de plus en plus indéterminée.⁷⁸ Ce principe peut être compris de manière différente selon qu'on interprète la mécanique quantique comme une théorie indéterministe ou comme une théorie déterministe.

Selon l'interprétation indéterministe dominante, soutenue par les physiciens de l'école de Copenhague⁷⁹, les valeurs exactes des propriétés d'un système microscopique, par exemple de la position et de la quantité de mouvement d'une particule quantique, ne peuvent pas en principe être connues : ces propriétés et, plus généralement, la notion classique d'état d'un système ne figurent pas dans le cadre de la mécanique quantique.

Selon les interprétations déterministes de la mécanique quantique, comme la théorie à variables cachées non-locales de Bohm⁸⁰, le principe introduit par Heisenberg est plutôt un principe d'incertitude exprimant notre incapacité théorique à connaître de manière précise toute propriété d'un système quantique. En effet, selon cette théorie, un système quantique possède un certain nombre de propriétés que nous ignorons (des variables cachées) et qui sont régies par des lois déterministes : il est donc en principe possible de décrire et de prédire l'évolution du système quantique par une théorie sous-jacente qui n'a rien de probabiliste mais qui est tout à fait déterministe. Bien que la théorie des variables cachées non-locales de Bohm, qui avait d'abord été élaborée par de Broglie dans les années 1920, ne soit défendue que par une minorité de physiciens, la question de l'interprétation de la théorie quantique est encore aujourd'hui au centre d'une controverse.⁸¹

Une interprétation indéterministe de la mécanique quantique implique l'impossibilité, en pratique et en principe, de connaître l'état complet d'un système microscopique (quantique), à tout instant, à partir de ses conditions initiales et avec les lois régissant son

⁷⁸ Nous renvoyons à Hilgevoord & Uffink, 2008, « The Uncertainty Principle ».

⁷⁹ Voir, par exemple, Heisenberg 1958[1971], *Physique et philosophie*.

⁸⁰ Bohm, 1952, « A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables ». Voir Ortoli & Pharabod, 1998, *Le cantique des quantiques*.

⁸¹ Pour une introduction aux deux interprétations majeures, l'une indéterministe et l'autre déterministe, de la théorie quantique, nous renvoyons respectivement à Faye, 2008, « Copenhagen Interpretation of Quantum Mechanics » et à Goldstein, 2008, « Bohmian Mechanics ».

évolution.⁸² La doctrine du déterminisme scientifique, plus précisément son volet épistémologique, perd donc son bien-fondé dans ce cadre. Si l'on considère la théorie quantique comme une représentation vraie du monde microscopique, le volet métaphysique du déterminisme semble aussi être mis en cause dans le cadre d'une telle interprétation. Les relations d'indétermination entre les propriétés des quantités caractérisant un système quantique impliquent que, à partir de l'état présent du système – qui d'ailleurs ne peut pas être défini de manière complète et infiniment précise car il est fondamentalement indéterminé –, son état à tout moment passé ou futur n'est pas déterminé de manière univoque et nécessaire. Au contraire, pour un moment donné dans le futur, plusieurs états du système sont possibles avec une certaine probabilité.⁸³ Quant au déterminisme méthodologique, il est compatible avec une interprétation indéterministe de la mécanique quantique, mais semble être dénué de sens dans ce cadre.

L'interprétation indéterministe et réaliste de la mécanique quantique est au fondement de la notion de hasard objectif en tant que hasard pur selon laquelle, au moins au niveau des phénomènes physiques microscopiques (quantiques), il existe des propriétés ou des entités intrinsèquement probabilistes. Popper a bien illustré cette conception du hasard objectif dans sa réflexion sur les propensions, justement élaborée pour rendre compte de l'indéterminisme propre aux phénomènes quantiques. Bien que Popper n'ait jamais affirmé que la notion de hasard fait référence aux propensions, que nous allons définir ci-dessous,⁸⁴ la manière dont il les caractérise du point de vue métaphysique et ontologique et le rôle qu'il leur attribue dans

⁸² Quant aux implications d'une interprétation déterministe et réaliste de cette théorie, elles pourraient bien être les suivantes : d'une part, une vision déterministe, au sens métaphysique, du monde microscopique ; d'autre part, une conception du hasard subjectif en tant qu'ignorance des causes, afin de rendre compte des relations d'incertitude.

⁸³ Cependant, il faut rappeler que, selon la théorie de Schrödinger, l'évolution dans le temps d'un système mécanique quantique, par exemple un électron, est régie par une équation linéaire, continue et déterministe (l'équation de Schrödinger). Ce n'est que lorsqu'on interagit avec un système quantique (par exemple, lorsqu'on l'observe pour réaliser une mesure) que son évolution perd son caractère déterministe : ce genre d'interactions ne peut être décrite que par des lois statistiques. C'est ce qu'on appelle le « problème de la mesure ».

⁸⁴ En revanche, dans *Le réalisme et la science* (1990[1982]) Popper affirme que sa théorie propensionniste des probabilités (que nous évoquons dans le chapitre 2 de ce travail) permet de résoudre le « problème fondamental de la théorie du hasard », sans pour autant indiquer quel est la signification du terme « hasard ». « Ce problème troublant peut se formuler ainsi : Comment pouvons-nous expliquer que toute suite enregistrée de jets d'une pièce, ou de lancers d'un dé présentent d'un côté un caractère typiquement aléatoire, et de l'autre une fréquence relative stable tendant apparemment vers une limite ? » (p. 410). Selon Popper, ces deux caractéristiques empiriques des suites de répétitions d'une expérience aléatoire découlent directement des propensions : d'une part, le caractère dispositionnel des propensions est à l'origine du caractère irrégulier des événements qu'elles tendent à réaliser et qui peuvent ou pas être effectivement actualisés ; d'autre part, le fait que le même ensemble de propensions est à l'œuvre au début de chaque répétition d'une expérience aléatoire produit la tendance à la convergence vers une fréquence stable. Popper conclut que « Presque toutes les suites infinies d'événements indépendants se produisant dans des conditions constantes tendront à présenter des fréquences relatives ayant des limites égales aux probabilités (ou propensions) des événements singuliers correspondants » (p. 411). Dans la mesure où la théorie propensionniste des probabilités résout le « problème fondamental de la théorie du hasard », elle constitue une théorie des phénomènes aléatoires.

un monde fondamentalement indéterministe, nous suggère qu'elles représentent une signification potentielle de la notion de hasard pur.

Dans toute la production philosophique de Popper depuis *Le réalisme et la science* (1990[1982]) jusqu'au recueil de conférences qu'est *Un univers de propensions* (1992[1990]), les propensions sont caractérisées de plusieurs manières difficilement compatibles entre elles. Popper les présente tantôt comme des propriétés dispositionnelles des conditions d'engendrement des événements et les identifie alors aux probabilités⁸⁵ ; tantôt comme des entités métaphysiques, c'est-à-dire (dans ce contexte) inobservables mais douées de réalité physique, comparables aux forces newtoniennes⁸⁶ ou résultant de la combinaison de ces dernières⁸⁷ ; tantôt comme des « possibilités pondérées » qui ont une réalité physique⁸⁸ ou comme des mesures de ces possibilités, donc comme des probabilités⁸⁹. Bien qu'il soit difficile de comprendre de manière univoque une telle variété de définitions, les propensions sont toujours caractérisées par les deux aspects suivants⁹⁰. Premièrement, elles sont quelque chose – une propriété, une entité, une force, une possibilité – d'objectif, c'est-à-dire qu'elles sont propres à la structure du monde, en particulier du monde quantique, indépendamment de toute connaissance humaine. Deuxièmement, elles sont quelque chose d'intrinsèquement probabiliste : il s'agit, soit de propriétés dispositionnelles correspondant aux probabilités des événements relativement aux conditions physiques de leur engendrement ; soit d'entités physiques dispositionnelles proches des forces newtoniennes ; soit de possibilités pondérées ou de mesures de possibilités, qui ont une réalité potentielle, donc probabiliste, et non pas actuelle. C'est pourquoi nous concluons que les propensions, selon la caractérisation que Popper leur donne, représentent un très bon exemple de la notion de hasard pur, étant donné que cette dernière renvoie à une propriété ou une entité métaphysique intrinsèquement probabiliste existant dans le monde.

⁸⁵ Popper, 1992[1990], *Un univers de propensions. Deux études sur la causalité et l'évolution*.

⁸⁶ Popper, 1990[1982], *Le réalisme et la science*.

⁸⁷ Popper, 1992[1990], *op. cit.*

⁸⁸ *Ibid.*

⁸⁹ Popper, 1990[1982], *op. cit.*

⁹⁰ Drouet, 2004, « La théorie propensionniste des probabilités ».

1.6. Notre agnosticisme métaphysique

Notre choix d'évacuer les questions de l'interprétation déterministe ou indéterministe de la théorie quantique et de la nature fondamentalement déterministe ou indéterministe des phénomènes se produisant au niveau quantique est à l'origine du fait que, dans le cadre de cette recherche sur la notion de hasard dans la caractérisation des sources causales de la variation biologique, nous n'évoquerons pas la notion de hasard pur que nous venons de présenter. Nous justifions ce choix par les raisons suivantes. D'abord, en ce qui concerne plus strictement le niveau des théories scientifiques, nous avons souligné que le débat au sujet de l'interprétation déterministe ou indéterministe de la mécanique quantique est ardu et loin de sa résolution. Par conséquent, bien que certains biologistes et philosophes de la biologie aient proposé des arguments basés sur une interprétation indéterministe de la mécanique quantique afin de rendre compte du caractère probabiliste de la théorie de l'évolution⁹¹, nous estimons que, dans le cadre de notre recherche, il est légitime d'évacuer la question de l'interprétation de la mécanique quantique (et donc la notion de hasard émergeant du cadre d'une interprétation indéterministe). À notre avis, ce problème doit plutôt trouver une solution dans les pratiques théoriques et expérimentales des physiciens.

En deuxième lieu, en ce qui concerne plus strictement le niveau du monde réel, nous justifions l'évacuation d'une réflexion sur la nature déterministe ou indéterministe des phénomènes quantiques pouvant influencer les sources causales de la variation biologique par le choix philosophique de ne pas essayer de répondre à la question métaphysique sur la nature fondamentalement déterministe ou indéterministe des phénomènes dans le monde physique et biologique. En effet, ce problème nous paraît insoluble du point de vue scientifique et sa solution d'un point de vue purement philosophique ne représente à notre avis que de la spéculation. D'abord, vu que les théories scientifiques sont notre seule voie d'accès à la réalité et qu'elles ne peuvent décrire les phénomènes réels que de manière partielle et simplifiée, la question de savoir si le monde ou un système particulier est fondamentalement déterministe ou indéterministe paraît indécidable. Ensuite, un même phénomène est susceptible d'être décrit à la fois par un modèle déterministe et par un modèle stochastique ou probabiliste, ce qui rend encore plus difficile la possibilité de discerner ses propriétés de manière objective et d'évaluer s'il s'agit d'un phénomène fondamentalement déterministe ou

⁹¹ En particulier, nous renvoyons à l'article de Stamos, 2001, « Quantum Indeterminism and Evolutionary Biology », sur la possible origine quantique des changements tautomériques des bases nucléotidiques de l'ADN, et donc des mutations génétiques.

indéterministe. Par conséquent, nous estimons légitime de diriger notre enquête vers la caractérisation d'une notion de hasard qui soit indépendante de tout choix quant à la nature métaphysique des phénomènes quantiques, des phénomènes biologiques et de tout phénomène dans le monde, c'est-à-dire nous en tenant à notre agnosticisme métaphysique en ce qui concerne la nature déterministe ou indéterministe du monde.⁹²

En troisième lieu, nous proposons une dernière raison justifiant l'évacuation de l'analyse de la notion de hasard pur, qui émerge d'une interprétation indéterministe de la théorie quantique⁹³. Nous prenons position contre le réductionnisme explicatif – et, d'une certaine manière, contre le réductionnisme théorique – et nous suspendons notre jugement quant au réductionnisme métaphysique (ou constitutif) qui nous semble, aussi bien que la question de la nature fondamentalement déterministe ou indéterministe du monde, une question indécidable.⁹⁴ Plus spécifiquement, nous soutenons qu'il n'est pas nécessaire de ramener les phénomènes que l'on étudie – dans le cadre de cette recherche, les sources causales de la variation biologique – aux phénomènes physiques du niveau quantique afin d'en donner une description et une explication adéquates et complètes. En particulier, nous estimons qu'il n'est pas nécessaire de ramener la notion de hasard utilisée en biologie à la notion de hasard pur émergeant de la mécanique quantique afin de rendre compte de manière correcte et complète de ce que cette notion désigne lorsqu'elle est utilisée pour caractériser les sources causales de la variation biologique (entre autres, les mutations génétiques et le bruit dans l'expression des gènes).

⁹² Suppes (1993, « The Transcendental Character of Determinism ») soutient une position proche de la nôtre. Il affirme que, pour un grand nombre de phénomènes empiriques, il est impossible de décider de manière scientifique si la théorie 'ultime' qui les décrits doit être déterministe ou indéterministe. Il conclut en affirmant que la question du caractère fondamentalement déterministe ou indéterministe de l'univers dans sa totalité transcende les faits et les théories scientifiques disponibles, dans un seul mot, l'expérience.

⁹³ Cependant, nous soulignons que la notion de hasard pur pourrait aussi bien émerger, indépendamment de la théorie quantique, d'une interprétation indéterministe d'une théorie biologique donnée, comprise comme une description vraie des dynamiques du monde naturel.

⁹⁴ En suivant la distinction faite par Mayr (1988, *Towards A New Philosophy of Biology*) et, plus récemment, par Sarkar (2007, *Molecular Models of Life: Philosophical Papers On Molecular Biology*), nous suggérons de définir les trois formes de réductionnisme évoquées ci-dessus de la manière suivante. Le réductionnisme explicatif est la thèse selon laquelle le comportement et les propriétés d'un système (physique, chimique ou biologique) peuvent être expliqués en termes de propriétés des parties du système et de leurs interactions (Wimsatt, 2000, « Emergence as Non-Aggregativity and the Biases of Reductionisms »). Selon une autre version plus radicale et formulée relativement à la biologie par Rosenberg (2006, *Darwinian Reductionism, or How to Stop Worrying and Love Molecular Biology*), le réductionnisme explicatif est la thèse selon laquelle tous les phénomènes et les propriétés biologiques peuvent, en principe, être analysés et expliqués de manière complète en termes moléculaires. La thèse du réductionnisme théorique porte sur les relations entre théories scientifiques, plus précisément sur la possibilité de réduire une théorie (par exemple, la biologie moléculaire) à une autre (la physique ou la chimie) et à partir de laquelle on peut dériver la première (Nagel, 1961, *The Structure of Science. Problems in the Logic of Scientific Explanation*). Le réductionnisme métaphysique (ou constitutif) affirme que la composition matérielle de toute entité, par exemple de toute entité biologique, ne diffère en rien de ce que décrivent la chimie et la physique.

Dans le prochain chapitre, nous commençons par introduire d'autres notions de hasard qui ont émergé du cadre de la théorie des probabilités, de la théorie algorithmique de l'information, et qui ont été définies en termes mathématiques. Le choix de les introduire dans un chapitre à part réside dans le fait que ces notions, contrairement à celles que nous venons de présenter, sont toujours évoquées en faisant référence à leurs définitions formelles et non de manière métaphorique ; de plus, elles sont toutes relativement neutres quant à la question de la nature subjective ou objective du hasard. Nous analyserons ces notions mathématiques de hasard en focalisant notre attention sur leur relation avec les doctrines du déterminisme et de l'indéterminisme, sous leurs formes diverses. Nous montrerons que la plupart d'entre elles ne sont pas pertinentes pour répondre à la question qui se trouve au centre de notre enquête, c'est-à-dire celle de savoir quelle est la notion de hasard appropriée pour caractériser les sources causales de la variation biologique. Ensuite, nous procéderons à une introduction aux théories philosophiques des probabilités, ces dernières étant au cœur de certaines notions mathématiques de hasard. En particulier, nous mettrons en évidence la différence entre l'interprétation épistémique et objective des probabilités. Sur cette base, nous avancerons une réponse possible à la question de la relation entre la notion de hasard et la notion de probabilité.

Chapitre 2 – La mathématisation du hasard

La notion de hasard n'a reçu de définition formelle que très tard par rapport à ses premières évocations au sein des réflexions de nombreux philosophes de la Grèce antique. La date à laquelle on fixe habituellement l'entrée de cette notion en mathématiques correspond à la première approche formelle des jeux de hasard, dont on considère généralement qu'il s'agit de la naissance du calcul des probabilités. C'est en 1654 que, afin de résoudre un problème de jeu de hasard soulevé par le cavalier de Méré, le scientifique et philosophe Pascal s'engage dans une correspondance avec le mathématicien Fermat et pose les bases de la future théorie des probabilités. Mais ce n'est qu'en 1933 que Kolmogorov propose une axiomatisation du calcul des probabilités qui le promeut au rang des autres théories mathématiques.¹

Dans ce chapitre consacré à la mathématisation du hasard, nous allons introduire quatre définitions formelles de cette notion. D'abord, nous verrons que, du cadre de la théorie des probabilités, ont émergé la notion de hasard au sens d'« équiprobabilité » des événements possibles, qui est au fondement même du calcul des probabilités, et la notion de « stabilité des fréquences à la limite », qui a été élaborée par Von Mises et améliorée par Church au cours des années 1930-1940. Ensuite, nous analyserons une troisième notion de hasard mathématiquement définie : c'est le hasard en tant qu'« incompressibilité » d'une suite binaire de 0 et 1, qui a trouvé son origine au sein de la théorie algorithmique de l'information, plus précisément dans les réflexions de Solomonov, de Kolmogorov et de Chaitin dans les années 1960. Dix ans plus tard, cette même notion a été développée par Chaitin, par Levin et par Schnorr et s'est révélée à peu près équivalente à la notion de hasard en tant que « satisfaction (par une suite) de tous les tests statistiques effectifs » que Martin-Löf avait développée en 1966. Cette dernière est aujourd'hui considérée de manière consensuelle comme la notion fournissant la bonne définition mathématique de ce qu'est le hasard.

La décision de consacrer un chapitre à part à ces quatre notions de hasard réside, d'une part, dans le fait qu'elles ont toutes émergé d'un cadre théorique purement mathématique (la théorie des probabilités et la théorie algorithmique de l'information). Au contraire, les notions

¹ Dans le cadre de ce deuxième chapitre, nous ne comptons pas tracer l'histoire de la théorie des probabilités. Nous renvoyons à Hacking, 2002[1975], *L'Émergence de la probabilité* et Daston, 1998, *Classical Probability in the Enlightenment*. Pour une introduction technique au calcul des probabilités, voir Feller, 1950, *An Introduction to Probability Theory and Its Applications*.

de hasard subjectif et de hasard objectif que nous avons analysées dans le chapitre précédent ont plutôt trouvé leur origine dans le contexte de réflexions philosophiques plus ou moins anciennes au sujet du hasard et/ou au sein de théories physiques diverses (par exemple, la physique des systèmes chaotiques pour la notion de hasard en tant que sensibilité aux conditions initiales et la mécanique quantique pour la notion de hasard pur). Elles peuvent aussi être définies formellement, en termes mathématiques, de même que les notions de hasard qui font l'objet de ce chapitre ; cependant, de manière générale et plus particulièrement au sein des disciplines biologiques, elles sont souvent évoquées pour caractériser le hasard de façon métaphorique. D'autre part, nous avons décidé de présenter dans un même chapitre les notions mathématiques de hasard car, bien qu'elles aient émergé à des moments différents et sous des formes distinctes, elles partagent un certain nombre de caractéristiques communes. En particulier, nous remarquerons que ces notions diverses ne sont pas fortement connotées du point de vue métaphysique : elles n'impliquent aucune prise de position par rapport aux doctrines du déterminisme et de l'indéterminisme métaphysique ; de plus, la plupart d'entre elles oscillent entre une notion de hasard subjectif et de hasard objectif.

Dans ce deuxième chapitre, nous présentons quatre notions mathématiques du hasard : en tant qu'« équiprobabilité », en tant que « stabilité des fréquences à la limite », en tant qu'« incompressibilité » d'une suite, et en tant que « satisfactions (par une suite) de tous les tests statistiques effectifs ». Pour chacune de ces notions de hasard, nous essaierons de répondre à la question de savoir si elles désignent un hasard subjectif ou objectif, c'est-à-dire si elles reflètent notre connaissance des phénomènes réels ou une propriété inhérente à ces derniers. Sur cette base, nous examinerons la relation de compatibilité ou d'incompatibilité entre chaque notion mathématique de hasard et les doctrines diverses du déterminisme et de l'indéterminisme.

Nous ne manquerons pas d'évoquer, bien que de manière succincte, l'utilisation de certaines de ces notions mathématiques de hasard en biologie. Nous le ferons pour les notions de hasard en tant qu'équiprobabilité, sur laquelle se fonde la théorie des probabilités, et en tant qu'incompressibilité d'une suite, qui a émergé de la théorie algorithmique de l'information. Notre but sera de mettre en évidence que la plupart de ces notions ne sont pas appropriées pour notre propos. En effet, le fait qu'elles sont presque toutes définies en termes de résultat et qu'elles ne peuvent donc être utilisées que pour rendre compte du caractère « aléatoire » d'un événement donné, et non pas du processus qui l'engendre, en réduit la portée descriptive et explicative dans le cas qui nous intéresse, c'est-à-dire dans la caractérisation des sources causales de la variation biologique.

En conclusion de ce chapitre, nous focaliserons notre attention sur la notion de probabilité et, plus précisément, sur les diverses théories philosophiques qui ont été formulées au cours du siècle dernier afin d'interpréter le calcul des probabilités. En effet, notre analyse aura mis en évidence la place centrale que ces dernières occupent au sein des définitions de certaines des notions mathématiques de hasard. Notre but sera donc de souligner que le hasard et les probabilités entretiennent une relation privilégiée et d'en suggérer une caractérisation possible. C'est en vertu de cette relation que nous suggérerons de répondre à la question de la signification et du rôle de la notion de hasard, qui est utilisée pour caractériser un phénomène donné, à la lumière de l'interprétation choisie pour les probabilités mobilisées pour décrire ce dernier.

Bien que ce deuxième chapitre porte sur des notions de hasard définies en termes purement mathématiques, nous procéderons à leur présentation sans faire appel à aucune équation ou formulation technique, ce qui caractérise de manière générale le style que nous avons décidé d'adopter dans notre travail de recherche.

2.1. La notion de hasard en tant qu'« équiprobabilité »

La notion de hasard est habituellement invoquée lorsqu'on est confronté à un événement qui est imprédictible, du moins de manière univoque et précise. En dépit de son imprédictibilité, on peut parfois évaluer *a priori* la probabilité d'apparition d'un événement par rapport à d'autres événements possibles et qui sont mutuellement exclusifs. Cela peut être réalisé, soit sur la base des symétries caractérisant notre ignorance quant à tous ces événements possibles, c'est-à-dire en invoquant le principe de raison insuffisante (ou principe d'indifférence)² ; soit à partir de la connaissance des symétries du dispositif physique pouvant produire ces événements, c'est-à-dire en utilisant « la méthode des fonctions arbitraires »³.

De manière plus précise, d'une part, selon le principe de raison insuffisante (ou principe d'indifférence), en l'absence d'aucune raison pour attribuer des probabilités

² Bernoulli, 1713, *Ars Conjectandi* ; Laplace, 1986[1814], *Essai philosophique sur les probabilités* ; Keynes, 1963[1921], *Treatise on Probability*.

³ Poincaré, 1987[1912], *Calcul des probabilités*. Voir aussi, von Kries, 1886, *Die Principien der Wahrscheinlichkeitsrechnung. Eine logische Untersuchung* ; Borel, 1924[1909], *Éléments de la théorie des probabilités* ; 1914, *Le hasard* ; Fréchet, 1938, *Méthode des fonctions arbitraires, théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles*. La méthode des fonctions arbitraires a été récemment ré-évaluée par von Plato (1983, « The Method of Arbitrary Functions ») et par Strevens (1998, « Inferring Probabilities From Symmetries »).

différentes à des événements mutuellement exclusifs – autrement dit, sur la base d’une incertitude égale à leur égard –, on doit considérer ces événements comme équipossibles et leur attribuer la même probabilité. Il s’agit donc d’un principe méthodologique et épistémique car fondé sur le manque d’informations d’un sujet connaissant, ce qui a soulevé des objections importantes quant à sa légitimité dans le cadre du calcul des probabilités⁴. D’autre part, il est possible d’identifier des événements équipossibles et de leur attribuer une même valeur de probabilité en vertu des symétries propres au dispositif physique pouvant les engendrer et sur la base de certaines hypothèses mathématiques quant au caractère continu et régulier (c’est-à-dire, constant sur de petits intervalles) de la fonction de probabilité. Par exemple, les symétries physiques d’une roulette, notamment sa symétrie circulaire et l’alternance rapide de petites bandes rouges et noires de taille identique, permettent d’inférer que sur l’ensemble des vitesses avec lesquelles on peut la faire tourner, la probabilité de l’événement « rouge » est la même que celle de l’événement « noir ». Cette valeur est égale à $\frac{1}{2}$.⁵ De manière générale, indépendamment de la méthode utilisée, une fois qu’on a identifié l’ensemble des cas équipossibles et qu’on leur a attribué la même probabilité, on peut associer une valeur de probabilité à l’événement auquel on s’intéresse. C’est la notion classique de probabilité, supposant l’équiprobabilité dans sa définition, selon laquelle la probabilité d’un événement est le rapport du nombre des cas favorables (c’est-à-dire, des cas qui en sont des instances) au nombre total des cas équipossibles, donc équiprobables.⁶

Afin d’explicitier le rôle de la notion d’équiprobabilité dans le cadre du calcul *a priori* des probabilités, considérons un jeu de hasard, par exemple le jeu de dés. Tous les résultats possibles d’un lancer de dé ($\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$), qu’on appelle « événements élémentaires »⁷, sont mutuellement exclusifs : dans le cas d’un dé non pipé, ils peuvent être considérés comme des éventualités équipossibles donc équiprobables, soit sur la base du principe d’indifférence, soit en utilisant la méthode des fonctions arbitraires. L’ensemble de tous les événements élémentaires (équi)possibles est ce qu’on appelle « l’espace fondamental »⁸ ($\{1, 2, 3, 4, 5, 6\}$). Imaginons que nous voulions calculer la probabilité qu’un nombre pair apparaisse lors d’un lancer du dé. Les résultats du lancer (ou événements élémentaires) caractérisés par cette propriété, et donc appartenant à l’événement « un nombre pair apparaît lors d’un lancer du dé », sont au nombre de trois ($\{2\}$, $\{4\}$, $\{6\}$) : on appelle

⁴ Voir Bertrand, 1889, *Calcul des probabilités*.

⁵ Pour plus de détails, nous renvoyons à Strevens, 1998, *op. cit.*

⁶ Laplace 1878[1812], *Théorie analytique des probabilités*.

⁷ Voir le Glossaire.

⁸ *Id.*

« événement »⁹ l'ensemble de tous les événements élémentaires partageant une propriété commune (dans ce cas, « être un nombre pair » : $\{2, 4, 6\}$). Sur la base de l'équiprobabilité des événements élémentaires (équi)possibles, on peut conclure que la probabilité *a priori* de l'événement « un nombre pair apparaît lors d'un lancer du dé » est égale au rapport entre le nombre d'événements élémentaires appartenant à l'événement ($\{2\}, \{4\}, \{6\}$) et le nombre de tous les événements élémentaires possibles ($\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$). On pourrait aussi vouloir estimer la probabilité qu'au moins un nombre pair apparaisse après un certain nombre n de lancers successifs du même dé. Dans ce cas, il ne faudrait pas simplement établir l'espace des événements élémentaires équipossibles, donc équiprobables ; il serait aussi nécessaire d'établir que chaque lancer (ou chaque répétition de la même expérience aléatoire ou épreuve) est indépendant(e) des précédent(e)s. De cette manière, la règle de la multiplication des probabilités pourrait être appliquée¹⁰.

Imaginons maintenant qu'on veuille tester empiriquement (*a posteriori*) si la probabilité qu'on a attribuée (*a priori*) à l'événement « un nombre pair apparaît lors d'un lancer du dé » est correcte. À cette fin, on lance plusieurs fois et de manière indépendante le même dé ; après un nombre assez élevé de lancers (qui, idéalement, tend vers l'infini), on estime la fréquence relative de l'événement considéré. Si, sur le long terme, elle était beaucoup plus élevée que la fréquence attendue à la lumière de la probabilité calculée *a priori*, sur la base de l'hypothèse d'équiprobabilité des événements élémentaires (par exemple, si elle était égale à $3/4$ à la place de $1/2$), on serait amené à penser que le dé utilisé n'est pas bien équilibré, mais biaisé. On tendrait donc à conclure que l'événement dont on veut estimer la probabilité n'est pas le fruit du hasard ; au contraire, on affirmerait que certaines caractéristiques physiques du dé utilisé ont favorisé l'apparition des nombres pairs (c'est-à-dire, qu'elles privilégient l'événement $\{2, 4, 6\}$).¹¹

C'est du cadre de ce genre de raisonnements et, plus généralement, des principes et des méthodes fondant le calcul *a priori* des probabilités, qu'émerge la notion mathématique de hasard en tant qu'équiprobabilité. Si des résultats sont équipossibles, donc équiprobables, cela signifie qu'aucun d'entre eux n'est favorisé par une ou plusieurs causes. Sur cette base, on peut affirmer que le résultat particulier ayant lieu est le fruit du hasard. De la même

⁹ Voir le Glossaire.

¹⁰ Selon la règle de la multiplication des probabilités, si les événements A et B sont indépendants, alors $\Pr(A \& B) = \Pr(A) \times \Pr(B)$.

¹¹ La fréquence relative des résultats obtenus peut aussi être estimée en se plaçant, non pas au niveau des événements mutuellement exclusifs $\{2, 4, 6\}$ et $\{1, 3, 5\}$, mais au niveau des événements élémentaires possibles ($\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$) qui sont considérés *a priori* comme équiprobables. Dans ce cas, on n'invoquera le hasard que si la fréquence de ces événements élémentaires est distribuée uniformément, ce qui signifie qu'aucun événement élémentaire n'a été favorisé par rapport aux autres.

manière, si l'on calcule les probabilités d'un ensemble d'événements *a posteriori*, c'est-à-dire sur la base des fréquences observées, on affirmera que ces événements sont le fruit du hasard si, sur le long terme, ils ont eu lieu approximativement avec la même fréquence : aucun biais n'a favorisé un événement par rapport aux autres mais tous avaient la même probabilité d'avoir lieu.¹²

L'hypothèse d'équiprobabilité a été introduite pour calculer la probabilité d'événements mutuellement exclusifs et indépendants ; cependant, elle peut être généralisée et appliquée à un ensemble d'événements qui ne sont pas mutuellement exclusifs tout en gardant leur indépendance. Considérons, par exemple, la distribution d'une maladie comme le cancer des poumons au sein d'une population donnée. Si l'on observait que cette maladie n'est pas distribuée uniformément au sein d'une population d'individus habitant la même ville, mais atteint plus fréquemment les personnes de cette population qui fument, on n'affirmerait pas que le cancer des poumons est un événement se produisant au hasard. On avancerait plutôt l'hypothèse selon laquelle fumer favorise le cancer des poumons chez certaines personnes, à savoir les fumeurs.¹³ Au contraire, si cette maladie était distribuée uniformément sur la population considérée, indépendamment du fait que les individus la constituant fument ou ne fument pas, on affirmerait que la probabilité qu'elle atteigne une personne est la même pour tout individu, c'est-à-dire qu'elle est invariante au sein de la population.¹⁴ Sur cette base, on conclurait que le fait que l'individu *x* ait un cancer des poumons plutôt qu'un autre individu *y* l'ait est le fruit du hasard.

Dans la prochaine sous-section, nous allons définir cette même notion de hasard en tant qu'équiprobabilité mais en des termes différents, relativement au processus pouvant engendrer un ensemble d'événements équipossibles et donc équiprobables.

¹² Hacking, 2004[2001], *L'Ouverture au probable*, évoque implicitement la notion de hasard en tant qu'équiprobabilité lorsqu'il fournit la définition de l'absence de tout biais pouvant influencer un générateur aléatoire (par exemple, une machine générant des suites de 0 et 1). Il écrit : « Un générateur aléatoire n'est pas biaisé si et seulement si tous les événements ont la même fréquence relative dans le long terme », c'est-à-dire si chaque événement se produit aussi souvent que les autres. Alors, le générateur utilisé n'est pas biaisé et tous les événements sont donc équiprobables.

¹³ On arriverait à la même conclusion si l'on observait que le cancer des poumons atteint plus fréquemment les personnes résidant, par exemple, à Paris par rapport aux autres villes de France. Dans ce cas, on avancerait l'hypothèse selon laquelle le taux élevé de pollution à Paris favorise le cancer chez ses habitants.

¹⁴ De manière analogue, si la fréquence relative de cette maladie était distribuée uniformément sur l'ensemble des villes françaises, on conclurait que la probabilité qu'elle atteigne une personne est la même dans toute ville, c'est-à-dire qu'elle est invariante au sein du territoire urbain français.

2.1.1. L'équiprobabilité en termes de processus : la notion d'« échantillonnage indiscriminé »

Nous proposons de considérer un autre exemple qui nous permettra d'identifier de manière plus précise le sens de cette notion. Notre but est de fournir une définition de la notion mathématique de hasard en tant qu'équiprobabilité qui ne soit pas simplement une description (*a priori* ou *a posteriori*) des caractéristiques des cas possibles d'une expérience aléatoire – notamment une description des symétries caractérisant, soit notre ignorance au sujet du dispositif considéré, soit les propriétés physiques de ce dernier, soit les fréquences relatives observées. Nous visons à identifier une définition de l'équiprobabilité d'un ensemble de cas (équi)possibles qui renvoie au processus causal pouvant les engendrer : une telle notion ne serait pas simplement descriptive mais explicative car elle expliciterait pourquoi certaines symétries (épistémiques ou physiques) concernant l'expérience aléatoire considérée impliquent l'équiprobabilité (épistémique ou objective) de ses résultats possibles.

Imaginons que nous demandions à un aveugle de piocher dans une urne remplie d'une dizaine de boules de couleurs différentes mais de même volume et de même texture, donc indiscernables au toucher. Étant aveugle, cette personne pioche à chaque fois une boule indépendamment de ses propriétés physiques visibles, notamment la couleur, et donc sans être influencée par les différences de couleur entre les boules présentes dans l'urne. La notion mathématique de hasard en tant qu'équiprobabilité peut être invoquée pour décrire cette situation, plus précisément, afin d'exprimer le fait que, à chaque répétition, l'opération d'échantillonnage est indifférente à la couleur et, par conséquent, ne favorise ou discrimine aucune boule : elles ont toutes la même probabilité d'être tirées. En effet, bien que les boules diffèrent en ce qui concerne la couleur de leur surface, leur indiscernabilité ou invariance pour l'auteur de l'échantillonnage (une personne aveugle) explique pourquoi aucune n'est favorisée et donc chacune a la même probabilité d'être échantillonnée. Autrement dit, vu que, dans cette situation, la couleur (comme d'ailleurs toute autre propriété visible des dix boules) est indiscernable pour l'auteur de l'échantillonnage, les différences dans la couleur des boules n'ont aucun rôle causal dans le processus d'échantillonnage et dans la détermination du résultat de ce dernier.

Nous proposons de parler de « processus d'échantillonnage indiscriminé » pour désigner ce genre de situations dans lesquelles aucune différence dans les caractéristiques des éléments pouvant être échantillonnés ne joue de rôle causal dans le processus d'échantillonnage et donc dans la détermination de la fréquence relative de chaque élément au

sein de l'échantillon obtenu. Ce dernier est produit par des tirages répétés où l'élément pioché est systématiquement remplacé ; par exemple, dans une série de tirages de boules dans une urne, cela signifie que la boule tirée est remise systématiquement dans l'urne.¹⁵

Nous proposons la notion d'échantillonnage indiscriminé comme une version en termes de processus de la notion de hasard en tant qu'équiprobabilité : elle assume un rôle explicatif par rapport à cette dernière qui, au contraire, semble s'arrêter à la description des symétries concernant les cas possibles d'une expérience aléatoire donnée. Plus précisément, la notion d'échantillonnage indiscriminé explique en termes causaux pourquoi les résultats possibles d'un processus donné ont tous la même probabilité (épistémique ou objective) de se produire ou pourquoi, dans le long terme, on s'attend à observer une distribution uniforme des fréquences relatives sur les résultats possibles.¹⁶

Dans la prochaine sous-section, nous allons examiner la relation entre la notion mathématique de hasard en tant qu'équiprobabilité et les doctrines du déterminisme et de l'indéterminisme.

2.1.2. Déterminisme, indéterminisme et équiprobabilité

Quelles sont les implications de la notion mathématique de hasard en tant qu'équiprobabilité par rapport aux doctrines du déterminisme et de l'indéterminisme ? Au début du XIX^e siècle, la notion d'équiprobabilité fut introduite de manière systématique dans le cadre de la théorie classique (*a priori*) de la probabilité par Laplace. Bien que ce dernier ait conçu les probabilités de manière épistémique en tant que mesures de notre ignorance partielle des causes sous-jacentes des phénomènes¹⁷, la notion d'équiprobabilité est neutre par rapport à la distinction entre hasard subjectif et hasard objectif. De plus, sa signification n'a aucune implication directe quant au caractère déterministe ou indéterministe de la structure du monde, quant à la prévisibilité en principe de l'évolution de ce dernier ou des modèles utilisés pour le décrire, ou quant à l'approche méthodologique qu'on doit utiliser en science.

¹⁵ Au contraire, dans le cadre d'un « processus d'échantillonnage discriminé » avec remplacement, les différences dans les caractéristiques des éléments pouvant être échantillonnés jouent un rôle causal dans le processus d'échantillonnage et donc dans la détermination de la fréquence relative de chaque élément au sein de l'échantillon obtenu sur le long terme.

¹⁶ Dans le chapitre 7, nous évoquerons la notion d'échantillonnage indiscriminé au sein des définitions de ce que nous appellerons « aléatoire (au sens) fort » et « aléatoire (au sens) faible ».

¹⁷ Voir la section 2.6.1 de ce chapitre.

Par ailleurs, cette notion mathématique de hasard joue aussi un rôle central dans le cadre de la méthode des fonctions arbitraires qui, nous le rappelons, fut conçue au tournant du XX^e siècle : elle consiste à inférer l'équiprobabilité d'un ensemble de résultats possibles et mutuellement exclusifs à partir des symétries du dispositif physique pouvant les engendrer. Contrairement à l'approche classique laplacienne et en dépit de quelques aspects communs avec le principe d'indifférence¹⁸, la méthode des fonctions arbitraires semble suggérer que l'équiprobabilité des résultats possibles est objective car elle est due aux propriétés du dispositif physique considéré. Par conséquent, dans ce cadre, la notion de hasard en tant qu'équiprobabilité semble plutôt renvoyer à un hasard objectif. Cependant, le hasard subjectif en tant que reflet des limites de nos connaissances n'est pas exclu. En effet, le dispositif physique étudié pourrait être caractérisé non seulement par les symétries physiques qu'on a pu identifier, mais aussi bien par d'autres symétries et/ou par des asymétries qu'on ignore et qui modifieraient la distribution de probabilité sur l'ensemble des résultats possibles.

C'est pourquoi nous estimons que la question de savoir si la notion d'équiprobabilité désigne un hasard subjectif ou un hasard objectif dépend directement de la manière dont on interprète les probabilités (égales) d'un ensemble de cas (équi)possibles. Notamment, elles peuvent être interprétées, soit de manière épistémique en tant que mesures de notre ignorance, de notre connaissance ou de notre croyance au sujet du système (physique, biologique, etc.) considéré ; soit de manière objective en tant que degrés de possibilité, en tant que propensions ou en tant qu'idéalisations de fréquences relatives des événements possibles.¹⁹

Nous suggérons que, une fois qu'on s'est prononcé sur l'interprétation (épistémique ou objective) des probabilités, on peut en déduire si la notion d'équiprobabilité renvoie à un hasard subjectif ou objectif. Ensuite, on peut procéder à l'identification des relations de compatibilité ou d'incompatibilité que cette notion entretient avec les différentes doctrines déterministes et indéterministes.

À titre d'illustration, si l'on interprétait objectivement les probabilités des résultats possibles d'une expérience aléatoire (par exemple, en tant que mesures des degrés de possibilité de ces résultats), on ne pourrait pas adhérer à une vision laplacienne du déterminisme de la structure de l'univers dans sa totalité. De plus, on ne serait pas obligé de prendre position en faveur de l'indéterminisme (global ou local). En effet, le déterminisme métaphysique mais local pourrait rendre pleinement compte de l'équiprobabilité objective de

¹⁸ Nous renvoyons à voir Raidl, à paraître, « A method of arbitrary functions: A rehabilitation of classical probability? ».

¹⁹ Pour une introduction aux différentes interprétations, épistémiques et objectives, des probabilités, nous renvoyons à la section 2.6 de ce chapitre.

résultats qui sont dus à la rencontre imprédictible de chaînes causales nécessairement déterminées, du point de vue nomique, mais indépendantes entre elles.

Avant de procéder à la présentation et à l'analyse d'une autre notion mathématique de hasard qui est définie en termes de probabilités, nous précisons que la notion de hasard en tant qu'équiprobabilité est amplement utilisée en science, en particulier dans le cadre de la modélisation théorique de phénomènes divers (physiques, chimiques, biologiques, etc.). D'une part, supposer l'équiprobabilité d'un ensemble d'événements possibles, bien que cela ne soit pas nécessairement le cas dans la réalité, peut se révéler une approximation avantageuse du point de vue méthodologique : en effet, lorsqu'il s'agit d'une simplification légitime, elle peut faciliter la réalisation de prédictions au sujet du phénomène étudié. D'autre part, une telle démarche peut être utile dans le processus de découverte scientifique. En effet, la réfutation d'un modèle faux, que l'on a construit en postulant l'équiprobabilité d'un ensemble d'événements possibles, est une méthode efficace pour identifier les caractéristiques du phénomène réel qu'on tâche de décrire, de prédire et d'expliquer.

2.2. Qu'est-ce qu'une suite aléatoire ?

Autour des années 1930, de nombreux efforts furent dirigés vers la formulation d'une définition rigoureuse, en termes mathématiques, de la notion de hasard ou, pour utiliser un terme plus répandu dans ce cadre technique, d'« aléatoire »²⁰. Plus précisément, ce genre de recherches, qui a perduré tout au long du siècle dernier, veut résoudre le problème de la bonne définition d'un modèle mathématique idéalisé, à savoir une suite infinie aléatoire. Sur cette base, il est possible d'identifier les propriétés constitutives du hasard dans ce cadre et de les repérer ensuite dans les événements physiques réels. Ainsi, une définition rigoureuse de la notion de hasard (ou d'aléatoire) peut permettre de saisir dans un cadre formel les propriétés qu'on lui associe de manière intuitive, par exemple, l'irrégularité, le désordre, le défaut de motif (*pattern*), l'absence de structure, l'absence de loi, l'imprédictibilité. De plus, elle permet de se débarrasser d'idées trompeuses au sujet de cette notion, qui sont principalement dues aux facteurs psychologiques affectant notre perception.²¹

Nous avons déjà remarqué que, de manière générale, la notion de hasard peut être

²⁰ Dans le cadre de ce chapitre, nous utiliserons les deux termes « hasard » et « aléatoire » de manière interchangeable.

²¹ Voir Gauvrit, 2005, « Les irrémédiables structures du hasard humain ».

définie, soit en termes de processus, soit en termes de résultats. De manière semblable et afin de mettre de l'ordre dans la confusion entourant cette notion, Earman²² introduit la distinction entre deux significations de ce qu'il appelle plutôt l'«aléatoire», l'une en termes de processus et l'autre en termes de performance (ou de résultat). D'abord, il définit un processus aléatoire comme un processus «qui n'opère pas de manière complètement capricieuse ou désordonnée mais en accord avec des lois stochastiques ou probabilistes»²³. Ensuite, il ne fournit pas de définition précise de «performance aléatoire», mais annonce qu'il ne s'agit pas d'un concept absolu car il dépend de ce qu'on connaît ou qu'on croit au sujet du processus engendrant un certain résultat. Enfin, il souligne que le caractère aléatoire d'une performance²⁴ ne garantit pas qu'elle a été engendrée par un processus aléatoire et, inversement, que le caractère aléatoire d'un processus n'implique pas qu'il produit une performance aléatoire²⁵. Ce point soulevé par Earman rappelle nos propres réflexions sur la valeur respective d'une notion de hasard caractérisant un résultat (par exemple, la fréquence relative d'un certain attribut dans une suite d'événements uniformes) et d'une notion de hasard caractérisant le processus causal qui est à l'origine de ce résultat. Cette dernière fournit à la fois une description du caractère aléatoire d'un processus et une explication du caractère aléatoire, si c'est bien le cas, de son résultat.

Les notions mathématiques du hasard que nous allons analyser dans les trois prochaines sections (les notions de hasard en tant que stabilité des fréquences à la limite, en tant qu'incompressibilité d'une suite et en tant que satisfaction (par une suite) de tous les tests statistiques effectifs) ont toutes émergé de l'étude d'un modèle très simple de phénomène aléatoire qu'est une suite aléatoire infinie. Il s'agit d'une suite infinie d'événements uniformes (c'est-à-dire qui sont produits par la répétition d'épreuves indépendantes et identiques), ces derniers pouvant différer par des attributs divers, par exemple des numéros (comme dans le cas des suites binaires de 0 et 1), des couleurs ou n'importe quelle autre caractéristique. Durant le siècle dernier, la notion de suite aléatoire infinie, considérée par von Mises comme une condition nécessaire pour qu'on puisse assigner des probabilités à des événements, fut l'objet de plusieurs réflexions au sein de la théorie des probabilités et dans la théorie algorithmique de l'information²⁶ visant toutes à identifier ses propriétés et à en fournir une

²² Earman, 1986, *A Primer on Determinism*.

²³ «[...] which does not operate wholly capriciously or haphazardly but in accord with stochastic or probabilistic laws», *ibid.*, p. 137 (notre traduction).

²⁴ La définition de ce qu'est une «performance aléatoire» est l'objet central du chapitre VIII de ce livre d'Earman.

²⁵ *Ibid.*, p. 138, p. 146.

²⁶ La théorie des probabilités doit sa première axiomatisation à Kolmogorov, 1956[1933], *Foundations of the*

définition purement mathématique.

Nous verrons que, dans le cadre de la plupart des définitions de ce qu'est une suite aléatoire, la notion de hasard utilisée ne décrit pas les caractéristiques du processus pouvant engendrer une telle suite, mais est plutôt définie en référence aux caractéristiques de cette dernière (c'est-à-dire, en référence aux propriétés du résultat ou de la réalisation d'un processus). En revanche, nous remarquerons aussi que la notion de hasard en tant qu'incompressibilité d'une suite – dont nous analyserons la formulation proposée indépendamment et presque simultanément par Solomonov, par Kolmogorov et par Chaitin dans les années 1960 – définit le caractère aléatoire d'une suite, non seulement en décrivant ses caractéristiques, mais aussi en termes de processus, plus précisément par les caractéristiques du programme d'instructions pouvant l'engendrer. Enfin, nous montrerons que toutes ces notions mathématiques de hasard, introduites au cours du siècle dernier afin de trouver une définition purement mathématique de la notion de suite aléatoire, ne sont pas appropriées pour caractériser les phénomènes biologiques au centre de notre recherche, c'est-à-dire les sources causales de la variation biologique. Cela n'est pas seulement dû au fait que la plupart d'entre elles sont définies en termes de résultat, mais surtout au caractère lacunaire des connaissances dont les biologistes disposent au sujet des phénomènes qui sont à l'origine de la variation biologique.

2.3. La notion de hasard en tant que « stabilité des fréquences à la limite »

Von Mises fut un représentant paradigmatique de la théorie fréquentiste des probabilités²⁷. Il était convaincu, contrairement par exemple à d'autres fréquentistes comme Reichenbach, qu'il n'était pas sensé de parler de la probabilité d'un événement singulier. Au contraire, il écrivait :

« En utilisant le langage de la physique, on peut affirmer qu'afin d'appliquer la théorie des probabilités on doit avoir une suite pratiquement illimitée d'observations uniformes »²⁸.

theory of probability. Dans les années 1930, ce même mathématicien fut l'un des initiateurs de la théorie algorithmique de l'information.

²⁷ Pour une introduction à la théorie fréquentiste des probabilités, nous renvoyons à la section 2.6.2 de ce chapitre.

²⁸ « *Using the language of physics, we may say that in order to apply the theory of probability we must have a practically unlimited sequence of uniform observations* », von Mises, 1961[1928], *Probability, Statistics and Truth*, p. 11 (notre traduction).

Une telle suite est ce que von Mises appela « collectif » et qu'il considéra comme le fondement de la théorie des probabilités : il considérait l'existence d'un collectif comme la condition de possibilité de toute attribution de probabilité²⁹.

Qu'est-ce qu'un collectif selon von Mises ? Il s'agit d'une suite infinie aléatoire d'événements uniformes qui diffèrent par des attributs observables pouvant être de nature diverse³⁰. Plus précisément, une suite binaire infinie $x = x_1 x_2 x_3 \dots$ (c'est-à-dire une suite infinie de 0 et 1) est un collectif si elle remplit les deux conditions suivantes : les fréquences relatives des attributs des événements la composant convergent vers une valeur à la limite ; elle est aléatoire. La première propriété faisant d'une suite binaire infinie un collectif n'est rien d'autre que la satisfaction de la « loi des grands nombres »³¹ qui, bien que cela représente une hypothèse forte de la part de von Mises, est considérée comme valable pour la plupart des suites. Considérons en détail la deuxième propriété qu'est le caractère aléatoire de la suite et qui nous intéresse particulièrement ici. Selon von Mises, une suite binaire infinie est aléatoire si les fréquences à la limite des 0 et des 1 la constituant sont préservées au sein des suites extraites par un « procédé de sélection de places admissible »³², par exemple, la sélection de sous-parties de la suite originelle sur la base de la position et indépendamment des attributs (0 et 1) de ses éléments, ou sur la base des attributs des membres qui précèdent les éléments en question.³³

Au cours des années 1930, la définition de la notion de suite infinie aléatoire proposée par von Mises fut l'objet de nombreuses critiques, surtout au sujet du caractère arbitraire et flou de la notion de « sélection admissible », que von Mises n'arriva jamais à définir de manière appropriée.³⁴ En 1940, Church³⁵ proposa d'améliorer la notion de collectif de la manière suivante : il affirma que les opérations admissibles de sélection sont définies par des

²⁹ Von Mises considérait sa vision fréquentiste des probabilités comme une avancée importante par rapport à la théorie classique basée sur la notion d'équiprobabilité des cas possibles (Laplace), l'équiprobabilité étant selon lui une propriété absente en dehors du domaine restreint des jeux de hasard.

³⁰ Von Mises, 1961[1928], *op. cit.*, p. 12.

³¹ La « loi des grands nombres » exprime le fait que la fréquence d'un événement converge en probabilités vers sa probabilité lorsque le nombre d'épreuves indépendantes augmente indéfiniment. Voir le Glossaire.

³² *Ibid.*, p. 24-25.

³³ *Ibid.*, p. 25-26, parlait aussi du « principe de l'impossibilité d'un système de pari » pour désigner la deuxième condition pour qu'une suite binaire infinie soit un collectif (c'est-à-dire, son caractère aléatoire). Dans le cas de tout jeu de hasard, la propriété d'être aléatoire empêche toute prédictibilité de la suite et donc la possibilité de tout pari gagnant contre elle. Cette idée fut ensuite développée, même si ce fut de manière insatisfaisante, par Ville (1939, *Etude critique de la notion de collectif*) et, plus tard et de manière conclusive, par Schnorr (1971, « A Unified Approach to the Definition of Random Sequence » ; 1977, « A Survey of the Theory of Random Sequences »). Voir Delahaye, 1999, *Information, complexité et hasard*, p. 38-39.

³⁴ Voir Salmon, 1977, « Objectively Homogeneous Reference Classes ».

³⁵ Church, 1940, « On the concept of a random sequence ».

fonctions effectivement calculables, c'est-à-dire des fonctions récursives³⁶. Pour Church, si la valeur à la limite de la fréquence relative des 0 et des 1 composant une suite binaire infinie ne change pas dans les suites extraites par l'application de ce type de sélections admissibles car effectivement calculables, alors la suite originelle est aléatoire. La notion de suite infinie aléatoire développée par Church, de même que celle proposée par von Mises, fut ensuite démontrée insuffisante à fournir une bonne définition de la notion mathématique de hasard (ou d'aléatoire).³⁷

2.3.1. Déterminisme, indéterminisme et stabilité des fréquences à la limite

Bien qu'au cours du siècle dernier les faiblesses de la notion de hasard en tant que stabilité des fréquences à la limite aient été mises en évidence et que cette notion ait montré son caractère inapproprié, nous estimons intéressant de consacrer quelques réflexions à la relation que cette notion entretient avec les doctrines du déterminisme et de l'indéterminisme.

³⁶ Selon la notion de calculabilité définie formellement par Church en 1934, une fonction est effectivement calculable si elle est récursive, à savoir s'il s'agit d'une fonction définie à partir des fonctions de base (comme la fonction constante zéro et la fonction successeur) en utilisant la composition et la récursion. En 1936, Turing montra que cette notion de calculabilité effective proposée par Church est équivalente à la calculabilité par une machine de Turing universelle, cette dernière étant un objet mathématique abstrait ~~que~~ introduit par Turing. On parle de la « thèse Church-Turing » pour désigner cette identification mathématique entre fonctions récursives ou effectivement calculables et fonctions calculables par une machine de Turing. Plus précisément, une machine de Turing est constituée par un ruban infini divisé en cases consécutives, par une tête de lecture/écriture qui peut lire et écrire les symboles dans le ruban et se déplacer à gauche et à droite le long de ce dernier, par un registre d'états mémorisant l'état courant de la machine et par une table de règles de transition indiquant à la machine ce qu'elle doit écrire, comment elle doit se déplacer et quel est le nouvel état. L'action d'une machine de Turing (qui consiste à écrire un symbole dans la case où se trouve la tête de lecture/écriture, à déplacer la tête d'une cellule à droite ou à gauche, etc.) dépend de son état présent, du symbole scanné par la tête et de la table des règles de transition qui joue le rôle d'un programme pour la machine. En termes non formels, une fonction est calculable par une machine de Turing s'il existe un ensemble d'instructions (un algorithme) permettant à la machine de calculer la fonction donnée indépendamment du temps requis pour le faire. Vu que le temps à disposition pour le calcul et la mémoire de la machine sont infinis, toute fonction non calculable par une machine de Turing l'est non pas à cause d'une insuffisance de temps et/ou de mémoire, mais à cause d'une limite de pouvoir computationnel de la machine. Une machine de Turing est dite « universelle » si elle peut simuler le comportement de n'importe quelle autre machine de Turing, c'est-à-dire qu'elle a un pouvoir de calcul maximum : à partir du ruban et des règles de transition de n'importe quelle autre machine, elle produit le même résultat que cette dernière.

³⁷ La définition que Church proposa de ce qu'est un procédé de sélection admissible fut critiquée, entre autres, par Martin-Löf (1966a, « On the Concept of a Random Sequence » ; 1966b, « The definition of random sequences » ; 1969, « The Literature on von Mises' Kollektivs Revisited »), par Schnorr (1971, *op. cit.*) et par van Lambalgen (1987, « Von Mises' Definition of Random Sequences Reconsidered »). Pour une présentation et une analyse détaillées des définitions de ce qu'est un collectif qui ont été avancées respectivement par von Mises et par Church, des critiques qui leur ont été adressées et des solutions proposées, voir Salmon, 1977, « Objectively Homogeneous Reference Classes » ; Earman, 1986, *A Primer On Determinism* ; Delahaye, 1999, *op. cit.* ; Volchan, 2002, « What Is a Random Sequence? » ; Galavotti, 2005, *A Philosophical Introduction to Probability*.

La notion de hasard en tant que stabilité des fréquences à la limite porte sur les propriétés du résultat (ou de la performance) d'un processus, par exemple sur les propriétés d'une suite de « pile » et « face » engendrée par la répétition, dans les mêmes conditions, du lancer d'un même dé un nombre indéfini de fois. Elle peut être considérée comme une notion opérationnelle car elle implique la formulation et l'application de procédures de sélection sur un résultat donné afin d'en détecter les caractéristiques, notamment le caractère aléatoire. Cette notion de hasard ne fournit aucune information sur les propriétés du processus qui est à l'origine d'un résultat donné, le premier pouvant donc être intrinsèquement déterministe ou indéterministe. Elle n'implique aucun engagement en ce qui concerne le caractère fondamentalement déterministe ou indéterministe du processus ayant engendré une suite aléatoire (c'est-à-dire, en ce qui concerne les doctrines du déterminisme et de l'indéterminisme métaphysiques) : elle ne donne aucun indice sur la façon dont une telle suite a été produite, alors qu'une suite aléatoire peut être engendrée par un processus intrinsèquement indéterministe aussi bien que par un processus parfaitement déterministe (par exemple, par un générateur quantique aussi bien que par un générateur déterministe de nombres aléatoires).³⁸

En revanche, la notion de hasard en tant que stabilité des fréquences à la limite pourrait mettre en question la doctrine du déterminisme scientifique³⁹. En effet, bien que le caractère aléatoire d'une suite donnée (c'est-à-dire l'invariance des fréquences limites par rapport à un procédé de sélection admissible) n'implique pas qu'elle soit le produit d'un processus fondamentalement indéterministe, il a été montré qu'aucun collectif ne peut être engendré par un algorithme. Cela revient à dire qu'aucun collectif n'est calculable (selon les notions équivalentes de calculabilité effective de Church et de calculabilité par une machine de Turing universelle), et qu'on ne peut donc pas prédire la suite d'événements le constituant et les fréquences relatives, à la limite, de leurs attributs respectifs. La non calculabilité de tout collectif semble donc ébranler la thèse épistémologique du déterminisme scientifique dans le

³⁸ Dans un article consacré à la causalité probabiliste et à la distinction entre hasard subjectif et hasard objectif, Glennan (1997, « Probable Causes and the Distinction between Subjective and Objective Chance ») distingue ce qu'il appelle la « disposition comportementale externe » et la « structure interne » d'un mécanisme, c'est-à-dire ce qu'un mécanisme fait (son résultat) et comment un mécanisme fonctionne (le processus). Glennan introduit cette distinction afin de souligner qu'un comportement aléatoire, qu'il désigne par le terme « stochastique », peut être dû à un mécanisme dont la structure interne est parfaitement déterministe aussi bien que par un mécanisme indéterministe (dans ses termes, stochastique). Glennan utilise aussi l'exemple de l'engendrement de nombres aléatoires qui peuvent être produits par des générateurs déterministes aussi bien que par des générateurs quantiques.

³⁹ En ce qui concerne le déterminisme méthodologique, la notion de hasard en tant que stabilité des fréquences à la limite n'implique pas que les méthodologies prescrivant de formuler des hypothèses d'allure déterministe ou de faire l'hypothèse de l'uniformité des lois de la nature sont privées de valeur en science.

cas d'une suite infinie aléatoire au sens de von Mises et de Church, c'est-à-dire la prédictibilité, en principe, d'un collectif.⁴⁰ Cependant, on pourrait imaginer qu'une suite donnée ne soit pas calculable par une machine de Turing, même universelle, mais qu'elle le soit par un autre dispositif qu'on pourrait, en principe, construire et qui serait capable de remplir d'autres instructions et de mener à bien d'autres tâches qu'une machine de Turing universelle.⁴¹ Dans ce cas, on pourrait soutenir sans contradiction la thèse épistémologique du déterminisme scientifique sur la calculabilité ou prédictibilité, en principe, d'une suite donnée.

La neutralité de la notion de hasard en tant que stabilité des fréquences à la limite relativement aux doctrines du déterminisme et de l'indéterminisme métaphysiques implique-t-elle que cette notion ne renvoie ni au hasard subjectif, ni au hasard objectif ? En d'autres termes, cette notion correspond-elle au hasard subjectif ou au hasard objectif selon le cadre dans lequel elle est utilisée et, plus spécifiquement, selon le phénomène auquel elle est appliquée ? Dans la lignée de Salmon⁴², nous soulignons que la définition de collectif proposée par von Mises rend son caractère aléatoire relatif à un langage particulier dans lequel les règles de sélection sont formulées : dans un autre langage, une même suite pourrait ne pas remplir les conditions pour être un collectif et donc ne pas être aléatoire. Par conséquent, la notion de stabilité des fréquences à la limite selon la formulation de von Mises peut être considérée comme une notion de hasard subjectif car elle le définit relativement à un aspect propre à la connaissance humaine : le langage utilisé.

C'est afin d'éviter que le caractère aléatoire d'une suite soit relatif au choix d'un langage particulier que Church a affiné la définition de collectif : il a soutenu que les sélections admissibles doivent être caractérisées en termes de fonctions effectivement calculables. En effet, la calculabilité des opérations de sélection admissible est souvent considérée comme un critère d'objectivité car elle est évaluée sur la base des propriétés inhérentes aux fonctions mobilisées pour les calculer ou, selon l'équivalence avec la calculabilité de Turing, sur la base du pouvoir de calcul d'une machine de Turing qui est

⁴⁰ Dans ce cas, la thèse métaphysique, sur laquelle la version plus exigeante du déterminisme scientifique se base, ne serait pas nécessairement mise en cause. En effet, bien qu'une suite ne soit pas calculable (effectivement ou par une machine de Turing universelle) et qu'elle soit donc imprédictible, elle peut toujours être le produit d'un programme déterministe (par exemple, d'un programme consistant en l'énumération, l'un après l'autre, des événements la constituant).

⁴¹ On parle d'hyper-calcul ou de calcul super-Turing pour désigner des modèles théoriques de calcul capables de réaliser simultanément une infinité de calculs ou de procéder à des opérations de calcul qu'une machine de Turing universelle n'est pas capable de réaliser. Voir, par exemple, Copeland & Proudfoot, 1999, « Alan Turing's Forgotten Ideas in Computer Science ».

⁴² Salmon, 1977, « Objectively Homogeneous Reference Classes ».

universelle (thèse Church-Turing). C'est pourquoi la notion de stabilité des fréquences à la limite pourrait être considérée comme une notion de hasard objectif.⁴³

En conclusion de cette section, soulignons que la notion de hasard formulée par von Mises et ensuite développée par Church n'est satisfaisante ni du point de vue mathématique pour définir ce qu'est une suite infinie aléatoire, ni pour caractériser comme aléatoires des phénomènes physiques – et donc en particulier des phénomènes biologiques qui sont les sources causales de la variation au sein des populations naturelles (entre autres, les mutations génétiques et le bruit dans l'expression des gènes). La première raison pour laquelle la notion étudiée ne permet de caractériser comme aléatoires des phénomènes physiques est la suivante : elle ne fournit qu'une description des propriétés d'un résultat donné, mais est muette à propos du processus l'ayant engendré et ne fournit donc aucune explication du phénomène ayant eu lieu. En second lieu, comme von Mises le soulignait, cette notion n'est pas applicable à un événement singulier mais requiert qu'un même événement se répète indéfiniment de manière à avoir « une suite pratiquement illimitée d'observations uniformes », ce qui n'est pas très vraisemblable dans le cas de nombreux phénomènes naturels. Considérons, par exemple, une distribution particulière d'altérations génétiques (de mutations) le long d'une séquence d'ADN. Premièrement, il nous semble inintéressant de procéder à la sélection de sous-parties de cette séquence nucléotidique afin de voir si la fréquence relative des mutations est insensible par rapport à des opérations de sélection admissibles : cela ne nous dirait rien quant au processus causal qui est à l'origine des mutations observées et, par conséquent, quant au fait qu'elles sont ou pas des phénomènes se produisant au hasard. Deuxièmement, même si le recours à la notion de hasard en tant que stabilité des fréquences à la limite se révélait intéressant, il faudrait pouvoir analyser une séquence d'ADN indéfiniment longue ou un nombre très élevé de séquences identiques pour pouvoir l'appliquer et pour évaluer la fréquence relative des mutations génétiques à la limite.

2.4. La notion de hasard en tant qu'« incompressibilité » d'une suite

La notion de hasard en tant qu'incompressibilité d'une suite fut proposée de manière indépendante autour des années 1960 par Solomonov, par Kolmogorov et par Chaitin⁴⁴. Elle

⁴³ Cependant, on pourrait toujours avancer l'objection selon laquelle les notions de calculabilité et de hasard proposées par Church sont définies relativement aux capacités à remplir des instructions et à mener à bien des tâches qui sont celles d'une machine universelle de Turing, et non pas celles d'un autre modèle de calcul plus puissant qu'on pourrait, en principe, construire (les modèles d'hyper-calcul ou de calcul super-Turing).

émergea du cadre de la théorie algorithmique de l'information⁴⁵ qui, en se posant comme alternative à la théorie classique de l'information⁴⁶ et en complément à la théorie classique de la calculabilité développée dans les années 1930 par Church, par Post et par Turing, vise à définir le degré de complexité d'un objet ou d'un énoncé mathématique au moyen de la quantité minimale d'information nécessaire pour l'engendrer.⁴⁷

La théorie algorithmique de l'information fournit une définition rigoureuse de ce qu'est une suite binaire aléatoire de manière à rendre compte des intuitions concernant les caractéristiques du hasard, notamment l'irrégularité, le défaut de tout motif (*pattern*), de toute structure et de toute règle ou loi, et l'imprédictibilité. La notion de hasard est définie en termes d'incompressibilité d'un objet en format numérisé (c'est-à-dire une fois ramené sous la forme d'une suite binaire de 0 et 1) : de cette manière, la théorie algorithmique de l'information ne vise pas simplement à fournir une description du caractère aléatoire d'un objet ; elle vise aussi à l'expliquer en évoquant les caractéristiques du processus (du programme d'instructions ou algorithme) le plus court, en termes de quantité d'information, pouvant l'engendrer.

La notion de complexité algorithmique est la base de l'idée selon laquelle une suite aléatoire est incompressible. Dans le cadre de la théorie algorithmique de l'information, la complexité d'une suite binaire s pour un ordinateur de référence qu'est une machine de Turing M (ce qu'on appelle le plus souvent la « complexité de Kolmogorov » $K_M(s)$) est la taille du plus petit programme en M capable de produire la suite s , la taille du programme étant mesurée en bits ou en digits binaires. Par exemple, une suite très structurée car constituée par une centaine de 1 peut être décrite par le programme « imprimez 1 une centaine de fois » ; par conséquent, sa complexité algorithmique est beaucoup moins élevée que celle, par exemple, d'une suite très désordonnée de 0 et de 1 qui, au contraire, n'est caractérisée par aucune structure et ne peut donc être décrite par un programme que de manière extensive.

⁴⁴ Solomonov, 1964, « A Formal Theory of Inductive Inference » ; Kolmogorov, 1965, « Three Approaches for Defining the Concept of Information Quantity » ; Chaitin, 1966, « On the Length of Programs for Computing Finite Binary Sequences, Statistical Considerations ». Voir aussi Chaitin, 1987, *Information, Randomness and Incompleteness: Papers on Algorithmic Information Theory*.

⁴⁵ Pour une introduction à la théorie algorithmique de l'information, nous renvoyons à Li & Vitanyi, 1997[1993], *An Introduction to Kolmogorov complexity and its applications* ; Delahaye, 1999, *Information, complexité et hasard*.

⁴⁶ Shannon, 1948, « A Mathematical Theory of Communication ».

⁴⁷ La théorie algorithmique de l'information connut deux versions. La première correspond aux idées introduites indépendamment et presque simultanément par Solomonov, par Kolmogorov et par Chaitin dans les années 1960. La deuxième version, formulée entre autres par Chaitin et par Levin dans les années 1970, ne diffère de la première que sur un point : les algorithmes ou programmes informatiques considérés (c'est-à-dire, des ensembles d'instructions codées en chaînes de 0 et 1) doivent être « auto-délimités » c'est-à-dire qu'ils doivent contenir eux-mêmes l'indication de leur fin.

La définition de la complexité algorithmique est considérée comme pratiquement indépendante de la machine universelle de Turing utilisée. En effet, il a été montré, d'abord par Solomonov et ensuite par Chaitin et Kolmogorov, que la différence entre les complexités d'une même suite pour deux machines universelles différentes est toujours inférieure à une certaine constante c (négligeable) : c'est ce qu'on appelle le « théorème d'invariance » et qui justifie l'idée d'utiliser une machine universelle particulière pour estimer la complexité (objective) d'une suite donnée⁴⁸. La complexité de Kolmogorov d'une suite binaire est donc définie comme la taille du plus petit programme pour une machine universelle de Turing capable de la produire au moyen d'un formalisme universel de la computation. Cependant, la constance c caractérisant la différence entre la complexité d'une même suite calculée par des machines de Turing différentes n'est négligeable que dans le cas des suites infinies. Par conséquent, ce n'est que dans ce cas que la complexité algorithmique peut être considérée comme une propriété objective d'une suite, non pas au sens de la notion métaphysique d'objectivité absolue mais en termes d'invariance (ou robustesse). Au contraire, la complexité d'une suite finie est toujours relative à la machine universelle particulière utilisée pour l'estimer et donc, dans ce sens, peut être considérée comme subjective.⁴⁹

Dans le cadre de la théorie algorithmique de l'information, une suite aléatoire est définie comme une suite dont le plus petit programme écrit en binaire (qu'on appelle le « programme minimal »⁵⁰) pouvant la décrire a au moins la même taille que la suite même, c'est-à-dire qu'il consiste en la transcription complète de chacun de ses éléments. Par conséquent, une telle suite (aléatoire) est considérée comme incompressible car elle ne peut pas être réduite sous une forme plus compacte et économique : sa complexité algorithmique est au moins égale à sa taille en digits binaires. Cette définition du caractère aléatoire d'une suite porte sur les suites finies ; une suite infinie aléatoire est alors définie comme une suite dont les segments initiaux sont aléatoires au sens de la complexité de Kolmogorov.⁵¹

⁴⁸ Voir Calude (ed.), 2007, *Randomness and complexity from Leibniz to Chaitin*.

⁴⁹ Voir Delahaye & Zenil, 2007, « On the Kolmogorov-Chaitin complexity for short sequences ». Le fait que la complexité d'une suite calculée par deux machines universelles particulières diffère toujours d'une certaine constante c est à l'origine de l'impossibilité de déterminer de manière stable et précise la complexité de Kolmogorov d'une suite donnée. Sur ce point, nous renvoyons à Delahaye & Zenil, 2008, « Towards a stable definition of Kolmogorov-Chaitin complexity ».

⁵⁰ Sur la base de la définition de la notion de hasard en théorie algorithmique de l'information, tout programme minimal est aléatoire. En effet, si le programme P_i était considéré comme le programme minimal d'une suite binaire donnée mais n'était pas aléatoire, cela signifierait qu'il existe un autre programme P_{ii} plus petit que P_i et qui peut le générer. Par conséquent, P_i ne serait pas le programme minimal de la suite binaire donnée. Voir Chaitin, 1975b, « Randomness and Mathematical Proof ».

⁵¹ La notion de complexité a connu des développements intéressants pendant les quarante dernières années. Entre autres, Bennett (1988a, « Démons, machines et thermodynamique » ; 1988b, « Logical Depth and Physical Complexity » ; 1990, « How to define complexity in physics, and why ») a distingué ce qui est complexe car

2.4.1. Déterminisme, indéterminisme et incompressibilité

La notion de hasard en tant qu'incompressibilité permet de discerner entre les suites – et, plus généralement, les objets – qui peuvent être produites par un ordinateur de référence (c'est-à-dire, par une machine universelle de Turing) et celles qui ne le sont pas mais qui ne peuvent être décrites que par leur transcription complète. En d'autres termes, c'est une notion permettant de distinguer ce qui est prédictible au moyen d'un modèle de calcul universel, de ce qui ne l'est pas : une suite incompressible est une suite dont on ne peut pas deviner à l'avance l'un des éléments sur la base des éléments le précédant. Analysons dans l'ordre comment la notion de hasard en tant qu'incompressibilité d'une suite se situe par rapport aux conceptions du hasard subjectif et objectif et par rapport aux doctrines du déterminisme et de l'indéterminisme.

D'abord, nous remarquons que, si l'on considère que la complexité algorithmique calculée par une machine universelle de Turing est une propriété d'une suite qui est objective car indépendante, à une constante c près considérée comme négligeable, de toute machine universelle particulière, alors la notion d'incompressibilité peut être considérée comme une notion de hasard objectif, non pas au sens de la notion d'objectivité absolue, mais en termes d'invariance (ou robustesse). Cependant, comme nous l'avons déjà souligné plus haut, le caractère négligeable de cette constante c n'est pas l'objet d'un consensus : on estime que la complexité calculée par une machine universelle de Turing est une propriété indépendante de la machine particulière utilisée dans le cas des suites infinies, mais que la constante c devient de moins en moins négligeable au fur et à mesure qu'on a à faire à des suites de plus en plus courtes. Par conséquent, la nature objective de la notion de hasard en tant qu'incompressibilité, non pas au sens absolu mais en termes d'invariance de cette propriété des suites, ne peut être affirmée que dans le cas des suites infinies. Enfin, de manière

aléatoire au sens de la complexité de Kolmogorov (par exemple, un gaz) de ce qui est complexe car très organisé (par exemple, un être vivant). Il a proposé l'expression de « profondeur logique » pour désigner la complexité organisée qu'il rattache au temps de calcul requis pour que le programme minimal, sur une machine de Turing universelle, puisse générer un objet donné, ce dernier ayant été préalablement numérisé. Par conséquent, un objet caractérisé par une grande profondeur logique est un objet dont l'origine la plus probable est un long calcul. Le fait qu'il ait une grande complexité ordonnée n'implique pas que sa complexité aléatoire soit aussi élevée. Par exemple, les 100 000 premières décimales de Pi sont, contrairement à l'intuition commune, hautement organisés : pour les décrire, on sait construire des programmes relativement courts mais qui requièrent beaucoup de calculs. Par conséquent, la suite constituée par les 100 000 décimales de Pi a une grande profondeur logique mais une petite complexité algorithmique. Bennett a montré que sa définition de la profondeur logique (ou complexité organisée) est relativement robuste car pratiquement indépendante de la machine universelle choisie pour faire tourner le programme : il en conclut qu'il s'agit d'une notion absolue car elle se réfère à des mécanismes universels de calcul. Pour une analyse critique du caractère absolu de la notion de complexité de Kolmogorov et de la notion de profondeur logique de Bennett, voir Delahaye, 1999, *Information, complexité et hasard*.

analogue au cas de la notion de hasard en tant que stabilité des fréquences à la limite formulée par Church, on pourrait objecter que la notion d'incompressibilité d'une suite est une notion de hasard subjectif car elle est relative à la capacité de calcul d'une machine de référence qu'est une machine universelle de Turing – par opposition à l'ensemble de tous les modèles qu'on pourrait en principe construire.

En deuxième lieu, nous remarquons que cette notion est indifférente aux doctrines du déterminisme et de l'indéterminisme métaphysiques. D'une part, une suite peut être aléatoire au sens où le programme minimal pouvant l'engendrer est aussi long que la suite même et, en même temps, elle peut avoir été engendrée par un processus complètement déterministe au sens métaphysique de ce terme. Des exemples paradigmatiques de ce genre d'objets émergent de la théorie du chaos : en effet, l'instabilité des systèmes chaotiques rend leur évolution imprédictible sur le long terme alors qu'elle est parfaitement déterministe, c'est-à-dire bien que l'état initial d'un système chaotique, avec les lois régissant son évolution, détermine de manière univoque et précise tous ses états futurs.⁵² C'est pourquoi les physiciens doivent faire face à l'absence de tout chemin de calcul économique en faisant recours aux simulations par ordinateur : c'est seulement de cette manière qu'ils peuvent étudier l'évolution de ce genre de systèmes très instables mais déterministes, sans pour autant en fournir une explication⁵³. D'autre part, une suite aléatoire car incompressible peut aussi bien être le résultat d'un processus fondamentalement indéterministe. Nous ne discuterons pas ce point car l'imprédictibilité entretient une relation beaucoup moins problématique avec l'indéterminisme qu'avec le déterminisme métaphysique.

En ce qui concerne le déterminisme scientifique, la notion de hasard en tant qu'incompressibilité pourrait poser problème car, comme nous l'avons souligné ci-dessus, elle désigne l'imprédictibilité, en termes de calcul par une machine universelle, d'une suite donnée. Le fait qu'il n'existe aucun programme plus petit que la suite même pouvant l'engendrer signifie qu'il n'existe aucune théorie permettant de prédire, à partir d'un élément de la suite, ses éléments ultérieurs. Cela représente une mise en cause du déterminisme scientifique, doctrine selon laquelle il est possible, en principe, de prédire l'évolution d'un système donné à partir de son état initial et avec les lois le régissant. Une suite aléatoire car incompressible est en effet considérée comme privée de toute loi : la seule règle régissant sa

⁵² Pour une introduction à la notion de hasard mobilisée dans la caractérisation des systèmes chaotiques, nous renvoyons au chapitre 1 de ce travail, plus précisément à la section 1.5.3 consacrée à la notion de hasard en tant que sensibilité aux conditions initiales.

⁵³ Le statut des simulations par ordinateur en science et les implications épistémologiques de leur utilisation sont actuellement au centre d'un débat philosophique controversé. Entre autres, voir Humphreys, 2004, *Extending Ourselves. Computational Science, Empiricism and Scientific Method*.

structure (ou plutôt l'absence de toute structure d'une telle suite) est l'absence de toute règle ; par conséquent, une suite aléatoire n'est pas susceptible de recevoir une explication possible.

On pourrait toutefois avancer les deux objections suivantes. D'une part, bien qu'une machine universelle de Turing ait un pouvoir de calcul maximum et soit donc censée fournir une définition objective, en termes d'invariance, de la propriété d'incompressibilité d'une suite, cette dernière n'est indépendante de la machine particulière utilisée pour faire tourner un programme que dans le cas des suites infinies. Sur cette base, on pourrait conclure que la notion de hasard en tant qu'incompressibilité (d'une suite) ne met en cause la thèse épistémologique du déterminisme scientifique que dans le cas des suites infinies. D'autre part, on pourrait argumenter en faveur du caractère relatif, donc épistémique, de l'imprédictibilité d'une suite par rapport à une machine universelle de Turing, et non pas relativement à toute machine d'hyper-calcul qu'on pourrait, en principe, construire. On pourrait donc en conclure que la notion de hasard en tant qu'incompressibilité d'une suite ne met pas en cause la validité du déterminisme scientifique.

2.5. La notion de hasard en tant que « satisfaction de tous les tests statistiques effectifs »

Avant de préciser quel est le consensus actuel quant à la notion mathématique de hasard, nous procédons à l'introduction d'une dernière définition de ce qu'est une suite infinie aléatoire : c'est la notion de suite satisfaisant tous les tests statistiques effectifs d'aléatoirité. Elle fut proposée par Martin-Löf en 1966, alors que la théorie algorithmique de l'information était encore à ses débuts. Ensuite, elle s'est révélée être équivalente à la notion de suite infinie aléatoire en termes de complexité algorithmique et a fourni ce qui est aujourd'hui considéré comme la bonne définition mathématique de ce que signifie être aléatoire.

De manière plus précise, sans pour autant rentrer dans les détails techniques⁵⁴, une suite infinie aléatoire au sens de Martin-Löf est une suite qui réussit un nombre dénombrable de tests statistiques effectifs d'aléatoirité que seul un ensemble de mesure nulle de suites ne

⁵⁴ Nous nous limitons à une brève introduction de la notion de « suite aléatoire » proposée par Martin-Löf, dans les années 1960. Martin-Löf, 1966a, « On the Concept of a Random Sequence » ; 1966b, « The definition of random sequences ».

satisfait pas.⁵⁵ Pour une suite donnée, chaque test permet d'éliminer une certaine propriété de régularité effectivement testable, qu'on considère n'être pas compatible avec la propriété d'être aléatoire, et que les suites aléatoires au sens de Martin-Löf ne vérifient donc pas. Par exemple, on peut construire un test statistique pour chacune des propriétés suivantes, considérées comme incompatibles avec la propriété d'être aléatoire : « ne pas satisfaire à la loi forte des grands nombres », « ne pas vérifier la loi du logarithme itéré », « avoir toujours plus de 0 que de 1 dans les n premiers digits », « avoir toujours la différence entre le nombre de 1 et le nombre de 0 inférieure à une quantité fixée ».⁵⁶

En ce qui concerne les caractéristiques de la notion de hasard en tant que satisfaction de tous les tests statistiques effectifs par une suite infinie, en particulier sa relation avec les doctrines du déterminisme et de l'indéterminisme, nous soulignons brièvement les points suivants. Premièrement, c'est une notion définie en termes de résultat et qui ne fournit donc qu'une description, mais aucune explication, du caractère aléatoire d'une suite et, plus généralement, d'un objet donné. Cependant, il faut tout de même rappeler que cette notion de hasard introduite par Martin-Löf s'est révélée être équivalente (du moins dans le cas des suites infinies) à la notion de hasard en tant qu' incompressibilité (cette dernière définissant le caractère aléatoire d'une suite en termes des caractéristiques du programme minimal pouvant l'engendrer). Deuxièmement, la notion de satisfaction (par une suite) de tous les tests statistiques effectifs d'aléatoirité pourrait être considérée comme une notion de hasard objectif car elle présuppose la satisfaction, par une suite donnée, d'un test universel donc maximal. Cependant, nous rappelons que la construction même théorique d'un tel test universel s'est révélée impossible ; c'est pourquoi le caractère aléatoire d'une suite peut, au contraire, être considérée comme subjectif car relatif à un ensemble de tests statistiques particuliers, c'est-à-dire à un ensemble limité de propriétés statistiques connues.⁵⁷

⁵⁵ Ce qu'on appelle le « test U » et le « test Die » sont des tests statistiques incluant l'ensemble fini de tests statistiques que l'on connaît aujourd'hui. Cependant, ils ne comprennent pas l'ensemble complet et infini des « tests statistiques effectifs » d'aléatoirité évoqué dans la définition de la notion de hasard proposée par Martin-Löf. Voir L'Ecuyer & Simard, 2007, « TestU01: A C Library for Empirical Testing of Random Number Generators ».

⁵⁶ Voir aussi Delahaye, 1999, *Information, complexité et hasard*, p. 51. De manière analogue à la machine de Turing universelle, on pourrait imaginer un test universel résumant tous les tests statistiques effectifs et mettant donc à l'épreuve une suite donnée en ce qui concerne toutes les propriétés de régularité qu'une séquence aléatoire ne possède pas. Cependant, la construction de ce test s'est révélée impossible même d'un point de vue théorique. De plus, même si l'on pouvait construire ce genre de test, son utilisation ne serait pas praticable - car il s'agirait de soumettre une suite donnée, dont on veut évaluer si elle est aléatoire ou pas, à un ensemble infini de tests statistiques - et son résultat ne pourrait pas être testé à son tour - car cela ne pourrait pas être fait à la main, mais nécessiterait l'utilisation d'un autre test statistique universel d'aléatoirité. Pour une analyse critique de ce problème, nous renvoyons à Schnorr, 1973, « Process Complexity and Effective Random Tests ».

⁵⁷ Nous soulignons que, si un sous-ensemble fini de propriétés statistiques, et donc de tests statistiques effectifs, se révélait être approprié et fiable pour déterminer le caractère aléatoire d'une suite, on pourrait montrer qu'une

Troisièmement, de la même manière que la notion de hasard en tant qu'incompressibilité et de la notion de hasard de von Mises et de Church, la notion introduite par Martin-Löf n'a aucune implication quant au caractère fondamentalement déterministe ou indéterministe du processus à l'origine d'une suite aléatoire et, plus généralement, d'un objet aléatoire donné. Enfin, l'existence d'un test universel d'aléatoirité mettrait en cause le déterminisme scientifique car il éliminerait toute propriété de régularité pour les suites qui le satisferaient. Cependant, nous rappelons une dernière fois qu'un tel test ne peut pas être construit, et que cette impossibilité est à la fois pratique et théorique.

2.5.1. Le consensus autour de la bonne définition mathématique de hasard

Indiquons brièvement les relations d'équivalence – bien que dans la plupart des cas, elles ne soient valides que dans le cas de suites aléatoires infinies – entre les notions mathématiques de hasard que nous avons introduites dans ce chapitre. Ensuite, nous énoncerons les propriétés qui, selon le consensus qui prévaut aujourd'hui en mathématiques, caractérisent tout ce qui est aléatoire.

Nous avons déjà souligné que, lorsqu'on considère des suites infinies, la notion de hasard en tant que stabilité des fréquences à la limite est équivalente à la notion d'incompressibilité (d'une suite) de la théorie algorithmique de l'information. Cette dernière se révèle aussi être équivalente, bien que de manière approximative, à la notion mathématique de hasard en tant qu'équiprobabilité : il a en effet été montré que, si une suite binaire est aléatoire au sens de Kolmogorov, alors les fréquences relatives des 0 et des 1 dans n'importe quel échantillon de la suite est voisine de $\frac{1}{2}$, ce qui en d'autres termes signifie que les digits 0 et 1 sont équiprobables. De plus, vu que n'importe quel échantillon d'une suite aléatoire au sens de l'incompressibilité tend à être caractérisé par les mêmes propriétés que sont les fréquences relatives de 0 et 1 égales à $\frac{1}{2}$, la notion de hasard en tant qu'équiprobabilité semble recouper aussi la notion de von Mises et de Church. Cependant, cette affirmation fut contredite par Ville qui, en 1939, montra l'existence de collectifs qui ne remplissent pas la condition pour pouvoir parler du hasard au sens d'équiprobabilité⁵⁸. Enfin, nous rappelons

suite aléatoire l'est objectivement, non pas au sens absolu, mais selon une notion d'objectivité plus appropriée en science qui est définie en termes d'invariance (ou robustesse). Voir la troisième partie de ce travail.

⁵⁸ Ville (1939, *Etude critique sur la notion de collectif*) montra l'existence de séquences considérées par von Mises comme des suites aléatoires de 0 et de 1, mais qui sont caractérisées par un biais pour les 1 ou pour les 0 : les 0 et les 1 n'ont donc pas la même fréquence à la limite ou, en d'autres termes, ne sont pas équiprobables

que la notion de hasard en tant qu'incompressibilité d'une suite est équivalente à la notion de hasard introduite par Martin-Löf en 1966, bien que ce soit seulement dans le cas de suites infinies.

Sur la base de ces équivalences, le consensus actuel au sujet de ce que signifie « être aléatoire » consiste à affirmer que l'une des propriétés suivantes, qui sont équivalentes, doit être vérifiée⁵⁹ :

- 1) Ne satisfaire aucune propriété de régularité effectivement testable, c'est-à-dire passer tous les tests statistiques effectifs d'aléatoire⁶⁰ ;
- 2) Avoir un contenu en information incompressible, c'est-à-dire avoir une complexité algorithmique maximale⁶¹ ;
- 3) Être imprédictible, c'est-à-dire être tel qu'il est impossible de construire un système de pari qui garantisse un gain infini⁶².

2.5.2. Les notions mathématiques de hasard en biologie

En dépit du consensus actuel quant à ce que signifie être aléatoire en termes mathématiques, nous soulignons encore une fois que, de manière générale, une notion de hasard appropriée pour caractériser un phénomène physique ou biologique ne doit pas se limiter à décrire les caractéristiques de ce dernier au niveau du résultat – c'est-à-dire, en utilisant les termes de Glennan, la disposition comportementale externe qu'on peut observer. Elle doit aussi décrire les propriétés du processus à l'origine du phénomène physique ou biologique considéré afin de l'expliquer. C'est pourquoi nous suggérons que, au premier abord, la notion d'équiprobabilité définie en termes de processus (c'est-à-dire la notion d'échantillonnage indiscriminé) et la notion de hasard en tant qu'incompressibilité (d'une suite) semblent être plus pertinentes que les autres notions mathématiques de hasard pour caractériser un phénomène physique ou biologique aléatoire.

Contrairement à cette première impression, au cours de notre travail, nous montrerons que, du point de vue empirique, la notion d'échantillonnage indiscriminé n'est que

⁵⁹ Voir Delahaye, 1999, *Information, complexité et hasard*.

⁶⁰ Martin-Löf, 1966a, « On the Concept of a Random Sequence » ; 1966b, « The definition of random sequences ».

⁶¹ Levin, 1974, « Laws of Information Conservation and Aspects of the Foundation of Probability Theory » ; Chaitin, 1975a, « A Theory of Program Size Formally Identical to Information Theory ».

⁶² Von Mises, 1961[1928], *Probability, Statistics and Truth* ; Schnorr, 1971, « A Unified Approach to the Definition of Random Sequence » ; 1977, « A Survey of the Theory of Random Sequences ».

partiellement appropriée pour décrire les phénomènes qui sont au centre de notre recherche, c'est-à-dire les sources causales de la variation biologique (entre autres, les mutations génétiques et le bruit dans l'expression des gènes).⁶³ De plus, nous soutenons que la notion d'aléatoire en termes de complexité algorithmique n'est pas une notion adéquate pour décrire et expliquer ce genre de phénomènes biologiques. En effet, il est à notre avis paradoxal d'affirmer qu'un phénomène biologique ne peut pas être comprimé (c'est-à-dire prévu au moyen d'un programme informatique), et que dans ce sens il est le fruit du hasard, si l'on ne connaît que très peu à son sujet (ce qui est le cas, par exemple, pour les mutations génétiques et pour les fluctuations stochastiques affectant l'expression des gènes). Par exemple, une fois ramené à une suite binaire (de 0 et 1), un phénomène biologique pourrait être compressible et donc effectivement calculable par une machine de Turing universelle mais n'avoir jamais lieu dans la nature, alors qu'un autre, bien qu'incalculable par une machine universelle de Turing, pourrait avoir lieu très souvent et être le résultat d'un processus naturel que les biologistes ignorent à l'état de leurs recherches. De plus, un phénomène biologique pourrait être calculable par un algorithme particulier mais cela ne nous fournirait aucune raison de penser que, dans le monde biologique, il est aussi généré par la réalisation du même ensemble d'instructions.

Nonobstant notre scepticisme à l'égard de l'utilisation des notions de hasard en tant qu'équiprobabilité et en tant qu'incompressibilité en biologie, nous rappelons que, de l'autre côté, ces deux notions sont invoquées pour décrire certains phénomènes biologiques. Par exemple, dans le chapitre 7, nous verrons que, dans les modèles que les généticiens des populations construisent pour décrire et prédire le processus évolutif, les mutations génétiques pouvant affecter un site donné de la séquence d'ADN sont représentées comme étant équiprobables à tout moment au cours d'une génération et indépendantes des autres mutations affectant tout autre site de la même séquence.

Quant à la notion d'aléatoire en termes de complexité de Kolmogorov, elle est actuellement utilisée avec succès pour analyser des séquences d'ADN, plus particulièrement afin de comparer des séquences génomiques diverses et, sur cette base, reconstruire des arbres phylogénétiques.⁶⁴ Cependant, l'utilisation de cette notion mathématique de hasard en biologie⁶⁵, et encore plus en physique⁶⁶, n'empêche pas que les scientifiques dont la recherche

⁶³ En particulier, voir les chapitres 7 et 9.

⁶⁴ Entre autres, voir Delahaye, 1999, *op. cit.*

⁶⁵ La théorie algorithmique a été utilisée par Chaitin (1979, « Toward a Mathematical Definition of Life ») et par Küppers (1990, *Information and the Origin of Life*) pour caractériser de manière formelle un être organisé.

⁶⁶ Voir Zurek, 1989, « Algorithmic Randomness and the Physical Entropy »; Lloyd & Pagels, 1988,

porte sur ce sujet manifestent une certaine prudence quant à son application à des phénomènes biologiques et physiques. Entre autres, Chaitin⁶⁷ a affirmé que la théorie de la complexité qu'il propose se place au niveau des mathématiques pures et n'est donc pas une théorie réaliste et pratique pouvant être appliquée sans problème au monde de la biologie et de la physique. Delahaye a écrit⁶⁸ qu'il n'y a pas encore eu de réponse satisfaisante à la question de savoir si les théories mathématiques de l'information – entre autres, la théorie algorithmique – peuvent servir à la biologie. Par exemple, il est selon lui évident que les processus étudiés par la biologie moléculaire comme la réplication de la molécule d'ADN impliquent de l'information ; cependant, il estime qu'il n'y a aucune raison de penser que le génome d'un organisme individuel est écrit sous la forme la plus compressée et économique possible. Nous partageons pleinement cette réflexion de Delahaye. Nous ajoutons que, bien que très probablement le génome d'un individu possède une valeur de complexité, au sens de la profondeur logique de Bennett, car il est le résultat d'une longue histoire évolutive (en termes algorithmiques, d'une sorte de long calcul ayant eu lieu au cours du temps), d'une part, cette valeur est très difficile à déterminer ; d'autre part, l'existence de cette valeur ne nous garantit pas qu'elle nous dise quelque chose d'intéressant sur l'origine et les propriétés du génome considéré.

Enfin, il est intéressant de remarquer qu'une idée très proche de la notion de hasard en tant qu'incompressibilité (d'une suite) figure dans *Le hasard et la nécessité* (1970) de Monod où ce dernier vise à faire la part entre ce qui est le fruit du hasard et ce qui est régi par la nécessité dans l'évolution biologique.

Dans la section intitulée « Ontogenèse moléculaire », Monod suggère la célèbre hypothèse selon laquelle les suites de radicaux d'acides aminés (de 100 à 10 000) constituant les séquences polypeptidiques des protéines ne sont pas prédictibles. Il affirme que les séquences des protéines sont régies par la loi générale du hasard. Plus précisément, il écrit :

« ces structures sont 'au hasard' en ce sens que, connaissant exactement l'ordre de 199 résidus dans une protéine qui en comprend 200, il est impossible de formuler aucune règle, théorique ou empirique, qui permettrait de prévoir la nature du seul résidu non encore identifié par l'analyse. »⁶⁹

Nous n'entendons pas évaluer si cette affirmation est effectivement vraie. Nous souhaitons simplement souligner que Monod ne fait aucune référence à la notion de hasard introduite

« Complexity as Thermodynamic Depth ».

⁶⁷ Chaitin, 2006, « Complexité, logique et hasard. Entretien avec Gregory Chaitin ».

⁶⁸ Delahaye, 1999, *Information, complexité et hasard*.

⁶⁹ Monod 1970, *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, p. 127.

indépendamment par Solomonov, par Kolmogorov et par Chaitin dans les années 1960, alors qu'au moment de la publication de *Le hasard et la nécessité* la théorie de la complexité algorithmique avait déjà vu le jour. Cependant, l'idée qu'il introduit en ce qui concerne les séquences polypeptidiques des protéines, qu'il considère comme « indéchiffrables », correspond à la notion d'incompressibilité (d'une suite) selon laquelle une suite est aléatoire si le programme minimal pour la générer est approximativement aussi long que la suite même.

Monod ne se limite pas à décrire une telle conception du hasard, mais en revendique la nature objective. Il affirme :

« Dire de la séquence des amino-acides dans un polypeptide qu'elle est 'au hasard' ne revient nullement, il faut insister là-dessus, à un aveu d'ignorance, mais exprime une constatation de fait : à savoir que, la fréquence moyenne avec laquelle tel résidu est suivi de tel autre dans les polypeptides est égale au *produit* des fréquences moyennes de chacun des deux résidus dans les protéines en général »⁷⁰.

Dans ce passage, Monod veut souligner avant tout que la notion de hasard désigne une propriété des séquences des protéines, notamment leur imprédictibilité, qui n'est pas due à notre ignorance car elle trouve son origine dans le défaut, propre aux séquences polypeptidiques, de toute structure et de toute règle. Cependant, à aucun moment Monod n'oppose de manière explicite une telle conception du hasard au caractère fondamentalement déterministe des processus biochimiques pouvant être l'origine causale de ces séquences.

2.6. Les théories philosophiques des probabilités

Dans le premier chapitre, nous avons d'abord distingué deux conceptions majeures du hasard : le hasard subjectif et le hasard objectif. Ensuite, nous avons introduit et analysé un certain nombre de notions de hasard qui sont caractérisées par des significations et des rôles différents et dont chacune représente une des déclinaisons que les conceptions du hasard subjectif et objectif ont reçues au cours de l'histoire de la pensée philosophique et scientifique⁷¹. Enfin, dans le présent chapitre, nous avons introduit plusieurs notions

⁷⁰ Monod, 1970, *op. cit.*, p. 127.

⁷¹ Du côté des notions de hasard subjectif, nous avons identifié le hasard en tant qu'ignorance des causes ; du côté des notions du hasard objectif figurent le hasard en tant qu'absence d'un dessein, le hasard en tant que

mathématiques de hasard⁷² et nous avons montré que leur appartenance à une conception du hasard en tant que subjectif ou objectif est plus ouverte à discussion qu'en ce qui concerne les notions analysées précédemment.

Le caractère double de la notion de hasard (subjectif et objectif)⁷³ trouve son correspondant dans ce que Hacking a appelé « les deux visages de Janus » propres aux probabilités qui, comme nous allons le voir, représentent deux manières différentes d'interpréter⁷⁴ le calcul de probabilité.

« D'un côté, elle [la probabilité] est statistique et s'applique aux lois stochastiques des phénomènes aléatoires. De l'autre, elle est épistémique et concerne l'évaluation des degrés auxquels il est raisonnable de croire en la vérité de propositions n'ayant rien de statistique. »⁷⁵

Dans ce passage, Hacking montre que la notion moderne de probabilité englobe deux idées distinctes selon lesquelles la probabilité est, d'une part, l'inclination des événements à se produire et, d'autre part, le motif rationnel de croire, c'est-à-dire sur la base des informations disponibles, que des événements vont se produire. Daston⁷⁶ affirme que la distinction de ces deux manières de concevoir les probabilités date de la réflexion de Poisson, de Cournot et d'Ellis au début des années 1840.

Dès l'axiomatisation par Kolmogorov dans les années 1930, le calcul des probabilités a été interprété en privilégiant l'un ou l'autre de ce que Hacking appelle les deux visages de la probabilité. En suivant la distinction soulignée par ce dernier et la terminologie utilisée par Gillies⁷⁷, nous identifions deux groupes distincts qui rassemblent respectivement, pour l'un les interprétations épistémiques, pour l'autre les interprétations objectives du calcul des

rencontre de séries causales indépendantes, le hasard en tant que sensibilité aux conditions initiales et le hasard pur. Nous renvoyons au chapitre 1.

⁷² Le hasard en tant qu'équiprobabilité, le hasard en tant que stabilité des fréquences à la limite, le hasard en tant qu'incompressibilité d'une suite et le hasard en tant que satisfaction (par une suite) de tous les tests statistiques effectifs.

⁷³ Voir la section 1.2 du premier chapitre.

⁷⁴ Hajek (2008, « Interpretations of Probability ») souligne que parler d'« interprétation » est trompeur car des quantités diverses, qui n'ont rien à voir avec la probabilité (par exemple la longueur, le volume, la masse), satisfont les axiomes proposés par Kolmogorov du calcul des probabilités et, dans ce sens, sont des « interprétations » des probabilités. Il remarque aussi que certaines conceptions de ce qu'est la probabilité ne sont pas compatibles avec l'axiomatisation de Kolmogorov. C'est pourquoi Hajek suggère qu'on devrait plutôt parler de l'« analyse des divers concepts de probabilités ». Bien que la remarque de Hajek soit pertinente, dans le cadre de ce travail nous avons choisi d'utiliser le terme « interprétation » selon son usage le plus répandu en philosophie des probabilités : ce terme désigne la réponse à la question « qu'est-ce que la probabilité ? ».

⁷⁵ Hacking, 2002[1975], *L'Émergence de la probabilité*, p. 38.

⁷⁶ Daston, 1988, *Classical Probability in the Enlightenment*, p. 191.

⁷⁷ Gillies, 2000, *Philosophical theories of probability*.

probabilités.⁷⁸ Selon le premier groupe d'interprétations, la probabilité concerne la connaissance, l'ignorance ou la croyance humaines. Elle est donc épistémique, dyadique car relative à un ensemble de connaissances, et est le degré de connaissance, d'ignorance partielle, de croyance rationnelle ou de croyance personnelle. Selon les interprétations objectives, la probabilité est au contraire une propriété des événements du monde réel, indépendante de toute connaissance, ignorance ou croyance humaines. Elle est donc de nature objective, monadique, et est l'idéalisation de la fréquence relative des événements, la mesure de leur degré de possibilité, ou la mesure de propensions.

Au sein des deux visages de la probabilité, c'est-à-dire des interprétations épistémiques et des interprétations objectives, toujours en suivant Gillies⁷⁹ nous distinguons cinq « théories philosophiques » des probabilités selon lesquelles la notion de probabilité a des significations différentes. D'une part, la théorie classique (Laplace), la théorie logique (Keynes), et la théorie subjective (Ramsey, de Finetti), où les probabilités sont conçues comme quelque chose qui est de nature épistémique. D'autre part, la théorie fréquentiste (Reichenbach, von Mises) et la théorie propensionniste (Popper, Miller, Fetzer), où les probabilités sont conçues comme quelque chose d'objectif. À ces deux dernières théories, nous ajoutons une version objective de la théorie classique des probabilités.

Décrivons de manière synthétique et simple les principales théories philosophiques des probabilités, toujours en suivant la présentation qu'en fournit Gillies et en introduisant dans l'ordre celles correspondant à une interprétation épistémique et celles correspondant à une interprétation objective des probabilités.⁸⁰ Nous précisons que les définitions et les distinctions que nous allons présenter ci-dessous se révéleront utiles dans le cadre de notre recherche. Plus précisément, nous devons les considérer lorsque nous essaierons de répondre à la question de savoir comment on doit interpréter les probabilités figurant dans les descriptions formelles (mathématiques) des sources causales de la variation biologique (les mutations génétiques et le bruit dans l'expression de gènes). Nous poserons et analyserons cette question afin de résoudre le problème qui se trouve au centre de notre recherche, notamment afin de clarifier la signification et du rôle de la notion de hasard dans la caractérisation de ce genre de phénomènes biologiques.

⁷⁸ Nous rappelons que d'autres couples d'appellations ont été utilisés pour caractériser la distinction entre interprétations épistémiques et interprétations objectives des probabilités : par exemple, les couples subjectif/objectif (Popper, 1934, *Logic of Scientific Discovery*), épistémique/aléatoire (Hacking, 2004[2001], *L'Ouverture au probable*), numéro1/numéro2 (Carnap, 1962, *Logical Foundations of Probability*).

⁷⁹ Gillies 2000, *op. cit.*

⁸⁰ Pour une introduction philosophique et exhaustive aux théories des probabilités, nous renvoyons aussi à Galavotti, 2005, *A Philosophical Introduction to Probability* ; Mellor, 2005, *Probability. A Philosophical Introduction* ; Hajek, 2008, « Interpretations of Probability ».

2.6.1. Les interprétations épistémiques des probabilités

Du côté des interprétations épistémiques des probabilités figurent les trois théories philosophiques suivantes. Premièrement, selon la théorie classique développée par Laplace au début du XIX^e siècle, les probabilités sont relatives, en partie à l'ignorance, en partie à la connaissance humaine. On doit faire appel au calcul des probabilités lorsqu'on ne connaît pas de manière précise et complète les conditions initiales d'un système physique donné et/ou les lois de la nature qui le régissent et qu'on ne peut donc pas prédire ses états à tout instant futur et passé, bien que son évolution soit parfaitement déterministe. Plus précisément :

« La théorie des hasards consiste à réduire tous les événements du même genre à un certain nombre de cas également possibles, c'est-à-dire tels que nous soyons également indécis sur leur existence, et à déterminer le nombre de cas favorables à l'événement dont on cherche la probabilité. Le rapport de ce nombre à celui de tous les cas possibles est la mesure de cette probabilité, qui n'est ainsi qu'une fraction dont le numérateur est le nombre des cas favorables, et dont le dénominateur est le nombre de tous les cas possibles. »⁸¹

Nous rappelons que, selon la version laplacienne de la théorie classique, la probabilité d'un événement est évaluée *a priori* par la considération d'un ensemble d'événements mutuellement exclusifs dont on n'a aucune raison de penser qu'ils ne sont pas équipossibles et donc, sur la base du principe de raison insuffisante (ou principe d'indifférence), qu'ils n'ont pas la même probabilité d'avoir lieu. Dans ce cadre, la probabilité n'est donc pas la mesure d'une propriété des événements mêmes se produisant dans le monde, ce dernier étant considéré comme un tout complètement déterministe. Cette notion désigne plutôt la mesure de notre degré d'ignorance partielle quant aux propriétés physiques et causales des événements.

Deuxièmement, la théorie logique des probabilités, qui se développa au début du XX^e siècle à Cambridge (Russel, Moore, Wittgenstein, Keynes et Jeffreys) et aussi au Cercle de Vienne (entre autres, Carnap), est considérée comme faisant partie des interprétations épistémiques. Selon Keynes⁸², qui en est l'un des exposants majeurs, la théorie des probabilités est une théorie logique de l'implication partielle : la probabilité est conçue comme la mesure du degré d'implication partielle d'une proposition par une autre (d'une

⁸¹ Laplace, 1986[1814], *Essai philosophique sur les probabilités*, p. 35.

⁸² Keynes, 1963[1921], *Treatise on Probability*.

conclusion par une évidence donnée) et réside donc dans la relation logique entre un ensemble de propositions constituant les prémisses d'un argument et un ensemble de propositions en constituant sa conclusion. Il s'agit d'une notion de probabilité non pas absolue, mais relative à un certain corpus de connaissances : la valeur de probabilité d'une conclusion, c'est-à-dire le degré d'implication partielle d'une conclusion, est relative à l'évidence considérée. Keynes exprime cette même notion de probabilité en des termes différents mais équivalents : il affirme que la probabilité est la mesure du degré de croyance que tous les êtres humains rationnels devraient avoir, étant donné une certaine évidence, quant à une hypothèse ou à une prédiction donnée⁸³. Nous soulignons que, bien que la théorie logique des probabilités soit ordinairement considérée comme une interprétation épistémique, la notion de probabilité qu'elle invoque est caractérisée par un certain type d'objectivité car elle porte sur la relation d'implication logique entre propositions et sur la croyance partagée car rationnelle quant à la vérité d'hypothèses ou de prédictions.

Troisièmement, la théorie subjective des probabilités, développée indépendamment par Ramsey⁸⁴ en Angleterre et par de Finetti⁸⁵ en Italie, utilise aussi une notion épistémique, et non pas objective, de la probabilité. Elle diffère de la théorie logique qu'elle critique au sens où la probabilité désigne la mesure du degré de croyance, non pas intersubjective, mais d'un individu particulier quant à une hypothèse ou à une prédiction.⁸⁶ Par conséquent, la théorie subjective de la probabilité ne suppose pas que tous les êtres humains rationnels ont le même degré de croyance quant à une proposition donnée à partir d'une même évidence ; au contraire, elle admet des estimations différentes par des individus différents de la valeur d'une même probabilité. Bien qu'apparemment le degré de croyance personnelle soit une notion totalement subjective, selon cette théorie l'assignation de la valeur de probabilité doit respecter certaines conditions, notamment un individu doit être cohérent, c'est-à-dire qu'il doit respecter les axiomes du calcul des probabilités, pour éviter une perte certaine.

⁸³ Keynes, 1963[1921], *op. cit.*, p. 4.

⁸⁴ Ramsey, 1964[1931], « Truth and Probability ».

⁸⁵ De Finetti, 1989[1931], « Probabilism » ; 1933[1931], « On the Subjective Meaning of Probability ».

⁸⁶ De Finetti (aussi bien que Ramsey) identifie la probabilité subjective d'un agent, c'est-à-dire son degré de croyance personnelle, à son comportement de pari : si et seulement si un agent est disposé à parier pour un événement donné, alors on peut parler de la probabilité subjective de cet événement. De cette manière, les situations de pari rendent la notion de probabilité opérationnelle.

2.6.2. Les interprétations objectives des probabilités

Du côté des interprétations objectives des probabilités figurent les trois théories philosophiques suivantes. Premièrement, nous proposons de parler d'une version objective de la théorie classique des probabilités, selon laquelle l'équipossibilité des cas, d'où découle le fait qu'ils ont la même probabilité d'avoir lieu, n'est pas évaluée *a priori* sur la base du principe de raison insuffisante (ou principe d'indifférence), mais à partir de la considération des symétries physiques caractérisant le dispositif physique utilisé pour les produire⁸⁷. Il s'agit de ce qu'on appelle la méthode des fonctions arbitraires qui fut formalisée, entre autres, par Poincaré⁸⁸ dans le cas de dispositifs mécaniques simples et idéalisés (par exemple, la roulette). Nous rappelons que, contrairement à la vision laplacienne, la probabilité estimée sur la base des symétries propres à un dispositif physique n'est pas la mesure du degré d'ignorance partielle au sujet des événements que ce dernier peut produire, mais d'une propriété de ces événements physiques mêmes, plus précisément de leur degré respectif de possibilité étant donné le dispositif physique pouvant les produire. Bien qu'on ait tendance à interpréter cette version de la théorie classique de manière objective, on pourrait opposer qu'il s'agit plutôt d'une interprétation épistémique des probabilités car la probabilité n'est pas absolue, mais relative à un ensemble de symétries physiques que l'on connaît. Nous acceptons une telle objection mais estimons que, si les symétries physiques d'un phénomène sont évaluées de manière très précise, ce qui n'est possible que dans le cas d'un dispositif physique idéalisé, ou si elles sont l'objet d'une connaissance robuste, c'est-à-dire qu'elles se révèlent être des propriétés invariantes sous des descriptions différentes du même phénomène réel, les probabilités évaluées sur la base de telles symétries peuvent être considérées comme des mesures objectives du degré de possibilité de chaque événement possible.

Deuxièmement, la théorie fréquentiste trouva son origine dans la réflexion de Jacques Bernoulli, plus précisément dans la formulation de la loi des grands nombres ; elle fut ensuite développée au milieu du XIX^e siècle par Ellis et par Venn et, au cours de la première moitié du siècle dernier, par von Mises⁸⁹ et par Reichenbach⁹⁰. Elle est habituellement considérée comme une interprétation objective des probabilités, ces dernières étant considérées comme indépendantes de l'individu particulier qui les estime *a posteriori* sur la base des fréquences observées. En dépit des différences entre les résultats des réflexions de ces mathématiciens

⁸⁷ Voir la section 2.1 de ce chapitre.

⁸⁸ Entre autres, voir Poincaré, 1987[1912], *Calcul des probabilités*.

⁸⁹ Von Mises, 1961[1928], *Probability, Statistics and Truth*.

⁹⁰ Reichenbach, 1949, *The Theory of Probability*.

divers⁹¹, de manière générale on peut affirmer que, selon la théorie fréquentiste, la probabilité est dans une relation d'identité avec la fréquence relative, à la limite ou sur le long terme. Plus précisément, la probabilité d'un événement est la valeur vers laquelle converge sa fréquence relative lorsque les mêmes conditions sont répétées de manière indépendante un nombre élevé (qui est idéalement infini) de fois. Par exemple, la probabilité de « pile » lors du lancer d'une pièce non biaisée est égale à la fréquence relative du résultat « pile » dans une longue (idéalement, indéfiniment longue) série de lancers indépendants de la même pièce.

L'un des problèmes majeurs au sein de la théorie fréquentiste concerne l'évaluation et le caractère objectif de la probabilité d'un événement. En effet, d'une part, dans les situations réelles on n'a jamais à faire avec des séries infinies d'événements et ne peut donc calculer la fréquence relative à la limite d'un événement donné. D'autre part, si l'on considère les probabilités comme identiques aux fréquences relatives sur le long terme, il se pose aussi le problème de la longueur appropriée d'une série d'événements uniformes pour qu'on puisse identifier la probabilité d'un événement donné à sa fréquence relative. Un autre défi pour toute version de la théorie fréquentiste des probabilités, en particulier pour l'interprétation objective des probabilités la caractérisant, est ce qu'on appelle « le problème de la classe de référence ». En effet, la fréquence relative d'un événement est toujours relative à une classe de référence, c'est-à-dire au type d'événements auquel on le rattache ; de plus, un même événement peut appartenir à plusieurs classes de références à la fois et avoir donc des fréquences relatives différentes car relatives à des classes de référence différentes. Par conséquent, le problème réside dans l'identification de la classe de référence appropriée pour évaluer de manière objective la fréquence relative, et donc la probabilité, d'un événement donné.⁹²

Troisièmement, la théorie propensionniste de la probabilité fut développée par Popper⁹³ autour des années 1950, en particulier afin de faire face à l'un des inconvénients majeurs, à son avis, de la théorie fréquentiste, qui est l'impossibilité d'attribuer des probabilités objectives à des événements singuliers : Popper visait à fournir une notion de la probabilité appropriée pour caractériser les phénomènes étudiés en mécanique quantique. Lorsque dans le chapitre précédent nous avons introduit la notion de hasard pur en tant

⁹¹ Par exemple, von Mises niait l'existence de la probabilité d'un événement singulier et considérait l'existence des collectifs comme la condition de possibilité pour toute attribution de probabilité, alors que Reichenbach était plus flexible et admettait la possibilité d'attribuer des probabilités à des événements singuliers.

⁹² Pour une analyse détaillée et critique du problème de la classe de référence, nous renvoyons à Hajek (2007, « The Reference Class Problem Is Your Problem Too »), où l'auteur montre qu'il s'agit d'un problème concernant non seulement la théorie fréquentiste, mais toute théorie philosophique des probabilités.

⁹³ Popper, 1959, « The Propensity Interpretation of Probability ».

qu'entité métaphysique intrinsèquement probabiliste dans le monde, nous avons souligné que la notion de propension est confuse et changeante chez Popper : ce dernier parle des propensions pour désigner soit des propriétés dispositionnelles d'ensembles de conditions physiques, soit des entités métaphysiques, soit des possibilités pondérées. En dépit d'une telle diversité, nous avons montré que Popper conçoit toujours les propensions comme quelque chose d'objectif et d'intrinsèquement probabiliste : en tant que propriétés dispositionnelles d'ensembles de conditions physiques, elles correspondent aux probabilités ; en tant qu'entités métaphysiques, elles sont à l'origine causale des probabilités ; en tant que possibilités pondérées, elles sont des grandeurs qui mesurent des possibilités et, dans ce sens, correspondent aux probabilités.

Alors qu'il existe des versions plus ou moins canoniques des autres théories des probabilités, au cours du siècle dernier le propensionnisme a été développé par des philosophes des sciences différents (entre autres, Popper, Miller, Fetzer)⁹⁴ dans des directions différentes. Il est quand même possible de distinguer deux types de propensionnisme, la théorie propensionniste des cas singuliers (« *single case* ») et la théorie propensionniste de long terme (« *long run* »), qui s'opposent au sujet de l'ensemble des conditions physiques déterminant les propensions aussi bien que relativement au type d'événements que les propensions tendent à réaliser. D'une part, selon le propensionnisme des cas singuliers, les propensions sont associées soit à l'état actuel et complet de l'univers, soit à l'ensemble complet des conditions (nomiquement et/ou causalement) pertinentes dans le monde à un instant donné et qui tendent à réaliser des résultats particuliers ; les probabilités sont donc la mesure d'une telle tendance⁹⁵. D'autre part, selon le propensionnisme de long terme, les propensions sont associées à des conditions répétables et tendent à produire, dans de longues séries de répétitions de ces conditions, des fréquences d'événements qui sont approximativement égales aux probabilités de ces derniers⁹⁶. Ces deux théories propensionnistes interprètent de manière objective les probabilités, ces dernières étant considérées comme la mesure des tendances propres à certaines conditions physiques du monde réel à produire des événements singuliers ou à produire des fréquences d'événements. Cependant, ces deux propensionnismes diffèrent quant à leur statut respectif en tant que théories : le propensionnisme des cas singuliers a toute l'apparence d'une théorie métaphysique qui n'est pas susceptible d'être testée empiriquement, alors que le

⁹⁴ Gillies, 2000, *Philosophical theories of probability*, p. 113.

⁹⁵ Popper, 1992[1990], *Un univers de propensions. Deux études sur la causalité et l'évolution* ; Miller, 1994, *Critical Rationalism. A Restatement and Defence* ; Fetzer, 1982, « Probabilistic Explanations » .

⁹⁶ Popper, 1959, *op. cit.* ; Gillies, 2000, *op. cit.*

propensionnisme de long terme, vu qu'il associe les propensions à des conditions répétables, peut être testé de manière empirique par l'observation des fréquences obtenues.

En conclusion de cette section, nous signalons que Gillies⁹⁷ introduit une manière supplémentaire de concevoir la probabilité qu'il place entre une interprétation épistémique et une interprétation objective des probabilités. Plus précisément, il introduit l'expression « probabilité artefactuelle » pour désigner un type de probabilité qui se situe entre l'objectivité au sens fort d'indépendance totale par rapport à toute connaissance humaine et la subjectivité selon laquelle la probabilité représente le degré de croyance d'un individu particulier. Contrairement à la probabilité intersubjective qui mesure le degré de croyance d'un groupe social atteignant un consensus, la probabilité artefactuelle ne porte pas sur la connaissance humaine du monde matériel mais est objective en tant que propriété de ce dernier : elle est artefactuelle au sens où elle est le résultat de l'interaction entre les êtres humains et le monde réel ; cependant, elle peut être testée de manière empirique et, si corroborée, peut être considérée comme objective.⁹⁸

2.6.3. La relation entre hasard et probabilité

Dans notre introduction, nous avons annoncé que les probabilités semblent entretenir une relation privilégiée avec la notion de hasard. Nous avons avancé l'hypothèse selon laquelle les probabilités pourraient être l'expression mathématico-formelle de la notion de hasard, quelle que soit la manière dont cette dernière est conçue, et comme un indice fiable de l'utilisation de la notion de hasard dans la caractérisation du phénomène considéré. De plus, nous avons remarqué que, si cette hypothèse se révélait être correcte, il serait intéressant de comprendre s'il est possible de déterminer, d'une part, quelle notion de hasard doit être utilisée pour caractériser des phénomènes physiques ou biologiques donnés à partir de la manière appropriée d'interpréter les probabilités quant à ces derniers et, inversement, comment on doit interpréter les probabilités attribuées à des phénomènes physiques ou biologiques donnés sur la base de la notion de hasard appropriée pour les caractériser. Enfin,

⁹⁷ Gillies, 2000, *op. cit.*, chapitre 8.

⁹⁸ Nous sommes particulièrement intéressée à une telle interprétation de probabilités car elle n'est ni absolument objective, ni purement épistémique. Au contraire, elle essaie rendre compte de la pratique scientifique, en particulier du fait que toute connaissance humaine des phénomènes réels est partielle et approximative. C'est pourquoi elle nous semble plus appropriée pour interpréter les probabilités qui sont utilisées en science. En ce qui concerne notre réflexion sur l'interprétation appropriée des probabilités dans les descriptions formelles (mathématiques) des sources causales de la variation biologique, nous renvoyons au chapitre 10.

dans le présent chapitre, nous avons d'abord identifié une notion de hasard qui est définie en termes purement probabilistes : c'est la notion d'équiprobabilité⁹⁹. Puis, nous avons introduit et analysé d'autres notions mathématiques de hasard qui ont toutes un lien, plus ou moins explicite, avec les probabilités et, plus particulièrement, avec la notion d'équiprobabilité à laquelle elles sont, en partie, équivalentes.

Dans cette sous-section, nous allons suggérer une réponse possible à la question philosophiquement très dense de la relation entre la notion de hasard et les probabilités, plus précisément, entre leur signification et leur rôle respectifs. La solution de ce problème que nous proposons ci-dessous correspond à une hypothèse directrice pour notre travail de recherche, au cours duquel nous serons confrontée à la fois à des descriptions verbales de divers phénomènes biologiques, où la notion de hasard est largement évoquée, et à des modèles mathématiques de ces mêmes phénomènes, où les probabilités occupent une place non négligeable.

Sur la base de la distinction des deux conceptions majeures du hasard (subjectif et objectif) et de sa correspondance avec la distinction de deux groupes d'interprétations (épistémiques et objectives) des probabilités, nous avançons l'hypothèse d'une concordance entre la notion de hasard, subjectif ou objectif, et l'interprétation des probabilités, épistémique ou objective, lorsqu'elles concernent un même phénomène. De manière plus explicite, la caractérisation d'un événement par une notion de hasard subjectif pourrait appeler à une interprétation épistémique des probabilités qui lui sont attribuées, alors que sa caractérisation par une notion de hasard objectif pourrait appeler à une interprétation objective des probabilités le concernant.

En ce qui concerne, plus spécifiquement, la signification des notions de hasard et de probabilité, nous suggérons que quelques-unes des notions de hasard que nous avons analysées dans le premier chapitre semblent recouper certaines théories philosophiques des probabilités, c'est-à-dire la signification qui est attribuée à la notion de probabilité au sein de ces dernières. Pour en donner quelques exemples, considérons la notion de hasard subjectif en tant qu'ignorance des causes : elle semble trouver son correspondant, en termes de signification, au sein de la théorie classique des probabilités développée par Laplace où les probabilités, interprétées de manière épistémique, sont définies comme la mesure du degré d'ignorance partielle quant aux événements. En ce qui concerne la notion de hasard pur en tant qu'entité métaphysique intrinsèquement probabiliste dans le monde, elle trouve de

⁹⁹ Voir la section 2.1 de ce chapitre.

manière assez évidente son correspondant, en termes de signification, dans les théories propensionnistes où les probabilités sont conçues comme la mesure de propriétés dispositionnelles et objectives du monde, indépendantes de toute connaissance humaine. Quant à la notion de hasard en tant que rencontre de séries causales indépendantes, il est intéressant de souligner que Cournot¹⁰⁰ l'introduit dans le cadre de son ouvrage sur le calcul des probabilités au moment où il argumente en faveur d'une théorie fréquentiste des probabilités. Sur cette base, nous suggérons l'hypothèse selon laquelle, pour Cournot, les probabilités renvoient aux fréquences relatives de rencontres de séries causales indépendantes.

En conclusion, une telle correspondance entre hasard (subjectif ou objectif) et interprétation (épistémique ou objective) des probabilités et, plus spécifiquement, entre les significations respectives des notions de hasard et de probabilité, suggère que, au sein de toute théorie philosophique, la probabilité est conçue comme la mesure d'un certain hasard, ce dernier pouvant être compris selon des notions différentes, c'est-à-dire qui sont caractérisées par des significations et par des rôles différents.

Dans ces deux premiers chapitres, nous n'avons pas traité explicitement le rôle des notions de hasard et de probabilité qui, comme nous l'avons déjà souligné, dépend directement de leurs significations respectives. À ce propos, nous avançons l'hypothèse selon laquelle les probabilités ont le rôle de mesurer le degré de hasard des événements probabilisés, indépendamment du fait qu'il s'agisse d'un hasard subjectif dû aux bornes de notre connaissance ou d'un hasard objectif inhérent aux phénomènes réels. De plus, nous suggérons que, selon le rôle de la notion de hasard dont les probabilités sont une mesure, ces dernières peuvent avoir un rôle simplement descriptif ou un rôle explicatif. Elles jouent ce rôle relativement à certaines caractéristiques de nos connaissances ou à certaines propriétés objectives des phénomènes réels, selon que la notion utilisée pour caractériser ces derniers désigne un hasard subjectif ou objectif. En effet, les diverses notions de hasard peuvent assumer un rôle descriptif ou un rôle explicatif : cela semble dépendre du cadre dans lequel elles sont mobilisées. Nous suggérons que, d'une part, si l'on vise simplement à identifier les caractéristiques d'un événement donné, on utilisera plutôt une notion de hasard définie en termes de résultat (par exemple, les notions d'équiprobabilité et de stabilité des fréquences à la limite), qui fournit une description des propriétés de ce qui est considéré comme le fruit du hasard. Si, d'autre part, on veut expliquer pourquoi un événement est le fruit du hasard, on

¹⁰⁰ Cournot, 1984[1843], *Exposition de la théorie des chances et des probabilités*.

utilisera plutôt une notion définie en termes de processus (par exemple, les notions d'absence d'un dessein, de rencontre de séries causales indépendantes, de sensibilité aux conditions initiales ou la notion d'échantillonnage indiscriminé) qui, de cette manière, joue un rôle explicatif relativement aux caractéristiques du résultat considéré.

Dans cette première partie, nous avons tracé les grandes lignes de l'histoire de la notion de hasard en philosophie et en science. Tout en essayant de fournir une introduction critique et satisfaisante au questionnement philosophique autour de la notion de hasard, la ligne directrice que nous avons essayé de poursuivre reflète les buts suivants. D'abord, nous avons voulu donner forme au cadre philosophique et scientifique général où s'insère notre réflexion, plus restreinte et spécifique, sur la notion de hasard en biologie.

Deuxièmement, nous avons visé à mettre en évidence les questions principales qu'on soulève depuis l'Antiquité au sujet de la notion de hasard. Notamment, depuis longtemps, on cherche à savoir si cette notion doit être définie en termes subjectifs ou objectifs, à identifier sa signification précise et son rôle dans le contexte où elle est utilisée, et à caractériser les relations de compatibilité ou d'incompatibilité qu'elle entretient avec des notions diverses, qui posent aussi problème, comme le déterminisme, l'indéterminisme et la probabilité.

Troisièmement, nous avons conçu cette première partie comme le lieu pour pouvoir introduire et clarifier un certain nombre de notions et de distinctions qui nous serviront par la suite, au cours de notre recherche sur la notion de hasard en biologie, en tant qu'outils conceptuels. Notamment, nous avons présenté de manière critique la distinction entre hasard subjectif et hasard objectif, entre des formes diverses de déterminisme et d'indéterminisme, et entre les interprétations épistémiques et objectives des probabilités. De plus, nous avons souligné l'importance de ces distinctions afin de comprendre les relations entre ces notions différentes. Une telle opération de clarification nous permettra de mieux expliciter les résultats de notre recherche et de les placer de manière appropriée au sein de la réflexion plus générale, philosophique et scientifique, sur la notion de hasard.

Enfin, cette première partie nous a permis de faire la part entre ce dont nous parlerons par la suite et ce dont nous ne parlerons pas, et surtout de justifier ce choix. Par exemple, bien qu'on ait consacré un chapitre assez important à ce que nous avons appelé les notions mathématiques de hasard, nous avons conclu que certaines d'entre elles, notamment la notion de hasard en tant qu'incompressibilité d'une suite, ne sont pas appropriées pour notre propos. De manière précise, nous avons avancé quelques arguments afin de montrer que ces notions, qui ont d'ailleurs été l'objet de réflexions et de définitions très techniques au cours du XX^e

siècle, ne fourniraient pas une caractérisation particulièrement intéressante des phénomènes à l'origine causale de la variation biologique. D'une part, en effet, il s'agit de notions de hasard qui sont définies en termes de résultat, qui se limitent à fournir une simple description du phénomène qu'elles caractérisent, et qui n'ont donc aucun pouvoir explicatif. D'autre part, elles ne peuvent être appliquées de manière pertinente que si l'on a une bonne connaissance du phénomène considéré, auquel ces notions mathématiques de hasard s'appliquent.

Nous avons donc posé les bases pour pouvoir nous engager dans une analyse critique de la signification et du rôle de la notion de hasard lorsqu'elle est utilisée pour caractériser les phénomènes qui sont à l'origine causale de la variation biologique. Nous consacrerons la deuxième partie de notre travail aux mutations génétiques, sources causales de la variation génétique et phénotypique, et la troisième partie au bruit affectant le processus d'expression génétique qui, de cette manière, est une source de variation phénotypique.

Partie II – Hasard et mutations génétiques

Les mutations génétiques représentent une source causale importante de variation génétique et phénotypique au sein des populations naturelles. Depuis la découverte de la structure physico-chimique du matériel génétique dans les années 1950¹, une mutation est définie, au sens large, comme une altération de la séquence nucléotidique de l'ADN². Plus précisément, elle correspond à toute modification du matériel génétique exception faite des changements dus au processus de recombinaison³. La notion actuelle de mutation est le résultat d'une longue évolution conceptuelle qui a eu lieu durant la période qui va de la fin du XIX^e siècle jusqu'au tournant moléculaire en biologie (1953). Notamment, nous verrons que, autour du début du siècle dernier, les premiers généticiens mendéliens utilisaient le terme « mutation » pour désigner quelque chose de très différent d'une modification du matériel génétique au sein d'un organisme individuel. Par conséquent, le rôle que cette notion jouait dans l'explication de l'évolution naturelle était très éloigné de la place occupée par la notion de mutation au sein de la Théorie Synthétique Moderne et par la notion actuelle de mutation, définie en termes moléculaires, au sein de la théorie moderne de l'évolution.

Dans notre introduction générale, nous avons souligné que la variation, dont les mutations génétiques sont une source, représente l'une des conditions de possibilité pour qu'une population évolue via le processus de sélection. En effet, le squelette logique de la théorie darwinienne de l'évolution comprend les trois principes suivants. 1) Les individus différents d'une population ont des morphologies, des physiologies et des comportements différents (variation phénotypique). 2) Des phénotypes différents ont des taux différents de survie et de reproduction dans des environnements différents (fitness différentielle). 3) Il y a une corrélation entre les parents et les descendants en ce qui concerne leur contribution respective à la génération future (la fitness est héritable).⁴ Donc, pour qu'une mutation génétique – qui est une source causale de la variation génétique et, même si ce n'est pas de manière nécessaire, de la variation phénotypique – puisse avoir des conséquences au niveau

¹ Watson & Crick, 1953, « A Structure for Deoxyribose Nucleic Acid ». Voir Morange, 2003[1994], *Histoire de la biologie moléculaire*.

² L'acide désoxyribonucléique (ADN) est le matériel génétique de la plupart des organismes vivants (certains virus utilisent un acide alternatif, l'acide ribonucléique) et, avec d'autres facteurs non génétiques, est transmis d'une génération à la suivante. Il consiste en deux chaînes polynucléotidiques (les brins) torsadées de façon à former une double hélice et orientées de manière anti-parallèle. Chaque brin est constitué par un squelette sucre-phosphoryle et des bases azotées disposées linéairement (la guanine et l'adénine, qui sont des bases puriques ; la cytosine et la thymine, qui sont des bases pyrimidiques). Les bases adjacentes appartenant au même brin sont associées par des liaisons hydrophobes qui assurent la stabilité de la double hélice. Les bases des brins opposés sont appariées de manière spécifique et complémentaire (A s'apparie à T et C s'apparie à G) par des liaisons hydrogènes plus faibles : cette complémentarité dérive de leurs caractéristiques physico-chimiques respectives.

³ La recombinaison génétique consiste en l'échange de fragments de matériel génétique entre deux brins d'ADN appariés. Le processus de recombinaison n'introduit aucune nouveauté dans la séquence d'ADN d'un organisme donné, mais se limite à la remanier et à produire des combinaisons nouvelles du matériel génétique.

⁴ Lewontin, 1970, « The units of selection ».

de l'évolution de la population à laquelle les organismes individuels porteurs appartiennent, elle doit pouvoir être transmise d'une génération à la suivante. En particulier, chez les organismes multicellulaires caractérisés par deux lignes cellulaires distinctes – l'une germinale, l'autre somatique – seules les mutations génétiques affectant le matériel génétique des cellules germinales ou reproductrices, c'est-à-dire susceptibles d'être impliquées dans la génération des descendants, peuvent être transmises au cours des générations et sont donc potentiellement importantes du point de vue de l'évolution⁵. Au contraire, chez les organismes qui ne sont pas caractérisés par cette distinction entre soma et germen, par exemple chez les procaryotes et chez plusieurs eucaryotes unicellulaires⁶, toute mutation génétique peut être transmise à la génération suivante et est donc susceptible de jouer un rôle dans l'évolution de la population des organismes concernés.

Selon la vision consensuelle au sein de la communauté des biologistes depuis les années de la Synthèse Moderne (1930-1950), les mutations génétiques sont des événements « aléatoires » ou qui sont le « fruit du hasard ». Parfois, cette même idée, qui vit le jour pendant la deuxième moitié du XIX^e siècle dans les réflexions de Darwin sur les causes de la variation au sein des populations naturelles, est exprimée en utilisant des termes différents : les mutations sont considérées comme des sources « aveugles », « spontanées », « accidentelles », « fortuites », « probabilistes », « stochastiques » et « imprédictibles » de la variation.

À première vue, la notion de hasard à laquelle les termes ci-dessus semblent tous renvoyer donne l'impression d'être intuitivement facile à comprendre. Cependant, c'est une notion qui, à la fois de manière générale et, plus particulièrement, dans un cadre scientifique, peut évoquer des idées très différentes, même si elles ne correspondent pas à des notions de hasard dont la signification et le rôle respectifs sont nécessairement incompatibles. Par exemple, le fait de parler de « hasard » relativement à un événement donné peut renvoyer à l'ignorance des causes sous-jacentes à la production de cet événement ; à l'absence d'un plan à son origine ; au fait que l'événement considéré est dû à la rencontre de deux ou plusieurs séries causales indépendantes ; à l'instabilité propre au système qui l'engendre ; à l'absence de régularité, statistiquement testable, dans l'occurrence de cet événement ; ou encore, à sa

⁵ Par exemple, chez les humains, une mutation génétique pouvant être transmise à la génération suivante est une mutation qui se produit au sein des gamètes féminins et/ou masculins, c'est-à-dire dans les cellules paternelles de sperme et/ou dans les œufs maternels. Une mutation génétique non héréditaire concerne, au contraire, le matériel génétique des cellules somatiques du corps des individus et ne peut donc pas être transmise à leurs descendants.

⁶ Chez certains eucaryotes unicellulaires, on retrouve la distinction entre les cellules somatiques et les cellules reproductives. Par exemple, chez le *Paramecium* (la paramécie), appartenant au phylum des Ciliés, on retrouve deux noyaux situés à l'intérieur d'une même unité cytoplasmique, dont l'un est germinale et l'autre est somatique.

complexité computationnelle et à son imprédictibilité dans ce sens. Ces manières diverses de concevoir le hasard correspondent à quelques-unes des déclinaisons que cette notion a reçues au cours de l'histoire de la pensée philosophique et scientifique, que nous venons de présenter dans la première partie de ce travail. À une telle multiplicité de formes conceptuelles de la notion de hasard s'ajoute le fait que la question de sa signification et de son rôle, de manière générale et en particulier dans un cadre scientifique précis, est encore un objet de controverse : elle a reçu des réponses assez satisfaisantes en termes mathématiques⁷ qui, cependant, méritent d'être analysées ultérieurement, notamment quant à leur pertinence dans chaque discipline scientifique. La notion de hasard invoquée dans la caractérisation des mutations génétiques n'est donc pas aussi facile à identifier et à définir qu'il ne semble au premier abord.

Un élément parlant quant au problème de l'utilisation de la notion de hasard dans la caractérisation des mutations génétiques et quant à la difficulté, à la fois conceptuelle et empirique, de sa définition, réside dans la longue histoire du débat autour de ces questions. En effet, elles sont depuis longtemps et sont encore aujourd'hui au centre d'une controverse au sein des communautés des biologistes, des historiens et des philosophes de la biologie. Dans cette deuxième partie, nous tracerons les grandes lignes de l'histoire de ce débat, dès la publication de *L'Origine des espèces* (1859) et l'opposition à laquelle Darwin et ses partisans durent faire face à cheval entre le XIX^e et le XX^e siècle, en particulier de la part des biologistes se disant « lamarckiens », jusqu'à ses éclats les plus récents autour des années 1990.

Le but de cette deuxième partie de notre travail est multiple. D'abord, nous visons à clarifier dans quel sens Darwin considérait que la variation individuelle était « spontanée », « accidentelle » ou le « fruit du hasard ». Ensuite, nous nous proposons d'identifier et de définir la notion de hasard que les biologistes de la Synthèse Moderne utilisaient pour caractériser cette source causale de la variation que sont les mutations génétiques. Enfin, nous visons à déterminer s'il s'agit d'une notion de hasard dont la signification et le rôle sont appropriés, à la fois du point de vue conceptuel et empirique, pour caractériser les mutations génétiques dans le cadre de la théorie moderne de l'évolution. Une telle analyse conceptuelle et historique nous permettra d'évaluer le bien fondé des critiques actuelles, que nous présenterons et avec lesquelles nous nous confronterons longuement au cours de cette partie, au caractère « aléatoire » de toute mutation génétique.⁸

⁷ Voir le chapitre 2.

⁸ Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic*

Nous verrons que la notion de hasard qui est au centre du débat sur le caractère des mutations génétiques depuis plus d'un siècle concerne la relation entre mutation, sélection et adaptation. C'est pourquoi, une fois définie, nous la dénommerons « hasard évolutionnaire ». Bien que la plupart de notre analyse soit focalisée sur cette notion car elle a posé et semble encore poser problème pour un certain nombre de biologistes et de philosophes de la biologie, nous nous occuperons aussi de la question de savoir dans quel sens les mutations génétiques sont « aléatoires » du point de vue moléculaire, c'est-à-dire relativement aux processus physico-chimiques qui les engendrent et non pas en termes adaptatifs. Nous aborderons cette question afin de fournir, par ce travail, une clarification exhaustive de la notion de hasard lorsqu'elle est utilisée pour caractériser les mutations génétiques.

Avant d'entrer dans le vif de cette deuxième partie, nous précisons que nous concentrerons notre attention sur les études expérimentales réalisées avec des bactéries, des années 1940 jusqu'à aujourd'hui. Nous ne serons donc pas confrontée directement au problème de savoir si une mutation affecte le matériel génétique d'une cellule germinale ou d'une cellule somatique, autrement dit, si une mutation génétique peut ou ne peut pas être transmise à la génération suivante au moment de la reproduction de l'organisme porteur. En effet, nous rappelons que cette question ne se pose que relativement aux organismes dont les cellules reproductives sont distinctes de celles du corps (par exemple, chez les animaux et chez les végétaux). Nous justifions le choix de focaliser notre enquête sur l'étude des mutations génétiques chez ces microorganismes par les deux raisons suivantes. En premier lieu, les développements importants que la recherche au sujet des mécanismes de mutation a connus en microbiologie au cours des trente dernières années nous fournissent une quantité considérable de résultats expérimentaux et de réflexions théoriques sur lesquels baser notre analyse. En second lieu, l'analyse de situations expérimentales où toute mutation génétique est susceptible d'être transmise à la génération suivante facilite la tâche d'une analyse critique de la vision lamarckienne de l'origine et de la transmission de la variation, plus précisément l'idée de l'hérédité des caractères acquis, par rapport à la vision consensuelle propre à la Théorie Synthétique Moderne de l'Evolution.

Voici la manière dont se déploiera notre parcours argumentatif. Dans le chapitre 3, nous tracerons d'abord l'histoire de la notion de mutation et des idées diverses sur l'origine et sur le caractère de la variation biologique, dès le XIX^e siècle jusqu'à aujourd'hui. Ensuite,

Variation in the History of Life ; Shapiro, 1995a, « Adaptive Mutation: Who's Really in the Garden? » ; Shapiro, 2005, « A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering » ; Sternberg, 2002, « On the Roles of DNA Repetitive Elements in the Context of a Unified Genomic-Epigenetic System » ; Fox Keller, 2003[2000], *Le siècle du gène*.

nous procéderons à une introduction au phénomène moléculaire de mutation et aux méthodes théoriques et expérimentales utilisées dans le cadre de son étude.

Dans le chapitre 4, nous examinerons la manière dont les acteurs majeurs de la Synthèse Moderne ont caractérisé les mutations génétiques et, sur cette base, nous identifierons la notion de hasard qui faisait, et qui fait encore, consensus : c'est ce que nous appellerons « hasard évolutionnaire ».

Le chapitre 5 présentera les avancées des recherches sur les mutations génétiques qui ont eu lieu au cours des trente dernières années. Notamment, nous procéderons à une présentation de quelques-uns des mécanismes de mutation – aussi dits « mécanismes mutateurs » – qui ont suscité un grand débat, à la fin des années 1980 et au cours des années 1990, autour du caractère « aléatoire » ou pas des mutations génétiques. La découverte de ces mécanismes motive encore un certain nombre de biologistes, d'historiens et de philosophes de la biologie à parler d'un défi majeur pour la Synthèse Moderne.

Dans le chapitre 6, nous nous confronterons à l'une des critiques les plus provocatrices de l'idée de la Synthèse Moderne, selon laquelle toute mutation génétique est le fruit du hasard du point de vue évolutionnaire.⁹ Nous montrerons que, contrairement à ce que les biologistes Jablonka et Lamb affirment dans leur dernier livre (2005), les découvertes récentes quant aux mécanismes de mutation ne mettent pas en cause le caractère approprié de la notion de hasard qui est invoquée au sein de la Théorie Synthétique Moderne pour caractériser les mutations génétiques.

Enfin, dans le chapitre 7, nous nous occuperons du caractère « aléatoire » des mutations génétiques du point de vue moléculaire et non pas en termes adaptatifs. À la lumière des notions de hasard qui ont vu le jour au cours de l'histoire de la pensée philosophique et scientifique¹⁰, nous concluons cette deuxième partie par une analyse de la signification et du rôle des deux notions de hasard – le « hasard évolutionnaire » et l'« aléatoire (au sens) faible » – que nous aurons identifiées en tant qu'appropriées, à la fois du point de vue conceptuel et empirique, pour caractériser les mutations génétiques.

⁹ Jablonka & Lamb, 2005, *op. cit.*

¹⁰ Voir les chapitres 1 et 2.

Chapitre 3 – Les mutations génétiques : une introduction

3.1. Les conceptions de l'origine de la variation biologique (XIX^e-XXI^e siècles)

3.1.1. Brève histoire de la notion de « mutation »

Retraçons l'histoire de la notion de « mutation » de la fin du XIX^e siècle jusqu'à nos jours. La reconstruction que nous proposons ici ne prétend pas être exhaustive, mais est plutôt une ébauche des grandes lignes historiques à ce propos : elle n'épuise donc pas toutes les notions de « mutation » qui ont vu le jour au cours du XIX^e et du XX^e siècle.¹

La définition des mutations génétiques en termes moléculaires est relativement récente : elle fut formulée autour des années 1950 sur la base de la découverte de la nature et de la structure physico-chimique du matériel génétique. Avant l'époque du tournant moléculaire en biologie, plus précisément à partir de la moitié du XIX^e siècle, la notion de « mutation » avait connu une longue évolution. En 1869,² le paléontologue allemand Waagen avait parlé de « mutation » en référence aux différentes formes de fossiles appartenant à une même lignée et étant apparues à des moments successifs dans le temps. Il utilisait donc ce terme pour désigner les plus petits changements perceptibles d'une forme à une autre au cours du temps. Ensuite, à cheval entre le XIX^e et le XX^e siècles, les premiers mendéliens – parmi lesquels figurent le botaniste hollandais de Vries et le zoologiste anglais Bateson – avaient adhéré à la pensée typologique, n'avaient attribué aucun rôle constructif à la sélection naturelle dans l'origine de l'adaptation et avaient conféré à la notion de mutation une signification nouvelle et différente de celle qu'on lui attribua plus tard en génétique et,

¹ Pour l'histoire de la notion de mutation au XIX^e siècle, voir Stubbe, 1972[1963], *History of Genetics. From Prehistoric Times to the Rediscovery of Mendel's Laws*. Voir aussi Dunn, 1965, *Short History of Genetics. The development of some of the main lines of thought: 1864-1939*, en particulier pour l'histoire de la notion de mutation entre la fin du XIX^e siècle et les quarante premières années du XX^e.

² La notion de mutation en tant que phénomène à l'origine de la variation des unités héréditaires figure déjà, en 1751, dans le *Système de la nature* de Maupertuis où ce dernier affirme que la mutation peut empêcher aux particules élémentaires (héréditaires) de retenir l'ordre qu'elles possédaient chez les animaux père et mère. Plus précisément, il écrit que chaque degré d'erreur pourrait produire une nouvelle espèce et que ce genre de déviations répétées pourrait être à l'origine de la diversité des animaux qu'on observe aujourd'hui. Nous verrons que l'idée avancée par Maupertuis se rapproche beaucoup de la notion de mutation propre à de Vries en tant qu'événement à l'origine d'une nouvelle espèce.

ensuite, en biologie moléculaire. Ils distinguaient la variation individuelle, continue et de petite ampleur de la variation au niveau de l'espèce, discontinue et de grande ampleur : selon de Vries et les autres mendéliens de l'époque, le premier type de variation n'avait pas de valeur du point de vue évolutionnaire alors que le deuxième, contrairement à la vision gradualiste de Darwin, pouvait provoquer l'apparition soudaine et sans transition d'une nouvelle espèce. Ils parlaient donc de « mutation » pour désigner une transformation majeure, spontanée, soudaine et même brutale, donnant lieu à la naissance d'une nouvelle espèce – autrement dit, d'un nouveau type – sans que ni la sélection naturelle ni l'environnement n'aient joué aucun rôle. Il s'agissait d'une vision « saltationniste » de l'évolution que de Vries formula et appela « Théorie Mutationniste »³.

Quelques années plus tard, les expériences réalisées par le généticien Morgan et son équipe avec la *Drosophila melanogaster* fournirent les bases de la notion moderne de « mutation »⁴. Morgan avait nié l'idée des premiers mendéliens selon laquelle les mutations étaient des événements majeurs – des « *changements presto* »⁵ – provoquant nécessairement l'apparition d'espèces nouvelles. Il avait affirmé que les mutations consistaient plutôt en des altérations plus ou moins importantes et spontanées – c'est-à-dire dues à des agents naturels inconnus⁶ – du matériel héréditaire (des gènes)⁷ dont la nature et la structure n'avaient pas encore été découvertes : ces altérations génétiques étaient à l'origine de l'apparition de nouveaux allèles, c'est-à-dire, de versions alternatives d'un même gène⁸. Même si Morgan avait relativisé le rôle des mutations génétiques dans le processus macroévolutif, tout comme de Vries et les autres premiers mendéliens il n'avait attribué à la sélection naturelle aucun rôle constructif dans le processus évolutif, en particulier dans l'origine de l'adaptation.

³ De Vries, 1901[1910], *The mutation theory. Experiments and observations of the origin of species in the vegetable kingdom*. De Vries proposa sa « théorie mutationniste » en ignorant complètement les causes et les types de changements génétiques à l'origine de ce qu'il appelait « mutation », autrement dit l'apparition d'une nouvelle espèce. Par la suite, il a été démontré que les variations que de Vries observa chez les plantes d'*Oenothera lamarckiana* et qu'il appela « mutations » correspondaient à des réarrangements chromosomiques et à des phénomènes de polyploïdie (voir Morgan, 1988[1926], *The Theory of The Gene*).

⁴ Morgan, 1988[1925], *The Genetics of Drosophila* ; 1988[1926], *op. cit.* Voir Dunn, 1965, *op. cit.*

⁵ Morgan, 1988[1926], *op. cit.*, p. 68.

⁶ Morgan (*ibid.*, p. 67) affirmait que : « [...] nous n'avons presque pas d'évidence quant à la vraie nature du processus de mutation » (notre traduction), « [...] *we have almost no evidence as to the real nature of the mutation process.* »

⁷ Morgan (1988[1925], *op. cit.*, p. 22) définissait une mutation comme « [...] un changement transmissible du plasma germinatif produisant un changement détectable dans les caractéristiques de l'individu » (notre traduction), « [...] *a transmissible change in the germplasm, producing a detectable change in the characteristics of the individual.* »

⁸ Le terme gène fut inventé par Johannsen en 1909. Ce dernier introduit aussi les termes « génotype » et « phénotype » en 1911.

C'est à Muller, un généticien de l'équipe de Morgan, qu'on doit attribuer la définition moderne de « mutation » comme altération d'un gène individuel. Il arriva à cette conclusion sur la base des expériences qu'il réalisa dans les années 1920 et 1930 avec la drosophile : il montra que l'exposition de ces organismes à des quantités variables de rayons X augmente de manière proportionnelle la fréquence des mutations génétiques⁹. Il est important de rappeler que, jusqu'au début de la deuxième moitié du XX^e siècle, la structure physico-chimique des gènes était restée inconnue. Autour des années 1920, il n'y avait aucune position consensuelle, parmi les généticiens, à propos de la nature des gènes. Cependant, de nombreux collègues de Morgan admettaient que les gènes étaient des corpuscules discrets situés en ordre linéaire et dans des positions précises sur les chromosomes¹⁰ : cette notion trouva son origine dans les arguments avancés par Morgan dans *The Theory of the Gene* (1988[1926]) et fut validée par les recherches et les méthodes de la cytologie¹¹. De plus, les résultats des expériences réalisées par Muller montrant qu'un changement du matériel génétique peut être produit par l'impact de rayons X (et aussi de rayons gamma, béta, cathodiques, ultraviolets) encouragea l'idée selon laquelle le gène est un petit corpuscule matériel.¹²

Au moment de la fondation de la Synthèse Moderne autour des années 1930, les biologistes avaient reconnu l'importance, d'une part, des processus de mutation et de recombinaison en tant que source de la variation au sein des populations naturelles et, d'autre part, de la sélection naturelle, dans le processus évolutif. L'idée proposée par l'équipe de Morgan, et soutenue en particulier par les travaux de Muller, selon laquelle une mutation est une altération d'un gène individuel commençait à devenir l'objet d'un consensus. De plus, les généticiens avaient réalisé que les mutations étaient plutôt de petites modifications des gènes pouvant produire, à la fois, des effets phénotypiques petits et continus ou grands et discrets.

⁹ En 1946, Muller reçut le prix Nobel pour avoir découvert que les mutations génétiques peuvent être induites par des rayons X.

¹⁰ L'histoire de la notion de gène n'est pas si simple et le débat autour de sa définition est encore ouvert aujourd'hui. Entre les années 1910 et les années 1960, la notion corpusculaire du gène était relativement répandue au sein de la communauté des biologistes, bien que certains (par exemple, Goldschmidt) imaginaient les gènes comme des portions bien définies des chromosomes, ces derniers étant des champs moléculaires, ou comme une substance continue et diffuse (voir Mayr, 1989[1982], *Histoire de la biologie. De Darwin à nos jours*). Nous renvoyons à Fox Keller (2003[2000], *Le siècle du gène*) pour l'histoire de la notion de gène et des débats à son sujet au cours du XX^e siècle. Pour une histoire précise et synthétique de la génétique classique, voir Dunn, 1965, *Short History of Genetics. The development of some of the main lines of thought: 1864-1939*.

¹¹ *Ibid.*

¹² Sur la base d'un certain nombre de données sur la fréquence des mutations et sur l'entrecroisement chromosomique, Muller estima la taille approximative d'un gène et le nombre total de gènes chez la drosophile (entre 1400 et 1800). Voir Mayr, 1989[1982], *op. cit.*.

Entre les années 1930 et 1950, les physiciens jouèrent un rôle important dans l'étude des phénomènes biologiques, en particulier au niveau moléculaire¹³. Ces chercheurs, parmi lesquels figurent entre autres les cristallographes Perutz, Wilkins et Astbury et les physiciens Bohr et Schrödinger, furent à l'origine d'un transfert des nouveaux concepts de la physique vers la biologie et, de cette manière, contribuèrent de manière décisive à la naissance d'une nouvelle discipline biologique, la biologie moléculaire (1953). C'est dans le cadre des recherches biologiques réalisées par ces physiciens qu'une notion de « mutation » en termes quantiques vit le jour et se répandit.

Dans son petit ouvrage intitulé *Qu'est-ce que la vie ?* (1986[1944]), Schrödinger présente les expériences du généticien Muller sur l'induction des mutations par les rayons X ; ensuite, il explique de manière très accessible une conception quantique de la structure du gène capable de rendre compte de sa stabilité ; enfin, il compare ce modèle, dû à Delbrück¹⁴, avec les données génétiques. Schrödinger affirme que le gène est une grande molécule (ce qu'il appelle aussi un « cristal ou solide aperiodique ») dont la structure repose sur une configuration moléculaire relativement stable caractérisée par un certain niveau énergétique. Sur cette base, il ajoute qu'une mutation est un événement brusque et rare provoquant un réarrangement des atomes constituant le gène et venant donc perturber la stabilité de ce dernier : elle est due à un saut quantique, c'est-à-dire à une transition entre deux états énergétiques correspondant à deux configurations moléculaires relativement stables d'un gène. Cette notion quantique de « mutation » trouva un certain consensus entre les années 1930 et 1950, en particulier chez les physiciens focalisant leur recherche sur le monde biologique. Cependant, le modèle quantique de la structure du gène échoua et, avec lui, la proposition de Schrödinger.

Enfin, c'est grâce à la découverte que les gènes sont formés d'ADN et à la description de la structure physico-chimique des molécules de cet acide nucléique autour des années 1950 qu'une mutation génétique a été définie comme toute altération de la séquence nucléotidique de l'ADN, exception faite pour les changements dus au processus de recombinaison. C'est à cette époque que les biologistes ont commencé à parler du « code¹⁵ génétique » pour désigner la correspondance entre les 64 triplets de nucléotides possibles et les 20 acides aminés. Plus

¹³ Voir Morange, 2003[1994], *Histoire de la biologie moléculaire* ; Fox Keller, 1990, « Physics and the emergence of molecular biology: A history of cognitive and political synergy ».

¹⁴ Pour une présentation de la conception du gène de Delbrück, voir Fischer & Lipson, 1988, *Thinking about Science: Max Delbrück and the Origins of Molecular Biology*.

¹⁵ Le mot « code » figure déjà dans le petit ouvrage de Schrödinger (1986[1944], *op. cit.*) pour désigner l'hypothèse selon laquelle « les chromosomes [...] contiennent sous la forme d'une espèce de code le modèle intégral du développement futur de l'individu et de son fonctionnement dans l'état adulte. » (p. 71).

précisément, c'est la succession des bases nucléotidiques le long de la séquence d'ADN qui est considérée comme le « code » pour les différents acides aminés formant la chaîne polypeptidique des protéines. Les biologistes ont donc commencé à concevoir les mutations, non pas simplement comme des altérations des gènes, mais comme la corruption de l'information ou des instructions dont ces derniers sont porteurs, c'est-à-dire de la spécification de la séquence d'acides aminés des protéines.¹⁶

La notion informationnelle de « programme génétique » a aussi été introduite à cette époque, suite à la découverte des mécanismes génétiques de régulation¹⁷. Elle a été appliquée au génome pour signifier, plus explicitement, qu'il contient les informations nécessaires au contrôle du processus de développement.¹⁸ Par conséquent, les biologistes ont aussi commencé à concevoir les mutations génétiques comme des modifications susceptibles d'affecter et de corrompre la structure fonctionnelle des protéines dont l'ADN porte, codée, l'information et, plus généralement, le développement dans sa totalité. Bien que la vision déterministe propre à la biologie moléculaire – qui l'a particulièrement marquée tout au début de son développement – ait été désormais remise en question¹⁹, l'idée qu'une mutation est une altération du programme pour le développement n'a pas encore complètement disparu aujourd'hui.

¹⁶ Watson & Crick, 1953, « A Structure for Deoxyribose Nucleic Acid » ; Crick, 1970, « Central Dogma of Molecular Biology ». Morange (2003[1994], *Histoire de la biologie moléculaire*, p. 156) remarque que l'idée selon laquelle « la révélation de la structure de l'ADN imposa, presque naturellement, la notion de code génétique » est accréditée dans de nombreux ouvrages. Cependant, bien que Watson et Crick parlèrent déjà en 1953 du « code » pour désigner la séquence des bases nucléotidiques, cela n'impliquait pas qu'ils eussent une conscience claire de son existence. Au cours des années 1950, de nombreux types de codes furent proposés ; ce n'est qu'à partir de 1961 que le code génétique fut décrypté. Voir aussi Woese, 1967, *The Genetic Code. The Molecular Basis for Genetic Expression* ; Olby, 1974, *The Path to The Double Helix* ; Portugal & Cohen, 1977, *A Century of DNA. A History of the Discovery of the Structure and Function of the Genetic Substance* ; Judson, 1979, *The Eighth Day of Creation. The Makers of the Revolution in Biology* ; Kay, 1993, *The Molecular Vision of Life: Caltech, The Rockefeller Foundation, and the Rise of the New Biology*.

¹⁷ Jacob & Monod, 1961, « Genetic Regulatory Mechanisms in the Synthesis of Proteins » ; voir Fox Keller 2003[2000], *Le siècle du gène*.

¹⁸ A ce propos, il serait intéressant d'analyser les influences réciproques entre les deux disciplines que sont la biologie moléculaire et l'informatique : elles virent le jour entre 1940 et 1960 et mobilisent des notions apparemment semblables. Voir Kay, 1993, *op. cit.* ; Fox Keller, 2003[2000], *op. cit.* ; Morange, 2003[1994], *op. cit.*

¹⁹ Nous renvoyons à Sterelny & Griffiths (eds.), 1999, *Sex and Death. An Introduction to Philosophy of Biology*, où la question du déterminisme génétique est traitée de manière critique dans plusieurs chapitres.

3.1.2. Lamarck, Darwin et les premiers lamarckiens

La notion de hasard est utilisée de manière systématique pour caractériser les mutations génétiques depuis plus d'un siècle. L'origine de cette caractérisation, qui au cours du XX^e siècle est devenue un élément essentiel de la théorie de l'évolution, réside dans la notion de variation « spontanée », « accidentelle » ou « fruit du hasard » que Darwin avança dans le cadre de son hypothèse de l'évolution par sélection naturelle²⁰. Tout en acceptant la validité de certaines idées largement répandues au XIX^e siècle comme l'acquisition de modifications des organes par leur usage et non-usage et la transmission de ces modifications individuelles d'une génération à la suivante (ce qu'on appelle, l'« hérédité des caractères acquis »²¹), Darwin proposa une vision et une explication alternatives du processus évolutif et, plus particulièrement, de l'origine de la variation et de l'adaptation. Il conçut l'évolution comme un processus en deux étapes, la première consistant en la production des variations qui peuvent être favorables ou défavorables car aléatoires par rapport aux besoins adaptatifs des organismes concernés ; la deuxième étant la perpétuation des variations (favorables) par le processus de sélection naturelle. Selon Darwin, la sélection naturelle tend à augmenter la fréquence des variations bénéfiques d'une génération à la suivante et est donc à l'origine de l'adaptation des populations naturelles concernées.

Un demi-siècle auparavant, Lamarck²² avait au contraire conçu la transformation des espèces de manière à coupler les phénomènes à l'origine de la production et de la perpétuation de la variation. Il avait distingué deux causes déterminant le cours adaptatif de l'évolution : en premier lieu, la tendance progressive et inhérente aux organismes vivants à devenir de plus en plus complexes ; en second lieu, l'hérédité des caractères acquis en tant que mécanisme d'adaptation à l'environnement changeant. En particulier, Lamarck avait résumé ce deuxième aspect de sa théorie de la transformation des espèces par la formulation de deux « lois [...] de la nature » : selon la première²³, le fait qu'un organisme individuel utilise de manière fréquente un de ses organes provoque le renforcement et l'agrandissement

²⁰ Darwin, 1859, *On The Origin of Species By Means of Natural Selection or The Preservation of Favoured Races in The Struggle For Life* ; 1868, *The Variation of Animals and Plants Under Domestication*.

²¹ En ce qui concerne l'histoire de l'expression « hérédité des caractères acquis » et de la notion qu'elle désigne, voir Zirkle, 1946, « The early history of the idea of the inheritance of acquired characters and of pangenesis » ; Gayon, 2006b, « Hérédité des caractères acquis ».

²² Lamarck, 1809, *Philosophie Zoologique, ou Exposition des considérations relatives à l'histoire naturelle des animaux*.

²³ « Première Loi. Dans tout animal qui n'a point dépassé le terme de ses développemens, l'emploi plus fréquent et soutenu d'un organe quelconque, fortifie peu à peu cet organe, le développe, l'agrandit, et lui donne une puissance proportionnée à la durée de cet emploi ; tandis que le défaut constant d'usage de tel organe, l'affoiblit insensiblement, le détériore, diminue progressivement ses facultés, et finit par le faire disparaître. », *ibid.*, p. 235.

de ce dernier alors que son non-usage l'affaiblit et le détériore. Selon la seconde loi²⁴, les modifications acquises par un organisme individuel au cours de son développement à travers l'usage et non-usage de ses organes sont transmises à sa descendance. Lamarck pensait que les variations acquises étaient adaptatives car elles consistaient en des réponses physiologiques des organismes individuels aux changements de l'environnement afin de faire face à leurs nouveaux besoins. De plus, en concevant ces variations adaptatives comme héréditaires, Lamarck n'avait besoin de postuler aucun processus de sélection. Il est en effet évident que, les variations étant produites par les organismes individuels mêmes²⁵ dans le but de s'adapter à leur environnement – autrement dit, les variations étant « dirigées » en vue de l'adaptation –, l'étape consistant en la perpétuation des variations par un processus de sélection était non seulement absente, mais même pas nécessaire dans le cadre du transformisme de Lamarck. Le caractère adaptatif des variations héréditaires suffisait pour expliquer le cours adaptatif de l'évolution.²⁶

Suite à la publication de *L'Origine des Espèces* (1859), ouvrage dans lequel Darwin proposa l'hypothèse de l'évolution par sélection naturelle, l'hérédité des caractères acquis continua à être considérée par les biologistes évolutionnistes, par exemple par Spencer et Haeckel, comme un mécanisme pouvant expliquer l'adaptation des organismes à leur environnement au même titre que le processus de sélection. À cette époque, les conceptions darwiniennes et lamarckiennes de l'évolution n'étaient donc pas forcément opposées, mais arrivaient à cohabiter même auprès d'un même biologiste : leur confrontation engendrait plutôt des débats autour de la question de l'importance relative de l'hérédité des caractères acquis et de la sélection naturelle dans le processus d'évolution et d'adaptation.

²⁴ « Deuxième Loi. Tout ce que la nature a fait acquérir ou perdre aux individus par l'influence des circonstances où leur race se trouve depuis long-temps exposée, et, par conséquent, par l' influence de l' emploi prédominant de tel organe, ou par celle d' un défaut constant d' usage de telle partie ; elle le conserve par la génération aux nouveaux individus qui en proviennent, pourvu que les changemens acquis soient communs aux deux sexes, ou à ceux qui ont produit ces nouveaux individus. », Lamarck, 1809, *op. cit.*, p. 235.

²⁵ En ce qui concerne les animaux, Lamarck pensait que l'acquisition des modifications adaptatives était due à la réponse volontaire des organismes individuels, en particulier à travers l'usage et le non-usage de leurs organes, aux changements de l'environnement ; au contraire, dans le cas de plantes, Lamarck parlait d'une réponse physiologique et adaptative mais involontaire et automatique aux modifications environnementales.

²⁶ En suivant Lewontin (1983, « The organism as the subject and object of evolution »), nous pouvons aussi distinguer la vision darwinienne et la vision lamarckienne de l'évolution dans les termes suivants. D'une part, la pensée de Darwin correspond à une vision *variationnelle* de l'évolution selon laquelle une espèce ou une population change au cours du temps à cause des modifications dans la proportion d'organismes différents la constituant. D'autre part, la pensée de Lamarck correspond à une vision *transformationnelle* de l'évolution selon laquelle les changements au niveau de l'espèce ou de la population sont dus aux changements au niveau des organismes individuels.

3.1.3. Les lamarckiens et les « néo-lamarckiens » (fin XIX^e - début XX^e)

Autour des années 1880, l'hypothèse de l'évolution par sélection naturelle de Darwin connut une bonne réception auprès de Weismann et des naturalistes qui, en même temps, s'opposèrent vivement à l'hérédité des caractères acquis. Plus précisément, sur la base d'un certain nombre d'arguments théoriques, qui étaient parfois très sophistiqués, Weismann²⁷ récusait l'idée de Lamarck selon laquelle les modifications apparues au cours de la vie des organismes individuels pouvaient être transmises à la progéniture.²⁸ Il démontra la séparation entre le plasma germinatif et le soma et, en particulier, l'impossibilité que des modifications au niveau des caractères somatiques des organismes individuels puissent affecter le matériel héréditaire, c'est-à-dire le plasma germinatif, et être donc transmises à la génération suivante.²⁹

La cohabitation des positions des darwiniens et des lamarckiens qui avait caractérisé les deux premières décennies suivant la publication de *L'Origine des Espèces* ne put durer que jusqu'aux années 1880, lorsque l'influence de Weismann augmenta et provoqua une polarisation radicale dans l'opinion quant aux processus sous-jacents au cours adaptatif de l'évolution. En effet, en réaction aux idées et aux résultats de Weismann, de la fin du XIX^e siècle jusqu'aux vingt premières années du XX^e siècle, une grande controverse vit s'opposer, d'une part, ceux qu'en 1888 Romanes appela les « néo-darwiniens », d'autre part, un certain nombre de lamarckiens et de « néo-lamarckiens » américains³⁰. Les premiers rejetaient l'hérédité des caractères acquis, soutenaient le processus de sélection naturelle en tant que moteur principal de l'évolution et soulignaient le caractère aléatoire de la variation héréditaire par rapport à l'adaptation ; les deuxièmes défendaient l'hérédité des caractères acquis et refusaient le rôle positif et constructif de la sélection naturelle dans l'évolution et, plus précisément, dans l'origine de l'adaptation.

²⁷ Weismann, 1892[1883], *De l'Hérédité* ; 1893[1892], *The Germ-Plasm. A Theory of Heredity*.

²⁸ Weismann recourut aussi à quelques expériences afin de réfuter l'hérédité des caractères acquis. Cependant, il ne réalisa que des travaux expérimentaux concernant des cas très spécifiques et particuliers (par exemple, la très célèbre amputation de la queue d'un lézard afin de montrer que, sur plusieurs générations, la réduction graduelle de la taille de la queue n'allait pas se produire).

²⁹ En 1875, Galton avait déjà introduit une distinction semblable à la distinction weismannienne entre le soma et le germe : il avait utilisé le terme anglais « stirp » (du latin *stirpes*, qui signifie « racine ») pour désigner le matériel héréditaire et avait soutenu l'indépendance entre les cellules du corps et les cellules reproductives. Voir Dunn, 1965, *A Short History of Genetics. The development of some of the main lines of thought: 1864-1939*.

³⁰ Les termes « néo-lamarckien » et « néo-lamarckisme » sont des mots très spécifiques : ils furent proposés en 1885 par le scientifique américain Packard pour désigner un mouvement présent aux Etats-Unis qui visait à établir l'hérédité des caractères acquis comme l'alternative à la vision néo-darwiniste de Wallace et de Weismann de l'évolution. Voir Bowler, 1983, *The Eclipse of Darwinism. Anti-Darwinian Evolution Theories in the Decades around 1900* ; Jordanova, 1984, *Lamarck*. Nous signalons que Shanahan (2004, *The Evolution of Darwinism*) affirme, au contraire, que le terme « néo-lamarckisme » fut inventé par Lankester en 1885.

Il est intéressant de remarquer que beaucoup de différences caractérisaient et distinguaient les positions respectives défendues par les biologistes se proclamant lamarckiens et néo-lamarckiens : le fait de renvoyer à Lamarck en tant que père fondateur de sa propre vision de l'évolution n'impliquait pas nécessairement une bonne connaissance de l'oeuvre de Lamarck et une adhésion fidèle et complète aux idées de ce dernier³¹. De la fin du XIX^e siècle jusqu'à la fin de la première guerre mondiale, tout biologiste se déclarant lamarckien ou néo-lamarckien soutenait l'idée générale de l'hérédité des caractères acquis selon laquelle les variations acquises par les organismes individuels au cours de leur développement sont transmises à leurs descendants. Cependant, il n'y avait pas de consensus quant à la définition précise de cette notion qui pouvait assumer des formes différentes, en particulier en ce qui concerne l'origine de la variation et sa direction du point de vue de l'adaptation.³²

D'une part, les biologistes lamarckiens les plus fidèles à la pensée de Lamarck (par exemple, Jennings, Henslow et Kammerer) parlaient de l'hérédité des caractères acquis pour désigner le processus par lequel un changement se produisant au niveau du corps d'un organisme individuel est transféré dans le plasma germinatif par la production d'un changement correspondant au niveau de ce matériel héréditaire et peut donc être transmis à la génération suivante. De plus, ils soutenaient que les variations acquises par les organismes individuels et transmises à leur descendance sont dues à des réponses spécifiques et adaptatives des organismes individuels aux changements de l'environnement : le caractère adaptatif de la variation était en effet un élément constitutif et fondamental de l'explication de l'origine de l'adaptation chez Lamarck.

D'autre part, du côté des biologistes se déclarant néo-lamarckiens mais n'étant pas strictement fidèles à la pensée de Lamarck (en particulier, les biologistes de l'école américaine de paléontologie, par exemple Hyatt, Cope (même si ce dernier s'occupa aussi des tendances linéaires et adaptatives de l'évolution) et les disciples de Cope (Scott, Osborn)), on retrouve des positions diverses qu'on peut résumer de la manière suivante. D'abord, la plupart de ces biologistes s'accordaient à soutenir que les variations n'apportent pas nécessairement un avantage mais tendent à être non adaptatives : ils focalisèrent leurs études sur la transmission de ce genre de variations car leur but principal était de refuser la sélection

³¹ Voir Jordanova, 1984, *op. cit.*

³² La reconstruction historique que nous proposons de la pensée des lamarckiens et des néo-lamarckiens à la fin du XIX^e et durant la première moitié du siècle dernier se base sur les ouvrages de Jordanova (1984, *op. cit.*) et de Bowler (1993, *op. cit.*). Ils sont consacrés respectivement à la pensée de Lamarck et à l'éclipse de la vision darwinienne de l'évolution dans la période qui va de la publication de *L'Origine des Espèces* (1859) aux premières décennies du XX^e siècle.

naturelle en tant que moteur de l'évolution, ce qui leur semblait une tâche plus facile à accomplir en montrant l'acquisition et l'hérédité des caractères non adaptatifs. De plus, les néo-lamarckiens paléontologues en particulier affirmaient que l'environnement, tout en étant à l'origine causale de la variation, ne dicte pas la direction du changement, qui est au contraire due à une prédisposition interne propre à l'organisme, en particulier au plasma germinatif : l'environnement n'exerce qu'une fonction de stimulus pour l'activité physiologique de production de la variation au sein de l'organisme individuel concerné. Enfin, les biologistes plutôt orientés vers une position qu'on appelle le « lamarckisme environnemental » (par exemple, Packard et Allen)³³ pensaient que l'environnement provoque directement les variations héréditaires au niveau du corps des organismes individuels, et même dans leur plasma germinatif (Stockard), sans qu'aucun effort ou réponse de ces derniers soient impliqués.

Il est évident que la vision soutenue par ce deuxième ensemble de biologistes se disant néo-lamarckiens s'éloignait de deux éléments constitutifs de la perspective théorique et originelle de Lamarck, qui sont le fait que l'environnement ne peut pas provoquer directement des variations et le caractère adaptatif de ces dernières. De plus, la vision de certains d'entre eux s'éloignait aussi de la notion d'hérédité des caractères acquis propre aux lamarckiens les plus fidèles à la pensée de Lamarck selon laquelle le corps des organismes individuels peut acquérir une variation sans qu'aucune modification préalable se soit produite au niveau du matériel héréditaire (du plasma germinatif). Leurs idées, en particulier celles des paléontologues américains, s'approchaient plutôt des théories orthogénétiques (très popularisées, entre autres, par Eimer) affirmant que l'évolution est progressive, mais ne se déploie pas nécessairement vers une augmentation de l'adaptation des organismes à leur environnement. Plus précisément, l'orthogénèse est la vision selon laquelle l'adaptation n'est pas une caractéristique du processus évolutif ; au contraire, ses partisans pensaient qu'une tendance linéaire, régulière et non adaptative, indépendante de l'environnement et de tout besoin des organismes vivants mais inhérente au développement de ces derniers, était à l'origine du parcours linéaire du processus évolutif.

Sur la base de ces éléments propres au lamarckisme de la moitié du XIX^e siècle jusqu'aux premières décennies du siècle dernier, nous considérons comme un choix pertinent à la fois du point de vue conceptuel et historique d'utiliser le terme « lamarckien » pour désigner la vision la plus fidèle à Lamarck qui regroupe les idées suivantes. En premier lieu,

³³ Voir Bowler, 1993, *The Eclipse of Darwinism. Anti-Darwinian Evolution Theories in the Decades around 1900*.

la notion de l'hérédité des caractères acquis porte sur les modifications du corps que les organismes individuels peuvent acquérir au cours de leur développement et qu'ils peuvent transmettre à leur descendance en vertu des modifications correspondantes au niveau du matériel génétique (c'est-à-dire, du plasma germinatif). En deuxième lieu, ces modifications acquises sont des variations adaptatives produites par les organismes individuels en réponse aux changements de l'environnement, et non pas directement par l'environnement (autrement dit, elles sont ce que nous appellerons par la suite des variations « dirigées »). En troisième lieu, les variations individuelles sont directement à l'origine de l'adaptation des organismes individuels et de leurs descendants au cours du temps, sans que le processus de sélection des variations joue un rôle constructif dans ce cadre.

3.1.4. De la Synthèse Moderne (1930-1950) à aujourd'hui

Au cours des premières décennies du XX^e siècle, les diverses formes de lamarckisme s'affaiblirent³⁴ à cause, d'une part, des avancées que la génétique connut, en particulier au sein de l'école mendélienne, et, d'autre part, à cause du fait qu'aucun biologiste lamarckien ne proposa une théorie complète de l'hérédité pouvant rivaliser avec la théorie de Mendel et pouvant donc représenter une alternative cohérente à cette dernière.

Au début du XX^e siècle, un débat controversé vit s'opposer les généticiens de l'école mendélienne et les naturalistes darwiniens au sujet du caractère discret ou continu et saltatoire ou graduel de la variation. Ensuite, la notion de variation (« spontanée », « accidentelle » ou « fruit du hasard ») proposée par Darwin fut enfin reprise et reformulée en termes de mutation par les biologistes de la Théorie Synthétique Moderne (1930-1950), cette dernière consistant en la réconciliation et en la consécutive synthèse de la génétique mendélienne et de l'évolutionnisme darwinien, auparavant considérés comme incompatibles. À cette époque, la question de l'origine de la variation phénotypique qu'on observe au sein des populations naturelles n'était plus la boîte noire que Darwin n'avait pas pu ouvrir et qui, pourtant, ne l'avait pas empêché d'avancer des hypothèses pertinentes quant aux causes de la variation. Les idées des biologistes à ce sujet commençaient à s'éclaircir : la mutation et la

³⁴ Cependant, les historiens du lamarckisme rapportent les efforts que certains biologistes ont déployé pendant la période suivant la première guerre, et même plus récemment au cours des années 1970-1980 (Steele, 1979, *Somatic Selection and Adaptive Evolution: On the Inheritance of Acquired Characters*), afin de fournir des preuves expérimentales en faveur de l'hérédité des caractères acquis. Pour plus de détails à ce sujet, voir Bowler (1983, *op. cit.*), en particulier p. 98-106.

recombinaison étaient considérées comme les deux principales sources causales de l'altération du matériel génétique héréditaire et donc de la variation génotypique et, même si ce n'est pas nécessairement, phénotypique, chez les organismes vivants. La notion de hasard était invoquée pour caractériser ces deux phénomènes à l'origine causale de la variation au sein des populations naturelles : en particulier, les mutations étaient considérées comme des changements génétiques qui se produisent « au hasard » par rapport aux besoins adaptatifs des organismes concernés.

À partir de la fin de la première moitié du siècle dernier, les résultats d'études expérimentales diverses (entre autres, la célèbre expérience de Luria et Delbrück en 1943) furent considérés comme des preuves en faveur de cette conception du caractère aléatoire des mutations génétiques, qui datait de la pensée de Darwin et qui fut ensuite reprise par les biologistes de la Synthèse Moderne. Sur cette base, elle devint la vision consensuelle au sein de la communauté des biologistes, en particulier en génétique classique, en génétique des populations, et plus tard en biologie moléculaire. Cependant, à la fin des années 1980, une longue controverse scientifique et philosophique fut déclenchée par un article dans lequel le microbiologiste Cairns et ses collègues³⁵ prétendaient avoir montré expérimentalement l'existence de « bactéries lamarckiennes », c'est-à-dire de bactéries capables de répondre de manière spécifique et adaptative aux changements de l'environnement où elles se trouvent et de transmettre la variation favorable acquise à leur descendance. Ils pensaient donc avoir trouvé un élément mettant en cause les démonstrations expérimentales des années 1940-1950 et donc la vision consensuelle quant au caractère aléatoire des mutations génétiques.

Au cours des vingt dernières années, des expériences ultérieures ont montré de manière définitive le caractère non fondé des conclusions de Cairns et ses collègues. Cependant, la notion de hasard invoquée par la plupart des biologistes pour décrire les mutations génétiques est encore aujourd'hui l'objet d'un défi conceptuel et empirique sur la base d'expériences et de résultats similaires à ceux qui avaient déclenché un débat si controversé à la fin des années 1980. Plus précisément, elle semble être mise en cause par la découverte relativement récente de ce qu'on appelle les « mécanismes mutateurs » et que nous analyserons en détail dans les chapitres suivants.

Quelle est donc cette notion de hasard au centre des controverses qui ont caractérisé la réflexion scientifique et philosophique en biologie de la fin du XIX^e siècle et tout au long du XX^e jusqu'à aujourd'hui ? C'est le but de cette deuxième partie d'identifier et de définir de

³⁵ Cairns *et al*, 1988, « The origins of mutants ».

manière précise la signification et le rôle de cette notion utilisée par Darwin, par les biologistes de la Synthèse Moderne et par la plupart des biologistes jusqu'à aujourd'hui pour caractériser cette source causale de la variation. Une fois cette notion identifiée, nous chercherons aussi à comprendre si elle est appropriée dans ce cadre, c'est-à-dire s'il s'agit d'une notion de hasard cohérente du point de vue conceptuel et pertinente du point de vue empirique relativement aux mutations génétiques. Ces deux buts que nous nous donnons correspondent à ce que nous appelons le volet descriptif et le volet normatif de notre enquête sur la notion de hasard dans la caractérisation des sources de la variation biologique.

3.2. Des altérations diverses de la séquence d'ADN

Afin d'entrer dans le vif de notre travail de clarification conceptuelle, il est d'abord nécessaire de présenter les connaissances dont les biologistes disposent quant à ces sources causales de la variation que sont les mutations génétiques. Nous nous intéresserons en particulier aux différents types de mutations possibles, aux mécanismes moléculaires dont l'activité consiste à éviter que des mutations se produisent et à la probabilité de ces dernières, à la fois au sein des populations naturelles et au niveau moléculaire.

Des altérations diverses peuvent affecter le matériel génétique : les mutations génétiques sont habituellement classifiées et définies relativement à leur origine, relativement aux types de changement de la séquence d'ADN, relativement à leurs conséquences au niveau de la synthèse des protéines et, enfin, relativement à leur valeur adaptative pour l'organisme individuel concerné.³⁶

Considérons d'abord les origines possibles des mutations génétiques. Les mutations dites « spontanées » sont des altérations de la séquence d'ADN dues aux processus physico-chimiques se déroulant à l'intérieur de la cellule. Elles peuvent trouver leur origine dans des transformations chimiques de la séquence nucléotidique, comme l'isomérisation tautomérique, qui est le changement de la forme chimique des bases de l'ADN et qui est une source potentielle d'appariements erronés ; la dépurination, consistant en l'élimination d'une base purique (adénine ou guanine) d'une double hélice intacte et qui peut se produire spontanément à des températures physiologiques ; la désamination, c'est-à-dire la perte d'un

³⁶ Pour une introduction claire et synthétique aux mutations génétiques, voir Futuyma, 1998[1979], *Evolutionary Biology* ; Ridley, 2004, *Evolution* ; Stearns & Hoekstra, 2005, *Evolution: An Introduction* ; Lewin, 2004, *Genes VIII* ; Klug *et al*, 2006, *Génétique*. Pour une description et une explication plus détaillées et techniques de tout type de mutation génétique, voir Friedberg *et al*, 2006, *DNA Repair and Mutagenesis*.

groupe aminé au niveau d'une base nucléotidique, qui provoque la conversion, par exemple, d'une cytosine en uracile et d'une adénine en hypoxanthine et donc l'altération de la spécificité d'appariement de la base nucléotidique affectée³⁷.

Les mutations dites « spontanées » peuvent aussi être dues à des lésions de l'ADN provoquées par des composants moléculaires présents naturellement dans le milieu intracellulaire. Par exemple, le métabolisme produit fréquemment des radicaux libres d'oxygène pouvant provoquer des lésions au niveau la séquence de l'ADN. D'autres lésions peuvent être dues à la réaction d'alkylation³⁸ : en entrant en réaction avec un brin de la molécule d'ADN, des agents chimiques (des groupes méthyl, éthyl et propyl)³⁹ peuvent l'endommager au niveau de son squelette sucre-phosphoryle ou des bases nucléotidiques et altérer ses propriétés d'appariement.

Les erreurs se produisant au cours du processus de réplication de l'ADN sont aussi une source de mutations « spontanées ». Les ADN polymérases, qui sont des complexes enzymatiques impliqués dans la synthèse d'un brin d'ADN à partir d'un brin matrice donné, peuvent insérer une base nucléotidique erronée au cours de la synthèse du nouveau brin : cette erreur peut se transformer en une mutation génétique si elle n'est pas détectée et réparée par les divers mécanismes de réparation se déployant avant, pendant et après le processus de réplication. Des erreurs de réplication peuvent se produire à cause du glissement des ADN polymérases sur le brin matrice d'ADN : ces complexes enzymatiques peuvent se déplacer le long de la séquence sans répliquer un nombre variable de nucléotides et, de cette manière, provoquer la délétion de ces derniers. Les ADN polymérases peuvent aussi produire des erreurs au cours de la réplication de la séquence d'ADN en introduisant dans le nouveau brin synthétisé des bases nucléotidiques qui ne correspondent pas à celles présentes sur le brin matrice : le résultat est l'addition de séquences d'ADN plus ou moins longues.

Les mutations dites « induites » sont des altérations de la séquence d'ADN dues à des agents mutagènes de nature physico-chimique. Ces mutagènes peuvent être des composants naturels de l'environnement où la cellule concernée se trouve (comme les toxines des champignons, les rayonnements ultraviolets et les rayonnements cosmiques) ou des produits physico-chimiques dus à l'activité humaine (comme les produits de l'industrie chimique ou

³⁷ La déamination est un processus spontané mais peut aussi être la conséquence d'un traitement par des agents mutagènes chimiques tels que l'acide nitreux.

³⁸ L'alkylation est une réaction chimique consistant en le transfert d'un groupe alkyl, c'est-à-dire d'un groupement d'un nombre variable d'atomes de carbone et d'hydrogène, d'une molécule à une autre. Elle provoque l'augmentation du nombre d'atomes de carbone d'un composant organique.

³⁹ Les groupes méthyl, éthyl et propyl sont des groupements d'un nombre variable d'atomes de carbone et d'hydrogène : ils sont des radicaux alkyls. Leurs formules respectives sont : CH₃ pour le groupe méthyl, C₂H₅ pour le groupe éthyl et C₃H₇ pour le groupe propyl.

les radiations provenant d'outils technologiques divers). Plus précisément, parmi les mutagènes chimiques figurent ce qu'on appelle les analogues de bases, c'est-à-dire des substances chimiques qui ressemblent aux bases azotées, qui peuvent être incorporées durant la synthèse de la séquence d'ADN et qui montrent un appariement différent des bases qu'ils remplacent ; les gaz moutarde, qui sont des agents alkylants ; les agents intercalants, comme les colorants acridiniques, qui sont des molécules capables de s'insérer entre les bases nucléotidiques appariées et qui peuvent provoquer des distorsions de la molécule d'ADN, des additions, des délétions et des décalages de la lecture de la séquence d'ADN lors de la production des protéines. Parmi les mutagènes physiques figurent les rayonnements ultraviolets et les radiations ionisantes⁴⁰ (les rayons X, les rayons gamma et les rayons cosmiques). À leur passage, ces dernières peuvent provoquer l'ionisation de certaines molécules présentes à l'intérieur de la cellule qui, en se transformant en radicaux libres et en ions réactifs, peuvent réagir avec l'ADN et l'endommager. Les radiations ultraviolettes et les radiations ionisantes peuvent avoir des origines naturelles (le soleil, les produits de la terre, l'atmosphère, etc.) et être le produit de l'activité humaine (les tests médicaux, la radio, la télévision, les tests nucléaires, les détecteurs de fumée, etc.).

En deuxième lieu, les mutations génétiques peuvent être distinguées et définies relativement aux types de changements de la séquence d'ADN. Les mutations dites « ponctuelles » consistent en la substitution d'une paire de bases en une autre dans la séquence de la molécule d'ADN⁴¹. Elles peuvent être « spontanées » ou « induites ». Elles sont appelées « transitions » lorsqu'une base pyrimidique est remplacée par une autre base pyrimidique ou une base purique est remplacée par une autre base purique. On parle de « transversion » lorsqu'une base purique et une base pyrimidique se substituent l'une à l'autre.

On désigne les insertions et les délétions d'une ou plusieurs bases nucléotidiques à un endroit de la séquence d'ADN par l'expression « mutations de décalage du cadre de lecture ». Afin d'expliquer l'appellation de ce genre de mutations, il faut rappeler que la structure primaire d'une protéine est une chaîne plus ou moins longue d'acides aminés : sa synthèse consiste donc en la production d'un certain nombre d'acides aminés à partir d'une séquence

⁴⁰ L'ionisation consiste en la perte ou l'ajout de charges (d'électrons) à un atome ou à une molécule. Elle peut provoquer la rupture des liaisons que cet atome – ou cette molécule – a avec les atomes – ou les molécules – l'entourant. L'atome – ou la molécule – ionisé est appelé un ion.

⁴¹ Parfois les biologistes utilisent l'expression « mutation ponctuelle » pour désigner une altération affectant un petit nombre de paires de bases de la séquence d'ADN, comme de petites insertions et délétions. Lorsqu'une séquence de deux ou plusieurs nucléotides est remplacée par une autre séquence, nous préférons utiliser l'expression « substitution de séquence ».

d'ARN qui, à son tour, a été synthétisée à partir d'un brin d'ADN donné. Chaque triplet de trois bases nucléotidiques adjacentes (ou codon) le long de la séquence d'ARN correspond à un certain type d'acide aminé ; de plus, certains codons ne correspondent à aucun acide aminé mais jouent un rôle spécifique dans le processus de synthèse des protéines en régulant son initiation ou sa terminaison⁴². Par conséquent, une mutation consistant en l'insertion⁴³ ou en la délétion d'une base ou davantage peut changer la constitution des codons en les décalant le long de la séquence nucléotidique et, par conséquent, peut provoquer la synthèse d'acides aminés différents ou une altération au niveau de la régulation du processus de synthèse protéique.

Des mutations peuvent aussi affecter de grandes régions chromosomiques et même des chromosomes entiers. On parle d'« amplification » pour désigner la duplication de segments chromosomiques contenant un gène ou davantage (plus spécifiquement, on parle de « duplication génétique »). Les chromosomes peuvent aussi être altérés par délétion, ce qui est à l'origine de la perte des gènes situés dans les régions supprimées ; par inversion, qui consiste en la réorientation de segments chromosomiques ; par translocation, c'est-à-dire par l'échange de morceaux de matériel génétique entre des chromosomes non homologues. Au niveau du génome, des modifications importantes peuvent se produire par l'addition ou la perte d'un ou plusieurs chromosomes entiers (on parle d'« aneuploïdie »).

En troisième lieu, les mutations génétiques peuvent être distinguées et définies sur la base de leurs conséquences au niveau de la synthèse des protéines. On parle de mutation « faux-sens » lorsqu'une mutation, en modifiant la constitution d'un codon, provoque le changement de l'acide aminé produit et est donc susceptible d'affecter la structure et la fonction de la protéine produite. Ce genre de mutation peut impliquer la perte partielle ou totale de la fonction propre à la protéine synthétisée à partir de la séquence d'ADN modifiée. Au contraire, une mutation « silencieuse » ou « synonyme » est une mutation qui, tout en modifiant la composition d'un codon, ne provoque pas de changement au niveau de la séquence d'acides aminés constituant la protéine synthétisée. Une mutation « non-sens » est

⁴² Dans certains cas, plusieurs codons différents correspondent à un même acide aminé (par exemple, les codons UCA et UCG correspondent à la leucine) ou jouent le même rôle dans la régulation du processus de synthèse des protéines (par exemple, les triplets UAA et UAG provoquent l'arrêt du processus de traduction). Cette redondance est due à la différence entre le nombre de triplets possibles de bases nucléotidiques (64) et les nombres d'acides aminés (20) ; elle est aussi due au fait qu'un certain nombre de codons ne sont impliqués dans la production d'aucun acide aminé mais dans la régulation – dans l'initiation ou dans la terminaison – de la synthèse des protéines.

⁴³ Les insertions sont dues à ce qu'on appelle les « éléments transposables » : ce sont des séquences nucléotidiques qui, après ou sans leur duplication préalable, sont capables de s'exciser d'un site de l'ADN, de se déplacer et de s'insérer au niveau d'un autre site de l'ADN.

une mutation qui transforme un codon originellement impliqué dans la production d'un certain acide aminé en un codon de terminaison du processus de synthèse des protéines : en engendrant l'arrêt prématuré de la synthèse d'une protéine donnée, ce type de mutation génétique peut provoquer le dysfonctionnement partiel ou total de cette dernière.

En quatrième lieu, les mutations génétiques peuvent aussi être distinguées et définies selon une perspective évolutionnaire, c'est-à-dire relativement à leur valeur adaptative pour l'organisme individuel concerné dans un environnement donné. En effet, une mutation peut être favorable, défavorable (même létale) ou neutre en termes de survie et de reproduction de l'organisme porteur. Sa valeur adaptative ne dépend pas simplement du type de changement produit au niveau de la séquence d'ADN ; elle dépend aussi du contexte moléculaire dans lequel cette altération se produit, de son effet phénotypique⁴⁴, et des caractéristiques de l'environnement dans lequel l'organisme porteur se trouve.⁴⁵

3.2.1. La stabilité dynamique de l'ADN : les mécanismes de réparation

En dépit du fait que le matériel génétique peut subir toutes ces altérations et beaucoup d'autres dues, par exemple, au processus de recombinaison, l'ADN est une molécule assez stable au cours du temps⁴⁶. Sa stabilité n'est pas simplement due à sa nature chimique et à sa structure physique ; il s'agit plutôt d'une propriété dynamique qui est assurée par un certain nombre d'enzymes hautement spécifiques impliquées dans les activités de protection, de relecture, de contrôle et de réparation du matériel génétique au cours de sa réplication. Ces complexes enzymatiques, qui diminuent la fréquence des erreurs dans la réplication de l'ADN

⁴⁴ L'effet phénotypique d'une mutation est déterminé par un ensemble complexe de facteurs environnementaux (au sens large) qui sont internes et externes à l'organisme individuel et qui affectent son développement. Pour une analyse critique de la relation complexe entre le génotype et le phénotype et du rôle central du processus de développement et de l'environnement dans la genèse de la variation phénotypique à partir de la variation au niveau du génome, voir Hallgrímsson & Hall (eds.), 2005, *Variation. A Central Concept in Biology*.

⁴⁵ À ce propos, il faut souligner que l'environnement d'un organisme individuel ne peut pas simplement être conçu comme l'ensemble des facteurs abiotiques composant le milieu physique où ce dernier se trouve. L'environnement comprend aussi l'ensemble des facteurs biotiques dont l'organisme individuel donné se nourrit et tous les autres organismes de sa population et d'autres populations qui sont situées dans le même milieu physique. En ce qui concerne les problèmes liés à la définition de l'environnement d'un organisme ou d'une population donnée, voir Brandon, 1984, « On the concept of environment in evolutionary biology ».

⁴⁶ Sur la base de la comparaison des séquences d'ADN de différentes espèces, le taux moyen des mutations spontanées par paire de bases par réplication a été estimé autour de 10^{-9} . Cependant, nous verrons que les taux de mutation varient énormément entre des organismes d'espèces différentes et, de plus, diffèrent aussi chez un même organisme selon les gènes et les séquences nucléotidiques intragéniques considérés. Voir Futuyma, 1998[1979], *Evolutionary Biology*, p. 272-273.

à environ une sur dix milliards⁴⁷, sont des produits du processus évolutif et leur découverte a apporté une explication satisfaisante à la question de l'origine de la stabilité et de la fidélité génétiques.

On peut distinguer quatre types de mécanismes moléculaires assurant la stabilité de l'ADN et la fidélité de sa réplication⁴⁸ : l'activation de chacun de ces mécanismes dépend de manière spécifique du type de lésion ou d'erreur affectant la séquence de l'ADN. De manière générale, on utilise l'expression « mécanisme de réparation » pour désigner un mécanisme intracellulaire qui entre en jeu en réponse à des lésions ou à des erreurs de la séquence d'ADN : son activité consiste à restaurer la séquence nucléotidique et la structure de l'ADN normales. Cependant, nous verrons que l'activité de quelques mécanismes dits « de réparation » ne consiste pas à restaurer l'ADN endommagé ou modifié, mais à tolérer des lésions afin que la réplication de la séquence nucléotidique puisse continuer et que la cellule concernée puisse survivre.

Un premier type de mécanisme de réparation consiste en la réversion de certaines lésions de la séquence d'ADN par l'action de complexes enzymatiques spécifiques sans que cela requière la rupture du squelette sucre-phosphoryle du brin affecté. Par exemple, la photoréactivation est un mécanisme de réparation de ce genre et consiste en la réversion des lésions dues à des radiations ultraviolettes (des dimères de thymine) grâce à l'activité d'une enzyme de photoréactivité (la photolyase) ; l'action de cette dernière est induite par une brève exposition des cellules affectées à la lumière bleue du spectre visible. D'autres lésions pouvant être réparées par réversion grâce à l'activité de mécanismes hautement spécifiques sont la méthylation des bases de guanine, d'adénine et de cytosine.

Un deuxième type de mécanisme de réparation est activé lorsqu'une ou plusieurs bases de la séquence d'un seul brin d'ADN sont affectées par une lésion ou une erreur : l'autre brin peut donc être utilisé pour guider la réparation du brin endommagé ou modifié. Ces mécanismes consistent en l'excision de bases ou de nucléotides⁴⁹ endommagés ou erronés et en leur remplacement par des bases ou des nucléotides appropriés. Le mécanisme de réparation par excision d'une base (BER) repose sur l'excision et sur le remplacement d'une seule base nucléotidique lésée ou erronée. Le mécanisme de réparation par excision nucléotidique (NER) entre en jeu lors de l'occurrence de lésions affectant des séquences de 2

⁴⁷ Radman & Wagner, 1988, « The high fidelity of DNA duplication ».

⁴⁸ Pour une introduction aux mécanismes de réparation de l'ADN, voir Friedberg *et al*, 2006, *DNA Repair and Mutagenesis*.

⁴⁹ Un nucléotide est une sous-unité de la molécule d'ADN composée par une base azotée (adénine, guanine, cytosine, thymine), une molécule de sucre (le désoxyribose) et un groupe phosphate.

à 30 bases : il repose sur l'excision et sur le remplacement de nucléotides lésés ou erronés. Le mécanisme de réparation des mésappariements (MRS) consiste en la correction d'erreurs qui se sont produites au cours des processus de réplication et de recombinaison de l'ADN et qui ont provoqué des mésappariements entre les deux brins.

Un troisième type de mécanisme de réparation entre en jeu lors de la rupture d'un brin ou des deux brins de l'ADN : ce genre de lésions peut avoir des conséquences importantes (par exemple, le blocage de la réplication de l'ADN) pouvant mettre en jeu la survie de la cellule concernée. Ces mécanismes de réparation consistent en l'activation du processus de recombinaison homologue, qui implique l'échange de fragments d'ADN entre deux molécules d'ADN, ou du processus de recombinaison non homologue et raccordement (*joining*) qui, au contraire, ne requiert aucun échange de matériel génétique.

Enfin, un dernier type de mécanismes de réparation de l'ADN ne repose pas sur l'excision des lésions et sur leur remplacement par des séquences nucléotidiques appropriées : on parle de « mécanismes de tolérance » pour désigner des mécanismes qui gardent et entérinent les lésions au niveau de la séquence d'ADN afin de permettre à la cellule concernée de survivre. Parmi ces mécanismes figure ce qu'on appelle la « réparation par recombinaison » : elle permet d'initier à nouveau la réplication de la séquence d'ADN en aval d'une lésion qui en provoque l'arrêt et de synthétiser la séquence manquante par des échanges de matériel génétique entre le brin intact et le brin lésé (c'est-à-dire, par recombinaison).

Un autre mécanisme de tolérance est ce qu'on appelle la « synthèse translésionnelle » qui se fait grâce à l'activité de complexes enzymatiques hautement spécifiques (des ADN polymérases spécialisées). Ce mécanisme permet de faire face à de forts stress affectant la molécule d'ADN (par exemple, encore une fois l'arrêt de sa réplication dû à une lésion au niveau de la séquence nucléotidique) mais, en même temps, ne garantit pas sa réplication fidèle. En effet, les ADN polymérases impliquées dans la synthèse translésionnelle peuvent continuer la réplication d'une séquence d'ADN endommagée, contrairement à d'autres ADN polymérases, mais ont tendance à provoquer des erreurs ultérieures au cours de sa synthèse.⁵⁰

⁵⁰ Dans le chapitre 5, nous verrons que les ADN polymérases impliquées dans les mécanismes de réparation dits « de tolérance » sont aussi impliquées dans certains « mécanismes mutateurs » (notamment, ce qu'on appelle le « système SOS ») et jouent un rôle central dans l'augmentation du taux de mutation dans des cellules en situation de stress environnemental.

3.3. Le taux de mutation

Depuis les premières études en génétique au début du siècle dernier, les biologistes utilisent des approches diverses, mais toutes basées sur l'analyse statistique de données, afin d'évaluer de manière quantitative l'occurrence des mutations. En termes actuels, ce qu'on appelle le taux de mutation, c'est-à-dire la probabilité qu'une mutation se produise en un laps de temps donné et dans une certaine portion de la séquence d'ADN (ou dans le génome entier), dépend strictement des caractéristiques physico-chimiques de la molécule d'ADN considérée, des caractéristiques des processus de réplication, de recombinaison et de réparation la concernant, de l'efficacité des mécanismes de sa réparation et des conditions environnementales où l'organisme/la cellule concerné/e se trouve. De plus, comme le taux de mutation est obtenu en mesurant la fréquence de mutations par unité de temps et par séquence génomique choisies, sa valeur dépend aussi de la longueur de la période de temps et de la longueur de la séquence d'ADN prises en considération⁵¹.

Les biologistes ont évalué le taux de mutation⁵² dans le cas d'organismes appartenant à des espèces diverses, des rétrovirus jusqu'aux humains : ils ont identifié des *patterns* de mutation différents qui dépendent donc de la taille du génome, de la longueur des générations, et des caractéristiques des processus de réplication, de recombinaison et de réparation chez les organismes de chaque espèce considérée. Ces études ont amené les biologistes à conclure que les taux de mutation sont à la fois le résultat de l'histoire évolutive commune à tout organisme vivant et de l'histoire adaptative spécifique à chaque espèce particulière.⁵³

⁵¹ Par exemple, le taux de mutation par génome par génération dépend de la longueur temporelle d'une génération chez les organismes étudiés et du nombre de paire de bases constituant la séquence d'ADN analysée.

⁵² La plupart des estimations du taux de mutation portent sur l'occurrence des mutations spontanées, indépendamment des altérations produites au niveau de la séquence nucléotidique, c'est-à-dire indépendamment du fait qu'il s'agisse de mutations ponctuelles, d'insertions ou de délétions, ou même d'altérations majeures au niveau chromosomique. Voir Drake, 1991a, « Spontaneous mutation » ; 1991b, « A constant rate of spontaneous mutation in DNA-based microbes » ; 1992, « Mutation rates » ; 1993, « Rates of spontaneous mutation among RNA viruses » ; Drake *et al.*, 1998, « Rates of spontaneous mutation ». Certaines études consistent aussi en l'évaluation du taux de chaque type de mutation, par exemple par génome ou par gène par réplication, chez des espèces d'organismes dont l'ADN a été séquencé. Voir Schaaper & Dunn, 1991, « Spontaneous Mutation in the *Escherichia coli lacI* Gene » ; Oller & Schaaper, 1994, « Spontaneous mutation in *Escherichia coli* containing the *dnaE911* DNA polymerase antimutator allele ». Cependant, nous soulignons que le calcul du taux de mutation devient plus difficile et s'expose davantage à des erreurs d'estimations lorsqu'il porte sur des mutations plus importantes que les mutations ponctuelles (par exemple, des insertions et des délétions de plusieurs nucléotides) ; dans ces cas, il est donc nécessaire de considérer avec précaution les estimations obtenues.

⁵³ Kimura (1967, « On the evolutionary adjustment of spontaneous mutation rates ») a procédé à une analyse des forces appliquées sur le taux de mutation et, sur cette base, a proposé l'idée selon laquelle, de manière générale, le taux de mutation le plus faible est sélectionné. Le généticien moléculaire Drake a interprété l'analyse de Kimura comme reflétant un taux minimal de mutation dont le seul frein réside dans le coût des mécanismes de réparation de la séquence d'ADN. Dans le chapitre suivant, nous verrons que l'hypothèse du taux minimal de mutation a été mise en cause par la découverte des mécanismes moléculaires de mutation dits « mutateurs » car

3.3.1. Les méthodes d'évaluation du taux de mutation

Le taux de mutation peut être évalué à différents niveaux en utilisant des méthodes théoriques et expérimentales diverses et à plusieurs échelles spatio-temporelles.⁵⁴

En ce qui concerne le niveau d'évaluation, depuis les premières études en génétique classique, le taux de mutation est estimé sur la base des effets phénotypiques détectables qu'elles produisent⁵⁵. Autour des années 1940, Luria et Delbrück⁵⁶ inventèrent et utilisèrent la méthode expérimentale dite « test (ou analyse) de fluctuation » afin de montrer que les mutations génétiques ne se produisent pas en réponse à, mais avant toute pression de sélection⁵⁷. Bien que de nombreuses critiques aient été adressées au test de fluctuation au cours des trente dernières années, en particulier au sujet de ses hypothèses de départ⁵⁸ et de l'évacuation de certains phénomènes pouvant affecter son résultat final⁵⁹, il figure parmi les méthodes statistiques utilisées encore aujourd'hui pour évaluer le taux de mutation.

Le test de fluctuation consiste à déterminer le taux de mutation en analysant la distribution du nombre de mutants dans des cultures parallèles. Il repose sur la démarche expérimentale suivante. On inocule un petit nombre d'individus isogéniques dans plusieurs cultures parallèles et on les laisse pousser jusqu'au stade de saturation (dans le cas de l'expérience de Luria et Delbrück, il s'agissait de cultures indépendantes de bactéries, toutes provenant de la croissance d'un même nombre d'individus). Ensuite, un certain nombre de

ils produisent une augmentation du taux de mutation dans des conditions environnementales de stress.

⁵⁴ Notre présentation des méthodes d'évaluation du taux de mutation suit Futuyma, 1998[1979], *Evolutionary Biology*, p. 271-272 ; Ridley, 2004, *Evolution*, p. 31-32 ; Friedberg *et al*, 2006, *DNA Repair and Mutagenesis*.

⁵⁵ Voir Morgan, 1988[1925], *The Genetics of Drosophila*, p. 22-23.

⁵⁶ Luria & Delbrück, 1943, « Mutations of Bacteria From Virus Sensitivity to Virus Resistance ».

⁵⁷ Déjà autour des années 1920, les généticiens mendéliens essayaient de déterminer à quel moment du cycle de vie les mutations se produisent en analysant le nombre des premiers organismes mutants dans une population donnée (Voir Morgan, 1988[1925], *op. cit.*). Dans le cadre du chapitre 5, nous analyserons la pertinence du test de fluctuation proposé par Luria et Delbrück relativement à leur projet de déterminer si les mutations génétiques se produisent avant ou suite à une pression de sélection.

⁵⁸ Il s'agit des hypothèses de ce qu'on appelle le modèle Lea-Coulson sur l'expansion des clones mutants. Ce modèle avait été originairement décrit par Luria et Delbrück en 1943 ; ensuite, il avait été repris et étendu par Lea et Coulson en 1949 (« The distribution of the numbers of mutants in bacterial populations »). Dans le dernier chapitre de cette deuxième partie, nous considérerons certaines des nombreuses hypothèses de départ de ce modèle lorsque nous analyserons la notion de hasard dans la caractérisation des mutations génétiques au niveau moléculaire.

⁵⁹ Cairns *et al*, 1988, « The origin of mutants » ; Stewart *et al*, 1990, « Fluctuation Analysis: The Probability Distribution of the Number of Mutants Under Different Conditions » ; Sarkar, 1991a, « Lamarck *contre* Darwin, Reduction *versus* Statistics: Conceptual Issues in the Controversy Over Directed Mutagenesis in Bacteria » ; Jaeger & Sarkar, 1995, « On the distribution of bacterial mutants: The effects of differential fitness of mutants and non-mutants ».

cultures sont étalées séparément dans un milieu sélectif relativement à un trait donné permettant aux cellules mutantes de produire des colonies (dans le cas de l'expérience de Luria et Delbrück, le facteur sélectif était un phage lytique, donc le trait pris en considération était la résistance à ce virus bactérien). Le résultat de cette étape de l'expérience permet de relever le nombre de mutants qui ont pu survivre et se dupliquer dans un tel environnement et donc d'estimer leur nombre moyen par culture. En même temps, un petit nombre de cultures sont étalées séparément dans un milieu non sélectif afin de déterminer le nombre total de cellules après la phase de croissance. Sur la base des paramètres fondamentaux dans ce type de démarche expérimentale que sont le nombre de mutants par culture (r), le nombre initial (N_0) et le nombre final (N_T) de cellules, le volume de la culture (V), le nombre de culture parallèle (C) et le nombre de mutations par culture (m) (qui est calculé à partir de r et qui dépend strictement de V et C), le taux de mutation (μ), c'est-à-dire le nombre moyen de mutations se produisant durant la vie d'une cellule, peut être estimé. Plus précisément, il est égal au rapport entre le nombre de mutations par culture (m) et un paramètre mesurant le nombre de cellules susceptibles de muter, par exemple N_t ($\mu = m/N_t$)⁶⁰.

Une autre méthode basée sur l'analyse de la variation au niveau phénotypique est celle qu'on appelle l'« accumulation des mutations » et qui consiste en l'estimation du taux de mutation à travers la quantification du changement du nombre d'individus mutants par culture quant à un trait phénotypique donné au cours des générations. Cependant, les nombreuses limites atteignant cette deuxième méthode rendent le test de fluctuation préférable⁶¹. De manière générale, la limite majeure de toute estimation du taux de mutation sur la base des effets phénotypiques détectables consiste en la sous-estimation du taux de mutation car, d'une part, un même effet phénotypique peut être dû à une seule mutation dans un certain site mais aussi à plusieurs altérations de la séquence d'ADN ; d'autre part, de nombreuses mutations peuvent n'avoir aucune conséquence au niveau phénotypique.

Une deuxième manière d'estimer le taux de mutation consiste à se placer au niveau

⁶⁰ Nous renvoyons à Rosche & Foster, 2000, « Determining Mutation Rates in Bacterial Populations », et à Foster, 2006, « Methods for Determining Spontaneous Mutation Rates », en particulier pour une analyse exhaustive et critique des méthodes pour calculer le nombre de mutations par culture (m) et pour estimer le taux de mutation (μ). Dans ces deux articles figurent aussi une présentation détaillée des problèmes concernant les méthodes statistiques pour calculer le taux de mutation et une variété de solutions possibles. En effet, les estimations de m et de μ obtenues via le test de fluctuation ne sont ni distribuées selon une loi normale, ni non biaisées; par conséquent, il n'est pas approprié de calculer la moyenne et l'écart type de ces résultats. Voir aussi Asteris & Sarkar, 1996, « Bayesian procedures for the estimation of mutation rates from fluctuation experiments » ; Stewart, 1994, « Fluctuation Tests: How Reliable Are the Estimates of Mutation Rates? » ; Jones *et al.*, 1994, « Luria-Delbrück fluctuation experiments: design and analysis ».

⁶¹ Voir Foster, 2006, *op. cit.*

des protéines produites : après les avoir séparées par électrophorèse⁶², il faut procéder au séquençage des protéines afin de détecter les modifications qui ont affecté leur structure primaire, c'est-à-dire les changements au niveau des chaînes d'acides aminés. Cette méthode connaît aussi un certain nombre de limites. Nous signalerons simplement qu'elle est contrainte par la même limite que la quantification sur la base des effets phénotypiques détectables, même si c'est de manière moins importante : en effet, certaines altérations de la séquence d'ADN, notamment les substitutions de plusieurs bases nucléotidiques situées en troisième position au sein des codons, ne provoquent aucun changement au niveau de la chaîne d'acides aminés constituant une protéine.

Les technologies moléculaires de séquençage de l'ADN⁶³, qui sont des avancées récentes et qui désormais sont devenues des outils indispensables pour la recherche en biologie moléculaire, permettent d'estimer directement le taux de mutation en analysant les altérations au niveau de la séquence nucléotidique de l'ADN sur plusieurs générations. Plus précisément, les biologistes arrivent à construire ce qu'ils appellent des spectres mutationnels qui représentent graphiquement la probabilité (ou le taux) de l'occurrence des mutations, et même de chaque type d'altération possible, le long de la séquence d'ADN d'un groupe donné d'organismes⁶⁴.

Cette méthode moléculaire est aujourd'hui beaucoup plus utilisée que celles basées sur l'analyse de données au niveau phénotypique ; cependant, elle n'est pas exempte de limites. Une source possible d'erreur d'estimation réside dans les principes méthodologiques qui guident les biologistes dans l'interprétation des données obtenues par le séquençage des génomes d'un ensemble d'organismes à des générations successives. Un cas éclairant de cette limite est la mobilisation du principe de parcimonie⁶⁵ qui, dans le cas de la quantification du taux de mutation, consiste à ne faire l'hypothèse que du nombre minimum d'altérations

⁶² L'électrophorèse des protéines est une technique chimique utilisée pour séparer un mélange de protéines et les caractériser. Elle repose sur l'observation du déplacement des protéines chargées positivement ou négativement à travers les pores d'un gel lorsqu'on applique un courant électrique. La révélation de leur déplacement est possible grâce à des colorations spécifiques. Plus une protéine est grande et constituée par une longue séquence polypeptidique, moins elle migre car elle est retenue par les mailles du gel.

⁶³ Voir, par exemple, Alpey (1997), *DNA Sequencing: From Experimental Methods to Bioinformatics*.

⁶⁴ Schaaper & Dunn, 1987, « Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: The nature of *in vivo* DNA replication errors » ; 1991, « Spontaneous Mutation in the *Escherichia coli lacI* Gene » ; Oller & Schaaper, 1994, « Spontaneous mutation in *Escherichia coli* containing the *dnaE911* DNA polymerase antimutator allele ».

⁶⁵ Le principe de parcimonie est un principe heuristique, appelé aussi « rasoir d'Occam » du nom du philosophe et théologien auquel on l'attribue. Il consiste à n'expliquer l'inconnu que par le connu ou, autrement dit, à ne pas multiplier les hypothèses tant que celles déjà annoncées suffisent pour résoudre un problème donné. De nombreuses publications de Sober, des années 1970 jusqu'à son dernier livre publié en 2008, fournissent une analyse philosophique très approfondie du principe de parcimonie, en particulier lorsqu'il est appliqué en phylogénétique. Voir, par exemple, Sober, 2006, « Parsimony ».

possibles de la séquence d'ADN et qui, par conséquent, peut induire en erreur et fausser de manière importante l'estimation du taux de mutation. Par exemple, lorsque les biologistes observent qu'au niveau d'un même site nucléotidique et à plusieurs générations d'écart figure la même base, le principe de parcimonie les amènera à conclure qu'aucune mutation ne s'est produite à cet endroit de la séquence d'ADN, alors qu'ils ne connaissent pas les étapes intermédiaires impliquées dans la détermination de ce résultat, donc ils ne peuvent pas savoir si plusieurs mutations ont affecté ce site et se sont compensées au cours du temps.

Le taux de mutation peut donc être estimé sur la base de l'analyse de la variation à des niveaux différents – le niveau des traits phénotypiques, le niveau des protéines et le niveau de la séquence nucléotidique – et, de plus, à diverses échelles spatio-temporelles. Du point de vue spatial, on peut calculer la probabilité qu'une mutation se produise par paire de bases ou par nucléotide, par séquence intragénique donnée, par gène et par génome. Du point de vue temporel, le taux de mutation peut être estimé à des échelles très différentes : par réplication, par génération, sur plusieurs générations, sur des laps de temps plus long (par exemple, un an).

De manière générale, l'estimation du taux de mutation est une opération à la fois exigeante du point de vue expérimental car elle requiert des conditions de mesure précises et bien contrôlées, et problématique d'un point de vue théorique car elle implique inévitablement des hypothèses de départ parfois éloignées des caractéristiques du processus mutationnel dans la réalité.

L'un des nombreux aspects problématiques pouvant rendre difficile l'estimation du taux de mutation, en particulier l'accord entre le modèle théorique et les conditions expérimentales, concerne les conditions de mesure du taux de mutation : les biologistes visent à éliminer ou, du moins, minimiser l'action de la sélection naturelle au sein du groupe d'organismes étudié. En effet, les mutations génétiques peuvent être favorables, défavorables ou neutres pour l'organisme porteur relativement à l'environnement où ce dernier se trouve ; en d'autres termes, elles peuvent augmenter la probabilité de survie et de reproduction de l'organisme porteur dans un environnement donné, elles peuvent la diminuer, ou elles peuvent n'avoir aucun effet en termes de survie et de reproduction de l'organisme porteur. Dans des conditions naturelles, les biologistes ont observé que beaucoup de mutations sont défavorables, et même létales, car un organisme est généralement assez bien adapté aux conditions environnementales où il se trouve. Le caractère défavorable des mutations implique que les organismes porteurs peuvent mourir avant qu'elles soient détectées, soit au niveau de leurs effets phénotypiques, soit au niveau des protéines produites, soit au niveau de

la séquence l'ADN : par conséquent, le risque commun à toute méthode de quantification du taux de mutation en présence de la sélection naturelle consiste en sa sous-estimation. De plus, la sélection naturelle est très difficile à neutraliser complètement ; les estimations des taux de mutation sont donc toujours des valeurs non pas réelles mais approchées⁶⁶.

3.4. Notre projet : clarifier la notion de hasard dans la caractérisation des mutations génétiques

La démarche à la fois descriptive et normative que nous allons suivre dans les chapitres suivants afin de comprendre quels sont la signification et le rôle de la notion de hasard dans la caractérisation des mutations portera sur le débat controversé, qui a explosé à plusieurs reprises au cours du XX^e siècle, sur le caractère aléatoire des mutations génétiques du point de vue évolutionnaire. Il n'est pas dans nos projets de suivre et de décrire en détail cette controverse. Plutôt, nous focaliserons notre analyse sur le défi que, selon certains biologistes, historiens et philosophes de la biologie contemporains⁶⁷, les résultats relativement récents des recherches sur les mutations génétiques posent à la notion de mutation « aléatoire » propre au paradigme de la Théorie Synthétique Moderne.

Durant les trente dernières années, de nombreuses recherches expérimentales en biologie moléculaire, en particulier en microbiologie, ont montré que le taux de mutation peut être régulé à travers certains mécanismes moléculaires, dits mécanismes « mutateurs », en relation aux forces sélectives en action⁶⁸. C'est la découverte de ce lien causal entre le taux de mutation et le substrat sélectif qui a amené certains à affirmer que les mutations génétiques héréditaires dues aux mécanismes mutateurs sont « non aléatoires » car elles se produisent de manière spécifique au moment et à l'endroit de la séquence d'ADN où elles sont avantageuses pour la survie et la reproduction des organismes individuels concernés. En particulier, nous nous confronterons directement aux arguments que Jablonka et Lamb (2005)⁶⁹ ont avancés contre le caractère « aléatoire » des mutations dues à un certain type de mécanismes

⁶⁶ Ridley, 2004, *Evolution*, p. 31.

⁶⁷ Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life* ; Shapiro 1995a, « Adaptive Mutation: Who's Really in the Garden? » ; Shapiro, 2005, « A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering » ; Sternberg, 2002, « On the Roles of DNA Repetitive Elements in the Context of a Unified Genomic-Epigenetic System » ; Fox Keller, 2003[2000], *Le siècle du gène*.

⁶⁸ Giraud *et al*, 2001, « The rise and the fall of mutator bacteria ».

⁶⁹ Jablonka & Lamb, 2005, *op. cit.*

mutateurs (les mécanismes dont l'activité est induite temporairement par un stimulus de l'environnement et provoque une augmentation locale du taux de mutation). Selon elles, ces mécanismes sont à la fois des processus darwiniens et lamarckiens car à la fois sélectifs et instructifs, et donc mettent en cause la validité de l'idée propre à la Synthèse Moderne selon laquelle toute mutation génétique se produit « au hasard » par rapport à la direction de l'adaptation.

Afin de comprendre si la découverte de ces mécanismes mutateurs pose véritablement un défi au caractère « aléatoire » des mutations génétiques au sens de la Théorie Synthétique Moderne, il est d'abord nécessaire de clarifier ce que les expressions « mutation aléatoire », « mutation qui se produit au hasard », etc., signifient selon la vision consensuelle de cette théorie. Nous allons donc essayer de comprendre quelle est la notion de hasard invoquée dans la caractérisation des mutations génétiques du point de vue évolutionnaire sur la base de ce qu'un certain nombre des pères fondateurs de la Synthèse Moderne ont écrit à ce sujet et à la lumière des résultats empiriques de l'époque sur les mutations génétiques, en particulier sur les mécanismes à leur origine et sur leur relation avec l'environnement et les forces sélectives en action. C'est en se souciant de garder une certaine cohérence entre ces deux aspects, l'un théorique et l'autre plutôt expérimental, que nous identifierons et redéfinirons de manière empiriquement et conceptuellement appropriée la notion de hasard utilisée au sein de la Théorie Synthétique Moderne dans la caractérisation des mutations génétiques du point de vue évolutionnaire. C'est cette notion donc, que nous désignerons par l'expression « hasard évolutionnaire », qui selon Jablonka et Lamb ne s'appliquerait pas au cas des mutations dues à certains des mécanismes mutateurs récemment découverts. C'est seulement après une telle clarification qu'il nous sera possible de saisir le vrai impact des résultats récents en biologie moléculaire et leurs conséquences en ce qui concerne le caractère aléatoire des mutations génétiques.

Une fois identifiée la notion de « hasard évolutionnaire » qui, selon notre analyse, correspond à la vision consensuelle de la Synthèse Moderne quant au caractère aléatoire des mutations génétiques et la notion de variation « spontanée », « accidentelle » ou « fruit du hasard » de Darwin, nous présenterons en détail les divers types de mécanismes mutateurs qui ont été découverts durant les trente dernières années. Ensuite, afin de comprendre si Jablonka et Lamb ont raison lorsqu'elles affirment que les mutations dues à certains de ces mécanismes ne sont pas aléatoires par rapport au cours de l'évolution contrairement à la façon dont les biologistes de la Synthèse Moderne les concevaient, nous testerons la notion de « hasard évolutionnaire » dans le cas d'un des mécanismes mutateurs parmi les plus controversés. Le

parcours argumentatif que nous proposerons vise à montrer que les mutations provoquées de manière plus ou moins spécifique par tous ces mécanismes mutateurs en réponse à des conditions environnementales de stress sont aussi des mutations aléatoires au sens de la notion de « hasard évolutionnaire » propre à la Synthèse Moderne.

Notre argument consistera essentiellement, d'abord, en l'observation critique selon laquelle Jablonka et Lamb ont attaqué cette notion sans l'avoir comprise et définie préalablement et sans avoir une idée claire et historiquement correcte de ce qu'elles proposent, c'est-à-dire de ce que c'est une vision lamarckienne de l'origine causale de la variation héréditaire et de l'adaptation. En second lieu, en une mise en évidence de certaines distinctions qui jouent un rôle fondamental dans la définition des mutations aléatoires par la notion de « hasard évolutionnaire ». Nous concluons par une défense de la validité actuelle de la Théorie Synthétique Moderne qui, bien qu'elle soit encore aujourd'hui l'explication la plus cohérente et complète de l'évolution naturelle, a besoin de quelques mises à jour sur des points précis concernant par exemple les mécanismes moléculaires de mutation.

Nous défendrons l'idée – qui était plutôt un postulat au moment de la Synthèse dans les années 1930-1950 – selon laquelle *toute* mutation est une altération du matériel génétique qui se produit au hasard du point de vue évolutionnaire en montrant à la fois ce que les avancées récentes en biologie moléculaire modifient et ne modifient pas par rapport aux résultats empiriques et au cadre théorique de la vision consensuelle de la Synthèse Moderne. La notion de « hasard évolutionnaire » se révélera donc comme la notion appropriée, à la fois du point de vue empirique et conceptuel, dans la caractérisation de toute mutation génétique observée par les biologistes jusqu'à maintenant et comme une notion dont la définition est spécifique à la biologie.

L'analyse de ce débat sur le caractère aléatoire des mutations génétiques du point de vue évolutionnaire sera suivie par une enquête au sujet de la notion de hasard appropriée pour décrire le caractère aléatoire des mutations au niveau moléculaire, autrement dit non pas relativement à la relation entre mutations, sélection et adaptation, mais par rapport aux processus physico-chimiques qui les engendrent⁷⁰. Dans le dernier chapitre de cette deuxième partie, nous rappellerons ce que les biologistes de la Synthèse reconnaissaient quant aux causes des mutations génétiques et à leur distribution le long de la séquence génomique, et nous présenterons ce que les généticiens moléculaires connaissent aujourd'hui à ce sujet. Sur cette base, nous caractériserons la manière dont les biologistes de la Synthèse Moderne

⁷⁰ Voir le chapitre 7.

concevaient le caractère aléatoire des mutations génétiques du point de vue moléculaire. Ensuite, nous proposerons une définition de ce que nous appellerons la notion d'« aléatoire (au sens) faible » qui, selon notre analyse, est la notion conceptuellement et empiriquement appropriée pour décrire le hasard des mutations génétiques au niveau de la séquence d'ADN et relativement aux résultats les plus récents des recherches en biologie moléculaire. Nous mettrons donc en évidence les nouveautés apportées par ces avancées quant à la conception du caractère aléatoire des mutations du point de vue moléculaire par rapport à la vision de la Synthèse Moderne.

Une partie importante de notre enquête portera sur l'examen analytique des définitions des notions de « hasard évolutionnaire » et d'« aléatoire (au sens) faible ». Nous focaliserons notre attention sur leur formulation respective, sur l'extension propre à chacune dans le cas des mutations génétiques, et nous réfléchirons sur la relation entre elles et avec les notions de hasard qui ont marqué l'histoire de la pensée philosophique et scientifique⁷¹. Notre but est de comprendre, d'abord, si les notions de « hasard évolutionnaire » et d'« aléatoire (au sens) faible », mobilisées dans la caractérisation des mutations génétiques, sont définies en termes objectifs ou subjectifs ; ensuite, si leur signification respective correspond à une des notions philosophico-scientifiques de hasard ; enfin, si elles ont un rôle purement descriptif ou explicatif.

⁷¹ Voir la première partie de ce travail.

Chapitre 4 – La notion de « hasard évolutionnaire » de Darwin à la Synthèse Moderne

4.1. Darwin

L'idée centrale du paradigme de la Théorie Synthétique Moderne (1930-1950) au sujet des mutations génétiques est souvent exprimée en affirmant que les mutations sont « aléatoires » ou le fruit du « hasard »¹ par rapport à l'adaptation, autrement dit relativement aux besoins adaptatifs des organismes individuels concernés. Cette idée remonte à la pensée de Darwin à propos de ce qu'il appelait les variations « spontanées », « accidentelles » ou fruit du « hasard » chez les organismes vivants. Dans les ouvrages de ce dernier², on peut identifier et distinguer trois manières différentes de concevoir la notion de hasard.³

En premier lieu, Darwin s'approprie de l'idée des théologiens naturels du XVIII^e et du XIX^e siècles selon laquelle un événement causal est le fruit du hasard s'il n'est pas le résultat d'un projet divin :

¹ Les biologistes de la Synthèse Moderne utilisaient principalement les deux termes « hasard » et « aléatoire » pour caractériser les mutations génétiques ; parfois ils mobilisaient aussi d'autres mots comme « aveugle », « accidentel », « stochastique », « probabiliste » et « imprédictible ». Dans le cadre de cette analyse, nous mobiliserons les deux premiers termes en réservant le mot « hasard » pour désigner le caractère des mutations génétiques du point de vue évolutionnaire et le mot « aléatoire », qui est un terme pouvant être interprété selon des connotations mathématiques formelles et précises, pour désigner le caractère des mutations au niveau moléculaire, relativement aux processus physico-chimiques qui les engendrent.

² Darwin, 1859, *On The Origin of Species By Means of Natural Selection or The Preservation of Favoured Races in The Struggle For Life* ; 1868, *The Variation of Animals and Plants Under Domestication* ; Darwin & Seward (eds.), 1903, *More letters of Charles Darwin*.

³ Voir Beatty, 1984, « Chance and Natural Selection » ; 1987, « The probabilistic revolution in evolutionary biology. An overview » ; 2006, « Chance Variation: Darwin on Orchids » ; 2008, « Chance Variation and Evolutionary Contingency: Darwin, Simpson (*The Simpsons*) and Gould », p. 189-210 ; Eble, 1999, « On the dual nature of chance in evolutionary biology and palaeology » ; Lennox, 2008, « Darwinism ». Beatty (2006, *op. cit.*) distingue deux significations de l'idée de variation « spontanée », « accidentelle » ou fruit du « hasard » chez Darwin. La première exprime le fait que les variations au sein des populations naturelles ne sont pas des réponses adaptatives à l'environnement, donc elle fait référence à la notion de hasard au sens d'absence d'un dessein et à ce que nous appellerons la notion de « hasard évolutionnaire ». La deuxième signification renvoie à l'imprédictibilité des relations causales entre l'environnement et les variations : on retrouve ici la notion de hasard au sens de l'ignorance des causes sous-jacentes des variations. Cependant, Beatty semble suggérer deux manières, l'une subjective l'autre objective, de concevoir l'origine de cette ignorance chez Darwin. En effet, ce dernier distinguait deux classes de variations : les variations qui sont imprédictibles en principe, donc à cause de leurs caractéristiques objectives, et les variations qui ne sont pas prédictibles présentement, donc à cause de l'ignorance humaine.

« Je ne peux voir aucune preuve d'un dessein bienfaisant ni même d'un quelconque dessein, dans les détails. »⁴

Selon Darwin, les variations individuelles qu'on observe au sein des populations naturelles sont donc « spontanées », « accidentelles » ou le fruit du « hasard » au sens où elles ne sont pas dues à un dessein divin ou à un projet finalisé de n'importe quelle autre nature.⁵

En deuxième lieu, Darwin conçoit la notion de hasard de manière subjective (ou épistémique) et affirme qu'elle désigne notre ignorance des causes précises des variations héréditaires :

« Jusqu'à présent j'ai parfois parlé comme si les variations – [...] – étaient dues au hasard. Cela est évidemment une expression totalement incorrecte, mais elle est utile afin d'admettre pleinement notre ignorance des causes de chaque variation particulière. »⁶ :

Darwin considère donc que les variations sont régies par des lois spécifiques de la nature et déterminées par des causes compliquées mais précises⁷ que, cependant, nous ne connaissons pas ; par conséquent, il caractérise les variations comme « spontanées », « accidentelles » ou le fruit du « hasard » en faisant référence à l'ignorance humaine des lois et des causes sous-jacentes.⁸

En troisième lieu, Darwin caractérise les variations héréditaires en utilisant une notion de hasard qui fait référence à la relation que les variations entretiennent avec le processus évolutif, en particulier avec l'action de la sélection naturelle et l'adaptation des organismes vivants. Il ne s'agit pas d'une notion métaphysique de hasard pur selon laquelle le hasard est

⁴ « *I can see no evidence of beneficent design or indeed of design of any kind, in the details.* », Darwin & Seward (eds.), 1903, *op. cit.*, p. 321 (notre traduction).

⁵ On retrouve ici la notion de hasard objectif au sens de l'absence d'un dessein. Nous renvoyons au chapitre 1.

⁶ « *I have hitherto sometimes spoken as if the variations – [...] – had been due to chance. This, of course, is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation.* », Darwin, 1859, *op. cit.*, p. 131 (notre traduction).

⁷ Beatty (2006, *op. cit.* ; 2008, *op. cit.*) montre que Darwin, tout en s'opposant à l'idée selon laquelle les différences parmi les formes des organismes sont le résultat d'un projet divin, ne s'oppose pas à l'attribution d'un rôle à Dieu dans la création des organismes vivants. Plus précisément, Darwin pense que les lois de la variation, de l'hérédité et de l'évolution ont été conçues par Dieu. Par contre, la manière d'atteindre l'adaptation – ce qu'il appelle « les détails » – est laissé au domaine du hasard : « J'ai tendance à regarder toute chose comme le résultat de lois planifiées, avec les détails, bons ou mauvais, laissés au travail de ce qu'on pourrait appeler hasard », Darwin à Gray, 22 May 1860, *In Darwin & Seward* (eds.), 1903, *op. cit.*, p. 223-224, (notre traduction), « *I am inclined to look at everything as resulting from designed laws, with the details, whether good or bad, left to the working out of what we may call chance* ».

⁸ Cette deuxième manière de concevoir le hasard de la variation selon Darwin mobilise la notion de hasard subjectif désignant l'ignorance humaine des causes sous-jacentes d'un événement (voir le chapitre 1). Nous renvoyons à la note 3 de ce chapitre pour plus de détails quant à la distinction entre imprédictibilité en principe des variations, donc due aux caractéristiques objectives de ces dernières, et imprédictibilité limitée au moment présent, donc due à l'ignorance humaine des causes des variations.

une entité réelle et active dans le monde ou qui renvoie à l'absence radicale de causes. En effet, Darwin affirme :

« Je ne peux pas voir l'univers comme le résultat d'un hasard aveugle [...] »⁹.

Il s'agit plutôt d'une notion de hasard strictement liée à la première au sens d'absence d'un dessein divin ou d'un projet finalisé : elle désigne l'absence d'une détermination causale et spécifique des variations en vue de la production de résultats adaptatifs. Darwin écrit :

« Imaginons un architecte devant construire un édifice avec des pierres non taillées, tombées d'un précipice. La forme de chaque fragment peut être appelée accidentelle. La forme de chacun a été déterminée [...] par des événements et des circonstances, tous dépendants de lois naturelles, mais il n'y a aucune relation entre ces lois et le but pour lequel chaque fragment est utilisé par le constructeur. De la même manière, les variations de chaque créature sont déterminées par des lois fixes et immuables ; mais elles n'entretiennent aucune relation avec la structure vivante qui est lentement construite à travers la puissance de la sélection [...] »¹⁰.

Encore une fois, Darwin souligne que les variations sont déterminées causalement et selon des lois précises ; cependant, elles ne sont pas déterminées causalement de manière spécifique et en vue de l'adaptation des organismes individuels concernés et de leur descendance dans l'environnement où ils se trouvent. Selon Darwin, les variations sont donc « spontanées », « accidentelles », autrement dit elles sont le fruit du « hasard », par rapport au cours adaptatif de l'évolution : elles sont des variations « non dirigées » du point de vue évolutionnaire ou, comme nous les caractériserons à la fin de ce chapitre, des variations se produisant au « hasard évolutionnaire ».¹¹

⁹ « *I cannot look at the universe as the result of blind chance [...]* », Darwin & Seward (eds.), 1903, *More letters of Charles Darwin*, p. 321 (notre traduction).

¹⁰ « *Let an architect be compelled to built an edifice with uncut stones, fallen from a precipice. The shape of each fragment may be called accidental. Yet the shape of each has been determined [...] by events and circumstances, all of which depend on natural laws ; but there is no relation between these laws and the purpose for which each fragment is used by the builder. In the same manner the variations of each creature are determined by fixed and immutable laws ; but these bear no relation to the living structure which is slowly built up through the power of selection [...]* », Darwin, 1886, *The Variation of Animals and Plants Under Domestication*, p. 249 (notre traduction).

¹¹ Il est historiquement et conceptuellement intéressant de rappeler que Darwin admettait la plausibilité de certaines idées propres à Lamarck et généralement reconnues tout au long du XIX^e siècle. Il reconnaissait la possibilité de l'hérédité des caractères acquis (ou hérédité flexible), c'est-à-dire le fait que le corps des organismes individuels puisse acquérir, au cours de son développement, des variations et que ces dernières puissent être transmises à la génération suivante. Plus précisément, Darwin considérait l'influence de l'environnement et l'usage et non-usage des organes comme deux causes potentielles des variations héréditaires au sein des populations naturelles. Nous partageons l'opinion de Jablonka et Lamb (2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, p. 15) selon

Le développement de la théorie des probabilités et des statistiques au cours du XIX^e siècle influença très faiblement Darwin dans la formulation des trois déclinaisons de l'idée des variations « spontanées », « accidentelles » ou fruit du « hasard »¹². Cependant, il est intéressant de rappeler que Darwin caractérisa aussi les variations en termes de probabilités, plus précisément en ce qui concerne leur occurrence dans des laps de temps plus ou moins longs et dans des populations de tailles différentes. En effet, il pensait qu'une variation favorable a plus de chances de se produire au sein d'une population de grande taille et après beaucoup de générations ; autrement dit, une variation favorable, selon Darwin, est plus probable dans ce genre de situation que dans le cas d'une population caractérisée par un petit nombre d'individus et de naissances.¹³

Le début du XX^e siècle vit s'affronter deux approches dans le cadre de l'étude des variations génétiques, l'école biométrique et l'école Mendélienne, ensuite réconciliées par Fisher en 1925. Au sujet des différences distinguant ces deux écoles, on peut rappeler que, d'une part, les partisans de l'approche biométrique (parmi lesquels figure le fondateur de l'école biométrique et eugénique Galton, qui était le cousin de Darwin) adhéraient à la vision darwinienne selon laquelle la variation consiste en des changements graduels et continus ; de plus, ils évaluaient les caractéristiques phénotypiques des organismes vivants en les mesurant et en les analysant avec les outils des statistiques quantitatives. D'autre part, l'approche de l'école Mendélienne (entre autres, Bateson, de Vries, Correns) reposait sur l'idée selon laquelle la variation est, au contraire, discrète et qualitative : chaque caractère phénotypique présente des versions alternatives qualitativement bien distinctes. En dépit de ces différences au sujet de la nature de la variation, l'école biométrique et l'école Mendélienne partageaient une caractéristique commune du point de vue méthodologique qu'est la mobilisation, beaucoup plus importante en biométrie, des outils des statistiques et, plus généralement, des raisonnements probabilistes dans le cadre de l'étude de la variation.¹⁴

lesquelles l'ouverture de Darwin à des causes de la variation dépendant de l'environnement et au mécanisme de l'hérédité flexible – qui, à partir de la fin du XIX^e siècle, étaient plutôt considérées comme des idées lamarckiennes – n'affaiblit pas la validité de son hypothèse de l'évolution par sélection naturelle. Darwin ne considérait pas cette dernière comme la seule explication de l'évolution des populations naturelles, mais comme une explication alternative, plus puissante, et plus cohérente que les autres avec ce qu'il observait dans le monde naturel.

¹² Beatty, 1987, « The probabilistic revolution in evolutionary biology. An overview ».

¹³ Nous partageons l'interprétation de Beatty (2006, « Chance Variation: Darwin on Orchids ») selon laquelle Darwin, lorsqu'il parlait des variations en ces termes probabilistes, pensait à quelque chose de proche à la loi des grands nombres.

¹⁴ Voir Gigerenzer *et al* (eds.), 1989, *The Empire of Chance. How Probability Changed Science and Everyday Life*.

En dépit de l'influence et de l'apport de plus en plus importants des statistiques et de la théorie des probabilités en biologie, aucun changement ne se produisit dans la conception de la notion de hasard relativement au phénomène de la variation : les biologistes du début du siècle dernier suivirent l'idée de Darwin selon laquelle les variations sont « aléatoire » ou « non dirigées » par rapport au cours adaptatif de l'évolution et continuèrent donc à définir la notion de hasard en faisant référence à la relation des variations avec l'adaptation, c'est-à-dire au résultat du processus de sélection au niveau des populations naturelles. En particulier, la plupart des biologistes considérés comme les acteurs de la Synthèse Moderne (1930-1950) mobilisèrent un certain nombre de reformulations semblables, mais parfois ambiguës, de la conception de Darwin des variations « spontanées », « accidentelles » ou fruit du « hasard » relativement au cours de l'adaptation et l'appliquèrent aux sources de la variation, en particulier aux mutations génétiques. Une telle variété se révèle problématique car ces reformulations ne sont pas toutes très explicites ni claires quant à la notion de hasard à laquelle elles renvoient. En effet, selon la signification qu'on attribue aux termes « hasard » et « aléatoire », qui ne sont presque jamais définis explicitement, et selon la caractérisation précise des éléments auxquels d'autres expressions comme « absence de corrélation », « indépendance », « manque de référence », etc., renvoient, certaines formulations peuvent être interprétées de façons multiples et parfois semblent même entrer en contradiction avec ce que les biologistes de la Synthèse Moderne savaient et reconnaissaient quant à la variation génétique et à ses causes possibles.

L'analyse à laquelle nous consacrerons cette deuxième partie de notre recherche vise à identifier quelle est la notion de hasard utilisée par les biologistes de la Synthèse Moderne pour caractériser les mutations génétiques et, plus spécifiquement, à en comprendre la signification et le rôle au sein de la théorie de l'évolution. Une fois cette notion identifiée, nous visons aussi à en fournir une formulation appropriée qui puisse représenter de manière précise la vision consensuelle de la Synthèse Moderne et ensuite de la communauté des biologistes jusqu'à aujourd'hui. Nous nous demanderons aussi si cette notion correspond véritablement à la pensée de Darwin à ce sujet, c'est-à-dire à l'idée que les variations sont « non dirigées » par rapport au cours évolutif de l'adaptation et si, de manière générale, elle est appropriée, à la fois du point de vue conceptuel et empirique, pour caractériser les mutations génétiques.

4.2. La Synthèse Moderne

Selon la Théorie Synthétique de l'Evolution, il y a deux sources causales de la variation génétique – qui, à cette époque, était considérée comme la seule variation pouvant être transmise à la génération suivante – chez les organismes vivants : la mutation et la recombinaison. Ces deux processus, sources potentielles d'innovation génétique et phénotypique, sont considérés comme des phénomènes imprédictibles en ce qui concerne le moment dans le temps où ils se produisent, les gènes qu'ils affectent, et les organismes individuels concernés¹⁵. En nous plongeant dans les ouvrages majeurs des acteurs de la Synthèse Moderne, nous avons identifié des formulations légèrement différentes qui étaient toutes utilisées de manière systématique par les biologistes de la Théorie Synthétique pour exprimer le caractère « aléatoire » des mutations génétiques ou, en d'autres termes, le fait que les mutations sont le fruit du « hasard ». Sur la base de quelques aspects semblables entre ces formulations, nous les avons réparties en trois groupes.

Le premier groupe rassemble des formulations qui évoquent explicitement la notion de hasard par les termes « aléatoire » et « hasard » et qui définissent le caractère aléatoire d'une mutation génétique sans préciser s'il s'agit d'une caractérisation en termes du processus physico-chimique de mutation ou de la modification même que ce dernier produit au niveau du matériel génétique. Ces formulations consistent à affirmer qu'une mutation est « aléatoire » ou le fruit du « hasard »

« *relativement à l'évolution* »¹⁶, « *du point de vue de l'adaptation et de l'intégration fonctionnelle* »¹⁷, « *par rapport à la direction de l'adaptation* »¹⁸, « *en référence à leur (sa) valeur fonctionnelle ou adaptative* »¹⁹.

La formulation avancée par Dobzhansky représente une légère exception au sein de ce groupe car plus précise du point de vue de la caractérisation de la mutation génétique en termes de processus :

¹⁵ Dobzhansky, 1951[1937], *Genetics and The Origin of Species*, p. 45 ; Simpson, 1953, *The Major Features of Evolution*, p. 86 ; Dobzhansky, 1970, *Genetics of the Evolutionary Process* p. 92.

¹⁶ « [...] *random with regard to evolution* », Huxley, 1948[1942], *Evolution. The Modern Synthesis*, p. 54 (notre traduction).

¹⁷ « [...] *random from the point of view of adaptation and functional integration* », Simpson, 1984[1944], *Tempo and Mode in Evolution*, p. 55-56 (notre traduction).

¹⁸ « [...] *random with respect to the direction of adaptation* », Stebbins, 1971, *Processes of Organic Evolution*, p. 35 (notre traduction).

¹⁹ « [...] *random with reference to their functional or adaptive value* », Stebbins, 1982, *Darwin to DNA, Molecules to Humanity*, p. 69 (notre traduction).

« une mutation est un processus aléatoire *par rapport aux* besoins adaptatifs de l'espèce »²⁰.

Dans les formulations du deuxième groupe, les termes « hasard » et « aléatoire » ne sont presque jamais employés. Cependant, d'autres expressions et d'autres idées qui accompagnent souvent la notion de hasard (par exemple, les idées d'absence de corrélation ou d'indépendance) sont utilisées pour caractériser les mutations génétiques. Dans ce cadre, le caractère aléatoire de ces dernières est défini, tantôt en faisant référence à leur occurrence, donc au processus mutationnel, tantôt en renvoyant aux effets ou aux résultats finaux du processus, donc aux mutations mêmes. Comme dans le cas du premier groupe, ces formulations figurent dans plusieurs ouvrages des auteurs de la Synthèse Moderne et correspondent aux affirmations suivantes :

« *Aucune corrélation* n'a été trouvée entre les conditions externes et la direction de la mutation »²¹ ; « Les mutations sont des changements se produisant *indépendamment* du fait qu'ils sont ou qu'ils pourraient être potentiellement adaptatifs »²² ; « La direction des changements produits par elles (les mutations) semble n'avoir *aucune relation* avec la direction du changement évolutionnaire qu'on observe dans le type [d'organisme] ni avec les besoins adaptatifs ou fonctionnels de l'organisme »²³ ; « Leurs effets [des mutations] n'ont *aucune relation* avec les besoins de l'organisme ni avec les conditions où il se trouve. Elles se produisent *sans aucune référence* à leurs conséquences ou à leurs utilisations biologiques possibles »²⁴ ; « Il n'existe *aucune corrélation* connue entre un groupe particulier de conditions de l'environnement et, parmi les nombreux possibles, l'allèle particulier auquel mutera le gène »²⁵ ; « Les mutations sont des changements *aléatoires* car elles se produisent *indépendamment* du fait qu'elles sont favorables ou défavorables [...] »²⁶ ; « La mutation n'a *rien à voir* avec l'adaptation »²⁷.

²⁰ « [...] *mutation is a random process with respect to the adaptive needs of the species* », Dobzhansky, 1970, *op. cit.*, p. 65 (notre traduction).

²¹ « *No correlation has been found between external conditions and direction of mutation [...]* », Wright, 1931, « Evolution in Mendelian Populations », p. 142 (notre traduction).

²² « *Mutations are changes that occur regardless of whether they are or may be potentially useful* », Dobzhansky, 1951[1937], *op. cit.*, p. 74 (notre traduction).

²³ « [...] *the directions of the changes produced by them appear to be unrelated either to the direction of the evolutionary change to be observed in the type, or to the adaptive or functional needs of the organism* », Huxley, 1948[1942], *op. cit.*, p. 54 (notre traduction).

²⁴ « *Their effects are not related to the needs of the organism, or to the conditions in which it is placed. They occur without reference to their possible consequences or biological uses* », Huxley, 1953, *Evolution in Action*, p. 39 (notre traduction).

²⁵ Mayr, 1974[1970], *Populations, espèces et évolution*, p. 117.

²⁶ « *Mutations are random changes because they occur independently of whether they are beneficial or harmful, and therefore they are a disordering process* », Dobzhansky et al., 1977, *Evolution*, p. 66 (notre traduction).

²⁷ « *Mutation has nothing to do with adaptation* », Mayr 1976, *Evolution and the Diversity of Life*, p. 12 (notre traduction).

Enfin, l'aspect commun aux formulations que nous avons rassemblées dans un troisième groupe réside dans le fait qu'il ne s'agit que de définitions négatives du caractère aléatoire des mutations génétiques : elles consistent à opposer la conception des mutations propre à la Synthèse Moderne à celle des théories lamarckiennes de l'évolution et, en particulier, à la conception de la variation héréditaire qui avait caractérisée la pensée de Lamarck. De plus, dans le cadre de ces formulations, les mutations, et en particulier leur caractère aléatoire, sont décrits et définis en faisant référence de manière plus ou moins explicite à certaines caractéristiques causales du processus mutationnel et à la valeur adaptative de l'effet final de ce dernier, donc en concevant les mutations à la fois en termes de processus et de résultat, et plus généralement en renvoyant aux caractéristiques du processus évolutif. Ce genre de formulations correspond aux affirmations suivantes :

« La nature des mutations observées *n'est pas* compatible avec la vision selon laquelle l'évolution est dirigée par ses buts »²⁸ ; « En ce qui concerne la mutation génétique, l'observation [...] révèle des caractéristiques qui semblent être *les plus éloignées possibles* de celles requises par un processus évolutif directement adaptatif »²⁹ ; « La génétique affirme que l'organisme *n'a pas* l'habileté providentielle de répondre aux demandes de l'environnement par la production de mutations adaptées à ces demandes » ; « une situation idéale serait que les organismes puissent répondre aux défis posés par les changements de l'environnement en ne produisant que des mutations favorables où et quand ils (les organismes) en ont besoin. Mais la nature *n'a pas été* assez généreuse pour fournir à ses créations une habileté si providentielle »³⁰ ; « Un gène *ne mute jamais* en réponse à un besoin »³¹ ; « D'autre part, lorsqu'il est appliqué à la mutation le terme 'aléatoire' fait souvent référence au *manque de correspondance* de l'effet phénotypique avec le stimulus et avec la direction actuelle ou adaptative de l'évolution. Des mutations induites par la chaleur *ne produisent pas* un changement phénotypique lié à la tolérance à la chaleur »³² ; « Les mutations *ne sont pas dirigées* par rapport aux besoins adaptatifs de l'espèce. [...] Le fait que la

²⁸ « *The nature of the mutations observed is not compatible with the view that evolution is directed by their means* », Fisher, 1999[1930], *The Genetical Theory of Natural Selection: A complete variorium edition*, p. 20 (notre traduction).

²⁹ « *As to gene mutation, observation [...] reveals characteristics which seem as far as possible from those required for a directly adaptive evolutionary process* », Wright, 1931, *op. cit.*, p. 99 (notre traduction).

³⁰ « *Genetics does assert that the organism is not endowed with providential ability to respond to the requirement of the environment by producing mutations adapted to these requirements.* » ; « *An ideal situation would be if the organism were to respond to the challenge of the changing environment by producing only beneficial mutations where and when needed. But nature has not been kind enough to endow creations with such a providential ability* », Dobzhansky, 1951[1937], *Genetics and The Origin of Species*, p. 51, 74 (notre traduction).

³¹ « *A gene never mutates in response to a need* », Mayr, 1942, *Systematics and the Origin of Species*, p. 68 (notre traduction).

³² « *On the other hand, the term 'randomness' as applied to mutation often refers to the lack of correspondence of phenotypic effect with the stimulus and with the actual or the adaptive direction of evolution. Heat-induced mutations do not produce phenotypic change related to heat tolerance* », Simpson, 1953, *The Major Features of Evolution*, p. 87 (notre traduction).

mutabilité *n'est pas* limitée aux changements qui augmentent l'adaptation de ses porteurs peut sembler une déplorable imperfection de la nature. Cependant, seul un Panglossien vitaliste pourrait imaginer que les gènes savent où et quand il est bon pour eux-mêmes de muter »³³ ; « Non seulement les mutations *ne semblent jamais* se produire en réponse au besoin d'un changement génétique particulier, mais encore une mutation particulière *n'est pas* non plus lancée par une constellation particulière de facteurs environnementaux »³⁴.

Comme nous l'avons anticipé plus haut, toutes ces formulations que nous avons identifiées et distinguées en trois groupes semblent retenir l'idée de variation « non dirigée » du point de vue évolutionnaire de Darwin car elles caractérisent les mutations génétiques, non pas en tant que phénomènes physico-chimiques se produisant au niveau du génome, mais en termes de leur relation avec le cours du processus évolutif et, plus précisément, avec l'adaptation. Cependant, certaines de ces formulations ne nous permettent pas de comprendre de manière incontestable ce que les biologistes de la Synthèse Moderne entendaient en affirmant que les mutations génétiques sont « aléatoires » ou sont le fruit du « hasard ».

Dans le cas du premier groupe, le sens du terme « aléatoire », que les biologistes de la Synthèse Moderne utilisaient beaucoup plus souvent que le terme « hasard », n'est pas défini et, par conséquent, laisse ces formulations ouvertes à des interprétations diverses dont certaines erronées. Les formulations du deuxième groupe semblent plus claires à ce sujet car elles substituent au terme « aléatoire » d'autres expressions qui apparemment en révèlent le sens ; cependant, dans le cadre de ces formulations reste quand même une légère ambiguïté en ce qui concerne les éléments précis entre lesquels il n'existe aucun lien ou dépendance et surtout quant à la nature statistique ou causale de ce lien qui manque entre eux. Enfin, les formulations du troisième groupe sont beaucoup plus révélatrices que les précédentes car elles expriment de manière suffisamment claire comment les biologistes de la Synthèse Moderne ne concevaient pas les mutations génétiques, c'est-à-dire pas selon une vision et une explication lamarckiennes de l'origine et du caractère de la variation héréditaire. Les formulations de ce dernier groupe se révéleront donc très précieuses dans le cadre de notre enquête qui consiste à fournir une bonne définition de la notion de hasard correspondant à la vision consensuelle de la Synthèse Moderne au sujet des mutations génétiques et qui vise à

³³ « *Mutations are undirected with respect to the adaptive needs of the species. [...] It may seem a deplorable imperfection of nature that mutability is not restricted to changes that enhance the adaptedness of their carriers. However, only a vitalist Pangloss could imagine that the genes know how and when it is good for them to mutate* », Dobzhansky, 1970, *Genetics of the Evolutionary Process*, p. 92 (notre traduction).

³⁴ « *Mutations never seem to happen in response to a need for a particular genetic change, nor is a particular mutation released by a particular constellation of environmental factors* », Mayr, 1976, *Evolution and the Diversity of Life*, p. 32 (notre traduction).

comprendre si cette notion est mise en cause par la découverte relativement récente de mécanismes régulant le taux de mutation en relation avec les conditions environnementales. En effet, ce n'est que sur la base de l'identification et de la définition de la vision consensuelle de la Synthèse Moderne quant aux mutations génétiques qu'il est possible de déterminer si les avancées en génétique moléculaire, en particulier la découverte de certains parmi les mécanismes moléculaires de mutation dits « mutateurs », représentent un véritable défi au paradigme de la Synthèse Moderne, en particulier au caractère aléatoire de toute mutation génétique.

4.2.1. Les « biais de mutation » à l'époque de la Synthèse Moderne

La première étape de notre quête de la vision consensuelle du caractère aléatoire des mutations chez les biologistes de la Synthèse Moderne consiste à rappeler et à souligner ce que ces derniers savaient et reconnaissaient quant aux mutations génétiques, en particulier en ce qui concerne leurs causes possibles, leur probabilité d'occurrence et leurs caractéristiques relativement aux conditions environnementales et au processus d'évolution. Les sources bibliographiques auxquelles nous renvoyons pour toute information sur ces sujets correspondent aux ouvrages principaux des acteurs de la Synthèse Moderne³⁵.

Tout d'abord, en ce qui concerne l'origine causale des mutations génétiques, à la lumière des recherches de la génétique mendélienne, les biologistes de la Synthèse Moderne reconnaissaient que les mutations peuvent être « spontanées », c'est-à-dire dues à des erreurs dans le processus de copie du matériel génétique ou simplement dues à des causes inconnues, ou « induites » par des agents physico-chimiques mutagènes (par exemple, des radiations, des changements de température et des substances chimiques). Suite à la découverte de l'ADN en tant que support du matériel génétique autour des années 1950, ils avaient aussi reconnu que certains sites du génome sont plus mutables que d'autres à cause des caractéristiques physico-chimiques de la séquence nucléotidique ; par exemple, ils savaient que la probabilité qu'une mutation se produise est plus élevée dans des régions de la séquence riches en cytosine à

³⁵ Fisher, 1999[1930], *The Genetical Theory of Natural Selection: A complete variorium edition* ; Wright, 1931, « Evolution in Mendelian Populations » ; Haldane, 1966[1932], *The Causes of Evolution* ; Dobzhansky, 1951[1937], *Genetics and the Origin of Species* ; 1970, *op. cit.* ; Huxley, 1948[1942], *Evolution. The Modern Synthesis.* ; 1953, *Evolution in Action.* ; Mayr, 1942, *Systematics and the Origin of Species.* ; 1974[1970], *Populations, espèces et évolution* ; 1976, *op. cit.* ; Simpson, 1984[1944], *Tempo and Mode in Evolution*, 1953, *The Major Features in Evolution* ; Stebbins, 1971, *Processes of Organic Evolution* ; 1982, *Darwin to DNA, Molecules to Humanity*; Dobzhansky *et al.*, 1977, *Evolution.*

cause de l'instabilité chimique de cette base nucléotidique. De plus, toujours à la lumière de l'étude des mutations suivant la découverte de la double hélice, ils avaient aussi admis que certains agents mutagènes ont tendance à provoquer des mutations dans certaines régions du génome plutôt que d'autres ; par exemple, dans le cas des radiations ultraviolettes, ces dernières tendent à produire des mutations dans des séquences nucléotidiques riches en répétitions de thymine.

En ce qui concerne la probabilité des mutations au niveau du génome, même avant la découverte de la nature et de la structure physico-chimique du matériel génétique, les biologistes de la Synthèse Moderne reconnaissaient qu'une mutation n'a pas la même probabilité de se produire le long du génome: ils admettaient que le taux de mutation varie selon le gène concerné. Un peu plus tard, ils avaient aussi reconnu l'existence de séquences génomiques beaucoup plus mutables que d'autres (au moins dix fois plus), qui furent observées pour la première fois par le physicien et biologiste américain Benzer à la fin des années 1950³⁶ et qui sont encore aujourd'hui appelées « points chauds de mutation ». Selon les biologistes de la Synthèse Moderne, le taux de mutation ne différait pas simplement selon le gène concerné, mais aussi en fonction de la présence d'agents mutagènes pouvant augmenter le taux global de mutation et relativement à l'étape du développement où l'organisme individuel concerné se trouve. De plus, ils savaient aussi que des types différents de mutations ont des probabilités différentes de se produire, soit dans le génome entier soit dans un gène particulier.

En ce qui concerne la probabilité des mutations au niveau des organismes concernés et non à l'échelle du génome, les biologistes de la Synthèse Moderne reconnaissaient que le taux de mutation diffère entre espèces différentes, entre différentes souches de la même espèce et même entre organismes appartenant à la même souche. De plus, ils admettaient que ces différences étaient le résultat à la fois de l'histoire évolutive commune à tout organisme vivant et de l'histoire évolutive particulière à chaque espèce, à chaque souche ou à chaque population d'organismes vivants. De la même manière, ils pensaient aussi que la mutabilité du génome est une caractéristique déterminée par la structure physico-chimique du matériel génétique qui, à son tour, est le résultat d'un long processus évolutif dans lequel l'environnement et la sélection naturelle ont joué un rôle majeur. Les biologistes de la Synthèse Moderne reconnaissaient donc que le taux de mutation est contrôlé génétiquement,

³⁶ Benzer & Freese, 1958, « Induction of specific mutations with 5-bromouracil » ; Benzer 1959, « On the topology of genetic fine structure » ; 1961, « On the topology of genetic fine structure » ; 1962, « The fine structure of the gene ».

parfois même par des gènes précis qui augmentent la mutabilité et qu'ils appelaient « modificateurs ». C'est pourquoi certains d'entre eux avançaient même l'hypothèse selon laquelle des mutations pourraient se produire de manière dirigée, non pas du point de vue de leur valeur adaptative, mais en ce qui concerne l'endroit de leur occurrence et leur abondance, c'est-à-dire dans des sites déterminés du génome et avec une probabilité spécifique³⁷. Cette possibilité, même si elle n'était qu'hypothétique, les faisait douter davantage quant au caractère strictement aléatoire de toute mutation génétique³⁸. Cependant, ils tendaient toujours à être plutôt prudents à ce sujet et à mettre en avant leur ignorance de la physiologie des gènes et de la chimie du processus de mutations.

Cela nous amène à focaliser notre attention sur ce que les biologistes de la Synthèse Moderne savaient et reconnaissaient quant aux mutations génétiques du point de vue plus spécifiquement évolutionnaire. En ce qui concerne le taux de mutation, ils admettaient donc qu'il est sous contrôle génétique et qu'il peut évoluer (augmenter ou diminuer) en étant l'objet de la sélection naturelle selon les conditions environnementales dans lesquelles les organismes concernés se trouvent. L'idée d'une régulation adaptative du taux de mutation relativement à l'environnement commençait donc à prendre forme mais encore de manière hypothétique et parfois controversée³⁹. Au sujet de la valeur adaptative des mutations génétiques, les biologistes de la Synthèse Moderne ne pensaient pas qu'une mutation a la même probabilité d'être favorable, défavorable et neutre, mais qu'elle a plus de chances d'avoir une certaine valeur adaptative plutôt qu'une autre en dépendance du contexte génomique où la modification génétique se produit et des conditions environnementales où l'organisme individuel porteur se trouve. Ils avaient aussi compris que la probabilité de diverses valeurs adaptatives d'une mutation change selon l'importance de la modification produite au niveau du génome : Fisher⁴⁰ avait démontré mathématiquement que plus une mutation consiste en une petite modification du matériel génétique plus il y a de chances qu'elle soit favorable, alors que plus une mutation consiste en un changement important du génome plus il y a de chances qu'elle soit défavorable et même létale.

Enfin, sur la base des estimations des taux de mutation, les biologistes de la Synthèse Moderne affirmaient que, d'une part, dans un environnement stable et homogène il est beaucoup plus probable que des mutations nouvelles soient défavorables que favorables ou neutres car un organisme est normalement bien adapté au milieu où il se trouve. D'autre part,

³⁷ Dobzhansky, 1951[1937], *Genetics and The Origin of Species*, p. 62-63.

³⁸ Mayr, 1942, *Systematics and the Origin of Species*, p. 68.

³⁹ Dobzhansky, 1951[1937], *op. cit.*, p. 60-61.

⁴⁰ Fisher, 1999[1930], *The Genetical Theory of Natural Selection: a complete variorium edition*.

lorsqu'un organisme change d'environnement ou lorsque l'environnement où il se trouve subit des modifications, la probabilité relative que des mutations nouvelles soient favorables augmente par rapport à la situation environnementale précédente, car l'organisme se trouve dans un environnement auquel il n'est pas bien adapté.⁴¹

Les biologistes de la Synthèse Moderne admettaient donc toutes ces caractéristiques propres au phénomène de mutation génétique – que nous proposons d'appeler des « biais de mutation » – au niveau génomique, ensuite moléculaire, et du point de vue du processus évolutif ; en même temps, ils ne renonçaient pas à l'idée selon laquelle toute mutation génétique est « aléatoire » ou le fruit du « hasard ». De quelle manière considéraient-ils les mutations génétiques comme des phénomènes aléatoires ? C'est la question de savoir quelle est la notion de hasard invoquée dans le cadre de la Synthèse Moderne pour caractériser les mutations ; sa réponse consiste à identifier et à définir la vision consensuelle de la Théorie Synthétique à ce sujet.

La deuxième étape de notre quête de la vision consensuelle du caractère aléatoire des mutations consiste à prendre en considération une situation empirique admise par les biologistes de la Synthèse Moderne pour ensuite analyser les trois groupes de formulations à la lumière de cette situation et de tout ce que les biologistes savaient et reconnaissaient à cette époque quant aux mutations génétiques. Une telle analyse simultanée des résultats empiriques sur l'occurrence des mutations et des définitions verbales du caractère aléatoire de ces dernières nous semble un parcours potentiellement efficace afin de comprendre, d'abord, comment les biologistes de la Synthèse Moderne ne pouvaient pas concevoir la notion de hasard dans la caractérisation des mutations génétiques car elle se serait révélée en contradiction avec certaines situations empiriques connues et admises ; ensuite, comment ils la concevaient effectivement.

Considérons donc la situation empirique, admise par les biologistes de la Synthèse Moderne, où des conditions environnementales adverses – par exemple, un soudain changement de température – provoquent une augmentation du taux de mutation dans tout le génome des organismes individuels d'une population donnée. D'abord, il est intéressant de remarquer que, dans la situation décrite, il y a un lien causal entre les facteurs de

⁴¹ Ce dernier point reconnu par les biologistes de la Synthèse Moderne nous semble très discutable et, en tout cas, nécessite des études expérimentales ultérieures pour être corroboré. En effet, ce genre de généralisation nous semble simpliste car elle évacue le caractère complexe de chaque situation particulière. Plus précisément, elle ne semble pas prendre en compte le fait que la valeur adaptative d'une mutation ne dépend pas simplement de l'état adapté ou non adapté de l'organisme individuel concerné à l'environnement où il se trouve. La valeur adaptative d'une mutation dépend d'un ensemble de facteurs parmi lesquels figurent l'altération se produisant au niveau du génome, le contexte génomique où elle se produit, la variation que la mutation provoque au niveau du phénotype de l'organisme porteur et les conditions environnementales où ce dernier se trouve.

l'environnement, qui peuvent être considérés comme des facteurs sélectifs, et l'augmentation de la probabilité qu'une mutation se produise. En effet, les biologistes de la Synthèse Moderne admettaient la possibilité que les mutations soient induites par des agents mutagènes de l'environnement et que ces derniers puissent provoquer une augmentation globale du taux de mutation. Dans le cas particulier décrit ci-dessus, le soudain changement de température, qui fait partie de l'environnement et qui joue le rôle de facteur de sélection, est impliqué dans l'origine causale de l'augmentation du taux de mutation.

En deuxième lieu, dans ce genre de situation où l'environnement change et de manière subite, un taux élevé de mutation peut se révéler avantageux pour les organismes individuels concernés : en effet, par rapport à une situation où les mutations se produisent moins fréquemment, un taux élevé de mutation fournit beaucoup de variabilité et augmente donc la probabilité qu'une mutation favorable se produise dans un laps de temps plus court. Cela implique l'augmentation de la probabilité que les organismes concernés arrivent à survivre et à se reproduire, et donc à transmettre la variation génétique favorable à leur descendance, plus rapidement que si leur taux de mutation était faible. Les biologistes de la Synthèse Moderne reconnaissaient ce point : ils admettaient donc la possibilité qu'il existe au moins une corrélation, qu'elle exprime ou non un lien causal, entre la probabilité des mutations et le besoin de variabilité de la population concernée.⁴²

En troisième lieu, si les biologistes synthétistes étaient interrogés au sujet de la situation décrite, ils affirmeraient que, comme le soudain changement des conditions environnementales constitue une situation de stress pour les organismes individuels concernés, la probabilité qu'une mutation soit avantageuse est plus élevée que la même probabilité mais dans une situation environnementale stable et homogène. Ce point, qui porte sur le lien entre la valeur adaptative des mutations et l'environnement où l'organisme concerné se trouve, était largement reconnu par les biologistes de la Synthèse Moderne, en particulier par Simpson⁴³. Ce dernier affirmait qu'une mutation affectant un organisme individuel mal adapté à son environnement, ou situé dans un environnement changeant, a plus de chances d'être avantageuse qu'une mutation affectant un organisme bien adapté à son

⁴² Il est important de mettre en évidence que, même si dans la situation décrite ci-dessus une mutation avantageuse a plus de chances de se produire dans un laps de temps plus court, une telle augmentation du taux de mutation n'implique pas nécessairement que la probabilité qu'une mutation soit favorable est plus grande que la probabilité qu'une mutation soit défavorable ou que la probabilité qu'une mutation soit neutre. La probabilité relative qu'une mutation soit favorable peut être la même que dans une situation où le taux de mutation est plus bas : en effet, même si beaucoup plus de mutations se produisent, le rapport entre le nombre de mutations avantageuses possibles et le nombre de mutations défavorables et neutres possibles ne varie pas nécessairement, mais peut être le même que dans une situation caractérisée par un taux bas de mutation.

⁴³ Simpson 1984[1944], *Tempo et Mode in Evolution*, p. 55-56.

environnement ou situé dans un environnement stable. Cependant, nous soulignons encore une fois que cette affirmation nous semble controversée et en besoin d'analyses ultérieures des données expérimentales à ce sujet.

La question qui nous intéresse est la suivante : les mutations se produisant dans une telle situation à cause d'un changement soudain des conditions environnementales étaient-elles aléatoires selon les biologistes de la Synthèse Moderne et, si c'est bien le cas, dans quel sens ? La réponse à cette question se situe au croisement entre les éléments mis en évidence par la situation empirique que nous avons décrite, ce que ces biologistes savaient et reconnaissaient quant aux mutations génétiques et les formulations diverses du caractère aléatoire des mutations qu'ils ont mobilisées.

4.2.2. La vision consensuelle de la Synthèse Moderne quant au hasard des mutations

D'abord, considérons le premier groupe de formulations et choisissons-en une qui soit représentative de ce genre de définition du caractère aléatoire des mutations. La formulation proposée par Dobzhansky selon laquelle « une mutation est un processus aléatoire par rapport aux besoins adaptatifs de l'espèce »⁴⁴ a le mérite de faire explicitement référence au processus à l'origine des mutations, et non pas seulement au résultat de ce processus et au caractère aléatoire des mutations produites du point de vue de leur valeur adaptative. Cette manière de définir le caractère aléatoire des mutations nous paraît intéressante et positive. En effet, une définition qui ne porte que sur le résultat du processus mutationnel – par exemple, sur le caractère aléatoire de la distribution des mutations le long du génome ou de la distribution du nombre de mutations au niveau d'un gène, du génome et même d'une population donnée – ne fournit aucune explication, mais une simple description du *pattern* probabiliste qu'on observe au niveau du résultat. Au contraire, c'est n'est qu'en définissant la notion de hasard relativement au processus de mutation et non pas seulement au niveau de son résultat qu'on peut, d'abord, espérer d'expliquer quelle est l'origine du caractère aléatoire de la distribution des mutations et, en second lieu, poser et essayer de répondre à la question de la signification et du rôle appropriés de la notion de hasard lorsqu'elle est utilisée pour caractériser les mutations génétiques.⁴⁵ C'est pourquoi nous prenons en considération et examinons la

⁴⁴ Dobzhansky, 1970, *Genetics of The Evolutionary Process*, p. 92.

⁴⁵ En effet, il est important de remarquer qu'il peut y avoir une relation qu'on pourrait appeler d'« un-à-plusieurs » entre un résultat observé et les processus pouvant l'engendrer. De manière générale, un même

formulation de Dobzhansky à la lumière de ce que les biologistes de la Synthèse Moderne savaient et acceptaient quant aux mutations génétiques et de la situation empirique décrite plus haut.

D'abord, cette formulation ne fournit pas de réponse claire à la question de la notion de hasard utilisée dans le cadre de la définition du caractère aléatoire des mutations à cause de l'ambiguïté du terme « aléatoire ». Si l'on comprend ce terme comme s'il faisait référence à la notion de hasard en tant qu'équiprobabilité⁴⁶ et à la notion d'indépendance⁴⁷ (statistique et causale), Dobzhansky semble nous dire deux choses : d'abord, qu'une mutation est un processus aléatoire s'il tend à produire avec la même probabilité des mutations favorables, des mutations défavorables et des mutations neutres ; en second lieu, qu'une mutation est un processus aléatoire s'il n'existe aucun type de lien, ni statistique ni causal, entre sa direction adaptative – autrement dit, la valeur adaptative des résultats que le processus de mutation produit – et les besoins adaptatifs de l'espèce concernée. Cependant, cette compréhension de la notion de hasard n'est pas admissible dans le cadre de cette formulation car cette dernière entrerait en contradiction avec les résultats empiriques obtenus et reconnus à l'époque de la Synthèse Moderne.

Nous rappelons que l'idée de l'équiprobabilité qu'une mutation soit favorable, défavorable ou neutre ne figurait pas parmi ce que biologistes de la Théorie Synthétique pensaient quant aux mutations génétiques ; donc, le hasard en tant qu'équiprobabilité ne correspond pas à la notion que les biologistes utilisaient pour définir le caractère aléatoire des mutations. Nous rappelons aussi que les biologistes de la Synthèse Moderne affirmaient que la probabilité qu'une mutation soit favorable peut augmenter dans des conditions environnementales adverses auxquelles un organisme individuel n'est pas bien adapté et où la variabilité peut donc se révéler décisive pour sa survie et sa reproduction ; ils reconnaissaient donc la possibilité d'un lien, du moins statistique, entre la valeur adaptative des mutations et les besoins adaptatifs des organismes porteurs. L'idée d'indépendance sans aucune

résultat aléatoire (par exemple, une certaine distribution aléatoire des mutations le long du génome) peut être dû à des processus divers qui peuvent être déterministes ou stochastiques (par exemple, la même distribution aléatoire des mutations le long du génome peut être due à un phénomène quantique considéré comme fondamentalement indéterministe ou à la rencontre déterministe entre la séquence nucléotidique et un agent mutagène de nature physique ou chimique). De plus, le caractère déterministe ou stochastique d'un processus peut recevoir des interprétations épistémologiques et métaphysiques diverses. Par exemple, il peut être considéré comme une propriété intrinsèque et objective du processus, comme une caractéristique du modèle représentant ce processus, ou comme une propriété subjective donc liée simplement à la connaissance et à la compréhension que nous en avons (du processus).

⁴⁶ Au sujet de la notion de hasard en tant qu'équiprobabilité, voir le chapitre 2, section 2.1.

⁴⁷ Voir le Glossaire.

spécification ultérieure n'est donc non plus une idée à laquelle les biologistes pensaient lorsqu'ils parlaient du caractère aléatoire des mutations.

Nous pouvons aboutir aux mêmes conclusions en analysant cette même formulation à la lumière de la situation que nous avons brièvement décrite plus haut et que les biologistes de la Synthèse Moderne considéraient comme possible. En particulier, cette situation empirique met en évidence le fait que ces biologistes admettaient, d'abord, la possibilité d'un lien causal entre les facteurs sélectifs de l'environnement et la probabilité des mutations. En deuxième lieu, la possibilité d'un lien, du moins statistique, entre la probabilité qu'une mutation se produise et le besoin de variabilité de la population concernée. Enfin, le fait que, lorsque l'environnement change, la probabilité qu'une mutation soit avantageuse pour l'organisme concerné peut être plus grande par rapport à la même probabilité mais dans une situation où l'environnement est stable. Sur la base de toutes ces considérations à propos de la formulation proposée par Dobzhansky et appartenant au premier des trois groupes que nous avons distingués, nous concluons que les biologistes de la Synthèse Moderne n'employaient, ni la notion mathématique de hasard au sens d'équiprobabilité, ni la notion d'indépendance, pour décrire le caractère aléatoire des mutations.

Considérons maintenant deux formulations appartenant au deuxième groupe :

« Les mutations sont des changements se produisant *indépendamment* du fait qu'ils sont ou qu'ils pourraient être potentiellement adaptatifs »⁴⁸ ; « Les mutations sont des changements *aléatoires* car elles se produisent *indépendamment* du fait qu'elles sont favorables ou défavorables [...] »⁴⁹.

Nous n'examinerons pas en détail ces formulations, qui ont été proposées, la première par Dobzhansky dans les années 1930, la deuxième par Dobzhansky avec Ayala, Stebbins et Valentine après le début des recherches en biologie moléculaire mais toujours en restant fidèles au paradigme de la Théorie Synthétique. En effet, il est assez évident que leur analyse à la lumière des résultats expérimentaux dont les biologistes de la Synthèse Moderne disposaient déjà au cours de la première moitié du siècle dernier et sur la base de la situation empirique décrite plus haut nous amènerait à la même conclusion que l'examen du premier groupe de formulations. De manière plus explicite, le terme « aléatoire » ne peut pas être interprété selon la notion mathématique de hasard au sens d'équiprobabilité⁵⁰ et selon la

⁴⁸ Dobzhansky, 1951[1937], *Genetics and The Origin of Species*, p. 74.

⁴⁹ Dobzhansky *et al*, 1977, *Evolution*, p. 66.

⁵⁰ De plus, quelques lignes au-dessus de la deuxième formulation, Dobzhansky et ses collègues (*ibid.*, p. 65) écrivent explicitement que « les mutations ne sont pas aléatoires dans le sens où n'importe quelle mutation aurait la même probabilité de se produire que n'importe quelle autre [...] » (notre traduction), « *Mutations are not*

notion d'indépendance (statistique et causale) car une telle interprétation ferait entrer en contradiction ces deux formulations avec ce que les biologistes savaient et reconnaissaient au sujet des mutations à l'époque de la Synthèse Moderne.

Avant de nous occuper de l'analyse du troisième groupe de formulations avancées par les biologistes de la Synthèse Moderne pour définir le caractère aléatoire des mutations, il nous semble intéressant de souligner que, de la même manière que la formulation de Dobzhansky appartenant au premier groupe que nous avons analysée plus haut, les deux formulations proposées par Dobzhansky et par Dobzhansky avec Ayala, Stebbins et Valentine ont le mérite de mettre l'accent sur le processus de production des mutations génétiques. Ces formulations semblent aussi nous fournir, de manière plus explicite, un élément ultérieur pour comprendre le caractère des mutations génétiques : c'est le processus à l'origine des mutations qui est indépendant de la valeur adaptative (favorable, défavorable, neutre) de ces dernières ; plus précisément, il s'agit de l'absence d'une certaine relation de dépendance causale entre la valeur adaptative d'une mutation particulière et la probabilité qu'elle soit produite par le processus mutationnel.

Considérons enfin quelques formulations du dernier groupe qui, comme nous l'avons déjà remarqué, consistent en des définitions négatives du caractère aléatoire des mutations : elles opposent la conception des mutations de la Synthèse Moderne à la vision et à l'explication lamarckiennes de l'origine et du caractère de la variation héréditaire (et de l'origine du cours adaptatif de l'évolution). L'analyse de ce dernier groupe de formulations nous permettra de mieux comprendre et de clarifier ce que les formulations du deuxième groupe que nous venons de considérer semblaient nous suggérer : la notion de hasard utilisée par les biologistes de la Synthèse Moderne pour caractériser les mutations génétiques désigne le manque d'une certaine relation de dépendance causale qui va de la valeur adaptative des mutations au processus qui les engendre.

La formulation suivante proposée par Simpson⁵¹ est particulièrement éclairante :

« D'autre part, lorsqu'il est appliqué à la mutation le terme 'aléatoire' fait souvent référence au *manque de correspondance* de l'effet phénotypique avec le stimulus et avec la direction actuelle ou adaptative de l'évolution. Des mutations induites par la chaleur *ne produisent pas* un changement phénotypique lié à la tolérance à la chaleur. »

random in the sense that any mutation would be equally likely to occur as any other mutation [...] ».

⁵¹ Simpson, 1953, *The Major Features of Evolution*, p. 87.

Cette manière de définir et d'expliquer le caractère aléatoire des mutations ne semble entrer en contradiction, ni avec la situation décrite plus haut où un soudain changement de l'environnement provoque une augmentation du taux de mutation, ni plus généralement avec ce que les biologistes de la Synthèse Moderne savaient et reconnaissaient quant aux mutations génétiques. En effet, Simpson ne nie pas le fait qu'il puisse y avoir un lien, du moins statistique, entre la probabilité qu'une mutation se produise et sa valeur adaptative ou entre la probabilité qu'une mutation soit favorable et les besoins adaptatifs des organismes individuels concernés. De plus, juste au-dessus de ce passage, il affirme explicitement que le terme « aléatoire » dans la description des mutations ne peut pas être utilisé de manière réaliste selon les notions d'équiprobabilité et d'indépendance⁵².

Simpson définit le caractère aléatoire des mutations génétiques en regardant à leur origine causale : il affirme que, même si une mutation est le résultat d'un processus causal qui implique un facteur sélectif de l'environnement (par exemple, un soudain changement de température), elle n'est pas provoquée de manière spécifique et en vue de l'adaptation de l'organisme individuel concerné et de sa descendance aux conditions environnementales où ils se trouvent. Il n'y a donc pas de relation spécifique qui va de l'ordre des causes à l'ordre des effets entre, d'une part, la valeur adaptative d'une mutation se produisant dans un organisme individuel situé dans un environnement donné et, d'autre part, le fait que cette mutation se produise dans cet environnement. En termes probabilistes, la définition du caractère aléatoire des mutations génétiques proposée par Simpson désigne donc l'absence d'un lien causal et spécifique entre la probabilité qu'une mutation soit favorable (dans un environnement donné) et la probabilité qu'elle se produise (dans cet environnement).

Une formulation proposée par Dobzhansky et appartenant au même groupe que celle de Simpson que nous venons d'analyser exprime cette même idée. Dobzhansky écrit :

« Une situation idéale serait que les organismes puissent répondre aux défis posés par les changements de l'environnement en ne produisant que des mutations favorables où et quand ils

⁵² « Les mutations ne sont pas aléatoires au sens plein et habituel du mot ou de la même manière que les premiers Darwiniens considéraient de manière non réaliste la variation disponible à la sélection naturelle comme complètement aléatoire » ; « il y a, d'une part, l'aléatoire en ce qui concerne où et quand une mutation se produira » : les mutations peuvent être considérées comme « statistiquement aléatoires », en dépit du fait que leur source n'est « pas une réaction complètement aléatoire » (Simpson, 1953, *op. cit.*, p. 86-87, notre traduction), « *mutations are not random in the full and usual sense of the word or in the way some early Darwinists unrealistically considered as fully random the variation available for natural selection* » ; « *there is, on the one hand, a randomness as to where and when a mutation will occur [...] statistically random [...] not a wholly random reaction* »).

[les organismes] en ont besoin. Mais la nature n'a pas été assez généreuse pour fournir à ses créations une habileté si providentielle. »⁵³

Une telle manière d'expliquer ce que les mutations génétiques ne sont pas et de définir donc de manière négative leur caractère aléatoire en opposition à la vision lamarckienne de l'origine et du caractère de la variation (et de l'origine de l'adaptation) est parfaitement cohérente avec ce que les biologistes de la Synthèse Moderne admettaient quant aux mutations, en particulier en ce qui concerne leur origine causale et les différences de leur taux d'occurrence dans des situations environnementales et sélectives diverses. Nous concluons donc que ces deux formulations, l'une de Simpson et l'autre de Dobzhansky, sont celles qui expriment de la manière la plus fidèle la vision consensuelle de la Synthèse Moderne quant au caractère aléatoire des mutations qui, au cours de la deuxième moitié du siècle dernier, est devenue un objet de consensus au sein de la communauté des biologistes. De plus, cette manière de caractériser les mutations génétiques correspond à la conception de Darwin des variations « non dirigées » du point de vue évolutionnaire que nous avons présenté au début de ce chapitre.⁵⁴

4.3. La notion de « hasard évolutionnaire »

Nous suggérons de parler de la notion de « hasard évolutionnaire »⁵⁵ pour désigner cette conception du caractère aléatoire des mutations génétiques qui correspondait à la vision des biologistes de la Synthèse Moderne et qui correspond présentement au consensus au sein de la communauté des biologistes. Nous proposons de définir cette notion en reformulant la définition de « mutation aléatoire » que la philosophe américaine Millstein⁵⁶ a proposée et qui l'oppose à la notion, apparemment lamarckienne, de « mutation dirigée »⁵⁷: une mutation

⁵³ Dobzhansky, 1951[1937], *Genetics and The Origin of Species*, p. 74.

⁵⁴ Nous laissons au lecteur le loisir de retourner aux autres formulations du troisième groupe afin d'en apprécier la clarté par rapport à celles du deuxième et du premier groupes et d'évaluer leur correspondance avec l'idée de Darwin des variations « non dirigées » du point de vue évolutionnaire.

⁵⁵ Dans un article très intéressant sur les notions de hasard en biologie évolutionnaire et en paléobiologie, Eble (1999, « On the dual nature of chance in evolutionary biology and palaeology ») a proposé cette même expression pour désigner la notion de hasard mobilisée par Darwin pour caractériser les variations du point de vue évolutionnaire.

⁵⁶ Millstein, 1997, *The Chances of Evolution: An Analysis of the Roles of Chance in Microevolution and Macroevolution*.

⁵⁷ « [...] une mutation est dirigée si et seulement si elle est causée par un stress environnemental de manière spécifique et exclusivement adaptative », *ibid.*, p. 151, (notre traduction), « [...] a mutation is directed if and only if it is specifically caused by environmental stress in an exclusively adaptive manner ». Dans le chapitre

génétique est le fruit du hasard du point de vue évolutionnaire – ou est le fruit du « hasard évolutionnaire » – si et seulement si elle n'est pas causée de manière spécifique et (exclusivement)⁵⁸ adaptative par un processus physico-chimique en réponse aux conditions de l'environnement.

D'une part, une mutation est causée « de manière spécifique » si et seulement si elle fait partie d'une augmentation du taux de mutation locale, c'est-à-dire ciblée dans une région ou un site particulier du génome, et non globale, c'est-à-dire concernant toute la séquence génomique. D'autre part, une mutation est produite « de manière (exclusivement) adaptative » si et seulement si sa probabilité en tant que mutation avantageuse est nettement plus élevée que la probabilité d'autres mutations défavorables ou neutres dans un même environnement donné, cela étant dû aux caractéristiques du processus causal qui l'engendre. Une mutation est donc produite « de manière spécifique et (exclusivement) adaptative » – autrement dit, une mutation est « dirigée » – si et seulement si elle fait partie d'une augmentation locale du taux de mutation et le processus physico-chimique qui la cause rend sa probabilité en tant que mutation avantageuse nettement plus élevée que la probabilité d'autres mutations défavorables ou neutres dans le même environnement. Par conséquent, comme Millstein⁵⁹ l'a déjà souligné, si une mutation n'est pas causée par un processus physico-chimique de manière spécifique, ou si une mutation est causée par un processus physico-chimique de manière spécifique mais non (exclusivement) adaptative, ou si une mutation n'est pas causée par un processus physico-chimique en réponse à un changement de l'environnement, alors elle n'est

suivant, nous mettrons particulièrement en évidence le fait que, pour qu'une mutation « dirigée » puisse être considérée comme due à un processus véritablement lamarckien, il ne suffit pas qu'elle soit produite par un processus physico-chimique de manière spécifique et (exclusivement) adaptative en réponse aux conditions de l'environnement. De plus, sa « directionnalité » spécifique et adaptative doit reposer sur une tendance interne à l'organisme individuel porteur et non pas sur un mécanisme de mutation qui est le résultat adaptatif de l'histoire évolutive et sélective de la population concernée.

⁵⁸ Le terme « exclusivement » ne doit pas être interprété de manière littérale et n'est pas essentiel dans le cadre de la définition de la notion de « hasard évolutionnaire ». Millstein (1997, *op. cit.*) l'utilise simplement afin d'éviter un des problèmes caractérisant, selon son analyse, la définition « faible » de « mutation dirigée » proposée par Sarkar (1991a, « Lamarck *contre* Darwin, *Reduction versus* Statistics: Conceptual Issues in the Controversy Over Directed Mutagenesis in Bacteria », p. 239). En effet, en affirmant qu'une mutation est dirigée si (et non « si et seulement si ») elle se produit (ou se produit plus souvent) dans un environnement où le phénotype qui lui est associé a une fitness plus élevée, Sarkar risque de considérer comme dirigée une mutation qui se révèle avantageuse dans un environnement donné mais qui n'était pas nettement plus probable que des mutations neutres ou défavorables. La définition de « mutation dirigée » de Sarkar risque aussi de s'appliquer à une mutation qui se produit plus souvent dans un environnement où elle est avantageuse que dans un environnement où elle ne l'est pas, mais non pas à cause de son caractère favorable ou sans que cela soit dû aux caractéristiques du processus causal qui l'engendre (par exemple, à un processus physico-chimique qui la cause de manière spécifique et adaptative en réponse aux conditions de l'environnement). En réalité, ces mutations favorables ne sont pas « dirigées » mais le fruit du « hasard évolutionnaire » car dues à un processus causal « non dirigé » du point de vue du cours adaptatif de l'évolution. Par la suite, le lecteur pourra remarquer que la limite de la définition de Sarkar consiste à ne considérer qu'une seule des deux conditions pour que, selon notre analyse, une mutation soit dirigée et à ne pas justifier cette seule condition.

⁵⁹ Millstein, 1997, *op. cit.*, p. 151.

pas dirigée mais aléatoire du point de vue évolutionnaire au sens de la notion de « hasard évolutionnaire ».

Cette même notion de « hasard évolutionnaire » relativement aux mutations génétiques peut être exprimée autrement en explicitant d'abord les conditions pour qu'une mutation soit « dirigée » et en définissant ensuite, de manière négative, les conditions pour qu'une mutation soit le fruit du « hasard évolutionnaire ». Une mutation est une « mutation dirigée » si et seulement si elle remplit les deux conditions (nécessaires et suffisantes) suivantes : 1) elle est plus probable dans un environnement E_1 où elle est avantageuse que dans un environnement E_2 où elle est défavorable ou neutre ; 2) elle est nettement plus probable dans un environnement donné E_1 où elle est avantageuse que d'autres mutations neutres ou défavorables.⁶⁰ La condition 1) porte sur la probabilité d'une même mutation dans des situations environnementales différentes : elle renvoie au fait que le processus causal produisant une mutation « dirigée » est à l'origine d'une augmentation ciblée du taux de mutation dans la région ou dans le site où la mutation adaptative potentielle devrait se produire. La condition 2) porte sur la probabilité d'une mutation avantageuse par rapport à la probabilité d'autres mutations défavorables ou neutres dans une même situation environnementale donnée: elle renvoie au fait que le processus causal à l'origine d'une mutation « dirigée » tend à produire la mutation qui est adaptative relativement aux conditions environnementales où l'organisme concerné se trouve.⁶¹

⁶⁰ La définition de « mutation dirigée » proposée par Millstein, et à partir de laquelle nous définissons de manière négative la notion de « hasard évolutionnaire » dans le cas des mutations génétiques, rappelle la définition de « mutation dirigée » proposée par Lenski et Mittler (1993, « The Directed Mutation Controversy and neo-Darwinism »). En effet, ces derniers affirment qu'une mutation dirigée est une mutation qui « se produit avec un taux plus élevé spécifiquement quand (et même car) elle est avantageuse pour l'organisme concerné, alors que des augmentations comparables du taux ne se produisent pas i) dans le même environnement pour des mutations semblables mais qui ne sont pas avantageuses, et ii) pour la même mutation dans des environnements similaires où elle n'est pas avantageuse », p. 188 (notre traduction), « [...] *occurs at a higher rate specifically when (and even because) it is advantageous to the organism, whereas comparable increases in rate do not occur i) in the same environment for similar mutations that are not advantageous, and ii) for the same mutation in similar environments where it is not advantageous* ». Les conditions i) et ii) proposées par Lenski et Mittler correspondent respectivement aux conditions 2) et 1) de notre définition. En revanche, la notion de « hasard évolutionnaire » que nous proposons met en avant de manière plus explicite le rôle central de la causalité et, plus précisément, de la manière dans laquelle une mutation est causée, afin de comprendre si une mutation est dirigée ou le fruit du hasard du point de vue de l'évolution (du « hasard évolutionnaire »).

⁶¹ Dans les chapitres suivants, nous montrerons et soulignerons que la notion de « mutation dirigée » ainsi définie ne représente qu'un aspect de la conception lamarckienne de l'origine et du caractère de la variation et n'est donc pas suffisante pour la caractériser. En effet, nous avons déjà rappelé que la notion de « directionnalité » soutenue par Lamarck était beaucoup plus forte que celle mobilisée dans la caractérisation des « mutations dirigées » que nous venons de présenter et à laquelle les biologistes de la Synthèse Moderne s'opposaient. Lamarck affirmait que les nouvelles caractéristiques acquises par les organismes au cours de leur développement déterminent le cours de l'évolution et sont à l'origine de l'adaptation, à la fois au niveau individuel et au niveau de l'espèce, sans qu'aucun processus de sélection ne soit impliqué.

Sur la base de cette caractérisation de ce qu'est une mutation « dirigée », nous pouvons donc formuler les conditions pour qu'une mutation soit le produit du « hasard évolutif » de la manière suivante : une mutation est le fruit du « hasard évolutif » si et seulement si elle n'est pas une mutation « dirigée », autrement dit, si et seulement si elle ne remplit pas au moins une des deux conditions suivantes : 1) être plus probable dans un environnement E_1 où elle est avantageuse que dans un environnement E_2 où elle est défavorable ou neutre ; 2) être nettement plus probable dans un environnement donné E_1 où elle est avantageuse que d'autres mutations neutres ou défavorables.

Notre analyse nous a donc amené à comprendre que, selon la Synthèse Moderne et selon la plupart des biologistes aujourd'hui, même si un stimulus environnemental est impliqué dans la détermination causale d'une mutation, cette dernière est aléatoire au sens de la notion de « hasard évolutif » si et seulement s'il n'y a pas de lien causal entre la valeur adaptative de la mutation produite dans une situation donnée et les besoins adaptatifs, dans cette situation environnementale, des organismes individuels concernés. Par conséquent, selon cette définition, qui est focalisée sur la valeur adaptative du résultat phénotypique dû à une altération de la séquence génomique, les mutations ne sont pas aléatoires par rapport aux besoins adaptatifs des organismes porteurs au sens de la notion de hasard en tant qu'équiprobabilité et de la notion d'indépendance (statistique et causale). Au contraire, dans ce cadre la notion de hasard assume plutôt un sens spécifique à la biologie et désigne le caractère « non dirigé » du processus causal à l'origine des mutations par rapport à l'adaptation : les mutations qui sont le fruit du « hasard évolutif » ne sont pas le résultat d'un processus physico-chimique qui les cause de manière spécifique dans « la direction de l'amélioration de l'adaptation »⁶².

Nous concluons ce chapitre en soulignant que la notion de « hasard évolutif » nous semble appropriée du point de vue conceptuel car sa définition est consistante et semble évoquer certaines des notions de hasard qui ont marqué l'histoire de la pensée philosophico-scientifique de l'antiquité jusqu'à aujourd'hui. Dans le dernier chapitre de cette partie (7), nous analyserons en détail cette relation et, plus généralement, la pertinence conceptuelle de la notion de « hasard évolutif ». Nous soulignons aussi que cette notion semble appropriée du point de vue empirique car elle décrit et explique le caractère aléatoire des mutations sans entrer en contradiction avec ce que nous savons de ces phénomènes moléculaires ou, du moins, sans contredire ce que les biologistes de la Synthèse Moderne

⁶² « [...] *the direction of improved adaptation* », Ridley, 2004, *Evolution*, p. 89 (notre traduction).

savaient et reconnaissaient quant aux mutations génétiques. Nous consacrerons les deux chapitres suivants (5 et 6) à l'analyse de la pertinence empirique de la notion de « hasard évolutionnaire » dans le cas des mutations dues à certains des mécanismes dits « mutateurs » qui ont été découverts au cours des trente dernières années et qui, selon un certain nombre de biologistes et de philosophes de la biologie, posent un défi au paradigme de la Synthèse Moderne.

Chapitre 5 – Avancées et débats sur l’origine et le caractère des mutations génétiques, de la Synthèse Moderne jusqu’à nos jours

5.1. Du test de fluctuation de Luria et Delbrück (1943) à l’hypothèse des « bactéries lamarckiennes » (Cairns *et al*, 1988)

Durant les années 1930, l’idée propre à la Synthèse Moderne selon laquelle toute mutation génétique est le fruit du hasard du point de vue évolutionnaire¹ était un postulat accepté de manière consensuelle par les généticiens et les biologistes de l’évolution, en particulier en ce qui concerne l’origine de la variation héréditaire chez les organismes multicellulaires. Cependant, cette idée n’était encore appuyée par aucune preuve expérimentale directe². La première étude prétendant fournir une confirmation décisive à cette caractérisation des mutations génétiques vit le jour dans le cadre des expériences sur les bactéries, alors que l’origine de la variation chez ces microorganismes, et plus généralement chez les organismes unicellulaires, représentait encore un cas controversé entre la vision lamarckienne et la conception de la Synthèse Moderne.

5.1.1. La preuve fournie par le test de fluctuation

En 1943, Luria et Delbrück³ observèrent que, si une population d’*Escherichia coli* est étalée sur une boîte de Petri contenant des virus (des phages lytiques), des bactéries résistantes à l’infection virale apparaissent après quelques heures. Ils distinguèrent donc deux hypothèses possibles pour expliquer ce phénomène. Selon la première (a), qu’ils appelèrent l’« hypothèse de l’immunité acquise et héréditaire » et qui représentait la vision lamarckienne de l’origine, du caractère et de la transmission de la variation, l’apparition de bactéries

¹ Sur la base de notre analyse de la notion de hasard mobilisée par les biologistes de la Synthèse Moderne pour caractériser les mutations génétiques (voir le chapitre précédent), nous parlons de la vision consensuelle de la Synthèse Moderne quant au caractère des mutations génétiques pour désigner l’idée du « hasard évolutionnaire » (et vice versa).

² Sniegowski & Lenski, 1995, « Mutation and Adaptation: The Directed Mutation Controversy in Evolutionary Perspective ».

³ Luria & Delbrück, 1943, « Mutations of Bacteria From Virus Sensitivity to Virus Resistance ».

résistantes aux phages lytiques était le résultat de mutations induites de manière spécifique et adaptative par la présence des agents sélectifs (les virus). On supposait donc que ces mutations étaient apparues, non pas en phase de croissance des populations, mais après l'étalement des bactéries dans le milieu sélectif létal. Selon la deuxième hypothèse (b), que Luria et Delbrück appelèrent l'« hypothèse de la mutation » et qui était censée représenter la conception de la Synthèse Moderne, la présence des phages lytiques n'avait aucun rôle dans l'apparition des bactéries résistantes. On supposait donc que les mutations génétiques à l'origine de la variation héréditaire pour la résistance aux virus étaient, dans ce sens, spontanées car indépendantes des agents sélectifs : elles étaient apparues en phase de croissance de la population de bactéries avant son étalement dans le milieu sélectif létal.

Luria et Delbrück procédèrent au test de fluctuation⁴ afin de comprendre, sur la base des données statistiques sur la distribution du nombre de bactéries résistantes, si les mutations s'étaient effectivement produites avant ou après l'exposition de la population de bactéries aux virus, autrement dit si elles étaient susceptibles d'une explication lamarckienne ou d'une explication conforme au paradigme de la Synthèse Moderne. En effet, ils avaient aussi déterminé la distribution attendue du nombre de mutants respectivement dans le cas des deux hypothèses. D'une part, si l'« hypothèse de l'immunité acquise et héréditaire » (a) se révélait être la bonne, c'est-à-dire si les mutations génétiques conférant la résistance aux bactéries étaient apparues après l'exposition aux virus car induites par ces derniers, la distribution du nombre de mutants aurait dû être une distribution de Poisson, donc caractérisée par une moyenne égale à la variance. D'autre part, si l'« hypothèse de la mutation » (b) était, au contraire, la bonne, c'est-à-dire si ces mutations étaient apparues en phase de croissance avant que les bactéries ne soient étalées sur le milieu sélectif létal sans qu'aucune induction de la part des virus intervienne, cette même distribution aurait dû être caractérisée par une variance beaucoup plus grande que celle impliquée par une distribution de Poisson, donc par un rapport entre la variation et la moyenne plus grand que 1.

⁴ Pour une description du test de fluctuation réalisé par Luria et Delbrück, nous renvoyons au chapitre 3.

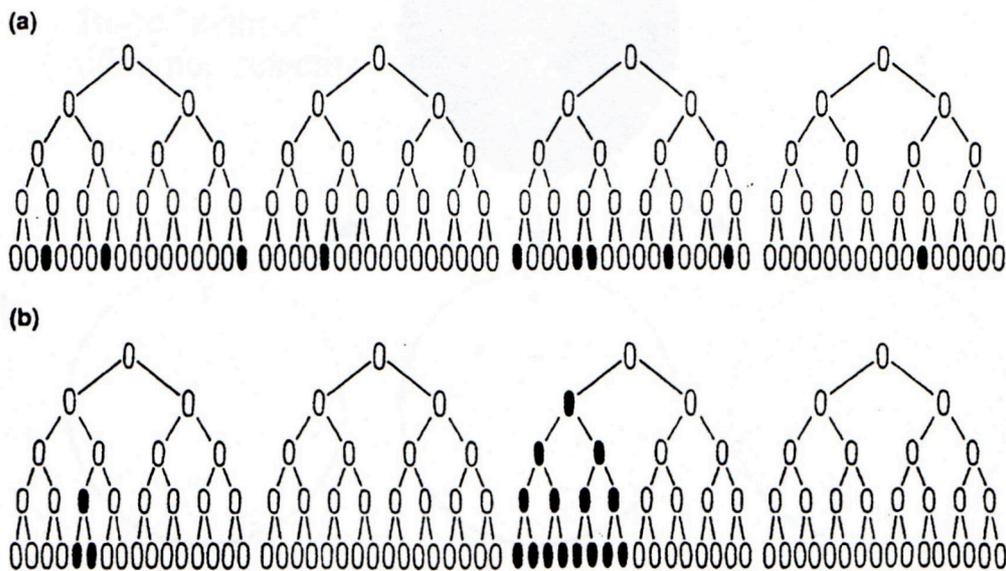


Figure 1 : Le test de fluctuation de Luria et Delbrück (1943). Distribution des mutants (représentés par des cellules pleines) dans 4 populations distinctes, fondées par 4 bactéries. a) Hypothèse de l'« immunité acquise et héréditaire » : la distribution du nombre de mutants suit une loi de Poisson ; b) Hypothèse de la « mutation » : la distribution du nombre de mutants est caractérisée par une variance beaucoup plus grande que celle prévue par la loi de Poisson. Figure tirée de Taddei, 1995, *Environnement et contrôle de la variabilité génétique chez Escherichia coli*, p. 23.

La distribution que Luria et Delbrück observèrent était caractérisée par une variance nettement supérieure à la moyenne. Par conséquent, ils conclurent, en faveur du paradigme de la Synthèse Moderne, que les mutations génétiques avaient eu lieu en phase de croissance des populations de bactéries, avant l'exposition de ces dernières au milieu sélectif contenant des phages lytiques, et donc de manière spontanée au sens où la présence des virus n'avait joué aucun rôle inducteur dans leur occurrence.

Le résultat obtenu par Luria et Delbrück à travers la démarche expérimentale et l'analyse statistique décrites ci-dessus fut immédiatement considéré par la communauté des biologistes comme une preuve de la validité de la vision consensuelle de la Synthèse Moderne quant au caractère des mutations génétiques⁵. Durant les années 1940 et 1950, d'autres expériences, reposant sur des démarches expérimentales différentes mais toujours basées sur l'analyse statistique de la distribution du nombre de mutants ou d'autres caractères phénotypiques des bactéries, furent réalisées. Leurs résultats confirmèrent la conclusion de Luria et Debrück et furent considérés à leur tour comme des preuves ultérieures de la vision

⁵ L'expérience de Luria et Delbrück en 1943 est habituellement considérée comme la première à fournir une preuve du fait que les mutations génétiques se produisent au hasard du point de vue évolutionnaire, bien qu'auparavant, plus précisément en 1934, Yang et White fussent arrivés à cette même conclusion sur la base de l'observation des variants dans des colonies de *Vibrio cholerae*. Voir Sarkar, 1991a, « Lamarck contre Darwin, Reduction versus Statistics: Conceptual Issues in the Controversy Over Directed Mutagenesis in Bacteria ».

consensuelle de la Synthèse Moderne quant au fait que les mutations génétiques se produisent au hasard du point de vue évolutionnaire. Parmi ces travaux expérimentaux figurent les expériences de ré-étalement réalisées par Newcombe (1949)⁶, les expériences en chemostat (c'est-à-dire dans des conditions où la concentration des nutriments est maintenue constante) effectuées par Novick et Szilard (1950)⁷, les expériences de répliques sur boîtes réalisées par Lederberg et Lederberg (1952)⁸ et les expériences de sélection indirecte de Cavalli-Sforza et Lederberg (1956).⁹

5.1.2. Les limites des preuves en faveur de la vision consensuelle de la Synthèse Moderne

Bien qu'à l'époque les résultats obtenus par Luria et Delbrück, et par les réalisateurs des expériences suivantes, fussent considérés comme des preuves en faveur de la vision consensuelle de la Synthèse Moderne quant au caractère des mutations génétiques héréditaires, il est important de souligner les limites propres à deux aspects centraux, à la fois méthodologiques et théoriques, à la base de ces travaux expérimentaux.

D'abord, il est intéressant de se demander si le test de fluctuation et, plus généralement, toute démonstration reposant sur l'analyse de données statistiques, représentent véritablement une bonne méthode scientifique pour décider si les mutations génétiques se produisent avant toute pression sélective ou à cause de cette dernière. Sarkar¹⁰ a soulevé cette problématique et est arrivé à la conclusion que la simple analyse des données statistiques sur la distribution du nombre de mutants ne peut fournir aucune solution conclusive à la controverse sur le caractère des mutations génétiques. Il affirme que le test de fluctuation, comme toute analyse statistique, est une méthode scientifique non réductionniste visant à comprendre ce qui se produit à un niveau inférieur – à l'intérieur des bactéries individuelles, au niveau moléculaire du processus physico-chimique à l'origine causale des mutations – à travers l'observation de ce qui se produit à un niveau supérieur – au niveau du phénotype des bactéries individuelles et de la population dont elles font partie. C'est pourquoi, selon Sarkar,

⁶ Newcombe, 1949, « Origin of bacterial variants ».

⁷ Novick & Szilard, 1950, « Experiments with the chemostat on spontaneous mutations of bacteria ».

⁸ Lederberg & Lederberg, 1952, « Replica plating and indirect selection of bacterial mutants ».

⁹ Cavalli-Sforza & Lederberg, 1956, « Isolation of pre-adaptive mutants in bacteria by sib-selection ». Pour une analyse détaillée de toutes ces expériences diverses, voir Sarkar, 1991a, *op. cit.* ; Taddei, 1995, *Environnement et contrôles de la variabilité génétique chez Escherichia coli* ; Sniegowski & Lenski, 1995, « Mutation and Adaptation: The Directed Mutation Controversy in Evolutionary Perspective » ; Jablonka & Lamb, 1995, *Epigenetic Inheritance and Evolution*.

¹⁰ Sarkar, 1991a, *op. cit.*

cette méthode scientifique n'est pas appropriée : ce n'est que grâce à la découverte d'un mécanisme¹¹ de mutation pouvant fournir une explication causale de la production des mutations au niveau moléculaire qu'il est possible de soutenir de manière fondée et rigoureuse l'hypothèse lamarckienne quant à l'origine de la variation héréditaire.

Sarkar¹² rappelle les nombreuses limites du test de fluctuation, en particulier la variété de facteurs pouvant affecter la distribution finale du nombre de mutants et pouvant donc induire en erreur les biologistes en ce qui concerne l'origine et le caractère des mutations¹³. En particulier, il focalise sa critique du test de fluctuation, et de toute méthode statistique utilisée pour déterminer le caractère des mutations génétiques, sur son utilisation par les partisans de l'hypothèse lamarckienne des mutations « dirigées ». En effet, ces derniers, dans les années qui suivirent la publication des résultats de Luria et Delbrück, répétèrent le test de fluctuation et interprétèrent toute déviation de la distribution du nombre de mutants observée par Luria et Delbrück vers une distribution de Poisson comme une preuve du caractère « dirigé », du point de vue évolutionnaire, des mutations génétiques.

Nous rejoignons la critique de Sarkar et, de plus, nous soulignons qu'elle doit aussi être adressée à Luria et Delbrück et à tout biologiste prétendant démontrer par le test de fluctuation, ou par d'autres méthodes semblables basées sur l'analyse statistique, que les mutations génétiques se produisent au hasard du point de vue évolutionnaire. En ce qui concerne la démarche de Luria et Delbrück, il est en effet important de rappeler que ces derniers ne regardèrent pas le mécanisme moléculaire à l'origine des mutations. Ils examinèrent simplement la distribution statistique du nombre de mutants dans des populations de bactéries et, sur cette base, conclurent que les mutations s'étaient produites en phase de croissance, avant l'exposition des bactéries au substrat sélectif, et que les virus n'étaient donc pas impliqués dans leur origine causale. Cette méthode ne nous semble pas appropriée pour comprendre de manière concluante si les mutations sont des phénomènes qui se produisent au hasard du point de vue évolutionnaire. En effet, comme nous l'avons déjà souligné dans le chapitre précédent, un même résultat statistique – dans ce cas, une même distribution du nombre de mutants dans une population donnée de bactéries – peut être dû à des processus

¹¹ Nous utilisons le terme « mécanisme » pour désigner la manière dont un processus, c'est-à-dire une suite temporelle d'états liés causalement, se déroule au cours du temps, de ses conditions initiales jusqu'à son état final. Pour une définition de ce qu'est un mécanisme, nous renvoyons à Machamer *et al*, « Thinking About Mechanisms », 2000, p. 3, p. 6.

¹² Sarkar, 1991a, « Lamarck *contre* Darwin, Reduction *versus* Statistics: Conceptual Issues in the Controversy Over Directed Mutagenesis in Bacteria », p. 253-255.

¹³ Sarkar souligne notamment l'effet que peuvent avoir sur la distribution, par exemple, une différence de fitness entre mutants et non mutants, le décalage phénotypique, l'efficacité de l'étalement sur la boîte de Petri et la présence de mutations multiple et de la sélection.

moléculaires différents, constitués par des facteurs physico-chimiques divers, et peut donc être interprété et expliqué en invoquant des mécanismes caractérisés par des propriétés moléculaires et par une signification évolutionnaire différentes. Par conséquent, de la même manière que les partisans du caractère « dirigé » des mutations génétiques par rapport à l'adaptation, tout biologiste visant à montrer que les mutations sont le fruit du hasard du point de vue évolutionnaire ne peut pas se limiter à l'interprétation des données statistiques dont il dispose. Il doit aussi imaginer un mécanisme plausible pouvant expliquer le résultat statistique qu'il observe – un mécanisme expliquant l'origine de la distribution statistique du nombre de mutants caractérisée par une variance beaucoup plus grande que la moyenne –, le décrire à la fois en ce qui concerne ses caractéristiques moléculaires et du point de vue de son rôle au sein du processus évolutif, et ensuite en fournir une corroboration expérimentale¹⁴.

En second lieu, il est aussi intéressant de s'interroger sur la valeur des résultats obtenus dans le cadre de ces études expérimentales, qui furent réalisées dans les années 1940-1950 afin de démontrer empiriquement la vision consensuelle de la Synthèse Moderne quant aux mutations génétiques. En ce qui concerne ce deuxième point, nous focalisons notre attention sur le test de fluctuation de Luria et Delbrück. En dépit de la critique que nous venons d'adresser ci-dessus à l'utilisation de cette méthode d'analyse statistique dans ce cadre, supposons que la distribution de Luria et Delbrück sur le nombre de mutants soit véritablement une preuve du fait que les mutations génétiques se sont produites en phase de croissance avant l'étalement des bactéries dans le milieu sélectif, donc qu'elles n'ont pas été induites par la présence des virus. Se pose une question centrale dans le cadre de notre enquête sur la notion de hasard : ce fait, correspond-il véritablement à la vision consensuelle de la Synthèse Moderne au sujet du caractère des mutations génétiques ? En d'autres termes, correspond-il à l'idée selon laquelle les mutations se produisent au hasard par rapport au cours adaptatif de l'évolution au sens précis de la notion que nous avons appelée « hasard évolutionnaire » ?¹⁵

Rappelons d'abord le résultat de l'analyse historique et conceptuelle que nous avons menée dans le chapitre précédent et qui concerne la notion de hasard utilisée par les biologistes de la Théorie Synthétique de l'Évolution pour caractériser les mutations génétiques. Dans le cadre de la Synthèse Moderne, toute mutation est considérée comme le fruit du « hasard évolutionnaire » au sens où elle n'est pas causée de manière spécifique et

¹⁴ Sarkar 1991a, *op. cit.*, p. 255.

¹⁵ Nous pourrions aussi poser la question de savoir si l'hypothèse de l'« immunité acquise et héréditaire » formulée par Luria et Delbrück correspond véritablement à la vision lamarckienne de l'origine de la variation, ce que nous avons décidé de ne pas traiter explicitement.

(exclusivement) adaptative par un processus physico-chimique en réponse aux conditions de l'environnement. Cette conception ne semble pas exclure que des mutations puissent se produire lorsque les organismes concernés se trouvent sous la pression d'un facteur de sélection ; la notion de « hasard évolutionnaire » requiert simplement que les mutations ne soient pas dirigées par un processus physico-chimique en réponse aux forces sélectives en action et en vue de l'adaptation des organismes porteurs.

La comparaison de la vision consensuelle de la Synthèse Moderne avec la conclusion de Luria et Delbrück nous révèle que cette dernière est parfaitement compatible avec la notion de « hasard évolutionnaire », mais qu'elle est caractérisée par une extension différente. L'idée selon laquelle toute mutation se produit au hasard, non seulement au sens où elle se produit de manière non dirigée par rapport aux besoins des organismes porteurs, mais aussi au sens où elle a lieu avant que ces derniers soient confrontés avec des agents sélectifs, est plus exigeante, car moins étendue, que l'idée selon laquelle toute mutation n'est pas causée de manière spécifique et (exclusivement) adaptative par un processus physico-chimique en réponse aux conditions de l'environnement. En effet, la première idée nie la possibilité que des mutations se produisent lorsqu'un organisme se trouve dans un milieu sélectif, alors que la notion de « hasard évolutionnaire » ne l'exclut pas.

Nous concluons donc que les méthodes statistiques utilisées par Luria et Delbrück et par de nombreux biologistes autour des années 1940-1950 ne sont pas appropriées pour décider si les mutations génétiques se produisent de manière « dirigée » ou au hasard du point de vue évolutionnaire. De plus, ce que Luria et Delbrück ont prétendu avoir démontré est compatible avec le paradigme de la Synthèse Moderne mais ne correspond pas à la vision de cette dernière quant au caractère des mutations génétiques, c'est-à-dire au « hasard évolutionnaire » des mutations. Plutôt, Luria et Delbrück ont démontré ce que nous suggérons d'appeler une interprétation « naïve » de la conception des mutations propre à la Synthèse Moderne, et qui est malheureusement assez répandue depuis plus d'un demi siècle au sein de la communauté des biologistes. C'est une interprétation qui se base sur la manière simplifiée et idéalisée dans laquelle le processus de mutation génétique était – et est encore aujourd'hui – représenté dans le cadre des modèles de l'évolution en génétique des populations¹⁶. Elle ne

¹⁶ En génétique des populations, plus précisément dans le cadre des modèles de l'évolution des populations naturelles, le processus de mutation génétique est représenté par un facteur (u) désignant le taux de mutation. Ce dernier correspond à la probabilité qu'une mutation se produise, par exemple au niveau d'un locus particulier, à chaque génération. Lorsque les généticiens des populations appliquent un modèle théorique à un cas particulier pour lequel ils disposent de données expérimentales, la valeur qu'ils attribuent au taux de mutation dépend des hypothèses suivantes. En premier lieu, une mutation génétique est considérée comme un événement assez rare, donc le taux de mutation est relativement bas (de l'ordre de 10^{-5} en ce qui concerne les mutations visibles au

correspond pas à la vision consensuelle de la Synthèse Moderne quant à l'origine et au caractère des mutations génétiques car elle ne prend pas en compte tous les biais de mutation que les biologistes de la Synthèse Moderne reconnaissaient, à la fois du point de vue moléculaire et évolutionnaire¹⁷, mais qu'ils ne considéraient pas dans le cadre des descriptions formelles (mathématiques) du processus évolutif pour des raisons pratiques de modélisation.

5.2. Le défi lancé au paradigme de la Synthèse Moderne : des mutations « dirigées » (1988) aux mutations « adaptatives » (1990-2008)

À la fin des années 1980, un groupe de biologistes de Harvard (John Cairns, Julie Overbaugh et Stephan Miller) prétendit remettre en question les résultats des expériences réalisées dans les années 1940-1950 et qui étaient considérés comme des preuves concluantes en faveur de la Synthèse Moderne. En particulier, ces biologistes prétendirent avoir obtenu des données contredisant le résultat du test de fluctuation réalisé par Luria et Delbrück et donc mettant en cause le paradigme de la Synthèse Moderne au sujet du caractère des mutations génétiques. Bien que l'argument qu'ils avancèrent en faveur du caractère « dirigé », selon eux au sens lamarckien, des mutations génétiques fût fondamentalement basé sur l'analyse de données statistiques, plus précisément sur la déviation de la distribution du nombre de mutants qu'ils observèrent de la distribution de Luria et Delbrück, nous verrons qu'ils proposèrent aussi des mécanismes moléculaires, plus ou moins plausibles, afin de renforcer et de rendre plus précise leur hypothèse quant à l'origine de la variation génétique.

Cairns et ses collègues¹⁸, en rapportant des résultats apparemment en contradiction avec le paradigme dominant¹⁹, déclenchèrent un débat controversé qui, malgré les preuves

niveau phénotypique et de l'ordre 10^{-9} au niveau nucléotidique). En deuxième lieu, la probabilité qu'une mutation se produise dans un locus particulier est considérée comme indépendante de la présence d'autres mutations dans d'autres loci. En troisième lieu, la probabilité de l'occurrence d'une mutation par locus par génération est considérée comme constante durant la croissance de la population concernée. En quatrième lieu, aucune mutation ne se produit lorsqu'une population se trouve sous une pression de sélection. Sur la base de ces hypothèses, la distribution des mutations génétiques au sein d'une population donnée est supposée suivre une loi de Poisson (voir le chapitre 3).

¹⁷ Nous renvoyons au chapitre 4.

¹⁸ Cairns *et al*, 1988, « The origin of mutants ».

¹⁹ Quelques années plutôt, les expériences réalisées par Shapiro (1984, « Observations on the formation of clones containing araB-lacZ cistron fusion ») avec une souche de bactéries *Escherichia coli* avaient déjà fourni des résultats pouvant suggérer l'occurrence de mutations « dirigées » par rapport à l'adaptation et qui étaient donc apparemment en contradiction avec le paradigme de la Synthèse Moderne.

expérimentales que les biologistes ont fournies au cours des années en faveur de la conception de l'origine des mutations propre à la Synthèse Moderne, perdure encore aujourd'hui bien que sous des termes différents. Ils suggérèrent que les bactéries individuelles sont capables de « sentir » l'environnement et de produire des mutations bénéfiques, c'est-à-dire de diriger la production des mutations génétiques en relation aux conditions environnementales où elles se trouvent afin de survivre et de se reproduire : c'est ce qu'on a appelé l'hypothèse des « mutations dirigées » ou, de manière en partie impropre, l'hypothèse des « bactéries lamarckiennes ».

De manière plus précise, Cairns et ses collègues répétèrent le test de fluctuation qui avait été effectué par Luria et Delbrück en 1943. Ils n'imposèrent pas de sélection létale qui, selon eux, avait empêché Luria et Delbrück de détecter les mutations causées en réponse à la pression de sélection, plus précisément qui avaient été induites par la présence des phages lytiques. Ils étalèrent les bactéries sur des boîtes de Petri où la seule source de carbone était le lactose et que les individus de la souche d'*Escherichia coli* qu'ils utilisèrent ne pouvaient dégrader et utiliser qu'en mutant (de Lac⁻ à Lac⁺)²⁰. Ils analysèrent la distribution obtenue du nombre de mutants et ils observèrent que la variance était beaucoup plus faible que celle qui avait été observée par Luria et Delbrück en présence d'un agent sélectif létal. Sur cette base, Cairns et ses collègues conclurent que la distribution du nombre de mutants dans des conditions de sélection non létale était due à la fois à des mutations ayant eu lieu au cours de la croissance de la population de bactéries – donc avant leur étalement sur le milieu sélectif riche en lactose – et à des mutations ayant eu lieu après l'exposition des bactéries à l'agent sélectif – donc produites de manière « dirigée » par un mécanisme moléculaire en réponse aux besoins adaptatifs des bactéries. À leur avis, il s'agissait d'une distribution hybride, autrement dit de la somme d'une distribution de Luria et Delbrück et d'une distribution de Poisson du nombre de mutants.

Cairns et ses collègues crurent donc avoir montré que l'idée du paradigme de la Théorie Synthétique, selon laquelle toute mutation génétique est le produit du hasard du point de vue évolutionnaire, est erronée. Leurs résultats suggérèrent que, dans certains cas, l'agent sélectif est indispensable au déclenchement du processus causal à l'origine des mutations et que certaines mutations génétiques peuvent être produites de manière « dirigée », c'est-à-dire

²⁰ Shapiro (1984, *op. cit.*) avait déjà observé que, lorsque des bactéries se trouvent dans un milieu où la seule source de carbone est le lactose ou l'arabinose, des mutations, plus précisément des excisions, permettant aux bactéries d'utiliser le lactose ou l'arabinose (de Lac(Ara)⁻ à Lac(Ara)⁺) se produisent après plusieurs jours, alors qu'elles n'ont pas lieu lorsque les bactéries sont étalées dans d'autres milieux.

de manière spécifique et (exclusivement) adaptative, par rapport au cours adaptatif de l'évolution.²¹

Nous avons déjà anticipé que Cairns et ses collègues imaginèrent et décrivent de manière plus ou moins détaillée quelques mécanismes moléculaires pouvant être à l'origine causale des mutations « dirigées ». Plus précisément, afin d'expliquer l'origine causale des mutations, à leur avis « dirigées », ayant lieu après l'étalement des bactéries dans un milieu où la seule source de carbone est le lactose et leur permettant de survivre et de se reproduire, ils proposèrent deux mécanismes apparemment lamarckiens : la « transcription reverse spécifique » et la « transcription reverse non spécifique »²².

Le premier mécanisme repose sur la production d'un ensemble très variable de molécules d'ARN messager et sur la successive transcription reverse en ADN de celle appropriée du point de vue adaptatif, c'est-à-dire de la molécule d'ARN messager impliquée dans la production de la protéine capable de dégrader le lactose et permettant donc à la bactérie porteuse d'utiliser cette source de carbone, de survivre et de se reproduire. Ce mécanisme requiert la présence d'une organelle spéciale contenant de la transcriptase reverse et d'un facteur de monitoring de la fonction des protéines produites à partir des molécules d'ARN messager. Le deuxième mécanisme proposé par Cairns et ses collègues repose aussi sur la production d'un ensemble très variable de molécules d'ARN et sur la présence de la transcriptase reverse ; cependant, aucun facteur de monitoring du rôle fonctionnel des protéines n'est impliqué. Dans ce cas, la transcription reverse n'est activée que lors que les bactéries commencent à se reproduire, c'est-à-dire lorsque la protéine fonctionnellement appropriée pour dégrader et utiliser le lactose a été synthétisée et a donc permis aux bactéries porteuses de survivre et de se reproduire dans un milieu où la seule source de carbone est le lactose. Cairns et ses collègues imaginèrent aussi d'autres mécanismes possibles à l'origine des « mutations dirigées » – par exemple, le mécanisme des « copies variantes d'ADN » – mais ils n'en proposèrent pas de description détaillée.

²¹ Nous rappelons que, pour qu'une mutation génétique « dirigée » du point de vue évolutionnaire puisse être considérée comme due à un processus (véritablement) lamarckien, son caractère spécifique et adaptatif doit trouver son origine dans un mécanisme physiologique qui se déclenche en réponse à un changement de l'environnement, et non pas dans un mécanisme adaptatif apparu au cours de l'histoire évolutive et sélective de la population à laquelle l'organisme concerné appartient. Cependant, dans le cadre du débat sur la possibilité des mutations « dirigées » chez les bactéries, aucun biologiste n'a jamais considéré et discuté cet aspect jouant un rôle fondamental dans l'attribution d'une mutation génétique à un processus lamarckien. C'est pourquoi nous caractériserons comme « apparemment lamarckiens » les mécanismes proposés par les partisans de l'hypothèse des « mutations dirigées ».

²² Durant les années 1970, Temin découvrit la transcription reverse : il considérait ce processus moléculaire, qu'il opposa au dogme central de la biologie moléculaire, comme un mécanisme évolutif lié au fonctionnement. Voir Temin & Mizutani, 1970, « RNA-dependent DNA polymerase in virions of Rous sarcoma virus ».

Quelques années plus tard, Cairns et Foster²³ reconnurent le manque de corroborations expérimentales du mécanisme de « transcription reverse spécifique », dont l'équipe de Cairns avait davantage imaginé les détails physico-chimiques. Ils proposèrent donc un autre mécanisme moléculaire pouvant être à l'origine des mutations « dirigées », du point de vue évolutionnaire, chez les bactéries situées dans un environnement sélectif non léthal où la seule source de carbone est le lactose : ils l'appelèrent « amplification non aléatoire des gènes mutants bénéfiques ». C'est encore une fois un mécanisme apparemment lamarckien : il repose sur la production de beaucoup de copies du gène qui, en mutant, rend les bactéries capables d'utiliser le lactose en tant que nutriment et leur permet donc de survivre et de se reproduire. Le rôle de l'amplification du gène dont la mutation est avantageuse réside dans le fait d'augmenter de manière considérable la probabilité de cette dernière par la création d'une large cible pour le processus de mutation.²⁴

5.2.1. Le débat controversé sur le caractère « dirigé », puis « adaptatif », des mutations génétiques

L'article que Cairns et ses collègues publièrent en 1988 fit éclater un grand débat qui se déroula durant les années 1990 et qui n'est pas encore complètement résolu aujourd'hui. Cette controverse se nourrit de la répétition des expériences présentées par l'équipe de Cairns et de nombreuses expériences semblables qui furent réalisées surtout avec des bactéries²⁵ : elle vit s'opposer les partisans de l'hypothèse des mutations « dirigées » (ou des « bactéries lamarckiennes ») et les partisans de la vision consensuelle de la Synthèse Moderne quant au caractère des mutations génétiques. Bien que les limites des analyses statistiques réalisées par Luria et Delbrück (1943) et par d'autres microbiologistes dans les années qui suivirent aient commencé à être l'objet de réflexions critiques²⁶, initialement les positions des acteurs du débat déclenché par l'article de l'équipe de Cairns (1988) reposaient simplement sur

²³ Cairns & Foster, 1991, « Adaptive Reversion of a Frameshift Mutation in *Escherichia coli* ».

²⁴ Pour plus de détails sur ce dernier mécanisme et sur tous les autres mécanismes de mutation proposés par les partisans de l'idée de « mutation dirigée », voir Sarkar, 1991a, « Lamarck contre Darwin, Reduction versus Statistics: Conceptual Issues in the Controversy Over Directed Mutagenesis in Bacteria » ; Sniegoski & Lenski, 1995, « Mutation and Adaptation: The Directed Mutation Controversy in Evolutionary Perspective » ; Darden, 2006, *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly*.

²⁵ Pour une liste et une description détaillée de toutes ces expériences réalisées entre 1984 et 1995, voir les références dans la note précédente.

²⁶ Entre autres, voir Stewart *et al.*, 1990, « Fluctuation Analysis: The Probability Distribution of the Number of Mutants Under Different Conditions » ; Stewart, 1994, « Fluctuation Tests: How Reliable Are the Estimates of Mutation Rates? » ; Sarkar, 1991a, *op. cit.* ; 1991b, « Haldanes's Solution of the Luria-Delbrück Distribution ».

l'observation et l'interprétation des distributions du nombre de mutants. Ensuite, les études expérimentales sur le caractère des mutations et les échanges entre les deux camps opposés devinrent assez rapidement beaucoup plus intéressants car l'objet de recherche et de débat fut focalisé sur l'existence et sur les caractéristiques de mécanismes moléculaires plausibles, pouvant être à l'origine causale des mutations et des distributions observées des mutants.

Dans le cadre de la présente recherche, nous ne présenterons et n'examinerons pas en détail tous ces travaux expérimentaux et le débat controversé qu'ils nourrirent car ces événements ont plusieurs fois été très clairement exposés et analysés auparavant²⁷. Cependant, il est très important de rappeler que l'hypothèse des mutations « dirigées », avancée en 1988 par l'équipe de Cairns, n'a jamais trouvé de démonstrations expérimentales convaincantes. Ce manque de preuves a assez rapidement contraint Cairns lui-même et ses collaborateurs à reconnaître que les mutations favorables qu'ils avaient observées chez *Escherichia coli* n'étaient pas causées de manière spécifique et (exclusivement) adaptative en réponse aux forces sélectives en action²⁸. De plus, dans les années suivant immédiatement la publication de l'article de l'équipe de Cairns, les biologistes favorables à la Synthèse Moderne fournirent plusieurs explications alternatives d'un certain nombre de résultats expérimentaux pouvant suggérer l'existence de mutations « dirigées »²⁹ : tout en s'opposant à la vision lamarckienne de l'origine et du caractère de la variation héréditaire et en étant cohérents avec l'idée darwinienne, et propre à la Synthèse Moderne, du « hasard évolutionnaire » de toute mutation génétique³⁰, ces biologistes imaginèrent et décrirèrent un certain nombre de mécanismes moléculaires pouvant expliquer les résultats statistiques

²⁷ Sarkar, 1991a, *op. cit.* ; Lenski & Mittler, 1993, « The Directed Mutation Controversy and neo-Darwinism » ; Sniegowski & Lenski 1995, *op. cit.* ; Jablonka & Lamb, 1995, *Epigenetic Inheritance and Evolution* ; Millstein, 1997, *The Chances of Evolution: An Analysis of the Roles of Chance in Microevolution and Macroevolution* ; Darden, 2006, *op. cit.*

²⁸ Foster & Cairns, 1992, « Mechanisms of Directed Mutation ».

²⁹ Cairns *et al.*, 1988, « The Origin of Mutants » ; Cairns & Foster, 1991, *op. cit.* ; Hall, 1988, « Adaptive evolution that requires multiple spontaneous mutations. I. Mutations involving an insertion sequence » ; 1990, « Spontaneous point mutations that occur more often when advantageous than when neutral » ; Benson *et al.*, 1988, « Mutations that alter the pore function of the OmpF porin of *Escherichia coli* K12 » ; Dijkmans *et al.*, 1994, « Poisson-like fluctuation patterns of revertants of leucine auxotrophy (*leu-500*) in *Salmonella typhimurium* caused by delay in mutant cell division ».

³⁰ Voir Davis, 1989, « Transcriptional bias: A non-Lamarckian mechanism for substrate-induced mutations » ; Stahl, 1988, « A unicorn in the garden » ; 1992, « Unicorns revisited » ; Boe, 1990, « Mechanism of induction of adaptive mutations in *Escherichia coli* » ; Lenski *et al.*, 1989a, « Mutation and selection in bacterial population: Alternatives to the hypothesis of directed mutation » ; 1989b, « Another alternative to directed mutation ». Il faut quand même rappeler que, de la même manière que dans le cas des mécanismes suggérés par Cairns *et al.* (1988, *op. cit.*) et par Cairns et Foster (1991, *op. cit.*), un certain nombre de ces mécanismes conformes à la conception de la Synthèse Moderne au sujet du caractère des mutations n'a été l'objet d'aucune confirmation expérimentale concluante.

obtenus par les partisans des mutations « dirigées ». Par conséquent, déjà au cours des premières années 1990, les biologistes arrêtaient de parler de « mutations dirigées ».

L'idée des « mutations dirigées » survécut cependant sous le nom de « mutations adaptatives » : cette expression fut introduite par Cairns et Foster (1991) et n'est pas sans ambiguïté car elle « se situe inconfortablement entre une connotation lamarckienne et une connotation darwinienne »³¹. Le débat sur le caractère des mutations génétiques continua donc en voyant s'opposer, d'une part, les biologistes définissant les mutations « adaptatives » en opposition avec le paradigme de la Synthèse Moderne, mais qui ne parlaient pas forcément d'une origine lamarckienne de la variation ; d'autre part, les défenseurs de la compatibilité entre les mutations dites « adaptatives » et la vision orthodoxe de l'origine de la variation génétique.

En 1991, en collaboration avec Foster, Cairns publia³² les résultats d'expériences ultérieures qu'il avait réalisées, non pas en utilisant la même souche de bactéries *Escherichia coli* qu'il avait employée auparavant (SM195), mais la souche FC40 : d'autres biologistes (par exemple, Miller) avaient en effet observé et communiqué à Cairns et Foster que chez ces bactéries le taux de réversion de Lac⁻ vers Lac⁺ augmente après leur étalement dans un milieu contenant une quantité minimale de lactose. À partir de la publication de l'article de Cairns et Foster, les bactéries de cette souche (FC40) furent utilisées dans le cadre de nombreuses études sur les mutations « adaptatives » et devinrent l'un des systèmes biochimiques paradigmatiques du débat sur l'origine de ce genre de mutations et sur leur signification relativement au paradigme de la Synthèse Moderne³³.

Au cours des années 1990, des modèles mécanistiques variés pouvant expliquer l'occurrence de mutations favorables dans un milieu sélectif chez FC40 furent proposés ; des variantes récentes de ces modèles se confrontent encore aujourd'hui³⁴. Présentons-les de manière rapide afin d'en montrer la diversité.

Selon le modèle de l'hypermutation³⁵, le caractère apparemment dirigé des mutations se produisant lorsqu'une population de bactéries se trouve dans un milieu sélectif est dû à un mécanisme de contrôle qui, en réponse aux conditions de stress où ces microorganismes se trouvent, place un petit sous-ensemble de la population dans un état transitoire

³¹ « [...] sits uneasily between Lamarckian and Darwinian connotations », Sniegowski & Lenski, 1995, « Mutation and Adaptation: The Directed Mutation Controversy in Evolutionary Perspective », p. 566 (notre traduction).

³² Cairns & Foster, 1991, « Adaptive Reversion of a Frameshift Mutation in *Escherichia coli* ».

³³ Voir Foster, 1999, « Mechanisms of Stationary Phase Mutation: A Decade of Adaptive Mutation ».

³⁴ Voir Foster, 2004, « Adaptive Mutation in *Escherichia coli* » ; Roth & Andersson, 2004, « Adaptive Mutation: How Growth under Selection Stimulates Lac⁺ Reversion by Increasing Target Copy Number ».

³⁵ Hall, 1990, *op. cit.*

d'hypermutation (autrement dit, dans un état caractérisé par une augmentation globale du taux de mutation). Selon un modèle alternatif³⁶, les mutations favorables ayant lieu en présence d'agents sélectifs sont encore une fois des mutations se produisant dans une sous-population ; cependant, cette dernière n'est pas en état d'hypermutation comme dans le modèle précédent, mais en phase de croissance. Un troisième modèle³⁷ ne fait pas appel à une augmentation du taux de mutation pour expliquer l'occurrence de mutations favorables lorsqu'une population de bactéries se trouve dans un milieu sélectif ; il invoque l'activation du mécanisme d'amplification³⁸ des gènes en phase de croissance et l'occurrence des mutations dans le cadre d'un taux normal de mutation. Un dernier modèle³⁹ fait appel à l'activation des processus de réplication et de recombinaison et d'un mécanisme de réponse au stress (le système SOS, donc nous parlerons dans la section suivante) lorsqu'une population de bactéries se trouve dans une situation environnementale de stress (c'est-à-dire qu'elle se trouve confrontée avec des agents sélectifs) en phase stationnaire.

Aucun de ces modèles ne propose d'expliquer l'occurrence des mutations favorables dans un milieu sélectif par un mécanisme apparemment – et encore moins véritablement – lamarckien : à l'époque de leur première proposition, l'hypothèse des mutations « dirigées » allait être abandonnée et, par conséquent, aucun biologiste ne visait plus à démontrer que des mutations peuvent être produites de manière spécifique et (exclusivement) adaptative en réponse aux conditions sélectives de l'environnement. Cependant, ces modèles ne se positionnent pas tous de la même manière relativement au paradigme classique de la théorie de l'évolution. Bien que nous soutenions qu'ils reposent tous sur des mécanismes moléculaires de mutation parfaitement cohérents et compatibles avec la vision consensuelle de la Synthèse Moderne, les biologistes les proposant ne partageaient, et ne partagent pas encore aujourd'hui, notre analyse. En particulier, le premier et le dernier type de mécanismes sont présentés comme une nouveauté radicale par rapport aux processus de mutation admis par la Théorie Synthétique de l'Evolution.

Nous sommes convaincue que l'idée d'un défi au paradigme de la Synthèse Moderne posé par les résultats de ces expériences sur l'occurrence de mutations favorables dans un

³⁶ Lenski *et al*, 1989a, « Mutation and selection in bacterial population: Alternatives to the hypothesis of directed mutation ».

³⁷ Hendrickson *et al*, 2002, « Amplification Mutagenesis: evidence that growth with a selected gene amplification causes adaptative mutation and hyper-mutability » ; Roth & Andersson, 2004, *op. cit.*

³⁸ Nous rappelons que quelques années plus tôt Cairns et Foster (1992) avaient aussi proposé un mécanisme d'amplification à l'origine de l'occurrence de mutations favorables dans un milieu sélectif (« l'amplification non aléatoire des gènes mutants bénéfiques »).

³⁹ Foster & Rosche, 1999, « Increased episomal replication accounts for the high rate of adaptive mutation in *recD* mutants of *Escherichia coli* » ; Foster, 1999, *op. cit.* ; 2004, *op. cit.*

milieu sélectif trouve son origine dans l'interprétation, encore une fois « naïve », que ces biologistes donnent de la Synthèse Moderne : ils la réduisent à l'idée selon laquelle les mutations génétiques se produisent au hasard du point de vue évolutionnaire au sens où elles ont lieu en phase de croissance, et non pas en phase stationnaire, avant et indépendamment de toute pression de sélection. Nous avons brièvement analysé les limites de cette interprétation relativement aux expériences de Luria et Delbrück. Nous souhaitons simplement souligner que, déjà au début des années 1960, les biologistes savaient que les bactéries *Escherichia coli* peuvent accumuler des mutations lorsqu'elles ne sont pas en train de se diviser ou de répliquer leur ADN, c'est-à-dire en phase stationnaire⁴⁰.

5.2.2. La notion de « mutation adaptative »

Il est maintenant temps de clarifier comment, au cours des années 1990, les biologistes impliqués dans le débat sur le caractère des mutations génétiques définissaient ces mutations favorables, dites « adaptatives », se produisant dans un milieu sélectif et ayant pris la place des mutations « dirigées » au centre de cette controverse.

Le sens large et communément accepté de l'expression « mutations adaptatives » ne désignait aucun élément nouveau par rapport aux notions propres à la Synthèse Moderne : cette expression renvoyait aux « mutations qui confèrent un quelconque avantage sélectif en étant adaptatives (ou favorables), par contraste avec celles qui sont mal adaptatives (ou délétères) »⁴¹. Selon une définition plus précise en vogue aussi au cours des années 1990, les mutations « adaptatives » étaient considérées comme des mutations favorables se produisant dans des populations dites « mutatrices », c'est-à-dire des populations avec un taux élevé de mutation dans des conditions de stress environnemental⁴². Une telle caractérisation introduisait une spécification ultérieure par rapport à la première et n'entraînait pas non plus en contradiction avec le paradigme de la Synthèse Moderne. Elle consiste à définir les mutations

⁴⁰ Voir Cairns & Foster, 1991, « Adaptive Reversion of a Frameshift Mutation in *Escherichia coli* ». Il est presque étonnant de remarquer que nous avons trouvé cette information dans un article écrit par des biologistes, Cairns et Foster, qui tendent plutôt à mettre en évidence l'aspect innovateur et défiant de leurs résultats par rapport au paradigme de la Synthèse Moderne. Voir aussi Ryan *et al*, 1963, « Spontaneous mutation in spheroplasts of *Escherichia coli* » ; Ryan *et al*, 1961, « Is DNA replication a necessary condition for spontaneous mutation? ».

⁴¹ « [...] mutations that confer some selective advantage as being adaptive (or beneficial), in contrast with those that are maladaptive (or deleterious) », Sniegowski & Lenski, 1995, « Mutation and Adaptation: The Directed Mutation Controversy in Evolutionary Perspective » p. 285 (notre traduction).

⁴² Taddei, 1995, *Environnement et contrôles de la variabilité génétique chez Escherichia coli*.

« adaptatives », non seulement en termes de l'avantage qu'elles fournissent aux organismes porteurs, mais aussi en référence au type de situation environnementale où elles se produisent. Autrement dit, une mutation « adaptative » n'est pas simplement une mutation avantageuse pour l'organisme porteur : elle est une mutation favorable qui se produit lorsqu'un organisme est confronté, par exemple, à un changement soudain de son environnement et est caractérisé par un taux élevé de mutation.

Ce lien entre l'occurrence de mutations favorables et des conditions environnementales adverses amena un certain nombre de biologistes, parmi lesquels Shapiro⁴³, à proposer une définition plus controversée et ambiguë des mutations « adaptatives » : elles sont des mutations utiles qui se produisent plus fréquemment quand elles sont requises pour la survie et la reproduction des organismes concernés. En évoquant implicitement l'idée de « mutation dirigée » qui avait resurgi à la fin des années 1980 lors de la publication de l'article de l'équipe de Cairns, cette dernière définition nourrit le débat sur le caractère des mutations génétiques et, de plus, souleva une controverse ultérieure quant à l'origine des mutations dites « adaptatives » (au sens non controversé de cette expression⁴⁴).

Deux hypothèses furent proposées à ce sujet : l'hypothèse « pathologique » et l'hypothèse « adaptative » de l'origine des mutations « adaptatives ». Selon la première, le taux élevé de mutation caractérisant des populations situées dans des conditions de stress environnemental pourrait n'être qu'un « symptôme pathologique d'une cellule qui est en train de perdre le contrôle de sa propre intégrité génétique »⁴⁵ : les partisans de cette hypothèse visaient surtout à mettre en garde la communauté des biologistes contre la mobilisation parfois excessive des explications adaptationnistes. Selon l'hypothèse alternative, dite « adaptative », l'augmentation du taux de mutation chez des populations situées dans des situations de stress est une réponse adaptative induite par les conditions sélectives de l'environnement. Les mutations « adaptatives » ne sont pas simplement des mutations favorables se produisant chez des populations « mutatrices » en condition de stress environnemental ; elles sont des mutations favorables se produisant plus souvent au sein des populations situées dans ce genre de situation car susceptibles de permettre la survie et la reproduction des organismes les composant⁴⁶.

⁴³ Shapiro, 1995b, « 'Adaptive Mutation': The Debate Goes On », p. 286.

⁴⁴ Voir Taddei, 1995, *op. cit.*

⁴⁵ « [...] *pathological symptom of a cell that is losing control over its genetic integrity* », Lenski & Sniegowski 1995, « 'Adaptive Mutation': The Debate Goes On », p. 285 (notre traduction).

⁴⁶ Shapiro (Shapiro, 1995b, *op. cit.*, p. 287, notre traduction) écrit qu'un organisme est une espèce d'« ingénieur génétique » doté d'« une boîte à outils pleine de dispositifs moléculaires sophistiqués pour réorganiser les molécules d'ADN ». À son avis, cette boîte aurait été conservée par la sélection naturelle car elle peut fournir

L'étude des mutations dites « adaptatives » a fourni une meilleure connaissance de certains mécanismes moléculaires, dits « mutateurs », qui sont à l'origine de l'augmentation du taux de mutation chez des populations situées dans des conditions de stress environnemental (c'est-à-dire, des mécanismes à l'origine des mutations dites « adaptatives ») ; de plus, l'hypothèse « adaptative » a été confirmée expérimentalement dans le cas d'un certain nombre de ces mécanismes moléculaires de mutation. Cependant, le débat entre l'hypothèse « pathologique » et l'hypothèse « adaptative » est encore ouvert aujourd'hui. Bien que ces deux hypothèses soient parfaitement cohérentes et compatibles avec la vision consensuelle de la Synthèse Moderne au sujet des mutations génétiques, des définitions de « mutation adaptative » rappelant implicitement l'idée apparemment lamarckienne de « mutation dirigée », ou l'évoquant de manière explicite, sont soutenues encore aujourd'hui et sont présentées comme des nouveautés révolutionnaires par rapport au paradigme de la Synthèse Moderne. Plus précisément, sur la base des résultats expérimentaux qui accompagnèrent et nourrirent le débat sur le caractère des mutations de la fin des années 1980 jusqu'à aujourd'hui, et à la lumière des connaissances acquises au cours des trente dernières années au sujet de mécanismes moléculaires qui augmentent le taux de mutation en relation aux conditions de stress environnemental (les mécanismes « mutateurs »), un certain nombre de biologistes, d'historiens et de philosophes de la biologie persistent à affirmer que les mutations génétiques dues à ces mécanismes « mutateurs » représentent un défi important, et même une sorte de « révolution quantique »⁴⁷, pour la vision consensuelle de la Synthèse Moderne⁴⁸.

« la variabilité héréditaire nécessaire pour la survie à court et à long terme » (« *genetic engineer [...] an impressive toolbox full of sophisticated molecular devices for reorganizing DNA molecules [...] the hereditary variability needed for short- and long-term survival* »).

⁴⁷ Shapiro, 1999, « Genome System Architecture and Natural Genetic Engineering in Evolution ».

⁴⁸ Voici quelques exemples des affirmations de ces auteurs adressant un défi au paradigme de la Synthèse Moderne. Dans un article consacré à la découverte des mécanismes mutateurs, le microbiologiste Shapiro écrit (*ibid.*, p. 23, notre traduction) : « L'univers conceptuel de la génétique moléculaire est aussi différent de la génétique classique et de la théorie de l'évolution que la physique quantique de la mécanique classique » (« *The conceptual universe of molecular genetics is as different from classical genetics and evolutionary theory as quantum physics is from classical mechanics* »). Bien que la microbiologiste Wright et ses collègues soient fondamentalement d'accord avec la majeure partie de notre analyse empirique et conceptuelle des mutations dues aux mécanismes mutateurs (voir la section suivante), ils écrivent au sujet de ces derniers (1999, « Hypermutation in derepressed operons of *Escherichia coli* K12 », p. 5089 ; voir aussi p. 5094) : « Bien que les mutations soient *per se* aléatoires, [...], les mécanismes qui ciblent les opérons pour une augmentation du taux de transcriptions sont hautement spécifiques. Cette spécificité n'est pas compatible avec l'actuel dogme néo-darwinien », (« *Although the mutations per se are random, [...], the mechanisms that target operons for increased rates of transcription are highly specific. This specificity is not compatible with current neo-Darwinian dogma* »). L'historienne des sciences Fox Keller écrit (2003[2000], *Le siècle du gène*, p. 40) : « L'idée que les mécanismes de la faculté d'évolution (c'est-à-dire, les mécanismes mutateurs) pourraient eux-mêmes avoir évolué est un sérieux défi lancé à la théorie néo-darwinienne, car elle porte en elle la conséquence hérétique que les organismes ne constituent pas seulement le substrat passif de l'évolution, mais qu'ils

Les biologistes Jablonka et Lamb⁴⁹ sont deux représentantes emblématiques de ce défi actuellement lancé au paradigme de la Synthèse Moderne: elles soutiennent que les mutations dues à certains de ces mécanismes « mutateurs » ne sont pas « aléatoires » au sens du paradigme de la Synthèse Moderne. De plus, elles poussent leur argument critique jusqu'à conclure que, à la lumière des avancées récentes de la recherche sur les mutations génétiques, l'évolution darwinienne doit et peut inclure en son sein des processus lamarckiens instructifs afin d'expliquer ce genre de mutations. Il est important de garder bien à l'esprit au cours de notre enquête que le but de Jablonka et Lamb – et de tout autre biologiste et philosophe partageant les idées de ces dernières – n'est nullement de renverser la Théorie Synthétique de l'Evolution dans sa totalité. Elles ne mettent en cause que le postulat de la vision classique selon lequel toute mutation génétique est « aléatoire » ou le fruit du « hasard » et prétendent qu'une amélioration et un élargissement de la Synthèse Moderne requièrent l'incorporation en son sein des idées lamarckiennes quant à l'origine de la variation héréditaire.

Nous avons choisi de focaliser notre enquête sur le dernier livre de Jablonka et Lamb (2005), en particulier sur leur analyse empirique et conceptuelle des mécanismes controversés dits « mutateurs » et des mutations qu'ils produisent. Notre but est de comprendre si elles ont raison en affirmant que la découverte de ces phénomènes moléculaires met en cause la caractérisation des mutations génétiques propre à la Synthèse Moderne, c'est-à-dire ce que nous avons appelé la notion de « hasard évolutionnaire ». De plus, en tentant de montrer que leur argument est incorrect, nous visons à en identifier les failles à la fois du point de vue empirique et conceptuel.

5.3. Les mécanismes « mutateurs »

Présentons maintenant une variété de mécanismes moléculaires de mutation, dits « mutateurs », pour ensuite procéder dans le cadre du chapitre suivant à l'examen d'un des plus controversés : notre but est de comprendre pourquoi la découverte de certains d'entre eux

représentent aussi le moteur de leur propre changement ; cela laisse penser qu'ils ont été dotés d'une sorte de capacité d'intervention sur leur propre évolution ». En ce qui concerne les idées de Sternberg, voir en particulier les pages 170 et 171 de son article publié en 2002 (« On the Roles of DNA Repetitive Elements in the Context of a Unified Genomic-Epigenetic System ») et qui est presque entièrement consacré au rôle des éléments répétitifs dans la séquence d'ADN.

⁴⁹ Jablonka & Lamb, 1995, *Epigenetic Inheritance and Evolution* ; 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*.

suggère encore aujourd'hui l'idée d'un défi à la vision consensuelle de la Synthèse Moderne quant au caractère des mutations génétiques.

Durant les trente dernières années, les études expérimentales en génétique moléculaire au sujet des processus causaux à l'origine des mutations génétiques ont montré que, à travers l'activité d'un certain nombre de mécanismes moléculaires, le taux de mutation peut être l'objet d'une régulation génétique en relation avec les conditions de vie des organismes concernés. Les biologistes utilisent l'expression « mécanismes mutateurs » pour désigner ce genre de mécanismes qui, à travers la mobilisation de dispositifs enzymatiques complexes, sont à l'origine causale de l'augmentation des taux de mutation en relation, et même en réponse directe, aux forces sélectives en jeu dans l'environnement où les organismes concernés se trouvent.⁵⁰

C'est au sein des recherches en microbiologie, en particulier sur des espèces diverses de bactéries (*Escherichia coli*, *Neisseria meningitidis*, *Staphylococcus aureus*, *Streptococcus pneumoniae* et *Pseudomonas aeruginosa*), que la découverte des mécanismes mutateurs a vu le jour et que leur étude a connu des développements majeurs. Cela s'explique, d'abord, par la rapidité de croissance et d'évolution de ces microorganismes qui facilite donc la réalisation de beaucoup d'expériences en des temps relativement courts ; ensuite, par le fait que les bactéries sont des organismes qui vivent souvent dans des conditions environnementales changeantes et représentent donc un cas d'étude particulièrement approprié⁵¹. C'est pourquoi nous avons choisi de concentrer notre présentation et notre analyse sur ces mécanismes moléculaires qu'on retrouve, entre autres, chez les bactéries ; cependant, il est important de rappeler que des mécanismes analogues ont été observés et étudiés chez d'autres procaryotes et chez des organismes eucaryotes aussi⁵².

⁵⁰ La découverte des mécanismes mutateurs régulant le taux de mutation en relation avec les conditions de l'environnement et, plus précisément, de la possibilité que dans une situation de stress le taux de mutation évolue vers son augmentation, a représenté une objection importante à l'hypothèse du taux minimal de mutation. Cette hypothèse avait trouvé son origine dans l'idée de Kimura (1967, « On the evolutionary adjustment of spontaneous mutation rates ») selon laquelle, de manière générale, le taux de mutation le plus faible est sélectionné. Voir le chapitre 3.

⁵¹ Pour se rendre compte du type de conditions environnementales où les bactéries se trouvent habituellement, il suffit de penser aux souches de bactéries *Escherichia coli* qui ont fait de l'intestin des mammifères leur environnement et qui sont donc susceptibles de faire l'expérience d'environnements très différents au cours du temps.

⁵² Au sujet des mécanismes mutateurs observés et étudiés chez d'autres procaryotes que les bactéries et chez des organismes eucaryotes, voir Jablonka & Lamb (1995, *op. cit.*), et les références bibliographiques qu'elles suggèrent. En ce qui concerne les eucaryotes, il est important de rappeler que, dans la plupart des cas, les biologistes n'ont observé et étudié que des mécanismes mutateurs provoquant des mutations génétiques non transmissibles à la génération suivante, c'est-à-dire qui affectent les cellules somatiques, et non pas celles reproductives. Par exemple, en ce qui concerne les eucaryotes unicellulaires, les modifications de la chromatine se produisant durant le développement dans le macronoyau des protozoaires ciliés ; en ce qui concerne les eucaryotes multicellulaires, l'hypermutableté caractérisant les gènes impliqués dans la production des anticorps

Une présentation claire de ces phénomènes moléculaires consiste à distinguer les différents types de mécanismes mutateurs sur la base des caractéristiques suivantes. D'abord, un mécanisme mutateur peut être « constitutif » : les biologistes utilisent ce terme pour désigner le fait que, une fois que le mécanisme de mutation est apparu et s'est fixé génétiquement au sein d'une population située dans des conditions environnementales adverses, il reste actif chez les organismes individuels porteurs et chez leurs descendants même en absence de tout stimulus de stress. Un mécanisme mutateur peut aussi être « induit », c'est-à-dire qu'il n'est activé que de manière temporaire et en réponse à des stimuli de stress plus ou moins spécifiques. En second lieu, indépendamment du fait qu'un mécanisme mutateur soit constitutif ou induit, il peut provoquer une augmentation globale ou locale du taux de mutation dans des conditions de stress environnemental : soit le taux de mutation devient élevé le long du génome entier, soit son augmentation est ciblée dans une région ou dans un site particulier de la séquence d'ADN.

En référence aux distinctions présentées ci-dessus et sur la base de nombreuses études expérimentales, les biologistes ont identifié quatre types de mécanismes moléculaires à l'origine de l'augmentation du taux de mutation en condition de stress : a) des mécanismes mutateurs constitutifs, produisant une augmentation globale du taux de mutation ; b) des mécanismes mutateurs constitutifs, produisant une augmentation locale du taux de mutation ; c) des mécanismes mutateurs induits à l'origine causale d'un taux élevé de mutation dans tout le génome ; d) des mécanismes mutateurs induits à l'origine causale d'une augmentation locale du taux de mutation. Examinons d'abord les trois premiers types pour ensuite concentrer notre analyse sur les mécanismes mutateurs de type d) qui sont les plus controversés aujourd'hui dans le cadre du débat sur l'origine et le caractère des mutations génétiques.

5.3.1. Les mécanismes mutateurs constitutifs

a) L'inactivation d'une des étapes du système de contrôle, de relecture et de réparation des erreurs se produisant au niveau de la séquence d'ADN au cours de sa réplication (par exemple, le système de réparation méthyl-dirigé des mésappariements découvert chez *Escherichia coli*) est considéré comme un mécanisme mutateur constitutif qui augmente le

chez les mammifères.

taux de mutation dans tout le génome des organismes concernés⁵³. Généralement, ce type d'inactivation et l'augmentation du taux de mutation qui s'ensuit évoluent au sein de populations situées dans des conditions environnementales adverses : en effet, une variabilité élevée peut être avantageuse pour la survie d'un groupe d'organismes dans une situation de stress.

La diminution de l'efficacité du système de réparation est due à une mutation affectant un ou plusieurs gènes impliqués dans la synthèse des enzymes réparatrices (par exemple, dans le cas du système de réparation des mésappariement chez *Escherichia coli*, les gènes *mutS* et *mutL*). Le taux élevé de mutation (10^2 - 10^3 fois plus élevé que le taux de mutation chez des organismes non mutateurs ou sauvages, ce dernier étant de l'ordre de 10^{-6} à 10^{-7} par nucléotide répliqué⁵⁴) est donc dû à l'absence des facteurs catalytiques pouvant accomplir la fonction de réparation des erreurs qui se sont produites au cours de la réplication de l'ADN.⁵⁵

En vertu de son association physique avec une mutation favorable (ou davantage) que l'augmentation du taux de mutation a produite, l'allèle mutateur à l'origine de l'inactivation du système de réparation peut être sélectionné, et donc transmis à la génération suivante. Si cela se vérifie, la fréquence de cet allèle augmente au cours des générations au sein de la population concernée ; de cette manière, le nombre d'organismes mutateurs caractérisés par un taux élevé de mutation augmente aussi. Les biologistes parlent du phénomène de l'« autostop », ou de la « sélection de deuxième ordre », pour désigner ce processus de sélection indirecte du phénotype mutateur en vertu de l'avantage apporté par une ou plusieurs mutations favorables. Une fois que l'inactivation du système de réparation s'est fixée au niveau populationnel, c'est-à-dire que la fréquence de l'allèle mutateur est de 100%, le taux élevé de mutation devient un trait caractéristique de la population qui est donc dite « mutatrice ».⁵⁶

⁵³ Radman & Wagner, 1986, « Mismatch Repair in *Escherichia coli* » ; Taddei *et al*, 1997, « Genetic variability and adaptation to stress » ; Sniegowski *et al*, 1997, « Evolution of high mutation rates in experimental populations of *E. coli* » ; Sniegowski *et al*, 2000, « The evolution of mutation rates: separating causes from consequences » ; Radman *et al*, 2000, « Evolution-driving genes » ; Giraud *et al*, 2001, « The rise and the fall of mutator bacteria ».

⁵⁴ Glickman & Radman, 1980, « *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction ».

⁵⁵ Il est intéressant de remarquer que les mutations provoquées par l'inactivation du système de réparation sont essentiellement des transitions ou des mutations de décalage de lecture. Voir Schaaper & Dunn, 1987, « Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: The nature of *in vivo* DNA replication errors ».

⁵⁶ La sélection d'un taux élevé de mutation dû à la présence d'un allèle mutateur au sein d'une population dépend aussi de nombreux autres facteurs parmi lesquels figurent : le degré d'augmentation du taux de mutation chez les organismes mutateurs par rapport au taux de mutation des non mutateurs ; la taille de la population concernée et le rapport initial entre le nombre d'organismes mutateurs et d'organismes non mutateurs la constituant ; le degré de stabilité de l'environnement ; le nombre de mutations requises pour qu'un organisme

La diminution de l'efficacité du système de réparation due à la présence d'un allèle mutateur est considérée comme un mécanisme mutateur au sens où il s'agit d'une réponse consistant en l'augmentation de la fréquence des mutations et qu'une population peut acquérir, en termes évolutionnaires, en réponse aux défis posés par l'environnement. De plus, elle est considérée comme un mécanisme mutateur constitutif car, une fois acquise, l'inactivation du système de réparation perdure en présence de la mutation qui l'a provoquée, même en absence de toute source environnementale de stress. Il est aussi important de souligner que ce genre de mécanisme de mutation, en augmentant le taux de mutation, augmente la probabilité que des mutations avantageuses se produisent par rapport à une situation où la population n'est pas « mutatrice » et est donc caractérisée par un taux moins élevé de mutation. Cependant, la probabilité relative des mutations avantageuses par rapport aux mutations neutres et défavorables ne change pas nécessairement⁵⁷.

Comme nous l'avons déjà indiqué plus haut, on peut observer l'augmentation du taux de mutation due à l'inactivation du système de réparation lorsqu'une population d'organismes, par exemple de bactéries, se trouve dans une condition de stress, comme dans le cas d'un environnement caractérisé par des changements rapides et continuels. Pendant ces périodes de forte pression sélective, le taux de mutation évolue vers son augmentation pour les raisons présentées ci-dessus et qui sont liées à l'augmentation de la probabilité qu'une mutation avantageuse se produise dans un court laps de temps, donc à l'augmentation de la rapidité d'évolution. Une fois que la population est bien adaptée et que son environnement est devenu stable, un taux élevé de mutation peut se révéler défavorable, donc il peut changer et évoluer vers des valeurs plus basses jusqu'à sa valeur minimale possible⁵⁸. Cette dynamique évolutive explique la raison pour laquelle une population peut être caractérisée par une alternance de taux de mutation bas et élevés en dépendance de l'alternance de périodes où l'environnement est stable et de périodes de stress. L'avantage d'être mutateur n'est donc pas

s'adapte à l'environnement où il se trouve ; le degré d'avantage fourni par les mutations favorables sélectionnées au premier ordre et leur nombre ou, plus généralement, les bénéfiques qu'elles fournissent par rapport au coût dû à un taux élevé de mutation ; le taux de recombinaison. Voir Tenaillon *et al*, 2001, « Mutators and sex in bacteria: Conflict between adaptive strategies ». Il est intéressant de remarquer qu'un taux élevé de recombinaison (*crossing over*, chez les organismes à reproduction sexuée) tend à empêcher le phénomène de l'« autostop » en coupant le lien entre l'allèle mutateur et les allèles qui ont subi des mutations favorables. En ce qui concerne la présence de ce premier type de mécanisme mutateur constitutif, qu'est l'inactivation du système de réparation, chez des organismes à reproduction sexuée, voir Johnson, 1999, « Beneficial Mutations, Hitchhiking and the Evolution of Mutation Rates in Sexual Populations ».

⁵⁷ Voir le chapitre 4 du présent travail.

⁵⁸ Les seules contraintes pour un taux nul de mutation lorsqu'un organisme se trouve dans une situation environnementale favorable et stable sont la vitesse élevée du processus de réplication par rapport à l'activité de réparation et le coût énergétique des mécanismes antimutateurs. Voir Tenaillon *et al*, 2001, « Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation ».

fixe mais conditionnel : dans le cas d'un taux élevé de mutation, l'avantage des allèles favorables doit être plus grand que le coût, pour les organismes, d'être mutateur et d'avoir donc un système de réparation inactif.

b) Le processus à l'origine de ce que les biologistes appellent les « points chauds de mutation » représente un exemple d'un mécanisme mutateur constitutif provoquant une augmentation locale, dans des sites spécifiques du génome, du taux de mutation⁵⁹. Ces « points chauds » de mutabilité sont dus à la présence de séquences de nucléotides répétés – les « microsattellites », constitués par des répétitions de deux ou trois types de nucléotides (par exemple, ATATATAT) – qui peuvent déranger l'activité des enzymes de réplication (les ADN polymérases) en les induisant en erreur. Plus précisément, les microsattellites peuvent être à l'origine du glissement des ADN polymérases, c'est-à-dire de leur déplacement le long de la séquence d'ADN sans qu'elles procèdent à la réplication d'un nucléotide ou davantage : de cette manière, des mutations du cadre de lecture peuvent donc avoir lieu.⁶⁰

Des « point chauds » de mutation ont été observés chez les bactéries pathogéniques *Neisseria* et *Influenzae* : ils sont dus à la présence de répétitions nucléotidiques au niveau des gènes impliqués dans la production de facteurs protéiques – les protéines de surface – qui interagissent avec un environnement hautement imprédictible – les défenses des organismes hôtes. En effet, au moment de l'invasion, ces bactéries pathogéniques doivent faire face au système immunitaire des organismes qu'ils infectent. L'issue de l'interaction avec les défenses de ces derniers dépend de la structure des protéines de surface : un taux élevé de mutation au niveau des gènes impliqués dans la production des protéines de surface et, par conséquent, une grande variabilité au niveau de ces dernières, peut donc se révéler avantageuse pour ces envahisseurs pathogéniques car elle augmente la probabilité que le système immunitaire de l'hôte ne les reconnaisse pas et que l'hôte soit donc affecté.⁶¹

Moxon a proposé de parler des « gènes de contingence » pour désigner ces séquences nucléotidiques hypermutables observées chez les bactéries pathogéniques *Neisseria* et *Influenzae* : cette expression renvoie au fait que les produits de ces gènes ont souvent à faire avec un environnement hautement contingent et imprédictible. Il a aussi avancé et ensuite confirmé expérimentalement l'hypothèse selon laquelle les « points chauds de mutation » sont

⁵⁹ Moxon *et al*, 1994, « Adaptive evolution of highly mutable loci in pathogenic bacteria » ; Moxon *et al*, 2006, « Bacterial Contingency Loci: the Role of Simple Sequence DNA Repeats in Bacterial Adaptation ».

⁶⁰ Des répétitions de nucléotides au niveau de la séquence génomique peuvent aussi provoquer l'augmentation du taux de translocation, du taux de recombinaison et du taux de rupture de brin de l'ADN.

⁶¹ Les biologistes ont aussi observé et étudié des « point chauds de mutation » au sein du système immunitaire des hôtes des bactéries (des mammifères, donc des eucaryotes, parmi lesquels les humains), plus précisément au niveau des séquences nucléotidiques qui sont impliquées dans la production des anticorps. En effet, ce sont ces derniers qui entrent directement en interaction avec les protéines de surface des pathogènes.

un produit adaptatif de l'évolution par sélection naturelle : les répétitions de nucléotides dans les « gènes de contingences » ont été sélectionnées à cause de l'avantage qu'un taux élevé de mutation apporte au niveau de ces sites du génome.

Ce deuxième mécanisme mutateur constitutif est moins coûteux que l'inactivation du système de réparation car, dans ce cas, l'augmentation du taux de mutation ne concerne pas tout le génome mais est localisée dans des séquences précises : les gènes vitaux très sensibles aux mutations, comme ceux impliqués dans le système de contrôle et de réparation (*housekeeping genes*), ne sont pas concernés.

5.3.2. Les mécanismes mutateurs induits

c) Le système SOS est un mécanisme mutateur postulé par le microbiologiste croate Radman autour des années 1970 afin d'expliquer la mutagenèse observée chez la bactérie *Escherichia coli*⁶².

Il s'agit d'un mécanisme dont l'activation peut être induite de manière graduelle et transitoire⁶³ et qui peut provoquer une augmentation globale et/ou locale du taux de mutation en réponse à de forts stress environnementaux⁶⁴. Plus précisément, le système SOS est composé par un peu plus d'une vingtaine de gènes qui sont habituellement réprimés et dont l'expression peut être induite par l'exposition des cellules (des bactéries) à des conditions qui endommagent et interfèrent avec la réplication de leur ADN (par exemple, des lésions de la séquence nucléotidique, la présence de régions d'ADN simple brin, des ruptures du double-brin de l'ADN, l'irradiation par rayonnements UV, de forts stress métaboliques comme l'absence de nutriments dans l'environnement où les organismes se trouvent).

L'expression des gènes SOS donne lieu à la production d'une variété de facteurs protéiques d'activation et d'inhibition de l'expression génétique (par exemple, LexA et RecA chez *Escherichia coli*), d'un ensemble de protéines impliquées dans des mécanismes divers de

⁶² Radman (1973, « Phenomenology of an Inducible Mutagenic DNA Repair Pathway in *Escherichia coli*: SOS repair hypothesis »). Radman proposa le système SOS, non seulement pour rendre compte de la mutagenèse qu'il avait observée sur le chromosome des bactéries, mais aussi pour expliquer les phénomènes de réactivation et de mutation suivant l'irradiation par les rayonnements ultraviolets qui avaient été observés et étudiés chez des phages lytiques. Pour des références bibliographiques à ce sujet, voir Taddei, 1995, *Environnement et contrôles de la variabilité génétique chez Escherichia coli*.

⁶³ *Ibid.* Bien que le système SOS soit considéré comme un mécanisme mutateur induit, il peut arriver qu'il soit actif de manière constitutive, même en absence de sources exogènes de stress, à cause de la présence d'une mutation dans un gène particulier – le gène *recA*.

⁶⁴ *Ibid.* ; Taddei *et al.*, 1997, « Genetic variability and adaptation to stress » ; Tenailon *et al.*, 2004, « Evolutionary significance of stress induced mutagenesis ».

réparation de l'ADN (chez *Escherichia coli*, les protéines UvrA, UvrB, UvrC, Uvr(A)BC, UvrD, PolI et la ligase sont impliqués dans la réparation par excision-synthèse) et de quelques types particuliers d'enzymes de réplication et de réparation, les ADN polymérases, qui ont tendance à produire des erreurs et à contourner (*by-pass*) les lésions au niveau de la séquence d'ADN (l'ADN polymérase IV et V chez *Escherichia coli*).⁶⁵ Certaines de ces ADN polymérases (PolV chez *Escherichia coli*) impliquées dans la réponse SOS sont dotées de la capacité à continuer la réplication de l'ADN en dépit des lésions qui intéressent ce dernier ; cependant, elles ne sont pas caractérisées par une haute fidélité de copie et produisent donc une augmentation du taux de mutation au niveau des régions génomiques endommagées⁶⁶. D'autres (PolIV chez *Escherichia coli*) peuvent aussi provoquer une augmentation du taux de mutation au niveau des séquences privées de lésions et donc tout au long du génome entier.

Le résultat de l'activation du système SOS par un stimulus de stress est donc l'augmentation du taux de mutation, qui peut être globale et/ou locale en dépendance des types de facteurs mutagéniques (les ADN polymérases) induits qui, à leur tour, dépendent du type de stress ayant activé le système SOS.⁶⁷ Une fois que le stimulus de stress disparaît, les gènes SOS ne sont plus exprimés mais à nouveau inhibés et le taux de mutation diminue jusqu'à atteindre sa valeur moyenne normale.

⁶⁵ Le système SOS contrôle l'activation de trois types de mécanismes de réparation : l'excision-synthèse, la recombinaison et un troisième système qui implique l'activité des ADN polymérases à basse fidélité de réplication et dont l'activité provoque donc l'occurrence de mutations (Friedberg *et al*, 2006, *DNA Repair and Mutagenesis*).

⁶⁶ Dans ce cas, on peut affirmer que l'augmentation du taux de mutation ciblée au niveau des séquences endommagées de l'ADN est un effet secondaire des activités de réparation de ce type d'ADN polymérases qui contournent les lésions de l'ADN et, de cette manière, les incorporent dans la séquence génomique. De toute manière, comme Taddei et ses collègues l'affirment (1997, *op. cit.*, p. 280, notre traduction), on peut conclure que « la mutagenèse ciblée (sur les lésions de l'ADN) par l'induction du système SOS peut être considérée comme une fonction de réparation de l'ADN aussi bien qu'une fonction à l'origine de la variation génétique », « *SOS induction-stimulated targeted (to DNA lesions) mutagenesis can be considered a DNA repair function as well as a function that generates genetic variation* ».

⁶⁷ Taddei (1995, *op. cit.*, p. 37) précise que « la réponse SOS est graduelle en fonction de la quantité de lésions de l'ADN et de leur persistance [...] Les divers éléments de régulation assurent une adéquation de la réponse à l'ampleur des dommages causés à l'ADN ».

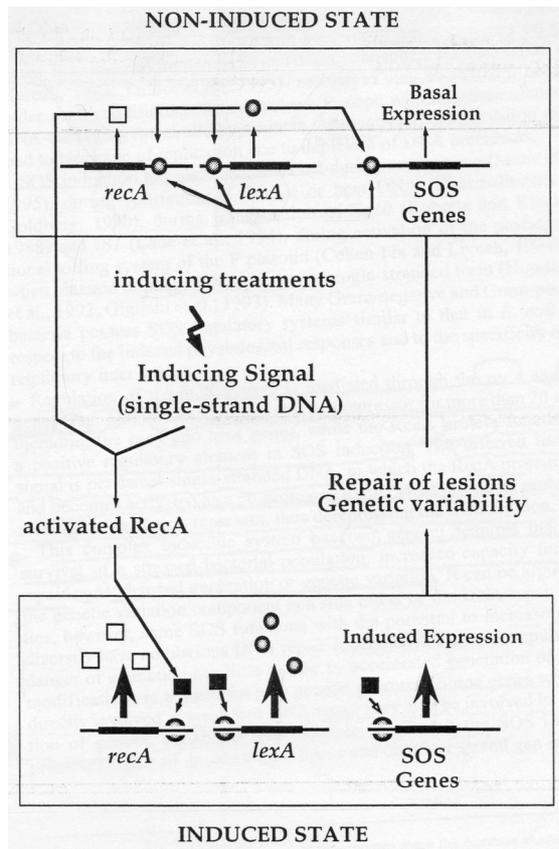


Figure 2 : Représentation schématique de la réponse SOS. Figure tirée de Taddei *et al*, 1997, « Genetic variability and adaptation to stress », p. 279.

D'un point de vue évolutif, l'origine du mécanisme mutateur qu'est le système SOS est débattue entre les biologistes doutant du fait que l'augmentation du taux de mutation en réponse au stress puisse être une adaptation, et soutenant donc l'hypothèse « pathologique », et les biologistes convaincus du fait qu'il s'agit, au contraire, d'une réponse adaptative évoluée par sélection naturelle, et qui sont donc partisans de l'hypothèse « adaptative » quant à son origine évolutive. À la lumière des résultats expérimentaux obtenus en étudiant *Escherichia coli*, le découvreur du système SOS, Radman, est convaincu que ce mécanisme mutateur est une stratégie adaptative due à l'action de la sélection naturelle. En effet, il écrit :

« la confirmation ultime de l'hypothèse SOS vint en 1999 avec l'identification de nombreuses enzymes clairement conçues (*designed*) pour produire des mutations. Ces enzymes, qu'aujourd'hui nous appelons mutases de l'ADN, appartiennent toutes à un groupe particulier d'ADN polymérases »⁶⁸.

⁶⁸ « The ultimate validation of the SOS hypothesis came in 1999 with the identification of several enzymes clearly designed to produce mutations. These enzymes, which we now call DNA mutases, all belong to a special

De plus, Radman ajoute que, même si le système SOS n'avait pas évolué par sélection naturelle – autrement dit, même si ce mécanisme mutateur n'avait évolué qu'en tant que sous-produit de la sélection pour un autre trait avantageux et non pas à cause de sa propre contribution à la fitness relative de ses organismes porteurs – « il a désormais évolué pour devenir un système de contrôle de la mutabilité, à travers l'augmentation de la fréquence des mutations dans des moments de stress »⁶⁹. Par conséquent, selon Radman, il y a des éléments en faveur du fait que la réponse SOS est du moins le résultat de l'histoire évolutive, plutôt que la simple manifestation d'une condition pathologique d'organismes situés dans des conditions environnementales de stress.⁷⁰

d) Un exemple de mécanisme mutateur induit produisant une augmentation locale, ciblée dans une région ou dans un site particulier du génome, du taux de mutation est représenté par le mécanisme physiologique que Wright et ses collègues⁷¹ ont observé et étudié chez la souche d'*Escherichia coli* K12.⁷²

D'abord, il est nécessaire de rappeler que, chez les bactéries, l'expression de certains gènes est active ou inhibée selon les conditions environnementales. Parmi les gènes qui sont réprimés en présence d'une abondante quantité de nutriments figurent les gènes impliqués dans la production des acides aminés car ces derniers sont habituellement disponibles dans la nourriture. Lorsqu'un acide aminé particulier vient à manquer, le gène pertinent pour sa production est activé et, de cette manière, les bactéries porteuses sont capables de synthétiser elles-mêmes ce produit et de l'utiliser pour survivre dans le milieu où elles se trouvent. Wright et ses collègues ont étudié une souche de bactéries *Escherichia coli* K12 LeuB⁻, c'est-à-dire caractérisée par une mutation au niveau du gène *leuB*, au sein de l'opéron⁷³ leucine

group of DNA polymerases », Radman *et al* 2000, « Evolution-driving genes », p. 93 (notre traduction). Voir aussi Radman, 1999, « Enzymes of evolutionary change ».

⁶⁹ « [...] it has now evolved to become a system to control mutability, increasing the mutation frequency in times of stress », Radman cité par Jablonka & Lamb 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, p. 322 (notre traduction).

⁷⁰ Il existe des preuves expérimentales en faveur du fait que les ADN polymérases impliquées dans la réponse SOS, ayant tendance à produire des erreurs et à contourner les lésions au niveau de la séquence nucléotidiques, sont des résultats adaptatifs de l'évolution par sélection naturelle : elles semblent avoir été sélectionnées car elles augmentent la probabilité que la molécule d'ADN, et donc l'organisme porteur, survivent (Olivier Tenaillon, communication personnelle).

⁷¹ Wright *et al*, 1999, « Hypermutation in derepressed operons of *Escherichia coli* K12 » ; Wright, 2000, « A Biochemical Mechanism for Nonrandom Mutations and Evolution ».

⁷² L'induction par des conditions environnementales de stress (par exemple, des chocs thermiques) de l'augmentation du taux de mutation (en particulier, du taux de délétion) dans une région spécifique du génome a été observée et étudiée chez la plante moutarde *Brassica nigra*, qui est un organisme eucaryote. Ce mécanisme mutateur semble impliquer le processus de recombinaison, mais il n'est pas encore connu dans les détails. Voir Jablonka & Lamb, 2005, *op. cit.*, p. 98.

⁷³ Le concept d'opéron fut introduit par Jacob et Monod (1961, « Genetic Regulatory Mechanisms in the

(l'opéron *leu*), qui les rend incapables de produire cet acide aminé. Ils ont placé et observé ces bactéries dans un milieu où la leucine est épuisée : ils ont remarqué que, comme prévu, le manque de cet acide aminé déclenche une série d'activités métaboliques spécifiques pour l'activation de l'opéron *leu*, pour l'augmentation de son taux de transcription et aussi pour l'augmentation du taux de mutation au niveau de tous les gènes de cet opéron, le gène *leuB* compris. En d'autres termes, Wright et ses collègues ont observé que, lorsque des bactéries incapables de synthétiser la leucine sont situées dans un environnement où cette acide aminé est absent, le taux de mutation au niveau de l'opéron *leu* impliqué dans la production de la leucine est plus élevé que dans des conditions environnementales normales. Ce phénomène moléculaire spécifique augmente la probabilité de toute mutation affectant les gènes de l'opéron *leu* et, en particulier, de la mutation provoquant la transformation des bactéries de $LeuB^-$ à $LeuB^+$: il augmente donc la probabilité que ces microorganismes arrivent à survivre et à se reproduire dans ce milieu adverse.

Les résultats obtenus par l'équipe de Wright semblent représenter un exemple de ce genre de mutations génétiques que la définition la plus controversée de la notion de « mutations adaptatives » désignait durant les années 1990. Nous rappelons que selon cette définition une mutation « adaptative » est une mutation favorable se produisant le plus souvent lorsqu'elle est requise pour la survie et la reproduction des organismes concernés⁷⁴. Afin d'arriver à des conclusions bien fondées quant au caractère des mutations observées chez *Escherichia coli* par Wright et ses collègues, une description claire et préalable du mécanisme que ces derniers étudièrent est nécessaire. De manière schématique, nous pouvons le résumer par les étapes suivantes⁷⁵.

Synthesis of Proteins») lorsqu'ils découvrirent le phénomène de régulation de l'expression des gènes, en particulier la structure et la régulation de ce qu'ils appelèrent l'opéron lactose. L'opéron est une unité de transcription. Plus précisément, il est composé par un ensemble linéaire de gènes de structure (c'est-à-dire, de gènes impliqués dans la production de facteurs protéiques), un promoteur, un terminateur et un opérateur.

⁷⁴ Shapiro, 1995b, « 'Adaptive Mutation': The Debate Goes On ».

⁷⁵ Voir Wright, 2000, *op. cit.*, p. 2998, figure 3.

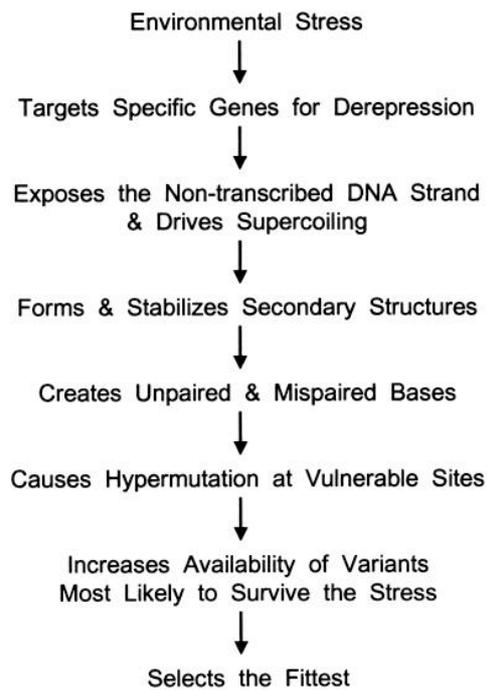


Figure 3 : Les étapes principales du mécanisme moléculaire à l'origine de l'hypermutabilité locale, au niveau de l'opéron *leu*, que Wright et ses collègues ont observée chez des populations d'*Escherichia coli* K12. Figure tirée de Wright, 2000, « A Biochemical Mechanism for Nonrandom Mutations and Evolution ».

D'abord, un stress environnemental se manifeste pour la population de bactéries mutatrices $LeuB^-$ dont l'opéron *leu* est réprimé : il s'agit de l'épuisement de l'acide aminé de la leucine dans le milieu où elle se trouve. Ensuite, ce stimulus de stress induit la dérèpression des gènes de l'opéron *leu* impliqué dans la production de la leucine et la conséquente activation de leur transcription : ce phénomène est à l'origine de la formation de la bulle de transcription, de la présence d'une séquence d'ADN simple brin et du super-enroulement (négatif) de l'ADN au niveau de l'opéron *leu*. L'activation de la transcription de ce dernier est aussi à l'origine de la formation de structures secondaires de l'ADN qui produisent à leur tour la formation de mésappariements de bases nucléotidiques toujours au niveau de ce site génomique impliqué dans la production de la leucine.

Tous ces phénomènes de modification de la structure du génome au niveau des gènes de l'opéron *leu* sont dus à sa dérèpression initiale : en rendant ces gènes plus instables, leur probabilité de subir des mutations augmente. Comme Wright l'indique, il est important de souligner que la spécificité caractérisant ce genre de réponse à un stress environnemental réside entièrement dans la première étape, c'est-à-dire dans le processus de dérèpression des gènes de l'opéron *leu* qui est induit par l'absence de cet acide aminé dans le milieu où les bactéries se trouvent.

Wright affirme que cet ensemble de mécanismes de *feedback*, qui va des conditions de l'environnement aux processus physiologiques à l'intérieur des bactéries, tend à faciliter et à accélérer l'adaptation de ces dernières à la nouvelle situation environnementale. En effet, l'augmentation locale du taux de mutation au niveau de l'opéron leucine augmente la probabilité que la mutation reverse pouvant rétablir la capacité à produire cet acide aminé (de LeuB⁻ à LeuB⁺) se produise dans un laps de temps plus court que si le taux de mutation était plus bas. De plus, le caractère local de ce taux élevé de mutation augmente aussi la probabilité de cette mutation favorable par rapport à une situation où le taux de mutation est élevé le long de tout le génome, et non pas de manière spécifique dans le temps et en ce qui concerne le site génomique intéressé. Une fois que la mutation reverse transformant les bactéries de LeuB⁻ à LeuB⁺ se produit, elle permet à ces dernières de survivre, de se reproduire et, par conséquent, elle est transmise à la génération suivante.

Wright va plus loin lorsqu'elle propose de considérer le processus métabolique induit par l'absence de la leucine comme un résultat adaptatif de l'évolution par sélection naturelle : il s'agit, selon elle, d'une réponse adaptative à des conditions environnementales adverses et, plus spécifiquement, en réponse au manque d'acides aminés.⁷⁶ Il est central de comprendre si Wright considère comme un produit de l'évolution par sélection naturelle, soit le processus d'augmentation du taux de mutation au niveau de l'opéron *leu*, soit l'étape spécifique initiale de cette cascade d'interactions physico-chimiques, qu'est la dérégulation de l'opéron *leu*. En effet, la dérégulation de l'opéron *leu* en réponse à l'épuisement de la leucine dans le milieu où les bactéries se trouvent est de manière peu controversée un mécanisme physiologique adaptatif qui doit son origine à l'histoire évolutive de la population de bactéries concernée. L'augmentation du taux de mutation peut aussi bien être considérée comme un produit adaptatif de l'évolution ; cependant, il pourrait simplement s'agir d'une conséquence inévitable et, heureusement, potentiellement avantageuse du mécanisme physiologique adaptatif consistant en la dérégulation de l'opéron leucine, car l'activation de la transcription rend cet ensemble de gènes plus vulnérables aux mutations.

Si Wright et ses collègues retiennent que l'augmentation du taux de mutation est elle-même une adaptation, c'est-à-dire qu'elle est le résultat du processus d'évolution par sélection

⁷⁶ Nous analyserons plus tard l'affirmation de Wright et ses collègues, reprise ensuite par Jablonka et Lamb (2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*), selon laquelle les résultats de leurs expériences sur l'hypermutation des opérons dérégulés chez *Escherichia coli* K12 ne sont pas compatibles avec le « dogme néo-Darwinien » (Wright *et al.*, 1999, « Hypermutation in derepressed operons of *Escherichia coli* K12 », p. 5089), c'est-à-dire le paradigme de la Synthèse Moderne.

naturelle, nous suggérons que leur conviction ne pourrait que dériver d'un biais caractérisant leur étude. Les expériences qu'ils ont réalisées utilisent des bactéries porteuses d'une mutation *leuB*⁻ empêchant ces microorganismes de produire de la leucine ; par conséquent, l'augmentation du taux de mutation au niveau de l'opéron *leu* augmente la probabilité que la mutation reverse avantageuse se produise par rapport à une situation où le taux de mutation augmente mais de manière globale le long de tout le génome ou bien là où le taux de mutation est moins élevé. Cependant, imaginons que Wright et ses collègues réalisent cette même expérience avec des bactéries *LeuB*⁺ capables de synthétiser de la leucine. Nous suggérons qu'il est très plausible qu'ils observent dans ce cas aussi une augmentation du taux de mutation au niveau de l'opéron *leu* lorsque ces bactéries *LeuB*⁺ sont placées dans un milieu caractérisé par l'épuisement de cet acide aminé : en effet, la dérégulation et l'activation de la transcription de l'opéron leucine rendraient ce site génomique beaucoup plus vulnérable que d'autres du point de vue physico-chimique, et donc plus mutable. Par conséquent, et contrairement à l'hypothèse adaptative de l'origine de ce mécanisme moléculaire de mutation avancée par l'équipe de Wright, dans la situation expérimentale que nous venons d'imaginer, on n'observerait pas l'augmentation de la probabilité d'une mutation avantageuse au niveau du gène *leuB* ; on remarquerait plutôt l'augmentation de la probabilité qu'une mutation défavorable ou neutre se produise au niveau de ce même site génomique (par exemple, de la mutation reverse transformant les bactéries de *LeuB*⁺ à *LeuB*⁻).

Avec le raisonnement ci-dessus nous ne visons pas à refuser la possibilité d'une origine adaptative du mécanisme moléculaire de mutation observé et étudié par Wright et ses collaborateurs. Nous voulons simplement mettre en garde ces biologistes quant à la possibilité qu'une explication adaptationniste se révèle erronée dans ce cas.

Dans le chapitre suivant, nous analyserons le bien fondé du défi que, selon certains, la découverte des mécanismes mutateurs a posé au paradigme de la Synthèse Moderne, notamment à l'idée que toute mutation génétique est le fruit du hasard du point de vue évolutionnaire. Nous focaliserons notre attention les mécanismes mutateurs de type d), c'est-à-dire qui sont induits en réponse à des stress environnementaux et qui provoquent une augmentation locale du taux de mutation : en effet, les détracteurs de la Synthèse Moderne quant au « hasard évolutionnaire » des mutations génétiques, en particulier Jablonka et Lamb (1995, 2005), s'appuient sur les études expérimentales de ce genre de mécanismes afin de corroborer leurs arguments critiques.

Chapitre 6 – Peut-on encore affirmer que *toute* mutation génétique est le fruit du « hasard évolutif » ?

6.1. Les mécanismes mutateurs : un véritable défi à la Synthèse Moderne ?

La découverte des mécanismes moléculaires de mutation dits « mutateurs », que nous avons présentés à la fin du chapitre précédent, correspond à la reconnaissance de la régulation génétique du taux de mutation en relation aux forces sélectives en action dans l'environnement où les organismes concernés se trouvent. Durant les dix dernières années, un certain nombre de biologistes, d'historiens et de philosophes de la biologie¹ ont soutenu que ces avancées sont en tension avec ce que la Théorie Synthétique Moderne de l'Evolution affirme au sujet de l'origine et du caractère des mutations génétiques : la conception de la Synthèse Moderne n'est, à leur avis, pas adéquate pour rendre compte des mutations que ces mécanismes moléculaires produisent, de manière plus ou moins spécifique, en relation, et même en réponse directe, à des conditions environnementales de stress. C'est plus particulièrement le lien entre l'augmentation du taux de mutation et le type de substrat sélectif qui, selon eux, fait problème et pose un véritable défi à l'idée classique des mutations qui se produisent au hasard par rapport au cours adaptatif de l'évolution.

Nous défendons dans ce chapitre une interprétation qui s'oppose à ces critiques du paradigme de la Synthèse Moderne. La découverte des mécanismes mutateurs représente certainement un progrès dans la recherche sur le processus d'évolution par rapport à la Théorie Synthétique de l'Evolution ; cependant, cette dernière peut intégrer ces avancées relativement récentes au sujet des mutations génétiques sans pour autant devoir abandonner ses postulats fondamentaux – parmi lesquels figurent l'idée du « hasard évolutif » des mutations – et en préservant donc sa validité empirique et son pouvoir explicatif.

¹ Shapiro, 1999, « Genome System Architecture and Natural Genetic Engineering in Evolution » ; 2005, « A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering » ; Wright *et al.*, 1999, « Hypermutation in derepressed operons of *Escherichia coli* K12 » ; Fox Keller, 2003[2000], *Le siècle du gène* ; Sternberg, 2002, « On the Roles of DNA Repetitive Elements in the Context of a Unified Genomic-Epigenetic System » ; Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*.

Un élément pouvant montrer, entre autres, que la découverte des mécanismes mutateurs ne représente pas un changement radical et révolutionnaire par rapport au paradigme classique réside dans la conception qu'avaient les biologistes de la Synthèse Moderne de l'origine des mutations génétiques et, en particulier, de leur taux d'occurrence². Nous pouvons rappeler que les principaux acteurs de cette synthèse entre la génétique classique et la vision darwinienne de l'évolution pensaient, d'abord, que la mutabilité du génome est à la fois due aux caractéristiques physico-chimiques du matériel génétique et le résultat du processus d'évolution par sélection naturelle concernant les organismes porteurs. De plus, ils savaient déjà que le taux de mutation est sous contrôle génétique. Enfin, ils avançaient même l'hypothèse selon laquelle les mutations pourraient se produire de manière dirigée, non pas du point de vue adaptatif, mais en ce qui concerne l'endroit de leur occurrence et leur abondance.

Cela nous pousse à considérer la question suivante, que les biologistes Jablonka et Lamb n'ont pas posée mais ont exprimée sous la forme d'une remarque ironique envers les défenseurs de la vision classique³ : pourquoi serait-il aussi révolutionnaire par rapport au paradigme de la Synthèse Moderne de penser que les mécanismes de mutation ont aussi été modifiés et ont évolué de manière adaptative via l'action de la sélection naturelle ? C'est le caractère rhétorique de cette question qui, dans le cadre de notre analyse, nous amènera à conclure que toutes les avancées récentes concernant les mécanismes moléculaires à l'origine des mutations génétiques sont compatibles avec la théorie classique de l'évolution et, plutôt que de la remettre en question, l'enrichissent.

6.2. Jablonka & Lamb (2005) : analyse empirique et conceptuelle des mutations dues aux mécanismes mutateurs

Jablonka et Lamb (2005)⁴ ont récemment repris certaines idées qu'elles avaient exposées auparavant (1995)⁵ au sujet de la découverte des mutations dues aux mécanismes mutateurs (qu'elles appellent mutations « interprétatives », « semi-dirigées » ou « non aléatoires ») et, en particulier, au sujet de sa portée en biologie : elles ont à nouveau avancé

² Voir la section 4.2.1 du chapitre 4.

³ Jablonka & Lamb, 2005, *op. cit.*, p. 101.

⁴ *Ibid.*

⁵ Jablonka & Lamb, 1995, *Epigenetic Inheritance and Evolution*.

des arguments poussant à considérer ces avancées comme un défi à la conception des mutations génétiques propre à la Synthèse Moderne.

Jablonka et Lamb ne s'attaquent pas au premier type de mécanismes mutateurs constitutifs⁶ produisant une augmentation du taux de mutation dans le génome entier (par exemple, l'inactivation constitutive du système de réparation). Elles reconnaissent⁷ que ce genre de mécanisme représente une stratégie biologique qu'elles appellent « exploratrice » car consistant en l'augmentation de la fréquence de mutations qui sont « aléatoires » en ce qui concerne le moment et le site du génome de leur occurrence et, plus spécifiquement, par rapport à leur caractère adaptatif du point de vue de l'évolution. Par conséquent, nous ne nous occuperons pas de ce premier type de mécanisme mutateur dans la suite notre enquête. Par contre, Jablonka et Lamb affirment que les mutations dues aux autres types de mécanismes mutateurs⁸ ne sont pas le fruit du hasard au sens suggéré par la Synthèse Moderne à cause de la spécificité caractérisant leur occurrence. Dans le cas des mécanismes mutateurs induits, cette spécificité concerne le moment où les mutations se produisent ; dans le cas des mécanismes produisant une augmentation locale du taux de mutation, elle concerne l'endroit où la séquence génomique est altérée et, plus précisément, la fonction des facteurs protéiques produits à partir des gènes mutés. De manière plus concise, Jablonka et Lamb voient dans ces mutations génétiques un défi pour la Synthèse Moderne à cause du lien qu'elles entretiennent avec le substrat sélectif où les organismes porteurs se trouvent: ces mutations ont lieu quand et/ou là où les organismes concernés en ont besoin pour survivre et se reproduire dans leur environnement.

Nous avons décidé de focaliser notre analyse du défi posé à la Synthèse Moderne, et donc à la pertinence empirique de la notion de « hasard évolutionnaire »⁹, sur les mutations dues aux mécanismes mutateurs induits et produisant une augmentation locale, ciblée au niveau d'un site particulier du génome, du taux de mutation (les mécanismes mutateurs du type d)¹⁰). Ce choix est motivé par le fait qu'il s'agit de l'un des cas les plus controversés par rapport au paradigme de la Synthèse Moderne car les mutations que ce type de mécanisme moléculaire produit sont non seulement provoquées en réponse à un stimulus de stress, mais aussi localisées au niveau des gènes dont la mutation pourrait se révéler avantageuse dans la

⁶ Les mécanismes mutateurs de type a). Voir la section 5.3.1 du chapitre 5.

⁷ Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, p. 92.

⁸ Les mécanismes mutateurs b), c), et d). Voir les sections 5.3.1 et 5.3.2 du chapitre 5.

⁹ Nous rappelons que, dans le chapitre 4, nous avons montré que la notion de « hasard évolutionnaire » correspond à la vision consensuelle de la Synthèse Moderne au sujet du caractère des mutations génétiques.

¹⁰ Voir la section 5.3.2 du chapitre 5.

situation environnementale adverse que les organismes doivent affronter. En particulier, Jablonka et Lamb affirment que ces mutations sont « potentiellement avantageuses »¹¹ car elles ne sont pas aveugles en ce qui concerne la fonction biologique qu'elles affectent : elles sont localisées au niveau des gènes pouvant permettre aux organismes porteurs de survivre ; par conséquent, en aucun sens elles ne peuvent être considérées comme des « mutations aléatoires ». C'est pourquoi Jablonka et Lamb affirment que les mutations faisant partie d'une augmentation induite et locale du taux de mutation sont « à la fois acquises et requises » et que, par conséquent, leur explication nécessite d'inclure au sein de l'évolution darwinienne des processus lamarckiens.¹²

Le but de la présente analyse est d'identifier les failles de cet argument avancé par Jablonka et Lamb, de souligner qu'il n'est pas approprié du point de vue conceptuel et historique de parler de processus « lamarckiens » relativement aux mécanismes mutateurs, et de montrer de manière concluante que ces derniers ne posent aucun défi à l'égard de la Théorie Synthétique de l'Evolution. Nous démontrerons que toute mutation génétique, même les mutations induites et ciblées en réponse à des stress environnementaux, sont le fruit du hasard au sens de la vision consensuelle de la Synthèse Moderne, c'est-à-dire selon la notion de « hasard évolutionnaire ».

6.2.1. Le paradigme de la Synthèse Moderne selon Jablonka & Lamb (2005)

Dans la première partie de leur dernier livre (2005) consacré aux quatre dimensions de l'évolution des organismes vivants – l'hérédité génétique, épigénétique, comportementale et symbolique –, Jablonka et Lamb proposent d'abord une présentation rapide de la théorie darwinienne de l'évolution, des idées de ses opposants lamarckiens et des transformations qu'elle a connues dès sa naissance (Darwin, 1859) jusqu'à la reformulation, en termes moléculaires, de la Synthèse Moderne autour des années 1960. Ensuite, elles consacrent le deuxième chapitre de leur livre à la relation complexe entre le matériel génétique et les caractères phénotypiques. Dans le troisième chapitre, elles développent une analyse claire et exhaustive des découvertes récentes quant aux mécanismes de mutation génétique, pour ensuite montrer le caractère inadéquat, selon elles, de la vision classique de l'origine de la variation génétique.

¹¹ « Likely to be adaptive », Jablonka & Lamb, 2005, *op. cit.*, p. 101.

¹² *Ibid.*, p. 97, 101-102.

Avant de nous plonger dans l'analyse empirique et conceptuelle des mutations produites par les mécanismes mutateurs que Jablonka et Lamb proposent, examinons d'abord comment elles décrivent la vision consensuelle de la Synthèse Moderne au sujet du caractère des mutations génétiques.

Dans le cadre de l'introduction et du premier chapitre, Jablonka et Lamb¹³ définissent la conception classique de l'origine de la variation génétique héréditaire comme l'idée selon laquelle les mutations sont « entièrement aléatoires » ou « aveugles » et la posent en contraste avec la conception lamarckienne de l'hérédité flexible ou, autrement dit, de l'hérédité des caractères acquis. Elles définissent cette dernière comme « l'hérédité des changements génomiques induits par les facteurs environnementaux »¹⁴.

Une vingtaine de pages plus loin, toujours dans le cadre du premier chapitre¹⁵, elles présentent en trois points les fondements de la Synthèse Moderne concernant l'hérédité, la variation et la sélection naturelle : y figure l'idée selon laquelle les mutations sont le résultat de « changements accidentels » – plus précisément, elles sont des « changements aléatoires » dans les séquences nucléotidiques de l'ADN situé à l'intérieur du noyau des cellules – qui n'ont rien à voir avec l'histoire développementale de l'organisme individuel concerné. Ensuite, Jablonka et Lamb présentent à nouveau la vision de la Synthèse Moderne quant à l'origine de la variation génétique héréditaire de manière un peu plus détaillée et en termes moléculaires¹⁶. Elles rappellent d'abord que les processus de mutation et de recombinaison sont considérés comme les deux sources causales des altérations de la séquence d'ADN et, plus généralement, de la variation héréditaire. Ensuite, elles spécifient – et elles soulignent

¹³ Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, p. 7.

¹⁴ *Ibid.*, « *the inheritance of genomic changes induced by environmental factors* », (notre traduction). Plus loin (p. 31) Jablonka et Lamb rappellent aussi le dogme central de la biologie moléculaire qui, bien qu'il ait été introduit au moment du tournant moléculaire dans les années 1950, est devenu partie intégrante de la vision de l'origine de la variation héréditaire propre à la Synthèse Moderne. Elles l'expriment en affirmant qu'il n'existe aucun moyen à travers lequel les changements phénotypiques induits par l'environnement puissent avoir un quelconque effet sur le matériel génétique. Cependant, la définition du dogme central proposée par Jablonka et Lamb déforme complètement l'idée originelle de Crick qui, en 1958 (« On Protein Synthesis », p. 153), écrivit : « [...] une fois que l'information est passée dans la protéine elle ne peut pas en sortir à nouveau. Plus en détail, le transfert d'information d'acide nucléique à acide nucléique, ou d'acide nucléique à protéine peut être possible, mais le transfert de protéine à protéine, ou de protéine à acide nucléique est impossible. Information signifie ici la détermination précise de la séquence de bases dans l'acide nucléique ou d'acides aminés dans la protéine. », (notre traduction), « [...] *once 'information' has passed into protein it cannot get out again. In more details, the transfert of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfert from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein.* » Voir aussi Crick, 1970, « Central Dogma of Molecular Biology ».

¹⁵ Jablonka & Lamb, 2005, *op. cit.*, p. 29.

¹⁶ *Ibid.*, p. 33-34.

tout au long de cette première partie de leur livre¹⁷ – que les mutations sont des erreurs ou des lésions se produisant durant la réplication du matériel génétique ou dues à des agents mutagènes. Enfin, elles rappellent que, toujours dans le cadre de la vision classique, les facteurs physico-chimiques (par exemple, les rayonnements UV, les gaz moutardes, etc.) pouvant induire une augmentation du taux de mutation ne peuvent qu'augmenter l'abondance de mutations « aléatoires » du point de vue de l'adaptation et non pas, de manière spécifique, l'occurrence de mutations « adaptatives » (au sens qu'elles fournissent un avantage sélectif).

Dans le troisième chapitre, Jablonka et Lamb rappellent encore une fois les idées caractérisant la Synthèse Moderne¹⁸. D'une part, elles présentent quelques biais de mutation reconnus par les biologistes partisans de la Synthèse Moderne – par exemple, la spécificité au niveau moléculaire des mutations produites par certains agents mutagènes et le fait que le taux moyen de mutation est le résultat d'un processus de sélection naturelle. D'autre part, elles soulignent que l'idée de l'occurrence des mutations génétiques en réponse aux besoins adaptatifs des organismes concernés est complètement étrangère à la vision classique de l'origine de la variation. Ce dernier rappel au sujet du paradigme de la Synthèse Moderne marque le point de départ de l'argument que Jablonka et Lamb avancent afin de mettre en doute le caractère aléatoire des mutations génétiques dues aux mécanismes mutateurs (exception faite de celles provoquées par les mécanismes constitutifs provoquant une augmentation globale du taux de mutation).

L'argumentation de Jablonka et Lamb appelle plusieurs remarques. En premier lieu, cette présentation de la vision classique de l'origine de la variation génétique que Jablonka et Lamb proposent est certes correcte, mais n'est pas suffisamment analytique pour pouvoir comprendre et cerner pleinement la conception des mutations génétiques propre à la Synthèse Moderne que nous avons recherchée, identifiée et analysée dans le chapitre précédent. Elle correspond plutôt à l'interprétation « naïve » du paradigme de la Synthèse Moderne : elle ne prend en compte que la représentation simplifiée et idéalisée du processus de mutation génétique dans le cadre des modèles utilisés par les biologistes de la Synthèse Moderne ; de plus, elle ne considère pas ce que ces derniers reconnaissaient déjà au sujet de l'origine évolutive et moléculaire (après 1953) de la variation génétique. En effet, dans les premières pages de leur livre, Jablonka et Lamb focalisent leur attention sur le caractère spontané des mutations génétiques selon la vision classique et tendent à la radicaliser : par exemple, elles soulignent l'indépendance totale du processus de mutation de tout autre

¹⁷ Par exemple, voir Jablonka & Lamb, 2005, *op. cit.*, p. 99.

¹⁸ *Ibid.*, p. 87.

événement développemental et évolutif concernant les organismes individuels porteurs. De plus, bien que Jablonka et Lamb rappellent que, selon la vision classique, les mutations ne sont pas des phénomènes complètement aléatoires, elles limitent leur présentation des biais de mutation reconnus ou soupçonnés par les biologistes de la Synthèse Moderne aux faits et aux idées les moins intéressantes par rapport au débat sur le caractère des mutations génétiques du point de vue évolutionnaire.

En second lieu, lorsque Jablonka et Lamb opposent la conception classique de l'origine de la variation héréditaire à la notion lamarckienne d'hérédité flexible ou d'hérédité des caractères acquis, elles fournissent une définition assez ambiguë et historiquement inadéquate de cette dernière. Le biologiste évolutionnaire Haig¹⁹, dans un article critique envers le (néo-)lamarckisme moderne auquel Jablonka et Lamb s'identifient, a récemment mis en évidence la confusion qui entoure actuellement la notion d'hérédité des caractères acquis. En particulier, il souligne que la définition que Jablonka et Lamb proposent, et selon laquelle l'hérédité des caractères acquis est la transmission d'un changement génétique induit par l'environnement, implique, par exemple, qu'une mutation génétique causée par l'exposition à des radiations ou à des agents chimiques doit être considérée comme un phénomène d'hérédité flexible ou d'hérédité des caractères acquis. De la même manière, cette définition implique aussi que d'autres phénomènes non strictement génétiques – par exemple, l'hérédité des *patterns* de méthylation de la séquence d'ADN – représentent des cas d'hérédité des caractères acquis.

Bien que Haig souhaite bannir cette expression (« hérédité des caractères acquis »), il pense que rien n'empêche de considérer ce genre d'influences de l'environnement sur ce qui est transmis d'une génération à la suivante comme une instance de l'hérédité des caractères acquis. Cependant, il souligne que, contrairement à ce que Jablonka et Lamb soutiennent, la notion de l'hérédité des caractères acquis ainsi définie n'entre pas en contradiction avec les idées fondamentales de la Synthèse Moderne : aucun des phénomènes génétiques et épigénétiques qu'elle englobe ne correspond à ce que la notion lamarckienne de l'hérédité des caractères acquis, critiquée par Weismann²⁰, désigne historiquement. Plus généralement, la manière dont Jablonka et Lamb définissent la notion d'hérédité des caractères acquis ne

¹⁹ Haig, 2006 « Weismann Rules! Epigenetics and Lamarckian temptation ».

²⁰ Weismann, 1892[1883], « Sur l'Hérédité » ; 1893[1892], *The Germ Plasm. A Theory of Heredity*. Sur la base d'un certain nombre d'arguments théoriques et de quelques études expérimentales, Weismann récusait l'hérédité des caractères acquis. Il soutint que les changements du soma, c'est-à-dire des cellules du corps, survenus au cours de la vie des organismes individuels (par exemple, le renforcement des muscles, la perte d'un membre, etc.) ne pouvaient pas induire des changements correspondants dans le plasma germinatif permettant au caractère phénotypique altéré d'être transmis aux générations suivantes.

renvoie pas à l'idée lamarckienne selon laquelle l'adaptation ontogénétique, concernant les organismes individuels, est la source de l'adaptation phylogénétique, c'est-à-dire au niveau des populations naturelles.²¹ En effet, d'une part, la notion de l'hérédité des caractères acquis, attaquée à la fin du XIX^e siècle par Weismann, fait référence à la transmission de changements somatiques adaptatifs que les organismes individuels ont acquis au cours de leur développement sans qu'aucune altération préalable se soit produite au niveau du matériel héréditaire. D'autre part, l'idée lamarckienne selon laquelle l'adaptation phylogénétique trouve son origine, non pas dans un processus de sélection mais dans l'adaptation ontogénétique, désigne l'hypothèse selon laquelle les réponses physiologiques et adaptatives des organismes individuels à leur environnement sont la source causale des changements des caractères héréditaires au cours du temps, de l'adaptation de ces organismes à leur environnement et, plus généralement, de la transformation des espèces.

Nous partageons la critique que Haig adresse à la définition proposée par Jablonka et Lamb de l'hérédité des caractères acquis et, plus généralement, à leur (néo-)lamarckisme moderne : Jablonka et Lamb se proclament en partie lamarckiennes et plaident la nécessité d'introduire des processus instructifs lamarckiens au sein de la théorie darwinienne de l'évolution sans pour autant fonder leur proposition sur une conception claire et historiquement pertinente des notions lamarckiennes auxquelles elles font appel. Dans le cadre de ce chapitre, nous visons donc à montrer que Jablonka et Lamb prétendent mobiliser des idées enracinées dans la pensée de Lamarck sans pour autant faire référence à leur signification originelle du point de vue historique.

6.2.2. Les mutations dues aux mécanismes mutateurs selon Jablonka & Lamb (1995, 2005)

La présentation et l'analyse empirique de ce que Jablonka et Lamb²² appellent les « mutations interprétatives »²³ (ou « semi-dirigées » ou « non aléatoires ») – qui sont les

²¹ En ce qui concerne la pensée de Lamarck et des lamarckiens les plus fidèles à ce dernier, nous renvoyons au chapitre 3 du présent travail.

²² Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, p. 92-102.

²³ Jablonka et Lamb (*ibid.*, p. 89-92) utilisent une expérience de pensée afin de distinguer trois stratégies biologiques différentes en ce qui concerne l'origine de la variation génétique. Elles imaginent trois tribus caractérisées par des styles de vie différents. Les membres de la première tribu s'appellent « les Conservatifs ». Leur attitude consiste à apprendre du passé : c'est pourquoi, par métaphore, ils correspondent à la stratégie biologique consistant à répondre à toute situation par des mécanismes physiologiques bien établis ou par des

mutations produites de manière plus ou moins spécifique par les mécanismes mutateurs du type b), c) et d) en relation avec les conditions environnementales de stress, et même en réponse directe aux changements de l'environnement – est très accessible et claire : elle fournit une analyse appropriée et exhaustive, du point de vue empirique, des découvertes relativement récentes sur l'origine de la variation génétique héréditaire.

Jablonka et Lamb introduisent trois types de situations où, selon elles, les mutations génétiques se situent entre les deux extrêmes représentés par ce qu'elles appellent les « mutations totalement aléatoires » et les « mutations totalement dirigées »²⁴. Premièrement, les mutations induites globalement dans tout le génome (qui correspondent aux mutations dues aux mécanismes mutateurs du type c) comme le système SOS). Deuxièmement, l'hypermutableté constitutive et localisée dans un endroit particulier du génome (qui correspond aux mutations dues aux mécanismes mutateurs constitutifs du type b)), par exemple les « points chauds » de mutation). Troisièmement, les mutations induites localement dans un site ou dans une région du génome (qui correspondent aux mutations dues aux mécanismes mutateurs du type d)) étudiés, par exemple, par Wright et ses collègues chez *Escherichia coli*)²⁵.

Il est intéressant de rappeler que Jablonka et Lamb traitent de manière approfondie la question de l'origine de ces mécanismes mutateurs : elles reconnaissent la plausibilité de l'hypothèse « adaptative » qui a été l'objet de plusieurs confirmations expérimentales et selon laquelle ils sont des produits adaptatifs de l'évolution par sélection naturelle²⁶. Par exemple, elles affirment que le caractère « interprétatif » des mutations dues à ces divers mécanismes de mutation réside dans le fait que le moment et l'endroit où elles se produisent dépendent du passé évolutif de la population concernée²⁷. Quelques pages plus loin, elles affirment explicitement que « les chances qu'une cellule trouve une solution mutationnelle sont élevées car le passé évolutif a construit un système fournissant des indices intelligents à propos d'où

mutations très spécifiques et précises. Les membres de la deuxième tribu sont « les Explorateurs ». Ils tendent à ne pas baser leur action sur les expériences passées mais à créer des solutions nouvelles et créatives : c'est pourquoi Jablonka et Lamb les associent aux stratégies biologiques consistant en l'augmentation globale du taux de mutations aléatoires. Les membres de la troisième tribu s'appellent « les Interprètes ». Ils sont caractérisés par une tendance à l'exploration tout en gardant un certain lien avec les pratiques et les règles traditionnelles : ils correspondent donc aux stratégies biologiques reposant sur la production d'une augmentation du taux de mutation quand et/ou à l'endroit du génome où des altérations pourraient apporter un avantage sélectif aux organismes porteurs.

²⁴ Dans cette section, nous allons procéder à la clarification de la signification attribuée à ces termes et à ces expressions par Jablonka et Lamb.

²⁵ Wright *et al*, 1999, « Hypermutable in derepressed operons of *Escherichia coli* K12 ».

²⁶ Jablonka & Lamb, 2005, *op. cit.*, p. 320-323.

²⁷ *Ibid.*, p. 92.

et quand générer des mutations »²⁸. Enfin, elles rappellent aussi la phrase de Pasteur, reprise et transformée par la généticienne Caporale – « le hasard favorise le génome préparé » - et concluent en affirmant que cette préparation propre aux systèmes génétiques est évidemment évolutionnaire²⁹.

Jablonka et Lamb semblent donc reconnaître l'importance des nouveautés apportées par la découverte des mécanismes mutateurs tout en considérant que le paradigme de la Synthèse Moderne est encore le meilleur modèle explicatif du processus de l'évolution. Cependant, leur analyse s'égaré lorsqu'elles adressent un défi considérable à la Théorie Synthétique en affirmant que la production de la variation génétique héréditaire est due à des processus qui ne sont ni « purement sélectifs » ni « purement instructifs » :

« Entre ces extrêmes on trouve la plupart des processus du monde réel, qui sont à des degrés différents à la fois instructifs et sélectifs »³⁰.

Par les termes « instructif » et « sélectif » Jablonka et Lamb désignent respectivement la vision lamarckienne et la vision propre à la Synthèse Moderne de l'origine de la variation au sein des populations naturelles. Elles soutiennent que les mécanismes mutateurs à l'origine d'une augmentation du taux de mutation au moment et/ou à l'endroit où un changement du génome pourrait se révéler avantageux sont en partie instructifs car les mutations qu'ils produisent ne sont pas totalement aléatoires mais ciblées et/ou induites en réponse à des stimuli environnementaux de stress. Par conséquent, selon elles, la variation héréditaire que ces mécanismes génèrent n'est pas simplement due à un processus de sélection des variants produits de manière aléatoire, mais à un processus qui est en partie lamarckien.

Nous voulons suggérer que, dans ce cadre, Jablonka et Lamb font référence à une conception de la vision lamarckienne de la variation héréditaire qui n'est pas historiquement appropriée ou, du moins, qui ne renvoie pas aux idées originelles de Lamarck. Lorsqu'elles parlent de « processus instructifs » en référence aux mécanismes mutateurs, elles semblent simplement désigner le caractère induit et biaisé des mutations génétiques produites par ces derniers, ce qui est parfaitement cohérent et compatible avec la Théorie Synthétique de l'Evolution et qui, comme elles le reconnaissent, est très vraisemblablement un résultat adaptatif

²⁸ « *The cell's chances of finding a mutational solution are enhanced because its evolutionary past has constructed a system that supplies intelligent hints about where and when to generate mutations* », Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, p. 101 (notre traduction).

²⁹ *Ibid.*

³⁰ « *Between these extremes we find the majority of the processes in the real world, which are to varying degrees both instructive and selective* », *ibid.*, p. 102 (notre traduction).

de l'évolution par sélection naturelle³¹. Cependant, elles précisent ensuite que ce caractère instructif fait de ces mécanismes de mutation des « processus lamarckiens », ce qui les placerait donc en opposition radicale avec le paradigme de la Synthèse Moderne. Leur position pourrait sembler contradictoire ; cependant, elle est parfaitement cohérente car fondée sur l'idée selon laquelle il est nécessaire et suffisant qu'un processus de mutation soit induit par un facteur de l'environnement et soit (fortement) biaisé – indifféremment, de manière adaptative ou non adaptative – pour qu'il affecte le cours de l'évolution et pour qu'on puisse donc parler d'une origine lamarckienne de la variation héréditaire.

Nous attaquons la pertinence historique d'une telle notion de « processus lamarckien » et de la vision de la Synthèse Moderne auxquelles Jablonka et Lamb font référence. En premier lieu, en ce qui concerne la vision lamarckienne, nous avons déjà souligné les limites de la définition de l'hérédité des caractères acquis que Jablonka et Lamb mobilisent et qui, contrairement à la notion originelle d'hérédité flexible, n'est nullement incompatible avec le paradigme de la Synthèse Moderne.

En deuxième lieu, nous rappelons que, selon Lamarck, une variation est une modification spécifique et adaptative d'un caractère qui est due à la réponse physiologique d'un organisme individuel aux changements de l'environnement et qui peut être transmise à la génération suivante sans qu'aucun processus de sélection ne soit impliqué. En termes actuels et moléculaires, nous soutenons qu'une mutation génétique doit être considérée comme résultant d'un processus lamarckien si et seulement si d'une part elle est « dirigée » du point de vue évolutionnaire (c'est-à-dire si et seulement si elle est causée de manière spécifique et (exclusivement) adaptative par un processus physico-chimique en réponse aux conditions de l'environnement³²) et si d'autre part la « directionnalité » spécifique et adaptative du processus qui l'engendre n'est pas le résultat d'une histoire évolutive et sélective, mais la manifestation d'une tendance inhérente à l'organisme individuel concerné. Sur cette base, nous contestons l'idée de Jablonka et Lamb selon laquelle un processus de mutation est lamarckien si et seulement si son activité est induite par l'environnement et les mutations qu'il produit sont (fortement) biaisées, pas nécessairement de manière adaptative. En effet, les biologistes se disant « néo-lamarckiens » qui, au début du XX^e siècle, focalisèrent leurs études expérimentales sur l'origine de la variation non adaptative afin de montrer le rôle des

³¹ Nous renvoyons au chapitre 4 de la présente recherche où nous avons présenté les biais de mutation reconnus par les biologistes de la Synthèse Moderne.

³² Voir le chapitre 4.

biais de mutation, et non pas des biais de sélection, au sein du processus évolutif³³ visaient surtout à contester le rôle attribué à la sélection naturelle par la théorie darwinienne de l'évolution – ce qui ne figure pas parmi les buts de l'analyse de Jablonka et Lamb³⁴. De plus, ces « néo-lamarckiens » professaient plutôt des idées propres aux théories de l'orthogénèse et non pas à la vision lamarckienne originelle de la variation et de l'adaptation.

En ce qui concerne la vision de la Synthèse Moderne, nous rappellerons simplement que toute mutation génétique est considéré comme le produit du hasard du point de vue évolutionnaire (ou le fruit du « hasard évolutionnaire ») si et seulement si elle n'est pas une mutation « dirigée » par rapport au cours adaptatif de l'évolution : une telle caractérisation nie qu'une mutation génétique puissent être produite par un processus physico-chimique de manière spécifique et (exclusivement) adaptative en réponse aux conditions de l'environnement, à la fois dans le cas où la directionnalité propre au processus serait le résultat de l'histoire évolutive et sélective de la population à laquelle l'organisme concerné appartient et dans le cas où cette même directionnalité serait lamarckienne car due à une tendance adaptative inhérente à l'organisme individuel porteur de la mutation, et non pas à l'évolution par sélection naturelle.³⁵ Dans le chapitre précédent, nous avons montré que, contrairement à ce que Jablonka et Lamb pensent, cette conception de l'origine de la variation héréditaire correspond à la vision consensuelle de la Synthèse Moderne et n'est nullement en contradiction avec une variété de biais moléculaires et évolutionnaires de mutation parmi lesquels certains étaient déjà reconnus dans les années 1930-1950.

Nous concluons donc que la faille principale de l'argument de Jablonka et Lamb réside dans les définitions historiquement inappropriées qu'elles donnent de leur position (néo-)lamarckienne et de la vision classique de l'origine de la variation génétique

³³ Voir Jablonka & Lamb, 1995, *Epigenetic Inheritance and Evolution*, p. 280.

³⁴ Au contraire, Jablonka et Lamb font appel à la vision lamarckienne pour désigner le caractère instructif de mécanismes moléculaires qui, comme elles le reconnaissent elles-mêmes, sont très probablement des produits adaptatifs de l'évolution. Cela est à nos yeux paradoxal car la directionnalité des processus lamarckiens, pour que ces derniers soient tels, ne doit dépendre d'aucun événement de sélection mais d'une tendance inhérente à chaque organisme individuel (voir Haig, 2006, « Weismann Rules! Epigenetics and Lamarckian temptation »). Déjà dans leur livre publié en 1995, Jablonka et Lamb affirmaient qu'elles « ne doutent pas que les mécanismes de base sous-jacents à l'hérédité des caractères acquis évoluèrent de manière darwinienne via la sélection de variations accidentelles » (Jablonka & Lamb, 1995, *op. cit.*, p. 25, notre traduction), « [...] *do not doubt that the basic mechanisms underlying the inheritance of acquired variations evolved in a Darwinian fashion by the selection of accidental variations.* »

³⁵ Cependant, il est important de souligner que l'observation d'une mutation « dirigée » du point de vue évolutionnaire car due, non pas à un processus lamarckien, mais à un mécanisme moléculaire dont le caractère spécifique et adaptatif réside dans des cycles précédents de variation et sélection représenterait un véritable défi pour la pertinence de la notion de « hasard évolutionnaire » dans la caractérisation de toute mutation génétique. Pourtant, une telle découverte empirique ne mettrait pas en cause la Synthèse Moderne dans sa totalité, en particulier son pouvoir explicatif quant à la production et la perpétuation de la variation.

héréditaire.³⁶ Ce point met en évidence que notre désaccord envers les conclusions de Jablonka et Lamb ne concerne pas les caractéristiques physico-chimiques des processus traités et les hypothèses avancées en ce qui concerne leur histoire évolutive et sélective. Il s'agit plutôt d'un désaccord d'ordre terminologique et conceptuel.³⁷

6.2.3. Jablonka et Lamb (1995, 2005) : des mutations « totalement aveugles » aux mutations « totalement dirigées »

Considérons les distinctions terminologiques et conceptuelles que Jablonka et Lamb introduisent et utilisent dans le cadre de leur analyse des mutations dues aux mécanismes mutateurs. Plus précisément, il s'agit de comprendre où, selon elles, se situe la ligne de démarcation entre le caractère « aléatoire » et « non aléatoire » des mutations et, plus en détail, entre ce qu'elles appellent les mutations « totalement aveugles » ou « aléatoires », les mutations « semi-dirigées » et les mutations « totalement dirigées ».

Dans leur premier livre, Jablonka et Lamb³⁸ affirment que l'expression « mutation dirigée » peut être utilisée pour désigner deux types différents de mutations. Selon une première acception, elle signifie simplement qu'« un défi environnemental particulier produit des changements répétables et spécifiques dans le matériel génétique. Les changements peuvent être avantageux, nuisibles, ou neutres »³⁹. Dans un deuxième sens, une « mutation dirigée » est une mutation « 'dirigée' vers une fin adaptative : les mutations produites en réponse à un défi environnemental sont à la fois spécifiques et sélectivement avantageuses »⁴⁰. Elles décident d'utiliser l'expression « mutation dirigée » dans la première acception selon laquelle une mutation, pour être considérée comme « dirigée », ne doit pas nécessairement apporter un avantage sélectif : elles justifient leur choix en affirmant qu'il est

³⁶ Par contre, nous tenons à souligner que leur analyse est claire et pertinente par rapport au sujet biologique traité.

³⁷ Haig (2006, *op. cit.*) a aussi admis que l'origine de son propre désaccord avec Jablonka et Lamb est plutôt une question de préférences conceptuelles, en particulier à propos de la manière de définir les termes. Cela ne diminue pas la valeur des critiques qu'il leur adresse : comme Haig affirme en conclusion de son article, les distinctions terminologiques et conceptuelles sont importantes pour promouvoir la clarté de la pensée.

³⁸ Jablonka & Lamb, 1995, *Epigenetic Inheritance and Evolution*, p. 57.

³⁹ « [...] a particular environmental challenge produces specific and repeatable changes in the hereditary material. The changes can be advantageous, detrimental, or neutral. », Jablonka & Lamb, 1995, *op. cit.*, p 57, (notre traduction).

⁴⁰ « [...] 'directed' towards an adaptive end: the mutations produced in response to an environmental challenge are both specific and selectively advantageous. », *ibid.*, (notre traduction).

dicté à la fois par la signification commune du terme « dirigé » et par les précédents historiques concernant l'utilisation de l'expression « mutation dirigée »⁴¹.

Dans leur dernier livre, Jablonka et Lamb (2005)⁴² introduisent une distinction supplémentaire entre ce qu'elles appellent les mutations « totalement dirigées » et les mutations « semi-dirigées ». Les premières sont « des changements adaptatifs reproductibles se produisant dans des sites spécifiques en réponse à des stimuli spécifiques »⁴³, comme les changements du matériel génétique se produisant au cours du développement de certains animaux⁴⁴. Les deuxièmes sont définies comme des altérations génétiques produites dans des conditions de stress, en réponse à des signaux environnementaux, mais qui ne conduisent pas à une réponse unique et nécessairement adaptative. Il s'agit donc des mutations « interprétatives » dues à l'activation des mécanismes mutateurs qui, selon Jablonka et Lamb, ne sont pas « aléatoires » au sens de la vision de la Synthèse Moderne et tombent « quelque part entre les variations totalement aveugles, qui ne sont spécifiques ni quant à leur nature ni en ce qui concerne le moment dans le temps et le site du génome où elles se produisent, et les variations totalement dirigées [...] »⁴⁵.

En lisant ce passage, on peut donc comprendre que Jablonka et Lamb parlent de mutations « totalement aveugles » (ou « aléatoires » ou « spontanées ») pour désigner des changements du matériel génétique qui ne sont spécifiques d'aucune façon, ni en ce qui concerne la séquence d'ADN qu'ils affectent, ni quant au moment où ils se produisent, ni du point de vue adaptatif ou, autrement dit, en ce qui concerne la fonction affectée. Il s'agit, selon elles, du seul type de mutation admis au sein du paradigme de la Synthèse Moderne.

⁴¹ Dans le cas de cette deuxième raison d'ordre historique justifiant le caractère non nécessairement avantageux du point de vue sélectif des « mutations dirigées », Jablonka et Lamb font sûrement référence aux expériences réalisées par les « néo-lamarckiens », en particulier les paléontologues de l'école américaine, du début du XX^e siècle, ces derniers visant à montrer l'idée orthogénétique selon laquelle la pression de mutation peut guider l'évolution dans des directions qui ne sont pas forcément adaptatives.

⁴² Jablonka & Lamb, 2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*.

⁴³ « [...] *reproducible adaptive changes that occur at specific sites in response to specific stimuli* », *ibid.*, p. 89 (notre traduction).

⁴⁴ Jablonka et Lamb (*ibid.*, p. 68-71) renvoient, par exemple, aux changements concernant l'ADN des anticorps lors de la maturation des lymphocytes chez les organismes vertébrés, au phénomène de diminution de la chromatine chez le parasite *Ascaris* vivant dans l'intestin des chevaux et à la réplication d'une partie des chromosomes ayant lieu pendant le développement de la mouche du vinaigre *Drosophila*. Dans le cadre du présent travail, nous ne considérerons pas ces mutations se produisant au cours du développement car, comme Jablonka et Lamb le spécifient, elles ne concernent que les cellules somatiques, et non pas la lignée germinale : ces altérations génétiques ne sont pas transmises à la génération suivante et donc ne jouent pas de rôle important dans le processus évolutif.

⁴⁵ *Ibid.*, p. 88-89 (notre traduction), « [...] *somewhere between totally blind variations, which are specific neither in their nature nor in the time and site in the genome where they occur, and totally directed variations [...]* ».

<i>Jablonka & Lamb 2005</i>	
Mutations “aléatoires”	Mutations “totalement aveugles”
Mutations “non aléatoires” (en partie lamarckienne)	Mutations “semi-dirigées” (dues à des mécanismes mutateurs divers) Mutations “totalement dirigées” (développementales)

Figure 4 : La distinction entre mutations « aléatoires » et mutations « non aléatoires » selon Jablonka et Lamb (2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*).

Les définitions des différents types de mutations proposées par Jablonka et Lamb nous informent que les critères sur la base desquels ces dernières tracent la ligne de démarcation entre ce qu’elles appellent les mutations « aléatoires » (« totalement aveugles ») et les mutations « non aléatoires » (« semi-dirigées » et « totalement dirigées ») sont, d’abord, la spécificité de l’occurrence des altérations de l’ADN, c’est-à-dire leur appartenance à une augmentation globale ou locale du taux de mutation ; en second lieu, le rôle causal de l’environnement dans leur production. Selon Jablonka et Lamb, pour qu’une mutation ne soit pas « aléatoire » il suffit qu’elle soit spécifique, au sens où elle fait partie d’une augmentation locale du taux de mutation, ou que l’environnement extérieur soit impliqué causalement dans sa production. Le caractère adaptatif (favorable, défavorable ou neutre) de la mutation ne joue aucun rôle en ce qui concerne son caractère « aléatoire » ou « dirigé », mais n’entre en jeu que pour faire la part entre les mutations « semi-dirigées », pouvant être favorables, défavorables ou neutres, et les mutations « totalement dirigées », ces dernières étant non seulement spécifiques, mais aussi sélectivement avantageuses.

Nous partageons le choix des critères de la spécificité de la mutation et du rôle causal de l’environnement dans sa production afin d’identifier et classifier le caractère d’une mutation génétique donnée. Cependant, suivant la vision consensuelle de la Synthèse Moderne que nous avons identifiée dans le chapitre précédent, de manière cohérente du point de vue historique avec les idées lamarckiennes les plus fidèles à la pensée originelle de Lamarck et en accord avec ce que Millstein⁴⁶ affirme à ce sujet, nous suggérons d’ajouter un troisième critère, à savoir le caractère adaptatif des mutations, afin de distinguer les mutations

⁴⁶ Millstein, 1997, *The Chances of Evolution: An Analysis of the Roles of Chance in Microevolution and Macroevolution*.

« aléatoires » des mutations « non aléatoires ». Dans les termes et selon les notions que nous avons définies dans le cadre du présent travail, c'est uniquement en considérant aussi ce troisième critère, que nous pourrions déterminer si une mutation est le produit du « hasard évolutionnaire » ou s'il s'agit d'une « mutation dirigée » par rapport au processus évolutif. En effet, ce qui caractérise historiquement ce dernier type de mutation est non seulement le fait qu'il s'agit de mutations spécifiques et qui sont produites en réponse à un stimulus de l'environnement, mais qu'en outre elles apportent un avantage sélectif aux organismes individuels concernés : ces trois critères représentent l'ensemble des conditions qu'une mutation doit remplir pour pouvoir être considérée comme « dirigée » du point de vue évolutionnaire.

Nous adoptons donc la deuxième acception de l'expression « mutation dirigée » que Jablonka et Lamb ont proposée dans leur premier livre et selon laquelle une mutation, pour être considérée comme « dirigée », doit nécessairement apporter un avantage sélectif. Cependant, nous ne conservons pas la formulation qu'elles ont suggérée mais préférons définir une « mutation dirigée » comme une mutation qui est causée de manière spécifique et (exclusivement) adaptative par un processus physico-chimique en réponse aux conditions de l'environnement⁴⁷. Cette définition s'oppose à la vision consensuelle de la Synthèse Moderne (qui invoque la notion de « hasard évolutionnaire » pour caractériser les mutations génétiques), elle évoque la caractérisation lamarckienne de la variation héréditaire et, si l'on ajoute la condition selon laquelle le caractère spécifique et adaptatif de la mutation dite « dirigée » est dû à une tendance inhérente à l'organisme individuel porteur, elle correspond à la vision originelle lamarckienne de l'origine de la variation et de l'adaptation.

Nous concluons donc que, si une mutation est causée par un processus physico-chimique sans que l'environnement joue aucun rôle causal dans sa production, ou si une mutation est causée par un processus physico-chimique en réponse aux conditions de l'environnement mais non pas de manière spécifique, ou si une mutation est causée de manière spécifique par un processus physico-chimique en réponse aux conditions de l'environnement mais non pas de manière (exclusivement) adaptative, alors elle n'est pas « dirigée » mais aléatoire du point de vue évolutionnaire au sens de la notion de « hasard évolutionnaire ».

Le point de controverse entre la position soutenue par Jablonka et Lamb et la nôtre ne concerne que les caractéristiques définissant les différents types de mutations génétiques et les

⁴⁷ Nous renvoyons au chapitre 4.

critères qu'il est approprié d'utiliser pour les distinguer. En effet, nous pensons également que les mutations génétiques produites par les mécanismes mutateurs en relation à des conditions environnementales de stress ne sont, ni strictement équiprobables et indépendantes des conditions de l'environnement, ni « dirigées » du point de vue évolutionnaire, mais se situent entre ces deux extrêmes⁴⁸. Cependant, contrairement à Jablonka et Lamb, nous ne concevons pas les mutations dues aux mécanismes mutateurs comme fondamentalement différentes des autres changements héréditaires de la séquence d'ADN : elles ne diffèrent que par le fait d'être biaisées par des facteurs physico-chimiques différents car l'origine de ces derniers est très probablement due à l'histoire évolutive et sélective de la population naturelle concernée.

Pour pouvoir conclure que la vision consensuelle de la Synthèse Moderne s'applique encore aujourd'hui à toute altération de la séquence d'ADN et que toute mutation génétique est le fruit, à différents degrés et de manières diverses, d'un « hasard biaisé », nous devons montrer que la notion de « hasard évolutionnaire » s'applique aux mutations dues à tout mécanisme mutateur, en particulier aux mécanismes les plus controversés dont l'activité est induite en réponse à des stimuli environnementaux de stress et qui produisent une augmentation locale du taux de mutation.

Dans la section suivante, nous allons chercher à déterminer si la notion de « hasard évolutionnaire » s'applique au cas des mutations avantageuses observées et étudiées par Wright et ses collègues chez les bactéries *Escherichia coli* de la souche FC40⁴⁹. Une fois démontrée la validité actuelle de la vision consensuelle de la Synthèse Moderne, nous pourrions conclure que cette dernière, contrairement aux affirmations de Jablonka et Lamb, ne doit affronter aucun défi : la Théorie Synthétique ne nécessite aucun apport de la part de la vision lamarckienne quant à l'origine de la variation héréditaire pour intégrer en son sein et expliquer l'origine et le caractère des mutations génétiques dues aux mécanismes mutateurs.

⁴⁸ Jablonka et Lamb (2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, p. 99-101, notre traduction) expriment cette idée de la manière suivante : « on peut remarquer qu'il y a plusieurs types d'altérations de l'ADN qui ne rentrent pas proprement dans les descriptions en termes de 'aléatoire' ou de 'dirigée' [...], mais il y a beaucoup de cas intéressants dans la zone floue entre les deux extrêmes [...]; les altérations de l'ADN résultant d'une augmentation du taux de mutation liée au stress sont très proches de l'extrémité aveugle de l'axe, alors que les mutations locales et régionales résultant de changements physiologiques spécifiques sont semi-dirigées [...], « you can see that there are many type of DNA alteration that do not fit tidily with descriptions in terms of 'random', or 'directed' [...], but there are many interesting cases in the twilight zone between the two extremes [...]; those (DNA alterations) resulting from a stress-related global increase in mutation, are very close to the blind end of the axis, whereas local and regional mutations that results from specific physiological changes are semi-directed [...]. »

⁴⁹ Wright *et al*, 1999, « Hypermutation in derepressed operons of *Escherichia coli* K12 » ; Wright, 2000, « A Biochemical Mechanism for Nonrandom Mutations and Evolution ».

6.3. Pour le « hasard évolutif » de toute mutation génétique

Considérons à nouveau l'exemple fourni par Jablonka et Lamb d'un mécanisme mutateur (de type d)) à l'origine de mutations génétiques qui en aucun sens, selon leur analyse conceptuelle, ne peuvent être considérées comme « aléatoires » : revenons donc sur l'étude, réalisée par Wright et ses collègues⁵⁰, des mutations génétiques chez des bactéries *Escherichia coli* appartenant à une souche qui, à cause d'une mutation au niveau de l'opéron *leu*, est incapable de produire de la leucine. Rappelons que, dans des conditions environnementales caractérisées par l'épuisement de cet acide aminé, Wright et ses collègues ont observé une augmentation du taux de mutation ciblée au niveau des gènes de l'opéron leucine. Par conséquent, ils ont observé une augmentation de l'occurrence de la mutation inverse (de $LeuB^-$ à $LeuB^+$) permettant à ces bactéries de produire cet acide aminé par rapport à une situation où la leucine est naturellement présente dans le milieu où la population d'*Escherichia coli* se trouve.

6.3.1. Trois distinctions en faveur du « hasard évolutif » de toute mutation génétique

Dans le cadre de notre démonstration de l'applicabilité de la notion de « hasard évolutif » aux mutations dues à ce type de mécanisme mutateur – qui figure parmi les plus controversés au sein du débat sur le caractère des mutations génétiques puisqu'il produit une augmentation locale du taux de mutation en réponse à des stress environnementaux –, nous estimons nécessaire de rappeler et de clarifier trois distinctions particulièrement importantes au sein de ce genre de discussion d'ordre conceptuel. Dans leur analyse des mécanismes mutateurs, Jablonka et Lamb (2005) considèrent certaines des distinctions que nous allons introduire et, sur cette base, avancent des observations importantes portant sur les caractéristiques des mutations produites de manière spécifique en réponse à un stress environnemental et des mécanismes moléculaires à leur origine. Cependant, probablement à cause de leur tendance à vouloir démontrer le caractère en partie lamarckien de ce genre de phénomènes, elles ne semblent ne pas remarquer que ces trois distinctions peuvent jouer un rôle important dans le cadre d'un argument en faveur de la vision de la Synthèse Moderne

⁵⁰ Wright *et al.*, 1999, *op. cit.*

quant au « hasard évolutif » des mutations génétiques. C'est pourquoi nous avons décidé de prendre en considération ces distinctions et d'en mettre en évidence toute l'importance, ce qui nous permettra de renforcer notre critique de l'analyse conceptuelle proposée par Jablonka et Lamb au sujet du caractère des mutations génétiques dues aux mécanismes mutateurs.

La première distinction souligne la différence entre le mécanisme moléculaire à l'origine de l'augmentation (globale ou locale) du taux de mutation et chaque mutation particulière produite. Le premier consiste en une série d'états physico-chimiques reliés causalement, se déroulant au cours du temps et aboutissant à un état physico-chimique final. Une mutation particulière n'est qu'un des résultats possibles dans le cadre de l'augmentation du taux de mutation produite par le mécanisme mutateur. Cette distinction peut sembler triviale ; cependant, elle se révèlera particulièrement intéressante dans la suite de notre analyse.

En deuxième lieu, nous suggérons de distinguer les mécanismes mutateurs sur la base de leur origine. D'une part, un mécanisme mutateur peut être le résultat du processus évolutif : il peut être une adaptation, au sens qu'il est un produit de l'évolution par sélection naturelle, ou il peut être un sous-produit de l'histoire évolutive de la population concernée, par exemple un résultat dérivé du processus de sélection pour d'autres caractéristiques phénotypiques qui sont complètement indépendantes de la présence et de l'activité du mécanisme mutateur. D'autre part, l'augmentation temporaire (globale ou locale) du taux de mutation peut aussi être la simple manifestation de l'état pathologique d'organismes devant faire face à des changements importants de l'environnement : dans ce cas, on parle d'un mécanisme de mutation qui n'est pas le résultat du processus évolutif.⁵¹ En combinant la distinction des mécanismes mutateurs (pouvant être un produit de l'évolution par sélection naturelle, un sous-produit de l'évolution, ou la manifestation d'un état pathologique) selon leur origine avec le fait que les mécanismes mutateurs peuvent être constitutifs ou induits et qu'ils peuvent produire une augmentation globale et/ou locale du taux de mutation, nous disposons d'un ensemble d'outils empiriques et conceptuels précis pour discerner et classifier chaque mécanisme moléculaire de mutation.

⁵¹ En effet, nous avons vu que l'hypothèse adaptative et, plus généralement, évolutive de l'origine des mécanismes mutateurs n'est pas la seule explication possible : en dépit de nombreuses preuves en faveur de la première, l'hypothèse pathologique n'a pas encore été refusée de manière concluante.

Mécanismes/ Biais (Exemples)	NON Evolués*	Evolués** (par SN)	Globaux	Locaux	Constitutifs	Induits
L'hypermutabilité ciblée dans un site ou une région génomique		X (selon Wright <i>et al</i> , 1999)		X		X
Les « points chauds » de mutation		X		X	X	
La réponse SOS		X	X			X
L'inactivation du système de réparation		X	X		X	
Les agents physico-chimiques	X			X		X
...

Figure 5 : Les mécanismes mutateurs et les biais de mutation.

*Mécanismes et biais de mutation « NON évolués », c'est-à-dire dus à des contraintes physico-chimiques ou à un état pathologique.

**Mécanismes et biais de mutation « Evolués (par SN) », c'est-à-dire qu'ils sont des sous-produits du processus évolutif ou des résultats adaptatifs de l'évolution par sélection naturelle.

En troisième lieu, il est important de garder toujours à l'esprit la distinction entre la « théorie sélective » et la « théorie instructive » de l'origine de l'adaptation⁵², autrement dit entre la conception darwinienne, adoptée par la Synthèse Moderne, et la conception proprement lamarckienne de l'adaptation et, plus généralement, de l'évolution.

6.3.2. La notion de « hasard évolutionnaire » dans le cas des mutations les plus controversées (Wright *et al*, 1999)

Analysons maintenant les mutations génétiques permettant aux bactéries *Escherichia coli* porteuses de la mutation *leuB*⁻ de survivre et de se reproduire dans un milieu où l'acide aminé de la leucine est absent. Examinons-les afin de comprendre s'il s'agit de « mutations dirigées » ou si, au contraire, elles sont le fruit du « hasard évolutionnaire », et cela en gardant à l'esprit les quatre trois distinctions que nous venons d'introduire ci-dessus.

Dans le chapitre 4 de ce travail, nous avons conclu qu'une mutation est le fruit du « hasard évolutionnaire » si et seulement si elle n'est pas une mutation « dirigée », autrement dit, si et seulement si elle ne remplit pas au moins une des deux conditions suivantes : 1) être

⁵² Dawkins, 1982, *The Extended Phenotype*.

plus probable dans un environnement E_1 où elle est avantageuse que dans un environnement E_2 où elle est défavorable ou neutre ; 2) être est nettement plus probable dans un environnement donné E_1 où elle est avantageuse que d'autres mutations neutres ou défavorables. Par exemple, si une mutation est plus probable dans un environnement E_1 où elle est avantageuse que dans un environnement E_2 où elle est défavorable ou neutre mais, en même temps, elle n'est pas nettement plus probable dans un environnement donné E_1 où elle est avantageuse que d'autres mutations neutres ou défavorables, elle peut être considérée comme le fruit du « hasard évolutionnaire », et non pas comme une mutation « dirigée » par rapport au processus d'évolution.

Que peut-on affirmer à ce propos dans le cas des mutations observées et étudiées par Wright et ses collègues (1999) ? Plus précisément, les mutations permettant à des bactéries *Escherichia coli* qui ne pouvaient à l'origine pas produire de la leucine de synthétiser cet acide aminé, et cela alors qu'elles doivent faire face à une situation environnementale caractérisée par l'épuisement de cet acide aminé, sont-elles causées de manière spécifique et (exclusivement) adaptative par un processus physico-chimique en réponse à ces conditions de l'environnement et peuvent-elles donc être considérées comme des mutations « dirigées » du point de vue évolutionnaire ? Ces mutations, représentent-elles un défi à l'idée fondamentale de la Synthèse Moderne selon laquelle toute mutation génétique est le produit du « hasard évolutionnaire » ? Jablonka et Lamb, ont-elles raison lorsqu'elles affirment que la théorie de l'évolution nécessite l'intégration de processus lamarckiens afin d'expliquer ce genre de phénomènes mutationnels à l'origine de la variation génétique héréditaire ?

Plaçons-nous au niveau des bactéries individuelles et considérons l'occurrence de la mutation inverse et avantageuse au niveau du gène *leuB*. Cette altération de la séquence d'ADN au sein de l'opéron leucine permet à ces microorganismes de produire de la leucine lorsqu'elles doivent faire face à l'épuisement de cet acide aminé (alors qu'elles n'en étaient pas capables) et leur permet donc de survivre, de se reproduire et de transmettre la variation favorable ($LeuB^+$) à leur descendance. Cette mutation est la conséquence de l'augmentation du taux de mutation au niveau de l'opéron leucine : elle est due à l'état physico-chimique de la molécule d'ADN au niveau de ce site de la séquence (plus précisément, l'augmentation du taux de mutation à cet endroit du génome est due à la présence de portions d'ADN simple brin, au super-enroulement négatif de la double hélice, à la présence de structures secondaires caractérisées par des mésappariement des bases nucléotidiques, etc.). Cet état physico-chimique est à son tour dû à la dérégulation des gènes de l'opéron *leu* et à l'activation de leur transcription. Nous pouvons donc affirmer, d'une part, que la mutation avantageuse

permettant à une bactérie individuelle de synthétiser de la leucine est causée par un processus physico-chimique déclenché en réponse à un stimulus de stress environnemental ; d'autre part, que cette mutation avantageuse est produite de manière spécifique, c'est-à-dire qu'elle fait partie d'une augmentation locale, ciblée sur l'opéron *leu*, du taux de mutation.

Sur la base de tous ces éléments, nous pouvons conclure que cette mutation avantageuse ne remplit que la première condition pour qu'une mutation soit « dirigée » et peut donc être considérée comme un produit du « hasard évolutif ». En effet, la mutation inverse transformant les bactéries de $LeuB^-$ à $LeuB^+$ est plus probable dans un environnement E_1 où cet acide aminé est épuisé et donc où elle est avantageuse, que dans un environnement E_2 , par exemple dans un milieu riche en leucine, où elle est neutre (ou défavorable)⁵³ : cela est dû à l'augmentation locale du taux de mutation au niveau de l'opéron *leu* en réponse à l'épuisement de la leucine dans l'environnement où les bactéries se trouvent. En revanche, cette mutation inverse et avantageuse n'est pas nettement plus probable que d'autres mutations neutres ou défavorables dans le même environnement E_1 caractérisé par l'absence de leucine.

À ce propos, il faut rappeler que Wright et ses collègues⁵⁴ n'ont pas observé une augmentation du taux de la mutation inverse au niveau du seul gène *leuB* qui, en mutant, permet aux bactéries de survivre et de se reproduire ; ils ont aussi relevé une augmentation du taux de mutation au niveau des tous les autres gènes constituant l'opéron leucine. La principale raison à l'origine de l'apparent caractère « dirigé » de cette mutation avantageuse chez *Escherichia coli* réside dans le fait que seule cette mutation au niveau du gène *leuB* peut permettre aux bactéries appartenant à cette souche de survivre et de se reproduire dans un environnement caractérisé par l'absence de cet acide aminé. Les bactéries individuelles porteuses d'altérations diverses au niveau des autres gènes de l'opéron *leu* croissent plus lentement ou sont sélectionnées négativement et, par conséquent, sont difficiles à détecter au cours du temps, alors que les bactéries individuelles porteuses de la mutation inverse avantageuse au niveau du gène *leuB* survivent, sont sélectionnées positivement et contribuent

⁵³ Cependant, nous avons déjà affirmé que, si nous réalisons la même étude expérimentale avec des bactéries $LeuB^+$ capables de synthétiser de la leucine, nous pourrions observer que le taux de mutation augmente dans ce cas aussi au niveau de l'opéron leucine à cause de sa dérégulation induite par l'absence de cet acide aminé. De plus, nous pourrions remarquer que la probabilité du même type d'altération génétique, dans ce cas la réversion transformant les bactéries de $LeuB^+$ à $LeuB^-$ (et non pas de $LeuB^-$ à $LeuB^+$), est plus élevée que dans une situation environnementale différente, alors qu'elle n'apporte aucun avantage sélectif à la population de bactéries concernée dans ce cas. Sur cette base, nous pourrions conclure que la condition 1) pour qu'une mutation soit « dirigée » n'est pas véritablement remplie non plus par les mutations avantageuses au centre de l'étude de Wright et ses collègues

⁵⁴ Wright *et al.*, 1999, « Hypermutation in derepressed operons of *Escherichia coli* K12 », p. 5093.

à la formation de la génération suivante ; elles peuvent donc être observées et quantifiées plus facilement.

De la même manière, nous pourrions aussi bien montrer que les mutations provoquées par tous les autres types de mécanismes mutateurs en relation aux conditions de l'environnement ne sont pas causées de manière spécifique et (exclusivement) adaptative (c'est-à-dire qu'elles ne sont pas « dirigées » par rapport au processus évolutif), mais sont le fruit du « hasard évolutionnaire ». Cependant, nous ne nous attarderons pas à montrer la validité de la vision consensuelle de la Synthèse Moderne quant au caractère des mutations pour chaque type de mécanisme mutateur : le fait que notre argument est valable dans le cas du mécanisme de mutation le plus controversé nous permet de prendre cette liberté.

Nous tenons à préciser que nous ne sommes nullement en train d'affirmer l'impossibilité physique des mutations « dirigées » en vue de l'adaptation des organismes individuels porteurs. En revanche nous remarquons tout d'abord que les mutations dues aux mécanismes mutateurs ne sont pas des mutations « dirigées » du point de vue évolutionnaire et que, de toute manière, ce dernier type de mutation n'a encore été observé chez aucun organisme vivant. Ensuite, l'existence ou l'apparition d'un mécanisme provoquant des mutations de manière spécifique et (exclusivement) adaptative (c'est-à-dire, d'un mécanisme provoquant des mutations « dirigées ») nous semble très improbable : en effet, le caractère adaptatif d'une mutation dépend d'un ensemble très complexe de facteurs environnementaux parmi lesquels figurent le contexte moléculaire intracellulaire où la mutation se produit, l'environnement interne et externe à l'organisme porteur, les autres organismes présents dans le même environnement, etc. De plus, l'environnement n'est pas stable ; il est plutôt composé par des facteurs qui changent sans arrêt et de manière plus ou moins rapide. Cela rend très peu plausible l'évolution d'un mécanisme de mutation « dirigée » via un processus de sélection cumulative au cours du temps.

Soulignons enfin que, si un jour une mutation « dirigée » était observée et faisait l'objet d'une confirmation expérimentale, elle ne serait le résultat d'un processus lamarckien de mutation que de manière apparente : son caractère spécifique et (exclusivement) adaptatif ne pourrait qu'être dû à l'histoire, évolutive et sélective, de la population à laquelle l'organisme individuel concerné appartient, et non pas à une sorte de tendance adaptative indépendante de tout processus évolutif, inhérente à l'organisme individuel concerné et à l'origine de l'adaptation de ce dernier et de ses descendants.

6.3.3. La valeur conceptuelle et empirique de la notion de « hasard évolutionnaire »

Pourquoi les définitions et la distinction que nous proposons des mutations « dirigées » et des mutations qui sont le fruit du « hasard évolutionnaire » devraient-elles être préférées aux définitions et à la distinction des mutations « aléatoires » et « non aléatoires » proposées par Jablonka et Lamb (2005) ?

<i>J&L 2005</i>	<i>J&L 2005</i>		<i>MERLIN</i>	<i>MERLIN</i>
Mutations « aléatoires » (Synthèse Moderne)	Mutations « totalement aveugles »		Mutations équiprobables et aucunement spécifiques	Physiquement possibles mais peu probables
Mutations « NON aléatoires » (en partie « lamarckiennes »)	Mutations « semi-dirigées » (dus à des mécanismes mutateurs divers)		Mutations au « hasard évolutionnaire »	Mutations au « hasard biaisé » par : -mécanismes/biais NON évolués (constitutifs/induits ; globaux/locaux) -mécanismes/biais évolués (par SN) (constitutifs/induits ; globaux/locaux)
	Mutations « totalement dirigées » (développementales)		Mutations « dirigées »	Pas observées mais physiquement possibles (en laboratoire)
			Mutations « lamarckiennes »	Pas possibles
Critères	Spécificité & Causalité		Critères	Spécificité & Causalité + Caractère Adaptatif

Figure 6 : Notre analyse des mutations génétiques en comparaison avec la vision de Jablonka et Lamb (2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*)⁵⁵.

⁵⁵ Nous soulignons que, dans ce tableau, un même terme peut assumer deux significations différentes, l'une dans l'analyse de Jablonka et Lamb, l'autre au sein de notre proposition. Pour plus de précisions quant à la définition de chaque terme, nous renvoyons au texte de ce chapitre.

Nous ne soutenons pas que la valeur des résultats de notre analyse conceptuelle réside dans le fait qu'ils correspondent à la vision consensuelle de la Synthèse Moderne ; cependant, nous soulignons à nouveau qu'il est nécessaire d'identifier et de définir de manière précise quelle est cette vision consensuelle quant au caractère des mutations génétiques avant de pouvoir la contester ou de la défendre : c'est ce que nous avons essayé de faire dans le chapitre 4. Les raisons nous ayant conduit à soutenir notre proposition et à évaluer de manière critique celle de Jablonka et Lamb sont liées aux trois distinctions que nous avons introduites dans la section précédente. Ces dernières nous paraissent constituer des éléments décisifs dans la résolution du débat sur le caractère des mutations génétiques du point de vue évolutionnaire.

En premier lieu, l'application de la notion de « hasard évolutionnaire » à toute mutation génétique met particulièrement en évidence la différence entre le mécanisme de mutation, qui peut être un produit adaptatif de l'évolution, et les mutations particulières que le mécanisme produit, qui au contraire ne peuvent pas être considérées comme des résultats adaptatifs de l'histoire évolutive. Jablonka et Lamb traitent aussi cette distinction et, en particulier, soulignent à plusieurs reprises la possibilité d'une origine évolutive et/ou sélective des mécanismes mutateurs et de la consécutive spécificité des mutations qu'ils provoquent. Cependant, elles ne considèrent pas la portée de la distinction entre mécanisme mutateur et mutations particulières dans une analyse conceptuelle du caractère des mutations génétiques.

De manière plus précise, la notion de « hasard évolutionnaire » rend compte des deux éléments suivants. D'une part, les mutations particulières produites à cause de l'activation d'un mécanisme de mutation en relation, et même en réponse, à des conditions environnementales de stress sont des phénomènes concernant l'histoire individuelle des organismes appartenant à la population concernée. Ces mutations sont le résultat d'un mécanisme physiologique ayant lieu à l'intérieur de chaque organisme individuel et peuvent être avantageuses, défavorables ou neutres, avec des probabilités différentes, en dépendance du contexte environnemental et de leur spécificité : elles ne sont donc pas des adaptations ou, autrement dit, des produits adaptatifs de l'histoire évolutive de la population concernée.

D'autre part, les mécanismes mutateurs produisant une augmentation (globale ou locale) du taux de mutation en relation, et même en réponse, à des stress environnementaux concernent aussi l'histoire développementale (ontogénique et physiologique) de chaque organisme individuel. De plus, si l'on se place du point de vue de l'hypothèse adaptative quant à l'origine de ces mécanismes moléculaires de mutation – qui est devenue de plus en plus plausible dans le cas de certains mécanismes mutateurs –, ces derniers et l'augmentation du taux de mutation qu'ils provoquent peuvent être considérés comme des résultats de

l'histoire évolutive de la population à laquelle les organismes concernés appartiennent et, par conséquent, comme des phénomènes concernant le niveau populationnel. Si l'hypothèse adaptative se révélait être correcte dans le cas de tout mécanisme mutateur, on pourrait donc conclure que leur origine et leur caractère adaptatifs expliquent l'apparent caractère « dirigé » des mutations qu'ils produisent. Plus précisément, on pourrait expliquer le caractère spécifique des mutations dues aux mécanismes mutateurs – en ce qui concerne le moment où elles ont lieu et/ou l'endroit de la séquence génomique concerné et/ou la fonction affectée – par le lien causal et adaptatif qui s'est instaurée au cours de l'histoire évolutive entre les mécanismes moléculaires de mutation (et non pas les mutations particulières) et certains types de stimuli de stress environnemental.

La deuxième raison en faveur du caractère approprié de la notion de « hasard évolutionnaire » dans la caractérisation de toute mutation génétique et, par conséquent, des mutations dues aux mécanismes mutateurs, renvoie à la distinction entre différents types de mécanismes et, plus généralement, de biais de mutation. Nous avons remarqué qu'un mécanisme mutateur peut être constitutif ou induit, qu'il peut provoquer une augmentation globale ou locale du taux de mutation et qu'il peut être un résultat adaptatif de l'évolution, un sous-produit du processus évolutif ou une simple manifestation d'un état pathologique. Ces distinctions, qui peuvent également être appliquées à tout biais de mutation, nous permettent de comprendre l'apparent caractère « dirigé » du point de vue évolutionnaire de certaines mutations qui sont au centre du défi posé par Jablonka et Lamb à la Synthèse Moderne et qui ont fait l'objet d'un débat controversé au cours des vingt dernières années.

Cette manière de distinguer les divers mécanismes et biais de mutation – en tenant compte du fait que, tout en étant physiquement possible, premièrement, aucune mutation « dirigée » au sens évolutionnaire (en apparence due à un processus lamarckien) n'a jamais été observée et, deuxièmement, aucune mutation n'est strictement équiprobable et indépendante des conditions de l'environnement au sens large –, nous permet d'identifier la place que chaque mutation occupe entre ces deux extrêmes⁵⁶. Nous voulons souligner que, non seulement toute mutation génétique est le fruit du « hasard évolutionnaire », mais que toutes les mutations génétiques sont, d'une manière ou d'une autre, biaisées : leurs différences résident dans le caractère constitutif ou induit, global ou local, adaptatif ou non adaptatif, des mécanismes mutateurs et, plus généralement, des processus physico-chimiques qui les

⁵⁶ Nous avons déjà vu que Jablonka et Lamb mobilisent aussi cette image, mais caractérisent les deux extrêmes de manière différente, l'un étant occupé par ce qu'elles appellent les mutations « totalement aléatoires » (qui correspondent, à leur avis, à la vision de la Synthèse Moderne), l'autre par les mutations « totalement dirigées » (qui ont lieu au cours du développement). Voir plus haut.

produisent de manière plus ou moins biaisée⁵⁷. Cette manière de caractériser les mutations est parfaitement cohérente et compatible avec la vision consensuelle de la Synthèse Moderne.

La troisième raison pour laquelle notre proposition en faveur de la validité actuelle de la vision de la Synthèse Moderne quant au caractère des mutations génétiques est préférable au défi que Jablonka et Lamb lui lancent nous permet de revenir sur la vision lamarckienne et sur les idées fondamentales sur lesquelles elle repose. Il est important de rappeler et de souligner la différence fondamentale entre ce que Dawkins⁵⁸ appelle « la théorie sélective » et la « théorie instructive » de l'origine de l'adaptation. Selon la première, qui correspond à la vision darwinienne adoptée par les biologistes de la Synthèse Moderne, l'adaptation des organismes à leur environnement augmente via le processus de sélection naturelle. Selon la « théorie instructive », qui correspond à la conception de l'origine de la variation et de l'adaptation avancée par Lamarck et qui a été reprise au cours des années par les lamarckiens les plus fidèles à ce dernier, l'environnement est impliqué causalement dans la production de la « bonne » variation, c'est-à-dire de la variation avantageuse. Plus précisément, selon cette deuxième conception, l'adaptation des organismes augmente au cours du temps en vertu du lien causal, spécifique et adaptatif, entre les changements affectant les caractéristiques des organismes individuels et les conditions environnementales où ces derniers se trouvent.

Jablonka et Lamb affirment que les mécanismes mutateurs, en particulier les mécanismes induits par un stimulus de stress environnemental et produisant une augmentation locale du taux de mutation, sont des exemples de mécanismes à la fois sélectifs, car ils doivent leur origine au processus d'évolution par sélection naturelle, et instructifs donc lamarckiens, car ils produisent des mutations au moment et à l'endroit du génome où ces dernières peuvent apporter un avantage sélectif aux organismes concernés. Cependant, sur la base de notre analyse, nous pouvons conclure que leur affirmation n'est pas correcte si par « instructif » et par « lamarckien » elles renvoient à la « théorie instructive » proprement lamarckienne de l'origine de l'adaptation. Nous avons remarqué à plusieurs reprises que ce n'est pas le sens que Jablonka et Lamb attribuent à ces deux termes et qu'elles les utilisent de manière inappropriée du point de vue historique, ce qui représente le point de départ de notre critique à leur analyse.

Ajoutons une dernière raison pour laquelle la notion de « hasard évolutionnaire » est préférable à la notion de « semi-directionnalité » proposée par Jablonka et Lamb pour

⁵⁷ C'est pourquoi nous proposons d'utiliser l'expression « hasard biaisé » pour caractériser toute mutation génétique. Davis (1989, « Transcriptional bias: A non-Lamarckian mechanism for substrate-induced mutations ») a déjà utilisé de manière semblable cette même expression.

⁵⁸ Dawkins, 1982, *The Extended Phenotype*, p. 173.

caractériser les mutations dues aux mécanismes mutateurs. Nous soulignons que, de la même manière que toute mutation génétique, les mutations produites de manière plus ou moins spécifique par ces mécanismes moléculaires en relation, et même en réponse directe, aux conditions de l'environnement n'influencent pas directement la trajectoire de l'évolution : ces mutations peuvent être avantageuses, défavorables ou neutres avec des probabilités différentes, mais sans que la probabilité qu'elles apportent un avantage aux organismes individuels porteurs soit nettement plus élevée que la probabilité qu'elles soient défavorables ou neutres.⁵⁹ Elles n'imposent aucune direction particulière et adaptative au processus évolutif d'une population donnée au cours du temps, ce qui serait vrai, au contraire, dans le cas d'une mutation « dirigée » au sens lamarckien. Par contre, nous ne critiquons pas Jablonka et Lamb lorsqu'elles affirment que les mécanismes mutateurs eux-mêmes, en tant que résultats adaptatifs de l'évolution par sélection naturelle et en tant qu'éléments susceptibles d'évoluer et de s'affiner ultérieurement, tendent à influencer le cours de l'évolution (même si ce n'est pas vers une direction nécessairement adaptative). En effet, ils tendent à produire plus de mutations quand et où la population concernée en a besoin pour pouvoir espérer ne pas disparaître à la génération suivante⁶⁰.

Nous avons donc montré dans ce chapitre que, de manière cohérente avec la vision consensuelle de la Synthèse Moderne, *toute* mutation génétique (y compris les mutations produites par les mécanismes mutateurs en relation aux conditions de l'environnement) est le fruit du « hasard évolutionnaire », ce dernier étant toujours, de manières différentes et à des degrés différents, un « hasard biaisé ». Nous avons défendu la vision consensuelle de la Synthèse Moderne contre les attaques d'un certain nombre de biologistes, d'historiens et de philosophes de la biologie, en particulier contre le défi posé par Jablonka et Lamb (2005). Nous avons exposé les raisons pour lesquelles les définitions et les distinctions qu'elles suggèrent ne sont pas appropriées du point de vue conceptuel : notre principale critique a visé le caractère historiquement inapproprié de la conception qu'ont Jablonka et Lamb de la Synthèse Moderne, d'une part, et de la vision lamarckienne de l'origine et du caractère de la variation héréditaire, d'autre part. Nous avons ensuite donné des arguments montrant que notre analyse de ces questions mène à des définitions et des distinctions empiriquement et conceptuellement plus appropriées.

⁵⁹ On pourrait certes nous objecter que les mutations génétiques influencent la trajectoire évolutive car elles représentent la matière brute à partir de laquelle l'évolution peut continuer son cours. Nous admettons et reconnaissons la validité de cette objection et, en même temps, nous soulignons qu'elle ne consiste pas à dire que les mutations génétiques influencent directement le cours du processus évolutif.

⁶⁰ Voir aussi Wright, 2000, « A Biochemical Mechanism for Nonrandom Mutations and Evolution ».

Nous concluons que la théorie classique de l'évolution nécessite l'intégration des avancées relativement récentes quant à l'origine de la variation héréditaire. En particulier, elle doit intégrer en son sein, premièrement, l'importance des variations génétiques se produisant au cours du processus développemental des organismes individuels et, deuxièmement, le rôle causal que l'histoire évolutive et sélective d'une population donnée peut jouer relativement au caractère plus ou moins spécifique et adaptatif du processus de mutation génétique. Nous concluons aussi que le paradigme de la Synthèse Moderne peut répondre de manière efficace aux défis qu'on lui a posés sur la base de ces avancées : il peut incorporer la découverte des mécanismes mutateurs et les implications qu'elle comporte tout en préservant ses notions fondamentales, en particulier la notion de « hasard évolutif », sa validité empirique et son pouvoir explicatif relativement à tout phénomène biologique observé jusqu'à maintenant.⁶¹

⁶¹ De la même manière que Jablonka et Lamb, nous reprenons l'affirmation de Pasteur dans la reformulation qu'en a proposée Caporale : « le hasard favorise le génome préparé ». En effet, elle exprime de manière efficace notre conclusion, qui est la suivante : sur la base de tous les éléments théoriques et expérimentaux dont nous disposons aujourd'hui, toute mutation génétique – même ces mutations provoquées de manière hautement spécifique, dans le temps et au niveau de la séquence d'ADN, en vertu de la préparation acquise par le génome au cours de son histoire évolutive – est le produit du hasard du point de vue du cours adaptatif de l'évolution.

Chapitre 7 – Le hasard des mutations génétiques : du niveau évolutif au niveau moléculaire

La notion de « hasard évolutif » correspond à la manière dont les mutations génétiques furent caractérisées dans le cadre de la Théorie Synthétique Moderne de l'Évolution (1930-1950) et, plus généralement, à la vision consensuelle de la communauté des biologistes des années 1930 jusqu'à aujourd'hui. Dans les chapitres précédents, nous avons montré que cette notion est centrée sur la relation entre mutation, sélection et adaptation, c'est-à-dire qu'elle définit le caractère des mutations génétiques du point de vue propre à la biologie de l'évolution, et nous avons défendu la thèse selon laquelle cette notion est appropriée, à la fois du point de vue conceptuel et empirique, pour caractériser toute mutation génétique.

Nous avons déjà annoncé à plusieurs reprises que la perspective évolutif n'est pas la seule manière dont les biologistes considèrent les mutations génétiques. Des termes comme « aveugle », « fortuit », « aléatoire », « stochastique », « probabiliste », « statistique », etc., caractérisent l'usage biologique d'autres notions de hasard qui sont plus générales que la notion de « hasard évolutif » car elles ne font pas nécessairement référence à la relation entre les mutations et leur utilité du point de vue de l'évolution. Par exemple, en génétique des populations¹, la notion de « hasard » est aussi utilisée pour décrire la manière dont les mutations génétiques sont distribuées au sein des populations naturelles. En génétique classique² et en génétique moléculaire³, elle caractérise la distribution du nombre de mutations le long du génome et les processus physico-chimiques à leur origine causale⁴.

¹ Le but principal de la génétique des populations est de comprendre la manière dont les fréquences géniques changent au cours du temps au sein d'une population donnée et de décrire en ces termes le processus évolutif.

² Nous utilisons l'expression « génétique classique » pour désigner les recherches sur la variation héréditaire qui ont eu lieu du début du XX^e siècle jusqu'à la fin des années 1930, notamment les travaux de l'équipe de Morgan (voir Dunn, 1965, *A Short History of Genetics. The development of some of the main lines of thought: 1864-1939*). De manière générale, la génétique classique est l'étude de la transmission des traits au cours des générations : elle explique la transmission des différences phénotypiques en termes de l'hérédité des différences génotypiques d'une génération à la suivante.

³ La génétique moléculaire, dont on peut situer la naissance au moment de la découverte de la structure physico-chimique de l'ADN autour des années 1950, vise à comprendre les mécanismes moléculaires sous-jacents à la variation génétique, à son processus d'expression, à sa réplication et à sa transmission.

⁴ Voir le chapitre 3.

Dans le cadre de ces trois disciplines biologiques, l'analyse statistique et le calcul des probabilités sont massivement mobilisés. D'une part, au sein des recherches en génétique des populations, ils sont utilisés pour décrire la distribution de la variation génétique au sein des populations naturelles et pour prédire le changement des fréquences relatives des gènes au niveau populationnel au cours du temps. D'autre part, en génétique classique et en génétique moléculaire, on utilise l'analyse statistique et le calcul des probabilités pour estimer les taux de divers types de mutations à des échelles spatio-temporelles diverses⁵, afin d'évaluer leur distribution dans le temps et le long du génome. Nous verrons que, dans ce cadre, les mutations génétiques sont caractérisées par des notions de hasard définissables en termes de probabilités.

Nous avons décidé de concentrer notre enquête sur la recherche en génétique moléculaire.⁶ C'est en effet à l'échelle moléculaire qu'on dispose aujourd'hui de descriptions et d'explications assez satisfaisantes, bien qu'encore lacunaires, des mécanismes causaux à l'origine des mutations. Notre enquête va donc porter sur les mutations définies à l'échelle moléculaire en tant qu'altérations de la séquence nucléotidique de l'ADN : le changement d'une seule paire de bases nucléotidiques, l'insertion ou la délétion d'une ou plusieurs bases, la duplication (ou amplification) d'un segment de la séquence du chromosome, etc. Nous voulons identifier la notion de hasard que les généticiens moléculaires utilisent pour caractériser les événements mutationnels, en définissant précisément sa signification et son rôle : notre but est de comprendre si cette notion est appropriée, à la fois d'un point de vue conceptuel et empirique, pour caractériser les mutations relativement aux processus physico-chimiques qui les causent. En effet, seule une notion de hasard décrivant les caractéristiques du processus causal à l'origine d'un certain résultat mutationnel (par exemple, d'une altération particulière de la séquence nucléotidique ou d'une certaine distribution des mutations sur un laps de temps donné et/ou le long de la séquence d'ADN) peut prétendre fournir une explication du caractère aléatoire de ce dernier. Une définition de la notion de hasard qui ne prendrait en compte que les propriétés d'un résultat donné n'en fournirait pas une explication mais une simple description.

Dans ce but, nous introduirons deux notions de hasard – que nous appellerons respectivement « aléatoire (au sens) fort » et « aléatoire (au sens) faible ». La première reflète la manière dont les mutations génétiques sont conçues par les généticiens moléculaires

⁵ Voir le chapitre 3.

⁶ Dans ce chapitre, nous évoquerons aussi et à plusieurs reprises la génétique des populations, en particulier tout ce que les généticiens des populations supposent quant au processus de mutation génétique.

lorsqu'ils procèdent à l'évaluation du taux de mutation par des méthodes statistiques⁷. La deuxième correspond à une conception moléculaire des mutations prenant en compte tous les facteurs physico-chimiques pouvant influencer le processus de mutation. À la lumière de tous les biais de mutation (que nous avons déjà présentés et analysés dans les chapitres précédents) et de quelques études expérimentales récentes sur les mutations génétiques, nous allons suggérer que la notion d'« aléatoire (au sens) faible » est appropriée pour caractériser toute mutation génétique selon une perspective moléculaire.

7.1. Les notions d'« aléatoire (au sens) fort » et d'« aléatoire (au sens) faible »

Dans le cadre de l'étude des mutations en génétique moléculaire, la notion de hasard semble aller de pair avec une certaine imprédictibilité de ces phénomènes physico-chimiques. Le site, le type, le moment où les mutations se produisent et leur taux d'occurrence sont habituellement estimés à partir des données expérimentales sur la variation au sein des populations naturelles. Les généticiens moléculaires peuvent prédire de manière assez fiable, mais probabiliste, les taux caractéristiques de différents types de mutations génétiques et sont confrontés à des problèmes non négligeables dans la prédiction, qui est toujours probabiliste, du site, du type et du moment d'occurrence des mutations.

Le caractère probabiliste de ces prédictions soulève la question suivante : est-il dû aux caractéristiques propres des processus physico-chimiques à l'origine causale des mutations, ou est-il relatif aux connaissances dont les généticiens moléculaires disposent des causes sous-jacentes aux mutations génétiques ? En d'autres termes, les probabilités doivent-elles être comprises de manière objective (par exemple, comme la mesure du degré de possibilité de ces événements), ou de manière épistémique (par exemple, comme la mesure du degré de croyance rationnelle des généticiens moléculaires quant au processus de mutation) ? Dans ce chapitre, nous ne chercherons cependant pas à répondre à cette question. Nous visons plutôt à identifier la signification précise et le rôle de la notion de hasard appropriée pour caractériser au niveau moléculaire les mutations génétiques.

⁷ Nous verrons que la notion d'« aléatoire (au sens) fort » correspond aussi à la manière dont les mutations génétiques sont conçues en génétique des populations, dans le cadre de la modélisation du changement des fréquences génétiques.

Introduisons et analysons les notions de hasard que les biologistes pourraient invoquer pour caractériser les mutations génétiques, non selon une perspective évolutionnaire, mais du point de vue moléculaire, c'est-à-dire en tant qu'altérations de la séquence nucléotidique de l'ADN et relativement au processus physico-chimique qui les cause.

D'abord, nous suggérons d'employer le terme « aléatoire » afin de distinguer ces notions de hasard de la notion de « hasard évolutionnaire ». Ensuite, nous proposons de distinguer deux notions de l'aléatoire sur la base de leur extension respective. D'une part, la notion d'« aléatoire (au sens) fort », qui est strictement liée à la notion mathématique de hasard en tant qu'équiprobabilité – ou, en termes de processus, en tant qu'échantillonnage indiscriminé⁸ – car elle la contient nécessairement dans sa définition. Nous allons montrer que cette notion n'est pas appropriée du point de vue empirique pour caractériser les mutations au niveau moléculaire (ou, en tout cas, la plupart des mutations génétiques physiquement possibles). D'autre part, la notion d'« aléatoire (au sens) faible »⁹, dont nous allons soutenir la pertinence pour caractériser toute mutation génétique au niveau moléculaire (ou, du moins, la plupart des mutations génétiques physiquement possibles).

Définissons de manière précise ces deux notions de l'aléatoire et déclinons-les dans le cas des mutations. Les deux notions définissent le caractère aléatoire d'un événement, non par la description de ses résultats, mais en référence aux caractéristiques du processus causal qui est à son origine.

1-Un événement est « *aléatoire (au sens) fort* » si et seulement si il est le résultat d'un processus stochastique¹⁰ causal qui remplit les deux conditions suivantes ($C1 \wedge C2$) :

C1 : C'est un processus d'échantillonnage indiscriminé où nulle différence dans les caractéristiques des éléments pouvant être échantillonnés ne joue un rôle causal. La probabilité d'être échantillonné est donc la même pour tout élément.

⁸ Voir le chapitre 2.

⁹ Sarkar (1991a, « Lamarck *contre* Darwin, Reduction *versus* Statistics: Conceptual Issues in the Controversy Over Directed Mutagenesis in Bacteria ») utilise l'expression « notion faible d'aléatoire » mais ne désigne pas ce que nous appelons l'« aléatoire (au sens) faible ». En mobilisant cette expression, Sarkar renvoie à sa définition de la notion de hasard caractérisant les mutations génétiques du point de vue évolutionnaire, c'est-à-dire par rapport à l'adaptation. Pour une analyse critique de la « notion faible d'aléatoire » proposée par Sarkar, voir Millstein, 1997, *The Chances of Evolution: An Analysis of the Roles of Chance in Microevolution and Macroevolution*.

¹⁰ Nous définissons un processus stochastique comme un processus qui, à partir d'un même ensemble de conditions initiales, produit plusieurs résultats possibles selon une certaine loi de probabilité. Dans ce cadre, nous ne soutenons aucune position particulière quant à la nature déterministe ou indéterministe, d'un point de vue métaphysique, d'un processus stochastique et, plus spécifiquement, du processus de mutation génétique.

C2 : C'est un processus invariant dans le temps. La probabilité qu'un élément soit échantillonné est donc la même à tout instant, ce qui implique son indépendance par rapport à tout autre événement (toute combinaison d'événements incluse).

En d'autres termes, un événement « aléatoire (au sens) fort » est le résultat d'un processus d'échantillonnage indiscriminé avec remplacement, c'est-à-dire analogue à une série de tirages de boules dans une urne, où toutes les boules ont la même probabilité d'être tirées et où la boule tirée est systématiquement remise dans l'urne.

2-Un événement est « *aléatoire (au sens) faible* » si et seulement si il est le résultat d'un processus stochastique causal qui ne remplit pas au moins une des deux conditions pour que l'événement soit « aléatoire (au sens) fort » ($\neg C1 \vee \neg C2$), c'est-à-dire dans les trois conditions suivantes (S1, S2, S3) :

S1 : c'est un processus d'échantillonnage discriminé¹¹ ET invariant dans le temps. Il s'agit donc d'un processus d'échantillonnage discriminé avec remplacement, c'est-à-dire analogue à une série de tirages de boules dans une urne, où certaines boules ont une probabilité plus grande que d'autres d'être tirées à cause de leurs propriétés physiques (par exemple, car elles sont collantes alors que les autres sont glissantes) et où la boule tirée est systématiquement remise dans l'urne.

S2 : C'est un processus d'échantillonnage indiscriminé ET variant dans le temps¹². Il s'agit donc d'un processus d'échantillonnage indiscriminé sans remplacement, c'est-à-dire analogue à une série de tirages de boules dans une urne, où toutes les boules ont la même probabilité d'être tirées et où la boule tirée n'est pas remise dans l'urne.

S3 : C'est un processus d'échantillonnage discriminé ET variant dans le temps. Il s'agit donc d'un processus d'échantillonnage discriminé sans remplacement, c'est-à-dire analogue à une série de tirages de boules dans une urne, où certaines boules (les collantes) ont une probabilité plus grande que d'autres (les glissantes) d'être tirées et où la boule tirée n'est pas remise dans l'urne.

Illustrons ces deux notions de l'aléatoire dans le cas du phénomène de mutation génétique. À cette fin, considérons une séquence entière d'ADN sur un certain laps de temps (par exemple, sur une génération de la cellule porteuse) et n'importe quel type d'altération pouvant l'affecter (par exemple, une mutation ponctuelle, l'insertion ou la délétion de

¹¹ Un processus d'échantillonnage discriminé est un processus où les différences dans les caractéristiques des éléments pouvant être échantillonnés jouent un rôle causal. La probabilité d'être échantillonné n'est donc pas la même pour tout élément.

¹² Dans le cadre d'un processus variant dans le temps, la probabilité qu'un élément soit échantillonné n'est pas la même à tout instant, ce qui implique sa dépendance d'autres événements (des combinaisons d'événements incluses).

plusieurs bases nucléotidiques, l'amplification d'une portion de la séquence nucléotidique, etc.). On voit alors que :

1- Une mutation est « *aléatoire (au sens) fort* » si et seulement si elle est le résultat d'un processus stochastique physico-chimique (causal) qui remplit les deux conditions suivantes ($C1 \wedge C2$) :

C1 : C'est un processus d'échantillonnage indiscriminé par rapport aux différences dans les caractéristiques physico-chimiques des sites de la séquence d'ADN, c'est-à-dire un processus où ces différences ne jouent aucun rôle causal dans le processus de mutation (dans le fait que certains sites mutent et d'autres ne mutent pas) et donc dans la détermination de la distribution des mutations le long de la séquence d'ADN. La probabilité qu'une mutation ait lieu est la même au niveau de tout site nucléotidique de la séquence.¹³

C2 : C'est un processus invariant dans le temps (par exemple, au cours d'une génération cellulaire). En d'autres termes, c'est un processus d'échantillonnage indiscriminé par rapport aux différences dans les conditions physico-chimiques, intra- et extra- cellulaires, au cours du laps de temps considéré (une génération cellulaire), c'est-à-dire où ces différences ne jouent aucun rôle causal dans la détermination de la distribution des mutations dans le temps. La probabilité qu'une mutation ait lieu en un site particulier (ou dans la séquence d'ADN entière) est la même à tout instant du laps de temps considéré, ce qui implique son indépendance par rapport à tout autre événement mutationnel et, de manière générale, par rapport à tout événement environnemental au sens large (intra- et extra- cellulaire).¹⁴

Le processus de mutation est donc analogue à un processus d'échantillonnage indiscriminé avec remplacement (voir plus haut).

2- Une mutation est « *aléatoire (au sens) faible* » si et seulement si elle est le résultat d'un processus stochastique physico-chimique (causal) qui ne remplit pas au moins une des deux conditions pour que la mutation soit « *aléatoire (au sens) fort* » ($\neg C1 \vee \neg C2$), c'est-à-dire dans les trois situations suivantes (S1, S2, S3) :

S1 : C'est un processus d'échantillonnage discriminé par rapport aux différences dans les caractéristiques physico-chimiques des sites nucléotidiques de la séquence d'ADN¹⁵ ET

¹³ C1 est la condition de site-indépendance du processus de mutation ou de symétrie spatiale des probabilités des événements mutationnels le long de la séquence d'ADN.

¹⁴ C2 est la condition de temps-indépendance du processus de mutation ou de symétrie temporelle des probabilités des événements mutationnels.

¹⁵ C'est un processus où ces différences jouent un rôle causal dans le processus de mutation et donc dans la détermination de la distribution des mutations le long de la séquence d'ADN. La probabilité qu'une mutation ait lieu n'est donc pas la même au niveau de tout site nucléotidique de la séquence.

invariant dans le laps de temps considéré. Le processus de mutation est donc analogue à un processus d'échantillonnage discriminé avec remplacement (voir plus haut).

S2 : C'est un processus d'échantillonnage indiscriminé par rapport aux différences dans les caractéristiques physico-chimiques des sites nucléotidiques de la séquence d'ADN ET variant dans le laps de temps considéré¹⁶. Les processus de mutation est donc analogue à un processus d'échantillonnage indiscriminé sans remplacement (voir plus haut).

S3 : C'est un processus d'échantillonnage discriminé à la fois par rapport aux différences dans les caractéristiques physico-chimiques des sites nucléotidiques de la séquence d'ADN ET dans les conditions physico-chimiques intra- et extra- cellulaires au cours du laps de temps considéré. Le processus de mutation est donc analogue à un processus d'échantillonnage discriminé sans remplacement (voir plus haut).

On peut aussi décliner les notions d'« aléatoire (au sens) fort » et d'« aléatoire (au sens) faible » en se plaçant à des niveaux différents, par exemple au niveau d'un site nucléotidique particulier ou d'une séquence nucléotidique d'une certaine longueur, etc., et en considérant des événements mutationnels mutuellement exclusifs à ce niveau, par exemple, pouvant affecter le site ou la séquence nucléotidiques considérés.

Considérons, par exemple, le site nucléotidique S_a de la séquence d'ADN d'une cellule donnée et posons l'hypothèse qu'au moment t_0 il est occupé par une adénine (A). Les altérations pouvant affecter ce site sont les suivantes : la substitution de l'adénine par une cytosine ($M_1 = A \rightarrow C$), par une thymine ($M_2 = A \rightarrow T$), par une guanine ($M_3 = A \rightarrow G$), la délétion d'une adénine (M_4), l'addition d'une adénine (M_5), d'une cytosine (M_6), d'une thymine (M_7) ou d'une guanine (M_8).¹⁷ Considérons l'état du site S_a à tout instant du laps de temps de t_0 à t_5 correspondant, par exemple, à une génération cellulaire et imaginons qu'au moment t_3 la mutation M_1 ait lieu. Nous soutenons qu'il s'agit d'une mutation « aléatoire (au sens) fort » si et seulement si le processus physico-chimique qui l'a produite est un processus stochastique d'échantillonnage indiscriminé qui ne favorise aucune altération possible sur les autres (les altérations possibles (M_1, M_2, \dots, M_8) et mutuellement exclusives ont donc la même probabilité d'avoir lieu) ET qui ne favorise aucun instant de la génération cellulaire

¹⁶ C'est un processus où les différences dans les conditions intra- et extra- cellulaires au cours du laps de temps considéré (une génération cellulaire) jouent un rôle causal dans la détermination de la distribution des mutations dans le temps. La probabilité qu'une mutation ait lieu dans un site particulier (ou dans la séquence d'ADN entière) n'est donc pas la même à tout instant du laps de temps considéré, ce qui implique sa dépendance d'autres événements mutationnels et environnementaux au sens large (intra- et extra- cellulaire).

¹⁷ Dans la situation que nous imaginons, le site nucléotidique S_a mute. Cependant, nous aurions pu considérer le fait que, en moyenne (par exemple, par nucléotide par génération), la probabilité qu'un site nucléotidique donné soit affecté par une mutation au cours d'un certain laps de temps est beaucoup plus petite que la probabilité qu'il ne le soit pas.

considérée (la probabilité de chaque mutation possible est donc la même à tout instant de cette génération)¹⁸. Au contraire, si le processus à l'origine causale de la mutation M_1 favorisait une altération possible (ou plus) sur les autres et/ou un instant de la génération cellulaire considérée, alors nous soutiendrions que la mutation M_1 est « aléatoire (au sens) faible ». En résumé, une mutation est « aléatoire (au sens) faible » si elle n'est pas « aléatoire (au sens) fort ».

7.1.1. Aucune mutation génétique n'est « aléatoire (au sens) fort »

Analysons d'abord la notion d'« aléatoire (au sens) fort ». Elle ne semble pas correspondre à la notion appropriée du point de vue empirique pour caractériser les mutations au niveau moléculaire, c'est-à-dire toute altération de la séquence nucléotidique de l'ADN. Il s'agit pourtant d'une notion que les généticiens moléculaires invoquent, notamment lorsqu'ils évaluent le taux de mutation par le test de fluctuation. Rappelons que cette méthode statistique fut utilisée pour la première fois par Luria et Delbrück dans les années 1940, fut ensuite élargie et généralisée par Lea et Coulson en 1949¹⁹ et représente encore aujourd'hui le modèle théorique et expérimental le plus utilisé pour estimer le taux de mutation²⁰.

Considérons donc les hypothèses constitutives de ce qu'on appelle le modèle Lea-Coulson afin de mettre en évidence où la notion d'« aléatoire (au sens) fort » intervient dans ce cadre. En particulier, concentrons notre attention sur ce que ce modèle suppose au sujet du processus de mutation.²¹ D'abord, la probabilité qu'une mutation²² se produise est supposée

¹⁸ Il est important de souligner que, une fois muté à t_3 , le site S_a n'est plus occupé par une adénine mais par une cytosine. Par conséquent, l'ensemble des altérations pouvant l'affecter au cours du temps restant (de t_3 à t_5) a aussi changé ; il correspond aux mutations ponctuelles suivantes : $M_1 = C \rightarrow A$, $M_2 = C \rightarrow T$, $M_3 = C \rightarrow G$, $M_4 =$ délétion de C, $M_5 =$ addition d'une C, $M_6 =$ addition d'une A, $M_7 =$ addition d'une T, $M_8 =$ addition d'une G. Cela montre que, dans cette situation, le processus de mutation est un processus stochastique Markovien (ou doté de la propriété de Markov), c'est-à-dire dont la distribution conditionnelle de probabilité sur ses états possibles dans le futur (à $t_{3+1} = t_4$) ne dépend que de son état actuel (à t_3) et est indépendante de ses états passés (à t_1 , à t_2 et à t_0). Il est important de préciser que l'ensemble des résultats possibles du processus de mutation peut changer dans le temps, non seulement au niveau d'un site nucléotidique particulier ou d'une séquence d'une certaine longueur, mais aussi le long de la séquence entière d'ADN. Par exemple, la délétion d'une base nucléotidique ou davantage, à un instant donné, provoque le changement de l'ensemble des mutations possibles, à l'instant suivant, le long de la séquence.

¹⁹ Lea & Coulson, 1949, « The distribution of the numbers of mutants in bacterial populations ».

²⁰ Voir le chapitre 3.

²¹ Voir Foster, 2006, « Methods for Determining Spontaneous Mutation Rates » ; Rosche & Foster, 2000, « Determining Mutation Rates in Bacterial Populations ». Notre présentation des hypothèses du modèle Lea-Coulson pour calculer le taux de mutation fait référence à la notion moléculaire de mutation (c'est-à-dire, à l'idée qu'une mutation est une altération de la séquence nucléotidique de l'ADN).

²² Le plus souvent, on calcule le taux d'occurrence des mutations spontanées et, encore plus spécifiquement, des mutations ponctuelles.

être indépendante des événements mutationnels précédents. En deuxième lieu, cette probabilité est supposée être constante au cours du temps de vie d'une cellule. En troisième lieu, la proportion de mutants dans une population de cellules clonales est supposée être très petite. En quatrième lieu, les mutations inverses, rétablissant le type sauvage à partir d'un mutant, sont considérées comme négligeables. En cinquième lieu, aucun mutant (donc aucune mutation) n'est supposé(e) se produire après que la population a été placée dans un milieu sélectif. En sixième lieu, la probabilité qu'une mutation ait lieu par cycle cellulaire est considérée comme constante durant la croissance de la culture de cellules.²³ Une hypothèse ultérieure est introduite afin de calculer le taux de mutation par gène (ou par nucléotide) par génération à partir du taux de mutation par génome par génération. Une mutation est supposée avoir la même probabilité d'avoir lieu au cours d'une génération dans tout gène (ou nucléotide) du génome considéré. Sur cette base, on fait correspondre le taux de mutation par gène (ou par nucléotide) par génération à une valeur moyenne, qui est calculée en divisant par le nombre total de gènes (ou de nucléotides) le taux par génome par génération (et donc en négligeant le fait que le taux de mutation par génération de chaque gène (ou nucléotide) peut être différent).

Il est intéressant de remarquer que, dans les modèles de l'évolution en génétique des populations, les mutations sont représentées sous la forme d'un paramètre, la pression de mutation, dont la valeur est estimée en posant les hypothèses du modèle Lea-Coulson. De manière plus spécifique, les généticiens des populations décrivent l'évolution des populations naturelles par le changement dans le temps de la fréquence relative des allèles, c'est-à-dire des formes alternatives d'un même gène situé dans un emplacement ou locus particulier d'un chromosome. Ils supposent que la probabilité d'une mutation transformant un allèle donné dans une forme allélique différente à un même locus²⁴ est indépendante du fait que d'autres mutations ont eu lieu auparavant dans d'autres loci du même chromosome. Ils font aussi l'hypothèse que la probabilité qu'une telle transformation allélique ait lieu est très petite et

²³ Les autres hypothèses du modèle Lea-Coulson sont les suivantes : les cellules croissent de manière exponentielle ; les taux de croissance des mutants et des non mutants sont les mêmes ; la mort cellulaire est négligeable ; tous les mutants sont détectés ; le nombre initial de cellules est négligeable par rapport au nombre final de cellules.

²⁴ Depuis la naissance de la génétique des populations, on suppose qu'un même locus d'un gène présente deux états alléliques et que les mutations entre les deux formes sont réversibles. Cependant, un tel modèle ne peut pas être appliqué pour estimer la variabilité allélique au niveau moléculaire car, de ce point de vue, n'importe quel locus d'un gène a de très nombreux allèles possibles. C'est pourquoi Kimura et Crow (1964, « The number of alleles that can be maintained in a finite population ») ont introduit le modèle des allèles infinis selon lequel toute altération possible de la séquence nucléotidique à un locus est à l'origine d'un allèle différent. Autrement dit, « le nombre d'isoallèles possibles à un locus est si grand que chaque nouveau mutant représente un état allélique qui n'existait pas auparavant dans la population », Kimura, 1990[1983], *La théorie neutraliste de l'évolution*, p. 259.

constante au cours du temps. De plus, ils supposent souvent que la probabilité est la même pour toute mutation transformant un allèle donné (par exemple, l'allèle considéré comme caractéristique du type sauvage ou normal) dans toute direction possible, c'est-à-dire dans toute forme possible du gène considéré à un même locus.

Il nous paraît évident que la notion d'« aléatoire (au sens) fort » rend compte d'un certain nombre d'hypothèses que les généticiens moléculaires posent pour évaluer le taux de mutation et sur la base desquelles les généticiens des populations modélisent les changements des fréquences géniques (notamment, l'équiprobabilité des mutations le long de la séquence d'ADN, l'invariance de la probabilité d'une mutation au cours du temps et donc son indépendance par rapport à d'autres événements mutationnels). Cependant, toutes ces hypothèses ne sont que des simplifications, plus précisément des idéalizations par rapport à ce qu'on savait déjà au cours de la première moitié du XX^e siècle et, de manière beaucoup plus importante, par rapport aux connaissances que les généticiens moléculaires possèdent aujourd'hui sur le processus de mutation.²⁵

Bien que les généticiens moléculaires et les généticiens des populations reconnaissent qu'au niveau moléculaire toute mutation est inévitablement biaisée par des facteurs physico-chimiques divers, ils font abstraction de ces faits et considèrent toutes les hypothèses présentées ci-dessus comme des simplifications admissibles du processus réel de mutation, en particulier lorsqu'on le décrit au niveau populationnel. En effet, leur but n'est pas de fournir une représentation qui soit la plus réaliste possible de ce que sont les mutations au niveau de la séquence nucléotidique de l'ADN. En posant cet ensemble d'hypothèses qui leur permet, entre autres, d'estimer des valeurs moyennes, ils visent plutôt à simplifier les calculs requis par la description et par la prédiction formelles des mutations : leur but principal est de disposer d'un modèle qui soit général et précis à la fois, c'est-à-dire qu'ils puissent appliquer pour décrire et pour prédire de manière quantitative le taux de mutation chez divers organismes.²⁶

²⁵ Nous faisons référence à tous les biais de mutation pouvant influencer, à différents degrés et de manières diverses, le site, le type, le moment où les mutations se produisent et leur taux d'occurrence. Voir les chapitres 4, 5 et 6.

²⁶ Levins (1966, « The Strategy of Model Building in Population Biology ») affirme que, lorsqu'on étudie un phénomène par sa modélisation, il est souhaitable qu'on dispose d'un modèle maniable maximisant la généralité, le réalisme et la précision. Cependant, ces trois desiderata ne correspondent pas à un même but et, par conséquent, ne peuvent être maximisés en même temps. Levins distingue donc trois stratégies de modélisation dont chacune sacrifie l'un des trois desiderata : on peut construire un modèle réaliste et précis, un modèle général et précis ou un modèle réaliste et général d'un même phénomène en fonction du but de notre étude. Pour une analyse critique de ces trois stratégies de modélisation suggérées par Levins, nous renvoyons à Weisberg, 2006b, « Forty Years of 'The Strategy'. Levins on Model Building and Idealization ».

La considération de certaines études récentes en génétique moléculaire met en évidence que les connaissances actuelles sur le processus de mutation s'éloignent des hypothèses simplificatrices du modèle Lea-Coulson. La notion d'« aléatoire (au sens) fort », à laquelle ce modèle renvoie, se révèle donc une notion idéalisant le processus physico-chimique de mutation : elle est inappropriée du point de vue empirique pour caractériser les mutations au niveau moléculaire. Si nous ne prétendons pas exclure de manière catégorique qu'une mutation « aléatoire (au sens) fort » puisse avoir lieu, nous soulignons à nouveau qu'il s'agirait d'une altération de la séquence nucléotidique extrêmement peu probable du point de vue physico-chimique à cause de la variété et de l'abondance de facteurs pouvant biaiser le processus de mutation.

7.1.2. De nouveaux biais de mutation

Analysons donc la notion d'« aléatoire (au sens) faible » afin de montrer qu'elle correspond à la notion appropriée du point de vue empirique pour caractériser au niveau moléculaire toute mutation génétique. À cette fin, l'examen de l'état actuel de la recherche au sujet du processus de mutation, en particulier en génétique moléculaire, est nécessaire.

Avant de présenter les études expérimentales réalisées au cours des dix dernières années et d'en examiner la portée, rappelons ce que les biologistes de la Synthèse Moderne reconnaissaient quant aux mutations génétiques, d'abord, durant la première moitié du XX^e siècle sur la base de la recherche en génétique classique, et ensuite, à partir des années 1950, au moment du tournant moléculaire en biologie.²⁷ Notre objectif est de mettre en évidence que, malgré le fait que des avancées importantes aient marqué la recherche sur les mutations au cours du siècle dernier, la manière dont il est approprié de concevoir la notion de hasard pour caractériser ces phénomènes n'a pas changé et correspond à ce que nous appelons la notion d'« aléatoire (au sens) faible ».

²⁷ Quant aux biais de mutation reconnus par les biologistes de la Synthèse Moderne et dont nous en rappellerons quelques-uns ci-dessous, voir Fisher, 1999[1930], *The Genetical Theory of Natural Selection* ; Wright, 1931, « Evolution in Mendelian Populations » ; Haldane, 1966[1932], *The Causes of Evolution* ; Dobzhansky, 1951[1937], *Genetics and The Origin of Species* ; 1970, *Genetics of the Evolutionary Process* ; Huxley, 1948[1942], *Evolution. The Modern Synthesis* ; 1953, *Evolution in Action* ; Mayr, 1942, *Systematics and the Origin of Species* ; 1974[1970], *Populations, espèces et évolution* ; 1976, *Evolution and the Diversity of Life. Selected Essays* ; Simpson, 1984[1944], *Tempo and Mode in Evolution* ; 1953, *The Major Features of Evolution* ; Stebbins, 1971, *Processes of Organic Evolution* ; 1982, *Darwin to DNA, Molecules to Humanity* ; Dobzhansky *et al*, 1977, *Evolution*. Nous renvoyons à la section 4.2.1 du chapitre 4.

Avant la découverte de la structure moléculaire du matériel génétique, les biologistes de la Synthèse Moderne décrivaient les sources causales de la variation génétique, que sont les processus de mutation des gènes et de recombinaison chromosomique, en utilisant le vocabulaire de la génétique classique, c'est-à-dire en termes de modifications des gènes et des chromosomes dont la structure moléculaire était inconnue à cette époque. Ils admettaient qu'une mutation n'a pas la même probabilité de se produire dans tout gène ; ensuite, dans les années 1950 et 1960, la découverte des « points chauds de mutation » confirma que certains gènes sont beaucoup plus mutables que d'autres²⁸. De plus, toujours avant le tournant moléculaire en biologie, les biologistes de la Synthèse Moderne reconnaissaient que les différents types de mutations n'ont pas la même probabilité d'avoir lieu, que des agents physico-chimiques (par exemple, les rayons ultraviolets, les gaz moutardes, etc.) peuvent induire une augmentation du taux de mutation, que ce dernier est sous contrôle génétique et peut subir une augmentation à cause de la présence de gènes spécifiques dits « modificateurs ».

Suite à la découverte de la structure physico-chimique du matériel génétique (1953) et à la lumière des premiers résultats de la recherche en génétique moléculaire, les biologistes de la Synthèse Moderne ont reformulé en termes moléculaires tous les biais physico-chimiques qu'ils admettaient déjà quant aux mutations génétiques. Ils ont donc reconnu que certains sites du génome sont plus mutables que d'autres à cause des caractéristiques physico-chimiques de la séquence nucléotidique (par exemple, dans le cas des « points chauds de mutation », le taux élevé de mutation est dû à la présence de répétitions de nucléotides induisant les enzymes de réplication en erreur) et à cause de la structure à double hélice de l'ADN. Ils ont également admis que certains agents physico-chimiques tendent à produire des mutations dans des séquences nucléotidiques spécifiques et que la présence d'altérations dans certains gènes impliqués dans la production d'enzymes réparatrices est à l'origine d'une augmentation globale du taux de mutation. Dans les années 1960, ils ont aussi reconnu que les mutations peuvent avoir lieu non seulement en phase de croissance des populations, mais aussi en phase stationnaire et que des taux inégaux de mutation peuvent caractériser les différentes étapes du processus de développement. Enfin, à partir des années 1990, de nombreux biologistes ont soutenu que la présence de mécanismes de mutation, spécifiques et adaptatifs, au sein de populations situées dans des conditions environnementales adverses ne

²⁸ Voir Benzer & Freese, 1958, « Induction of specific mutations with 5-bromouracil » ; Benzer 1959, « On the topology of genetic fine structure » ; 1961, « On the topology of genetic fine structure » ; 1962, « The fine structure of the gene ».

pose aucun défi au paradigme de la Synthèse Moderne et, en particulier, au caractère aléatoire de toute mutation²⁹.

D'autres biais concernant l'origine, le taux et la distribution des mutations n'ont été observés que très récemment (les premiers articles à ce sujet datent du début des années 1990³⁰) ; depuis dix ans, ils font l'objet de nombreuses études afin d'être confirmés expérimentalement.³¹ Il s'agit de la découverte de corrélations statistiques, probablement dues à des liens de type causal, entre les mutations génétiques affectant une même séquence d'ADN : elles concernent le moment dans le temps et le site de la séquence nucléotidique où les altérations se produisent. La confirmation expérimentale de ces biais représenterait sans doute une nouveauté considérable par rapport à la manière dont les mutations génétiques étaient conçues par les premiers généticiens, par les pères fondateurs de la génétique (mathématique) des populations et par les biologistes de la Synthèse Moderne. Elle apporterait aussi des éléments nouveaux en génétique moléculaire, en particulier relativement aux hypothèses sous-jacentes aux méthodes d'analyse statistique pour calculer le taux de mutation (notamment le modèle Lea-Coulson).

En 1991, Ninio observa des populations microbiennes caractérisées par une diminution transitoire de la fidélité de réplication de l'ADN. Au sein de ces populations, il remarqua une augmentation transitoire du nombre de mutants porteurs de deux mutations ou davantage (ce qu'il appela des mutations « doubles » ou « multiples »)³². Il proposa un scénario pouvant expliquer ses observations : c'est l'hypothèse des « mutateurs transitoires ou phénotypiques » selon laquelle les mutations doubles ou multiples sont dues à l'hypermutableté transitoire d'une sous-population microbienne. Bien que plausible et avancée encore aujourd'hui par d'autres biologistes, dans le cas de Ninio cette explication était plutôt spéculative car elle n'était basée sur aucun corps de preuves expérimentales.

Quelques années plus tard, Ninio³³ remarqua que des mutations multiples (qu'il désigna avec l'expression « événements mutationnels complexes ») peuvent aussi se produire chez les organismes supérieurs (de la drosophile à l'homme). Dans ce cas, il avança deux modèles explicatifs. Selon le premier, les mutations multiples se produisent de manière indépendante et simultanée, comme dans le cas des « mutateurs transitoires ou

²⁹ Voir le chapitre 5.

³⁰ Ninio, 1991, « Transient Mutators: A Semiquantitative Analysis of the Influence of Translation and Transcription Errors on Mutation Rates ».

³¹ Voir Drake, 2007a, « Mutations in clusters and showers » ; 2007b, « Too Many Mutants with Multiple Mutations ».

³² Ninio, 1991, *op. cit.*, p. 961.

³³ Ninio, 1996, « Gene conversion as focusing mechanisms for correlated mutations: a hypothesis ».

phénotypiques » chez les bactéries³⁴ ; cependant, ce scénario implique qu'un nombre trop élevé de mutations délétères se produise, ce qui ne permettrait pas aux organismes porteurs de survivre. Selon le deuxième modèle explicatif, préféré par Ninio, les mutations multiples sont des événements successifs et corrélées à la fois en ce qui concerne le moment dans le temps et l'endroit de la séquence d'ADN où ils ont lieu. Dans le cadre de ce deuxième modèle, Ninio proposa deux hypothèses quant au mécanisme à l'origine de mutations multiples. Selon la première, une marque épigénétique (par exemple, un groupe méthyl) se formerait au niveau du site où une mutation a eu lieu et, successivement, des mutations se produiraient autour de ce site à cause de la marque épigénétique. Selon la deuxième hypothèse, que Ninio semble préférer aujourd'hui, les mutations multiples auraient lieu l'une après l'autre au niveau de « points chauds de recombinaison » à cause d'hétérogénéités locales entre chromosomes homologues. En d'autres termes, les mutations multiples se produiraient durant le processus de « conversion génique » consistant en l'altération d'un brin d'ADN de manière à le rendre complémentaire en toute position au brin auquel il est apparié. Bien que les modèles et les hypothèses explicatives proposés par Ninio soient intéressants et anticipent les résultats de recherches expérimentales plus récentes sur le même sujet, ils n'ont fait l'objet d'aucune expérience et d'aucune preuve expérimentale. Ils ne peuvent donc être considérés que comme de la spéculation théorique.

Autour des années 2000, Drake analysa un grand nombre de mutants générés *in vitro* par l'ADN polymérase du bactériophage RB69. Il observa un nombre de mutants avec des mutations doubles ou multiples beaucoup plus grand que le nombre attendu selon la vision classique de la distribution des mutations (à la fois au niveau populationnel et au niveau de la séquence génomique), c'est-à-dire si les mutations avaient été distribuées selon une loi de Poisson³⁵ parmi les molécules analysées et dans la séquence de chaque molécule d'ADN. Drake attendit quelques années avant de soumettre un article à ce sujet intitulé « *Too many mutants with multiple mutations* » (« Trop de mutants avec des mutations multiples », notre traduction). Il se vit refuser cet article à cause du titre, qu'il changea en « *Clusters of mutations from transient hypermutability* » (« Agglomérats de mutations résultant de l'hypermutabilité transitoire », notre traduction), ce qui lui permit de publier l'article en 2005³⁶. De la même manière que Ninio, Drake proposa d'expliquer le grand nombre de

³⁴ Ninio, 1991, *op. cit.*

³⁵ Voir le Glossaire.

³⁶ Drake *et al.*, 2005, « Clusters of mutations from transient hypermutability ».

mutations doubles ou multiples par l'hypermutableté transitoire d'une partie de la population de cellules.

Durant les dix dernières années, des mutants porteurs de mutations multiples ont été observés chez des ribovirus, des virus à ADN, des organismes procaryotes, des organismes eucaryotes comme la levure et chez les cellules somatiques de plusieurs eucaryotes supérieurs³⁷. La souris transgénique appelée « Big Blue[®] » en est un cas paradigmatique³⁸. Chez de nombreux mutants de cet organisme, deux mutations ou plus se produisent dans des sites très rapprochés de leur génome : c'est pourquoi les généticiens moléculaires ont commencé à utiliser l'expression « agglomérats ou isolats mutationnels ». Plus précisément, ils ont observé des mutants de « Big Blue[®] » porteurs de mutations multiples (des doublets, triplets, etc.) qui sont distribuées de manière relativement dense (de l'ordre d'une mutation chaque 3 kilobases), non pas dans tout le génome, mais le long d'une région génomique spécifique et circonscrite (la région *lacI* et ses alentours ne dépassant pas les 30 kilobases).³⁹ Ils ont donc commencé à parler de « douches mutationnelles » pour désigner ce genre de *patterns* de mutation qui, en contraste avec l'hypothèse classique de la distribution (suivant une loi de Poisson) des mutations génétiques au sein des populations et dans la séquence génomique des organismes individuels, suivent plutôt une loi exponentielle⁴⁰.

Les mécanismes moléculaires à l'origine des mutations multiples ne sont pas encore connus. En ce qui concerne la distribution de ces mutations le long de la séquence d'ADN, les généticiens moléculaires retiennent comme moins plausible l'hypothèse selon laquelle il s'agit du résultat de plusieurs événements mutationnels survenus lors de réplifications successives de la même molécule. De plus, ils excluent que ces mutations puissent être survenues de manière indépendante au cours de la même répllication de la séquence génomique (ce qui impliquerait un taux de mutation trop élevé et donc délétère pour les organismes concernés). Ils favorisent plutôt l'idée selon laquelle il s'agirait de mutations ayant lieu de manière synchronique (c'est-à-dire, au cours de la même répllication) et qui seraient liées causalement. En ce qui concerne la distribution de ces mutations au niveau des populations, les généticiens moléculaires retiennent que le nombre élevé de mutants observés et porteurs de plusieurs mutations rapprochées n'est probablement pas dû à l'activité d'un

³⁷ Pour la référence précise de chaque étude, voir Drake, 2007b, « Too Many Mutants with Multiple Mutations ».

³⁸ Buettner *et al*, 2000, « Evidence that proximal multiple mutations in Big Blue[®] transgenic mice are dependent events » ; Wang *et al*, 2007, « Evidence for mutation showers ».

³⁹ Des mutations multiples de ce genre ont aussi été observées chez la bactérie *Escherichia coli* présente dans l'intestin de la souris (voir Hill *et al*, 2004, « Spontaneous multiple mutations show both proximal spacing consistent with chronocoordinate events and alterations with p53-deficiency »).

⁴⁰ Voir le Glossaire.

mécanisme mutateur constitutif, mais plutôt à un état transitoire d'hypermutation ne concernant qu'une sous-population (ce qui évoque l'hypothèse des « mutateurs transitoires ou phénotypiques » de Ninio, 1991).

Bien que les scénarios avancés jusqu'à aujourd'hui afin d'expliquer les mutations multiples n'aient encore fait l'objet d'aucune confirmation expérimentale, nous pouvons déjà indiquer quelle serait la portée d'une confirmation éventuelle pour la manière dont les généticiens moléculaires décrivent et prédisent actuellement les événements mutationnels (notamment, pour les hypothèses simplificatrices du modèle Lea-Coulson). Les généticiens moléculaires devraient reconnaître, premièrement, que la probabilité qu'une mutation ait lieu peut varier, non seulement selon les caractéristiques de chaque site particulier et du contexte nucléotidique où il est placé, mais aussi à cause de la présence d'autres altérations de la séquence d'ADN dans ses environs. Deuxièmement, ils devraient admettre que la probabilité qu'une mutation ait lieu n'est pas constante au cours du temps mais peut varier, non seulement en fonction de la présence transitoire d'agents mutagènes, de l'étape du processus de développement, mais aussi en fonction de l'activation temporaire de mécanismes mutateurs divers et des changements que la séquence d'ADN a subis précédemment. La confirmation expérimentale de tous ces biais de mutation allongerait la liste des facteurs physico-chimiques pouvant influencer l'occurrence des mutations génétiques et fournirait une raison de plus pour considérer les mutations, non comme des événements « aléatoires (au sens) fort », mais « aléatoires (au sens) faible ».

7.1.3. Toute mutation génétique est un événement « aléatoire (au sens) faible »

À la lumière de l'état actuel de la recherche en génétique moléculaire⁴¹, nous concluons que la notion de hasard la plus appropriée du point de vue empirique pour caractériser au niveau moléculaire toute mutation génétique (ou, du moins, la plupart des événements mutationnels physiquement possibles) est celle d'« aléatoire (au sens) faible ».

Trois objections pourraient être adressées à cette manière de concevoir le caractère aléatoire des mutations. La première est une objection d'ordre méthodologique : comment évalue-t-on la probabilité qu'une mutation ait lieu le long d'une séquence d'ADN et au cours du temps ? En effet, selon le niveau auquel on se place (à l'échelle des nucléotides, de

⁴¹ Nous faisons référence à tous les biais de mutation que nous avons examinés dans cette deuxième partie. En particulier, nous renvoyons au chapitre 6.

portions d'une longueur donnée de la séquence d'ADN, de ce qu'on appelle les gènes⁴², des chromosomes ou du génome entier) et/ou relativement à la période de temps considérée (une génération, plusieurs générations, un laps de temps arbitraire, etc.), une même mutation pourrait être « aléatoire (au sens) fort » et « aléatoire (au sens) faible » à la fois, ce qui serait contradictoire. Par exemple, la probabilité qu'une mutation ponctuelle ait lieu peut être la même pour tout nucléotide d'une séquence d'ADN donnée, mais correspondre à des valeurs différentes lorsqu'elle est évaluée au niveau des gènes⁴³. De plus, elle pourrait être constante ou changer au cours du temps selon l'intervalle temporel considéré.

Nous répondons à cette première objection en rappelant que, dans le cadre de ce chapitre, nous examinons le caractère aléatoire des mutations génétiques au niveau moléculaire, c'est-à-dire en tant qu'altérations de la séquence nucléotidique de l'ADN et relativement aux processus physico-chimiques qui les causent. Autrement dit, nous parlons de la probabilité qu'une mutation ait lieu en faisant référence à l'échelle biologiquement pertinente la plus petite et la mieux définie qu'est le niveau des nucléotides et à partir duquel nous suggérons d'évaluer les probabilités des mutations. Il est indéniable que, d'un point de vue pratique, il est impossible d'évaluer de manière précise et objective la probabilité spécifique de toute altération possible au niveau nucléotidique. Cependant, du point de vue théorique, en particulier dans le cadre des définitions de l'aléatoire, nous proposons de parler de la probabilité qu'une altération de la séquence nucléotidique de l'ADN ait lieu en référence au niveau moléculaire de la réalité biologique, c'est-à-dire en prenant en considération, en principe, toute propriété et tout facteur physico-chimiques pouvant jouer un rôle causal dans le processus physico-chimique de mutation.

En ce qui concerne la probabilité d'une mutation au cours du temps, nous suggérons de l'évaluer par génération cellulaire simplement parce qu'il s'agit d'un laps temporel assez petit et, du point de vue biologique, assez bien défini : une génération correspond au

⁴² Nous rappelons que parler des « gènes » pose problème car il n'existe pas encore, et peut-être n'existera jamais, de définition précise et univoque de ce qu'est un gène (voir Fox Keller, 2003[2000], *Le siècle du gène*). Nous ne souhaitons pas entrer dans ce débat. Nous utilisons ce terme en accord avec l'idée commune selon laquelle un gène est un segment d'ADN situé à un endroit bien précis (locus) sur le chromosome et qui produit une certaine molécule d'ARN fonctionnelle. Cependant, nous devons signaler que cette conception a été désavouée, entre autres, par la découverte du processus d'épissage génétique consistant en l'élimination de portions de la séquence nucléotidique (notamment les introns) lors de la synthèse des molécules d'ARN messager chez les organismes eucaryotes. En effet, vus les épissages divers dont une même séquence d'ADN peut être l'objet, un segment d'ADN peut correspondre à de multiples molécules d'ARN fonctionnelles. Voir Griffiths & Stotz, 2007, « Gene ».

⁴³ Cela serait dû aux différences dans la taille des gènes en termes de nombre de nucléotides. Inversement, la probabilité qu'une mutation ponctuelle ait lieu pourrait être la même pour tout gène d'une séquence d'ADN donnée, mais correspondre à des valeurs différentes lorsqu'elle est évaluée au niveau des nucléotides. Cette deuxième situation serait due au fait que, en dépit des différences dans la probabilité qu'une mutation se produise au niveau de chaque nucléotide, la somme de ces probabilités pour chaque gène serait la même.

processus temporel qui va de la naissance à la reproduction (duplication) d'une cellule d'un organisme vivant donné.

La deuxième objection qu'on pourrait nous adresser porte sur l'extension de la notion d'« aléatoire (au sens) faible » : il s'agit en effet d'une notion pouvant s'appliquer à un ensemble très grand de phénomènes physiques et, en particulier, biologiques, ce qui pourrait être considéré comme un problème pour sa valeur conceptuelle et son intérêt pour les caractériser.

En réponse à cette objection, nous estimons qu'une notion peut s'appliquer à un ensemble très grand d'objets tout en gardant sa pertinence pour les caractériser ; de plus, s'il s'agit d'une notion explicative, son rôle n'est pas non plus mis en question par sa large applicabilité. Par exemple, si la notion d'« aléatoire (au sens) faible » était appropriée pour caractériser tout phénomène biologique, cela ne signifierait pas que notre argument a été conduit incorrectement ou que nous n'avons pas bien définie cette notion, mais simplement que tout phénomène biologique est « aléatoire (au sens) faible » (ce qui nous semble très plausible).

En ce qui concerne plus particulièrement les mutations génétiques, la notion d'« aléatoire (au sens) faible » s'est montrée appropriée pour caractériser toute mutation (ou, du moins, la plupart des mutations physiquement possibles) du point de vue moléculaire car elle rend compte de tout facteur physico-chimique pouvant influencer le processus de mutation (c'est-à-dire, tout biais de mutation). En revanche, cela n'empêche pas qu'on puisse limiter l'extension de la notion d'« aléatoire (au sens) faible » (autrement dit, préciser cette notion) dans chaque étude particulière. À la lumière des données expérimentales dont on dispose – par exemple, quant à la probabilité des mutations dans le temps et le long de la séquence d'ADN et en ce qui concerne leurs relations d'indépendance ou de dépendance –, on pourrait déterminer si ces mutations remplissent ou ne remplissent pas les conditions C1 et C2. Sur cette base, on pourrait donc formuler une définition plus restreinte et spécifique de la notion d'« aléatoire (au sens) faible » qui serait appropriée du point de vue empirique pour caractériser les mutations étudiées.

La troisième objection rejoint la deuxième : elle concerne la signification de la notion d'« aléatoire (au sens) faible », plus précisément le fait que son application pour caractériser un événement donné – par exemple, une mutation – consiste en quelque sorte à affirmer que cet événement est dû à un processus stochastique ou, en d'autres termes, qu'il s'agit d'un événement probabiliste. On pourrait donc nous objecter que la notion d'« aléatoire (au sens) faible » est en effet une notion trop faible de hasard.

Il est vrai que le caractère stochastique du processus causal engendrant un événement – plus spécifiquement, un événement mutationnel – « aléatoire (au sens) faible » et le caractère probabiliste de ce dernier sont au cœur même de cette notion d'aléatoire. Nous estimons que l'aspect problématique d'une telle définition de ce qu'est l'aléatoire n'est pas représenté par sa réduction au caractère probabiliste d'un événement (ou résultat) et du processus qui l'engendre. Le problème majeur de la notion d'« aléatoire (au sens) faible » (comme d'ailleurs toute notion caractérisée en termes de probabilité) réside plutôt dans la question de la signification et du rôle de la notion de probabilité mobilisée dans le cadre de sa définition. Le problème de savoir comment interpréter les probabilités utilisées pour décrire et pour prédire un phénomène donné (dans ce cas, les mutations génétiques) se présente à nouveau et, nous le verrons, sa résolution se révèle centrale pour comprendre la signification et le rôle de la notion de hasard en tant qu'« aléatoire (au sens) faible ».

7.2. La dérive génétique « aléatoire (au sens) faible »

La notion d'« aléatoire (au sens) faible » est appropriée non seulement pour caractériser les mutations génétiques et, comme nous le verrons, d'autres phénomènes moléculaires à l'origine causale de la variation (par exemple, le bruit dans l'expression des gènes⁴⁴), mais aussi des phénomènes biologiques au niveau de l'évolution des populations naturelles. Notamment, cette notion révèle sa pertinence et reçoit une définition précise dans la caractérisation des changements des fréquences géniques dus à ce qu'on appelle la « dérive génétique aléatoire ». Après avoir fourni une introduction aux caractéristiques principales de ce facteur de l'évolution biologique⁴⁵, nous montrerons que la dérive génétique aléatoire est un processus d'échantillonnage indiscriminé sans remplacement. Les changements évolutifs qu'elle engendre sont donc des événements « aléatoires (au sens) faible ».

⁴⁴ Voir le chapitre 9.

⁴⁵ Voir David & Samadi, 2000, *La théorie de l'évolution. Une logique pour la biologie* ; Gillespie, 2004, *Population Genetics: A Concise Guide*.

7.2.1. Qu'est-ce que la dérive génétique aléatoire ?

D'abord, rappelons que toute population naturelle est caractérisée par de la variation génétique qui s'accompagne, même si ce n'est pas de manière nécessaire, par des effets phénotypiques visibles. En génétique des populations, on parle de « monomorphisme » lorsque tous les membres d'une population sont du même type, c'est-à-dire qu'ils possèdent tous le même allèle à un locus génétique donné⁴⁶. Au contraire, on parle de « polymorphisme » lorsqu'une population est caractérisée par la coexistence de plusieurs types d'organismes, porteurs d'allèles différents à un même locus. Les états polymorphiques des populations sont transitoires et sont définis en termes de fréquences alléliques. L'évolution d'une population est décrite par l'augmentation ou la diminution de ces fréquences au cours du temps : sur le long terme, elle conduit à la fixation d'un allèle à un locus donné, c'est-à-dire à sa présence dans le génome de tous les membres d'une population, ou à sa disparition.

La dérive génétique aléatoire est un processus d'échantillonnage aléatoire ayant lieu au sein d'une population naturelle à chaque passage d'une génération à la suivante : il correspond aux fluctuations des fréquences géniques dues à la limitation de l'effectif de la population.⁴⁷ Depuis les années 1930, ce processus est décrit de manière analogue au tirage à l'aveugle d'une urne remplie de boules de couleurs différentes⁴⁸. Les boules correspondent à des entités biologiques (par exemple, des organismes ou des gamètes) et leurs couleurs à des

⁴⁶ Nous rappelons que les allèles sont les classes (ou types) de gènes situés à un même locus. On suppose que chaque mutation introduit un allèle nouveau qui est donc défini par sa séquence nucléotidique. En ce qui concerne le modèle des allèles infinis (Kimura, 1990[1983], *La théorie neutraliste de l'évolution*), nous renvoyons à la note 24 de ce chapitre.

⁴⁷ Au cours des trente dernières années, de nombreux philosophes de la biologie ont focalisé leur réflexion sur la dérive génétique aléatoire. Entre autres, voir Beatty, 1984, « Chance and Natural Selection » ; Sober, 1984, *The Nature of Selection. Evolutionary Theory in Philosophical Focus* ; Hodge, 1987, « Natural Selection as a Causal, Empirical, and Probabilistic Theory » ; Rosenberg, 1988, « Is the Theory of Natural Selection Really a Statistical Theory? » ; Brandon, 1990, *Adaptation and Environment* ; 2005, « The Difference Between Selection and Drift: A Reply to Millstein » ; Shanahan, 1992, « Selection, Drift, and the Aims of Evolutionary Theory » ; Brandon & Carson, 1996, « The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables' Proof But No Room for Determinism Either » ; Millstein 1996, « Random Drift and the Omniscient Viewpoint » ; 2002, « Are Random Drift and Natural Selection Conceptually Distinct? » ; Millstein & Skypper, 2007, « Population Genetics ». Parmi les nombreux problèmes analysés figure la question de savoir si la dérive génétique aléatoire est un processus stochastique (Beatty, 1984, *op. cit.* ; Millstein, 2002, *op. cit.*) ou une simple erreur statistique au niveau du résultat (Brandon, 1990, *op. cit.* ; 2005, *op. cit.*). Nous exprimons notre penchant pour une conception de la dérive génétique aléatoire en tant que processus stochastique de l'évolution.

⁴⁸ Dubinin & Romaschoff, 1932, « The Genetic Structure of Species and their Evolution » ; Dobzhansky, 1951[1937], *Genetics and The Origin of Species*. Voir Beatty, 1984, *op. cit.*

types différents, notamment à des formes alternatives d'un même gène (allèles) à un locus donné. En dépit de leurs différences, les boules sont équivalentes relativement au processus d'échantillonnage car ce dernier est supposé être réalisé par un agent aveugle ; en d'autres termes, les entités biologiques ont la même probabilité de donner lieu à la génération suivante. La composition en couleur de l'urne correspond à la composition génétique de la population naturelle à chaque génération : elle change au cours du temps par l'échantillonnage à l'aveugle d'un nombre fini de boules (c'est-à-dire d'entités biologiques) qui est réitéré à chaque génération.⁴⁹

Imaginons une situation très simple où, dans une population d'organismes diploïdes⁵⁰, la fréquence des allèles A et a à un même locus est la suivante : 60% de A et 40% de a . Supposons que les trois génotypes possibles (AA , Aa ou aa) soient sélectivement équivalents, c'est-à-dire que les organismes AA , Aa et aa soient également bien adaptés à l'environnement où la population se trouve : il n'y a donc pas de sélection. Si l'effectif de la population est infini, ou très grand, et en absence de tout phénomène de mutation, d'accouplement préférentiel et de migration, on attend qu'à la génération suivante la fréquence relative des allèles A et a ne change pas, c'est-à-dire qu'elle soit égale à 60% de A et 40% de a . La population est à ce qu'on appelle l'« équilibre d'Hardy-Weinberg ».⁵¹ Au contraire, si l'effectif de la population est fini⁵², la fréquence allélique réelle s'écartera de la fréquence

⁴⁹ L'expression « dérive génétique aléatoire » est une expression très large qui désigne des processus biologiques divers : ils ont tous un impact sur le changement des fréquences géniques, mais peuvent se dérouler à des niveaux biologiques différents. Par exemple, on parle plus spécifiquement d'« échantillonnage indiscriminé des parents » pour désigner le processus ayant lieu au niveau de populations d'organismes et déterminant, parmi les organismes d'une génération donnée, quels seront les parents de la génération suivante et combien de descendants ils auront respectivement. On parle aussi d'« échantillonnage indiscriminé des gamètes » pour désigner le processus ayant lieu au niveau de populations de cellules reproductives et déterminant, parmi les nombreux gamètes de deux types génétiquement différents produits par des parents hétérozygotes, lequel contribuera à la formation de chacun de leurs descendants. Ce qu'on appelle « l'effet du fondateur » est aussi considéré comme un cas de dérive génétique aléatoire : il s'agit de l'établissement d'une nouvelle population par un petit nombre d'organismes fondateurs qui ne sont pas représentatifs de la variation génétique propre à la population parentale (Mayr, 1954, « Change of genetic environment and evolution » ; 1963, *Animal Species and Evolution*). Pour une présentation des types différents de dérive génétique aléatoire, voir Beatty, 1992, « Random Drift » ; Brandon, 1990, *op. cit.* ; Millstein, 2002, *op. cit.*, Millstein & Skipper, 2007, *op. cit.* ; Lamotte, 1988, « Phénomènes fortuits et évolution ».

⁵⁰ Un organisme diploïde possède deux exemplaires de chaque chromosome.

⁵¹ Le principe fondamental de la génétique des populations, qu'est le « principe Hardy-Weinberg », spécifie les conditions sous lesquelles les fréquences géniques ne changent pas d'une génération à la suivante, c'est-à-dire en absence de ce qu'on considère les causes de l'évolution : mutation, migration, sélection naturelle et dérive génétique aléatoire. Plus précisément, si p est la fréquence d'allèle A et q est la fréquence d'allèle a , l'équilibre Hardy-Weinberg correspond aux fréquences génotypiques suivantes : p^2AA : $2pqAa$: q^2aa . Voir Hartl & Clark, 1989, *Theoretical Population Genetics*.

⁵² Le taux de changement des fréquences alléliques dû au processus de dérive génétique aléatoire est considéré comme nul ou, en tout cas, négligeable au sein des populations idéales de taille infinie. En ce qui concerne les populations réelles qui sont toujours de taille finie, il est élevé dans des populations d'effectif restreint et bas dans des populations de grande taille. En effet, imaginons que nous ayons une urne contenant 1000 boules dont 600 rouges et 400 noires. En accord avec la loi des grands nombres, il est plus probable que la fréquence relative

attendue : il s'agit d'une variable aléatoire. Autrement dit, comme la population de descendants n'est due qu'à un échantillon de gamètes parmi tous les possibles, la fréquence allélique fluctuera sans direction précise à chaque passage d'une génération à la suivante. Lorsque l'un de deux allèles (A ou a) se sera fixé au sein de la population, les fréquences alléliques ne pourront plus changer sauf à cause d'autres phénomènes que la dérive génétique aléatoire (par exemple, des mutations, des événements de migration, etc.).

Le modèle mathématique de base utilisé en génétique des populations pour décrire et quantifier les effets de la dérive aléatoire sur la composition génétique des populations est la loi binomiale⁵³. À partir des fréquences alléliques initiales au sein d'une population et de la taille de cette dernière, ce modèle ne permet pas de prédire de manière précise les fréquences alléliques, mais d'estimer la distribution de probabilité sur les fréquences alléliques possibles dans la première génération de descendants. Il est important de préciser que la distribution des fréquences alléliques à une génération donnée ne dépend que de ces mêmes fréquences à la génération précédente : la dérive génétique aléatoire est donc un processus stochastique Markovien⁵⁴.

7.2.2. La controverse autour de la dérive génétique aléatoire

Les généticiens des populations étudient la dérive génétique aléatoire depuis les travaux pionniers de Fisher et de Wright dans les années 1920-1930⁵⁵. Son importance évolutionnaire par rapport à celle de la sélection naturelle a été au centre d'un débat controversé entre ces deux généticiens des populations et, plus généralement, au sein de la communauté des biologistes au cours du siècle dernier.

de boules rouges et de boules noires s'écarte de leur fréquence initiale dans l'urne (60% de boules rouges et 40% de boules noires) dans un échantillon de dix boules que dans un échantillon de mille.

⁵³ Voir le Glossaire.

⁵⁴ *Id.* De manière générale, les modèles mathématiques permettant de calculer la distribution statistique des fréquences alléliques révèlent que la probabilité de fixation d'un allèle par dérive aléatoire est égale à sa fréquence initiale.

⁵⁵ Fisher (1999[1930], *The Genetical Theory of Natural Selection: a complete variorum edition*), ne parlait pas encore de « dérive génétique aléatoire » mais d'« effet Hagedoorn », du nom de ceux qui, en 1921, avaient suggéré l'« extinction aléatoire » des allèles comme un facteur majeur de l'évolution (voir Gayon, 1992, *Darwin et l'après-darwin : une histoire de l'hypothèse de sélection naturelle*, p. 338). Wright, 1922, « The Effects of Inbreeding and Crossbreeding on Guinea Pigs. I. Decline of Vigor. II. Differentiation among Inbred Families. III. Crosses between highly Inbred Families » ; 1931, « Evolution in Mendelian Populations ». Depuis les années 1930, on parle souvent de « l'effet de Sewall Wright » pour désigner le phénomène de dérive génétique aléatoire.

Nous ne comptons pas analyser en détail cette controverse et les différences entre l'approche de Fisher et l'approche de Wright dans l'étude du changement évolutif.⁵⁶ Nous rappellerons simplement que, si ces deux généticiens des populations étaient en accord en ce qui concerne la théorie mathématique de l'évolution⁵⁷, ils ne l'étaient pas, entre autres, quant à la question empirique de la taille et de la structure des populations naturelles. D'une part, le raisonnement de Fisher⁵⁸ portait sur des populations très grandes, virtuellement infinies, ce qui lui permettait de soutenir le rôle prépondérant de la sélection dans la détermination de leur parcours évolutif (autrement dit, le rôle négligeable de la dérive génétique aléatoire dans l'évolution). D'autre part, Wright⁵⁹ concentrait son attention sur les effets dus à la limitation des effectifs et raisonnait donc sur des sous-populations de petite taille où les effets de la dérive génétique aléatoire ne peuvent pas être négligés. Ce dernier s'opposait surtout au fait d'attribuer tout changement évolutif à un seul phénomène, que ce soit la sélection naturelle ou la dérive génétique aléatoire. Il affirmait que ces deux facteurs de l'évolution ne sont pas exclusifs mais peuvent contribuer en même temps ou à des moments différents à la détermination du parcours évolutif d'une même population naturelle : c'est pourquoi Wright promouvait l'étude de l'importance relative de la sélection naturelle et de la dérive génétique aléatoire dans des situations biologiques particulières.⁶⁰

À la même époque, la plupart des biologistes évolutionnaires partisans de la Synthèse Moderne minimisèrent le rôle de la dérive génétique aléatoire dans l'évolution. En accord avec Fisher, la position dominante consistait à concevoir la sélection naturelle comme un phénomène omniprésent : elle procédait à l'élimination des nombreuses mutations délétères et à la conservation du petit nombre de mutations avantageuses. La dérive génétique aléatoire était donc considérée comme un facteur d'importance évolutionnaire négligeable⁶¹.

La dérive génétique aléatoire fut admise à plein titre en tant que facteur de l'évolution dans les années 1960-1970, lorsque son rôle fut reconnu et formalisé en ce qui concerne la variation moléculaire, c'est-à-dire au niveau de la séquence nucléotidique de l'ADN et de la

⁵⁶ Nous renvoyons à Gayon, 1992, *op. cit.*, chapitre IX.

⁵⁷ Au sujet de Fisher, Wright (1930, « The Genetical Theory of Natural Selection: A Review », p. 352, notre traduction), écrivit : « nos résultats mathématiques sur la distribution des fréquences géniques sont actuellement en accord complet, dans la mesure où ils sont comparables », « *our mathematical results on the distribution of gene frequencies are now on complete agreement, as far as comparable* ».

⁵⁸ Voir aussi Fisher & Ford, 1950, « The 'Sewall Wright Effect' ».

⁵⁹ Wright, 1948, « On the Roles of Directed and Random Changes in Gene Frequency in the Genetics of Populations ».

⁶⁰ Sur ce débat, voir Beatty 1984, « Chance and Natural Selection », Provine, 1985, « The R.A. Fisher-Sewall Wright Controversy and its Influence Upon Modern Evolutionary Biology ». Nous renvoyons aussi à Skipper, (2002, « The persistence of the R.A. Fisher-Sewall Wright controversy ») qui a montré l'actualité de cette controverse.

⁶¹ Entre autres, voir Mayr, 1963, *Animal Species and Evolution*.

séquence en acides aminés des protéines. Grâce aux avancées de la biologie moléculaire et à l'introduction de nouvelles techniques de détection et d'évaluation de la variation inter- et intra- spécifique et populationnelle au niveau moléculaire (par exemple, la technique de l'électrophorèse), la « théorie neutraliste de l'évolution moléculaire » (ou théorie de l'évolution par dérive aléatoire d'allèles sélectivement neutres) vit le jour⁶². En même temps, la grande querelle entre neutralistes et sélectionnistes éclata⁶³.

Selon la théorie neutraliste de l'évolution moléculaire, la grande majorité des changements évolutifs au niveau moléculaire ne résultent pas du processus de sélection naturelle mais de la dérive aléatoire d'allèles neutres ou presque neutres⁶⁴. Cette théorie n'est pas en contradiction avec l'existence et le rôle de la sélection naturelle dans la détermination du cours de l'évolution adaptative, en particulier au niveau phénotypique. Elle affirme plutôt que, au niveau moléculaire, la plupart des allèles mutants sont sélectivement neutres ou presque neutres, c'est-à-dire ni avantageux ni défavorables mais équivalents quant à leur valeur adaptative dans les conditions environnementales prédominantes de la population concernée.⁶⁵ Ils ne sont donc pas soumis au processus de sélection ; au contraire, ils se fixent ou disparaissent de manière aléatoire en raison de la taille finie de la population porteuse, c'est-à-dire à cause du fait que seulement un nombre restreint d'individus contribuera à la génération suivante.

⁶² Kimura, 1968a, « Evolutionary rate at the molecular level » ; 1968b, « Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles » ; Kimura & Ohta, 1969a, « The average number of generations until fixation of a mutant gene in a finite population » ; 1969b, « The average number of generations until extinction of an individual mutant gene in a finite population » ; King & Jukes, 1969, « Non-Darwinian Evolution ». Pour une introduction à la théorie neutraliste de l'évolution moléculaire, nous renvoyons à Kimura 1990[1983], *La théorie neutraliste de l'évolution*. Pour une analyse critique de l'histoire de la théorie neutraliste, voir Dietrich, 1994, « The Origins of the Neutral Theory of Molecular Evolution » ; 1998, « Paradox and Persuasion: Negotiating the Place of Molecular Evolution Within Evolutionary Biology » ; Dietrich & Millstein, 2008, « The Role of Causal Processes in the Neutral and Nearly Neutral Theories ».

⁶³ Voir Kimura 1990[1983], *op. cit.* : Crow, 1972, « The dilemma of nearly neutral mutations: How important are they for evolution and human welfare? » ; 1981, « The neutralist-selectionist controversy: an overview » ; Harris, 1976, « Molecular evolution: the neutralist-selectionist controversy » ; Calder, 1973, *The Life Game* ; Lewontin, 1974, *The Genetic Basis of Evolutionary Change* ; Ruffié, 1976, *De la biologie à la culture* ; 1982, *Traité du vivant*.

⁶⁴ Kimura souligne qu'il n'est pas nécessaire que des allèles soient strictement neutres, c'est-à-dire complètement équivalents quant à leur valeur adaptative, pour qu'ils se fixent au sein d'une population via le processus de dérive aléatoire. Plus précisément, si l'ordre de grandeur de l'avantage ou du désavantage sélectif qu'ils apportent est très petit, alors on peut les considérer comme « quasiment neutres » dans un environnement donné. Voir Kimura 1990[1983], *op. cit.*

⁶⁵ Pour les données empiriques en faveur de la neutralité de la variation au niveau moléculaire et donc de la théorie neutraliste de l'évolution, voir Kimura, 1991, « Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics ».

7.2.3. La dérive génétique aléatoire : un processus d'échantillonnage indiscriminé sans remplacement

Considérons quelques caractéristiques générales de toute population d'entités biologiques qui évolue via le processus de dérive génétique aléatoire. Notre but est de montrer que la dérive est un processus d'échantillonnage indiscriminé sans remplacement et que les changements évolutifs qu'elle provoque sont donc des événements « aléatoires (au sens) faible ».

Premièrement, à chaque génération, tous les membres d'une population ont la même probabilité d'être échantillonnés, c'est-à-dire de contribuer à la génération suivante, en dépit de leurs différences physico-chimiques (notamment de leurs différences quant à l'allèle qu'ils portent à un même locus). Donc, à fréquence égale d'entités biologiques de types différents (c'est-à-dire d'allèles différents) au sein d'une population à une génération donnée, la probabilité d'être représentée génétiquement à la génération suivante est la même pour toute entité biologique. Lorsqu'un certain type (c'est-à-dire un certain allèle) est plus fréquent qu'un autre, la probabilité qu'une entité biologique de ce type soit représentée génétiquement à la génération suivante est plus élevée par rapport à la même probabilité pour une entité biologique de type différent.

Deuxièmement, à chaque génération, les fréquences alléliques au sein d'une population ont la même probabilité d'augmenter ou de diminuer ; par conséquent, elles ne changent pas de manière systématique dans une direction particulière.

Troisièmement, la composition génétique (en termes de fréquences alléliques) d'une population change à chaque passage d'une génération à la suivante et peut même comporter la fixation ou la disparition irréversible d'un certain type d'entités biologiques (c'est-à-dire, d'un certain allèle) de la population.

Sur cette base, nous concluons que la dérive génétique aléatoire est un processus stochastique qui ne remplit qu'une des deux conditions ($C1 \wedge \neg C2$) pour qu'on puisse invoquer la notion d'« aléatoire (au sens) fort » pour caractériser les changements évolutifs qu'elle provoque :

C1 : C'est un processus d'échantillonnage indiscriminé par rapport aux différences dans les caractéristiques physiques, plus spécifiquement génétiques, des entités biologiques (organismes, gamètes, etc.) de la population concernée. En d'autres termes, ces différences ne jouent aucun rôle dans le processus d'échantillonnage, c'est-à-dire dans la détermination des fréquences (géniques) à la génération suivante. En effet, la probabilité, à une génération

donnée, qu'une entité biologique soit échantillonnée, c'est-à-dire qu'elle contribue à la génération suivante, est la même pour tout membre de la population, indépendamment de ses caractéristiques physico-chimiques. Par conséquent, à fréquence égale d'allèles différents, la probabilité d'être représentée à la génération suivante est la même pour tout allèle.⁶⁶

→C2 : La dérive génétique est un processus qui varie dans le temps, c'est-à-dire un processus d'échantillonnage discriminé relativement aux différences dans la composition génétique de la population à chaque génération au cours du temps évolutionnaire considéré. En d'autres termes, ces différences jouent un rôle causal dans le processus d'échantillonnage au cours du temps, c'est-à-dire dans la détermination de la distribution temporelle des changements des fréquences géniques d'une population. En effet, la probabilité qu'un allèle soit représenté à la génération suivante varie au cours du temps en fonction de la composition génétique de la population, qui change à chaque génération.⁶⁷

Nous venons de montrer que la notion d'« aléatoire (au sens) faible » est pertinente, non seulement pour caractériser au niveau moléculaire les mutations génétiques, mais aussi des phénomènes à forte signification évolutionnaire comme la transmission et la fixation par dérive aléatoire de mutations ou variants génétiques sélectivement neutres. Autrement dit, cette notion peut aussi être utilisée pour décrire et pour expliquer de manière appropriée comment les changements des fréquences géniques ont lieu au niveau populationnel au cours du temps.

Il est intéressant de remarquer que la notion d'« aléatoire (au sens) faible » relativement au processus de dérive génétique aléatoire est définie de manière plus spécifique et précise que cette même notion dans le cas des mutations génétiques. En effet, la dérive génétique aléatoire est un processus d'échantillonnage indiscriminé sans remplacement car elle remplit toujours la condition C1 et ne remplit jamais la condition C2 (voir ci-dessus). Au contraire, les mutations génétiques sont causées par un processus d'échantillonnage qui est discriminé et/ou sans remplacement. La définition de l'« aléatoire (au sens) faible » que nous avons suggérée pour caractériser toute mutation génétique correspond à la manière la plus large d'entendre cette notion et revient à nier la notion d'« aléatoire (au sens) fort ».⁶⁸ En

⁶⁶ Il est évident que, à une génération donnée, la probabilité de fixation d'un allèle dont la fréquence relative au sein d'une population est égale à 80% est plus élevée que la probabilité de fixation de l'allèle alternatif dont la fréquence est égale à 20%. Inversement, la probabilité de disparition de l'allèle le plus fréquent sera plus basse par rapport à celle de l'allèle alternatif.

⁶⁷ La dérive génétique aléatoire est un processus markovien qui, à chaque moment dans le temps (à chaque génération), dépend de ce qui s'est passé à l'instant immédiatement précédent (à la génération précédente).

⁶⁸ En effet, nous avons montré que toute mutation est « aléatoire (au sens) faible » car tout processus de mutation ne remplit pas au moins une des deux conditions, C1 et C2, pour qu'on puisse parler de mutation « aléatoire (au sens) fort ».

revanche, notre proposition n'exclut pas et, au contraire, invite à spécifier la notion d'« aléatoire (au sens) faible » sur la base des caractéristiques physico-chimiques de chaque processus mutationnel particulier. C'est à la lumière d'une étude détaillée du processus de mutation qu'il est possible (et souhaitable) d'identifier de manière plus précise quelles conditions il remplit et, sur cette base, de restreindre la notion d'« aléatoire (au sens) faible » appropriée pour caractériser les mutations génétiques qu'il engendre.

7.3. La signification et le rôle de la notion d'« aléatoire (au sens) faible »

Dans ce chapitre, nous avons introduit et défini la notion d'« aléatoire (au sens) faible » et nous l'avons identifiée à la notion de hasard appropriée du point de vue empirique pour caractériser toute mutation génétique au niveau moléculaire. Quels sont la signification et le rôle de cette notion ? Essayons de répondre à cette question en analysant le cas du phénomène de mutation génétique.

Le fait qu'une mutation est « aléatoire (au sens) faible » signifie qu'elle est le résultat d'un processus stochastique physico-chimique qui n'est pas indifférent aux caractéristiques physico-chimiques de la séquence d'ADN ET/OU aux différences dans les conditions environnementales, au sens large, au cours d'une génération cellulaire donnée (ou, plus généralement, au cours du laps de temps considéré). Cette formulation revient à affirmer qu'une mutation est « aléatoire (au sens) faible » simplement si elle est due à un processus stochastique ou, autrement dit, simplement si elle est un événement pouvant avoir lieu avec une certaine probabilité. Sa probabilité peut être la même que celle d'autres événements mutationnels possibles OU elle peut être invariante dans le temps et donc indépendante de tout autre événement mutationnel et environnemental. Par contre, si la probabilité qu'une mutation ait lieu est à la fois égale à la probabilité d'autres mutations possibles et invariante dans le temps, alors il ne s'agit pas d'une mutation « aléatoire (au sens) faible », mais « (au sens) fort » c'est-à-dire due à un processus stochastique d'échantillonnage indiscriminé à la fois du point de vue spatial et temporel. C'est là la signification de cette notion très générale d'aléatoire.

La notion d'« aléatoire (au sens) faible » peut être définie en termes subjectifs ou objectifs selon la façon dont les probabilités qu'elle évoque sont conçues. La correspondance dont nous avons fait l'hypothèse entre les interprétations épistémique et objective des

probabilités et les notions de hasard subjectif et objectif⁶⁹ nous suggère que la résolution du problème de l'interprétation des probabilités dans les descriptions et dans les prédictions formelles des mutations génétiques pourrait nous permettre de comprendre si la notion d'« aléatoire (au sens) faible » doit être comprise, dans ce cas, comme une notion de hasard subjectif ou objectif.⁷⁰

Le caractère probabiliste des mutations génétiques, en particulier en ce qui concerne le type, le site, le moment et le taux de leur occurrence, peut être conçu en évoquant une interprétation épistémique ou objective des probabilités. De plus, au sein de chaque interprétation, la notion de probabilité peut prendre des significations différentes (selon la théorie philosophique des probabilités adoptée), chacune renvoyant à une notion de hasard caractérisée par une signification précise⁷¹. Considérons quelques exemples. Premièrement, le caractère probabiliste des descriptions et des prédictions formelles des mutations pourrait être dû aux limites des connaissances que les biologistes ont des processus physico-chimiques à leur origine causale : les probabilités pourraient donc être interprétées de manière épistémique comme la mesure du degré d'ignorance partielle à ce sujet ou, en d'autres termes, comme la mesure du hasard en tant qu'ignorance des causes sous-jacentes aux mutations.

Deuxièmement, il pourrait aussi être dû à l'indéterminisme fondamental des processus physico-chimiques de mutation : les probabilités seraient alors objectives, consistant par exemple en propensions de certaines conditions physico-chimiques à produire certaines altérations de la séquence nucléotidique, c'est-à-dire du caractère intrinsèquement probabiliste du hasard conçu comme une entité métaphysique existant et agissant dans le monde.

Troisièmement, le caractère probabiliste des descriptions et des prédictions des mutations génétiques pourrait aussi trouver son origine dans d'autres caractéristiques objectives des processus physico-chimiques les produisant, par exemple dans l'instabilité dynamique qui leur est propre ou encore dans le fait que les mutations sont des événements situés à la rencontre de plusieurs séries causales déterministes mais indépendantes. Dans ces deux cas, les probabilités pourraient être interprétées de manière objective, par exemple, comme la mesure du degré de possibilité des mutations relativement à un processus physico-chimique particulièrement sensible aux conditions initiales, ou en les identifiant aux fréquences relatives auxquelles ces séries causales indépendantes à l'origine des mutations se

⁶⁹ Dans la section 2.6.3 du chapitre 2, nous avons suggéré que les probabilités sont la mesure du hasard et que leur signification relativement à un événement donné est strictement liée à celle de la notion de hasard qui caractérise ce même événement.

⁷⁰ Nous précisons que la notion d'« aléatoire (au sens) faible » est compatible avec toutes les notions de hasard que nous avons introduites dans les chapitres 1 et 2 mais n'implique aucune d'entre elles de manière nécessaire.

⁷¹ Voir les chapitres 1 et 2.

croisent (c'est-à-dire, comme la mesure du hasard en tant que sensibilité aux conditions initiales et du hasard en tant que rencontre de séries causales indépendantes).

Enfin, le fait qu'on mobilise des probabilités pour décrire et pour prédire l'occurrence des mutations peut aussi être attribué à l'imprédictibilité des événements mutationnels par un modèle de calcul universel (par une machine de Turing universelle) : leur interprétation, épistémique ou objective, serait ouverte à débat. Quant à leur signification, ces probabilités pourraient être considérées comme la mesure du caractère aléatoire des mutations en termes de la théorie algorithmique de l'information, donc de complexité. Cependant, nous avons déjà remarqué que la notion algorithmique de hasard, par exemple en tant qu'incompressibilité, ne nous semble pas appropriée dans le cas d'un phénomène comme les mutations génétiques dont beaucoup de propriétés restent encore dans l'ombre⁷². De plus, nous avons vu que, au cours des trente dernières années, la recherche en génétique moléculaire a révélé l'existence de nombreux biais de mutation, c'est-à-dire de nombreuses corrélations (statistiques et causales) entre la probabilité qu'une mutation ait lieu et la présence de divers facteurs physico-chimiques dans le milieu intra- et extra- cellulaire. Ces avancées semblent suggérer qu'il est possible de prédire (autrement dit, de calculer), bien que toujours de manière probabiliste, l'occurrence des mutations génétiques.

De manière générale, les biologistes ne semblent s'intéresser, ni à la question de l'interprétation épistémique ou objective des probabilités, ni à celle de savoir si la notion de hasard qu'il est approprié d'évoquer dans les descriptions et dans les prédictions des mutations génétiques désigne un hasard subjectif ou objectif⁷³. Nous avons, au contraire, décidé de placer ces questions au centre de ce travail. Ce n'est cependant que dans la troisième partie que nous proposerons une solution possible, avec la caractérisation, au niveau moléculaire, d'une autre source causale de variation biologique, à savoir le bruit (ou les fluctuations stochastiques) dans l'expression des gènes. Dans cette troisième partie, nous montrerons d'abord que cette source causale de la variation phénotypique peut, aussi bien que les mutations génétiques, être caractérisée par la notion d'« aléatoire (au sens) faible » (et par la notion de « hasard évolutionnaire ») ; par conséquent, se poseront les mêmes questions sur l'interprétation des probabilités. Ensuite, nous proposerons une piste possible (l'analyse de robustesse) en faveur d'une interprétation objective des probabilités figurant dans les modèles stochastiques de l'expression des gènes. Enfin, nous expliciterons la notion d'objectivité que nous utilisons et qui ne correspond pas à la notion philosophique d'objectivité absolue.

⁷² Voir le chapitre 2.

⁷³ Beatty, 1984, « Chance and Natural Selection ».

Nous n'avons pas encore traité le rôle de la notion d'« aléatoire (au sens) faible », en particulier dans la caractérisation des mutations génétiques. Rappelons brièvement que la place occupée par la notion de hasard dans un cadre théorique dépend directement de la signification propre à cette notion. Par exemple, admettons que les probabilités utilisées dans les descriptions et dans les prédictions formelles des mutations génétiques sont la mesure de notre ignorance partielle et, sur cette base, identifions le caractère « aléatoire (au sens) faible » des mutations au hasard subjectif en tant qu'ignorance des causes. Dans ce cas, le rôle de la notion d'« aléatoire (au sens) faible » serait de témoigner d'un état d'ignorance ou de connaissance partielle, c'est-à-dire un rôle épistémologique et descriptif. Cependant, nous pourrions aussi la considérer comme une notion de hasard subjectif mais explicative car, en identifiant les lacunes de notre connaissance quant au processus de mutation, elle rendrait compte du caractère probabiliste, d'un point de vue épistémique, des descriptions et des prédictions formelles au sujet des mutations.

Au contraire, si l'on interprète ces mêmes probabilités, par exemple, comme la mesure du degré de possibilité des mutations ou comme la mesure des propensions qui tendent à les produire, on identifiera le caractère « aléatoire (au sens) faible » des mutations à un hasard objectif (par exemple, au hasard en tant que sensibilité aux conditions initiales ou au hasard pur) : le rôle de la notion d'« aléatoire (au sens) faible » sera alors d'expliquer le caractère objectivement probabiliste des mutations par la description des caractéristiques du processus qui les cause. Plus précisément, elle explique pourquoi une mutation n'a pas à la fois la même probabilité d'avoir lieu tout au long d'une séquence d'ADN ET au cours d'une génération cellulaire. Du point de vue du généticien moléculaire qui étudie les mutations par des méthodes statistiques, la notion d'« aléatoire (au sens) faible » en tant que hasard objectif explique pourquoi il ne s'attend pas à observer – et, la plupart du temps, il n'observe pas – une distribution uniforme des fréquences relatives des mutations à la fois le long de la séquence d'ADN et au cours d'une génération.

7.4. La signification et le rôle de la notion de « hasard évolutionnaire »

En guise de conclusion, il nous reste à analyser de plus près la notion de « hasard évolutionnaire » que nous avons introduite et définie dans le chapitre 4 et dont nous avons

montré la pertinence conceptuelle et empirique pour caractériser toute mutation génétique selon une perspective évolutionnaire⁷⁴.

Rappelons qu'une mutation génétique est le fruit du hasard au sens de la notion de « hasard évolutionnaire » si et seulement si elle n'est pas causée de manière spécifique et (exclusivement) adaptative par un processus physico-chimique en réponse aux conditions de l'environnement, ce qui signifie qu'elle n'est pas une mutation « dirigée » par rapport au cours adaptatif de l'évolution. Plus précisément, il s'agit d'une mutation qui ne remplit pas au moins une des deux conditions suivantes : 1) être plus probable dans un environnement E_1 où elle est avantageuse que dans un environnement E_2 où elle est défavorable ou neutre ; 2) être nettement plus probable dans un environnement E_1 où elle est avantageuse que d'autres mutations neutres ou défavorables.

Avant de conclure sur la signification et sur le rôle de la notion de « hasard évolutionnaire », il est intéressant de remarquer qu'une mutation génétique qui n'est pas « dirigée » en vue de l'adaptation est aussi une mutation « aléatoire (au sens) faible ». En effet, bien que la notion de « hasard évolutionnaire » ait une extension plus grande et n'exclue pas la possibilité, au moins logique, qu'une mutation puisse être « aléatoire (au sens) fort »⁷⁵, nous venons de montrer que les notions de « hasard évolutionnaire » et d'« aléatoire (au sens) faible » correspondent aux notions appropriées pour caractériser toute mutation génétique, respectivement selon une perspective évolutionnaire et au niveau moléculaire. La pertinence de cette conclusion serait mise en cause si l'on montrait qu'il y a une contradiction entre ces deux notions, ce qui n'est pas le cas.

Nous suggérons la possibilité de décliner la notion d'« aléatoire (au sens) faible » en termes évolutionnaire, c'est-à-dire en référence à la relation entre l'occurrence des mutations et leur utilité du point de vue de l'évolution, de la manière suivante. Une mutation est « *aléatoire (au sens) faible* » si et seulement si elle est le résultat d'un processus stochastique physico-chimique (causal) qui ne remplit pas au moins une des deux conditions suivantes ($\neg C1 \vee \neg C2$) :

C1 : C'est un processus d'échantillonnage indiscriminé par rapport aux différences dans la valeur adaptative des mutations génétiques possibles pouvant être plus ou moins favorables, défavorables, ou sélectivement neutres.

⁷⁴ Voir le chapitre 6.

⁷⁵ Par contre, nous avons vu que des mutations « aléatoires (au sens) fort » sont physiquement très peu probables.

C2 : C'est un processus d'échantillonnage indiscriminé par rapport aux différences dans les caractéristiques de l'environnement où l'organisme concerné se trouve au cours du temps.

La déclinaison en termes évolutionnaires de la notion d'« aléatoire (au sens) faible » vient donc préciser la notion de « hasard évolutionnaire » et, dans sa version relative aux mutations au niveau moléculaire, complète notre compréhension de cette source causale de la variation biologique.

La notion de « hasard évolutionnaire » évoque deux notions de hasard objectif qui sont conceptuellement et historiquement très liées : les notions de hasard en tant qu'absence d'un dessein et en tant que rencontre de séries causales indépendantes⁷⁶. En premier lieu, de même que ces dernières, le « hasard évolutionnaire » est une notion de hasard objectif car elle porte sur des propriétés inhérentes aux phénomènes dans le monde et non pas sur la relation épistémique qu'on entretient avec ces derniers. Dans le cas des mutations, la notion de « hasard évolutionnaire » porte sur la manière dont elles sont engendrées par un processus physico-chimique ou, en d'autres termes, sur le type de liens que la valeur adaptative des mutations dans une situation environnementale donnée peut entretenir avec les besoins adaptatifs des organismes concernées et situés dans cet environnement.⁷⁷

En deuxième lieu, la signification de la notion de « hasard évolutionnaire » peut être comprise à la fois comme la déclinaison en termes spécifiques, biologiques et évolutionnaires, de la signification de la notion de hasard en tant que rencontre de séries causales indépendantes et en tant qu'absence d'un dessein, cette dernière étant double car elle comprend la première, qu'est la notion cournotienne de hasard, en son sein⁷⁸. Plus précisément, le fait qu'une mutation génétique soit décrite en termes de « hasard évolutionnaire » signifie, d'une part, qu'il n'y a pas de relation spécifique qui va de l'ordre des causes à l'ordre des effets entre la valeur adaptative d'une mutation se produisant dans un organisme individuel situé dans un environnement donné et le fait que cette mutation se produise dans cet environnement (autrement dit, qu'il n'y a pas de causalité finale impliquée dans la production d'une mutation). D'autre part, que la mutation se situe à la rencontre de deux séries causales causalement indépendantes, l'une déterminant l'occurrence d'une

⁷⁶ Voir le chapitre 1.

⁷⁷ Il faudrait plutôt reformuler cette idée de manière négative, c'est-à-dire en affirmant que la notion de « hasard évolutionnaire » porte sur la manière dont les mutations ne sont pas engendrées par un processus physico-chimique ou, en d'autres termes, sur les types de liens que leur valeur adaptative, dans une situation environnementale donnée, ne peut pas entretenir avec les besoins adaptatifs des organismes porteurs dans cet environnement. Il est intéressant de remarquer que la plupart des notions de hasard introduites et analysées dans le présent travail sont définies de manière négative (notamment le hasard en tant qu'ignorance des causes, qu'absence d'un dessein, que rencontre de séries causales non dépendantes et qu'incompressibilité).

⁷⁸ Voir le chapitre 1.

certaine mutation (ou la probabilité qu'une certaine mutation ait lieu) au sein d'un organisme individuel situé dans un certain environnement, l'autre déterminant les besoins adaptatifs de l'organisme porteur, et donc la valeur adaptative de la mutation qui se produit, dans cet environnement.

En troisième lieu, nous soutenons que la notion de « hasard évolutionnaire », de même que les notions de hasard en tant qu'absence d'un dessein et en tant que rencontre de séries causales indépendantes, a un rôle explicatif dans la caractérisation des mutations génétiques du point de vue du processus évolutif et, plus spécifiquement, de l'adaptation. En étant définie en termes de processus, elle explique pourquoi une mutation n'est pas à la fois plus probable dans un environnement où elle est avantageuse que dans un environnement où elle ne l'est pas ET nettement plus probable dans un environnement où elle est avantageuse que d'autres mutations neutres ou défavorables. Du point de vue du biologiste qui observe et étudie le phénomène de mutation, la notion de « hasard évolutionnaire » explique pourquoi il ne s'attend pas à ce qu'une mutation soit plus fréquente dans un environnement E_1 où elle est avantageuse que dans un environnement E_2 où elle ne l'est pas ET qu'elle ait une fréquence nettement plus élevée par rapport à d'autres mutations délétères ou neutres dans un environnement E_1 où elle est avantageuse. Cette notion explique aussi pourquoi, la plupart du temps, le biologiste n'observe pas de telles fréquences de mutations favorables.

On pourrait poser la question de savoir si la notion de « hasard évolutionnaire », bien que spécifique à la biologie évolutionnaire – et, dans notre formulation, au phénomène de mutation – a la même extension que les deux notions de hasard qu'elle évoque, c'est-à-dire si elle est équivalente à ces dernières. Nous nous limiterons à remarquer que, premièrement, de même que les notions de hasard en tant qu'absence d'un dessein et en tant que rencontre de séries causales indépendantes, la notion de « hasard évolutionnaire » est compatible avec la notion de hasard en tant que sensibilité aux conditions initiales mais ne l'implique pas pour autant. Cependant, imaginons que le processus physico-chimique d'expression génétique et, plus généralement, de développement – qui va du matériel génétique des organismes individuels aux traits phénotypiques de ces derniers – soit instable au sens où de petites modifications dans ses conditions initiales au niveau de la séquence nucléotidique (c'est-à-dire des mutations) provoquent des conséquences importantes au niveau du résultat phénotypique final. Si cette situation s'avérait, le fait qu'une mutation puisse être causée de manière spécifique et (exclusivement) adaptative en réponse aux conditions de l'environnement, c'est-à-dire qu'elle puisse être « dirigée » du point de vue évolutionnaire, serait fortement improbable (du point de vue physico-chimique).

Deuxièmement, la notion de « hasard évolutionnaire » aussi bien que les notions qu'elle évoque n'excluent pas que des notions de hasard subjectif puisse entrer en jeu, notamment dans la caractérisation de l'état de nos connaissances sur les causes sous-jacentes aux mutations génétiques. Il est en effet indéniable que, parfois, les biologistes utilisent à la fois la notion de « hasard évolutionnaire », désignant une propriété objective, et d'autres notions d'ordre épistémique (par exemple, celle de hasard en tant qu'ignorance des causes) pour rendre compte à la fois des propriétés des mutations (ou d'autres phénomènes biologiques) et des limites des connaissances actuelles à leur sujet.

Troisièmement, relativement aux notions mathématiques de hasard, nous avons déjà remarqué que la notion de « hasard évolutionnaire » et les notions d'absence d'un dessein et de rencontre de séries causales indépendantes sont compatibles avec l'équiprobabilité des événements possibles qu'elles caractérisent aussi bien qu'avec une notion de hasard en termes d'information algorithmique (par exemple, l'idée d'incompressibilité par une machine de Turing universelle). Cependant, d'une part, nous avons montré que la notion d'« aléatoire (au sens) faible » délimite l'extension de la notion de « hasard évolutionnaire » et, de cette manière, enrichit notre compréhension du phénomène de mutation génétique. En particulier, nous avons affirmé que, du point de vue empirique, aucune mutation (ou, du moins, que la plupart des mutations physiquement possibles) n'est due à un processus stochastique qui est indiscriminé à la fois par rapport au site de l'ADN ET par rapport au moment dans le temps où il produit des mutations (autrement dit, qu'aucune altération de la séquence nucléotidique de l'ADN est à la fois équiprobable par rapport aux autres altérations possibles ET au cours du temps). D'autre part, nous avons soutenu que la notion de hasard définie en termes de complexité algorithmique n'est pas pertinente, du moins en l'état actuel de la recherche en biologie, pour caractériser les mutations génétiques. En effet, les connaissances à ce sujet sont encore lacunaires ; de plus, au fur et à mesure que les recherches avancent, des liens de plus en plus forts se révèlent entre les dynamiques évolutives au niveau populationnel et les caractéristiques des mutations génétiques, ce qui s'oppose à l'idée de leur imprédictibilité par un modèle de calcul universel.

Il reste à traiter la question de ce que la notion de « hasard évolutionnaire » dans le cas des mutations génétiques implique quant à la nature fondamentalement déterministe ou indéterministe du processus physico-chimique à leur origine, quant à la possibilité de prédire, en principe, leur occurrence et quant à la valeur du déterminisme méthodologique dans l'étude de ces phénomènes biologiques. De la même manière que les notions de hasard en tant qu'absence d'un dessein et en tant que rencontre de séries causales indépendantes, la notion

de « hasard évolutionnaire » est compatible à la fois avec le déterminisme et l'indéterminisme métaphysique, sur lesquels nous ne nous prononçons pas pour les raisons que nous avons déjà données plus haut⁷⁹. C'est pourquoi elle peut mettre en cause, mais pas de manière nécessaire, la thèse épistémologique du déterminisme scientifique (notamment, dans le cas où le processus de mutation est considéré comme fondamentalement indéterministe) et ne représente aucun défi pour des pratiques favorisant des hypothèses déterministes en science.

⁷⁹ Voir la première partie du présent travail.

Partie III – Hasard et bruit dans l'expression des gènes

Le bruit dans l'expression des gènes est une des sources causales potentielles de la variation au sein des populations naturelles. Il s'agit de petites fluctuations, qu'on considère être stochastiques et imprédictibles, et qui peuvent perturber toutes les étapes de ce processus intracellulaire impliqué dans le développement des organismes vivants. Même en l'absence de toute modification au niveau du matériel génétique comme au niveau de l'environnement externe (à savoir, de l'ensemble des facteurs abiotiques et biotiques constituant le milieu où les organismes se trouvent), le bruit dans l'expression des gènes peut engendrer de la variation phénotypique au sein d'une population ou chez un même individu au cours du temps.

Cette possibilité d'une variation phénotypique qui ne serait ni d'origine génétique ni d'origine environnementale, mais qui serait due à des perturbations du processus de développement, est un objet de recherche en biologie depuis la moitié du siècle dernier, époque à laquelle on parlait déjà de « bruit développemental » en tant que phénomène de perturbation aléatoire et imprédictible. Cependant, ce n'est qu'à partir de la fin des années 1970 que l'étude de cette source causale de variation biologique s'est développée et a apporté des résultats intéressants. En particulier, à la fin des années 1990, l'étude du bruit s'est focalisée sur l'analyse des fluctuations dans le processus intracellulaire d'expression génétique. Des méthodes expérimentales nouvelles et des modèles théoriques probabilistes ont été introduits puis utilisés pour étudier le caractère aléatoire et imprédictible de ces fluctuations et pour rendre compte de son impact sur le déroulement et le résultat final de l'expression des gènes.

Nous avons décidé de consacrer la troisième partie de notre travail au bruit dans le processus d'expression génétique, car il s'agit d'un phénomène dont on ne commence à considérer et à connaître le rôle en tant que source causale de variation que depuis deux décennies. Par ailleurs, cette notion nous intéresse parce que les biologistes décrivent systématiquement le bruit dans l'expression des gènes par des termes comme « aléatoire », « stochastique », « probabiliste », « imprédictible », etc., qui évoquent la notion de hasard. Notons à cet égard que ces fluctuations intracellulaires impliquent une modélisation mathématique du processus d'expression génétique mobilisant des probabilités.

Dans cette troisième partie, nous examinerons trois questions strictement liées. Premièrement, quels sont la signification et le rôle de la notion de hasard à laquelle les biologistes renvoient lorsqu'ils utilisent les termes « aléatoire », « stochastique », « probabiliste », « imprédictible », etc., pour caractériser le bruit dans l'expression des gènes ? Deuxièmement, s'agit-il d'une notion conceptuellement et empiriquement appropriée et, si ce

n'est pas le cas, quelle notion de hasard doit-on invoquer dans ce cadre ? Troisièmement, comment doit-on interpréter les probabilités figurant dans les modèles mathématiques du processus d'expression génétique ? Nous répondrons à ces trois questions en laissant de côté la question de la nature déterministe ou indéterministe, d'un point de vue métaphysique, de l'expression génétique et du bruit l'affectant. Nous rappellerons néanmoins les principales motivations de notre point de vue agnostique sur cette question. Nous soulignerons aussi que la question d'ordre épistémologique sur la possibilité et la légitimité d'argumenter en faveur du déterminisme ou de l'indéterminisme du monde biologique ne peut que recevoir une réponse négative.

Le chapitre 8 introduit les connaissances reçues sur le bruit dans l'expression des gènes. Nous fournirons les bases nécessaires pour notre analyse critique de la notion de hasard dans la caractérisation de cette source causale de la variation phénotypique. Au chapitre 9, nous analysons plus en détail le bruit intracellulaire en focalisant notre attention sur les facteurs physico-chimiques pouvant influencer ses propriétés en fonction du gène concerné et du moment où l'expression génétique a lieu. Nous montrerons que le bruit affectant ce processus intracellulaire est biaisé, à la fois du point de vue évolutionnaire et du point de vue plus strictement physico-chimique. Nous concluons que, de la même manière que dans le cas des mutations génétiques, les notions de hasard conceptuellement et empiriquement appropriées pour caractériser le bruit dans l'expression des gènes sont les notions de « hasard évolutionnaire » et d'« aléatoire (au sens) faible ». Au chapitre 10, nous examinerons la question de savoir comment on doit interpréter les probabilités figurant dans les modèles stochastiques du processus d'expression génétique. Nous présenterons un argument possible, fondé sur l'analyse de robustesse, en faveur d'une interprétation objective de ces probabilités. Nous concluons en montrant que la notion d'objectivité à laquelle nous faisons référence, et que nous proposons en tant que pertinente en science, ne correspond pas à la notion d'objectivité absolue.

Chapitre 8 – Le bruit dans l’expression des gènes : une introduction

8.1. L’étude du bruit en science

Le terme « bruit » est souvent employé pour désigner une notion qui, à la fois selon son sens commun et son sens technique en science, est connotée de manière plutôt négative. Dans plusieurs disciplines scientifiques et dans son sens le plus général, ce terme désigne de la variation désordonnée, irrégulière, aléatoire et imprédictible, de n’importe quelle quantité mesurable. Par exemple, cette définition du bruit est adoptée en physique : on parle de bruit externe lorsque la variation aléatoire est due à des sources extérieures au système physique considéré (par exemple, des sources naturelles, l’utilisation de technologies expérimentales et d’instruments de mesure, etc.) et de bruit interne lorsqu’elle est due à des sources au sein du système physique même (par exemple, l’agitation thermique des molécules)¹. En particulier, en acoustique et dans le contexte des technologies du son, le terme « bruit » est utilisé pour désigner une perturbation sonore désagréable et désordonnée, en d’autres termes, de la variation nuisible par rapport à un son ordonné et harmonieux (en ce sens, le bruit est synonyme de « cacophonie »).

Un autre domaine scientifique où la notion de « bruit » joue un rôle important et reçoit une définition précise est la théorie de l’information². Dans le cadre de cette théorie, qui s’occupe de la transmission du signal d’un point de vue quantitatif, le terme « bruit » désigne une perturbation aléatoire et imprédictible du signal pouvant se produire pendant sa transmission (dans le canal) ou au niveau du récepteur, ce dernier devant décoder le signal

¹ Il n’y a pas de consensus quant au moment où le bruit est devenu un concept désignant de manière précise des phénomènes physiques comme le mouvement erratique d’une molécule dans un milieu liquide (ce qu’on appelle une molécule brownienne). La réponse standard des historiens des sciences renvoie aux recherches du physicien allemand Schottky sur le bruit électrique, c’est-à-dire la variation aléatoire de l’intensité du courant électrique dans un tube à vide (1917-1918). Cependant, Einstein avait déjà parlé du bruit en 1905, plus précisément lorsqu’il décrivit mathématiquement le mouvement perpétuel et erratique des molécules browniennes. Ce dernier avait été observé pour la première fois par le botaniste écossais Brown à la fin du XIX^e siècle ; cependant, il n’a été l’objet d’une formalisation mathématique (Einstein), suivie par une vérification expérimentale (Perrin), qu’au début du XX^e siècle. Pour une bibliographie complète sur l’histoire du bruit, nous revoyons à Cohen (2005, « The history of noise [on the 100th anniversary of its birth] »), selon lequel Einstein aurait « inventé » le bruit afin de montrer l’existence des atomes.

² Shannon, 1948, « A Mathematical Theory of Communication ».

reçu et reproduire le signal d'origine. Le bruit est considéré comme de l'information mais de l'information insignifiante, non pertinente, sans structure, alors que le signal représente une information utile, significative et ordonnée. Plus précisément, il est défini comme une erreur aléatoire ajoutée au signal. Shannon³ propose de visualiser le bruit de la manière suivante : si le signal transmis correspond à un point, le bruit ajouté au signal, donc externe relativement au signal, produit une petite région d'incertitude autour de ce point dans l'espace. Enfin, le bruit se distingue d'un autre type de perturbation, la distorsion : cette dernière est une opération de perturbation fixe appliquée au signal ; le bruit, au contraire, est une perturbation statistique et imprédictible. Donc, contrairement à la distorsion, lorsqu'un même bruit vient à perturber le signal en s'ajoutant à lui, il ne produit pas toujours le même changement.

8.2. L'étude du bruit en biologie

En biologie, la notion de « bruit » a gardé pendant longtemps la connotation négative qui lui vient de la physique théorique et appliquée mais aussi de la théorie de l'information : elle fait référence à une force désordonnée, imprédictible et destructrice de tout processus naturel et artificiel. De manière générale, on peut utiliser l'expression « bruit biologique » pour désigner toute perturbation ou fluctuation aléatoire affectant n'importe quel phénomène biologique.

Cette conception du bruit est présente de manière transversale dans plusieurs disciplines biologiques. Par exemple, en génétique, en particulier à partir de la découverte de la structure physico-chimique de l'ADN, le bruit a longtemps été conçu en opposition à la notion de programme génétique déterministe qui, en tant que dépositaire d'une information constructive et ordonnée nécessaire au développement et à l'évolution des organismes vivants, doit être transmis⁴. Au cours de la deuxième moitié du XX^e siècle, les généticiens moléculaires ont donc eu tendance à parler du bruit pour désigner tout phénomène aléatoire et imprédictible perturbant, soit la transmission fidèle du matériel génétique d'une génération à la suivante (les mutations, la recombinaison, la transposition, etc.), soit son expression et la production du phénotype qui lui est normalement associé (par exemple, des erreurs dans les processus de transcription de l'ADN en ARN et de traduction de l'ARN en protéines).

³ Shannon, 1998, « Communication in the Presence of Noise », p. 449.

⁴ Voir Samoilov *et al*, 2006, « From Fluctuations to Phenotypes: The Physiology of Noise ».

Une autre discipline biologique qui, depuis plus de cinquante ans, mobilise le terme « bruit » est la biologie du développement⁵. Ce domaine de recherche nous intéresse particulièrement dans le cadre de ce travail sur la notion de hasard et les sources causales de la variation biologique. Plus précisément, notre attention est attirée par l'étude des différences phénotypiques qui se produisent au cours du développement d'un même organisme situé dans un environnement homogène et constant au cours du temps ou d'organismes individuels isogéniques (c'est-à-dire avec le même matériel génétique) situés dans des environnements (hypothétiquement) identiques. En effet, l'objet au centre de cette troisième partie est un phénomène biologique qui ne provoque de modifications, ni dans le matériel génétique ni dans l'environnement externe, mais à l'intérieur des organismes individuels mêmes au niveau du processus de développement et, plus spécifiquement, de l'expression des gènes. Il s'agit de ce que les biologistes appellent le « bruit développemental ».

8.2.1. « La face noire de la variation phénotypique »⁶ : le bruit développemental

Avant de tracer les grandes lignes de la recherche sur le bruit développemental, de la moitié du siècle dernier jusqu'à aujourd'hui, rappelons quelles sont les sources causales des différences phénotypiques entre des organismes isogéniques. Deux sortes de modifications non génétiques doivent être mentionnées : les changements de l'environnement dans lequel les organismes se trouvent ; les changements se produisant dans les organismes mêmes ou, plus précisément, concernant tout processus physico-chimique impliqué dans leur développement. Les modifications environnementales sont omniprésentes dans l'espace et dans le temps et donc affectent tout organisme vivant : elles peuvent être dues à des facteurs physiques et chimiques divers et aux autres organismes situés dans le même environnement que l'organisme individuel concerné. Les modifications non génétiques se produisant à l'intérieur des organismes mêmes peuvent affecter des processus divers contrôlant le développement à différents niveaux d'organisation, du niveau moléculaire à l'intérieur de la cellule jusqu'au niveau de l'organisme individuel (ou organismique). Bien que ces modifications soient aussi inévitables que les changements de l'environnement, les biologistes

⁵ La biologie du développement est l'étude de tout changement se produisant au cours de la vie des organismes vivants (la croissance, la différenciation, la maturation, la reproduction, le vieillissement, la mort), plus spécifiquement l'étude du processus de transformation progressive d'une cellule unique à un organisme multicellulaire complet (c'est-à-dire, le développement).

⁶ West-Eberhard, 2003, Forword *In* Polak (ed.), *Developmental Instability. Causes and Consequences*.

n'ont commencé à les étudier de manière systématique qu'à la fin de la première moitié du XX^e siècle, en particulier afin d'expliquer les asymétries morphologiques entre le côté droit et le côté gauche du corps des organismes bilatéraux (les « asymétries fluctuantes »)⁷.

Autour des années 1950-1960, les biologistes ont commencé à publier des articles dans lesquels l'expression « bruit développemental »⁸ ou « instabilité développementale »⁹ était utilisée pour désigner ces changements qui ne concernent ni le matériel génétique, ni l'environnement, et qui sont à l'origine de la variation phénotypique au sein des populations naturelles¹⁰. Les recherches sur l'impact du bruit ont donc commencé à privilégier l'étude du développement d'organismes génétiquement identiques (par exemple, des clones) qui, même s'ils sont situés dans un environnement homogène et constant ou dans des environnements (hypothétiquement) identiques, présentent des différences phénotypiques (par exemple, morphologiques)¹¹. En effet, dans ce cas, l'absence de toute autre source de variation rend

⁷ Nous reviendrons sur ce sujet ci-dessus.

⁸ Waddington semble avoir été l'un des premiers à utiliser cette expression, plus précisément en 1957 dans *The Strategy of the Genes* (1957, p. 39-41). Il caractérise le bruit développemental de manière figurative en affirmant que « si la canalisation est représentée par une vallée dans le paysage épigénétique, le bruit affectant un système pourrait être symbolisé par l'imperfection de la sphéricité de la balle qui descend la vallée », p. 40, (notre traduction), « *If canalisation is represented as a valley in an epigenetic landscape, the noisiness of the system might perhaps be symbolized by the imperfection of the sphericalness of the ball which runs down the valley.* »

⁹ L'expression « instabilité développementale » semble remonter à une époque relativement plus ancienne que celle de « bruit développemental », plus précisément au développement de la biométrie (l'étude quantitative des êtres vivants à l'aide des méthodes statistiques) au cours de la première moitié du siècle dernier.

¹⁰ Voir Reeve & Robertson, 1953, « Analysis of environmental variability in quantitative inheritance » ; 1954, « Studies in quantitative inheritance. VI. Sternite chaeta number in *Drosophila*: a metamerie quantitative character » ; Mather, 1953a, « The genetical structure of populations » ; 1953b, « Genetical control of stability in development » ; Tebb & Thoday, 1954, « Stability in development and relational balance of X chromosomes in *Drosophila melanogaster* » ; Waddington, 1957, *op. cit.* ; Thoday, 1958, « Homeostasis in a selection experiment » ; Beardmore, 1960, « Developmental stability in constant and fluctuating temperatures » ; Reeve, 1960, « Some genetic tests on asymmetry of sternopleural chaeta number in *Drosophila* » ; Bradshaw, 1965, « Evolutionary significance of phenotypic plasticity in plants » ; Soulé & Baker, 1968, « Phenetics of natural populations. IV. The population asymmetry parameter in the butterfly *Coenonympha tullia* » ; Soulé, 1982, « Allomeric Variation. I. The Theory and Some Consequences ». Cependant, les premières études sur la variation phénotypique, plus précisément morphologique, due à des sources causales qui ne sont ni génétiques ni environnementales externes datent déjà de la fin du XIX^e siècle et du début du XX^e siècle : voir Willey, 1894, *Amphioxus and the ancestry of the vertebrates* ; Pearson, 1901, « Mathematical contributions to the theory of Evolution. IX. On the Principle of Homotyposis and its relation to heredity, to the variability of the individual and to that of the Race. Part 1. Homotyposis in the vegetable Kingdom » ; 1911, *The Grammar of Science* ; Astaurov 1974 [1930], « Study of hereditary deviations from bilaterla symmetry, associated with variation of metamerie structures » ; Ludwig, 1932, *Das Rechts-Links Problem im Tierreich und beim Menschen*. Ces premières études s'insèrent dans les domaines de l'anatomie comparative, de l'embryologie et de la systématique, et parfois touchent à des questions d'écologie et d'adaptation. Pour une bibliographie complète sur l'étude de l'instabilité développementale (ou bruit développemental) au cours du siècle dernier, voir Markow (ed.), 1994, *International conference on developmental instability: its origin and evolutionary implications* ; Polak (ed.) 2003, *Developmental Instability. Causes and Consequences* ; Hall & Hallgrimsson (eds.), 2005, *Variation. A Central Concept in Biology*.

¹¹ Un cas exemplaire est représenté par les jumeaux homozygotes qui, même s'ils sont élevés dans un même environnement physique, social, culturel, etc., contrôlé, homogène et constant dans le temps, ne sont pas phénotypiquement identiques.

l'impact des perturbations dans le développement de chaque organisme particulièrement manifeste au niveau du phénotype.

Une grande partie de ces recherches ont porté sur ce qu'on appelle, depuis le début du XX^e siècle, les « asymétries fluctuantes »¹². Dans ce cadre, les biologistes essaient de comprendre l'origine des asymétries entre les parties gauche et droite d'un même organisme ou dans un échantillon d'organismes génétiquement identiques caractérisés par un plan d'organisation à symétrie bilatérale et situé(s) dans un environnement homogène et constant dans le temps. Ils présupposent que, comme les parties droite et gauche d'un même organisme partagent le même environnement et le même matériel génétique, les effets des changements environnementaux sur les caractéristiques morphologiques des deux parties du corps sont identiques. De même, ils supposent aussi que, si des organismes appartiennent à une même population isogénique et font l'expérience d'un même environnement, ils ne diffèrent pas du point de vue morphologique. Or, contrairement à ces présupposés, les biologistes observent depuis longtemps que, même dans un environnement homogène et constant au cours du temps, la morphologie du corps d'un organisme bilatéral peut être caractérisée par des asymétries entre les parties gauche et droite ; de même, des organismes génétiquement identiques peuvent avoir des traits morphologiques différents. Par exemple, de nombreuses études ont porté sur les asymétries dans le nombre de poils entre les côtés gauche et droit du corps de la *Drosophile*¹³ et dans les motifs des ailes gauche et droite des papillons¹⁴. Déjà autour des années 1950 et 1960, les biologistes voyaient dans ces asymétries, dites fluctuantes car petites et distribuées de manière aléatoire, à la fois une manifestation et une mesure de l'instabilité développementale (ou bruit développemental), sans pourtant en connaître le mécanisme causal sous-jacent¹⁵.

¹² Pour les références à ce sujet, nous renvoyons à la note 10 de ce chapitre. La littérature contemporaine sur les asymétries fluctuantes s'inspire des recherches datant de la moitié du siècle dernier et a connu un essor au cours des quinze dernières années. Pour une bibliographie détaillée des études sur les asymétries fluctuantes et, plus généralement, de l'instabilité développementale, voir Palmer & Strobeck, 1986, « Fluctuating asymmetry: measurement, analysis, patterns » ; Palmer, 1994, « Fluctuating Asymmetry Analyses: A Primer » ; 1996, « Waltzing with asymmetry » ; Møller & Swaddle, 1997, *Asymmetry, Developmental Stability, and Evolution* ; Markow (ed.), 1994, *op. cit.* ; Polak (ed.), 2003, *op. cit.* ; Hall & Halgrimsson (eds.), 2005, *op. cit.*

¹³ Pour des références bibliographiques à ce sujet, voir Soulé, 1982, *op. cit.* ; Wright, 1952, « The genetic of quantitative variability » ; Thoday, 1956, « Balance, heterozygosity, and developmental stability » ; Latter, 1964, « Selection for a threshold character in *Drosophila*. I. An analysis of the phenotypic variance of the underlying scale » ; Reeve & Robertson, 1953, *op. cit.* ; 1954, *op. cit.*

¹⁴ Nous renvoyons encore une fois à Soulé, 1982, *op. cit.* ; Mason *et al.*, 1967, « The population biology of the butterfly, *Euphydryas editha*. V. Character clusters and asymmetry ».

¹⁵ Ce n'est que depuis les trente dernières années que les biologistes ont commencé à comprendre, même si ce n'est que de manière partielle, les causes des asymétries fluctuantes. Voir Klingenberg, 2005, « Developmental Constraints, Modules, and Evolvability », *In* Hallgrimsson & Hall (eds.), *op. cit.*, p. 219-247.

Jusqu'à la fin des années 1970, voire jusqu'aux années 1990, les expressions « bruit développemental » et « instabilité développementale » ne sont pas associées à une notion positivement définie, caractérisée par des contours nets, et désignant un phénomène biologique précis. Au contraire, elles sont utilisées pour donner un nom à des phénomènes au visage indéfini, à des perturbations considérées comme aléatoires et imprédictibles¹⁶ qui dévient le cours normal du développement des organismes et se manifestent, sans pourtant montrer leur nature, dans la variation phénotypique qu'elles produisent.

Tâchons d'attribuer une signification précise à la notion de « bruit développemental » (ou d'« instabilité développementale ») qui a caractérisé la biologie pendant une grande partie du siècle dernier. Le bruit développemental désigne de petites perturbations se produisant au cours du développement, dont l'origine est inconnue, dont la distribution est aléatoire, qui sont supposées être indépendantes du point de vue temporel et relativement aux autres sources de la variation, qui sont donc difficilement prédictibles, et qui sont susceptibles de bouleverser le cours normal du développement. L'étiologie de ces perturbations est restée mystérieuse pendant très longtemps et, en dépit des avancées récentes à ce sujet, elle est encore imprécise et partiellement inconnue. En revanche, les biologistes affirment depuis longtemps que ces perturbations sont des sources causales potentielles de la variation phénotypique au sein des populations naturelles et qu'elles peuvent donc jouer un rôle important dans leur évolution¹⁷. Tout au long de la deuxième moitié du XX^e siècle, l'étude du bruit développemental mobilisera beaucoup d'efforts et nombreux seront les biologistes à vouloir mieux comprendre ce que West-Eberhardt a appelé cette « face noire de la variation phénotypique »¹⁸.

Autour des années 1950, l'hypothèse selon laquelle le bruit développemental trouve son origine première dans les fluctuations moléculaires du milieu intracellulaire prend

¹⁶ Pearson (1901, « Mathematical contributions to the theory of Evolution. IX. On the Principle of Homotyposis and its relation to heredity, to the variability of the individual and to that of the Race. Part 1. Homotyposis in the vegetable Kingdom » ; 1911, *The Grammar of Science*) et Astaurov (1974 [1930], « Study of hereditary deviations from bilaterla symmetry, associated with variation of metameric structures ») parlaient déjà de variation stochastique pour désigner «une variation indépendante, qui ne peut se réduire ni aux différences génétiques ni à l'effet direct de l'environnement : « nous pouvons mesurer cette variation, nous pouvons la décrire quantitativement, mais sa cause reste un mystère de la même manière que la cause de la gravitation », Polak (ed.), 2003, *Developmental Instability. Causes and Consequences*, p. 343, (notre traduction), « *We can measure this variation, we can describe it quantitatively, but its cause stays a mystery like the casue of gravitation (Pearson 1911, p. 530).* »).

¹⁷ En biologie, la notion de « bruit développemental » n'est pas connotée que de manière négative. Les biologistes lui attribuent aussi une valeur positive car les perturbations qu'elle désigne sont des sources potentielles de la variation phénotypique au sein des populations naturelles. En d'autres termes, le bruit développemental est un phénomène pouvant garantir l'une des trois conditions pour qu'il puisse y avoir évolution par sélection naturelle (Lewontin, 1970, « The Units of Selection ») – notamment, la présence de la variation au niveau populationnel – et pouvant donc jouer un rôle positif dans le processus évolutif.

¹⁸ West-Eberhardt, 2003, Forword, *In* Polak (ed.), *op. cit.*

forme¹⁹. La fin des années 1970 est marquée par la publication des premiers modèles stochastiques²⁰ détaillés des processus moléculaires et cellulaires impliqués dans le développement (en particulier, l'expression des gènes)²¹ : c'est le début d'un courant de recherche utilisant de manière systématique des modèles mathématiques stochastiques afin de rendre compte de la présence du bruit à l'intérieur de la cellule et de son impact au niveau phénotypique. À partir de la fin des années 1990, la modélisation stochastique de ces processus connaît des avancées importantes²² : de nos jours, les modèles du bruit développemental, en particulier de celui affectant le processus d'expression des gènes, nous fournissent des descriptions et des explications plus précises de l'origine physico-chimique de ces fluctuations, de leurs conséquences sur le développement et des moyens cellulaires pour les contrôler, les réduire ou les exploiter²³.

C'est sur la base de ces avancées que, durant les quinze dernières années, l'expression « bruit développemental » a enfin trouvé un sens assez précis, en particulier grâce à la contribution des biophysiciens et des biologistes moléculaires à l'étude de ce phénomène biologique. Elle désigne toute perturbation aléatoire ou stochastique ne concernant ni la séquence du matériel génétique, ni l'environnement externe aux organismes individuels, mais qui a lieu à différentes échelles à l'intérieur des organismes mêmes (aux niveaux moléculaire, cellulaire, des gènes et des réseaux de gènes, du système développemental, organismique). Ces perturbations intra-organismiques peuvent affecter la dynamique et le résultat des

¹⁹ Reeve & Robertson, 1953, « Analysis of environmental variability in quantitative inheritance ».

²⁰ Un modèle stochastique est un modèle qui, à partir d'un même ensemble de conditions initiales, prédit plusieurs résultats possibles selon une certaine loi de probabilité. Nous parlons aussi de modèle stochastique dans un sens plus large et conformément à l'utilisation de cette expression dans les études du bruit dans l'expression des gènes pour désigner un modèle qui permet de prédire un seul résultat final mais probabiliste : ce dernier correspond au résultat associé à la valeur moyenne d'une distribution de probabilité.

²¹ Rigney & Schieve, 1977, « Stochastic model of linear, continuous protein-synthesis in bacterial populations » ; Rigney, 1979, « Stochastic model of constitutive protein levels in growing and dividing bacterial cells » ; Berg, 1978, « A model for statistical fluctuations of protein numbers in a microbial-population ». Voir Paulsson, 2005, « Models of stochastic gene expression ».

²² McAdams & Arkin, 1997, « Stochastic mechanisms in gene expression ».

²³ Samoilov *et al*, 2006, « From Fluctuations to Phenotypes: The Physiology of Noise » ; Mahesri & O'Shea, 2007, « Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression ». Selon Klingenberg (2005, « Developmental Constraints, Modules, and Evolvability » *In* Hallgrímsson & Hall (eds.), *Variation. A Central Concept in Biology*, p. 14), l'étude du bruit dans les processus moléculaires et cellulaires du développement (par exemple, l'expression génétique et le transport macromoléculaire) a apporté des informations plus spécifiques quant à l'origine du bruit développemental et des avancées importantes dans son explication. En effet, l'étude des perturbations affectant le développement s'est toujours limitée à en fournir des explications phénoménologiques, c'est-à-dire de simples descriptions de la variation aléatoire au niveau du résultat du développement et donc des traits phénotypiques des organismes concernés. Au contraire, les études réalisées depuis la fin des années 1990 s'intéressent aux causes du bruit développemental et fournissent des explications mécanistiques de la variation phénotypique qu'il produit.

processus physico-chimiques impliqués dans le développement et, par conséquent, elles peuvent affecter le résultat du développement lui-même.²⁴

8.2.2. L'étude du bruit au cœur de la cellule

L'étude de l'origine et des conséquences du bruit dans l'expression des gènes caractérise un domaine de recherche théorique et expérimentale relativement récent en biologie qui, à partir de la fin des années 1990, a connu un essor au niveau des publications²⁵. Les disciplines biologiques les plus impliquées dans cette étude sont la génétique moléculaire du développement et la microbiologie. Ces dernières s'appuient sur des outils théoriques et expérimentaux empruntés à d'autres domaines de recherche au croisement entre la biologie et des disciplines comme la biophysique, la biologie statistique et la bioinformatique.

Considérons d'abord la génétique moléculaire du développement. Elle entend étudier les propriétés physico-chimiques, structurelles et fonctionnelles des mécanismes qui gouvernent les étapes fondamentales du développement. Plus précisément, elle examine la manière dont le processus développemental est contrôlé et régulé au niveau génétique, en particulier chez les organismes multicellulaires. En effet, on ne peut parler de développement au sens strict, au sens d'un processus de transformation progressif qui part d'une cellule unique pour aboutir à un organisme complet (l'ontogénie ou ontogenèse), que relativement aux organismes multicellulaires au sein desquels le processus de différenciation (c'est-à-dire de spécialisation) des cellules a lieu. Cependant, on peut aussi parler de développement au sens large relativement aux organismes unicellulaires qui peuvent aussi être caractérisés, par exemple, par une phase de croissance du volume cellulaire et par un processus de vieillissement (c'est le cas des bactéries).

Les généticiens moléculaires du développement ont étudié l'origine du bruit affectant l'expression des gènes, son impact sur le processus de différenciation des cellules et, plus généralement, sur le développement et sur les caractéristiques phénotypiques des organismes concernés. Un cas d'étude paradigmatique est le développement de l'oeil chez les métazoaires, et notamment le développement du système de vision en couleurs chez la

²⁴ Hallgrímsson & Hall (eds.), 2005, *op. cit.*, Chapitre 20 ; Palmer, 1996, « Waltzing with asymmetry ».

²⁵ McAdam & Arkin (1997, *op. cit.*) ont publié un article pionnier à ce sujet. Pour une mise à jour claire et synthétique des études théoriques et expérimentales sur le bruit dans l'expression des gènes et, plus généralement, dans le développement, nous renvoyons à Mahesri & O'Shea, 2007, *op.cit.* ; Losick & Desplan, 2008, « Stochasticity and cell fate ».

Drosophile : il a été montré que la variation (ou bruit) dans le niveau d'expression du facteur de transcription *Spineless* joue un rôle causal dans la différenciation des cellules de l'oeil de la *Drosophile*²⁶. D'autres recherches sur le bruit dans le développement des organismes pluricellulaires ont analysé, par exemple, l'origine et l'impact du bruit dans l'expression des gènes impliqués dans la production des récepteurs olfactifs chez les mammifères²⁷ et dans la production des cônes verts et rouges de la rétine chez l'homme²⁸.

Un autre exemple de ce genre de recherches porte sur l'impact que la variation dans l'expression des gènes a au niveau des comportements chez les animaux. Les généticiens moléculaires du développement ont découvert que le taux d'expression d'un récepteur de la vasopressine – une hormone qui, entre autres, se trouve impliqué dans le contrôle de la rétention d'eau et dans l'adaptation au stress – joue un rôle dans la détermination du comportement sexuel des mâles de campagnols. Selon le niveau d'abondance de ce récepteur, un individu mâle peut adopter deux comportements sexuels différents : soit il s'occupe de sa femelle et de ses petits, soit il se désintéresse complètement de sa progéniture et s'accouple avec plusieurs femelles²⁹.

La deuxième discipline à s'être particulièrement impliquée dans l'étude du bruit ces vingt dernières années, c'est la microbiologie (ou l'étude des organismes vivants microscopiques, procaryotes et eucaryotes, et des virus). En effet, de nombreuses recherches sur l'origine et l'impact du bruit au niveau moléculaire et sur son influence au niveau phénotypique portent sur l'expression des gènes chez les micro-organismes unicellulaires, les procaryotes (notamment chez des bactéries diverses), les eucaryotes (par exemple, chez la levure), et aussi les virus (par exemple, le virus bactérien lambda). En ce qui concerne les bactéries, les biologistes ont notamment découvert que chez *Bacillus subtilis*, le bruit dans le niveau d'expression de la protéine ComK était à l'origine d'un phénomène de bi-stabilité phénotypique, qui consiste dans un basculement (*switch*) stochastique de l'état de compétence pour la transformation (c'est-à-dire, l'intégration d'un fragment d'ADN étranger³⁰) vers l'état

²⁶ Chou *et al*, 1996, « Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells » ; Mikeladze-Dvali *et al*, 2005, « The growth regulators *warts/lats* and *melted* interact in a bistable loop to specify opposite fates in *Drosophila* R8 photoreceptors ».

²⁷ Mombaerts, 2004, « Odorant receptor gene choice in olfactory sensory neurons: the one receptor-one neuron hypothesis revisited ».

²⁸ Nathans, 1999, « The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments ».

²⁹ Voir Morange, 2005, *Les secrets du vivant : Contre la pensée unique en biologie*, p. 113 : Lim *et al*, 2004, « Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene » ; Hammock & Young, 2005, « Microsatellite instability generates diversity in brain and sociobehavioral traits ».

³⁰ Le phénomène de la transformation génétique fut découvert en 1928 par le médecin anglais Griffith.

de non compétence, et vice versa³¹. Concernant les eucaryotes unicellulaires, le bruit dans l'expression des gènes chez la levure *Saccharomyces cerevisiae*, en particulier la perturbation de l'efficacité des processus de transcription et de traduction, a été l'objet de plusieurs études expérimentales et de modélisations mathématiques³². Enfin, le bruit dans l'expression des gènes impliqués dans la production de certains facteurs de transcription chez le virus bactérien lambda est à l'origine du passage aléatoire de l'état lytique (dans lequel le virus se réplique de façon intensive à l'intérieur de l'hôte, l'amène à la mort et le perce pour aller infecter d'autres bactéries) à l'état lysogénique (dans lequel le chromosome du virus s'intègre dans le matériel génétique de l'hôte bactérien et reste à l'état dormant) chez un même virus individuel au cours du temps. Ce même bruit explique aussi la présence simultanée de virus en état lytique et en état lysogénique au sein de populations isogéniques de virus lambda³³.

Toutes ces études ont mis en évidence que l'une des premières sources de la variation phénotypique, en particulier des différences entre des individus génétiquement identiques situés dans un même environnement homogène et constant, c'est le bruit au niveau de l'expression des gènes. Cette dernière est un processus commun à tout organisme vivant, unicellulaire et pluricellulaire, procaryote et eucaryote : il représente donc une cible privilégiée dans le cadre d'une analyse des causes non génétiques et non environnementales de variation au sein des populations naturelles. C'est là une des raisons qui nous ont amenée à choisir le bruit dans l'expression génétique afin de procéder à une deuxième analyse de la notion de hasard dans la caractérisation des sources de la variation biologique.

8.3. Une nouvelle approche pour l'étude des processus intracellulaires

La focalisation des recherches en génétique moléculaire du développement et en microbiologie sur l'étude du bruit dans l'expression génétique est en grande partie liée à l'intérêt que les physiciens ont commencé à montrer pour les phénomènes biologiques de niveau moléculaire pendant la deuxième moitié du XX^e siècle. L'influence des physiciens et de leurs outils d'analyse sur l'orientation et le développement de la recherche en biologie n'est pas une nouveauté : il suffit de penser aux idées précieuses qu'ils ont apportées lors de la découverte de la structure de la molécule d'ADN et, plus généralement, de leur contribution

³¹ Maamar *et al*, 2007, « Noise in gene expression determines cell fate in *Bacillus subtilis* ».

³² Blake *et al*, 2003, « Noise in eukaryotic gene expression » ; Raser & O'Shea, 2004, « Control of Stochasticity in Eukaryotic Gene Expression ».

³³ Ptashne, 2007, « On the use of the word 'epigenetic' » ; 2003, *A Genetic Switch: Phage Lambda Revisited*.

à la naissance de la biologie moléculaire³⁴. En ce qui concerne l'étude du bruit et de son impact sur les processus moléculaires impliqués dans le développement, le rôle des physiciens a consisté en une critique des modèles déterministes³⁵ utilisés en biologie pour représenter, entre autres, l'expression des gènes. Leur but était de mettre en évidence la présence et même l'impact des fluctuations stochastiques qui peuvent affecter les processus biologiques au niveau moléculaire. Les biologistes étaient conscients, depuis longtemps (au moins depuis la fin de la deuxième moitié du XX^e siècle) de la présence inévitable de fluctuations thermiques à l'intérieur de la cellule et, par conséquent, du caractère stochastique des réactions chimiques impliquant peu de molécules³⁶. Cependant, ils ne les avaient prises en compte que très rarement dans leurs descriptions des processus intracellulaires du développement : ils ne construisaient et n'utilisaient que des modèles déterministes et qualitatifs provenant de la cinétique chimique³⁷ classique et décrivant de manière mécaniste et déterministe les types d'interactions entre espèces moléculaires.

L'apport des physiciens n'aurait pas eu de conséquences importantes sur l'étude des processus intracellulaires impliqués dans le développement, en particulier de l'expression de gènes, s'il n'avait pas été accompagné par des innovations technologiques permettant d'observer l'impact du bruit sur les résultats de ces processus. En particulier, c'est l'utilisation de marqueurs biologiques fluorescents découverts autour des années 1960 et 1970 (par exemple, différents variants de protéines fluorescentes) et de systèmes très sensibles de détection des signaux lumineux (par exemple, la cytométrie en flux et la microscopie à fluorescence³⁸) qui a permis le développement et ensuite favorisé l'essor des études sur le bruit dans l'expression des gènes.

Quelles sont les nouveautés majeures apportées par l'introduction de l'analyse du bruit dans la modélisation d'une variété de processus intracellulaires et, en particulier, de l'expression des gènes ? D'abord, on peut remarquer un changement d'attitude de la part des biologistes concernant la représentation qu'ils se font d'une population d'organismes génétiquement identiques situés dans un même environnement homogène et constant dans le

³⁴ Voir Morange, 2003 [1994], *Histoire de la biologie moléculaire*.

³⁵ Un modèle déterministe est un modèle qui, à partir d'un même ensemble de conditions initiales, décrit de manière univoque l'évolution d'un système donné ou, autrement dit, ne prédit qu'un seul résultat final.

³⁶ Le traitement stochastique des réactions chimiques fut introduit par Kramers en 1940 (voir Turner *et al*, 2004, « Stochastic approaches for modelling in vivo reactions »).

³⁷ La cinétique chimique est l'étude de l'évolution au cours du temps des réactions chimiques.

³⁸ La cytométrie en flux est une technique qui permet d'étudier des particules isolées (molécules, cellules, etc.) entraînées par un flux liquide en les faisant défiler à grande vitesse dans le faisceau d'un laser. La microscopie à fluorescence est une technique de microscopie optique permettant d'observer et d'étudier des objets physiques et biologiques divers grâce aux phénomènes de fluorescence.

temps. Dans l'état actuel de la recherche sur le bruit, les biologistes admettent que des organismes isogéniques peuvent différer phénotypiquement sans qu'aucune modification de leur matériel génétique ou de leur environnement externe ne se produise, mais simplement à cause des fluctuations diverses affectant l'expression de leurs gènes et, plus généralement, se produisant au cours de leur développement. Cette représentation contraste avec celle qui avait cours à la fin des années 1980, où les biologistes avaient tendance à considérer ce genre de situation avec un regard assez différent. Monod représente un cas paradigmatique de cette attitude, lui qui expliquait « volontiers que son intérêt pour la culture des *Escherichia coli* en milieu liquide provient de ce que ce modèle expérimental est proche à ses yeux de l'étude d'un gaz parfait. Les différences morphologiques entre les bactéries, leur aspect, leur allure, leur mobilité – tout cela passe au second plan »³⁹. Un tel regard négligeait le rôle souvent décisif, à la fois pour la survie individuelle et pour l'évolution au niveau populationnel, de la variation phénotypique due à des sources non génétiques, le bruit intracellulaire en étant une.

La deuxième nouveauté apportée par l'étude du bruit dans l'expression génétique concerne les méthodes théoriques que les biologistes utilisent pour décrire ce processus et pour essayer d'en prédire le résultat final. Sous l'impulsion des physiciens, les biologistes ont commencé à préférer aux modèles déterministes provenant de la cinétique chimique classique (les équations du taux de réaction⁴⁰) des modèles et des simulations stochastiques provenant de la cinétique chimique stochastique et fondés sur les mathématiques des processus stochastiques⁴¹. Cette avancée dans la modélisation mathématique leur a permis de rendre compte du fait que l'expression génétique consiste en une série de réactions biochimiques discrètes et très sensibles aux fluctuations dues à l'agitation thermique du milieu intracellulaire car elles engagent un petit nombre de molécules⁴².

Enfin, le troisième apport de ces recherches sur le bruit concerne également les modèles employés dans la description de l'expression des gènes. Ces études marquent le passage de l'utilisation de modèles qualitatifs (et déterministes), propres à la biologie moléculaire, à l'utilisation de modèles quantitatifs (et stochastiques), propres à la physique. L'intérêt de ces derniers ne réside pas simplement dans les résultats quantitatifs qu'ils apportent, mais aussi et surtout dans les nouveautés qualitatives qu'ils prédisent au niveau des résultats. Dans le cas de l'expression génétique, les biologistes ont montré que la variation

³⁹ Debré, 1996, *Jacques Monod*, p. 188.

⁴⁰ Nous y reviendrons plus tard.

⁴¹ Les équations maîtresses, de Fokker-Planck, de Langevin, sur lesquelles nous reviendrons plus tard ; des simulations informatiques utilisant les méthodes de Monte Carlo.

⁴² On parle de « l'effet du petit nombre ». Nous y reviendrons plus tard.

quantitative dans le niveau de transcription de l'ADN en ARN et de traduction de l'ARN en protéines peut être à l'origine d'une variation qualitative, au niveau phénotypique, au sein d'une population isogénique ou chez un même individu au cours du temps. En effet, le niveau d'expression des facteurs protéiques à l'intérieur d'une cellule donnée peut être décisif pour le destin de cette cellule, c'est-à-dire dans la détermination de sa trajectoire particulière de différenciation. Plus généralement, les perturbations (ou bruit) affectant des réactions biochimiques diverses semblent jouer un rôle très important au niveau de l'évolution des systèmes biochimiques, en particulier lorsque plusieurs trajectoires sont possibles avec une certaine probabilité. La comparaison de modèles déterministes et stochastiques de l'évolution d'un même système biochimique a montré que, par exemple, lorsqu'un modèle déterministe ne prévoit qu'un seul état stable dénué d'oscillations, un modèle stochastique peut prévoir, au sein du même système biochimique, soit des bi- ou des multi-stabilités, soit des bifurcations, soit des oscillations régulières⁴³. Le bruit semble donc avoir un pouvoir presque créateur de nouvelles stabilités, de bifurcations, d'oscillation, dans les systèmes biochimiques qu'il affecte.

8.4. L'origine et les conséquences du bruit dans l'expression génétique

Le processus⁴⁴ d'expression génétique est un mécanisme⁴⁵ intracellulaire dont le résultat final est la production d'un nombre variable de protéines diverses à partir d'une séquence d'ADN donnée. Il a lieu à l'intérieur de la cellule et consiste en une série d'interactions physico-chimiques spécifiques (des collisions, des associations, des réactions) entre un certain nombre d'espèces moléculaires (l'ADN, les ARN messagers, les ARN de transfert, les ARN polymérases, les ribosomes, des facteurs de transcription et de traduction, des répresseurs et des activateurs, des protéines et des enzymes diverses). Ces molécules et leurs interactions sont responsables du déroulement du processus d'expression génétique au cours du temps, des conditions initiales jusqu'au résultat final.

⁴³ Voir Samoilov *et al*, 2005, « Stochastic and signaling in enzymatic futile cycles through noise-induced bistability with oscillations ».

⁴⁴ Le terme « processus » désigne de manière générale une suite temporelle d'états liés causalement.

⁴⁵ Le terme de « mécanisme » désigne ici une notion plus spécifique que celle de processus désignant la manière selon laquelle ce dernier se produit. Nous renvoyons à la définition du concept de mécanisme en termes d'entités et d'activités par Machamer *et al* (2000, « Thinking About Mechanisms », p. 3, 6).

Les étapes principales du processus d'expression génétique sont au nombre de deux : 1) la synthèse d'une quantité variable de molécules d'ARN sur la base d'une molécule d'ADN simple brin (ou transcription) ; 2) la synthèse d'une quantité variable de polymères d'acides aminés, qui sont la structure de base des protéines, à partir des molécules d'ARN messenger simple brin (ou traduction). La transcription et la traduction comprennent à leur tour plusieurs étapes intermédiaires consistant en des interactions physico-chimiques spécifiques entre deux espèces moléculaires ou davantage (par exemple, l'association des ARN polymérases à l'ADN simple brin au niveau du promoteur ; les interactions des facteurs de transcription, inhibiteurs et activateurs, entre eux et au niveau de la séquence d'ADN ; l'appariement entre les codons des ARN messagers et les anticodons des ARN de transfert ; les associations des ARN de transfert avec des sous-unités spécifiques des ribosomes ; la polymérisation des séquences d'acides aminés constituant des protéines diverses, etc.).⁴⁶

Des événements microscopiques divers peuvent se produire et engendrer des fluctuations stochastiques et imprédictibles dans toute étape de la transcription et de la traduction. De cette manière, ils peuvent donc avoir un impact sur le résultat de l'expression des gènes, notamment sur le nombre de protéines produites⁴⁷. Par exemple, des erreurs peuvent se produire au cours de la transcription d'une molécule d'ADN en une molécule d'ARN et de la traduction d'une molécule d'ARN en protéines.⁴⁸ À cela s'ajoute le fait que la première étape – la transcription – est, en utilisant une expression répandue parmi les biologistes, un « processus intrinsèquement probabiliste ». En effet, il a été montré⁴⁹ que les gènes sont activés et réprimés par des événements aléatoires d'association-dissociation des facteurs de transcription. La production des molécules d'ARN messagers est donc régulée par des mécanismes probabilistes de *switch on-off* (activation-répression) qui dépendent de la concentration, de la localisation et de l'état des espèces moléculaires diverses présentes à l'intérieur de la cellule. Toutes ces variables peuvent changer de manière aléatoire et imprédictible dans le temps car les molécules impliquées dans l'expression des gènes sont peu nombreuses, donc leur diffusion peut engendrer des retards temporels dans les réactions chimiques impliquées dans le processus de transcription et, plus généralement, dans toute étape du processus d'expression génétique.

⁴⁶ Pour une introduction au processus d'expression génétique, voir Lewin, 2004, *Genes VIII* ; Alberts *et al*, 2002 [1983], *Molecular Biology of the Cell*.

⁴⁷ Voir Hall & Hallgrimsson (eds.), 2005, *Variation. A Central Concept in Biology*, ch. 10.

⁴⁸ Il est intéressant de remarquer que le taux d'erreur de ces deux étapes du processus d'expression génétique est trois fois plus grand que celui de la réplication de l'ADN. Voir Alberts *et al*, 2002[1983], *op. cit.*

⁴⁹ Voir Maheshri & O'Shea 2007, « Living with Noisy Genes: How cells Function Reliably with Inherent Variability in Gene Expression ».

Ces mécanismes probabilistes de *switch on-off* sont à l'origine de la production de quantités variables d'ARN messagers sur des intervalles de temps distribués de manière aléatoire (on parle d'explosions (*bursts*) aléatoires dans la production d'ARN messagers, distribuées dans le temps de manière géométrique⁵⁰). Cette variation aléatoire au niveau du processus de transcription peut avoir un impact sur la synthèse des protéines (la traduction), plus précisément, sur l'abondance de protéines produites dans le temps (on parle aussi d'explosions (*bursts*) aléatoires dans la production des protéines, distribuées dans le temps de manière exponentielle⁵¹). De plus, en se propageant par cascades moléculaires, elle peut aussi affecter l'efficacité de tout autre processus impliqué dans le développement (au sens large), du niveau moléculaire jusqu'au niveau organismique.

Les biologistes concentrent leur attention sur la variation phénotypique, d'une part, qu'on peut observer au sein de populations clonales de cellules situées dans un même environnement et, d'autre part, pouvant se produire au cours du temps chez une même cellule dans un environnement homogène et constant. En effet, la présence à la fois de perturbations génétiques et non génétiques externes (environnementales) et internes rendrait difficile la tâche de déterminer quelle est la contribution relative de chaque source causale dans la variation phénotypique. C'est pourquoi les biologistes ont conçu des situations expérimentales propres à éliminer les perturbations génétiques et environnementales externes et qui rendent donc l'identification des effets du bruit (des perturbations internes ou intracellulaires) ainsi que sa quantification plus aisées.

Mais avant de décrire ces situations expérimentales et les méthodes utilisées dans leur analyse, il convient de clarifier soigneusement la manière dont les biologistes définissent le bruit dans l'expression des gènes et l'origine de ces fluctuations intracellulaires.

Le terme « bruit » désigne la variation aléatoire autour de la valeur moyenne du processus d'expression génétique, c'est-à-dire autour de son résultat moyen. Ce dernier peut être quantifié à différents niveaux, du niveau des protéines produites, et même relativement à la production des molécules d'ARN, jusqu'au niveau des traits phénotypiques. Dans le cadre des études sur lesquelles nous concentrons notre attention, les biologistes quantifient le bruit en termes d'abondance dans la production de protéines ; par conséquent, ils le définissent comme la variation aléatoire autour du nombre moyen de protéines produites, soit dans une population de cellules génétiquement identiques situées dans le même environnement à un

⁵⁰ Voir le Glossaire.

⁵¹ *Id.*

moment donné dans le temps, soit dans une cellule située dans un environnement homogène et constant au cours du temps.

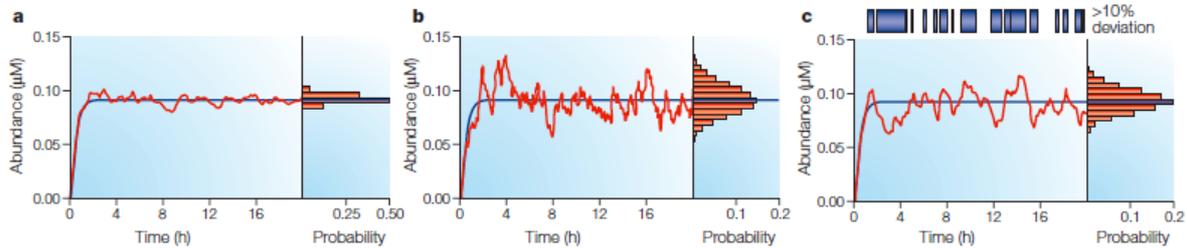


Figure 7 : Trois séries temporelles de la concentration de protéines au sein d’une cellule au cours du temps, l’une générée par simulation déterministe (courbe bleue), l’autre par simulation stochastique (courbe rouge). Les histogrammes montrent la probabilité qu’une cellule ait en son sein une certaine concentration intracellulaire de protéines. Figure tirée de Kærn *et al*, 2005, « Stochasticity in gene expression: from theories to phenotypes ».

En termes mathématiques, le bruit est quantifié par ce qu’on appelle le « coefficient de variation » (η), c’est-à-dire l’écart-type (σ) divisé par la moyenne (μ) : ces deux valeurs sont estimées relativement au nombre de protéines produites, soit dans une population à un moment donné dans le temps, soit chez un même individu au cours du temps⁵².

La variation aléatoire autour de l’abondance moyenne de protéines produites (ou « bruit protéique ») est due à ce qu’on appelle le « bruit biochimique », c’est-à-dire le caractère probabiliste des réactions chimiques impliquées dans l’expression des gènes. Plusieurs facteurs, situés à des niveaux physico-chimiques différents, en rendent compte. D’abord, l’expression des gènes se déroule dans un microenvironnement intracellulaire qui n’est pas homogène et bien mélangé : la diffusion des molécules n’est donc pas libre mais contrainte par l’organisation du volume cellulaire (en particulier, du cytoplasme) et limitée par la présence de plusieurs obstacles de nature moléculaire⁵³. Deuxièmement, les espèces moléculaires engagées dans l’expression des gènes

⁵² Le coefficient de variation ($\eta = \sigma/\mu$) est une mesure du bruit et fournit une bonne estimation de la variation dans une population clonale ou chez une même cellule au cours du temps. Les biologistes utilisent aussi d’autres quantités mathématiques afin de mesurer le bruit : par exemple, la variance normalisée ($\eta^2 = \sigma^2/\mu^2$), ou le carré du bruit, qui se révèle très utile en présence de plusieurs sources de bruit afin de les distinguer et de quantifier leur contribution relative ; l’intensité du bruit (σ^2/μ), souvent utilisée dans la comparaison des données expérimentales avec les résultats des modèles mathématiques stochastiques ; le temps d’autocorrélation du bruit, qui correspond à l’intervalle caractéristique pendant lequel une cellule individuelle maintient un même niveau d’expression génétique (voir Raser & O’Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control », Supporting Online Material).

⁵³ Cependant, la plupart des modèles et des simulations stochastiques du processus d’expression génétique ne prennent pas en compte ces caractéristiques du microenvironnement intracellulaire et, au contraire, supposent qu’il soit parfaitement homogène et bien mélangé. Voir Turner *et al*, 2004, « Stochastic approaches for modelling in vivo reactions ».

(les molécules d'ADN, d'ARN, un certain nombre de protéines et d'enzymes diverses) ressentent l'influence de l'agitation thermique du milieu intracellulaire. Par conséquent, leurs déplacements suivent des trajectoires erratiques et imprédictibles que l'on modélise, de manière continue ou discrète – mais de toute façon idéalisée – comme un mouvement brownien⁵⁴. De plus, le fait qu'elles soient présentes en petit nombre à l'intérieur de la cellule rend les réactions biochimiques très sensibles aux effets des fluctuations thermiques sur la concentration, la localisation, les collisions et l'état des molécules⁵⁵. Enfin, l'expression des gènes est susceptible de ressentir les effets des fluctuations quantiques qui sont intrinsèquement probabilistes et qui se produisent, en particulier, au niveau des liaisons chimiques entre molécules.

8.5. Les études expérimentales du bruit dans l'expression des gènes

Considérons plus en détail deux méthodes expérimentales que les biologistes ont privilégiées afin de comprendre quelles sont les sources du bruit dans le processus d'expression génétique : la méthode « à deux rapporteurs » et la méthode « à un rapporteur », un rapporteur étant un promoteur à partir duquel on observe l'expression d'une seule ou de plusieurs protéines⁵⁶.

La première méthode a été utilisée pour l'étude du bruit dans l'expression des gènes chez les bactéries *Escherichia coli* et *Bacillus subtilis*⁵⁷, ainsi que chez l'organisme eucaryote *Saccharomyces cerevisiae*⁵⁸. Elle consiste dans l'observation, par cytométrie en flux ou par microscopie à fluorescence, du niveau d'expression de deux protéines fluorescentes distinctes,

⁵⁴ Processus stochastique markovien à accroissements disjoints indépendants suivant une loi normale de moyenne nulle (une densité de probabilité en forme de cloche centrée sur zéro). Il peut être décrit de manière continue comme un processus de Wiener, qui est un processus stochastique continu markovien, ou comme une marche aléatoire par sauts discrets, qui est un processus stochastique discret markovien totalement non corrélé dans le temps (c'est-à-dire, dont l'état présent est complètement indépendant des états passés).

⁵⁵ On parle de « l'effet du petit nombre » : une diminution dans l'abondance de molécules (N) provoque une augmentation du bruit selon la relation suivante : $1/\sqrt{N}$ (le coefficient de variation $\eta \sim 1/\sqrt{N}$ et la variance normalisée $\eta^2 \sim 1/N$).

⁵⁶ Un promoteur est une région de la séquence d'ADN située à proximité d'un gène et sur laquelle se fixent les ARN polymérase, ces dernières étant des complexes enzymatiques indispensables à la transcription. Pour une présentation générale de ces deux méthodes et de leur utilisation expérimentale, nous renvoyons à Raser & O'Shea, 2005, *op. cit.* ; Maheshri & O'Shea, 2007, « Living with Noisy Genes: How cells Function Reliably with Inherent Variability in Gene Expression ».

⁵⁷ Elowitz *et al*, 2002, « Stochastic Gene Expression in a Single Cell » ; Ozbudak *et al*, 2002, « Regulation of noise in the expression of a single gene ».

⁵⁸ Raser & O'Shea, 2004, « Control of Stochasticity in Eukaryotic Gene Expression ».

l'une verte et l'autre jaune, qui sont synthétisées à partir de deux copies d'un même promoteur situées sur le même chromosome.

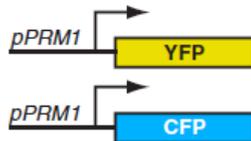


Figure 8 : La méthode « à deux rapporteurs ». Figure tirée de Maheshri & O'Shea, 2007, « Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression ».

Cette méthode permet de quantifier le bruit affectant l'expression génétique, soit dans une même cellule au cours du temps, soit dans plusieurs cellules d'une population isogénique à un instant donné. Autrement dit, elle permet d'estimer la variation aléatoire autour du nombre moyen de protéines produites à partir de deux promoteurs identiques et en l'absence de toute autre source (génétique et environnementale) de variation.

La méthode « à un rapporteur », quant à elle, a été utilisée pour l'étude du bruit dans l'expression des gènes particulièrement chez l'organisme eucaryote *Saccharomyces cerevisiae*⁵⁹. Elle consiste dans l'observation, là encore par cytométrie en flux ou par microscopie à fluorescence, du niveau d'expression d'une protéine fluorescente synthétisée à partir d'un seul promoteur, soit dans une même cellule au cours du temps, soit dans plusieurs cellules génétiquement identiques.

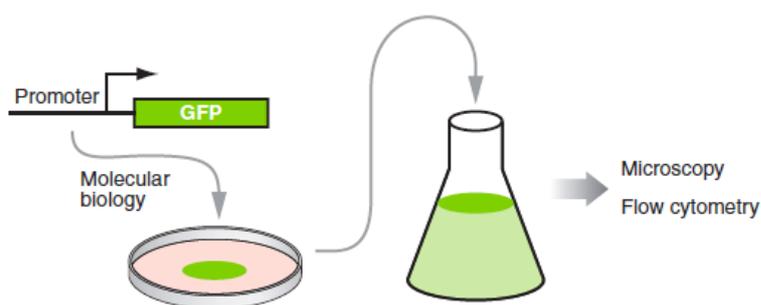


Figure 9 : La méthode « à un rapporteur ». Figure tirée de Maheshri & O'Shea, 2007, *op. cit.*

Les résultats quantitatifs de cette deuxième situation expérimentale ont corroboré les estimations obtenues par la méthode « à deux rapporteurs »⁶⁰.

⁵⁹ Blake *et al.*, 2003, « Noise in eukaryotic gene expression ».

⁶⁰ Nous rappelons que l'étude du bruit dans l'expression génétique n'a pas été réservée aux organismes unicellulaires. Par exemple, ce phénomène stochastique a aussi été l'objet d'études théoriques et expérimentales

La visée de ces méthodes expérimentales est d'identifier les différentes sources de bruit au niveau de l'expression des gènes. En effet, à l'issue de ces résultats expérimentaux, les biologistes ont distingué deux catégories de bruit qu'ils appellent le « bruit intrinsèque » et le « bruit extrinsèque »⁶¹. Nous pouvons identifier deux manières de les distinguer que nous appelons, l'une « théorique », l'autre « opérationnelle ».

D'abord, d'un point de vue « théorique », le bruit intrinsèque est défini comme la variation dans le niveau d'expression (autrement dit, dans l'abondance des protéines produites) due aux fluctuations stochastiques *inhérentes* au système d'expression génétique considéré, c'est-à-dire concernant la concentration, la localisation, l'état et les interactions des facteurs moléculaires impliqués dans les différentes étapes du processus d'expression génétique. Par exemple, les fluctuations stochastiques dans l'activation-répression des gènes sont considérées comme une source de bruit intrinsèque. Le bruit extrinsèque est en revanche défini comme la variation dans le niveau d'expression due aux fluctuations stochastiques *externes* au système d'expression génétique considéré, c'est-à-dire concernant la concentration, la localisation, l'état et les interactions d'autres facteurs moléculaires qui ne sont pas des composants du processus d'expression génétique mais qui, en étant situés dans le microenvironnement intracellulaire où l'expression des gènes a lieu, peuvent l'influencer. Par exemple, les fluctuations stochastiques dans la concentration des enzymes de transcription (les ARN polymérases) sont considérées comme une source de bruit extrinsèque, alors que leur rôle dans l'expression des gènes est indispensable.⁶²

En second lieu, les biologistes distinguent le bruit intrinsèque et le bruit extrinsèque d'une manière plus « opérationnelle » en utilisant les deux méthodes expérimentales « à deux rapporteurs » et « à un rapporteur ». Dans ce cas, le bruit intrinsèque correspond aux fluctuations stochastiques produisant des différences dans l'expression de deux protéines distinctes à partir de deux promoteurs identiques situés sur le même chromosome dans la même cellule. En effet, comme les deux rapporteurs sont supposés se trouver dans le même microenvironnement intracellulaire, la variation autour de la valeur moyenne de protéines produites ne peut qu'être due aux fluctuations inhérentes à chaque promoteur. Le bruit extrinsèque correspond aux fluctuations du microenvironnement qui affectent de la même manière deux promoteurs identiques situés sur le même chromosome dans la même cellule,

sur les cellules de *Drosophila melanogaster*, de diverses espèces de mammifères, et même sur les cellules humaines. Pour des références précises, voir Raser & O'Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control ».

⁶¹ Raser & O'Shea, 2005, *op. cit.* ; Kærn *et al.*, 2005, « Stochasticity in gene expression: from theories to phenotypes ».

⁶² Voir Raser & O'Shea, 2005, *op. cit.*

mais qui créent des différences entre des copies d'un même promoteur dans des cellules isogéniques distinctes (ou dans une même cellule au cours du temps).

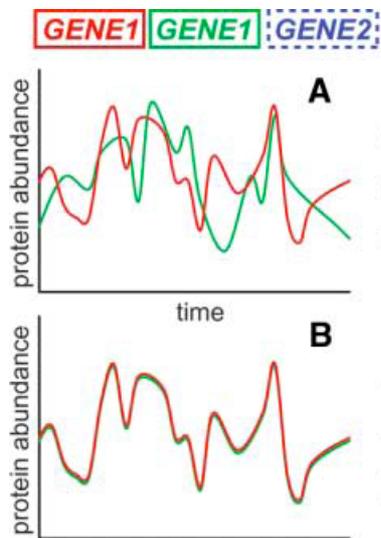


Figure 10 : Représentation graphique des définitions opérationnelles et des caractéristiques du bruit intrinsèque (A) et du bruit extrinsèque (B). Dans la situation représentée ci-dessus, les deux rapporteurs identiques (**GENE1** et **GENE1**) se trouvent dans la même cellule. Figure tirée de Raser & O'Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control ».

Ces deux manières de définir et de distinguer le bruit intrinsèque du bruit extrinsèque présentent néanmoins certaines limites. D'abord, sur le plan de la distinction « théorique », le bruit intrinsèque et le bruit extrinsèque sont deux notions relatives⁶³ : le bruit est considéré comme intrinsèque ou extrinsèque relativement à un système donné. Cela signifie que les mêmes fluctuations stochastiques (ou bruit) peuvent être considérées à la fois comme intrinsèques et extrinsèques relativement à deux systèmes d'expression génétique distincts (par exemple, relativement à l'expression de deux gènes différents). Cette ambivalence est encore plus apparente si l'on tient compte des niveaux d'organisation différents : par exemple, les fluctuations dans la concentration et dans l'activité de métabolites peuvent être dites extrinsèques relativement à l'expression d'un gène, au sens où elles ne font pas partie ni ne dépendent de ce processus, mais elles peuvent être considérées néanmoins comme du bruit intrinsèque relativement au cycle de la cellule dans laquelle l'expression de ce gène a lieu. Les notions de bruit intrinsèque et de bruit extrinsèque sont donc des notions relatives au sens où elles ne désignent pas deux types de phénomènes radicalement différents, mais peuvent renvoyer aux mêmes fluctuations relativement à deux systèmes différents, en fonction de la cible privilégiée par telle ou telle étude.

⁶³ Voir Kærn *et al*, 2005, *op. cit.*

Dans le cas de la distinction et des définitions que nous avons appelées « opérationnelles », il faut aussi préciser que les biologistes définissent le « bruit total » affectant un certain système, par exemple le système d'expression génétique, comme la somme du bruit intrinsèque et du bruit extrinsèque. Pourtant, il a été montré que les fluctuations intrinsèques et extrinsèques relativement à un même système d'expression génétique ne sont pas indépendantes mais peuvent s'influencer réciproquement⁶⁴. Mais qu'est-ce qui conduit les biologistes à distinguer et à caractériser ces fluctuations comme s'il s'agissait de deux composants du bruit dont les effets seraient additifs ? Essayons de clarifier ce point, douteux selon nous, à travers deux exemples.

En premier lieu, les biologistes considèrent les différences dans l'expression de deux protéines à partir de deux promoteurs identiques situés sur le même chromosome dans la même cellule comme la manifestation du bruit intrinsèque. Cependant, ces différences pourraient aussi bien être dues à du bruit extrinsèque. En effet, en dépit du fait que les fluctuations dans les variables considérées comme extrinsèques (par exemple, la concentration des ARN polymérase) soient des caractéristiques du microenvironnement commun aux deux rapporteurs, elles pourraient néanmoins les affecter différemment. Par exemple, les ARN polymérase pourraient être distribuées de manière non homogène dans le milieu intracellulaire de façon à être plus abondantes à proximité du promoteur d'un des deux gènes. Un biologiste répondrait à notre objection en soulignant que le caractère homogène et bien mélangé du microenvironnement intracellulaire est soigneusement contrôlé expérimentalement. De plus, un tel contrôle va de pair avec les modèles mathématiques de ce processus, dont fait partie l'hypothèse du caractère homogène et bien mélangé du milieu intracellulaire. Cependant, vu qu'il s'agit d'une simplification non négligeable de ce que l'on sait de l'environnement intracellulaire, il est tout de même important de souligner cette limite dans la distinction et les définitions « opérationnelles » du bruit intrinsèque et du bruit extrinsèque.

Par ailleurs, les biologistes considèrent comme une manifestation du bruit extrinsèque les différences dans l'expression de deux protéines à partir d'un même promoteur chez deux cellules isogéniques dans un environnement homogène et constant (ou dans une même cellule au cours du temps). Cependant, ces différences pourraient également être dues à du bruit intrinsèque, c'est-à-dire aux fluctuations inhérentes à l'expression de chaque gène.

⁶⁴ En particulier, le bruit extrinsèque semble pouvoir influencer le bruit intrinsèque de l'expression des gènes ; voir Kærn *et al*, 2005, « Stochasticity in gene expression: from theories to phenotypes ».

Ces limites suggèrent l'intérêt de remettre en cause la pertinence conceptuelle et empirique de ces définitions et distinctions « théoriques » et « opérationnelles » du bruit, et en particulier leur prétendue utilité dans l'étude de l'expression des gènes.

8.6. La modélisation stochastique de l'expression des gènes

Les biologistes utilisent des modèles mathématiques afin de décrire le processus d'expression génétique, de l'analyser (bien que ce ne soit que de manière indirecte) et d'en prévoir le résultat final, à savoir la quantité de protéines produites.

Les méthodes expérimentales que nous avons présentées ci-dessus accompagnent la modélisation mathématique et, dans ce cadre, jouent un double rôle : premièrement, elles fournissent les données à partir desquelles on attribue des valeurs aux paramètres dans les modèles ; deuxièmement, elles permettent de tester (de confirmer ou d'infirmer) les prédictions obtenues par modélisation mathématique et par simulation numérique⁶⁵. Nous utilisons l'expression « modélisation mathématique » pour désigner la représentation et l'analyse indirectes d'un phénomène réel (dans ce cas, le processus d'expression génétique) par des « modèles mathématiques », ces dernières étant des équations.⁶⁶

Nous rappelons que les modèles mathématiques de l'expression des gènes s'insèrent dans le cadre de la cinétique chimique. Cette discipline consiste à décrire l'état d'un système chimique à un moment donné et son évolution au cours du temps en termes du nombre ou de la concentration de ses composants moléculaires. En cinétique chimique classique, les réactions chimiques sont représentées d'un point de vue purement phénoménologique⁶⁷ comme des phénomènes macroscopiques, continus et déterministes. Une loi empirique (c'est-à-dire dérivée d'expériences *in vitro*), appelée « loi d'action de masse », fournit une relation de proportionnalité très simple entre les taux de réaction et les concentrations d'espèces moléculaires ; elle est exprimée par des équations différentielles ordinaires⁶⁸, les équations du taux de réaction, qui décrivent la dynamique des réactions chimiques. Par exemple, si l'on

⁶⁵ Dans le cadre de ce travail, nous ne traiterons pas ce genre de simulations.

⁶⁶ Le statut des équations en tant que modèles est un objet de controverse en philosophie des sciences. Pour une introduction critique au débat sur la notion de modèle, voir Frigg & Hartmann, 2008, « Models in Science ».

⁶⁷ Le terme « phénoménologique » renvoie à une approche purement descriptive, généralement au niveau macroscopique, qui se contente de rendre compte des phénomènes tels qu'on les observe.

⁶⁸ Une équation différentielle ordinaire est une équation mathématique décrivant une relation entre une fonction inconnue d'une seule variable et ses dérivées jusqu'à un certain ordre ($dx/dt = f(x)$). Une dérivée est la représentation d'un changement infinitésimal dans la fonction par rapport à sa variable ($f'(x) = df/dx$).

considère la réaction $A + B = C$ (A , B et C étant des réactants), son taux r est donné par l'équation suivante :

$$r = k(T) [A]^m [B]^n,$$

$[A]$ et $[B]$ étant l'expression de la concentration des réactants A et B ;

m étant l'ordre de la réaction par rapport au réactant A ⁶⁹ ;

n étant l'ordre la réaction par rapport au réactant B ;

t étant le temps ;

$k(T)$ étant une constante qui dépend de la température⁷⁰.

À partir des concentrations initiales des molécules, cette loi permet de décrire univoquement l'état du système chimique considéré tout au long de son évolution ; autrement dit, elle permet de déterminer ces concentrations de manière complète et à tout moment.

Nous concentrerons notre analyse sur les modèles mathématiques de l'expression des gènes empruntés à la cinétique chimique stochastique⁷¹. Cette approche, contrairement à celle de la cinétique chimique classique, décrit l'évolution dans le temps d'un système de réactions chimiques considérées comme ayant lieu dans un milieu bien mélangé⁷², en tenant compte du fait que les molécules sont des éléments discrets et que leur comportement est caractérisé par un certain degré de stochasticité⁷³. L'utilisation de la cinétique chimique stochastique pour décrire des systèmes biochimiques contenant un petit nombre d'espèces moléculaires réactives implique une déviation par rapport aux prévisions des équations différentielles déterministes utilisées par l'approche de la cinétique chimique classique⁷⁴.

⁶⁹ L'ordre de réaction, par rapport à un certain réactant, est la puissance à laquelle le terme de sa concentration est élevé dans l'équation du taux de réaction.

⁷⁰ On parle de la constante d'équilibre de Guldberg et Waage. Ces deux physiciens formulèrent pour la première fois la loi d'action de masse en 1864. La constante $k(T)$ correspond à la valeur reliant les concentrations des réactants de départ et des produits formés dans un système qui a atteint l'équilibre.

⁷¹ En effet, comme nous l'avons déjà annoncé plus haut, à partir de la fin des années 1990, les biologistes ont concentré leurs efforts sur la description du processus d'expression génétique par des modèles stochastiques et quantitatifs.

⁷² Dans un milieu « bien mélangé », les molécules sont continuellement en agitation et sont distribuées de manière aléatoire.

⁷³ Bartholomay (1957, *A stochastic approach to chemical reaction kinetics*) fut l'un des premiers chimistes à modéliser des réactions enzymatiques selon l'approche de la cinétique stochastique. Ensuite, au cours des vingt années suivantes, un certain nombre de réactions biochimiques, dont le mécanisme Michaelis-Menten, furent analysées de manière stochastique (voir Turner *et al*, 2005, « Stochastic approaches for modelling in vivo reactions »).

⁷⁴ Gillespie, 2007, « Stochastic Simulation of Chemical Kinetics ».

Analysons les modèles mathématiques stochastiques utilisés pour décrire l'expression des gènes⁷⁵, les équations maîtresses, de Fokker-Planck et de Langevin⁷⁶. Il s'agit d'équations assez abstraites pouvant décrire en termes mathématiques des processus stochastiques divers dans le cadre de plusieurs disciplines scientifiques (physique, biologie, économie, sociologie, etc.). Lorsqu'elles sont utilisées pour décrire l'évolution d'un système chimique et, plus spécifiquement, un système de réactants moléculaires impliqué dans l'expression des gènes, elles partagent les hypothèses suivantes.⁷⁷ Premièrement, les collisions moléculaires sont supposées être élastiques, c'est-à-dire qu'elles ne s'accompagnent d'aucune transformation de la forme des molécules ni de leurs vitesses. De plus, les positions des molécules à l'intérieur du volume cellulaire sont supposées être distribuées de manière aléatoire et leurs vitesses supposées suivre la distribution de Maxwell-Boltzmann⁷⁸. Cela repose forcément sur une certaine idéalisation, suivant laquelle le microenvironnement intracellulaire où l'expression des gènes a lieu est un milieu homogène et bien mélangé au cours du temps. Elle comporte aussi quelques simplifications : la probabilité qu'une molécule se trouve en un endroit quelconque du microenvironnement cellulaire est considérée comme uniforme dans tout son volume ; de plus, la probabilité qu'une molécule donnée entre en collision avec une autre molécule quelconque et la probabilité qu'elles produisent une réaction sont considérées comme indépendantes de la localisation des molécules à l'intérieur de la cellule⁷⁹.

Ces hypothèses communes à tout modèle mathématique de l'expression des gènes sont étroitement liées à la caractérisation statistique du bruit censé l'affecter : il correspond à un processus stochastique markovien⁸⁰ non corrélé dans le temps⁸¹. Lorsqu'il s'agit d'un bruit

⁷⁵ Nous nous limiterons à une présentation très simple de ces modèles stochastiques sans entrer dans les détails techniques de leurs formulations mathématiques. Pour une introduction aux mathématiques des processus stochastiques que ces modèles utilisent, voir Van Kampen, 1992, *Stochastic Processes in Physics and Chemistry*.

⁷⁶ Nous les introduisons ci-dessus.

⁷⁷ Voir Gillespie, 2007, *op. cit.*

⁷⁸ La distribution de Maxwell-Boltzmann est celle des vitesses des molécules dans un gaz à l'équilibre thermodynamique.

⁷⁹ Ces hypothèses simplificatrices sont de plus en plus discutées sur la base de considérations empiriques selon lesquelles, au contraire, le milieu intracellulaire (en particulier, le cytoplasme où a lieu la maturation des molécules d'ARN et la synthèse des protéines) est subdivisé en plusieurs compartiments et est rempli de manière non homogène par des molécules d'espèces diverses. La diffusion d'une molécule n'est donc pas libre mais limitée par tous ces obstacles moléculaires (Turner *et al*, 2004, « Stochastic approaches for modelling in vivo reactions »). Les biologistes essaient d'évacuer l'hypothèse d'un environnement homogène et bien mélangé en construisant et en utilisant des méthodes stochastiques qui prennent en compte les différences inhérentes au milieu intracellulaire, par exemple, les simulations Monte Carlo bi-dimensionnelles (El Samad *et al*, 2005, « Stochastic modelling of gene regulatory networks »).

⁸⁰ Voir le Glossaire.

⁸¹ Dans un processus stochastique non corrélé dans le temps, l'état présent est complètement indépendant des états passés. Par conséquent, aucune régularité ne caractérise ce genre de processus.

blanc gaussien, comme dans le cadre de l'équation heuristique de Langevin, c'est un processus stochastique qui suit une loi normale⁸² de moyenne nulle.

Considérons l'équation maîtresse qui, dans le cadre de la cinétique chimique stochastique, décrit de manière discrète l'évolution de la distribution de probabilité sur les états possibles d'un système chimique, ces derniers étant décrits en termes du nombre des espèces moléculaires le composant. Dans la modélisation de l'expression des gènes, l'équation maîtresse décrit l'évolution de la distribution de probabilité sur les valeurs possibles du nombre de protéines comprise dans le volume cellulaire de la manière suivante :

$$\frac{dP(X)}{dt} = kP(X-1) + \gamma(X+1)P(X+1) - [k + \gamma X]P(X)$$

X étant le nombre de protéines dans la cellule ;

$P(X)$ étant la probabilité qu'il y ait X protéines dans la cellule ;

t étant le temps ;

k et γ étant respectivement les taux de production et de dégradation de X .

Cette équation maîtresse décrit le processus de « création-destruction » d'un certain type de protéines dans une cellule donnée. La probabilité de transition⁸³ d'un état à un autre en termes de nombre de protéines dans le volume cellulaire dépend des taux de production et de dégradation des protéines, qui dépendent à leur tour d'un ensemble d'événements stochastiques impliqués dans l'expression génétique (par exemple, l'activation et la répression des gènes, le taux de transcription, le taux de dégradation des transcrits d'ARN, etc.). Les trois termes de l'équation décrivent respectivement : $kP(X-1)$, la probabilité de transition de l'état $X-1$ à l'état X étant donnée la probabilité que la cellule comprenne $X-1$ protéines ; $\gamma(X+1)P(X+1)$, la probabilité de transition de l'état $X+1$ à l'état X , étant donnée la probabilité que la cellule comprenne $X+1$ protéines ; $[k+\gamma X]P(X)$, la probabilité qu'il n'y ait aucun changement d'état, étant donnée la probabilité que le nombre de protéines comprises dans le volume cellulaire soit X .

L'équation maîtresse est considérée par les biologistes comme la description mathématique la plus appropriée de l'expression génétique car elle prend en considération le

⁸² Voir le Glossaire.

⁸³ La probabilité de transition est la probabilité qu'un système atteigne l'état X à l'instant $t+\tau$ étant donnée la probabilité qu'il se trouve à l'état $X-1$ ou à l'état $X+1$ à l'instant immédiatement précédent t .

caractère discret des espèces moléculaires impliquées dans ce processus et le caractère stochastique de leurs interactions. Bien qu'il s'agisse d'une équation assez facile à résoudre, elle trouve une limite lorsque le nombre des espèces moléculaires qui doivent être prises en compte augmente de manière considérable : sa solution analytique se révèle impossible, et même sa résolution numérique par simulation informatique est souvent trop coûteuse en termes de temps de calcul. Une alternative possible consiste à simuler numériquement le processus d'expression génétique au moyen d'algorithmes divers qui sont basés sur les méthodes de simulations stochastiques de Monte Carlo⁸⁴. Une autre solution pour faire face aux limites de résolubilité de l'équation maîtresse consiste à décrire l'expression des gènes en utilisant d'autres équations différentielles comme celles de Fokker-Planck et de Langevin.

L'équation de Fokker-Planck est un cas particulier de l'équation maîtresse qu'on obtient lorsque la variation entre différents états devient très petite et tend vers zéro, c'est-à-dire lorsqu'on passe d'une représentation discrète de l'évolution d'un système à une représentation continue⁸⁵. Il s'agit d'une équation de diffusion qui, en cinétique chimique stochastique, décrit l'évolution dans le temps de la densité de probabilité⁸⁶ sur les états possibles qu'un système chimique peut occuper, ces derniers étant décrits en termes du nombre ? des espèces moléculaires le composant. Dans son application au processus d'expression génétique, elle décrit l'évolution dans le temps de la densité de probabilité sur les valeurs possibles de la concentration de protéines dans le volume cellulaire :

$$\frac{\partial P(x,t)}{\partial t} = \frac{\partial}{\partial x} ((k' - \gamma x)P) + \frac{\partial^2}{\partial x^2} \left(\frac{(k' + \gamma x)}{2} P \right)$$

x étant la concentration de protéines (X/V , V étant le volume cellulaire) ;

$P(x)$ étant la probabilité, dépendant du temps, que la concentration de protéines dans la cellule soit x ;

t étant le temps ;

k et γ étant respectivement les taux de production et de dégradation ;

⁸⁴ Par exemple, l'algorithme de simulation stochastique (SSA), introduit par Gillespie en 1977. Les méthodes Monte-Carlo, dont le nom fut popularisé pas les mathématiciens Ulam, Fermi, von Neumann et Metropolis à la fin de la Deuxième Guerre Mondiale, sont des techniques numériques de simulation statistique. Elles utilisent des séquences de nombres aléatoires ou pseudo-aléatoires afin de simuler l'évolution d'un système donné et de calculer une valeur numérique approchée de son résultat final. Ces méthodes sont utilisées dans des disciplines scientifiques diverses, par exemple en physique, en optimisation, en chimie, en imagerie, en statistique, et en biologie. Voir Gillespie, 2007, « Stochastic Simulation of Chemical Kinetics ».

⁸⁵ Van Kampen, 1992, *Stochastic Processes in Physics and Chemistry*.

⁸⁶ Voir le Glossaire.

k' étant k/V .

L'équation Fokker-Planck ci-dessus comprend un premier terme dit « de dérive » (ou « de transport », ou « de convection », $\frac{\partial}{\partial x}((k' - \gamma x)P)$ qui décrit la tendance du changement dans la concentration de protéines et un deuxième terme dit « de diffusion » (ou « de fluctuation », $\frac{\partial^2}{\partial x^2} \left(\frac{(k' + \gamma x)}{2} P \right)$ afin de rendre compte des fluctuations stochastiques dues aux flux « entrant » et « sortant » relativement à la concentration des protéines.

L'équation de Langevin est une approximation de l'équation maîtresse et peut aussi être considérée comme une approximation de l'équation de Fokker-Planck⁸⁷. Cependant, elle est souvent construite heuristiquement en ajoutant à une équation différentielle déterministe à une variable aléatoire – un terme de bruit ξ – qui est habituellement un bruit blanc gaussien⁸⁸ et qui rend compte des perturbations stochastiques affectant le système considéré. Dans le cadre de la cinétique chimique stochastique et, plus particulièrement, dans le cas de l'expression des gènes, elle peut prendre la forme suivante :

$$\frac{dx}{dt} = k - \gamma x + \xi$$

x étant la concentration de protéines dans la cellule;

t étant le temps ;

k et γ étant respectivement les taux de production et de dégradation ;

ξ étant un bruit blanc gaussien.

L'équation de Langevin ci-dessus ne représente pas les protéines dans le volume cellulaire comme des éléments discrets, mais de manière continue en termes de leur concentration : elle décrit le processus d'expression génétique en termes de changements infinitésimaux dans la concentration des protéines et des fluctuations stochastiques affectant leur relative abondance.

Les équations de Fokker-Planck et de Langevin peuvent également être résolues par simulation numérique lorsque l'expression des gènes implique un nombre considérable

⁸⁷ Van Kampen, 1992, *op. cit.* Voir aussi Gillespie, 2000, « The chemical Langevin equation » ; Gillespie, 2007, *op. cit.*

⁸⁸ L'équation de Langevin est mathématiquement équivalente à l'équation de Fokker-Planck lorsque le terme « bruit » correspond à un bruit blanc gaussien.

d'espèces moléculaires. Cependant, les simulations numériques peuvent devenir très coûteuses en termes de temps et elles échouent parfois à résoudre ces équations.⁸⁹

⁸⁹ Dans le chapitre 10, nous considérerons de nouveau les aspects communs et les différences entre les équations maîtresses, les équations de Fokker-Planck et de Langevin.

Chapitre 9 – Le hasard et le bruit dans l’expression génétique

Dans les descriptions verbales et les formalisations mathématiques des fluctuations petites et désordonnées (bruit) qui affectent l’expression des gènes, sont utilisés, d’une part des termes comme « aléatoire », « stochastique », « probabiliste », « imprédictible » qui apparemment renvoient à la notion de hasard, d’autre part des probabilités qui semblent entretenir une relation étroite avec cette dernière. Dans ce chapitre, nous souhaitons montrer que les notions de « hasard évolutionnaire » et d’« aléatoire (au sens) faible » correspondent aux deux notions de hasard conceptuellement et empiriquement appropriées pour caractériser cette source causale de la variation biologique qu’est le bruit. À cette fin, nous considérerons un certain nombre de facteurs physico-chimiques pouvant influencer ses propriétés spatio-temporelles, plus précisément ceux qui sont à l’origine de la régulation du niveau de bruit¹ le long du génome et au cours du temps. Nous montrerons que le bruit intervenant dans l’expression génétique est biaisé, mais non pas dirigé, du point de vue évolutionnaire et aussi du point de vue plus strictement physico-chimique : il peut donc être considéré comme le fruit du « hasard évolutionnaire » et comme un phénomène « aléatoire », non pas « (au sens) fort » mais « (au sens) faible ».

9.1. Le bruit et la notion de « hasard évolutionnaire »

Analysons d’abord la notion de « hasard évolutionnaire » afin d’en montrer la pertinence dans la caractérisation du bruit dans l’expression des gènes. Cette tâche est apparemment plus simple que celle consistant à montrer le caractère approprié de cette même notion dans le cas des mutations génétiques². En effet, la présence du bruit à l’intérieur de la cellule et l’influence qu’il exerce sur les processus biochimiques s’y déroulant – entre autres, l’expression génétique – n’ont jamais été invoquées en faveur d’un retour des idées

¹ Lorsque nous utilisons le terme « niveau » relativement au bruit, nous voulons renvoyer au fait que ce dernier est toujours identifié quantitativement par des mesures statistiques (par exemple, le coefficient de variation). Voir aussi Raser & O’Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control », Support Online Material).

² Voir les chapitres 4, 5 et 6.

lamarckiennes quant à l'origine de la variation biologique. Premièrement, le bruit dans l'expression des gènes est considéré comme un exemple supplémentaire du caractère désordonné, irrégulier et imprédictible propre à toute source de variation biologique, que ce soit génétique ou phénotypique. Deuxièmement, ce phénomène intracellulaire est à l'origine d'une variation temporaire et facilement réversible³ qui, contrairement à la variation due aux mutations génétiques⁴, n'est pas transmise d'une génération à la suivante. Cependant, les recherches à ce sujet ont récemment mis en évidence qu'en dépit du caractère (qu'on dit) « aléatoire » des fluctuations affectant l'expression des gènes, leurs propriétés spatio-temporelles ne sont pas indépendantes de la situation environnementale et sélective où les organismes concernés se trouvent et, plus généralement, de l'histoire évolutive des populations auxquelles ils appartiennent. Les résultats de ces recherches ont apporté de nombreux éléments en faveur de l'idée selon laquelle le bruit dans l'expression génétique est un trait sujet à sélection qui peut évoluer au cours du temps et qui peut être transmis à la génération suivante.

Les biologistes qui se sont penchés sur l'origine et les conséquences du bruit dans l'expression des gènes⁵ ont montré, premièrement, que les fluctuations affectant ce processus intracellulaire sont régulées (réduites ou amplifiées) au sein d'une population de cellules génétiquement identiques et même chez une cellule individuelle au cours du temps⁶ ; deuxièmement, qu'une telle régulation est différente selon la fonction des gènes (et donc des protéines) concernés⁷ et selon les caractéristiques des conditions environnementales où la population cellulaire se trouve (en particulier, en fonction du type de changements environnementaux et de la fréquence avec laquelle ils se produisent⁸).

³ Thattai & van Oudenaarden (2004, « Stochastic Gene Expression in Fluctuating Environments ») qualifient la variation due au bruit intracellulaire d'« hétérogénéité dynamique ».

⁴ Ou « hétérogénéité statique » (*ibid.*).

⁵ La plupart des recherches sur le bruit dans l'expression des gènes sont réalisées par des biophysiciens, des généticiens moléculaires et des microbiologistes : elles portent sur la modélisation mathématique et/ou sur l'étude expérimentale de ce processus intracellulaire chez des microorganismes, notamment chez la bactérie *Escherichia coli* et chez la levure *Saccharomyces cerevisiae*. En effet, les microorganismes fournissent un cas idéal pour l'étude des dynamiques développementales et évolutives car ils croissent, évoluent et s'adaptent très rapidement et car ils peuvent être observés assez facilement en laboratoire dans des conditions hautement contrôlées. Cependant, l'origine et les conséquences du bruit dans l'expression génétique ont aussi été étudiées chez des organismes multicellulaires, notamment chez la drosophile, chez les mammifères et aussi chez l'homme où le bruit s'est révélé jouer un rôle central dans le processus de différenciation cellulaire, dans le processus de vieillissement et dans le développement de maladies comme le cancer. Pour des références à ce sujet, voir Kærn *et al*, 2005, « Stochasticity in gene expression: from theories to phenotypes » ; Raser & O'Shea, 2005, *op. cit.* ; Losick & Desplan, 2008, « Stochasticity and cell fate ».

⁶ Raser & O'Shea, 2005, *op. cit.* ; Samoilov *et al*, 2006, « From Fluctuations to Phenotypes: The Physiology of Noise » ; Maheshri & O'Shea, 2007, « Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression » ; Losick & Desplan, 2008, *op. cit.*

⁷ Fraser *et al*, 2004, « Noise Minimization in Eukaryotic Gene Expression ».

⁸ Thattai & van Oudenaarden, 2004, *op. cit.* ; Kussell *et al*, 2005, « Bacterial Persistence: A Model of Survival in

Les résultats de nombreuses études ont aussi soutenu que la régulation du bruit intracellulaire est une adaptation due à l'histoire évolutive de la population concernée, peut être transmise d'une génération à la suivante, et est susceptible d'évoluer de manière adaptative au cours du temps⁹. D'une part, le bruit intervenant dans l'expression des gènes est une source importante de variabilité phénotypique (dynamique car facilement réversible) au niveau populationnel : il est donc susceptible de favoriser la survie d'une population lorsqu'elle doit faire face à des conditions environnementales de stress, en particulier en l'absence de toute source causale de la variation génétique. D'autre part, ce bruit peut jouer un rôle décisif dans la détermination du destin de chaque cellule individuelle, c'est-à-dire dans la détermination de son état phénotypique et donc dans sa survie. Plus précisément, c'est l'organisation structurale et dynamique des réseaux de gènes et de protéines impliqués dans le processus d'expression génétique qui permet de réduire ou d'amplifier le bruit intracellulaire : elle est susceptible d'être transmise d'une génération à la suivante et est très probablement le résultat du parcours évolutif, plus précisément adaptatif, de la population cellulaire concernée.

9.1.1. Structures et mécanismes de contrôle et d'exploitation du bruit

Examinons les facteurs qui permettent une régulation du bruit (intrinsèque et extrinsèque) dans l'expression des gènes et, plus généralement, assurent un maintien du fonctionnement de la cellule en dépit des perturbations intracellulaires.¹⁰

En premier lieu, la présence de boucles de rétroaction (*feedback*) négatives, positives ou intégrales dans les réseaux très complexes de gènes et de protéines est à l'origine du contrôle et de l'exploitation du bruit dans l'expression des gènes, c'est-à-dire de la variation dans l'abondance de protéines produites au sein d'une population de cellules génétiquement identiques à un moment donné dans le temps ou dans une même cellule au cours du temps. Une boucle de rétroaction est un processus circulaire (ou circuit) permettant à un système d'évaluer son propre comportement dynamique et de l'ajuster en fonction du comportement souhaité. On parle de rétroaction négative lorsqu'un système produit une réponse qui va dans

Changing Environments » ; Lopez-Maury *et al*, 2008, « Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation ».

⁹ Voir les références dans la section 9.1.2 de ce chapitre.

¹⁰ Pour une liste exhaustive des publications au sujet de la régulation du bruit, voir Rao *et al*, 2002, « Control, exploitation and tolerance of intracellular noise » ; Stelling *et al*, 2004, « Robustness of Cellular Functions » ; Raser & O'Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control » ; Kærn *et al*, 2005, *op. cit.* ; Maheshri & O'Shea, 2007, *op. cit.*

la direction opposée par rapport à la perturbation qu'il a subie et, de cette manière, la réduit. En ingénierie et en économie, il est bien connu que ce type de mécanisme permet à un système d'opérer de manière robuste dans un environnement fluctuant. Lorsqu'un système produit une réponse qui va dans la même direction que la perturbation l'ayant affecté et, de cette manière, l'amplifie, on parle de boucle de rétroaction positive ou de mécanisme d'autocatalyse.¹¹ La rétroaction intégrale est un type particulier de boucle de rétroaction négative où un système régule son propre comportement sur la base d'une sorte de mémoire interne et en référence à un résultat souhaité.

Dans le cas du processus d'expression génétique, il a été montré que la rétroaction négative permet de réduire le bruit, c'est-à-dire la variation autour de la valeur moyenne de protéines produites¹². De plus, une boucle de rétroaction négative, en augmentant la fréquence des fluctuations au niveau de l'une des étapes de l'expression des gènes, peut minimiser l'impact du bruit sur ses étapes successives et, plus généralement, sur le développement de l'organisme individuel concerné. En effet, un processus n'est affecté par des fluctuations que si celles-ci ont des fréquences plus basses que la fréquence propre du processus ; donc, l'augmentation de la fréquence du bruit dans l'expression des gènes par une boucle de rétroaction négative réduit les effets de ce dernier car elle le rend invisible¹³. Un exemple de rétroaction négative jouant un rôle dans la réduction du bruit intracellulaire est le phénomène d'autorépression : un facteur de transcription inhibe sa propre synthèse, ce qui permet d'ajuster sa production à la demande et relativement à sa concentration au sein de la cellule¹⁴. Ce genre de boucles de rétroaction permet la stabilisation du niveau d'expression des gènes en dépit des perturbations intracellulaires (par exemple, en dépit d'une source de bruit intrinsèque concernant l'activité d'un facteur de transcription).

En ce qui concerne la rétroaction positive, il a été montré que sa présence amplifie le bruit et, de cette manière, peut générer de la variation phénotypique au sein de la population

¹¹ Dans un système non linéaire (qui, de manière simple, peut être défini comme un système instable caractérisé par une certaine disproportion entre ses conditions initiales et son état final), une boucle de rétroaction positive peut réduire le bruit par son amplification. En effet, un bruit élevé peut améliorer les performances d'un système car il augmente sa sensibilité à des signaux utiles et cohérents qui sont persistants mais faibles, donc difficilement détectables. C'est le phénomène de « résonance stochastique ». Voir Hänggi, 2002, « Stochastic Resonance in Biology. How Noise Can Enhance Detection of Weak Signals and Help Improve Biological Information Processing ».

¹² Becksei & Serrano, 2000, « Engineering stability in gene networks by autoregulation » ; Thattai & van Oudenaarden, 2001, « Intrinsic noise in gene regulatory networks » ; Orrell & Bolouri, 2004, « Control of internal and external noise in genetic regulatory networks ».

¹³ Pour une démonstration théorique de ce phénomène, voir Simpson *et al*, 2003, « Frequency domain analysis of noise in autoregulated gene circuits ».

¹⁴ Hasty *et al*, 2001, « Computational studies of gene regulatory networks: *in numero* molecular biology ».

cellulaire concernée¹⁵. En particulier, par l'amplification du bruit, les boucles de rétroaction positives (aussi bien que les boucles de rétroaction double-négatives et d'autres plus complexes encore) sont à l'origine de ce qu'on appelle le phénomène de *switch* (basculement). Au cours de son cycle de vie, une cellule individuelle peut passer d'un état phénotypique à un autre grâce aux fluctuations stochastiques affectant le niveau d'expression de l'un ou de plusieurs de ses gènes. En effet, ces fluctuations intracellulaires permettent de dépasser un certain seuil dans la concentration d'un ou de plusieurs facteurs protéiques et, de cette manière, activent le *switch* phénotypique.

Ce genre de phénomène a été observé et étudié chez la bactérie *Bacillus subtilis* dont l'état phénotypique de « compétence » pour la transformation – où les bactéries sont en phase stationnaire et sont capables d'incorporer dans leur propre chromosome de l'ADN trouvé dans l'environnement – est contrôlé par une boucle de rétroaction positive¹⁶. Les gènes requis pour qu'une bactérie soit dans l'état phénotypique de compétence ne sont exprimés qu'en présence du facteur protéique ComK, dont le taux de synthèse est très bas en phase de croissance où la boucle de rétroaction positive le concernant est inactive. Des fluctuations importantes dans le niveau d'expression du gène *comK* (ou bruit intrinsèque) peuvent provoquer le dépassement d'un certain seuil dans la concentration de ce facteur protéique et l'activation de la boucle de rétroaction positive. De cette manière, le facteur protéique ComK commence à autocatalyser sa propre synthèse, les gènes requis pour l'état de compétence sont exprimés et la bactérie individuelle concernée bascule (*switches*) de l'état de non compétence (ou état dit « végétatif ») à l'état de compétence, c'est-à-dire qu'elle devient capable d'incorporer de l'ADN étranger. Si, au cours du cycle de vie d'une bactérie individuelle, les fluctuations dans le niveau d'expression du gène *comK*, et donc la concentration du facteur protéique ComK, diminuent, la boucle de rétroaction positive est désactivée. La bactérie retourne à l'état végétatif et ne peut regagner l'état de compétence que dans le cas d'une augmentation de la concentration de ComK, de la réactivation de la boucle de rétroaction positive qui s'ensuit et donc de l'expression des gènes requis par cet état phénotypique.

¹⁵ Hasty *et al*, 2000, « Noise-based switches and amplifiers for gene expression » ; Becksei *et al*, 2001, « Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion » ; Isaacs *et al*, 2003, « Prediction and measurement of an autoregulatory genetic module ». Pour plus de références, voir Maheshri & O'Shea, 2007, « Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression » ; Rao *et al*, 2002, « Control, exploitation and tolerance of intracellular noise » ; Stelling *et al*, 2004, « Robustness of Cellular Functions ».

¹⁶ Voir Maamar *et al*, 2007, « Noise in gene expression determines cell fate in *Bacillus subtilis* » ; Leisner *et al*, 2008, « Stochastic switching to competence ». Voir aussi Losick & Desplan, 2008, « Stochasticity and cell fate » ; Samoilov *et al*, 2006, « From Fluctuations to Phenotypes: The Physiology of Noise ».

Enfin, le rôle de la rétroaction intégrale dans la régulation du bruit a été étudié dans le cas de l'expression des gènes aussi bien que d'autres processus biologiques, par exemple la chemotaxie chez les bactéries, qui est un processus par lequel ces microorganismes dirigent leur mouvement selon la concentration des gradients chimiques présents dans l'environnement¹⁷.

La redondance est une deuxième manière dont un système peut atténuer le bruit et l'impact sur ses performances. C'est le fait qu'il y ait plusieurs copies d'une même structure ou d'un même composant, ou encore le fait de posséder des structures ou des composants différents mais pouvant accomplir une même fonction. De cette manière, un système peut augmenter sa capacité à persister malgré des perturbations internes et externes ; en particulier, il peut réduire l'impact de ces perturbations sur la fonction accomplie par les structures ou les composants redondants et fonctionnellement équivalents. En ce qui concerne l'expression génétique, il a été montré que la présence de plusieurs copies d'un même gène au sein d'une cellule et, par conséquent, de plusieurs cascades parallèles de son expression, permet de minimiser le bruit, c'est-à-dire la variation autour de la valeur moyenne de protéines produites à partir de ce gène¹⁸.

L'existence de points de contrôle (*checkpoints*) représente un autre moyen par lequel un système peut réduire l'impact du bruit qui l'affecte : leur rôle est d'assurer la complète réalisation de chaque étape d'un processus donné avant de pouvoir procéder à l'étape suivante¹⁹. Dans le cas de l'expression des gènes, la présence de plusieurs mécanismes de contrôle des processus de transcription et de traduction²⁰ augmente la fidélité de la synthèse des protéines.

Enfin, des travaux de modélisation de l'expression des gènes ont révélé que le bruit est corrélé aux taux de transcription et de traduction²¹. Il a été montré qu'un taux élevé de transcription suivi par un taux très bas de traduction est associé à une faible variation dans l'abondance de protéines produites, alors qu'un taux bas de transcription suivi par un taux élevé de traduction est associé à une variation importante dans le niveau d'expression des

¹⁷ Yi *et al*, 2000, « Robust perfect adaptation in bacterial chemotaxis through integral feedback control ».

¹⁸ Cook *et al*, 1998, « Modeling stochastic gene expression: implications for haploinsufficiency » ; Kellis *et al*, 2004, « Proof and evolutionary analysis of ancient genome duplication in the yeast *Sccharomyces cerevisiae* ».

¹⁹ Hartwell & Weinert, 1989, « Checkpoints: controls that ensure the order of the cell cycle events ».

²⁰ Voir Alberts *et al*, 2002[1983], *Molecular Biology of the Cell* ; Lewin, 2004, *Genes VIII*.

²¹ Raser & O'Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control » ; McAdams & Arkin, 1997, « Stochastic mechanisms in gene expression » ; Kierzek *et al*, 2001, « The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression » ; Kepler & Elston, 2001, « Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations » ; Thattai & van Oudenaarden, 2001, « Intrinsic noise in gene regulatory networks ».

gènes mesuré en termes de nombre de protéines produites. Ces conclusions, d'abord suggérées sur la base de simulations numériques de modèles mathématiques, ont ensuite été vérifiées expérimentalement chez des organismes procaryotes (comme la bactérie *Bacillus subtilis*²²) et eucaryotes (comme la levure *Saccharomyces cerevisiae*²³).²⁴

9.1.2. La régulation adaptative du bruit

Sur la base de tous les résultats présentés ci-dessus, les biologistes soutiennent que les structures et les mécanismes propres aux réseaux de gènes et de protéines ne sont pas dus à de simples contraintes physico-chimiques, mais ont évolué au cours du temps. Nous faisons référence à la notion de « contrainte physique » selon laquelle une contrainte est une limitation de la variation biologique possible due aux lois physiques ou aux principes d'auto-organisation déterminant les caractéristiques des organismes vivants. Il s'agit d'une notion non historique de contrainte qui s'oppose à l'histoire évolutive des populations naturelles en tant qu'explication alternative de la variation²⁵. De plus, les biologistes soutiennent l'hypothèse « adaptative » selon laquelle ces structures et ces mécanismes, qui augmentent la robustesse du fonctionnement cellulaire face aux perturbations non pas génétiques ni

²² Ozbudak *et al*, 2002, « Regulation of noise in the expression of a single gene ».

²³ Blake *et al*, 2003, « Noise in eukaryotic gene expression ».

²⁴ Il est intéressant de souligner que le bruit intrinsèque et le bruit extrinsèque ne sont pas corrélés de la même manière au taux de synthèse des protéines. Par exemple, bien que les deux soient minimaux lorsque le taux de transcription est élevé, le bruit intrinsèque est maximal lorsque ce taux est très bas alors que le bruit extrinsèque l'est à un niveau intermédiaire du processus de transcription (Elowitz *et al*, 2002, « Stochastic Gene Expression in a Single Cell »). D'ailleurs, le bruit extrinsèque s'est révélé être la source principale de variation dans l'abondance de protéines produites (Raser & O'Shea, 2005, *op. cit.*).

²⁵ La notion de « contrainte physique » (Thompson, 1917, *On Growth and Form* ; Goodwin, 1994, *How The Leopard Changed Its Spots: The Evolution of Complexity* ; Kauffman, 1993, *The origins of order: Self-Organization and Selection in Evolution*) n'est pas la seule manière de définir ce qu'est une contrainte en biologie. Selon la notion historique de « contrainte évolutive », une contrainte est une limitation de la variation possible chez un organisme donné, due à la constitution physique et aux mécanismes de développement qui ont été retenus au cours de l'histoire évolutive (Maynard-Smith *et al*, 1985, « Developmental constraints and evolution » ; Wimsatt & Schank, 1988, « Two constraints on the evolution of complex adaptations and the means of their avoidance »). Selon la notion de « contrainte fonctionnelle », certaines structures développementales sont nécessaires car ce sont les seules capables d'assurer certaines fonctions biologiques : la relation de contrainte va donc des fonctions biologiques aux solutions structurales possibles chez les organismes vivants. Cette notion de contrainte n'est pas historique et ne se rattache pas aux lois physiques ou aux principes d'auto-organisation ; de plus, contrairement à la notion de « contrainte physique », elle n'est pas incompatible avec le principe de sélection naturelle qui, au contraire, peut être à l'origine des contraintes fonctionnelles (pour une analyse détaillée de la notion de contrainte fonctionnelle, voir Braillard, 2008, *Enjeux philosophiques de la biologie des systèmes*). À la lumière de la distinction entre ces trois notions de contrainte en biologie, nous suggérons qu'on pourrait considérer les structures et les mécanismes régulant le bruit dans l'expression génétique comme des contraintes évolutives ou comme des contraintes fonctionnelles, tout en niant qu'il s'agit de contraintes physiques.

environnementales externes mais intracellulaires, sont des adaptations au sens où ils sont le résultat du processus de sélection naturelle. Cette hypothèse et l'idée selon laquelle le bruit même peut être considéré comme un trait sujet à sélection, pouvant évoluer, a été soutenue, entre autres, par Fraser et ses collègues²⁶. Ces biologistes ont d'abord procédé à l'estimation du bruit intervenant dans l'expression de chaque gène chez la levure *Saccharomyces cerevisiae*. Ensuite, ils ont testé l'hypothèse selon laquelle le bruit dans l'expression des gènes essentiels, c'est-à-dire dont l'élimination serait létale pour cet organisme, ou des gènes impliqués dans la production des sous-unités de complexes multi-protéiques, est moins important que le bruit dans l'expression de gènes dont le rôle n'est pas vital. Fraser et ses collègues ont montré que, chez la levure *Saccharomyces cerevisiae*, un taux élevé de transcription suivi par un taux bas de traduction ne caractérise que l'expression de ces gènes essentiels, ce qui permet à cet organisme de minimiser les effets du bruit sur la production de protéines dont la variabilité ou l'absence pourraient se révéler défavorables et même létales. Ils ont donc interprété une telle modulation du bruit comme le résultat de l'histoire évolutive des organismes eucaryotes analysés ; de plus, ils ont soutenu que le bruit est très probablement un trait susceptible d'évoluer au cours du temps via le processus de sélection.²⁷

De nombreux biologistes ayant travaillé sur l'origine, les conséquences et le contrôle du bruit dans l'expression génétique ont soutenu que la variation quantitative dans la production des protéines est régulée de manière adaptative non seulement par rapport à la fonction de ces dernières (et donc des gènes impliqués dans leur synthèse), mais aussi en fonction des conditions environnementales que les populations cellulaires (et même d'organismes) ont rencontrées au cours du temps. Le bruit semble donc pouvoir être modulé de manière adaptative relativement aux caractéristiques de l'environnement (par exemple, par rapport à la fréquence avec laquelle l'environnement, et donc la pression de sélection, changent et par rapport à l'importance de ces changements²⁸). En effet, bien que le bruit dans

²⁶ Fraser *et al*, 2004, « Noise Minimization in Eukaryotic Gene Expression ».

²⁷ Newman et ses collègues (Newman *et al*, 2006, « Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise ») ont montré que l'expression des gènes impliqués dans la synthèse des facteurs de réponse aux stress est caractérisée par un bruit intrinsèque élevé alors que l'expression des gènes de croissance et de développement, c'est-à-dire impliqués dans la production des facteurs protéiques essentiels dans ces deux processus, est affectée par peu de fluctuations. De plus, ils ont montré que le bruit dépend des caractéristiques des processus de transcription et de traduction. Sur cette base, ils ont proposé de parler de « l'architecture du bruit biologique » et ils ont aussi avancé l'hypothèse « adaptative » selon laquelle toutes ces différences caractérisant le bruit sont le résultat du processus de sélection naturelle : elles réfléchissent les coûts et les avantages potentiels de la variation quantitative dans la synthèse des protéines.

²⁸ Kussel *et al*, 2005, « Bacterial Persistence: A Model of Survival in Changing Environments » ; Kussel & Leibler, 2005, « Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments » . Voir aussi Dhar & McKinney, 2007, « Microbial phenotypic heterogeneity and antibiotic tolerance » ; Lopez-Maury *et al*, 2008, « Tuning gene expression to changing environments: from rapid responses to evolutionary

l'expression des gènes puisse se révéler nuisible pour le fonctionnement cellulaire, il s'agit d'un phénomène qui, en générant de la variation phénotypique au cours du cycle de vie d'une cellule, peut lui permettre de survivre et de se reproduire en dépit d'un environnement instable. De plus, en tant que source causale de variation phénotypique au sein des populations naturelles, ce bruit peut se révéler avantageux au niveau populationnel : il offre une possibilité supplémentaire de diversification, bien que temporaire et facilement réversible, lorsqu'une population doit faire face à un environnement changeant et adverse (en particulier, en absence de toute variation génétique au sein de la population). Plus précisément, il a été montré qu'un bruit élevé dans l'expression des gènes, avec les structures et les mécanismes permettant son exploitation, tend à caractériser des populations situées dans des conditions environnementales instables et fluctuantes, alors qu'un bruit relativement bas, avec les structures et les mécanismes permettant son contrôle, tend à être une caractéristique des populations dont l'environnement est stable.

Considérons en détail une étude théorique sur ce sujet. Kussel et ses collègues²⁹ ont analysé le mécanisme stochastique de *switch* par lequel des bactéries individuelles et génétiquement identiques d'*Escherichia coli* peuvent passer, de manière réversible et sans qu'aucune mutation génétique ne se produise, d'un phénotype normal sensible aux antibiotiques (dit « non persistant ») à un phénotype résistant à ces derniers (dit « persistant »). Les bases moléculaires de ce mécanisme sont encore inconnues. Selon l'explication la plus accréditée, ce *switch* phénotypique de l'état de non persistance à celui de persistance aux antibiotiques est dû au même type de mécanismes régissant, par exemple, l'induction-répression de l'opéron lactose chez la bactérie *Escherichia coli*³⁰, le choix entre cycle lytique et cycle lysogénique chez le bactériophage lambda³¹ et le *switch* de l'état végétatif à l'état de compétence (et vice versa) chez la bactérie *Bacillus subtilis*³². Plus précisément, son origine résiderait dans la présence de fluctuations intracellulaires au sein de chaque bactérie individuelle et de structures dynamiques diverses permettant leur régulation (par exemple, des boucles de rétroaction négatives, positives ou intégrales). Selon cette explication en attente de confirmation expérimentale, le dépassement d'un certain seuil dans la concentration des facteurs protéiques produits et le *switch* consécutif de l'état de non

adaptation ».

²⁹ Kussel *et al*, 2005, *op. cit.* ; voir aussi Balaban *et al*, 2004, « Bacterial Persistence as a Phenotypic Switch » ; Kussel & Leibler, 2005, *op. cit.* ; Dhar & McKinney, 2007, *op. cit.*

³⁰ Ozbudak *et al*, 2004, « Multistability in the lactose utilization network of *Escherichia coli* » ; Mettetal *et al*, 2006, « Predicting stochastic gene expression dynamics in single cells ».

³¹ Oppenheim *et al*, 2005, « Switches in bacteriophage lambda development ».

³² Pour des références à ce sujet, voir la section 9.1.1 de ce chapitre.

persistance à l'état de persistance (et vice versa) chez la bactérie *Escherichia coli* sont dus au bruit intervenant dans l'expression des gènes impliqués dans la résistance aux antibiotiques (par exemple, les gènes de persistance TA impliqués dans la production des toxines et des antitoxines³³). Ce même bruit intracellulaire serait donc la source causale de la multistabilité phénotypique, c'est-à-dire de la présence de deux phénotypes cellulaires différents (ou davantage) au sein d'une population isogénique d'*Escherichia coli*³⁴. Par conséquent, le bruit dans l'expression des gènes et sa modulation par des structures et des mécanismes intracellulaires seraient à l'origine de sous-populations de bactéries, chaque sous-population étant caractérisée par des états phénotypiques différents (bien que ce soit de manière temporaire car cela dépend de l'expression fluctuante de certains gènes). Une telle variété peut apporter un avantage à la population entière et lui permettre de survivre, du moins sur le court terme³⁵, en dépit de conditions environnementales changeantes et adverses (par exemple, en présence d'antibiotiques).

Kussel et ses collègues ont aussi avancé l'hypothèse « adaptative » selon laquelle le mécanisme physiologique de *switch*, stochastique et réversible, du phénotype normal au phénotype résistant aux antibiotiques, est une adaptation à des environnements qui changent continuellement. Ils ont construit un modèle de la persistance chez la bactérie *Escherichia coli* afin de tester cette hypothèse et, plus précisément, afin de comprendre quand et comment ce mécanisme de *switch* peut conférer un avantage sélectif à la population de bactéries *Escherichia coli* concernée. Leur étude les a amenés à conclure que le taux de *switch* dépend très strictement de la fréquence avec laquelle l'environnement change. Un taux élevé de *switch*, apparemment dû à la présence d'un bruit intracellulaire suffisamment élevé, est sélectionné positivement et se fixe au sein d'une sous partie de la population lorsque cette

³³ Keren *et al*, 2004a, « Persister cells and tolerance to antimicrobials » ; Keren *et al*, 2004b, « Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli* » ; Lewis, 2005, « Persister cells and the riddle of biofilm survival » ; Gerdes *et al*, 2005, « Prokaryotic toxin-antitoxin stress response loci ». Selon d'autres études, le mécanisme de *switch* stochastique de l'état de non persistance à l'état de persistance, et vice versa, chez la bactérie *Escherichia coli* est dû aux fluctuations dans l'expression des gènes impliqués dans la synthèse d'autres facteurs protéiques. Voir Spoering *et al*, 2006, « GlpD and PlsB participate in persister cell formation in *Escherichia coli* » ; Vazques-Laslop *et al*, 2006, « Increased persistence in *Escherichia coli* caused by controlled expression of toxins or other unrelated proteins ».

³⁴ Voir Dhar & McKinney, 2007, « Microbial phenotypic heterogeneity and antibiotic tolerance ».

³⁵ L'« hétérogénéité dynamique » (Thattai & van Oudenaarden, 2004, « Stochastic Gene Expression in Fluctuating Environments ») due au bruit dans l'expression des gènes et caractérisant des populations isogéniques peut apporter un avantage sélectif à la population concernée, non pas sur le long terme mais sur le court terme. En effet, il s'agit d'une variation phénotypique très instable, facilement réversible et qui ne peut pas elle-même être transmise à la génération suivante. Pour qu'une population puisse acquérir un avantage sur le long terme, il est nécessaire que l'« hétérogénéité statistique », c'est-à-dire génétique, ou toute autre sorte de variation pouvant être transmise sur plusieurs générations, prenne le relais de l'« hétérogénéité dynamique » due au bruit intracellulaire. Voir Merlin (à paraître), « On Griffiths and Gray's concept of expanded and diffused inheritance ».

dernière se trouve dans un environnement fluctuant. Un taux plus bas de *switch*, apparemment dû à une réduction du bruit intracellulaire, caractérise plutôt une population située dans un environnement relativement stable. Le taux optimal d'activation de ce mécanisme stochastique de changement phénotypique, et le bruit conséquent à son origine, tendent donc à mimer la distribution des changements de l'environnement.

Le mécanisme stochastique de *switch* d'un état phénotypique à un autre par l'exploitation du bruit intracellulaire semble donc représenter une sorte de « police d'assurance »³⁶ pour une population de cellules dans un environnement instable. Lorsque ce dernier fluctue de manière relativement rapide, il s'agit d'une stratégie de survie préférable au mécanisme physiologique de réponse directe, spécifique et adaptative, aux changements environnementaux³⁷. D'une part, parce qu'elle n'implique aucun dispositif de détection et de réponse, ce qui la rend plus économique ; d'autre part, parce qu'elle peut permettre la préadaptation spontanée de sous-parties d'une population aux conditions environnementales adverses qu'elle pourra peut-être rencontrer ultérieurement.³⁸

9.1.3. Le bruit dans l'expression des gènes est le fruit du « hasard évolutif »

Les dynamiques évolutives à l'origine des caractéristiques spatio-temporelles du bruit et l'hypothèse « adaptative » à ce sujet nous renvoient à la deuxième partie de notre travail, plus précisément à l'analyse du fait que le taux de mutation caractérisant une population donnée peut évoluer au cours du temps relativement à la fonction spécifique de chaque gène et aux conditions environnementales rencontrées.

Il y a deux parallèles forts, l'un entre l'évolution d'un taux peu élevé de mutation au niveau des gènes vitaux et l'évolution d'un niveau peu élevé de bruit dans l'expression des

³⁶ Thattai & Oudenaarden, 2004, *op. cit.*

³⁷ Pour une présentation des recherches actuelles sur les mécanismes de réponse directe et adaptative aux changements de l'environnement, voir Lopez-Maury *et al.*, 2008, « Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation ».

³⁸ Kusser et Leibler (2005, « Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments ») ont procédé à une analyse détaillée, basée sur un modèle mathématique, de l'évolution du taux de *switch* phénotypique par rapport aux caractéristiques dynamiques de l'environnement. Ils ont fourni une évaluation très intéressante des avantages et des désavantages apportés au niveau populationnel respectivement par un mécanisme de *switch* stochastique et par un mécanisme de *switch* en réponse directe aux changements de l'environnement. En particulier, ils ont montré qu'un mécanisme activé de manière spécifique évoluerait dans un environnement qui fluctue très souvent : il serait préférable à un mécanisme stochastique, ce dernier étant moins coûteux mais aussi moins efficace dans des conditions très instables. Par contre, si l'environnement fluctuait mais avec une fréquence relativement basse, le *switch* stochastique serait sélectionné positivement ; de plus, il serait l'objet d'une modulation évolutive par rapport au type de changements environnementaux et à leur fréquence.

gènes essentiels ou impliqués dans la production de complexes multiprotéiques. L'autre parallèle est entre l'augmentation du taux de mutation dans les gènes impliqués dans les réponses au stress et le fait que ces mêmes gènes tendent à fluctuer de manière importante. Ces parallèles subsistent aussi par rapport aux caractéristiques de l'environnement. En effet, de même que pour le taux de mutation, le niveau de bruit tend à évoluer vers sa propre diminution lorsqu'une population se trouve dans des conditions environnementales stables, alors qu'il tend à augmenter dans un environnement fluctuant.

Cela suggère que les mécanismes mutateurs contrôlant le taux de mutation et les structures dynamiques régulant le bruit dans l'expression des gènes sont le résultat de l'histoire évolutive des populations naturelles. De plus, les deux hypothèses « adaptatives » respectives concernant l'origine de ces mécanismes de régulation des mutations et du bruit consistent à supposer exactement la même chose à propos de ces deux sources causales de variation biologique. Plus précisément, l'organisation structurale et dynamique modulant le bruit dans l'expression des gènes doit probablement son caractère adaptatif au processus d'évolution par sélection naturelle. Cela explique pourquoi la variation dans le nombre de protéines produites (au sein d'une population isogénique à un instant donné ou dans une même cellule au cours du temps) est différente selon le gène concerné et selon les conditions de l'environnement. Le bruit intracellulaire est informé au sujet du rôle fonctionnel des gènes et des caractéristiques des conditions environnementales, non pas de manière directe, mais par le passé évolutif de la population concernée.

Le parallèle entre la régulation du niveau de bruit et le contrôle du taux de mutation nous amène à conclure que, du point de vue évolutionnaire, le bruit affectant l'expression des gènes n'est pas le fruit du hasard au sens où il serait complètement indépendant du processus de sélection et d'évolution de la population concernée. Au contraire, il s'agit de fluctuations dont les propriétés spatio-temporelles sont biaisées par l'organisation structurale et dynamique de la cellule, c'est-à-dire par la présence de structures et de mécanismes cellulaires qui doivent leur origine à l'histoire évolutive, très probablement adaptative, de la population cellulaire concernée. Il s'agit donc d'un « bruit biaisé » ou, selon l'expression utilisée par Rao et ses collègues, d'un « bruit cultivé »³⁹. Tout cela est parfaitement compatible avec la notion de « hasard évolutionnaire » que nous avons introduite dans la deuxième partie de notre travail : cette notion se révèle en effet être des plus pertinentes pour

³⁹ Rao *et al*, 2002, « Control, exploitation and tolerance of intracellular noise ».

caractériser non seulement toute mutation génétique, mais aussi le bruit dans l'expression génétique selon une perspective évolutionnaire.

Rappelons donc la notion de « hasard évolutionnaire » utilisée dans le cas du phénomène de mutation génétique et regardons si elle s'applique au bruit intervenant dans l'expression des gènes.

Dans les chapitres 4 et 6, nous avons affirmé qu'une mutation est le fruit du « hasard évolutionnaire » si et seulement si elle n'est pas causée de manière spécifique et (exclusivement) adaptative par un processus physico-chimique en réponse aux conditions de l'environnement. Cela signifie que si une mutation est causée par un processus physico-chimique sans que l'environnement joue aucun rôle causal dans sa production, ou si une mutation est causée par un processus physico-chimique en réponse aux conditions de l'environnement mais non pas de manière spécifique, ou si une mutation est causée de manière spécifique par un processus physico-chimique en réponse aux conditions de l'environnement mais pas de manière (exclusivement) adaptative, alors elle n'est pas « dirigée » en vue de l'adaptation.

Nous avons aussi précisé qu'une mutation peut être considérée comme le fruit du « hasard évolutionnaire » si et seulement si elle ne remplit pas au moins une des deux conditions qu'une mutation « dirigée » remplit : 1) être plus probable dans un environnement E_1 où elle est avantageuse que dans un environnement E_2 où elle est défavorable ou neutre ; 2) être nettement plus probable dans un environnement E_1 où elle est avantageuse que d'autres mutations neutres ou défavorables. Par conséquent, si une mutation est plus probable dans un environnement E_1 où elle est avantageuse que dans un environnement E_2 où elle est défavorable ou neutre mais, que dans le même temps, elle n'est pas nettement plus probable dans un environnement donné E_1 où elle est avantageuse que d'autres mutations neutres ou défavorables, elle peut être considérée comme le fruit du « hasard évolutionnaire ».

À la lumière de notre analyse sur le biais du bruit dans l'expression génétique, la pertinence conceptuelle et empirique de la notion de « hasard évolutionnaire » apparaît clairement.

Considérons le bruit dans l'expression d'un gène, soit chez une population d'individus génétiquement identiques à un instant donné, soit chez un même individu au cours d'un certain laps de temps. On ne peut l'identifier qu'en termes quantitatifs et statistiques, par exemple en calculant le coefficient de variation (c'est-à-dire, l'écart type sur la moyenne synchronique des protéines produites par la population isogénique ou l'écart type sur la moyenne diachronique des protéines produites au sein de l'individu considéré). De manière

plus claire que dans le cas des mutations génétiques, ces fluctuations ne sont pas causées de manière spécifique et (exclusivement)⁴⁰ adaptative par un processus physico-chimique en réponse aux conditions de l'environnement (externe ou extracellulaire). Explicitons les raisons de notre affirmation ci-dessous.

Premièrement, le bruit dans l'expression génétique est dû au caractère discret et probabiliste des réactions biochimiques qui, à son tour, résulte de plusieurs facteurs physico-chimiques : non homogénéité du microenvironnement intracellulaire (contrairement à la manière dont il est décrit dans la plupart des modèles mathématiques de l'expression génétique) ; mouvement incessant et erratique des molécules du milieu intracellulaire ; présence d'un petit nombre de molécules à l'intérieur de la cellule ; événements de nature quantique. On peut donc affirmer que le bruit intracellulaire est le résultat d'un processus physico-chimique hétérogène, au sens où des facteurs divers interviennent dans sa réalisation.

Deuxièmement, ce processus n'est pas activé en réponse directe aux conditions environnementales ; cependant, l'environnement peut y jouer un rôle causal, du moins selon une perspective évolutionnaire. En effet, nous avons vu que le bruit dans l'expression des gènes peut faire l'objet d'une régulation relativement aux caractéristiques des conditions environnementales rencontrées par une population au cours du temps. C'est donc dans ce sens que l'environnement peut être considéré comme une cause des caractéristiques des fluctuations intervenant dans l'expression des gènes.

Troisièmement, le bruit peut être causé de manière spécifique : comme dans le cas des mutations génétiques faisant partie d'une augmentation locale du taux de mutation, des fluctuations peuvent faire partie d'une augmentation locale du bruit, c'est-à-dire qui ne concerne pas le génome entier mais l'expression d'un ou de certains de ses gènes. Nous rappelons que le caractère spécifique de certaines fluctuations est dû à la présence de structures et de mécanismes intracellulaires qui biaisent le bruit en le réduisant ou en l'amplifiant. Il s'agit d'une organisation structurale et dynamique qui est le résultat de l'histoire évolutive se déroulant au niveau populationnel. De plus, selon l'hypothèse « adaptative », cette organisation doit son origine au processus de sélection naturelle : en augmentant le bruit au niveau des gènes impliqués dans les réponses aux stress et en le diminuant au niveau des gènes essentiels, elle est susceptible d'apporter un avantage aux populations concernées.

⁴⁰ Pour l'interprétation du terme « exclusivement », nous renvoyons à la note 58 dans le chapitre 4.

Quatrièmement, le bruit lié à l'expression des gènes n'est pas causé de manière (exclusivement) adaptative. Bien qu'il puisse être plus ou moins élevé – donc, plus ou moins probable à niveau égal – le long de la séquence d'ADN et selon les conditions de l'environnement au cours du temps, et bien qu'une telle régulation soit très probablement le résultat adaptatif de l'évolution au niveau populationnel, ce bruit n'est pas produit en vue de l'adaptation, ni de la cellule individuelle ni de la population concernée. De plus, la variation phénotypique qu'il engendre n'étant pas transmise à la génération suivante, on peut évacuer l'idée d'un processus lamarckien à son origine.

Nous suggérons donc d'utiliser la notion de « hasard évolutionnaire » pour caractériser cette source causale de la variation phénotypique. En effet, par exemple dans le cas de l'expression d'un gène chez un même individu au cours d'un certain laps de temps, il n'y a pas de relation spécifique qui va de l'ordre des causes à l'ordre des effets entre, d'une part, la valeur adaptative du bruit dans l'expression de ce gène dans un certain environnement, et d'autre part, le fait que le bruit affecte l'expression de ce gène lorsque l'individu se trouve dans cet environnement. En termes probabilistes, il n'y a aucun lien causal et spécifique qui aille de la probabilité que le bruit soit favorable (dans un environnement donné) à la probabilité qu'il se produise (dans cet environnement).

On peut arriver à cette même conclusion en montrant que le bruit dans l'expression d'un gène, par exemple chez un même individu au cours du temps, ne remplit pas au moins une des deux conditions pour être « dirigé » par rapport au cours adaptatif de l'évolution. Autrement dit, on doit montrer qu'un certain niveau de bruit (1) n'est pas plus probable dans un environnement E1 où il est avantageux que dans un environnement E2 où il est défavorable ou neutre ET/OU (2) n'est pas nettement plus probable dans un environnement E1 où il est avantageux que d'autres fluctuations neutres ou défavorables.

D'abord, le bruit dans l'expression d'un gène peut être plus probable dans des conditions environnementales de stress (E1) – où il peut apporter un avantage à la fois au niveau individuel et au niveau populationnel – que dans un environnement stable (E2) – où il risque de perturber le fonctionnement cellulaire et la survie de la population concernée. Cependant, un tel biais n'est pas le résultat d'un processus physico-chimique en réponse directe aux défis posés par l'environnement ; au contraire, il est dû à la présence de structures et de mécanismes qui ont évolué, probablement de manière adaptative, par rapport aux caractéristiques des environnements rencontrés par une population au cours du temps.

En second lieu, le bruit dans l'expression d'un gène n'est pas nettement plus probable dans des conditions adverses (E1) où il est avantageux par rapport à d'autres fluctuations

défavorables ou neutres (par exemple, par rapport au bruit affectant l'expression d'un gène vital pour le fonctionnement cellulaire ou d'un gène fonctionnellement peu important). Cela ne signifie pas que toute fluctuation possible, qu'elle soit favorable, défavorable ou neutre, ait la même probabilité de se produire dans un certain environnement. Au contraire, il peut y avoir un lien, du moins statistique, entre la valeur adaptative du bruit affectant l'expression génétique et l'environnement où l'individu concerné se trouve (donc, avec les besoins adaptatifs de ce dernier). D'une part, cela s'explique par le fait que la valeur adaptative du bruit dépend strictement des conditions environnementales, et donc de la pression de sélection à laquelle la cellule concernée doit faire face ; d'autre part, par le fait que s'exercent des biais divers qui influencent le bruit dans l'expression des gènes et dont l'origine réside dans l'histoire évolutive, probablement adaptative, de la population concernée.

Nous pouvons donc conclure que la notion de « hasard évolutif » est une notion de hasard très appropriée, à la fois du point de vue conceptuel et empirique, pour caractériser cette source causale de la variation phénotypique qu'est le bruit dans le processus d'expression génétique.

9.2. Le bruit et la notion « aléatoire (au sens) faible »

Dans ce chapitre, il nous reste à voir si le bruit lié à l'expression génétique peut être considéré comme un phénomène « aléatoire (au sens) fort » ou « aléatoire (au sens) faible ». Nous posons donc la question de savoir si ces fluctuations sont dues à un processus stochastique qui intervient de la même manière dans l'expression de tout gène ET qui est invariant dans le temps, ou à un processus stochastique qui affecte différemment les gènes d'un génome donné ET/OU qui varie dans le temps.

En premier lieu, nous mettrons en évidence le décalage entre, d'une part, les hypothèses des modèles stochastiques et hautement idéalisés de l'expression des gènes, d'autre part, ce qu'on sait actuellement au sujet d'un certain nombre de facteurs biochimiques pouvant influencer le bruit le long du génome et dans le temps (par exemple, au cours d'une génération cellulaire)⁴¹. Ensuite, nous montrerons sur cette base que la notion conceptuellement et empiriquement appropriée pour caractériser le bruit affectant

⁴¹ En ce qui concerne les structures et les mécanismes régulant le bruit dans l'expression des gènes, nous renvoyons à la section 9.1.1 de ce chapitre.

l'expression des gènes est, de même que dans le cas des mutations, la notion d'« aléatoire (au sens) faible ».

9.2.1. Les limites des modèles de l'expression des gènes

Le bruit dans l'expression des gènes, c'est-à-dire la variation autour de la moyenne de l'abondance de protéines produites (au sein d'une population isogénique à un instant donné ou dans une même cellule au cours du temps) est fondamentalement dû au fait que les collisions moléculaires nécessaires à la production des réactions biochimiques sont des événements probabilistes. C'est pourquoi les biologistes ont commencé à utiliser des modèles mathématiques stochastiques (notamment, les équations matrices, de Fokker-Planck, de Langevin) pour décrire et prédire, en termes d'abondance de protéines produites, l'expression des gènes.⁴²

Nous rappelons que ces modèles stochastiques partagent les idéalizations suivantes : les collisions moléculaires sont élastiques, c'est-à-dire qu'elles n'entraînent aucune transformation des propriétés des molécules ; les vitesses des molécules suivent une distribution de Maxwell-Boltzmann ; l'environnement intracellulaire est homogène et bien mélangé ; les molécules sont distribuées de manière aléatoire à l'intérieur de la cellule (autrement dit, la probabilité qu'une molécule se trouve à un certain endroit est uniforme dans le volume unicellulaire) ; la probabilité de collision et de réaction est indépendante de la localisation des molécules dans la cellule. De plus, ces modèles sont presque tous focalisés sur les mêmes phénomènes de base et c'est donc en ces termes qu'ils essayent d'évaluer le bruit intracellulaire et de rendre compte de son impact sur l'expression des gènes.

Les étapes systématiquement prises en considération dans la modélisation stochastique de l'expression génétique sont l'activation des gènes, les processus de transcription et de traduction, la dégradation des transcrits ARN messagers et la dégradation des protéines.

⁴² Pour plus de détails sur les caractéristiques de ces descriptions mathématiques, nous renvoyons à la section 8.6 du chapitre 8.

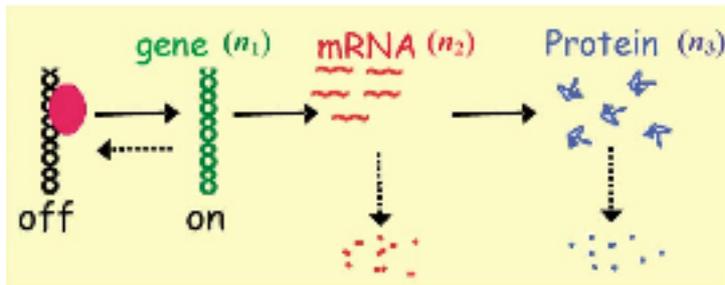


Figure 11 : Dessin décrivant l'activation génétique, la transcription, la traduction, la dégradation des ARN messagers et des protéines. Figure tirée de Paulsson, 2005, « Models of stochastic gene expression ».

La description de chacune de ces étapes permet de rendre compte des fluctuations d'origines diverses pouvant contribuer au bruit dans ce processus intracellulaire, notamment les fluctuations dans le nombre de gènes actifs, dans le nombre de molécules d'ARN messager et dans le nombre de protéines dans le milieu intracellulaire. Le choix de se limiter à la description de ces phénomènes moléculaires est controversé car il implique de ne rendre compte que du bruit intrinsèque – c'est-à-dire des fluctuations stochastiques inhérentes aux étapes du processus d'expression génétique. Le bruit extrinsèque – à savoir les fluctuations concernant d'autres facteurs moléculaires qui ne sont pas considérés comme des composantes de l'expression des gènes mais qui peuvent l'influencer – n'est pas pris en compte. La variation quantitative observée au niveau des protéines produites peut au contraire trouver son origine dans des phénomènes intracellulaires considérés comme des sources de bruit extrinsèque (par exemple, les fluctuations dans la concentration et dans la distribution des ribosomes, des ADN et des ARN polymérases, des RNAses et des protéases, ainsi que les fluctuations concernant toutes les enzymes qui jouent un rôle central dans ce processus intracellulaire ; les effets divers du cycle cellulaire ou, en ce qui concerne le bruit protéique au niveau populationnel, le caractère asymétrique des divisions cellulaires et la partition non équitable des composants moléculaires qui s'ensuit).

En suivant la première partie d'un article éclairant où Paulsson⁴³ met en évidence les aspects communs à la plupart des modèles stochastiques de l'expression des gènes, nous allons analyser la manière dont chaque étape de ce processus est caractérisée en termes mathématiques, plus spécifiquement probabilistes, au sein de ces descriptions théoriques. Notre but est de montrer que les modèles stochastiques de l'expression des gènes ne prennent pas en compte de nombreux facteurs qui peuvent avoir un impact sur le déroulement de l'expression génétique et sur son résultat final, et qui peuvent aussi biaiser le bruit qui l'affecte. Nous concluons que ces modèles ne sont pas appropriés dans la mesure où ils

⁴³ Paulsson, 2005, « Models of stochastic gene expression ».

décrivent le caractère stochastique de l'expression des gènes d'une manière trop approximative et idéalisée.⁴⁴

Paulsson présente dans l'ordre suivant les phénomènes moléculaires pris en compte par les divers modèles stochastiques de l'expression des gènes. D'abord, il y a l'étape d'activation des gènes. Il s'agit d'un processus probabiliste pouvant avoir des causes moléculaires multiples (entre autres, la dissociation des facteurs de répression, l'association des facteurs d'activation, le remodelage de la chromatine, etc.). En première approximation, l'activation des gènes est décrite de la même manière, chez les organismes procaryotes, chez les organismes eucaryotes, et pour tout gène, comme un *random telegraph process*, c'est-à-dire comme un processus stochastique markovien (sans mémoire)⁴⁵, continu dans le temps et caractérisé par deux valeurs distinctes : chaque gène passe de manière indépendante de l'état « *off* » (éteint ou réprimé) à l'état « *on* » (allumé ou activé) et vice versa avec un certain taux (λ_1^+ étant le taux d'activation et λ_1^- étant le taux de répression). En outre, l'activation et la répression des gènes sont supposées être indépendantes de la croissance cellulaire et du processus de réplication génétique. Enfin, à l'état stationnaire, le nombre de gènes actifs est distribué selon une loi binomiale⁴⁶ où la probabilité pour n'importe quel gène d'être dans l'état « *on* » (d'être activé et donc de s'exprimer) est $P_{on} = \lambda_1^+ / \lambda_1^+ + \lambda_1^-$. Une telle description de l'activation des gènes n'est pas très réaliste. D'abord, elle attribue à tout gène et à tout organisme les mêmes propriétés relativement à ce processus moléculaire. Deuxièmement, elle ne rend pas compte du fait que l'activation de certains gènes peut dépendre de la présence de produits protéiques dus à l'activation et à l'expression d'autres gènes qui en découle.

D'autres phénomènes moléculaires sont pris en compte et décrits par les modèles stochastiques de l'expression des gènes : le processus de transcription de la séquence d'ADN

⁴⁴ Notre critique des approximations et des idéalisations sur lesquelles se fondent les modèles stochastiques de l'expression des gènes ne vise nullement à mettre en cause, ni la valeur scientifique des recherches sur le bruit, ni la légitimité des modèles construits dans ce cadre. En effet, un modèle ne peut pas fournir une description qui soit à la fois générale, réaliste et précise d'un phénomène donné (voir Levins, 1966, « The Strategy of Model Building in Population Biology »). Le choix de sacrifier l'un de ces trois desiderata dépend du projet à la base de la modélisation d'un phénomène. Quant à l'étude du bruit intracellulaire, les biologistes visent à identifier son origine au niveau des étapes du processus d'expression génétique, à comprendre la manière dont une cellule peut réduire et exploiter ce bruit, et à prédire les conséquences de ce dernier au niveau de la variation individuelle et populationnelle, à la fois en termes moléculaires et en termes adaptatifs. Ce projet a comme objet un système dynamique physico-chimique qui est hautement complexe et qui est encore en partie inconnu. Dans l'état de nos connaissances, il n'est pas envisageable de parvenir à construire un modèle de l'expression des gènes qui soit à la fois précis (du point de vue quantitatif) et réaliste. C'est l'une des raisons pour lesquelles les biologistes en construisent des modèles quantitatifs et stochastiques mais généraux, donc approximatifs et idéalisés, qui sont plus faciles à traiter par rapport à des modèles précis et réalistes.

⁴⁵ Voir le Glossaire.

⁴⁶ *Id.*

en séquences d'ARN et le processus de traduction des séquences d'ARN messenger en protéines. Ils sont décrits comme des processus de Poisson⁴⁷ où la probabilité de production des transcrits d'ARN (le taux de transcription λ_2) est proportionnelle au nombre de gènes actifs (n_1) par unité de temps ($\lambda_2 n_1$) et la probabilité de production des protéines (le taux de traduction λ_3) l'est au nombre d'ARN messagers (n_2) par unité de temps ($\lambda_3 n_2$). Les taux de transcription et de traduction sont considérés comme constants au cours du temps. Mais il faut bien reconnaître qu'une telle description de la transcription et de la traduction n'est probablement pas une bonne approximation de ce qui se passe réellement à l'intérieur de la cellule, car de nombreux phénomènes moléculaires sont susceptibles d'affecter les taux de réalisation de ces processus. Donnons-en quelques exemples. D'abord, le processus de réplication de la séquence d'ADN peut provoquer un changement dans la structure de la chromatine : il peut avoir des conséquences sur la présence de facteurs d'activation et de répression de l'expression génétique, et peut donc modifier les taux de synthèse des protéines. Ensuite, le fait qu'une ARN polymérase se lie à la séquence d'ADN au niveau d'un gène donné peut changer la structure de ce dernier et, de cette manière, peut bloquer ou faciliter sa transcription. Enfin, le transport des ARN messagers hors du noyau chez les eucaryotes et la libération des transcrits d'ARN dans le milieu intracellulaire chez les procaryotes peuvent aussi affecter les dynamiques de l'expression génétique.

Les modèles stochastiques de l'expression des gènes incluent aussi la description des processus de dégradation des transcrits d'ARN messenger et des protéines. Une telle description se base sur les deux hypothèses suivantes : premièrement, chaque événement de dégradation est indépendant et sans mémoire ; deuxièmement, les molécules d'ARN et les protéines ont un temps de vie distribué exponentiellement⁴⁸. Cependant, il s'agit d'hypothèses restrictives qui font de ces modèles des descriptions non réalistes de l'expression des gènes. En effet, la dégradation des molécules diverses d'ARN messenger – ou la dégradation des protéines – peut dépendre de la compétition pour les enzymes de dégradation (les RNAses et les protéases) présentes à l'intérieur de la cellule. De plus, ces enzymes opèrent dans un état proche de la saturation : les processus de dégradation dépendent donc de la concentration du substrat qui, à son tour, peut changer pour des raisons diverses et de manière stochastique au cours du temps.

Les modèles stochastiques décrivant l'expression génétique par les étapes d'activation des gènes, de transcription, de traduction et de dégradation des ARN et des protéines,

⁴⁷ Voir le Glossaire.

⁴⁸ *Id.*

permettent d'évaluer le bruit intrinsèque (plus précisément, de calculer la variance normalisée⁴⁹ dans le nombre de protéines par cellule) : il est égal à la somme des bruits intervenant dans les étapes considérées comme constitutives de ce processus. Cette manière d'estimer le bruit intrinsèque ne peut également être considérée comme appropriée car les fluctuations liées aux différentes étapes ne sont pas indépendantes entre elles. Par exemple, les fluctuations dans l'activation des gènes influencent celles dans les processus de transcription et de traduction qui, à leur tour, peuvent influencer le nombre d'enzymes de dégradation présentes à l'intérieur de la cellule et donc les fluctuations dans la dégradation des ARN messagers et des protéines. De plus, nous soulignons à nouveau que le fait de ne prendre en considération que le bruit intrinsèque, et non pas le bruit extrinsèque, est une approximation regrettable. En effet, il a été montré, d'une part, que ces deux sources de bruit protéique ne sont pas indépendantes (plus précisément, que le bruit intrinsèque peut être influencé par le bruit extrinsèque) ; d'autre part, que le bruit extrinsèque représente, à la fois chez les procaryotes et chez les eucaryotes, la source principale des fluctuations dans l'expression des gènes.⁵⁰

Un dernier aspect de la modélisation stochastique de l'expression des gènes mérite d'être souligné. Ces modèles prédisent que les protéines sont produites par explosions (*bursts*) distribuées de manière aléatoire au cours du temps (par exemple, au cours d'une génération cellulaire). Plus précisément, ils prévoient que de petites quantités de protéines sont produites durant des périodes de temps assez longues et que le taux de leur synthèse devient très élevé (la taille des explosions en termes de nombre de protéines produites étant aussi considérée comme aléatoire) pendant des périodes très brèves. Ces explosions protéiques (aussi bien que les explosions dans la production des ARN messagers) ont été observées chez divers organismes (par exemple, chez la bactérie *Escherichia coli*, chez l'amibe social *Dictyostelium* et même chez des lignées cellulaires des mammifères⁵¹). Les biologistes impliqués dans la modélisation de l'expression des gènes ont ainsi sur quelques points annoncé la confirmation des hypothèses sur lesquelles se fondent les descriptions stochastiques de ce processus, notamment en ce qui concerne la distribution du nombre de gènes actifs, les taux de transcription et de traduction et le processus de dégradation des ARN et des protéines. Néanmoins, de telles conclusions nous semblent précipitées. En effet, comme nous venons de

⁴⁹ Nous renvoyons à la note 52 du chapitre 8.

⁵⁰ Elowitz *et al*, 2002, « Stochastic Gene Expression in a Single Cell » ; Blake *et al*, 2003, « Noise in eukaryotic gene expression » ; Raser & O'Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control ».

⁵¹ Voir Maheshri & O'Shea, 2007, « Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression ».

le montrer, la plupart des modèles ne prennent pas en considération de nombreux facteurs moléculaires pouvant affecter l'expression des gènes (par exemple, toute source de bruit extrinsèque) et la décrivent de manière approximative et idéalisée. La correspondance globale et approximative entre les prévisions de ces modèles mathématiques et les observations empiriques n'implique pas nécessairement que les premiers décrivent correctement les mécanismes véritablement en jeu à l'intérieur de la cellule ; au contraire, il pourrait s'agir d'une simple coïncidence.

9.2.2. Le bruit « aléatoire (au sens) faible » dans l'expression des gènes

Dans la section précédente, nous avons souligné que, contrairement aux hypothèses des modèles stochastiques de l'expression des gènes, de nombreux facteurs physico-chimiques peuvent influencer le déroulement de l'expression génétique et biaiser les propriétés spatio-temporelles du bruit. Dans le présent chapitre, nous avons aussi montré que le bruit intracellulaire est régulé par des structures et des mécanismes divers en fonction du gène concerné et relativement aux conditions environnementales rencontrées. Sur cette base, nous nous proposons de répondre à la question suivante : les fluctuations (ou bruit) dans le processus d'expression génétique⁵² sont-elles des événements « aléatoires (au sens) fort » ou « aléatoires (au sens) faible », selon le sens que nous avons attribué à ces notions⁵³ ? Rappelons-les dans leur formulation générale et par analogie avec l'opération d'échantillonnage.

Un événement « aléatoire (au sens) fort » est le résultat d'un processus d'échantillonnage indiscriminé avec remplacement, c'est-à-dire analogue à une série de tirages de boules dans une urne, où toutes les boules ont la même probabilité d'être tirées et où la boule tirée est systématiquement remise dans l'urne. Au contraire, un événement « aléatoire (au sens) faible » est le résultat d'un processus d'échantillonnage discriminé et/ou sans remplacement, c'est-à-dire analogue à une série de tirages de boules dans une urne, où certaines boules ont une probabilité plus grande que d'autres d'être tirées à cause de leurs propriétés physiques (par exemple, certaines seront collantes tandis que d'autres seront glissantes) ET/OU où la boule tirée n'est pas remise dans l'urne.

⁵² Nous rappelons qu'on mesure le bruit, soit relativement à la moyenne synchronique de protéines produites au sein d'une population isogénique, soit relativement à la moyenne diachronique dans une même cellule.

⁵³ Nous renvoyons à la section 7.1 du chapitre 7.

Revenons au bruit dans l'expression des gènes et à sa modélisation. Avant tout, nous trouvons intéressant de souligner à nouveau le caractère restrictif des hypothèses sur lesquelles se fondent la plupart des modèles stochastiques de ce processus. Premièrement, les collisions entre macromolécules ayant lieu à l'intérieur d'une cellule ne sont pas élastiques, mais peuvent entraîner des transformations des propriétés moléculaires (notamment, de la forme des molécules). Deuxièmement, les vitesses des macromolécules dans le volume cellulaire ne suivent pas véritablement une distribution de Maxwell-Boltzmann : d'une part, contrairement au cas d'un gaz parfait en équilibre thermodynamique, on ne peut pas supposer que dans le milieu intracellulaire aucune direction des vitesses n'est privilégiée ; d'autre part, contrairement au cas des molécules d'un gaz parfait ou à des molécules de même nature et de même taille, les macromolécules présentes à l'intérieur de la cellule ne sont pas identiques. Troisièmement, l'environnement intracellulaire n'est pas homogène et bien mélangé ; au contraire, la cellule est subdivisée en plusieurs compartiments remplis de manière non homogène par des macromolécules différentes.⁵⁴ Quatrièmement, la probabilité qu'une macromolécule se trouve à un certain endroit n'est pas uniforme dans le volume cellulaire à cause de son organisation en compartiments et de la présence de nombreux obstacles de nature moléculaire. Cinquièmement, une telle organisation du volume intracellulaire ne laisse pas les macromolécules se déplacer de manière libre mais contraint leur diffusion ; par conséquent, la probabilité de collision et de réaction n'est pas indépendante de leur localisation.

Le caractère idéalisé de ces hypothèses représente un premier élément contre le caractère « aléatoire (au sens) fort » des événements physico-chimiques se produisant à l'intérieur de la cellule lors de l'expression des gènes. Autrement dit, ces événements ne sont pas analogues aux résultats de l'échantillonnage, répété et avec remplacement, d'une urne contenant un certain nombre de boules identiques et donc indiscernables (c'est-à-dire d'un processus d'échantillonnage indiscriminé et invariant dans le temps). Au contraire, leur occurrence est influencée et favorisée par certaines caractéristiques physico-chimiques du milieu intracellulaire et des facteurs divers qu'il contient.

En second lieu, il nous faut reprocher aux autres hypothèses propres à la plupart des modèles stochastiques de l'expression des gènes leur caractère approximatif et idéalisé, et cela d'autant plus qu'elles concernent des étapes majeures de ce processus intracellulaire.

⁵⁴ Dans le chapitre 8, nous avons déjà remarqué que l'hypothèse simplificatrice d'un environnement intracellulaire homogène et bien mélangé est de plus en plus controversée. Voir Turner *et al*, 2004, « Stochastic approaches for modelling in vivo reactions » ; El Samad *et al*, 2005, « Stochastic modelling of gene regulatory networks ».

Nous avons affirmé plus haut que, contrairement à ce qu'on suppose dans le cadre de ces modèles, les taux des processus d'activation des gènes, de transcription, de traduction, de dégradation des transcrits d'ARN et de dégradation des protéines ne sont pas les mêmes dans tout organisme et dans le cas de tout génome ; plus précisément, ils varient selon le gène, la molécule d'ARN et la protéine concernés. De plus, les taux de ces processus moléculaires ne sont pas constants au cours du temps, mais changent selon les conditions physico-chimiques du milieu intracellulaire à des instants temporels différents. Nous avons aussi souligné que le bruit intrinsèque, dont la plupart des modèles rendent compte, n'est pas indépendant du bruit extrinsèque qui, au contraire, n'est pas considéré dans ce cadre. Ce dernier influence le bruit intrinsèque d'une manière tellement complexe et encore largement inconnue qu'il ne peut pas simplement lui être additionné.

Le caractère approximatif et idéalisé des descriptions mathématiques de l'expression des gènes nous fournit un deuxième argument contre le caractère « aléatoire (au sens) fort » des événements physico-chimiques d'activation, de transcription et de traduction des gènes, de dégradation des molécules d'ARN et de dégradation des protéines. En effet, aucun d'entre eux n'est dû à un processus physico-chimique invariant le long du génome et au cours du temps (par exemple, au cours d'une génération cellulaire).

Nous pouvons enfin montrer que le bruit affectant l'expression des gènes n'est pas (un phénomène) « aléatoire (au sens) fort ». En ce qui concerne ses propriétés spatiales, le bruit diffère selon le gène exprimé et donc la protéine synthétisée. Plus précisément, ces différences ne dépendent pas seulement du rôle fonctionnel du gène, et donc de la protéine qu'il produit, au sein de la cellule⁵⁵, mais aussi de la localisation physique du gène concerné le long du génome (qui dépend à son tour de l'état de la chromatine) et dans le volume cellulaire⁵⁶. En effet, il a été observé, d'une part, qu'il existe des corrélations entre les fluctuations dans l'expression des gènes contigus situés sur le même chromosome ; d'autre part, que le bruit intervenant dans chaque étape de ce processus intracellulaire est relatif à la concentration de facteurs protéiques divers au sein de la cellule (par exemple, les facteurs d'inhibition et d'activation des gènes, les ADN et les ARN polymérases, les RNAses et les protéases, les ribosomes, etc.). Il a aussi été observé que le bruit est caractérisé par des

⁵⁵ Fraser *et al*, 2004, « Noise Minimization in Eukaryotic Gene Expression ».

⁵⁶ Maheshri & O'Shea, 2007, « Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression » ; Becksei *et al*, 2005, « Contributions of low molecule number and chromosomal positioning to stochastic gene expression » ; Raj *et al*, 2006, « Stochastic mRNA synthesis in mammalian cells ».

niveaux différents en dépendance de la constitution nucléotidique du gène concerné et, plus particulièrement, de son promoteur⁵⁷.

Du point de vue de ses propriétés temporelles, le bruit change selon les conditions physico-chimiques caractérisant le milieu intracellulaire à des instants différents dans le temps. Par exemple, il dépend de la manière dont les macromolécules impliquées dans le processus d'expression génétique sont distribuées dans le volume intracellulaire, de leur concentration au sein de la cellule, de l'état de la chromatine, de l'activation ou de la répression des mécanismes de rétroaction pouvant le réguler, le réduire ou l'amplifier, de la présence ou absence de certaines structures dans les réseaux de gènes et de protéines impliqués dans l'expression génétique, etc.⁵⁸

Sur cette base, nous concluons que les fluctuations affectant l'expression des gènes, et qu'on mesure en termes de variation dans le nombre de protéines produites au sein d'une population isogénique à un instant donné ou dans une même cellule au cours du temps, ne sont pas des événements « aléatoires (au sens) fort » mais « aléatoires (au sens) faible ». Afin d'explicitier cette affirmation, procédons à la déclinaison de cette deuxième notion de l'aléatoire dans ce cas.

Considérons le bruit dans l'expression d'un gène chez un même individu au cours d'un certain laps de temps (par exemple, sur une génération de la cellule individuelle porteuse). Nous soutenons qu'il s'agit d'un événement « aléatoire (au sens) faible » car il n'est pas dû à un processus stochastique⁵⁹ physico-chimique (causal) qui remplit les deux conditions de la notion d'« aléatoire (au sens) fort » ($C1 \wedge C2$)⁶⁰. Au contraire, il peut s'agir de l'un des trois processus stochastiques suivants :

P1 : un processus d'échantillonnage discriminé par rapport aux différences dans les caractéristiques physico-chimiques des gènes le long du génome⁶¹ ET indiscriminé par

⁵⁷ Maheshri & O'Shea, 2007, *op. cit.* : Raser & O'Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control ».

⁵⁸ Voir Maheshri & O'Shea, 2007, *op. cit.*

⁵⁹ Nous rappelons que nous définissons un processus stochastique comme un processus qui, à partir des mêmes conditions initiales, peut produire des résultats divers selon une certaine loi de probabilité. De plus, nous soulignons à nouveau que nous ne soutenons aucune position particulière quant à la nature déterministe ou indéterministe, d'un point de vue métaphysique, de ce qu'on appelle un processus stochastique et, en particulier, du processus physico-chimique à l'origine des fluctuations (ou bruit) dans le niveau d'expression génétique.

⁶⁰ C1 : Être un processus d'échantillonnage indiscriminé où nulle différence dans les caractéristiques des éléments pouvant être échantillonnés ne joue un rôle causal. La probabilité d'être échantillonné est donc la même pour tout élément. C2 : Être un processus invariant dans le temps. La probabilité qu'un élément soit échantillonné est donc la même à tout instant, ce qui implique son indépendance de tout autre événement (toute combinaison d'événements incluse).

⁶¹ Ces différences jouent un rôle causal dans le processus à l'origine des fluctuations dans l'expression des gènes. La probabilité qu'une fluctuation ait lieu n'est donc pas la même au niveau de tout gène d'un génome donné.

rapport aux différences dans les caractéristiques physico-chimiques de l'environnement intracellulaire au cours du temps ;

P2 : un processus d'échantillonnage indiscriminé par rapport aux différences dans les caractéristiques physico-chimiques des gènes le long du génome ET discriminé par rapport aux différences dans les caractéristiques de l'environnement intracellulaire au cours du temps⁶² ;

P3 : un processus d'échantillonnage discriminé à la fois par rapport aux différences dans les gènes le long du génome ET par rapport aux différences dans l'environnement intracellulaire au cours du temps.

En résumé, le processus à l'origine du bruit dans le niveau d'expression génétique peut être analogue (P1) à un processus d'échantillonnage discriminé avec remplacement, (P2) à un processus d'échantillonnage indiscriminé sans remplacement, ou (P3) à un processus d'échantillonnage discriminé sans remplacement.

Comme nous l'avons dit, l'état actuel de la recherche montre le caractère idéalisé et inadéquat de la plupart des modèles stochastiques de l'expression des gènes. En effet, toute fluctuation dans le niveau d'expression génétique est inévitablement biaisée, de manières diverses et à des degrés différents, par un ou plusieurs facteurs physico-chimiques : il peut s'agir de simples contraintes physico-chimiques au sein de la cellule, de produits de l'histoire évolutive de la population à laquelle la cellule individuelle appartient et même de résultats du processus d'évolution par sélection naturelle. Toutes les connaissances actuelles sur le bruit intracellulaire supportent donc notre affirmation selon laquelle aucune fluctuation dans le niveau d'expression génétique ne peut être considérée comme « aléatoire (au sens) fort » ; autrement dit, *toute* fluctuation est « aléatoire (au sens) faible » et peut donc être caractérisée, au niveau moléculaire, par cette notion d'une manière conceptuellement et empiriquement appropriée.

De même que dans le cas des mutations génétiques⁶³, l'utilisation de la notion d'« aléatoire (au sens) faible » pour caractériser le bruit dans l'expression des gènes pourrait soulever plusieurs objections. Deux d'entre elles notamment semblent importantes.⁶⁴

⁶² Ces différences jouent un rôle causal dans la détermination du bruit au cours du temps (par exemple, au cours de la génération cellulaire considérée). La probabilité qu'une fluctuation affecte l'expression d'un gène particulier (ou d'un des gènes du génome considéré) n'est pas la même à tout instant du laps de temps/de la génération considéré/e, ce qui implique sa dépendance d'autres fluctuations et d'autres événements physico-chimiques se produisant à l'intérieur de la cellule.

⁶³ Voir le chapitre 7.

⁶⁴ Pour l'analyse de la signification et du rôle de la notion d'« aléatoire (au sens) faible », nous renvoyons à la section 7.3 du chapitre 7.

La première concerne l'extension de la notion d'« aléatoire (au sens) faible » : en effet, cette notion s'applique à un très grand nombre de phénomènes physiques et, plus spécifiquement, biologiques ; de plus, nous avons conclu qu'il s'agit de la notion la plus appropriée pour caractériser *toute* fluctuation intracellulaire intervenant dans l'expression des gènes. Sur cette base, on pourrait nous objecter que sa valeur conceptuelle et son intérêt dans la caractérisation du bruit intracellulaire sont affaiblis. Pour répondre à cette première critique possible, nous renvoyons au chapitre 7 où nous soutenons que la grande extension de la notion d'« aléatoire (au sens) faible » ne pose aucun problème. Au contraire, il s'agit d'une notion explicative qui prend en compte tout biais pouvant influencer les fluctuations dans le processus d'expression génétique. Plus précisément, elle explique pourquoi le bruit l'affectant n'est pas le même à la fois dans tout gène d'un génome donné et à tout instant d'un certain laps de temps (par exemple, d'une génération cellulaire). Du point de vue des biologistes, la notion d'« aléatoire (au sens) faible » explique pourquoi ces derniers ne s'attendent pas à observer – et pourquoi, la plupart du temps, ils n'observent pas – le même bruit dans l'expression de tout gène le long du génome et au cours du temps. De plus, l'extension très large de cette notion n'empêche pas qu'on puisse la restreindre et la préciser dans des cas particuliers sur la base des propriétés spatio-temporelles du bruit intracellulaire observé et étudié.

La deuxième objection pouvant être adressée à la notion d'« aléatoire (au sens) faible » consiste à affirmer que son utilisation pour caractériser le bruit dans l'expression des gènes n'exprime rien d'autre que le caractère probabiliste de ce phénomène moléculaire et du processus physico-chimique qui l'engendre ; sur cette base, on pourrait donc nous objecter qu'il s'agit d'une notion inintéressante. Encore une fois, nous renvoyons au chapitre 7. L'aspect problématique d'une telle manière de définir le caractère « aléatoire » d'un phénomène ne réside pas dans sa réduction à la notion de probabilité ; il concerne plutôt la signification et le rôle de cette dernière dans la définition de la notion d'« aléatoire (au sens) faible ». Ce problème ne peut être résolu qu'en essayant de répondre à la question de savoir comment on doit interpréter les probabilités figurant dans les descriptions mathématiques de l'expression des gènes : doit-on les interpréter de manière épistémique ou objective ? C'est cette question que nous abordons dans le chapitre suivant. Plus précisément, nous proposerons une méthode possible (l'analyse de robustesse) pour argumenter en faveur d'une interprétation objective, non pas au sens métaphysique mais dans un sens plus approprié en science, des probabilités mobilisées dans les modèles scientifiques. Ensuite, nous testerons l'applicabilité de cette méthode dans le cas des modèles stochastiques de l'expression des

gènes. Enfin, nous soulignerons la valeur de l'analyse de robustesse en science et nous expliciterons le sens où les résultats obtenus par cette méthode peuvent être considérés comme objectifs.

Chapitre 10 – Pour une interprétation objective des probabilités dans les modèles stochastiques de l’expression génétique

10.1. Une analyse philosophique des probabilités dans un cadre objectif

La notion de probabilité est au cœur de la définition de l’« aléatoire (au sens) faible » qui, selon notre analyse, correspond à la notion de hasard appropriée pour caractériser au niveau moléculaire les mutations génétiques et le bruit dans l’expression des gènes. De plus, le caractère « aléatoire (au sens) faible » de ces deux sources de variation biologique est pris en compte et décrit en termes formels via l’utilisation de modèles mathématiques stochastiques¹ : dans ce cadre, les probabilités sont indissociables des résultats prévus par le modèle qui, contrairement à un modèle déterministe, ne décrit pas de manière univoque l’évolution d’un système donné. La question de l’interprétation des probabilités relatives aux mutations génétiques et au bruit intracellulaire peut donc nous permettre de répondre à la question de savoir si la notion d’« aléatoire (au sens) faible » désigne un hasard subjectif ou objectif, c’est-à-dire les limites de notre connaissance au sujet des mutations et du bruit ou des propriétés inhérentes à ces phénomènes moléculaires.²

Dans ce chapitre, nous avons donc décidé d’examiner la question suivante : comment doit-on interpréter les probabilités figurant dans les modèles mathématiques de l’expression des gènes ?³ Sont-elles épistémiques au sens où elles seraient relatives aux connaissances des biologistes à ce sujet, ou objectives au sens où elles rendraient compte de certaines caractéristiques propres à ce processus intracellulaire ? Remarquons que les biologistes ne semblent pas s’être intéressés à cette question depuis l’essor des publications sur la modélisation stochastique de l’expression des gènes. Pourtant, il s’agit d’un problème qui est à la fois d’ordre empirique et méthodologique : un argument en faveur d’une interprétation objective ou épistémique des probabilités dans ces modèles mathématiques doit en effet se

¹ Nous rappelons qu’un modèle stochastique est un modèle qui, à partir d’un même ensemble de conditions initiales, permet de prédire plusieurs résultats possibles selon une certaine loi de probabilité.

² En ce qui concerne la correspondance entre les deux conceptions du hasard (subjectif et objectif) et le double visage des probabilités (épistémiques et objectives), nous renvoyons à la section 2.6.3 du chapitre 2.

³ Nous ne soulèverons pas cette même question en ce qui concerne les mutations génétiques. Cependant, notre analyse aboutira à un résultat qui est aussi valable pour toute source causale de la variation biologique et, plus généralement, pour tout phénomène physique pouvant être l’objet d’une formalisation mathématique.

fonder sur les résultats et sur les outils méthodologiques des études sur l'expression des gènes. D'abord, sur les caractéristiques physiques et causales de ce processus intracellulaire et sur l'étiologie (c'est-à-dire, l'ensemble des causes) des fluctuations (ou bruit) qui le perturbent ; ensuite, sur les méthodes théoriques et expérimentales utilisées pour inférer les valeurs des probabilités.

Deux éléments suggèrent que les propriétés des facteurs et des interactions physico-chimiques impliqués dans l'expression des gènes sont à l'origine des probabilités figurant dans les modèles de ce processus. C'est la raison pour laquelle nous avons décidé d'analyser la question de leur interprétation dans un cadre objectif.

Un premier élément motivant notre approche objective porte sur les différences, à la fois quantitatives et qualitatives, entre la modélisation mathématique déterministe et stochastique de ce processus et, plus généralement, de tout système de réactions chimiques. Les biologistes ont montré qu'un modèle stochastique d'un système biochimique donné prévoit des oscillations, des bifurcations et des stabilités multiples, alors qu'un modèle déterministe du même système ne prévoit aucune oscillation, aucune bifurcation et une seule stabilité.⁴ Relativement à notre propos, l'aspect le plus important de ces travaux réside dans le fait que, de plus, les prévisions quantitatives et qualitatives des modèles stochastiques ont été confirmées par des études expérimentales. Ces résultats représentent à notre avis un indice important en faveur d'une interprétation objective des probabilités que ces modèles utilisent. En effet, ils suggèrent que le caractère probabiliste de suites de réactions biochimiques, notamment l'expression des gènes, n'est pas dû aux limites de notre connaissance, mais est une propriété inhérente à ces processus : il s'agit d'une caractéristique jouant un rôle central du point de vue quantitatif et qualitatif au niveau de leur dynamique et de leur résultat final.

Un deuxième élément qui nous a décidée à analyser la question de l'interprétation des probabilités dans un cadre objectif concerne les caractéristiques quantitatives, plus précisément statistiques, du bruit. Des études récentes ont révélé que les fluctuations dans le niveau d'expression des gènes suivent des tendances prédictibles⁵ : la variation aléatoire autour

⁴ Pour une étude de ces différences entre modèles stochastiques et déterministes dans le cas d'un mécanisme biochimique omniprésent dans les systèmes biologiques (le cycle enzymatique futile), voir Samoilov *et al*, 2005, « Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations ». On parle de « cycle futile » lorsque deux voies métaboliques vont dans des directions opposées, comme dans le cas de la dégradation et de la production du glucose.

⁵ Bar-Even *et al*, 2006, « Noise in protein expression scales with natural protein abundance » ; Newman *et al*, 2006, « Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise » ; Mettetal *et al*, 2006, Predicting stochastic gene expression dynamics in single cells » ; Mettetal & van Oudenaarden, 2007,

de l'abondance moyenne de protéines au sein du volume cellulaire (ou « bruit protéique ») s'est révélée inversement proportionnelle au nombre de protéines produites par molécule d'ARN transcrite selon un facteur de proportionnalité d'à peu près 1200. L'étude de cette tendance caractéristique s'accompagne de la découverte d'une corrélation spécifique entre le bruit et les protéines concernées : les fluctuations stochastiques dans le niveau d'expression de gènes différents suivent des tendances spécifiques aux types de protéines produites. Par exemple, comme nous l'avons déjà indiqué dans le chapitre précédent, les fluctuations dans l'expression de gènes essentiels pour le fonctionnement cellulaire sont moins importantes que celles concernant l'expression de gènes impliqués dans la production des protéines de stress. La découverte de cette corrélation entre les propriétés statistiques du bruit et les caractéristiques physico-chimiques, causales et fonctionnelles, du processus d'expression génétique représente à notre avis un autre indice non négligeable en faveur d'une interprétation objective des probabilités dans les modèles stochastiques de ce processus intracellulaire. En effet, elle suggère que les tendances prédictibles du bruit trouvent leur origine dans les propriétés des facteurs et des interactions physico-chimiques impliqués dans ce processus, et non dans nos limites de connaissance à leur sujet.⁶

Dans la section suivante, nous allons introduire un argument pour montrer que les probabilités relatives à l'abondance de protéines produites par l'expression des gènes (dans une population de cellules isogéniques à un moment donné ou au sein d'une même cellule au cours du temps) peuvent être considérées comme des probabilités objectives.⁷ À cette fin, nous proposerons d'appliquer l'analyse de robustesse – qui est une méthode d'analyse scientifique permettant d'éliminer les abstractions et les idéalizations propres à des modèles particuliers d'un même phénomène – à la modélisation stochastique de l'expression des gènes. Nous ne soutiendrons pas que cette méthode permet de montrer l'objectivité des

« Necessary Noise ».

⁶ Ces propriétés peuvent être de simples contraintes physico-chimiques, des sous-produits de l'histoire évolutive des cellules concernées, ou des adaptations, c'est-à-dire des produits de l'évolution par sélection naturelle (voir le chapitre 9). Les études récentes sur la relation de proportionnalité inverse entre la variation aléatoire dans l'abondance de protéines produites et le nombre de protéines par RNA transcrit (voir la note qui précède) suggèrent que cette relation pourrait être un résultat évolutif d'optimisation du bruit par rapport à l'énergie cellulaire.

⁷ Nous ajoutons que, si l'on arrive à montrer que les probabilités au niveau du *résultat* de l'expression des gènes, c'est-à-dire de l'abondance de protéines produites, sont objectives, alors on dispose d'une bonne raison pour inférer que les probabilités au niveau de ce *processus*, c'est-à-dire qui sont relatives aux facteurs et aux interactions physico-chimiques impliqués dans son déroulement, sont aussi objectives. En effet, comment le modèle d'un processus qu'on ne connaît que de manière limitée, et qu'à cause de notre ignorance partielle on caractérise de manière probabiliste, pourrait-il donner lieu à un résultat objectivement probabiliste ?

probabilités au sens où elles figurent dans des représentations complètes et vraies de l'expression des gènes et rendent donc compte de l'indéterminisme intrinsèque de ce processus. Autrement dit, nous ne ferons pas référence à la notion philosophique d'objectivité désignant la vérité absolue du monde réel, qui est complètement indépendante de la connaissance que nous en avons. Nous soutiendrons plutôt la possibilité de démontrer, via l'analyse de robustesse, que les probabilités figurant dans les modèles de l'expression des gènes sont objectives selon une notion d'objectivité restreinte au cadre scientifique.

10.2. L'analyse de robustesse

Qu'est-ce que l'analyse de robustesse ? C'est une méthode d'analyse parmi d'autres utilisées en science pour comprendre quels modèles produisent des prédictions fiables et peuvent donc être utilisés pour décrire et pour expliquer des phénomènes donnés⁸. Elle est mise en œuvre par l'étude d'un ensemble de modèles particuliers, semblables mais différents, d'un même phénomène : ces modèles peuvent partager quelques hypothèses communes, mais doivent être caractérisés par des abstractions et des idéalizations différentes pouvant concerner, par exemple, les paramètres et les conditions initiales, la structure mathématique du modèle, ou le niveau de représentation. Si, en dépit de leurs différences, ces modèles prédisent tous un même résultat – c'est-à-dire qu'ils attribuent tous une même propriété au phénomène qu'ils sont censés représenter – alors ce résultat peut être considéré comme « robuste » : il est déterminé par la structure commune à l'ensemble de modèles particuliers analysés. Plus précisément, un « résultat robuste » est un résultat qui n'est pas affecté par les différences dans les abstractions et dans les idéalizations propres à chaque modèle particulier d'un même phénomène ; autrement dit, il s'agit d'un résultat qui est « invariant »⁹ face à certaines modifications au niveau de l'ensemble de modèles particuliers, mais qui changerait si l'on modifiait la structure commune à tous ces modèles.

⁸ Levins, 1966, « The Strategy of Model Building in Population Biology » ; Wimsatt, 1981, « Robustness, reliability, and overdetermination » ; Weisberg, 2006a, « Robustness Analysis » ; 2006b, « Forty Years of 'The Strategy'. Levins on Model Building and Idealization ».

⁹ En mathématique et en physique théorique, une propriété est considérée comme invariante si elle ne change pas en suivant une opération donnée. Autrement dit, si un certain groupe de transformations ne modifie pas une propriété, alors on a à faire à une symétrie. Weyl (1952, *Symmetry*) fournit de nombreux exemples – mathématiques, physiques et biologiques – de ce qu'est une symétrie en ce sens moderne. Voir aussi Brading & Castellani, 2008, « Symmetry and Symmetry Breaking ».

Wimsatt¹⁰ a caractérisé de manière particulièrement éclairante l'apport propre à l'analyse de robustesse et qui la rend plus efficace que d'autres méthodes d'analyse scientifique dans l'identification de résultats robustes et de modèles fiables et explicatifs. Il affirme que l'analyse de robustesse peut nous permettre de distinguer ce qui est réel de ce qui est illusoire, ce qui est objectif de ce qui est épistémique, ce sur quoi la science doit se focaliser de ce qui est de l'ordre des artefacts dus aux perspectives particulières et, enfin, ce qui est métaphysiquement et épistémologiquement fiable de ce qui est peu fiable, non généralisable, inutile et fugace. En caractérisant dans ces termes ce qu'il est possible d'identifier par l'analyse de robustesse, c'est-à-dire le résultat robuste et la structure qui l'engendre, Wimsatt met en évidence que la propriété d'être robuste implique une certaine objectivité.

Bien avant Wimsatt, Levins¹¹ s'est aussi beaucoup intéressé à l'analyse de robustesse, en particulier relativement à la modélisation en génétique des populations et en écologie. De même que Wimsatt, il a souligné la relation d'implication entre robustesse et objectivité. Plus précisément, Levins écrit que l'analyse de robustesse permet d'établir « [...] si un résultat dépend des aspects essentiels d'un modèle ou des détails des hypothèses simplificatrices. »¹² Puis, en ce qui concerne l'étude d'un ensemble de modèles particuliers, similaires mais distincts, il ajoute :

« [...], si ces modèles, en dépit de leurs hypothèses différentes, produisent des résultats similaires, on obtient ce qu'on appelle un théorème robuste¹³ qui est assez libre relativement aux détails du modèle. Donc, notre vérité se trouve à l'intersection de couches indépendantes »¹⁴.

Cette dernière phrase est ambiguë et ouverte à des interprétations diverses : en effet, Levins ne parle pas de « la vérité », mais de « notre vérité », ce qui semble mettre en discussion son caractère objectif, selon certains. Au contraire, nous allons montrer que l'objectivité propre à

¹⁰ Wimsatt, 1981, *op. cit.*, p. 128.

¹¹ Levins, 1966, *op. cit.*

¹² *Ibid.*, p. 423 (notre traduction), « [...] whether a result depends on the essentials of a model or on the details of the simplifying assumptions. »

¹³ Il est important de préciser que Levins n'utilise pas le terme « théorème » selon son sens conventionnel, c'est-à-dire pour désigner un énoncé déductible d'un ensemble d'axiomes. Il fait plutôt référence à un énoncé conditionnel qui relie un résultat ou une propriété robuste et une structure causale, cette dernière étant commune à un ensemble de modèles particuliers utilisés pour décrire un même phénomène. Cet énoncé conditionnel prend la forme suivante : *ceteris paribus*, si la structure causale commune est présente, alors le résultat ou la propriété robuste a lieu. Voir Weisberg, 2006a, *op. cit.* ; 2006b, *op. cit.*

¹⁴ Levins, 1966, *op. cit.*, p. 423 (notre traduction), « [...], if these models, despite their different assumptions, lead to similar results we have what we can call a robust theorem which is relatively free of the details of the model. Hence our truth is the intersection of independent lies. »

cette vérité – qui peut se révéler lors de l'utilisation de descriptions relativement indépendantes d'un même phénomène – ne correspond pas à la notion de « vérité objective » conçue comme absolue, complète et indépendante de la connaissance que nous en avons. Il s'agit plutôt d'une vérité relative à une perspective, cette dernière n'étant pas particulière et arbitraire mais scientifiquement pertinente et fiable, donc objective dans ce sens.¹⁵

10.2.1. Trois analyses de robustesse

Dans un article récent, Weisberg et Reisman¹⁶ ont proposé de distinguer trois types d'analyses de robustesse. La première (qu'ils appellent *parameter robustness analysis*) consiste à changer les valeurs attribuées aux paramètres et aux conditions initiales d'un modèle afin de vérifier si son résultat est affecté. Cette méthode d'analyse peut permettre de montrer l'indépendance d'un résultat par rapport à un ensemble particulier de valeurs des paramètres et des conditions initiales, donc sa robustesse relativement à ce genre de changements.¹⁷

Le deuxième type d'analyse de robustesse (que Weisberg et Reisman appellent *structural robustness analysis*) concerne la structure mathématique du modèle. Elle consiste, par exemple, à ajouter ou à enlever des variables au modèle, à changer certaines relations de dépendance ou d'indépendance entre elles, etc., afin de comprendre si un résultat est affecté par ces modifications ou, d'un autre point de vue, afin d'identifier les variables et les relations entre elles qui sont nécessaires pour qu'un certain résultat soit observé. Ce deuxième type d'analyse de robustesse permet de montrer l'indépendance d'un résultat par rapport à une structure mathématique particulière, donc sa robustesse par rapport aux changements apportés

¹⁵ L'idée selon laquelle ce qui est objectif ne doit pas dépendre d'une perspective particulière mais être invariant par rapport à des cadres de référence différents a été l'objet de nombreuses réflexions en science et en philosophie des sciences. Plus particulièrement, l'idée qu'invariance et objectivité sont strictement connectées a été formulée par des physiciens théoriciens comme Weyl (1952, *Symmetry*, p. 133). Des ouvrages plus récents, à la fois philosophiques et scientifiques, ont reconsidéré les implications épistémologiques et métaphysiques de l'invariance et, en particulier, son lien avec l'objectivité. Entre autres, nous signalons Nozick, 2001, *Invariances: The Structure of the Objective World* ; Debs & Redhead, 2007, *Objectivity, Invariance, and Convention: Symmetry in Physical Science*.

¹⁶ Weisberg & Reisman, 2008, « The Robust Volterra Principle ».

¹⁷ En d'autres termes, ce premier type d'analyse de robustesse consiste à étudier un ensemble de modèles qui ne diffèrent que quant aux valeurs des paramètres pris en considération et des conditions initiales. Si tous ces modèles prédisent un même résultat, alors ce dernier peut être considéré comme robuste par rapport aux changements dans les valeurs des paramètres et des conditions initiales. De manière générale, changer certains aspects d'un modèle pour évaluer les effets des modifications apportées sur le résultat prédit et construire des modèles qui ne diffèrent que sous certains aspects pour voir s'ils prédisent tous un même résultat sont deux procédures équivalentes.

au modèle à ce niveau. De plus, elle permet d'identifier quelle partie de la structure mathématique est essentielle pour l'observation d'un certain comportement du modèle, c'est-à-dire pour la production d'un certain résultat.

Un troisième type d'analyse de robustesse (que Weisberg et Reisman appellent *representational robustness*) consiste à changer le niveau auquel un modèle représente un phénomène donné, autrement dit à modifier l'ensemble des aspects sous lesquels ce dernier est décrit. Elle peut permettre de montrer l'indépendance d'un résultat par rapport à une perspective de représentation particulière, donc sa robustesse par rapport aux changements de perspective. Par exemple, on peut décrire un même phénomène, soit au niveau individuel en termes de paramètres et de variables caractérisant les éléments particuliers engagés dans le phénomène considéré, soit au niveau populationnel en termes de paramètres et de variables caractérisant les propriétés de l'ensemble de ces éléments individuels. Ce troisième type d'analyse de robustesse ne touche donc pas aux hypothèses d'un modèle, ni quant aux valeurs des paramètres et des conditions initiales, ni quant aux variables et à leurs relations ; plutôt, en changeant le niveau de représentation, elle permet de comprendre si la manière dont ces hypothèses sont décrites affecte le résultat prévu par le modèle.

Avant d'explicitier notre argument pour une interprétation objective des probabilités dans les modèles stochastiques de l'expression des gènes, il est important de souligner un point central au sujet de l'analyse de robustesse, en particulier à propos de ce qu'elle est censée apporter dans le cadre de la recherche en science.

Si l'on appliquait l'analyse de robustesse à un ensemble de modèles susceptibles de représenter un phénomène physique donné sans se préoccuper de montrer que ces modèles décrivent correctement, même si ce n'est que de manière partielle et à une certaine échelle, ce phénomène réel¹⁸, nous pourrions certes identifier un résultat indépendant des caractéristiques de chaque modèle particulier et une structure commune à l'ensemble de modèles ; cependant, il s'agirait d'un résultat robuste et d'une structure commune abstraits, confinés au niveau des modèles, sans aucune connexion avec le phénomène réel qu'ils sont censés représenter. En d'autres termes, l'application de l'analyse de robustesse à un ensemble de modèles sans qu'aucune connexion empirique avec la réalité soit établie ne nous permet pas de distinguer

¹⁸ Nous adhérons à la thèse suivante en ce qui concerne la notion de validité empirique d'un modèle : le fait qu'un modèle est une description correcte d'un phénomène réel signifie qu'il représente ce dernier et cela en vertu de leur relation de similarité spécifique. Le modèle et le phénomène réel ne sont pas isomorphes, mais similaires uniquement sous certains aspects qui dépendent du projet particulier à la base de la construction du modèle. Nous adhérons donc à une position proche du « réalisme perspectiviste » de Giere (2006, *Scientific Perspectivism*) selon lequel, même si tout modèle est partiel et approximatif, il peut représenter un phénomène réel à certains égards, autrement dit, il peut être vrai relativement à une certaine perspective qui, à son tour, dépend du contexte, plus précisément, du projet particulier.

ce qui est réel de ce qui est illusoire, ce qui est fiable de ce qui ne l'est pas, ce qui est objectif de ce qui est épistémique, etc.¹⁹, car dans ce cas on ne considérerait que les modèles et non pas le phénomène même. Donc, pour que l'analyse de robustesse puisse nous dire quelque chose quant au caractère fiable, non illusoire, réel, objectif, d'un résultat et de la structure à son origine, il faut interpréter empiriquement la structure mathématique du modèle, c'est-à-dire attribuer un contenu empirique à chaque élément (paramètre, variable, relation entre paramètres et entre variables, etc.) du modèle, afin d'établir une connexion entre l'ensemble de modèles utilisés et le phénomène réel que ces derniers sont censés décrire²⁰. Cette connexion peut être établie, par exemple, à travers l'investigation empirique, sur le terrain ou au laboratoire, sur la base de l'analyse des données expérimentales²¹.

10.2.2. L'analyse de robustesse et les modèles stochastiques de l'expression des gènes

Notre argument pour une interprétation objective des probabilités relatives à l'abondance de protéines produites par l'expression des gènes – soit dans une population de cellules isogéniques à un moment donné, soit au sein d'une même cellule au cours du temps – est le suivant.²²

- (1) Si l'on peut réaliser l'analyse de robustesse dans le cas de la modélisation stochastique du processus d'expression génétique ;
- (2) Si la distribution de probabilité sur les valeurs possibles de l'abondance de protéines produites (par exemple, au sein d'une cellule dans un certain laps de temps) se révèle être un résultat commun à l'ensemble divers de modèles stochastiques utilisés pour décrire l'expression génétique, donc un résultat robuste déterminé par la structure commune à ces modèles ;

¹⁹ Wimsatt, 1981, « Robustness, reliability, and overdetermination ».

²⁰ L'analyse de robustesse doit précéder l'investigation empirique ou l'accompagner : la connexion entre l'ensemble de modèles et le phénomène réel doit être établie au niveau du résultat robuste et de la structure commune à l'ensemble de modèles particuliers.

²¹ L'observation du phénomène qu'on veut représenter peut permettre d'identifier les similarités spécifiques entre le résultat robuste et le résultat observé et entre la structure commune aux modèles analysés et la structure physico-causale du phénomène dans la réalité. C'est en vertu de ces similarités ne concernant que certains aspects du phénomène réel que la structure commune à l'ensemble des modèles et le résultat robuste le décrivent correctement – autrement dit, le représentent.

²² Merlin, 2009, « Pour une interprétation objective des probabilités dans les modèles stochastiques de l'expression génétique ».

(3) Si l'on arrive à établir une connexion empirique, par exemple par l'investigation *in vivo* ou *in vitro*, entre les modèles de l'expression des gènes et le processus réel, à la fois au niveau du résultat robuste et de la structure qui lui est associée ;

(C) Alors, on a de bonnes raisons pour conclure que les probabilités au niveau du résultat prévu par ces modèles stochastiques de l'expression génétique sont fiables, réelles, objectives, car indépendantes des hypothèses et des simplifications de chaque modèle particulier et car elles ne sont dues qu'à la structure commune à l'ensemble des modèles analysés. Du point de vue du phénomène réel modélisé, l'objectivité des probabilités macroscopiques relatives aux protéines produites (par exemple, au sein d'une cellule dans un certain laps de temps) réside dans leur indépendance des détails microscopiques du processus d'expression génétique et, plus généralement, de l'ensemble d'événements physico-chimiques ayant lieu à l'intérieur de la cellule. C'est sur cette base qu'on peut soutenir une interprétation objective de ces probabilités.

Avant d'illustrer notre argument par un exemple, nous soulignons qu'il n'est spécifique, ni à l'étude du bruit dans l'expression des gènes, ni à l'étude des sources causales de la variation biologique (par exemple, les mutations génétiques), ni plus généralement aux sciences biologiques. Au contraire, il peut s'appliquer dans le cadre de toute discipline scientifique faisant appel aux probabilités pour décrire un phénomène donné. Cela est possible pourvu qu'on dispose d'un ensemble de modèles particuliers qui décrivent tous un même phénomène et qui diffèrent en ce qui concerne les détails de leurs abstractions et de leurs idéalizations de départ, tout en gardant une certaine structure commune. De plus, il faut qu'on puisse établir une connexion empirique entre cette structure commune aux modèles et la structure physico-causale du phénomène réel aussi bien qu'entre le résultat robuste au niveau des modèles et le résultat observé expérimentalement.

10.2.3. La condition de réalisabilité de l'analyse de robustesse

Un argument fondé sur l'analyse de robustesse en faveur de l'objectivité des probabilités requiert avant tout que cette méthode scientifique puisse être appliquée. Deux situations peuvent remplir cette condition de réalisabilité relativement à la modélisation de l'expression des gènes. Ou bien les modèles que les biologistes utilisent déjà pour décrire ce

processus forment un ensemble approprié tel qu'on puisse réaliser l'analyse de robustesse²³. Ou bien il est nécessaire de construire un ensemble de modèles stochastiques de l'expression des gènes qui soient reliés entre eux par des hypothèses communes mais qui diffèrent sous certains aspects, notamment quant au niveau d'abstraction et d'idéalisation par rapport au processus qu'ils sont censés représenter.

Si cette deuxième situation se révèle nécessaire, on peut procéder de la manière suivante. D'abord, on considère l'un des modèles mathématiques que les biologistes emploient pour décrire le processus d'expression génétique et en prévoir le résultat final. Ensuite, on lui apporte des modifications diverses afin de comprendre quel résultat est indépendant de quel changement, c'est-à-dire invariant sous un certain groupe de transformations de la description de l'expression des gènes. Par exemple, imaginons que le modèle considéré prédise une certaine distribution de probabilité sur les valeurs possibles de l'abondance de protéines produites au sein d'une même cellule au cours d'une génération. On peut tester la robustesse de ces probabilités, soit relativement aux changements dans les valeurs de certains paramètres (les constantes de diffusion, d'association, de dissociation, de réaction, des molécules) et des conditions initiales (la concentration, l'état et la localisation des espèces moléculaires diverses à l'intérieur de la cellule donnée au temps t_0) : il s'agit de ce que Weisberg et Reisman appellent *parameter robustness analysis*. Soit relativement à des changements dans les variables prises en considération et dans les relations de dépendance et d'indépendance entre elles et avec d'autres paramètres (c'est-à-dire, relativement à des modifications de la structure mathématique) : c'est la *structural robustness analysis*. Soit relativement à des changements dans le niveau de représentation de l'expression génétique : c'est la *representational robustness analysis*.²⁴ Par une telle procédure de modélisation, on peut comprendre si la distribution de probabilité prédite par le modèle est robuste et identifier la structure qui la détermine et qui lui est donc associée.

Avant de s'engager dans la construction d'un tel ensemble de modèles, se pose donc la question de savoir si les modèles mathématiques que les biologistes utilisent déjà pour décrire l'expression génétique, pour l'analyser (bien que ce ne soit que de manière indirecte) et pour

²³ Si c'est bien le cas, cela signifie que ces modèles sont caractérisés par des aspects communs établis sur des bases solides, théoriques et expérimentales (par exemple, les valeurs de certains paramètres, quelques aspects de la structure mathématique, etc.) et, en même temps, diffèrent quant à certaines hypothèses particulières (par exemple, relativement au niveau d'abstraction et d'idéalisation, au niveau de représentation, à certains aspects de la structure mathématique, aux valeurs de certains paramètres et des conditions initiales, etc.).

²⁴ Nous rappelons que le processus d'expression génétique peut être décrit de manière discrète ou continue, c'est-à-dire en termes du nombre ou de la concentration de protéines présentes dans le volume cellulaire.

en prédire le résultat final (les équations maîtresses, de Fokker-Planck et de Langevin²⁵) remplissent la condition de réalisabilité de l'analyse de robustesse.²⁶

Dans les chapitres précédents, nous avons vu que les modèles stochastiques de l'expression des gènes partagent un certain nombre d'hypothèses idéalisées²⁷ qui sont liées à la manière dont le bruit intracellulaire est habituellement caractérisé dans ce cadre : il correspond à un processus stochastique markovien, non corrélé dans le temps. De plus, dans le cadre de tous ces modèles mathématiques, seules certaines étapes de l'expression des gènes (l'activation et la désactivation des gènes, les processus de transcription et de traduction, la dégradation des transcrits d'ARN messenger, la dégradation des protéines) et donc seules certains composants moléculaires (les gènes, les transcrits d'ARN messenger, les protéines produites) sont pris en considération.

Du point de vue mathématique, les équations maîtresses, de Fokker-Planck et de Langevin, diffèrent entre elles relativement à leur degré d'approximation. En effet, nous rappelons que les équations de Fokker-Planck et de Langevin peuvent être dérivées de l'équation maîtresse par l'intermédiaire d'approximations successives. D'abord, l'équation de Fokker-Planck est un cas particulier de l'équation maîtresse qu'on obtient lorsque la variation entre états devient très petite et tend vers zéro, c'est-à-dire en passant d'une représentation discrète à une représentation continue de l'évolution du système considéré. Dans le cas de la modélisation de l'expression génétique, l'équation de Fokker-Planck est utilisée pour décrire l'évolution dans le temps de la densité de probabilité sur les valeurs possibles de la concentration de protéines dans le volume cellulaire²⁸ : l'état de ce processus intracellulaire n'est pas décrit par une variable aléatoire discrète mais continue. L'équation maîtresse et l'équation de Fokker-Planck diffèrent donc en ce qui concerne le niveau de représentation de l'expression des gènes : la première la décrit comme une série d'états discrets et en termes mésoscopiques, c'est-à-dire en référence au comportement statistique d'une espèce

²⁵ Nous renvoyons à la présentation de ces modèles dans la section 8.6 du chapitre 8.

²⁶ Sous une formulation différente qui tient compte des critiques pouvant être adressées à l'identification d'une équation à un modèle (~~voir le chapitre 8~~), la question que nous posons ici est celle de savoir si les équations maîtresses, de Fokker-Planck et de Langevin, décrivent toutes un même modèle de l'expression des gènes ou, au contraire, des modèles divers de ce processus de manière telle qu'on puisse mettre en œuvre l'analyse de robustesse.

²⁷ Les positions des molécules à l'intérieur du volume cellulaire sont supposées être distribuées de manière aléatoire et leurs vitesses selon la distribution de Maxwell-Boltzmann ; l'expression des gènes est supposée avoir lieu dans un milieu intracellulaire homogène et bien mélangé au cours du temps ; la probabilité qu'une molécule se trouve dans n'importe quel endroit du microenvironnement cellulaire est considérée comme uniforme dans tout le volume de la cellule ; la probabilité qu'une molécule donnée entre en collision avec une autre molécule quelconque et qu'elles donnent lieu à une réaction est considérée comme indépendante de la localisation des molécules à l'intérieur de la cellule.

²⁸ L'équation maîtresse décrit, au contraire, l'évolution dans le temps de la distribution de probabilité sur les valeurs possibles du nombre de protéines à l'intérieur de la cellule.

moléculaire ; la deuxième, au contraire, le décrit comme une série continue d'états et en termes macroscopiques, c'est-à-dire en référence à la concentration d'espèces moléculaires dans le milieu intracellulaire, donc au niveau populationnel.

En second lieu, l'équation de Langevin peut aussi être dérivée par approximation de l'équation maîtresse, mais aussi de l'équation de Fokker-Planck. Cependant, elle est souvent construite de manière heuristique en ajoutant un terme de « bruit » (habituellement, un bruit blanc gaussien) à une équation différentielle déterministe. De la même manière que l'équation de Fokker-Planck, l'équation de Langevin diffère donc de l'équation maîtresse quant au niveau de représentation de l'expression des gènes : elle décrit ce processus de manière continue et à l'échelle macroscopique, alors que l'équation maîtresse décrit son évolution par sauts discrets et à l'échelle mésoscopique. Enfin, l'équation de Langevin ne diffère de l'équation de Fokker-Planck que dans la manière dont les fluctuations sont représentées : dans la première équation, elles sont décrites comme du bruit additif et externe, alors que dans l'équation de Fokker-Planck il s'agit d'un bruit inhérent au processus d'expression génétique.

Tous ces éléments nous suggèrent que, lorsqu'elles sont utilisées dans la modélisation de l'expression des gènes, les équations maîtresses, de Fokker-Planck et de Langevin partagent un certain nombre d'hypothèses, d'approximations et d'idéalisations qui font que l'analyse de robustesse ne peut pas être réalisée de manière efficace dans ce cas. Ces équations forment un ensemble de modèles qui ne diffèrent que quant au niveau de représentation de l'expression des gènes : ils ne semblent pas être assez divers pour qu'on puisse conclure que leurs résultats communs, s'ils ont lieu, sont robustes et donc objectifs. S'ouvre donc la perspective d'un travail de construction d'un ensemble de modèles qui soit approprié pour mettre en œuvre l'analyse de robustesse et qui nous permette de fournir un argument en faveur de l'objectivité des probabilités au niveau du résultat de l'expression des gènes. Mais comment procéder ? Tâchons de répondre à cette question ci-dessous.

Nous suggérons de procéder au changement des hypothèses simplificatrices communes à tout modèle stochastique de ce processus et, sur la base des résultats récents obtenus au sujet du bruit intracellulaire²⁹, à leur remplacement par des hypothèses plus réalistes. En effet, ce n'est qu'en modifiant la manière approximative et idéalisée dont tout modèle caractérise l'environnement intracellulaire, les propriétés statistiques du bruit et les étapes du processus d'expression génétique, qu'on pourra identifier une distribution de probabilité sur les valeurs possibles de l'abondance de protéines produites³⁰ qui soit

²⁹ Pour plus de détails, nous renvoyons au chapitre 9.

³⁰ Soit dans une population isogénique à un moment donné, soit au sein d'un même individu au cours du temps.

véritablement robuste, c'est-à-dire qui soit commune à tout modèle de l'expression des gènes indépendamment du caractère plus ou moins idéalisé (ou plus ou moins réaliste) de ses hypothèses de départ. Par une telle démarche, d'une part, on peut montrer que l'absence de réalisme sur le comportement des éléments microscopiques à l'intérieur de la cellule, qui est propre à la plupart des modèles mathématiques de l'expression génétique, n'affecte aucunement le résultat probabiliste en termes d'abondance de protéines produites ; autrement dit, les probabilités macroscopiques sont indépendantes des détails des événements microscopiques du milieu intracellulaire. D'autre part, on peut identifier la structure essentielle qui est commune à tout modèle particulier de l'expression des gènes et qui détermine son résultat probabiliste robuste. Sur cette base, on peut suggérer une interprétation objective des probabilités au niveau du résultat de ce processus.

Voici une illustration simple de la manière dont on peut réaliser l'analyse de robustesse dans ce cas³¹. D'abord, considérons une description mathématique (par exemple, par un système d'équations maîtresses) de l'expression du gène G au sein de la cellule C au cours d'une génération. En particulier, concentrons notre attention sur l'une des hypothèses au fondement de ce modèle, par exemple, celle d'un milieu homogène et bien mélangé au sein de la cellule C où l'expression du gène G a lieu. Ensuite, substituons-la par une hypothèse plus réaliste qui prend en compte les différences caractérisant l'environnement intracellulaire (notamment, sa division en compartiments et la présence d'obstacles moléculaires en son sein).³² Enfin, contrôlons les effets de ce changement sur le résultat prédit par le modèle ainsi modifié, c'est-à-dire au niveau de la distribution de probabilité sur les valeurs possibles du nombre de protéines dans la cellule C au cours d'une génération. Si le changement apporté provoque une modification importante de ces probabilités et si ce résultat n'est pas confirmé expérimentalement, alors on a de bonnes raisons pour penser que l'hypothèse de l'homogénéité et du « bon mélange » du milieu intracellulaire est essentielle : elle fait partie de la structure commune à l'ensemble divers de modèles décrivant l'expression du gène G dans la cellule C au cours d'une génération. Si, au contraire, la substitution de cette hypothèse par une autre plus réaliste ne provoque aucune modification des probabilités relatives au nombre de protéines produites et qu'elles sont confirmées expérimentalement, alors on peut conclure que cette hypothèse n'est pas essentielle et ne fait donc pas partie de la structure

³¹ Nous n'entrerons pas dans les détails formels d'une telle procédure.

³² Nous rappelons que, dans le cadre des études récentes sur le bruit dans l'expression des gènes, les biologistes commencent à construire et à utiliser des simulations stochastiques, notamment les simulations Monte Carlo bi-dimensionnelles, qui évacuent l'hypothèse idéalisée d'un environnement intracellulaire homogène et bien mélangé. Voir El Samad *et al*, 2005, « Stochastic modelling of gene regulatory networks ».

commune aux modèles particuliers de l'expression du gène G.³³ Les probabilités pouvant être interprétées objectivement sont donc celles qui figurent au niveau du résultat engendré par la structure commune à l'ensemble de modèles particuliers analysés et qui sont confirmées expérimentalement.

10.3. L'objectivité des probabilités robustes

L'utilisation de la notion d'objectivité pour caractériser des probabilités qui se sont révélées robustes sur un ensemble de modèles relativement différents peut poser problème. Puisque la robustesse de ces probabilités n'est testée que par rapport à un nombre limité de changements dans la modélisation du phénomène étudié, il pourrait sembler illégitime de les interpréter objectivement vu qu'elles dépendent d'un nombre limité de représentations et, de manière plus générale, d'une perspective humaine. Nous souhaitons donc clarifier la notion d'objectivité à laquelle nous faisons appel dans la caractérisation de probabilités robustes et, plus généralement, d'un résultat quelconque qui s'est révélé robuste. Comme nous l'avons déjà suggéré à plusieurs reprises, il s'agit d'une notion d'objectivité propre à la science qui tient donc compte de l'inévitable caractère partiel et approximatif des modèles et des théories scientifiques et de l'utilisation incontournable de ces derniers afin d'appréhender la réalité.

De manière générale, si un résultat est robuste par rapport à des changements divers au niveau des modèles particuliers décrivant un même phénomène et s'il est aussi confirmé expérimentalement, cela ne signifie pas que ce résultat corresponde à une propriété intrinsèque du phénomène réel décrit par cet ensemble de modèles, indépendamment de la connaissance que nous en avons. Plus spécifiquement, si les probabilités relatives à l'abondance de protéines produites sont robustes par rapport à des modifications diverses, notamment au niveau des hypothèses restrictives propres aux modèles actuels de l'expression génétique, cela ne signifie pas qu'elles soient dues à la nature fondamentalement indéterministe de ce processus. Elles ne sont ni la mesure de l'indéterminisme propre au niveau moléculaire, ni des propensions objectives du système d'expression génétique à engendrer certains résultats. Fidèle à notre agnosticisme métaphysique³⁴, nous soutenons que

³³ Dans ce cadre, on doit mobiliser une méthode statistique qui permette de comparer les distributions de probabilité prédites par l'ensemble de modèles particuliers de l'expression du gène G et par laquelle on puisse spécifier un écart de confiance. Seulement sur cette base, il est possible d'évaluer si les probabilités relatives au nombre de protéines que ces modèles divers prédisent correspondent ou pas à un même résultat.

³⁴ Nous renvoyons à la section 1.6 du chapitre 1.

les probabilités objectives car robustes sont indifférentes par rapport au fait de concevoir le processus d'expression génétique et le bruit qui l'affecte comme fondamentalement déterministes ou indéterministes.

Nous proposons de caractériser l'objectivité des probabilités qui se sont révélées robustes en affirmant qu'elles sont objectives car déterminées (contraintes) par les caractéristiques physiques et causales du processus d'expression génétique, c'est-à-dire par les caractéristiques qui sont propres aux facteurs et aux interactions physico-chimiques impliqués dans ce processus. Du point de vue de la modélisation, l'ensemble des facteurs et des interactions qu'il est approprié de considérer pour représenter l'expression des gènes peut être identifié par la mise en œuvre de l'analyse de robustesse : il correspond à la structure commune d'un ensemble divers de modèles qui sont utilisés pour décrire ce processus et qui prédisent tous un même résultat. Le résultat robuste qui, dans le cas qui nous intéresse, est une certaine distribution de probabilité sur les valeurs possibles de l'abondance de protéines produites, est donc déterminé (contraint) par ce noyau commun.

Essayons de visualiser cette relation entre la structure commune et le résultat robuste au niveau des modèles et, au niveau de la réalité, entre le processus d'expression génétique et la distribution de probabilité sur les valeurs possibles de l'abondance de protéines produites.

Considérons un exemple très général : l'expression du gène G dans la cellule C au cours d'une génération. L'ensemble de facteurs moléculaires en interaction physico-chimique qui sont engagés dans le processus d'expression du gène G déterminent la variation dans l'abondance de protéines produites par la cellule C à partir de ce gène au cours d'une génération. Selon notre argument, le processus d'expression du gène G peut être identifié et caractérisé de manière appropriée au moyen de l'analyse de robustesse : il correspond à la structure commune aux modèles divers utilisés pour le décrire. Cette structure désigne l'ensemble des facteurs physiques et causaux pouvant représenter, sous certains aspects et à un certain niveau, le processus d'expression du gène G dans la cellule C au cours d'une génération. Par exemple, ce dernier peut consister, d'abord, en l'ensemble comprenant le gène G dont on étudie l'expression, des gènes situés sur la même molécule d'ADN, des molécules d'ARN messagers et de transfert, des ARN polymérases, des ribosomes, des facteurs de transcriptions et d'autres protéines et enzymes divers impliqués dans l'expression du gène G dans la cellule C au cours d'une génération. Ensuite, le processus d'expression de ce gène est aussi caractérisé par les probabilités relatives aux conditions initiales (par exemple, la concentration, nombre, localisation, état des espèces moléculaires au moment t_0), aux constantes d'évolution de certaines variables (par exemple, la concentration, nombre,

localisation, état des espèces moléculaires à l'intérieur de la cellule C) aux caractéristiques du milieu intracellulaire, etc. Selon notre argument, tous les facteurs physico-causaux et probabilistes définissant le processus d'expression du gène G dans la cellule C au cours d'une génération peuvent être identifiés par l'application de l'analyse à sa modélisation. Une fois qu'on les a ainsi identifiés, on peut les désigner comme l'ensemble des facteurs physiques et causaux pertinents pour décrire correctement, même si ce n'est toujours que de manière partielle et approximative, ce processus intracellulaire.

10.3.1. Vers une notion d'objectivité pour la science

Nous venons d'affirmer qu'un modèle peut représenter un phénomène réel donné s'il le décrit de manière correcte, même si ce n'est que sous certains aspects et à un certain niveau. Un modèle peut donc être considéré comme vrai (ou du moins partiellement vrai) relativement à une perspective qui n'est pas arbitraire, mais dont on a déterminé le caractère fiable et pertinent à travers l'application de l'analyse de robustesse. Sur cette base, nous suggérons que le résultat robuste produit par ce modèle est objectif, non pas du point de vue métaphysique, mais d'un point de vue scientifique, c'est-à-dire relativement à une description partielle mais fiable et pertinente d'un phénomène réel donné.

Essayons de préciser notre manière d'entendre la notion d'objectivité dans un cadre scientifique. Nous avons déjà remarqué à plusieurs reprises que tout modèle est une représentation partielle et simplifiée d'un phénomène réel. De plus, même si l'on pouvait imaginer construire un modèle complet et absolument objectif de l'univers dans sa totalité ou d'un phénomène réel donné, son exhaustivité ne serait pas explicative car le modèle construit ne serait qu'une copie parfaite de la réalité ou du phénomène qu'on vise à décrire et à expliquer³⁵. Sur cette base, nous avons donné une réponse négative à la question épistémologique suivante : est-il possible de résoudre le problème métaphysique de la nature fondamentalement déterministe ou indéterministe de la réalité ou d'un phénomène donné ? Cependant, nous soutenons qu'il est possible de conclure, par exemple sur la base du résultat de l'analyse de robustesse relativement à la modélisation d'un phénomène donné, que les probabilités qui le caractérisent sont objectives puisqu'elles renvoient à l'ensemble des facteurs physico-causaux qui les ont déterminées, c'est-à-dire à la structure du modèle

³⁵ Giere, 1999, « Using Models to Represent Reality ».

pertinent pour représenter ce phénomène. De plus, des probabilités robustes peuvent être considérées comme des probabilités objectives au sens où elles décrivent une portion de la réalité, même s'il ne s'agit pas d'une description exhaustive mais d'une description à un certain niveau et sous certains égards, ces derniers ayant été identifiés par l'analyse de robustesse en tant que pertinents. Les probabilités sont donc objectives car elles ne portent pas sur notre ignorance en ce qui concerne le phénomène réel considéré, mais désignent et caractérisent certains aspects de ce phénomène même.

En nous inspirant du « réalisme perspectiviste » de Giere³⁶, nous adoptons une position que nous appelons « réalisme local » en ce qui concerne les modèles : un modèle est une représentation vraie (ou du moins partiellement vraie) d'un phénomène réel s'il décrit ce dernier relativement à un ensemble de caractéristiques physico-causales qui s'est révélé fiable, pertinent et indépendant par rapport à un certain nombre de perspectives particulières. C'est dans ce cadre que nous concevons une notion d'objectivité propre à la science qui, nous le soulignons encore une fois, ne correspond pas à l'objectivité absolue et que nous invoquons relativement à l'interprétation des probabilités. Du point de vue de notre argument pour l'objectivité des probabilités robustes, si les trois conditions que nous avons évoquées sont remplies³⁷, nous pouvons conclure que les probabilités caractérisant le résultat de l'expression génétique sont objectives car « localement réelles/réelles au sens local » : elles font partie de la description correcte de ce processus intracellulaire sous certains égards et à un certain niveau pertinent de représentation.³⁸

Sur cette base, reconsidérons l'affirmation de Levins selon laquelle « notre vérité réside dans l'intersection de couches indépendantes »³⁹. Nous suggérons que, par cette phrase, Levins voulait exprimer cette même conception d'objectivité propre à la science que nous venons de présenter et que nous associons à la propriété de robustesse. Il parle de « *notre vérité* », ce qui met en évidence le fait que toute connaissance que nous pouvons avoir du

³⁶ Selon le « réalisme perspectiviste » de Giere (2006, *Scientific Perspectivism*), toute description scientifique capture seulement certains aspects spécifiques de la réalité qui sont relatifs à des projets et à des intérêts particuliers. Par conséquent, tout modèle d'un phénomène réel n'est pas une représentation objectivement vraie de ce dernier ; il s'agit plutôt d'une représentation approximativement vraie, relativement à une perspective humaine particulière.

³⁷ Nous renvoyons à la section 10.2.2 plus haut.

³⁸ Giere affirmerait que ces probabilités sont vraies, non pas indépendamment de la connaissance que nous en avons (donc, elles ne sont pas objectives du point de vue métaphysique) mais selon la perspective adoptée pour décrire le processus réel d'expression génétique qui, à son tour, est relative à un contexte donné. De manière générale, nous ajoutons que des probabilités dont on aurait montré la robustesse sont vraies relativement à une manière pertinente de représenter le phénomène réel auquel elles sont associées. Dans ce sens, elles ne sont pas objectives de manière absolue, mais objectives car relatives à et déterminées par un ensemble de facteurs pertinents pour décrire correctement – autrement dit, représenter – le phénomène réel donné.

³⁹ Levins, 1966, « The Strategy of Model Building in Population Biology », p. 423.

monde réel est relative à notre représentation de ce dernier : en effet, nous ne pouvons appréhender la réalité qu'en la représentant par des modèles toujours partiels et simplifiés ou, selon la terminologie de Giere, selon des perspectives diverses dépendant de nos intérêts particuliers et ne portant que sur certains aspects spécifiques du monde. Levins parle aussi de « notre *vérité* » : il aurait pu se passer de ce terme et en utiliser d'autres moins fortement connotés. La vérité que ce mot désigne dans le cadre de l'affirmation de Levins est une vérité certes non exhaustive, limitée à un certain niveau et à certains aspects de la réalité qui émergent lors de la rencontre de couches indépendantes – autrement dit, qui surgissent du noyau commun à plusieurs représentations d'un même phénomène réel –, mais pourtant *une vérité*, non pas complète mais « locale » et relative à une perspective qui s'est révélée fiable et pertinente. Nous retrouvons là le sens de la notion d'objectivité en termes de « réalisme local » : selon notre analyse, cette notion est pertinente pour caractériser les probabilités robustes au niveau du résultat dans les modèles du processus d'expression génétique ; plus généralement, elle correspond à la notion d'objectivité appropriée en science.

10.4. Réflexions conclusives sur l'analyse de robustesse

Dans ce dernier chapitre, nous avons essayé de répondre à la question de l'interprétation des probabilités dans les modèles mathématiques de l'expression des gènes qui rendent compte du bruit intracellulaire. Un tel questionnement a trouvé sa motivation principale dans la notion d'« aléatoire (au sens) faible » que nous avons introduite en tant que notion de hasard appropriée pour caractériser au niveau moléculaire les fluctuations stochastiques (ou bruit) dans l'expression des gènes aussi bien que les mutations génétiques⁴⁰. Plus précisément, nous avons soutenu que, en vertu de la relation privilégiée entre hasard et probabilité, il est possible de déterminer si la notion d'« aléatoire (au sens) faible » caractérisant le bruit intracellulaire désigne un hasard subjectif ou objectif à la lumière de l'interprétation des probabilités dans les descriptions formelles de ce phénomène ; en effet, la notion d'« aléatoire (au sens) faible » est définie en termes de probabilités.

Nous venons de fournir une méthode permettant d'argumenter en faveur d'une interprétation objective des probabilités dans les modèles mathématiques de l'expression des gènes : il s'agit de recourir à l'analyse de robustesse afin de montrer que la distribution de

⁴⁰ Voir respectivement les chapitre 9 et 7.

probabilité sur les valeurs possibles de l'abondance de protéines produites est robuste, donc objective car déterminée par la structure physico-causale du processus d'expression génétique. Plus généralement, nous soutenons que l'analyse de robustesse peut permettre de comprendre si la variation qu'on observe au sein des populations naturelles est due à des phénomènes qui sont probabilistes d'un point de vue épistémique, *et/ou ?* si elle peut être dite « objective » au sens de la notion d'objectivité que nous avons introduite plus haut et qui nous semble révéler toute sa valeur en science.

En guise d'exemple, considérons l'application de l'analyse de robustesse dans le cas d'une autre source de la variation biologique, les mutations génétiques. Plus particulièrement, considérons la méthode statistique la plus utilisée, à la fois en génétique moléculaire et en génétique des populations, pour estimer le taux de mutation : c'est le test de fluctuation ou modèle Lea-Coulson⁴¹. Nous rappelons qu'elle repose sur un certain nombre d'hypothèses qui s'éloignent considérablement de ce qu'on sait actuellement sur les biais pouvant affecter le processus de mutation. Dans ce cas, la mise en œuvre de l'analyse de robustesse peut nous permettre de déterminer si cette méthode statistique est fiable, autrement dit si la valeur du taux de mutation qu'elle prédit (par exemple, la probabilité qu'une mutation ait lieu dans une certaine cellule par génome par génération), est robuste donc objective.

Afin de tester le caractère objectif du taux de mutation calculé sur la base du modèle Lea-Coulson, nous suggérons de procéder de la manière suivante. D'abord, modifions l'une (ou plus) de ses hypothèses idéalisées (par exemple, l'hypothèse selon laquelle la probabilité qu'une mutation se produise est constante au cours d'une génération ou bien celle selon laquelle cette probabilité est indépendante des événements mutationnels précédents). Ensuite, évaluons ce que les changements apportés provoquent au niveau du résultat, c'est-à-dire au niveau des estimations statistiques du taux de mutation obtenues sur la base du modèle Lea-Coulson ainsi modifié.⁴² Si le changement d'une certaine hypothèse provoque une modification importante⁴³ de la probabilité qu'une mutation ait lieu, alors on a de bonnes raisons pour penser que cette hypothèse est essentielle et doit donc être prise en compte pour évaluer le taux de mutation. Si, au contraire, le changement de cette hypothèse ne provoque aucune modification d'une certaine importance au niveau du résultat, c'est-à-dire de la

⁴¹ Nous renvoyons aux chapitres 3 et 7.

⁴² L'importance de la modification provoquée au niveau du résultat probabiliste par le changement d'une des hypothèses du modèle doit être évaluée relativement à un intervalle de confiance, c'est-à-dire un intervalle qui contient, avec un certain degré de confiance, la valeur de référence à estimer. Par exemple, un intervalle de confiance de 95% a une probabilité égale à 0.95 de contenir la valeur qu'on cherche à estimer.

⁴³ Une modification importante est une modification qui excède l'intervalle de confiance auquel on se réfère.

probabilité qu'une mutation ait lieu, alors on peut conclure qu'elle n'est pas essentielle et ne doit donc pas être prise en considération afin d'estimer le taux de mutation.⁴⁴

La réalisation de l'analyse de robustesse dans le cas des mutations génétiques peut donc permettre d'identifier l'ensemble d'hypothèses essentielles dont toute méthode statistique doit tenir compte pour calculer de manière fiable le taux de mutation. En d'autres termes, elle peut permettre d'identifier un modèle robuste, donc pertinent et fiable, en fonction duquel l'estimation du taux de mutation pourra être dite robuste et donc objective.

L'analyse de robustesse peut aussi s'appliquer aux techniques expérimentales⁴⁵ utilisées pour collecter des données sur le bruit dans l'expression des gènes ou sur les mutations génétiques. Ces données seront ensuite utilisées pour attribuer des valeurs à des variables statistiques diverses concernant, par exemple, le processus d'expression génétique ou le processus de mutation.

L'importance de l'utilisation de plusieurs techniques expérimentales pour déterminer la fiabilité des observations qu'on réalise par des instruments scientifiques a été soutenue, entre autres, par Hacking⁴⁶. Plus récemment, Culp⁴⁷ a suggéré qu'on peut démontrer l'objectivité de certaines données expérimentales quand bien même toute donnée dépend inévitablement d'un ensemble de présupposés théoriques. Elle introduit ce qu'elle appelle le « critère de robustesse » selon lequel il faut construire un ensemble de techniques expérimentales tel qu'il est très improbable qu'elles produisent toutes des données comparables. Cet ensemble de techniques doit remplir les deux conditions suivantes. 1) S'il y a une cause commune, c'est-à-dire que les données obtenues par les différentes techniques à propos de la propriété P de l'objet x ont été contraintes par le fait que x possède P , alors toutes les techniques doivent produire des données comparables. 2) S'il n'y a pas de cause commune, alors au moins une technique doit produire des données qui ne sont pas comparables avec les données produites par les autres techniques. Autrement dit, si des techniques expérimentales, qui sont chargées théoriquement mais de manière différente,

⁴⁴ Une telle démarche correspond à ce qu'on appelle l'analyse de la « robustesse statistique » : elle permet d'identifier des valeurs statistiques robustes, c'est-à-dire qui ne changent pas en dépit de (petites) déviations dans les hypothèses du modèle utilisée pour les estimer. Pour une introduction à ce genre d'analyse de robustesse, voir Hampel *et al*, 1986, *Robust Statistics – The Approach Based on Influence Functions* ; Huber, 1981, *Robust Statistics*.

⁴⁵ Une technique expérimentale est une série de manipulations visant à rassembler des données brutes (Culp, 1995, « Objectivity in Experimental Inquiry: Breaking Data-Technique Circles »). Toute technique expérimentale dépend d'une théorie. Plus précisément, l'interprétation des données collectées dépend d'un ensemble de présupposés théoriques parmi lesquels figurent la théorie de l'objet d'investigation et la théorie sur laquelle repose la technique expérimentale utilisée.

⁴⁶ Hacking, 1983, *Representing and Intervening*.

⁴⁷ Culp, 1994, « Defending Robustness: the Bacterial Mesosome as a Test Case » ; 1995, *op. cit.*

produisent un même corps de données, cela est très probablement dû aux propriétés inhérentes à l'objet réel observé : ces données robustes, ou invariantes par rapport à un ensemble divers de techniques particulières, peuvent donc être considérées comme objectives en ce sens.

Dans le cas des mutations génétiques, on peut appliquer l'analyse de robustesse aux données expérimentales à partir desquelles on calcule le taux de mutation. Par exemple, on peut estimer la probabilité qu'une mutation ait lieu dans une certaine cellule par génome par génération en utilisant plusieurs techniques de séquençage de l'ADN qui consistent en des manipulations physico-chimiques différentes et donc ne reposent pas sur le même ensemble de présupposés théoriques (par exemple, le séquençage par dégradation chimique sélective et par synthèse enzymatique sélective⁴⁸). De cette manière, on peut identifier un corps de données robuste, c'est-à-dire un corps de données qui est invariant par rapport à un ensemble de techniques qui dépendent de théories différentes. En effet, si un même résultat (par exemple, un même nombre de mutations génétiques au sein d'une séquence d'ADN donnée au cours d'un certain laps de temps) est estimé en utilisant des techniques de séquençage différentes, alors on a de bonnes raisons pour penser qu'il ne s'agit pas d'une simple coïncidence. Au contraire, on doit conclure que le résultat obtenu est invariant car contraint par les caractéristiques propres aux séquences d'ADN analysées : il s'agit donc d'un corps de données objectif.

On peut aussi réaliser l'analyse de robustesse des données expérimentales dans le cas du bruit dans l'expression des gènes. Cela consiste à observer et à quantifier la variation autour de l'abondance moyenne de protéines produites, par exemple au sein d'une population de cellules isogéniques à un instant donné, en utilisant des techniques de détection des signaux lumineux qui, encore une fois, sont des manipulations physiques différentes et ne reposent donc pas sur un même ensemble de présupposés théoriques (par exemple, la cytométrie en flux et la microscopie à fluorescence⁴⁹). Si, par l'utilisation de techniques

⁴⁸ La technique de séquençage par dégradation chimique sélective fut développée par l'équipe de Walter Gilbert dans les années 1970. Elle consiste à dégrader l'ADN de manière sélective, c'est-à-dire au niveau des adénines, des cytosines, des guanines ou des thymines qui sont caractérisées par des réactivités chimiques différentes. Ensuite, on peut procéder à la reconstruction de la séquence nucléotidique de l'ADN en reconstituant l'ordre des coupures sélectives effectuées. La technique de séquençage par synthèse enzymatique sélective fut développée à la même époque par l'équipe de Sanger. Elle consiste à synthétiser de manière répétée un nouveau brin d'ADN à partir d'un ADN simple brin et à bloquer à chaque fois son élongation par l'incorporation d'un inhibiteur spécifique à chaque base nucléotidique différente. Une fois que cette opération a été répétée quatre fois, c'est-à-dire une fois pour chaque base nucléotidique, on peut procéder à la reconstruction de la séquence d'ADN à travers la séparation par électrophorèse des segments d'ADN obtenus, ce qui permet de repérer la position de chaque nucléotide le long de la séquence. Pour une présentation détaillée de ces deux techniques de séquençage de l'ADN et des différences dans leurs présupposés théoriques respectifs, voir Culp, 1995, *op. cit.*

⁴⁹ Nous rappelons que la cytométrie en flux est une technique qui permet d'étudier des particules isolées (molécules, cellules, etc.) entraînées par un flux liquide en les faisant défiler à grande vitesse dans le faisceau

différentes, l'on observe toujours une même variation dans la quantité de protéines au sein des cellules de la population isogénique considérée, alors on peut conclure qu'il ne s'agit pas d'une simple coïncidence. Au contraire, les données obtenues peuvent être considérées comme indépendantes des présupposés théoriques de chaque technique particulière utilisée : elles sont robustes, donc objectives, car déterminées (contraintes) par les propriétés physico-chimiques des cellules observées, notamment par la quantité de protéines effectivement produites au sein de chaque cellule de la population considérée.

En ce qui concerne l'utilisation de l'analyse de robustesse pour argumenter en faveur de l'objectivité d'un résultat qui s'est révélé robuste, qu'il soit probabiliste ou non, la question reste de savoir où réside la légitimité du lien entre robustesse et objectivité. En effet, le fait qu'un résultat soit robuste, par exemple dans le cas de la modélisation mathématique, dépend de deux facteurs : en premier lieu, la taille de l'ensemble de modèles utilisés pour décrire un phénomène donné ; en second lieu, le degré de différence entre les hypothèses et les abstractions/idéalisations propres à chaque modèle particulier. Se pose donc la question de savoir si ces deux facteurs doivent dépasser un certain seuil pour qu'on puisse attribuer la propriété d'être robuste, donc objectif, à un résultat qui s'est révélé invariant, ou si l'on doit parler de degrés différents de robustesse, et donc de degrés différents d'objectivité, selon la taille de l'ensemble de modèles analysés et selon le nombre de différences dans leurs hypothèses et dans leurs abstractions/idéalisations respectives.⁵⁰

Dans ce cadre, nous n'allons pas fournir de réponse à cette question car cela nous engagerait dans une analyse philosophique plus approfondie qui dépasse le but de notre travail. Un problème plus spécifique est, à notre avis, prioritaire : c'est la question de savoir comment on peut identifier les changements qu'il faut apporter lors de la réalisation de l'analyse de robustesse dans chaque cas particulier afin de pouvoir conclure qu'un certain résultat est robuste, donc objectif d'un point de vue scientifique. Nous avons posé cette question relativement à l'utilisation de l'analyse de robustesse dans la modélisation de l'expression génétique. Nous avons avancé une suggestion à ce propos. Cependant, seule une collaboration étroite avec des biologistes engagés dans l'étude du bruit intracellulaire pourra nous permettre de résoudre ce problème dans le cadre d'un travail de recherche ultérieur.

d'un laser, alors que la microscopie à fluorescence est une technique de microscopie optique permettant d'observer et d'étudier des objets physiques et biologiques divers grâce aux phénomènes de fluorescence.

⁵⁰ Entre autres, voir Culp, (1995, « Objectivity in Experimental Inquiry: Breaking Data-Technique Circles », p. 441) et Nozick (2001, *Invariances: The Structure of the Objective World*, p. 87), qui proposent de parler de degrés de robustesse (ou d'invariance) et de degrés d'objectivité.

Conclusion

Dans ce travail, nous avons proposé une clarification conceptuelle de la notion de hasard en biologie, en particulier dans la caractérisation des phénomènes à l'origine causale de la variation au sein des populations naturelles. Nous avons considéré deux sources de variation : les mutations génétiques et le bruit dans l'expression des gènes. Rappelons les questions que nous avons soulevées à ce sujet au début de notre enquête.

Quelle est la notion de hasard à laquelle les biologistes font référence lorsqu'ils caractérisent les sources causales de la variation génétique et phénotypique, notamment les mutations génétiques et le bruit dans l'expression des gènes ? S'agit-il d'une manière appropriée, à la fois du point de vue conceptuel et empirique, de concevoir le hasard et, si ce n'est pas le cas, quels sont la signification et le rôle de la notion de hasard qu'on doit utiliser pour caractériser de manière pertinente ces phénomènes à l'origine de la variation biologique ? Comment doit-on interpréter les probabilités figurant dans les descriptions formelles de ces phénomènes biologiques, en particulier dans la modélisation stochastique de l'expression des gènes et du bruit qui l'affecte ?

Nous avons analysé toutes ces questions à la lumière des réflexions sur la notion de hasard, à la fois en philosophie et en biologie, et sur la base de l'état actuel de la recherche sur les mutations génétiques et sur le bruit intracellulaire. Une telle analyse nous a conduite à défendre un certain nombre de thèses sur ces questions.

Une des principales thèses défendues dans ce travail concerne la notion de hasard et les mutations génétiques. Nous avons montré que, contrairement aux attaques qui ont été adressées à l'idée propre à la Synthèse Moderne selon laquelle les mutations sont le fruit du hasard par rapport à l'adaptation, la découverte de mécanismes moléculaires qui régulent le taux de mutation en réponse aux changements de l'environnement – les mécanismes mutateurs – ne pose aucun défi à cette idée.

Nous avons d'abord identifié et défini ce que nous appelons la notion de « hasard évolutionnaire » : nous avons montré qu'elle correspond à la notion utilisée par les biologistes de la Synthèse Moderne et qu'elle remonte à la pensée de Darwin. Elle porte sur la relation entre le phénomène de mutation, la sélection et l'adaptation : elle définit le fait qu'une mutation se produit « au hasard » par rapport à sa valeur adaptative pour l'organisme porteur.

Ensuite, nous avons montré que, contrairement à ce que, entre autres, Jablonka et Lamb ont affirmé, la notion de « hasard évolutionnaire » est appropriée, à la fois du point de vue conceptuel et empirique, pour caractériser toute mutation génétique, les mutations dues aux mécanismes mutateurs comprises. Nous avons soutenu que la Synthèse Moderne peut faire face au défi qu'un certain nombre de biologistes, d'historiens et de philosophes de la biologie lui ont lancé grâce à l'intégration des avancées importantes, mais non pas révolutionnaires, quant aux processus qui causent les mutations génétiques. En outre, nous avons souligné à plusieurs reprises que, sur la base d'une reconstruction fidèle de la pensée de Lamarck, les idées de ce dernier ne sont pas véritablement de retour.

En gardant à l'esprit les notions de hasard qui ont vu le jour au cours de l'histoire de la pensée philosophique et scientifique, nous avons conclu que le « hasard évolutionnaire » peut être considéré comme une version en termes spécifiques, biologiques et selon une perspective évolutionnaire, des notions aristotélicienne et cournotienne d'absence d'un dessein et de rencontre de séries causales indépendantes. De même que ces dernières, la notion de « hasard évolutionnaire » désigne un hasard objectif car elle porte sur les caractéristiques inhérentes au processus de mutation. En outre, étant définie en termes de processus, elle a un rôle explicatif relativement aux probabilités et aux fréquences observées des mutations génétiques. Enfin, cette notion de hasard est indifférente quant au caractère fondamentalement déterministe ou indéterministe des mutations qu'elle caractérise, et ne met donc pas nécessairement en cause la possibilité, en principe, de les prédire.

Nous avons reconnu que la pertinence de la notion de « hasard évolutionnaire » dans la caractérisation des mutations génétiques n'empêche pas que les biologistes puissent invoquer d'autres notions de hasard dans ce cadre, notamment celle de hasard subjectif comme ignorance des causes. En revanche, nous avons soutenu que les notions de hasard émergeant de la théorie algorithmique de l'information, notamment celle d'incompressibilité, ne sont pas appropriées dans l'état actuel de la recherche. En effet, les connaissances au sujet des mutations sont encore lacunaires. De plus, au fur et à mesure que la recherche avance, elle révèle des caractéristiques des mutations qui tendent à s'opposer à l'idée de leur imprédictibilité par une méthode universelle de calcul.

En ce qui concerne les mutations génétiques, nous avons aussi identifié une autre notion de hasard que nous appelons « aléatoire (au sens) faible » : nous avons soutenu qu'il s'agit de la notion conceptuellement et empiriquement pertinente pour caractériser les mutations au niveau moléculaire, relativement aux processus physico-chimiques qui les causent. Nous avons montré qu'elle ne correspond pas à la notion de hasard utilisée

habituellement par les biologistes, en particulier par les généticiens moléculaires et par les généticiens des populations lors de l'évaluation du taux de mutation sur la base de modèles approximatifs et idéalisés. Il s'agit plutôt d'une notion qui tient compte de tout biais pouvant avoir un impact sur le processus mutationnel, à la fois en ce qui concerne le moment et le site d'occurrence des mutations.

Nous avons défini la notion d'« aléatoire (au sens) faible » en termes de probabilités, plus spécifiquement par rapport aux caractéristiques du processus stochastique de mutation. C'est pourquoi nous avons soulevé la question de savoir si l'on doit interpréter de manière épistémique ou objective les probabilités dans le cadre des descriptions et des prédictions des mutations génétiques. En effet, selon l'une des hypothèses directrices de notre recherche, seule la réponse à cette question peut nous permettre de comprendre si cette notion de l'aléatoire désigne un hasard objectif ou subjectif et, sur cette base, si elle est compatible ou incompatible avec les doctrines du déterminisme et de l'indéterminisme. Indépendamment de ce point, nous avons soutenu que la notion d'« aléatoire (au sens) faible » joue un rôle explicatif en biologie : soit elle rend compte du caractère probabiliste des mutations en identifiant les lacunes de notre connaissance à leur sujet, soit elle fournit une explication des probabilités et des fréquences observées des mutations en caractérisant les processus physico-chimiques qui les engendrent.

En ce qui concerne la notion de hasard et le bruit dans l'expression des gènes, nous avons montré que les notions de « hasard évolutionnaire » et d'« aléatoire (au sens) faible » sont aussi les plus appropriées dans la caractérisation de cette source de la variation phénotypique. À la lumière des études récentes, théoriques et expérimentales, à ce sujet, nous avons en effet souligné le caractère simplifié et idéalisé des descriptions formelles du bruit intracellulaire et nous avons mis en évidence tout facteur pouvant biaiser, au niveau moléculaire et évolutionnaire, les fluctuations qui affectent l'expression génétique, à la fois du point de vue temporel et en fonction des gènes affectés. Nous sommes arrivés aux mêmes conclusions que dans le cas des mutations génétiques, en ce qui concerne la signification et le rôle de ces deux notions de hasard. Nous avons souligné qu'elles sont parfaitement compatibles lorsqu'elles sont employées relativement à un même phénomène (biologique) et, de plus, qu'ensemble elles en fournissent une description et une explication plus complètes car elles expriment deux points de vue différents.

En ce qui concerne la notion d'« aléatoire (au sens) faible » dans la caractérisation du bruit intracellulaire, nous avons donc soulevé à nouveau la question de savoir si elle désigne un hasard subjectif ou objectif. En suivant toujours l'une des hypothèses directrices de notre

recherche – qui est la correspondance de la signification et du rôle des notions de hasard et de probabilité lorsqu’elles sont utilisées relativement à un même phénomène –, nous avons soutenu la thèse suivante : si l’on arrive à montrer que les probabilités dans les modèles stochastiques de l’expression des gènes sont objectives, alors on peut conclure que le caractère « aléatoire (au sens) faible » du bruit intervenant dans ce processus renvoie à un hasard objectif. À cette fin, nous avons montré que la réalisation de l’analyse de robustesse au niveau des modèles stochastiques de l’expression des gènes permet d’argumenter en faveur d’une interprétation objective des probabilités au niveau du résultat de ce processus. Plus précisément, nous avons soutenu que cette méthode d’analyse permet d’identifier des probabilités robustes et donc objectives : il s’agit des probabilités communes à un ensemble divers de modèles stochastiques de l’expression des gènes. De plus, elle permet de savoir comment on doit décrire ce processus intracellulaire : sa représentation appropriée correspond à la structure commune à l’ensemble divers de modèles utilisés.

Nous avons précisé la notion d’objectivité à laquelle nous faisons référence. En accord avec notre position agnostique du point de vue métaphysique, notamment quant à la nature fondamentalement déterministe ou indéterministe des sources de la variation biologique, elle ne correspond pas à la notion métaphysique d’objectivité absolue. Il s’agit plutôt de l’objectivité au sens de l’invariance sous un ensemble de points de vue différents, c’est-à-dire l’objectivité par rapport à une perspective qui s’est révélée faible et pertinente pour décrire le phénomène d’intérêt. Nous avons soutenu qu’une telle notion d’objectivité est la plus appropriée en science car elle prend en compte le caractère partiel et approximé de tout modèle de la réalité (physique, biologique, etc).

Nous n’avons pas clos le dernier chapitre de notre recherche par la démonstration du fait que la notion d’« aléatoire (au sens) faible » désigne un hasard objectif, à la fois dans le cas des mutations et du bruit intracellulaire. Nous avons plutôt montré comment il est possible d’arriver à cette conclusion, notamment comment on peut appliquer l’analyse de robustesse aux modèles stochastiques de l’expression génétique afin de montrer la robustesse, et donc l’objectivité, des probabilités qui y figurent et, par conséquent, l’objectivité du caractère aléatoire des fluctuations qui affectent ce processus intracellulaire.

La signification et le rôle de la notion de hasard au sein des modèles et des théories en biologie font l’objet de réflexions et de controverses scientifiques et philosophiques au moins depuis la publication de *L’Origine des Espèces* (Darwin, 1859), dont on célèbre cette année le cent-cinquantième anniversaire. Cependant, cette notion fait encore l’objet qu’un questionnement passionnant, en biologie et en philosophie. Par ce travail, nous avons voulu

dissiper l'ambiguïté qui l'entoure lorsqu'elle est associée aux mutations génétiques et au bruit dans l'expression des gènes. Cela ne représente que le point de départ pour une analyse ultérieure de la notion de hasard dans la caractérisation d'autres sources de la variation biologique, génétique et phénotypique (par exemple, au niveau moléculaire, les phénomènes de recombinaison et de transposition, aussi bien que les changements dits épigénétiques comme les modifications des *patterns* de méthylation et, plus généralement, les altérations de la structure de la chromatine). Sur la base d'une telle analyse, il sera intéressant de voir si les notions de « hasard évolutionnaire » et d'« aléatoire (au sens) faible » sont effectivement appropriées dans la caractérisation de tout phénomène à l'origine de la variation, idée qui a pris forme et qui s'est renforcée au cours de notre travail. En outre, en suivant l'avancement de la recherche en biologie, il sera aussi intéressant de préciser la signification et le rôle de la notion de hasard dans la caractérisation de chaque source de la variation afin d'atteindre une compréhension plus fine de ces phénomènes biologiques.

Nous espérons que notre travail aura contribué à un projet beaucoup plus ambitieux, que des philosophes éminents poursuivent depuis longtemps : celui de fournir un cadre conceptuel, général et complet, qui rende compte du hasard dans toute discipline biologique. L'aboutissement de cette quête ne rendra pas cristallines les dynamiques au sein du monde biologique, mais nous permettra peut-être de donner une place au hasard en tant que facteur intervenant dans les phénomènes biologiques.

Glossaire

Notions de calcul des probabilités et de statistique

(en ordre alphabétique)

Ouvrages de référence :

DRESS, François (2004), *Probabilité et statistique de A à Z*, Paris, Dunod
CICHITELLI, G (2001), *Probabilità e statistica*, Rimini, Maggioli Editore

CORRELATION STATISTIQUE ρ :

Liaison ou dépendance linéaire entre deux variables aléatoires X et Y (ou deux caractères statistiques). Son intensité est quantifiée par le coefficient de corrélation linéaire ρ : il est égal au rapport de la covariance des deux variables aléatoires X et Y et du produit non nul de leurs écarts-types σ_X et σ_Y .

$$\rho_{X,Y} = \frac{Cov_{XY}}{\sigma_X \sigma_Y}$$

Le coefficient de corrélation est compris entre -1 et +1. Si le coefficient de corrélation est égal à -1 ou à +1, alors il existe une relation linéaire entre X et Y : $Y = \alpha + \beta X$. Si X et Y sont indépendantes, alors le coefficient de corrélation est égal à 0 (mais la réciproque est fausse). L'existence d'une corrélation entre deux variables aléatoires (ou deux caractères statistiques) n'est pas suffisante pour établir un lien causal entre ces deux variables (ou entre ces deux caractères).

COVARIANCE Cov :

Mesure de la variation simultanée de deux variables aléatoires. La covariance de deux variables aléatoires X et Y est

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E(XY) - E(X)E(Y)$$

$$-\infty < Cov(X, Y) < +\infty ;$$

E étant l'espérance mathématique.

Si $Cov(X, Y) > 0$ X et Y ont tendance à varier dans le même sens ;

Si $Cov(X, Y) < 0$ X et Y ont tendance à varier en sens contraire.

DENSITE DE PROBABILITE :

Fonction réelle positive continue f associée à toute variable aléatoire réelle continue X . L'intégrale de cette fonction est égale à la probabilité que X prenne une valeur appartenant à l'intervalle d'intégration.

$$P(a < X \leq b) = \int_a^b f(t)dt.$$

La densité de probabilité possède la signification intuitive suivante : elle peut être vue comme une version continue d'un histogramme. Si deux points x et $x + \delta$ sont « très voisins », comme la densité est une fonction continue, sa valeur est « à peu près constante » et égale à $f(x)$ sur l'intervalle $[x, x + \delta]$; alors la probabilité d'intervalle $P(x < X \leq x + \delta) = \int_x^{x+\delta} f(t)dt$ est « à peu près » égale au produit $(x + \delta - x)f(x) = f(x)\delta$ (cela est conforme à la signification standard du concept de densité).

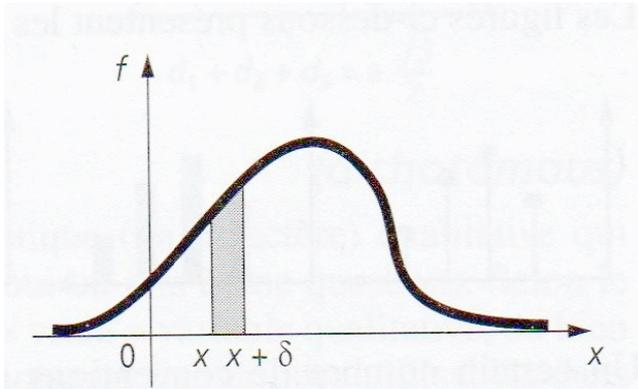


Figure 12 : Densité de probabilité. Figure tirée de Dress, 2004, *op. cit.*, p. 57.

La contrainte selon laquelle la somme des valeurs prises par une fonction de probabilité vaut 1 s'exprime par l'intégrale de la fonction de densité : $\int_{-\infty}^{+\infty} f(x)dx = 1$.

La densité de probabilité est la dérivée de la fonction de répartition : $f(x) = F'(x)$.

DISTRIBUTION DE PROBABILITE :

Dans un contexte probabiliste, synonyme de loi de probabilité.

ECART-TYPE σ :

Indicateur de dispersion attaché à une variable aléatoire réelle. Il est égal à la racine carrée de la variance (σ^2). On note $\sigma(X)$ ou σ_X l'écart-type de la variable aléatoire X .

ESPACE FONDAMENTAL Ω :

Ensemble de tous les événements élémentaires pouvant se produire comme résultats d'une expérience aléatoire. Par exemple, $\Omega = \{\text{Pile-Face, Face-Pile, Face-Face, Pile-Pile}\}$ est l'espace des résultats possibles du lancer de deux pièces.

L'espace fondamental peut contenir un nombre fini (comme dans l'exemple ci-dessus), une infinité dénombrable ou une infinité non dénombrable d'événements élémentaires. Les espaces contenant un nombre fini ou une infinité dénombrable d'éléments sont dits discrets ; ceux constitués par une infinité non dénombrable d'éléments sont dits continus.

ESPACE PROBABILISABLE (Ω, \mathcal{A}) :

Association d'un espace fondamental Ω et d'une famille \mathcal{A} de parties (événements), dont on pourra ultérieurement calculer la probabilité.

ESPACE PROBABILISE (Ω, \mathcal{A}, P) :

Association d'un espace fondamental Ω d'événements élémentaires, d'une famille \mathcal{A} de parties (événements) de Ω et d'une mesure de probabilités P sur \mathcal{A} .

ESPERANCE MATHÉMATIQUE ou valeur attendue d'une variable aléatoire **$E(X)$ ou μ** :

Principal indicateur numérique de tendance centrale attaché à une variable aléatoire réelle. Sa signification concrète est celle d'une moyenne des valeurs prises, pondérée par les probabilités.

Si X est une variable aléatoire discrète réelle caractérisée par l'ensemble (fini ou infini dénombrable) de valeurs $\{x_i\}$, avec les probabilités ponctuelles $p_i = P(X = x_i)$, alors on a :

$$E(X) = \sum_{x_i} x_i P(X = x_i) = \sum_{x_i} x_i p_i.$$

Si X est une variable continue, caractérisée par la densité de probabilité $f(x)$, on a :

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx.$$

L'analogie statistique de l'espérance mathématique est la moyenne μ .

EVENEMENT A :

Ensemble de résultats possibles (ou événements élémentaires) d'une expérience aléatoire, c'est-à-dire sous-ensemble de l'espace fondamental Ω .

Fixons un événement A (par exemple, « pair ») et réalisons une certaine expérience aléatoire (par exemple, le lancer d'un dé) : elle donne pour résultat un certain événement élémentaire $\omega \in \Omega$ (par exemple, « six »). Si $\omega \in A$, on dit que A s'est réalisé (ce qui est le cas dans l'exemple considéré) ; si $\omega \notin A$, on dit que A ne s'est pas réalisé.

Étant donné l'espace probabilisable (Ω, \mathcal{A}) associé à une expérience aléatoire, un événement est une partie A appartenant à l'ensemble \mathcal{A} de toutes les parties de l'espace fondamental Ω . Parmi les événements figurent l'événement vide \emptyset , les « singletons » qui correspondent à un seul événement élémentaire ω , et l'événement certain Ω .

Si p_i est la probabilité de l'événement élémentaire ω_i , autrement dit $p_i = P(\omega_i)$, alors la probabilité de l'événement A est

$$P(A) = \sum_A P(\omega_i) = \sum_A p_i.$$

Par exemple, pour l'événement « pair », on aura pour un dé non biaisé $P(\text{pair}) = p(6) + p(4) + p(2) = 3(1/6) = 1/2$.

EVENEMENT ELEMENTAIRE ω :

Résultat singulier d'une expérience aléatoire. En termes formels, un événement élémentaire est un élément ω de l'espace fondamental Ω . Par exemple, « six » est un événement élémentaire associé au lancer d'un dé.

EXPERIENCE ALEATOIRE ou EPREUVE :

Concept probabiliste qui modélise l'expérimentation ou l'observation d'un phénomène aléatoire (par exemple, le lancer d'un dé, le tirage d'une carte dans un jeu ou d'une boule dans une urne, etc.).

Étant donné un espace probabilisable (Ω, \mathcal{A}) , une expérience aléatoire est définie comme le tirage d'un événement élémentaire ω de Ω , qui est le résultat de l'expérience et qui, contrairement au cas d'une expérience déterministe, n'est pas prédictible. Une expérience aléatoire peut être répétée (une suite de lancers d'un dé, une suite de tirages d'une carte dans un jeu ou d'une boule dans une urne avec remise à chaque tirage, etc.) : ces expériences aléatoires successives sont représentées par des variables aléatoires X_i indépendantes et identiquement distribuées.

FONCTION DE PROBABILITE :

La fonction de probabilité f d'une variable aléatoire X associe à chaque valeur x de cette variable aléatoire sa probabilité.

$$f(x) = P(X = x).$$

Par exemple, dans le cas d'un lancer de dé non biaisé, la fonction de probabilité attribuée à l'événement de s'arrêter sur 5 la valeur $1/6$, soit $f(5) = P(X = 5) = 1/6$.

La fonction de probabilité est caractérisée par les propriétés suivantes :

$$\forall x \quad f(x) \geq 0$$

$$\sum_x f(x) = 1$$

Autrement dit, la fonction de probabilité est toujours positive ; de plus, la somme des probabilités est égale à 1.

FONCTION DE REPARTITION :

Fonction réelle F associée à toute variable aléatoire réelle X :

$$F(x) = P(X \leq x).$$

Il s'agit d'une fonction croissante définie sur \mathfrak{R} dont toutes les valeurs appartiennent à $[0 ; 1]$ et qui prend effectivement les valeurs 0 et 1. Si X est une variable aléatoire discrète, elle varie par sauts ; si X est une variable aléatoire continue, elle varie de manière continue.

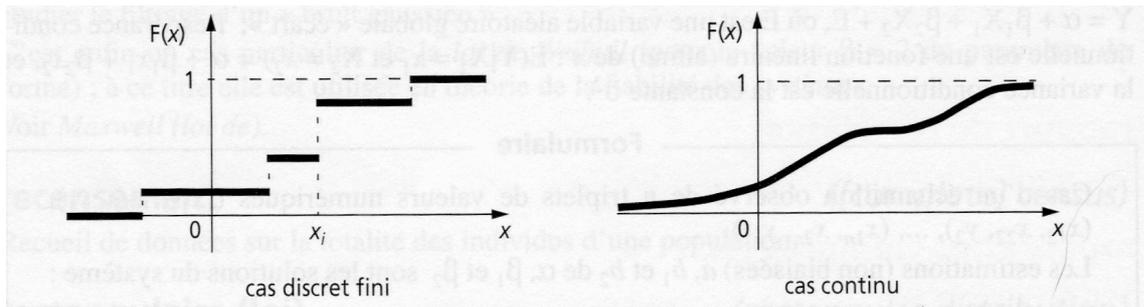


Figure 13 : Fonction de répartition. Figure tirée de Dress, 2004, *op. cit.*, p. 154.

La fonction de répartition est la primitive de la densité de probabilité : $F(x) = \int_{-\infty}^x f(t)dt$.

INDEPENDANCE STATISTIQUE (ou en probabilité) de deux événements :

On dit que deux événements A, B sont indépendants si

$$P(AB) = P(A) P(B).$$

On peut généraliser cette définition en considérant une suite $(A_n)_{n \in I}$ finie ou infinie d'événements : les A_n sont mutuellement indépendants si, pour tout sous-ensemble fini $\{i_1, i_2, \dots, i_k\}$ d'indices de I , on a :

$$P(A_{i1} A_{i2} \dots A_{ik}) = P(A_{i1}) P(A_{i2}) \dots P(A_{ik}).$$

INDEPENDANCE de deux variables aléatoires :

Soient deux variables aléatoires X et Y définies sur un même espace probabilisé (Ω, \mathcal{A}, P) , à valeurs respectivement dans $(\mathcal{E}_X, \mathcal{A}_X)$ et $(\mathcal{E}_Y, \mathcal{A}_Y)$. On dit qu'elles sont indépendantes si on a, pour tout événement $A \in \mathcal{A}_X$ et tout événement $B \in \mathcal{A}_Y$:

$$P(X \in A \text{ et } Y \in B) = P(X \in A) P(Y \in B).$$

On peut généraliser à une suite finie ou infinie de variables aléatoires qui sont définies sur le même espace probabilisé.

LOI DE BERNOULLI :

Une variable aléatoire discrète suit une loi de Bernoulli de paramètre réel p ($0 \leq p \leq 1$) (notation standard : $q = 1 - p$) si

1) elle peut prendre deux valeurs $x = 0, 1$;

2)

$$f(x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \end{cases}$$

p étant la probabilité de la valeur 1 et $1 - p$ étant la probabilité de la valeur 0.

Par exemple, le résultat d'un lancer de pièce codé, Pile par 1 et Face par 0, est une variable aléatoire qui suit une loi de Bernoulli (de paramètre $p = \frac{1}{2}$ si la pièce est bien équilibrée).

LOI BINOMIALE :

Une variable aléatoire discrète suit une loi binomiale de paramètres n (entier ≥ 1) et p réel ($0 \leq p \leq 1$) (notation standard : $q = 1 - p$) si

- 1) elle peut prendre les valeurs $x = 0, 1, \dots, n$;
- 2)

$$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

La loi binomiale sert à calculer la probabilité d'obtenir x succès au cours d'une suite de n épreuves indépendantes où, à chaque épreuve, la probabilité de succès est p et la probabilité d'échec $1 - p$. Cela peut s'appliquer à de nombreux jeux de hasard. Par exemple, cela permet de calculer la probabilité d'obtenir x piles (pile = succès) dans une série de n lancers à pile ou face, ou celle d'obtenir x boules blanches (blanc = succès) dans n tirages avec remise, etc.

LOI EXPONENTIELLE:

Une variable aléatoire continue suit une loi exponentielle de paramètre $\lambda \in \mathfrak{R}_+^*$ si

- 1) elle peut prendre les valeurs $0 \leq x < +\infty$;
- 2)

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \lambda e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

La loi exponentielle décrit le « temps d'attente » du premier événement ou l'intervalle entre deux événements successifs. Plus précisément, c'est la loi des situations concrètes modélisées par un processus de Poisson de taux λ (par exemple, la désintégration d'un atome radioactif, la dégradation de molécules d'ARN messenger et des protéines, etc.) ; dans ces cas, on parle plutôt de « durée de vie » (par exemple, d'un atome, d'un transcrit d'ARN ou d'une protéine).

LOI GEOMETRIQUE :

Une variable aléatoire discrète suit une loi géométrique de paramètre réel p ($0 \leq p \leq 1$) (notation standard : $q = 1 - p$) si

- 1) elle peut prendre les valeurs $x = 1, 2, 3, \dots$;
- 2)

$$f(x) = p(1-p)^{x-1}.$$

La loi géométrique décrit le « temps d'attente » du premier succès dans une suite d'épreuves répétées et indépendantes. La probabilité que le premier succès soit obtenu à la $x^{\text{ème}}$ épreuve quand p est la probabilité d'un succès est $p(1-p)^{x-1}$.

La loi géométrique est analogue à une distribution exponentielle lorsque le temps est représenté de manière discrète.

LOI DES GRANDS NOMBRES :

La « loi des grands nombres » exprime le fait que la fréquence d'un événement converge en probabilités vers sa probabilité lorsque le nombre d'épreuves indépendantes et identiquement distribuées augmente indéfiniment. Ce qui historiquement, à l'époque de Pascal, Fermat et

Huygens, n'était qu'un constat est devenu un théorème mathématique démontrable (Bernoulli, de Moivre). On parle toujours de la « loi des grands nombres ».

Dans sa version faible, elle correspond au théorème suivant :

On considère une suite (X_n) de variables aléatoires (réelles) indépendantes et identiquement distribuées, d'espérance mathématique μ . On définit les moyennes :

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Alors, pour tout $\varepsilon > 0$, on a :

$$P(|M_n - \mu| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Dans sa version forte, la « loi des grands nombres » correspond au théorème suivant :

On considère une suite (X_n) de variables aléatoires (réelles) indépendantes et identiquement distribuées, d'espérance mathématique μ . On définit les moyennes :

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Alors, la probabilité de l'événement $M_n \xrightarrow[n \rightarrow \infty]{} 0$ est égale à 1.

LOI NORMALE ou de GAUSS :

Une variable aléatoire continue suit une loi normale de paramètres $\mu \in \mathfrak{R}$ et $\sigma \in \mathfrak{R}_+^*$ si

- 1) elle peut prendre les valeurs $-\infty < x < +\infty$;
- 2)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ étant la moyenne et σ^2 étant la variance de la variable aléatoire.

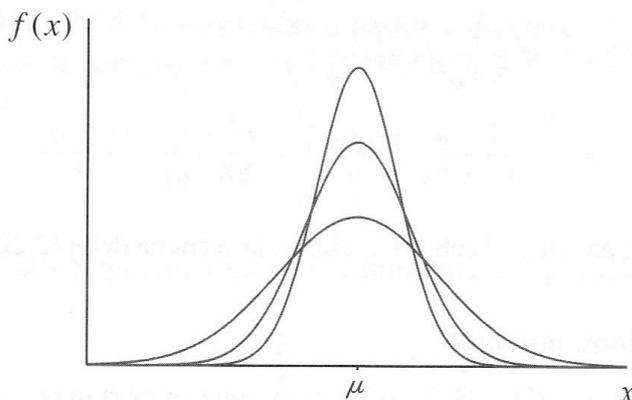


Figure 14 : Loi normale. Figure tirée de Cichitelli, 2001, *op. cit.*, p. 84.

La loi normale possède les propriétés suivantes :

- elle est symétrique, son axe de symétrie étant la droite $x = \mu$;
- elle est croissante dans l'intervalle $(-\infty, \mu)$ et décroissante dans l'intervalle $(+\infty, \mu)$;
- elle a deux points d'inflexion en $x = \mu - \sigma$ et en $x = \mu + \sigma$ (μ étant la moyenne et σ étant l'écart-type) ;
- elle est concave (vers le bas) dans l'intervalle $(\mu - \sigma, \mu + \sigma)$ et convexe ailleurs ;
- elle a comme asymptote l'axe des abscisses.

La loi normale est une approximation de la somme d'un très grand nombre de variables aléatoires qui sont indépendantes. Autrement dit, on parle de loi normale lorsqu'on a affaire à une variable aléatoire continue dépendant d'un très grand nombre de micro-causes dont les effets s'additionnent et dont aucune n'est prépondérante.

LOI DE POISSON :

Une variable aléatoire discrète suit une loi de Poisson de paramètre $\lambda \in \mathfrak{R}_+$ si

- 1) elle peut prendre les valeurs $x = 0, 1, 2 \dots$
- 2)

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}.$$

Elle décrit la probabilité d'un nombre x d'événements survenus jusqu'à un temps fixé dans ce qu'on appelle un processus de Poisson de taux λ , par exemple, la désintégration d'un ensemble d'atomes radioactifs, la dégradation d'un ensemble de molécules d'ARN messager, etc.

LOI DE PROBABILITE :

Dans un contexte probabiliste, synonyme de distribution de probabilité. En termes simples, lorsqu'on fait correspondre aux valeurs d'une variable aléatoire X des valeurs de probabilités, on a une distribution ou loi de probabilité.

Une distribution de probabilité peut être décrite de la manière suivante :

Valeurs de X	$x_1,$	$x_2 \dots$	x_i
Probabilités	$p_1,$	$p_2 \dots$	p_i

x_1, x_2, \dots étant les valeurs possibles de X ,

p_i étant les probabilités correspondantes, c'est-à-dire $p_i = P(X = x_i), i = 1, 2, \dots$

Par exemple, considérons l'expérience aléatoire consistant à lancer deux pièces et l'événement A « le nombre de Pile ». Soit X une variable aléatoire discrète :

Espace Ω	Valeurs prises par X
PP	2
PF	1
FP	1
FF	0

On peut donc faire correspondre les valeurs de X aux valeurs des probabilités : de cette manière, on obtient une distribution de probabilité sur les valeurs de la variable aléatoire X .

Valeurs prises par X	0	1	2 Pile(s)
Probabilités	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

La correspondance entre les valeurs de probabilités et les valeurs de la variable aléatoire X est exprimée par la fonction de probabilité

$$f(x) = P(X = x).$$

LOI UNIFORME :

Une variable aléatoire discrète suit une loi uniforme si

1) elle peut prendre les valeurs $x = 1, 2, \dots, N$;

2)

$$f(x) = 1/N.$$

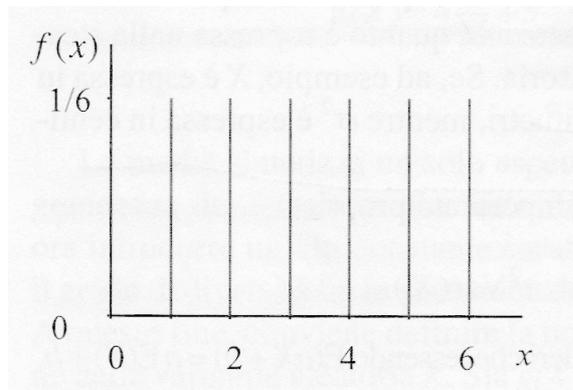


Figure 15 : Loi uniforme discrète. Figure tirée de Cichitelli, *op. cit.*, p. 58.

Par exemple, considérons la variable aléatoire du lancer d'un dé. On a :

$$f(x) = 1/6, x = 1, 2, \dots, 6.$$

Dans le cas d'une variable aléatoire continue, cette dernière suit une loi uniforme si

1) elle peut prendre les valeurs $a \leq x \leq b$;

2)

$$f(x) = 1/b - a.$$

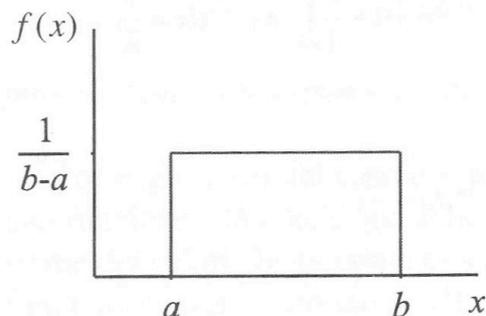


Figure 16 : Loi uniforme continue. Figure tirée de Cichitelli, 2001, *op. cit.*, p. 82.

MARKOV (propriété de) :

Un processus stochastique vérifie la propriété de Markov si et seulement si la distribution conditionnelle de probabilité sur ses états possibles dans le futur, étant donné les états passés et l'état présent, ne dépend que de son état présent et est indépendante de ses états passés.

Pour tout $n \geq 0$, pour toute suite d'états $[(X-n), \dots, (X-1), X]$,

$$P[(X+1) | (X-n), \dots, (X-1), X] = P[(X+1) | X].$$

Un processus markovien est donc un processus « sans mémoire ».

MESURE DE PROBABILITE $P(X)$:

Fonction à valeurs réelles, définie sur l'ensemble \mathfrak{F} , attribuant à chaque événement de l'espace fondamental Ω une valeur de probabilité. On dit qu'une application $P : \mathcal{A} \rightarrow \mathfrak{R}$ est une mesure de probabilité si elle satisfait les trois axiomes suivants :

- Pour tout $A \in \mathcal{A}$, $0 \leq P(A) \leq 1$ (les probabilités sont toujours positives, et inférieures ou égales à 1) ;
- $P(\Omega) = 1$ (l'unité des probabilités est la mesure de l'espace fondamental qui, en tant qu'événement, correspond à l'événement certain) ;
- $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ (si deux événements sont disjoints, leurs probabilités s'additionnent).

La mesure de probabilité P fournit donc la probabilité de chaque événement de \mathcal{A} .

PROBABILITE CONDITIONNELLE :

Probabilité définie en restreignant l'ensemble fondamental Ω à la partie constituant un événement A , de façon à modéliser convenablement les probabilités conditionnées par la survenue de A .

Soient deux événements A, B d'un espace probabilisé (Ω, \mathcal{A}, P) , avec $P(A) \neq 0$. On appelle probabilité conditionnelle de l'événement « B si A » (ou « B sachant A ») le quotient :

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

PROCESSUS DE POISSON :

Modèle mathématique décrivant la réalisation d'événements aléatoires indépendants se succédant au cours du temps (par exemple, des arrivées, des naissances, des incidents, des décès, des désintégrations d'atomes radioactifs, etc.). Plus précisément, il décrit un processus de dénombrement à accroissements indépendants par

- une suite croissante des temps d'arrivée $T_1, T_2, T_3 \dots$, à laquelle on associe les intervalles $T_1, T_2-T_1, \dots, T_{n+1}-T_n$ qui sont des variables aléatoires continues positives, indépendantes entre elles, et suivant une même loi exponentielle de paramètre λ (appelé le taux du processus) ;

- une fonction aléatoire croissante de comptage $N(t)$: il s'agit d'une variable aléatoire discrète à valeurs entières positives qui fait un saut de 1 chaque fois que t franchit une valeur T_n .

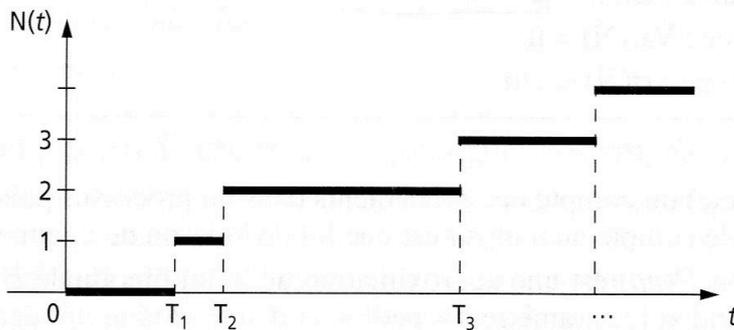


Figure 17 : Processus de Poisson. Figure tirée de Dress, 2004, *op. cit.*, p. 142.

Si le taux du processus est λ , la probabilité qu'un nombre x événements soient survenus au temps t est décrite par une loi de Poisson de paramètre λt .

VARIABLE ALEATOIRE (quantitative ou qualitative) X :

Variable dont la valeur dépend du résultat d'une expérience aléatoire. À chaque événement élémentaire ω de Ω correspond une valeur bien définie : cette correspondance est ce que, dans le vocabulaire ensembliste, on appelle une « fonction » ou une « application » (bien qu'elle ne se note jamais f ou g , mais X ou Y).

En termes formels, une variable aléatoire X est une fonction définie sur un espace Ω qui associe un nombre $X(\omega) = x$ à chaque événement élémentaire ω de Ω .

$$X: \quad \Omega \rightarrow \mathcal{E}$$

$$\omega \mapsto X(\omega)$$

\mathcal{E} étant l'ensemble des « valeurs prises » par la variable aléatoire X .

Par exemple, considérons le lancer de deux pièces et le nombre de Pile(s). L'espace $\Omega = \{PP, PF, FP, FF\}$ est constitué par les quatre événements élémentaires du lancer de deux pièces. Définissons l'application X de Ω dans \mathfrak{N} :

$X(\omega) =$ nombre de Pile(s) de l'événement ω . C'est une variable aléatoire. On a, par exemple, $X(FF) = 0$, $X(PP) = 2$, etc.

Étant donné un espace probabilisé (Ω, \mathcal{A}, P) et une variable aléatoire à valeurs dans \mathcal{E} , on peut obtenir une mesure de probabilité P sur \mathcal{E} .

Une variable aléatoire est discrète si l'espace Ω sur lequel elle est définie est discret ; dans ce cas, X peut se voir attribuer un nombre fini ou une infinité dénombrable de valeurs. Une variable aléatoire est continue si l'espace Ω sur lequel elle est définie est continu : Ω comprend alors une infinité indénombrable d'événements élémentaires.

VARIANCE d'un échantillon statistique Var_X ou s^2_X :

Principal indicateur numérique de dispersion. Elle est définie comme la moyenne numérique des carrés des écarts entre les valeurs observées et leur moyenne.

Formule pour n observations individualisées x_1, x_2, \dots, x_n , la moyenne \bar{x} ayant été préalablement calculée :

$$s^2 = \frac{(x_1 + \bar{x})^2 + (x_2 + \bar{x})^2 + \dots + (x_n + \bar{x})^2}{n} = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \bar{x})^2.$$

VARIANCE d'une variable aléatoire $\text{Var}(X)$ ou $\sigma^2(X)$:

Principal indicateur de dispersion attaché à une variable aléatoire réelle. Elle est définie comme la moyenne du carré de l'écart entre la variable et son espérance mathématique, pondérée par les probabilités.

Si la variable aléatoire réelle X est discrète, caractérisée par l'ensemble (fini ou dénombrable) de valeurs $\{x_i\}$, avec les probabilités ponctuelles $p_i = P(X = x_i)$, et si son espérance mathématique est $\mu = E(X)$, on a :

$$\text{Var}(X) = E((X - E(X))^2) = \sum_{x_i} p_i (x_i - \mu)^2.$$

Si X est continue, caractérisée par la densité de probabilité $f(x)$, et si son espérance mathématique est $\mu = E(X)$, on a :

$$\text{Var}(X) = E((X - E(X))^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx.$$

Bibliographie

- ALBERTS, B, JOHNSON, A, LEWIN, J, RAFF, M, ROBERTS, K & WALTER, P (2002[1983]), *Molecular Biology of the Cell*, New York, Garland Science Taylor & Francis Group.
- ALPHEY, L (1997), *DNA Sequencing: From Experimental Methods to Bioinformatics*, BIOS scientific publications, Oxford, Springer.
- ARISTOTE, *Les premiers analytiques*, traduction française par J Tricot, Paris, Vrin, 1936.
- ARISTOTE, *De l'interprétation*, traduction française par J Tricot, Paris, Vrin, 1966.
- ARISTOTE, *La métaphysique*, traduction française par J Tricot, Paris, Vrin, 1970.
- ARISTOTE, *La physique*, traduction française par A Stevens, Paris, Vrin, 1999.
- ASTAUROV, BL (1974[1930]), Study of hereditary deviations from bilateral symmetry, associated with variation of metameric structures, *In* PF Rokitskii (ed.), *Nasledstvennost' i razvities*, Moscow, Nauka Press, 54-109.
- ASTERIS, G & SARKAR, S (1996), Bayesian procedures for the estimation of mutation rates from fluctuation experiments, *Genetics* 142, 313-326.
- ATMANSPACHER, H & BISHOP, R (eds.) (2002), *Between Chance and Choice: Interdisciplinary Perspectives on Determinism*, Thorverton, Imprint Academic.
- BALABAN, NQ, MERRIN, J, CHAIT, R, KOWALIK, L & LEIBLER, S (2004), Bacterial Persistence as a Phenotypic Switch, *Science* 305, 1622-1625.
- BALIBAR, E & MACHEREY, P (2008), Determinisme, *In Encyclopaedia Universalis* 7, Paris.
- BARBEROUSSE, A (2000), *La physique face à la probabilité*, Paris, Vrin.
- BAR-EVEN, A, PAULSSON, J, MAHESHRI, N, CARMÍ, M, O'SHEA, E, PILPEL, Y & BARKAI, N (2006), Noise in protein expression scales with natural protein abundance, *Nature Genetics*, 38, 636-643.
- BARTHOLOMAY, AF (1957), *A stochastic approach to chemical reaction kinetics*, PhD Thesis, Harvard University.
- BEARDMORE, JA (1960), Developmental stability in constant and fluctuating temperatures, *Heredity* 14, 411-422.
- BEATTY, J (1984), Chance and Natural Selection, *Philosophy of Science* 51, 183-211.

- BEATTY, J (1987), The probabilistic revolution in evolutionary biology. An overview, *In* L Krüger, G Gigerenzer & MS Morgan, (eds.), *The Probabilistic Revolution*, vol. II *Ideas in the Sciences*, Cambridge, The MIT Press, 229.
- BEATTY, J (1992), Random Drift, *In* E Fox Keller & L Lloyd (eds.), *Keywords in Evolutionary Biology*, Cambridge, Harvard University Press.
- BEATTY, J (1995), The Evolutionary Contingency Thesis, *In* G Wolters & JG Lennox (eds.), *Concepts, Theories, and Rationality in the Biological Sciences*, The Second Pittsburgh-Konstanz Colloquium in the Philosophy of Science, Pittsburgh, University of Pittsburgh Press.
- BEATTY, J (2006), Chance Variation: Darwin on Orchids, *Philosophy of Science* 73, 692-641.
- BEATTY, J (2008), Chance Variation and Evolutionary Contingency: Darwin, Simpson (*The Simpsons*) and Gould, *In* M Ruse (ed.), *Oxford Handbook of the Philosophy of Biology*, Oxford, Oxford University Press.
- BECKSEI, A, SERAPHIN, B & SERRANO, L (2001), Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion, *EMBO Journal* 20, 2528-2535.
- BECKSEI, A & SERRANO, L (2000), Engineering stability in gene networks by autoregulation, *Nature* 405, 590-593.
- BECKSEI, A, KAUFMANN, BB & VAN OUDENAARDEN, A (2005), Contributions of low molecule number and chromosomal positioning to stochastic gene expression, *Nature Genetics* 37, 937-944.
- BENNETT, CH (1988a), Démons, machines et thermodynamique, *Pour La Science*, Janvier, 91-97.
- BENNETT, CH (1988b), Logical Depth and Physical Complexity, *In* R Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford, Oxford University Press, 227- 257.
- BENNETT, CH (1990), How to define complexity in physics, and why, *In* WH Zureck (ed.), *Complexity, Entropy and the Physics of Information*, SFI Studies in the Sciences of Complexity, Volume VIII, Redwood City, Addison-Wesley, 137-148.
- BENSON, SA, OCCI, JL & SAMPSON, BA (1988), Mutations that alter the pore function of the OmpF porin of *Escherichia coli* K12, *Journal of Molecular Biology* 203, 961-970.
- BENZER, S & FREESE, E (1958), Induction of specific mutations with 5-bromouracil, *Proceedings of the National Academy of Sciences USA* 44, 112-119.
- BENZER, S (1959), On the topology of genetic fine structure, *Proceedings of the National Academy of Sciences USA* 45, 1607-1520.

- BENZER, S (1961), On the topology of genetic fine structure, *Proceedings of the National Academy of Sciences USA* 47, 403-415.
- BENZER, S (1962), The fine structure of the gene, *Scientific American* 206, 70-84.
- BERG, OG (1978), A model for statistical fluctuations of protein numbers in a microbial-population, *Journal of Theoretical Biology* 71, 587-603.
- BERGSON, H (1948[1932]), *Les deux sources de la morale et de la religion*, Paris, Presses Universitaires de France.
- BERNARD, C (1865), *Introduction à l'étude de la médecine expérimentale*, Paris, Baillière & Fils.
- BERNARD, C (1966[1878]), *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux*, Paris, Vrin.
- BERNOULLI, Jacques (1713), *Ars Conjectandi*, Basileae.
- BERTRAND, J (1889), *Calcul des probabilités*, Paris, Gauthier-Villars.
- BISHOP, R (2003), On Separating Prediction from Determinism, *Erkenntnis* 58, 169-188.
- BISHOP, RC (2006), Determinism and Indeterminism, In DM Borchert (ed.), *Encyclopedia of Philosophy, Second Edition* 3, Farmington Hills, Macmillian Reference, 29-35.
- BLAKE, WJ, KÆRN, M, CANTOR, CR & COLLINS, JJ (2003), Noise in eukaryotic gene expression, *Nature*, 422, 633-637.
- BLOCH, O & VON WARTBURG W (1996), *Dictionnaire étymologique de la langue française*, Paris, Presses Universitaires de France.
- BOE, L (1990), Mechanism of induction of adaptive mutations in *Escherichia coli*, *Molecular Microbiology* 4, 597-601.
- BOHM, D (1952), A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables, I and II, *Physical Review* 85, 166-193.
- BOREL, E (1924 [1909]), *Éléments de la théorie des probabilités*, Paris, Hermann.
- BOWLER, PJ (1983), *The Eclipse of Darwinism. Anti-Darwinian Evolution Theories in the Decades around 1900*, Baltimore, The Johns Hopkins University Press.
- BRADING, K & CASTELLANI, E (2008), Symmetry and Symmetry Breaking, In EN Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <<http://plato.stanford.edu/archives/fall2008/entries/symmetry-breaking/>>.
- BRADSHAW, AD (1965), Evolutionary significance of phenotypic plasticity in plants, *Advances in Genetics* 13, 115-155.

- BRAILLARD, P-A (2008), *Enjeux philosophiques de la biologie des systèmes*, thèse de doctorat, Université Paris 1 Panthéon - Sorbonne.
- BRANDON, R (1984), On the concept of environment in evolutionary biology, *Journal of Philosophy* 81, 613-665.
- BRANDON, R (1990), *Adaptation and Environment*, Princeton, Princeton University Press.
- BRANDON, R (1997a), Does Biology has Laws? The experimental evidence, *Philosophy of Science* 1996, 2, 444-457.
- BRANDON, R (2005), The Difference Between Selection and Drift: A Reply to Millstein, *Biology and Philosophy* 20, 153-170.
- BRANDON, R & CARSON, S (1996) The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables' Proof But No Room for Determinism Either, *Philosophy of Science* 63, 315-337.
- BUETTNER, VL, HILL, KA, SCARINGE, WA & SOMMER, SS (2000), Evidence that proximal multiple mutations in Big Blue[®] transgenic mice are dependent events, *Mutation Research* 452, 219-229.
- CAIRNS, J & FOSTER, PL (1991), Adaptive Reversion of a Frameshift Mutation in *Escherichia coli*, *Genetics* 128, 695-701.
- CAIRNS, J, OVERBAUGH, J & MILLER, S (1988). The origin of mutants, *Nature* 355, 142-145.
- CALDER, N (1973), *The Life Game*, London, British Broadcasting Corporation.
- CALUDE, CS (ed.) (2007) *Randomness and complexity from Leibniz to Chaitin*, World Scientific Publishing Company.
- CARNAP, R (1962 [1950]), *Logical Foundations of Probability*, Chicago, The University of Chicago Press.
- CARRIER, M (1995), Evolutionary Change and Lawlikeness: Beatty on Biological Generalizations, In G Wolters & JG Lennox (eds.), *Concepts, Theories, and Rationality in the Biological Sciences*, Konstanz, Universitätsverlag Konstanz.
- CAVALLI-SFORZA, LL and LEDERBERG, J (1956), Isolation of pre-adaptive mutants of bacteria by sib-selection, *Genetics* 41, 367-381.
- CHAITIN, GJ (1966), On the Length of Programs for Computing Finite Binary Sequences, Statistical Considerations, *J.A.C.M.* 16, 145-159. Repris dans Chaitin (1987).
- CHAITIN, GJ (1975a), A Theory of Program Size Formally Identical to Information Theory, *J.A.C.M* 22, 329-340, Repris dans Chaitin (1987).

- CHAITIN, GJ (1975b), Randomness and Mathematical Proof, *Scientific American* 232, 47-52.
- CHAITIN, GJ (1979), Toward a Mathematical Definition of Life, In RD Levine & Tribus M (eds.), *The Maximum Entropy Formalism*, 477-498. Repris dans Chaitin (1987).
- CHAITIN, GJ (1987), *Information, Randomness and Incompleteness: Papers on Algorithmic Information Theory*, Singapore, World Scientific.
- CHAITIN, GL (2006), Complexité, logique et hasard. Entretien avec Gregory Chaitin, In R Benkirane (dir.), *La complexité, vertiges et promesses*, Paris, Le Pommier, 275-295.
- CHARPENTIER, E, GHYS, EJ-L & LESNE, A (dir.) (2006), *L'héritage scientifique de Poincaré*, Paris, Belin.
- CHOU, WH, HALL, KJ, WILSON, DB, WIDEMAN, CL, TOWSON, SM, CHADWELL, LV & BRITT, SG (1996), Identification of a novel Drosophila opsin reveals specific patterning of the R7 and R8 photoreceptor cells, *Neuron* 17, 1101-1115.
- CHURCH, A (1940), On the concept of a random sequence, *Bulletin of the American Mathematical Society* 46, 130-135.
- CICHITELLI, G (2001), *Probabilità e statistica*, Rimini, Maggioli Editore.
- COHEN, L (2005), The history of noise [on the 100th anniversary of its birth], *Signal Processing Magazine IEEE* 22, 20-45.
- CONCHE, M (2005), La métaphysique du hasard, In J-P Delahaye (dir.), *Le hasard. Une idée, un concept, un outil*, Paris, L'Harmattan.
- COOK, DL, GERBER, AN & TAPSCOTT, SJ (1998), Modeling stochastic gene expression: implications for haploinsufficiency, *Proceedings of the National Academy of Sciences USA* 95, 15641-15646.
- COPELAND, JB & PROUDFOOT, D (1999), Alan Turing's Forgotten Ideas in Computer Science, *Scientific American* 280, 98-103.
- COURNOT, A-A (1984[1843]), *Exposition de la théorie des chances et des probabilités*, Paris, Vrin
- COURNOT, A-A (1975[1851]), *Essai sur les fondements de nos connaissances et sur le caractère de la critique philosophique*, Paris, Hachette.
- COURNOT, A-A (1979[1875]), *Matérialisme, vitalisme, rationalisme*, Paris, Vrin.
- CRICK, FHC (1958), On Protein Synthesis, *Symposia of the Society for Experimental Biology*. 12, 138-163.
- CRICK, FHC (1970), Central Dogma of Molecular Biology, *Nature* 227, 561-563.

- CROW, JF (1972), The dilemma of nearly neutral mutations: How important are they for evolution and human welfare?, *Journal of Heredity* 63, 306-316.
- CROW, JF (1981), The neutralist-selectionist controversy: an overview, *In* EB Hook (ed.) *Population and Biological Aspects of Human Mutation*, New York, Academic Press, 3-14.
- CULP, S (1994), Defending Robustness: the Bacterial Mesosome as a Test Case, *In* D Hull, M Forbes & RM Burian (eds.), *PSA-1994* 1, East Lansing, The Philosophy of Science Association, 46-57.
- CULP, S (1995), Objectivity in Experimental Inquiry: Breaking Data-Technique Circles, *Philosophy of Science* 62, 438-458.
- DALMEDICO AD, CHABERT, J-L & CHEMLA, K (dir.) (1992), *Chaos & déterminisme*, Paris, Seuil.
- DARDEN, L (2006), *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly*, Cambridge, Cambridge University Press.
- DARWIN, F & SEWARD, AC (eds.) (1903), *More Letters of Charles Darwin*, Vol I, London, John Murray.
- DARWIN, C (1859), *On The Origin of Species By Means of Natural Selection or The Preservation of Favoured Races in The Struggle For Life*, London, John Murray (first edition).
- DARWIN, C (1868), *The Variation of Animals and Plants Under Domestication*, (2 vols), London, John Murray (first edition).
- DASTON, L (1988), *Classical Probability in the Enlightenment*, Princeton, Princeton University Press.
- DAVID, P & SAMADI, S (2000), *La théorie de l'évolution. Une logique pour la biologie*, Paris, Flammarion.
- DAVIS, BD (1989), Transcriptional bias: A non-Lamarckian mechanism for substrate induced mutations, *Proceedings of the National Academy of Sciences USA* 86, 5005-5009
- DAWKINS, R (1999 [1982]), *The Extended Phenotype. The Long Reach of the Gene*, Oxford, Oxford University Press.
- DEBRE, P (1996), *Jacques Monod*, Paris, Flammarion.
- DEBS, TA & REDHEAD, MLG (2007), *Objectivity, Invariance, and Convention: Symmetry in Physical Science*, Cambridge, Harvard University Press.
- DE FINETTI, B (1931[1989]), Probabilism, *In* (1989) *Erkenntnis*, 31, 169-223.

- DE FINETTI, B (1933[1931]), On the Subjective Meaning of Probability, English translation
In De Finetti, 1993, *Induction and Probability*, Bologna, Clueb, 291-321.
- DELAHAYE, J-P (1999), *Information, complexité et hasard*, Paris, Hermès.
- DELAHAYE, J-P (dir.) (2005), *Le hasard : une idée, un concept, un outil*, Collection « Les rendez-vous d'Archimède », Paris, L'Harmattan.
- DELAHAYE, J-P & ZENIL, H (2007), On the Kolmogorov-Chaitin complexity for short sequences, In CS Calude (ed.) *Randomness and Complexity: From Leibniz to Chaitin*, Singapore, World Scientific.
- DELAHAYE, J-P & ZENIL, H (2008), Towards a stable definition of Kolmogorov-Chaitin complexity, *Fundamenta Informaticae XXI*, 1-15.
- DELSOL, M (1989), *Cause, Loi, Hasard en Biologie*, Paris, Vrin.
- DE VRIES, H (1901[1910]), *The mutation theory. Experiments and observations of the origin of species in the vegetable kingdom*, London, Kegan Paul, Trench, Trubner & Co.
- DHAR, N & MCKINNEY, JD (2007), Microbial phenotypic heterogeneity and antibiotic tolerance, *Current Opinion in Microbiology* 10, 30-38.
- DIETRICH, M (1994), The Origins of the Neutral Theory of Molecular Evolution, *Journal of the History of Biology* 27, 21-59.
- DIETRICH, M (1998), Paradox and Persuasion: Negotiating the Place of Molecular Evolution Within Evolutionary Biology, *Journal of the History of Biology* 31, 289-291.
- DIETRICH, M & MILLSTEIN, RL (2008), The Role of Causal Processes in the Neutral and Nearly Neutral Theories, *Philosophy of Science*. 75, 548-559.
- DIJKMANS, R, KREPS, S & MERGEAY, M (1994), Poisson-like fluctuation patterns of revertants of leucine auxotrophy (*leu-500*) in *Salmonella typhimurium* caused by delay in mutant cell division, *Genetics* 137, 353-359.
- DOBZHANSKY, T (1951[1937]), *Genetics and The Origin of Species*, New York, Columbia University Press.
- DOBZHANSKY, T (1970), *Genetics of the Evolutionary Process*, New York, Columbia University Press.
- DOBZHANSKY, T, AYALA, FJ, STEBBINS, GL & VALENTINE, JW (1977), *Evolution*, San Francisco, WH Freeman & Company.
- DOBZHANSKY, T (1974), Chance and Creativity in Evolution, In FJ Ayala & T Dobzhansky (eds.), *Studies in the Philosophy of Biology*, London, The Macmillan Press Limited, 307-337.

- DRAKE, JW (1991a), Spontaneous mutation, *Annual Review of Genetics* 25, 125-146.
- DRAKE, JW (1991b), A constant rate of spontaneous mutation in DNA-based microbes, *Proceedings of the National Academy of Sciences USA* 88, 7160-7164.
- DRAKE, JW (1992), Mutation rates, *Bioessays* 14, 137-140.
- DRAKE, JW (1993), Rates of spontaneous mutation among RNA viruses, *Proceedings of the National Academy of Sciences* 90, 4171-4175.
- DRAKE, JW (2007a), Mutations in clusters and showers, *Proceedings of the National Academy of Sciences USA* 104, 8203-8204.
- DRAKE, JW (2007b) Too Many Mutants with Multiple Mutations, *Critical Reviews in Biochemistry and Molecular Biology* 42, 247-258.
- DRAKE, JW, BEBENEK, A, KISSLING, GE & PEDDADA, S (2005), Clusters of mutations from transient hypermutability, *Proceedings of the National Academy of Sciences USA* 102, 12849-12854.
- DRAKE, JW, CHARLESWORTH, B, CHARLESWORTH, D & CROW, JF (1998), Rates of spontaneous mutation, *Genetics* 148, 1667-1686.
- DRESS, F (2004), *Probabilités et statistique de A à Z*, Paris, Dunod.
- DROUET, I (2004), *La théorie propensionniste des probabilités*, mémoire de D.E.A., Université Paris 1 Panthéon - Sorbonne.
- DUBININ, NP & ROMASCHOFF, DD (1932), The Genetic Structure of Species and their Evolution, (traduction en anglais du russe) *Biologichesky Zhurnal* 1, 52-95.
- DUNN, LC (1965), *A Short History of Genetics. The development of some of the main lines of thought: 1864-1939*, McGraw-Hill Book Company.
- EARMAN, J (1986), *A Primer on Determinism*, Dordrecht, Reidel Publishing Company.
- EBLE, GJ (1999), On the dual nature of chance in evolutionary biology and palaeology, *Palaeobiology* 25, 75-87.
- EKELAND, I (1995), *Le chaos*, Paris, Flammarion/Dominos.
- ELOWITZ, MB, LEVINE, AJ, SIGGIA, ED & SWAIN, PS (2002), Stochastic Gene Expression in a Single Cell, *Science* 297, 1183-1186.
- EL SAMAD, H, KHAMMASH, M, PETZOLD, L & GILLESPIE, DT (2005), Stochastic modelling of gene regulatory networks, *International Journal of Robust and Nonlinear Control* 15, 691-711.

- FAYE, J (2008), Copenhagen Interpretation of Quantum Mechanics, *In* Zalta EN (ed.) *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <<http://plato.stanford.edu/archives/fall2008/entries/qm-copenhagen/>>.
- FELLER, W 1950, *An Introduction to Probability Theory and Its Applications*, New York, Wiley.
- FETZER, JH (1982), Probabilistic Explanations, *Philosophy of Science* 2, 194-207.
- FISHER, RA (1999[1930]), *The Genetical Theory of Natural Selection: a complete variorium edition*, Oxford, Oxford University Press.
- FISHER, RA & FORD, EB (1950), The 'Sewall Wright Effect', *Heredity* 4, 117-119.
- FISCHER, EP & LIPSON, C (1988), *Thinking about Science: Max Delbrück and the Origins of Molecular Biology*, New York, WW Norton.
- FOSTER, PL (1999), Mechanisms of Stationary Phase Mutation: A Decade of Adaptive Mutation, *Annual Review of Genetics* 33, 57-88.
- FOSTER, PL (2004), Adaptive Mutation in *Escherichia coli*, *Journal of Bacteriology* 186, 4846-4852.
- FOSTER, PL (2006), Methods for Determining Spontaneous Mutation Rates, *Methods of Enzymology* 409, 195-213.
- FOSTER, PL & CAIRNS, J (1992), Mechanisms of Directed Mutation, *Genetics* 131, 783-789.
- FOSTER, PL & ROSCHE, WA (1999), Increased episomal replication accounts for the high rate of adaptive mutation in *recD* mutants of *Escherichia coli*, *Genetics* 152, 15-30.
- FOX KELLER, E (1990), Physics and the emergence of molecular biology: A history of cognitive and political synergy, *Journal of the History of Biology* 23, 389-410.
- FOX KELLER, E (2003[2000]), *Le siècle du gène*, Paris, Gallimard.
- FOX KELLER, E & LLOYD EA (eds.) (1992), *Keywords in Evolutionary Biology*, Cambridge, Harvard University Press.
- FRANCESCHELLI, S, ROQUE, T & MICHEL, P (dir.) (2007), *Chaos et systèmes dynamiques. Éléments pour une épistémologie*, Paris, Hermann, Visions des sciences.
- FRASER, HB, HIRSH, AE, GIAEVER, G, KUMM, J & EISEN, MB (2004), Noise Minimization in Eukaryotic Gene Expression, *PloS Biology* 2, 0834-0838.
- FRECHET, MR (1938), *Méthode des fonctions arbitraires, théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles*, Paris, Gauthier-Villars.

- FRIEDBERG EC, WALKER GC, SIEDE W, WOOD RD, SCHULTZ RA & ELLENBERGER T (2006), *DNA Repair and Mutagenesis*, Second Edition, Washington, ASM Press.
- FRIGG, R & HARTMANN, S (2008), Models in Science, EN Zalta (ed.), *Models in Science, The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <<http://plato.stanford.edu/archives/fall2008/entries/models-science/>>.
- FUTUYMA, DJ (1998[1979]), *Evolutionary Biology*, Sunderland MA, Sinauer Associates.
- GALAVOTTI, MC (2005), *A Philosophical Introduction to Probability*, Stanford, CSLI Publications.
- GAUVRIT, N (2005) Les irrémédiables structures du hasard humain, In J-P Delahaye (dir.) *Le hasard : une idée, un concept, un outil*, Coll. « Les rendez-vous d'Archimède ». Paris, L'Harmattan, 128-145.
- GAYON, J (1992), *Darwin et l'après-darwin : une histoire de l'hypothèse de sélection naturelle*, Paris, Kimé.
- GAYON, J (1993), La biologie entre loi et histoire, *Philosophie* 38, 30-57.
- GAYON, J (1994), Le hasard dans la théorie évolutionniste moderne : une analyse philosophique, *Bulletin de la société d'histoire et d'épistémologie des sciences de la vie* 1, 4-16.
- GAYON, J (1998), Le déterminisme : origines d'un mot, évaluation d'une idée, In M Lesieur (dir.), *Turbulence et déterminisme*, Grenoble, EDP Sciences.
- GAYON, J (2004), Realism and Biological Science, In M Carrier, J Roggenhofer, G Küppers & Ph Blanchard (eds.), *Knowledge and the World: Challenges Beyond the Science Wars*, Heidelberg, Springer.
- GAYON, J (2005), Chance, Explanation, and Causation in Evolutionary Theory, *History and Philosophy of the Life Sciences* 27, 387-397.
- GAYON, J (2006a), Les réflexions méthodologiques de Claude Bernard : structure, contexte, origines, In M Bitbol & J Gayon (dir.), *L'épistémologie française 1830-1970*, Paris, Presses Universitaires de France, 431-463.
- GAYON, J (2006b), Hérité des caractères acquis, In P Corsi, J Gayon & G Gohau, *Lamarck, philosophe de la nature*, Paris, Presses Universitaires de France, 105-163.
- GERDES, K, CHRISTENSEN, SK & LOBNER-OLESEN, A (2005), Prokaryotic toxin antitoxin stress response loci, *Nature Reviews Microbiology* 3, 371-382.
- GIERE, RN (1999), Using Models to Represent Reality, In L Magnani, NJ Nersessian & P Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*, Dordrecht, Kluwer.
- GIERE, RN (2006), *Scientific Perspectivism*, Chicago, The University of Chicago Press.

- GIGERENZER, G, SWIJTINK, Z, PORTER, T, DASTON, L, BEATTY J & KRÜGER, L (eds.) (1989), *The Empire of Chance: How Probability Changed Science and Everyday Life*, Cambridge, Cambridge University Press.
- GILLESPIE, DT (2000), The chemical Langevin equation, *Journal of Chemical Physics* 113, 297-306.
- GILLESPIE, DT (2007), Stochastic Simulation of Chemical Kinetics, *Annual Review of Physical Chemistry* 58, 35-55.
- GILLESPIE, J (2004), *Population Genetics: A Concise Guide*, Baltimore, Johns Hopkins University Press.
- GILLIES, D (2000) *Philosophical theories of probability*, Nortfolk, Routledge.
- GIRAUD, A, RADMAN, M, MATIC, T & TADDEI, F (2001). The rise and the fall of mutator bacteria, *Current Opinion in Microbiology* 4, 582-585.
- GLEICK, J (1989), *La théorie du chaos*, Paris, Albin Michel.
- GLENNAN, S (1997), Probable Causes and the Distinction between Subjective and Objective Chance, *Noûs* 31, 496-519.
- GLICKMAN, BW & RADMAN, M (1980), *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction, *Proceedings of the National Academy of Sciences USA* 77, 1063-7.
- GOLDSTEIN, S (2008), Bohmian Mechanics, In ED Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2008) Edition*, URL = <<http://plato.stanford.edu/archives/spr2009/entries/qm-bohm/>>.
- GOODWIN, B (1994), *How The Leopard Changed Its Spots: The Evolution of Complexity*, Charles Scribner & Sons.
- GOULD, SJ (1989[1991]), *La vie est belle. Les surprises de l'évolution*, Paris, Seuil.
- GOULD, SJ & ELDREDGE, N (1972), Punctuated Equilibria: an Alternative to Phyletic Gradualism, In T Schopf (ed.) *Models in Paleobiology*, San Francisco, Freeman Cooper & Co.
- GOULD, SJ & LEWONTIN, RC (1979), The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme, *Proceedings of the Royal Society of London, Series B* 205, 1161, 581-598.
- GRAVES, L, HORAN, BL & ROSENBERG, A (1999), Is Indeterminism the Source of the Statistical Character of Evolutionary Theory?, *Philosophy of Science* 66, 140-157.
- GRIFFITHS, PE & GRAY, RD (1994), Developmental Systems and Evolutionary Explanation, *The Journal of Philosophy* 91, 277-304.

- GRIFFITHS, PE & GRAY, RD (2004), The Developmental Systems Perspective. Organism-environment systems as units of development and evolution, *In* M Pigliucci & K Preston (eds.), *Phenotypic Integration – Studying the Ecology and the Evolution of Complex Phenotypes*, Oxford, Oxford University Press.
- GRIFFITHS, PE & STOTZ, K (2007), Gene, *In* D Hull & M Ruse (eds.), *Cambridge Companion to Philosophy of Biology*, Cambridge, Cambridge University Press.
- HACKING, I (1983), *Representing and Intervining*, Cambridge, Cambridge University Press.
- HACKING, I (1995[1990]), *The Taming of Chance*, Cambridge, Cambridge University Press.
- HACKING, I (2002[1975]), *L'Emergence de la probabilité*, Paris, Seuil.
- HACKING, I (2004[2001]), *L'Ouverture au probable*, Paris, Armand Colin.
- HAIG, D (2006), Weismann Rules! Epigenetics and Lamarckian temptation, *Biology and Philosophy* 22, 415-428.
- HAJEK, A 2007, The Reference Class Problem Is Your Problem Too, *Synthèse* 156, 185-215.
- HAJEK, A (2008), Interpretations of Probability, *In* EN Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2008 Edition)*, URL = <<http://plato.stanford.edu/archives/win2008/entries/probability-interpret/>>.
- HALDANE, JBS (1966[1932]), *The Causes of Evolution*, New York, Cornell University Press.
- HALL, BG (1988), Adaptive evolution that requires multiple spontaneous mutations. I. Mutations involving an insertion sequence, *Genetics* 120, 887-897.
- HALL, BG (1990), Spontaneous point mutations that occur more often when advantageous than when neutral, *Genetics* 126, 5-16.
- HALLGRIMSSON B & HALL BK (eds.) (2005), *Variation. A Central Concept in Biology*, San Diego, Elsevier Academic Press.
- HAMMOCK, EAD & YOUNG, L 2005, Microsatellite instability generates diversity in brain and sociobehavioral traits, *Science* 308, 1630-1634.
- HAMPEL, FR, RONCHETTI, EM, ROUSSEUW, PJ & STAHEL, WA (1986), *Robust Statistics – The Approach Based on Influence Functions*, Wiley & Sons.
- HÄNGGI, P (2002), Stochastic Resonance in Biology. How Noise Can Enhance Detection of Weak Signals and Help Improve Biological Information Processing, *ChemPhysChem. A European Journal of Chemical Physics and Physical Chemistry* 3, 285-290.
- HARRIS, H (1976), Molecular evolution: the neutralist-selectionist controversy, *Federation Proceedings* 35, 2079-2082.

- HARTWELL, LH & WEINERT, TA (1989), Checkpoints: controls that ensure the order of the cell cycle events, *Science* 246, 629-634.
- HASTY, J, MCMILLEN, D, ISAACS, F & COLLINS, JJ (2001), Computational studies of gene regulatory networks: *in numero* molecular biology, *Nature Reviews Genetics* 2, 268-279.
- HASTY, J, PRADINES, J, DOLNIK, M & COLLINS, JJ (2000), Noise-based switches and amplifiers for gene expression, *Proceedings of the National Academy of Sciences USA* 97, 2075-2080.
- HEISENBERG, W (1958[1971]), *Physique et philosophie*, Paris, Albin Michel.
- HENDRICKSON, H, SLECHTA, ES, BERGTHORSSON, U, ANDERSSON, DI & ROTH, JR. (2002), Amplification Mutagenesis: evidence that growth with a selected gene amplification causes adaptative mutation and hyper-mutability, *Proceedings of the National Academy of Sciences USA* 99, 2164-2169.
- HILGEOORD, J & UFFINK, J (2008), The Uncertainty Principle, In EN Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <<http://plato.stanford.edu/archives/fall2008/entries/qt-uncertainty/>>.
- HILL, KA, WANG, J, FARWELL, KD, SCARINGE WA & SOMMER, SS (2004), Spontaneous multiple mutations show both proximal spacing consistent with chronocoordinate events and alterations with p53-deficiency, *Mutation Research* 554, 223-240.
- HODGE , MJS (1987), Natural Selection as a Causal, Empirical, and Probabilistic Theory, In L Krüger (ed.), *The Probabilistic Revolution*, Cambridge, The MIT Press, 233-270.
- HOEFER, C (2007), The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance, *Mind* 116, 549-596.
- HOEFER, C (2008), Causal Determinism, In EN Zalta (ed.), *The Stanford Eyclopedia of Philosophy (Winter 2008 Edition)*, URL = <<http://plato.stanford.edu/archives/win2008/entries/determinism-causal/>>.
- HUBER, PJ (1981), *Robust Statistics*, Wiley & Sons.
- HULL, D (1974), *The Philosophy of Biological Science*, Englewood Cliffs, Prentice-Hall.
- HUMPHREYS, P (2004), *Extending Ourselves: Computational Science, Empiricism and Scientific Method*, New York, Oxford University Press.
- HUXLEY, J (1948[1942]), *Evolution. The Modern Synthesis*, London, George Allen & Unwin Ltd.
- HUXLEY, J (1953), *Evolution in Action*, New York, Harper & Brothers.

- ISAACS, FJ, HASTY, J, CANTOR, CR & COLLINS, JJ (2003), Prediction and measurement of an autoregulatory genetic module, *Proceedings of the National Academy of Sciences USA* 100, 7714-7719.
- ISMAEL, J (2008), Quantum Mechanics, In EN Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Winter 2008 Edition), URL = <<http://plato.stanford.edu/archives/win2008/entries/qm/>>.
- JABLONKA, E & LAMB, M (1995), *Epigenetic Inheritance and Evolution*, Oxford, Oxford University Press.
- JABLONKA, E & LAMB, M (2005), *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, Cambridge, The MIT Press.
- JACOB, F (1981), *Le jeu des possibles*, Paris, Fayard.
- JACOB, F & MONOD, J (1961), Genetic Regulatory Mechanisms in the Synthesis of Proteins, *Journal of Molecular Biology* 3, 318-356.
- JAEGER, G & SARKAR, S (1995), On the distribution of bacterial mutants: The effects of differential fitness of mutants and non-mutants, *Genetica* 96, p. 217-223.
- JOHNSON, T (1999), Beneficial Mutations, Hitchhiking and the Evolution of Mutation Rates in Sexual Populations, *Genetics* 151, 1621-1631.
- JONES, ME, THOMAS, SM & ROGERS A (1994), Luria-Delbrück fluctuation experiments: design and analysis, *Genetics* 136:1209–1216.
- JORDANOVA, LJ (1984), *Lamarck*, Oxford, Oxford University Press.
- JUDSON, HF (1979) *The Eighth Day of Creation. The Makers of the Revolution in Biology*, New York, Simon & Schuster.
- KÆRN, M, ELSTON, TC, BLAKE, WJ & COLLINS, JJ (2005), Stochasticity in gene expression: from theories to phenotypes, *Nature Reviews Genetics* 6, 451-464.
- KAUFFMAN, S (1993), *The origins of order: Self-Organization and Selection in Evolution*, New-York, Oxford University Press.
- KAY, LE (1993) *The Molecular Vision of Life: Caltech, The Rockefeller Foundation, and the Rise of the New Biology*, Oxford, Oxford University Press.
- KELLIS M, BIRREN BW & LANDER ES (2004), Proof and evolutionary analysis of ancient genome duplication in the yeast *Sccharomyces cerevisiae*, *Nature* 428, 617-624.
- KEPLER, TB & ELSTON, TC (2001), Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations, *Biophysical Journal* 81, 3116-3136.
- KEREN, I, KALDALU, N, SPOERING, A, WANG, Y & LEWIS, K (2004a), Persister cells and tolerance to antimicrobials, *FEMS Microbiol Lett* 230, 13-18.

- KEREN, I, SHAH, D, SPOERING, A, KALDALU, N & LEWIS, K (2004b), Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*, *Journal of Bacteriology* 186, 8172-8180.
- KEYNES, JM (1963[1921]), *Treatise on Probability*, London, Macmillan.
- KIERZEK, AM, ZAIM, J & ZIELENKIEWICZ, P (2001), The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression, *Journal of Biological Chemistry* 276, 8165-8172.
- KIMURA, M, (1967), On the evolutionary adjustment of spontaneous mutation rates, *Genetic Research* 9, 23-34.
- KIMURA, M (1968a), Evolutionary rate at the molecular level, *Nature* 217, 624-626.
- KIMURA, M (1968b), Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles, *Genetics Research Cambridge* 11, 247-269.
- KIMURA, M (1990[1983]) *La théorie neutraliste de l'évolution*, Paris, Flammarion.
- KIMURA, M (1991), Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics, *Proceedings of the National Academy of Sciences USA* 88, 5969-5973.
- KIMURA, M & CROW, JF (1964), The number of alleles that can be maintained in a finite population, *Genetics* 49, 725-738.
- KIMURA, T & OHTA, T (1969a), The average number of generations until fixation of a mutant gene in a finite population, *Genetics* 61, 763-771.
- KIMURA, M & OHTA T (1969b), The average number of generations until extinction of an individual mutant gene in a finite population, *Genetics* 63, 701-709.
- KING, JL & JUKES, TH (1969) Non-Darwinian Evolution, *Science* 164, 788-798.
- KISTLER, M (2004), La causalité dans la philosophie contemporaine, *Intellectica* 38, 139-185.
- KISTLER, M (2006), La causalité comme transfert et dépendance nomique, *Philosophie* 89, 53-77.
- KISTLER, M (à paraître), La causalité, In A Barberousse, D Bonnay & M Cozic (dir.) *Précis de philosophie des sciences*, Paris, Vuibert, 2009.
- KLINGENBERG, CP (2005), Developmental Constraints, Modules, and Evolvability, In Hallgrímsson & Hall (eds.), *Variation. A Central Concept in Biology*, 219-247.

- KLUG W, CUMMINGS M & SPENCER C (2006), *Génétique*, Huitième édition, Paris, Pearson Education France.
- KOJEVE, A (1990[1932]), *L'idée du déterminisme dans la physique classique et dans la physique moderne*, Paris, Librairie générale française.
- KOLMOGOROV, AN (1956[1933]), *Foundations of the theory of probability*, New York, Chelsea Publishing Company.
- KOLMOGOROV, AN (1965), Three Approaches for Defining the Concept of Information Quantity, *Information Transmission* 1, 3-11.
- KRÜGER, L, DASTON, LJ & HEIDELBERG, M (eds.) (1987), *The Probabilistic Revolution*, Cambridge, The MIT Press.
- KÜPPERS, B-O (1990), *Information and the Origin of Life*, Cambridge, The MIT Press.
- KUSSELL, E, KISHONY, R, BALABAN, NQ & LEIBLER, S (2005), Bacterial Persistence: A Model of Survival in Changing Environments, *Genetics* 169, 1807-1814.
- KUSSELL, E & LEIBLER, S (2005), Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments, *Science* 309, 2075-2078.
- LALAND, KN, ODLING-SMEE, FJ & FELDMAN, MW (2003), *Niche Construction: The Neglected Process in Evolution*, Princeton, Princeton University Press.
- LALANDE, A (2002[1926]), *Vocabulaire technique et critique de la philosophie*, Paris, Presses Universitaires de France.
- LAMARCK, J-B (1809), *Philosophie zoologique, ou Exposition des considérations relatives à l'histoire naturelle des animaux*, Paris, Dentu.
- LAMOTTE, M (1988) Phénomènes fortuits et évolution, In M Marois (dir.), *L'évolution dans sa réalité et ses diverses modalités*, Paris, Masson, 241-268.
- LAPLACE, P-S (1878[1812]), *Théorie analytique des probabilités*, Paris, Gauthier Villars.
- LAPLACE, P-S (1986[1814]) *Essai philosophique sur les probabilités*, Paris, Bourgois.
- LATTER, BDH (1964), Selection for a threshold character in *Drosophila*. I. An analysis of the phenotypic variance of the underlying scale, *Genetic Research* 5, 198-210.
- LEA, DE & COULSON, CA (1949), The distribution of the numbers of mutants in bacterial populations, *Journal of Genetics* 49, 264-285.
- L'ECUYER, P & SIMARD, R (2007), TestU01: A C Library for Empirical Testing of Random Number Generators, *ACM Transactions on Mathematical Software* 33, 4, Article 22.

- LEDERBERG, J & LEDERBERG, E (1952), Replica plating and indirect selection of bacterial mutants, *Journal of Bacteriology* 63, 399-406.
- LEIBNIZ, GW (1696[1710]), *Essai de théodicée : sur la bonté de Dieu, la liberté de l'homme et l'origine du mal*, Paris, Garnier-Flammarion.
- LEISNER, M, STINGL, K, FREY, E & MAIER, B (2008), Stochastic switching to competence, *Current Opinion in Microbiology* 11, 553-559.
- LENAY, C (1989), *Enquête sur le hasard dans les grandes théories biologiques de la deuxième moitié du XIX^e siècle*, Thèse de doctorat, Université Paris 1 Panthéon - Sorbonne.
- LENAY, C (2003), Les limites naturelles de la durée de vie et la question de l'hérédité de l'acquis, *Etudes sur la mort. L'Esprit du temps* 124, 2, 43-58.
- LENNOX, J (2008), Darwinism, In EN Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <<http://plato.stanford.edu/archives/fall2008/entries/darwinism/>>.
- LENSKI, RE & MITTLER, JE (1993), The Directed Mutation Controversy and neo-Darwinism, *Science* 259, 188-194.
- LENSKI, RE & SNIEGOWSKI, PD (1995), 'Adaptive Mutation': The Debate Goes On, *Science* 269, 285-286.
- LENSKI, RE, MONTGOMERY, S & AYALA, FJ (1989a), Mutation and selection in bacterial population: Alternatives to the hypothesis of directed mutation, *Proceedings of the National Academy of Sciences USA* 86, 2775-2778.
- LENSKI, RE, MONTGOMERY, S & AYALA, FJ (1989b), Another alternative to directed mutation, *Nature* 337, 123-124.
- LESTIENNE, R (1993), *Le hasard créateur*, Paris, La Découverte.
- LEVIN, LA (1974), Laws of Information Conservation and Aspects of the Foundation of Probability Theory, *Problems of Information Transmission* 10, 206-210.
- LEVINS, R (1966), The Strategy of Model Building in Population Biology, *American Scientist* 54, 412-413.
- LEWIN, B (2004), *Genes VIII*, Upper Saddle River, Pearson Prentice Hall.
- LEWIS, D (1986), A Subjectivist's Guide to Objective Chance, *Philosophical Papers*, Vol. II, 83-132, Oxford, Oxford University Press.
- LEWIS, K (2005), Persister cells and the riddle of biofilm survival, *Biochemistry (Mosc)* 70, 267-274.

- LEWONTIN, R (1970), The Units of Selection, *Annual Reviews of Ecology and Systematics* 1, 1-18.
- LEWONTIN, R (1974), *The Genetic Basis of Evolutionary Change*, New York, Columbia University Press.
- LEWONTIN, RC (1983), The organism as the subject and object of evolution, *Scientia* 188, 65-82.
- LEWONTIN, RC (2000), *The Triple Helix: Gene, Organism, and Environment*, Cambridge, Harvard University Press.
- LI, M & VITANYI, P (1997), *An Introduction to Kolmogorov complexity and its applications*, New York, Springer-Verlag.
- LIM, MM, WANG, Z, OLAZABAL, DE, REN, X, TERWILLIGER, EF & YOUNG, LJ (2004), Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene, *Nature* 429, 754-755.
- LLOYD, S & PAGELS, H (1988), Complexity as Thermodynamic Depth, *Annals of Physics* 188, 186-213.
- LOEWER, B (2001), Determinism and Chance, *Studies in History and Philosophy of Modern Physics* 32, 609-620.
- LOPEZ-MAURY, L, MARGUERAT, S & BÄHLER, J (2008), Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation, *Nature Reviews Genetics* 9, 583-593.
- LOSICK, R & DESPLAN, C (2008), Stochasticity and cell fate, *Science* 320, 65-68.
- LUCRECE, *De rerum natura*, traduction française par B Pautrat, Paris, Librairie Générale Française, 2002.
- LUDWIG, W (1932), *Das Rechts-Links Problem im Tierreich und beim Menschen*, Berlin, Springer.
- LURIA, SE & DELBRÜCK, M (1943), Mutations of Bacteria From Virus Sensitivity to Virus Resistance, *Genetics* 28, 491-511.
- MAAMAR, H, RAJ, A & DUBNAU, D (2007), Noise in gene expression determines cell fate in *Bacillus subtilis*, *Science* 317, 526-529.
- MACHAMER, P, DARDEN, L & CRAVER, C (2000), Thinking About Mechanisms, *Philosophy of Science* 67, 1-25.
- MAHESHRI, N & O'SHEA, EK (2007), Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression, *Annual Review of Biophysics and Biomolecular Structures* 36, 413-434.

- MALATERRE, C & MERLIN, F (à paraître), L'(in)déterminisme de l'évolution naturelle : quelles origines pour le caractère stochastique de la théorie de l'évolution ?, In F Athané, T Heams, P Huneman, G Lecointre & M Silberstein, *Les mondes darwiniens*, Paris, Syllepse, 2009.
- MARKOW, TA (ed.) (1993), *International conference on developmental instability: its origin and evolutionary implications*, Tempe, Arizona, 14-15 June 1993, Dordrecht, Kluwer academic publications.
- MARTIN, T (1996), *Probabilités et critique philosophique selon Cournot*, Paris, Vrin.
- MARTIN-LÖF, P (1966a), On the Concept of a Random Sequence, *Theory of Probability and its Applications*, 11, 177-179.
- MARTIN-LÖF, P (1966b), The definition of random sequences, *Information and Control* 9, 602-619.
- MARTIN-LÖF, P (1969), The Literature on von Mises' Kollektivs Revisited, *Theoria* XXXV, 12-37.
- MASON, LG, EHRLICH, PR & EMMEL, TC (1967), The population biology of the butterfly, *Euphydryas editha*. V. Character clusters and asymmetry, *Evolution* 21, 85-91.
- MATHER, K (1953a), The genetical structure of populations, *Symposia for the Society of Experimental Biology* 7, 66-95.
- MATHER, K (1953b), Genetical control of stability in development, *Heredity* 7, 297-336.
- MAUPERTIUS, P-LM de (1751[1984]), *Système de la nature*, Paris, Vrin.
- MAYNARD-SMITH, JM, BURIAN, R, KAUFFMAN, S, ALBERCH, P, CAMPBELL, J, GOODWIN, B, LANDE, E, RAUP, D & WOLPERT, L (1985), Developmental constraints and evolution, *Quarterly Review of Biology*, 60, 265-287.
- MAYR, E (1942), *Systematics and the Origin of Species*, New York, Columbia University Press.
- MAYR, E (1954), Change of genetic environment and evolution, In J Huxley, AC Hardi & EB Ford, *Evolution as a Process*, London, George Allen & Unwin.
- MAYR, E (1963), *Animal Species and Evolution*, Harvard University Press.
- MAYR, E (1974[1970]), *Populations, espèces et évolution*, Paris, Hermann.
- MAYR, E (1976), *Evolution and the Diversity of Life. Selected Essays*, Belknap Press of Harvard University.
- MAYR, E (1988), *Towards A New Philosophy of Biology*, Cambridge, Harvard University Press.

- MAYR, E (1989[1982]), *Histoire de la biologie. De Darwin à nos jours*, vol. 1 et 2, Paris, Fayard.
- MCADAMS, HH & ARKIN, A (1997), Stochastic mechanisms in gene expression, *Proceedings of the National Academy of Sciences USA* 94, 814-819.
- MELLOR DH (1971) *The Matter of Chance*, Cambridge, Cambridge University Press.
- MELLOR, DH (2005), *Probability. A Philosophical Introduction*, London, Routledge.
- MENDEL, G (1866[1982]), Experiments in Plants Hybridization, traduit de l'allemand Versuch über Pflanzenghybriden, *Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1895, Adhandlungen*, 3-47, <http://www.netspace.org/MendelWeb/>, 1996.
- MERLIN, F (2009), Pour une interprétation objective des probabilités dans les modèles stochastiques de l'expression génétique, In J-J Kupiec, O Gandrillon, M Morange & M Silberstein (dir.), *Le hasard au cœur de la cellule. Probabilités, déterminisme, génétique*, Paris, Syllepse.
- MERLIN, F (à paraître), On Griffiths and Gray's concept of expanded and diffused inheritance, *Biology and Philosophy*.
- METTETAL, JT, MUZZEY, D, PREDAZA, JM, OZBUDAK, EM & VAN OUDENAARDEN, A (2006), Predicting stochastic gene expression dynamics in single cells, *Proceedings of the National Academy of Sciences USA* 103, 7304-7309.
- METTETAL JT & VAN OUDENAARDEN, A (2007), Necessary Noise, *Science*, 317, 46-47.
- MIKELADZE-DVALI, T, WERNET, MF, PISTILLO, D, MAZZONI, EO, TELEMAN, AA, CHEN, YW, COHEN, S & DESPLAN, C (2005), The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in Drosophila R8 photoreceptors, *Cell* 122, 775-87.
- MILL, JS (1866[1843]), *Système de logique déductive et inductive, exposé des principes de la preuve et des méthodes de recherche scientifique*, Paris, Ladrance.
- MILLER, D (1994), *Critical Rationalism. A Restatement and Defence*, Illinois, Open Court.
- MILLSTEIN, RL (1996), Random Drift and the Omniscient Viewpoint, *Philosophy of Science* 63 (3 Suppl), S10-S18.
- MILLSTEIN, RL (1997), *The Chances of Evolution: An Analysis of the Roles of Chance in Microevolution and Macroevolution*, PhD Thesis, University of Minnesota.
- MILLSTEIN, RL (2000), Chance and Macroevolution, *Philosophy of Science* 67, 603-624.
- MILLSTEIN, RL (2002), Are Random Drift and Natural Selection Conceptually Distinct?,

Biology and Philosophy 17, 33-53.

- MILLSTEIN, RL (2006), Discussion of 'Four Case Studies on Chance in Evolution: Philosophical Themes and Questions', *Philosophy of Science* 73, 678-687.
- MILLSTEIN, RL & SKIPPER, RA (2007), Population Genetics, In D Hull & M Ruse (eds.), *The Cambridge Companion to the Philosophy of Biology*, Cambridge, Cambridge University Press, 22-43.
- MØLLER, AP & SWADDLE, JP (1997), *Asymmetry, Developmental Stability, and Evolution*, Oxford, Oxford University Press.
- MOMBAERTS, P (2005), Odorant receptor gene choice in olfactory sensory neurons: the one receptor-one neuron hypothesis revisited, *Current Opinion in Neurobiology* 14, 31-36.
- MONOD, J (1970), *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, Paris, Seuil.
- MONOD, J (1974), On Chance and Necessity, In FJ Ayala & T Dobzhansky (eds.), *Studies in the Philosophy of Biology*, London, The Macmillan Press Limited.
- MONTAGUE, R (1974), Deterministic Theories, In RH Thomason (ed.), *Formal Philosophy*, New Haven, Yale University Press, 303-359.
- MORANGE, M (2003[1994]), *Histoire de la biologie moléculaire*, Paris, La Découverte.
- MORANGE, M (2005), *Les secrets du vivant : Contre la pensée unique en biologie*, Paris, La Découverte/Sciences et société.
- MORGAN, TH (1988[1925]), *The Genetics of Drosophila*, New York, Garland Publishing Inc.
- MORGAN, TH (1988[1926]), *The Theory of The Gene*, New York, Garland Publishing, Inc.
- MOXON, ER, RAINEY, PB, NOWAK, MA & LENSKI RE (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria, *Current Biology* 4, 1, 24-33.
- MOXON, ER, BAYLISS, C & HOOD, D (2006), Bacterial Contingency Loci: the Role of Simple Sequence DNA Repeats in Bacterial Adaptation, *Annual Review of Genetics* 40, 307-333.
- NAGEL, E (1961), *The Structure of Science. Problems in the Logic of Scientific Explanation*, Indianapolis, Cambridge, Hackett Publishing Company.
- NATHANS, J (1999), The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments, *Neuron* 24, 299-312.
- NEWCOMBE, H (1949), Origin of bacterial variants, *Nature* 164, 150-151.

- NEWMAN, JRS, GHAEMMAGHAMI, S, IHMELS, J, BRESLOW, DK, NOBLE, M, DERIS JL & WEISSMAN JS (2006), Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise, *Nature* 441, 840-846.
- NINIO, J (1991), Transient Mutators: A Semiquantitative Analysis of the Influence of Translation and Transcription Errors on Mutation Rates, *Genetics* 129, 957-962.
- NINIO, J (1996), Gene conversion as a focusing mechanism for correlated mutations: a hypothesis, *Molecular & General Genetics* 251, 503-508.
- NOVICK, A & SZILARD, L (1950), Experiments with the chemostat on spontaneous mutations of bacteria, *Proceedings of the National Academy of Sciences USA* 36, 708-719.
- NOZICK, R (2001), *Invariances: The Structure of the Objective World*, Cambridge, Belknap Press.
- OLBY, R (1974), *The Path to The Double Helix*, London, MacMillan.
- OLLER, AR & SCHAAPER, RM (1994), Spontaneous mutation in *Escherichia coli* containing the *dnaE911* DNA polymerase antimutator allele, *Genetics* 138, 263-270.
- OPPENHEIM, AB, KOBILER, R, STAVANS, J, COURT, DL & ADHYA, S (2005), Switches in bacteriophage lambda development, *Annual Review of Genetics* 39, 409-429.
- ORRELL, D & BOLOURI, H (2004) Control of internal and external noise in genetic regulatory networks, *Journal of Theoretical Biology* 230, 301-312.
- ORTOLI, S & PHARABOD, J-P (1998), *Le cantique des quantiques*, Paris, La Découverte.
- OYAMA, S, GRIFFITHS, PE & GRAY, RD (eds.) (2001), *Cycles of Contingency. Developmental Systems and Evolution*, Cambridge, The MIT Press.
- OZBUDAK, EM, THATTAI, M, KURTSEY, I, GROSSMAN, AD & VAN OUDENAARDEN, A (2002), Regulation of noise in the expression of a single gene, *Nature Genetics* 31, 69-73.
- OZBUDAK, EM, THATTAI, M, LIM, HN, SHRAIMAN, BI & VAN OUDENAARDEN, A (2004), Multistability in the lactose utilization network of *Escherichia coli*, *Nature* 427, 737-740.
- PALMER, AR (1994), Fluctuating Asymmetry Analyses: A Primer, In TA Markow (ed.) *International conference on developmental instability: its origin and evolutionary implications*, Tempe, Arizona, 14-15 June 1993, Dordrecht, Kluwer academic publications, 335-364.
- PALMER, AR (1996), Waltzing with asymmetry, *Bioscience* 46, 518-532.

- PALMER, AR & STROBECK, C (1986), Fluctuating asymmetry: measurement, analysis, patterns, *Annual Review of Ecology and Systematics* 17, 391-421.
- PAULSSON, J (2005), Models of stochastic gene expression, *Physics of Life Reviews* 2, 157-175.
- PEARSON, K (1901), Mathematical contributions to the theory of Evolution. IX. On the Principle of Homotyposis and its relation to heredity, to the variability of the individual and to that of the Race. Part 1. Homotyposis in the vegetable Kingdom, *Philosophical Transactions of the Royal Society London* 1 197, 285-379.
- PEARSON, K (1911), *The Grammar of Science*, London, A & C Black.
- PEIRCE, CS (2001[1923]), *Chance, love, and logic: philosophical essays*, London, Routledge.
- POINCARÉ, H (1987[1912]) *Calcul des probabilités*, Paris, Gauthiers-Villars.
- POINCARÉ, H (1920[1908]), *Science et méthode*, Paris, Flammarion, Paris.
- POLAK, M (ed.) (2003), *Developmental Instability. Causes and Consequences*, Oxford University Press US.
- POMIAN, K (dir.) (1990), *La querelle du déterminisme. Philosophie de la science d'aujourd'hui*, Paris, Gallimard.
- POPPER, K (1934), *Logic of Scientific Discovery*, London, Hutchinson.
- POPPER, K (1959), The Propensity Interpretation of Probability, *British Journal for the Philosophy of Science* 10, 25-42.
- POPPER, K (1984[1982]), *L'univers irrésolu. Plaidoyer pour l'indéterminisme*, Paris, Hermann.
- POPPER, K (1990[1982]), *Le réalisme et la science*, Paris, Hermann.
- POPPER, K (1992[1990]), *Un univers de propensions. Deux études sur la causalité et l'évolution*, Paris, L'Eclat.
- POPPER, K 1992, *Quantum Theory and the Schism in Physics*, New York, Routledge.
- PORTUGAL, FH & COHEN, JS (1977), *A Century of DNA. A History of the Discovery of the Structure and Function of the Genetic Substance*, Cambridge, The MIT Press.
- PROVINE, WB (1985), The R.A. Fisher-Sewall Wright Controversy and its Influence Upon Modern Evolutionary Biology, In R Dawkins & M Ridley (eds.), *Oxford Surveys in Evolutionary Biology*, New York, Oxford University Press, 197-219.
- PSILLOS, S (2003), *Causation and Explanation*, Ithaca, New York, McGill Queens University Press.

- PTASHNE, M (2003), *A Genetic Switch: Phage Lambda Revisited*, Cold Spring Harbor Laboratory Press.
- PTASHNE, M (2007), On the use of the word 'epigenetic', *Current Biology* 17, R233-6.
- RADMAN M (1973), Phenomenology of an Inducible Mutagenic DNA Repair Pathway in *Escherichia coli*: SOS repair hypothesis, *In* L Prakash, F Sherman, M Miller, C Lawrence, HW Tabor (eds.), *Molecular and Environmental Aspects of Mutagenesis*, Springfield, Charles C. Thomas, 128-142.
- RADMAN, M (1999), Enzymes of evolutionary change, *Nature* 401, 866-869.
- RADMAN, M, TADDEI, F & MATIC, I (2000), Evolution-driving genes, *Research in Microbiology* 151, 91-95.
- RADMAN, M & WAGNER, R (1986). Mismatch Repair in *Escherichia coli*, *Annual Review of Genetics* 20, 523-538.
- RADMAN, M & WAGNER, R (1988) The high fidelity of DNA duplication, *Scientific American* 259, p. 40-46.
- RAIDL, E (à paraître), A method of arbitrary functions: A rehabilitation of classical probability?
- RAJ, A, PESKIN, CS, TRANCHINA, D, VARGAS, DY & TYAGI, S (2006), Stochastic mRNA synthesis in mammalian cells, *PLoS Biology* 4, e309.
- RAMSEY, FP (1926), Truth and Probability, *In* FP Ramsey (1931), *The foundations of mathematics and other Logical Essays*, Routledge & Kegan Paul.
- RAO, CV, WOLF, DM & ARKIN, AP (2002), Control, exploitation and tolerance of intracellular noise, *Nature* 420, 231-237.
- RASER JM & O'SHEA EK (2004), Control of Stochasticity in Eukaryotic Gene Expression, *Science* 304,1811-1814.
- RASER JM & O'SHEA EK (2005), Noise in Gene Expression: Origins, Consequences, and Control, *Science* 309, 2010-2013 (Supporting Online Material: www.sciencemag.org/cgi/content/full/309/5743/2010/DC1).
- REEVE, ECR (1960), Some genetic tests on asymmetry of sternopleural chaeta number in *Drosophila*, *Genetic Research* 1, 151-172.
- REEVE, ECR & ROBERTSON, FW (1953), Analysis of environmental variability in quantitative inheritance, *Nature* 171, 874-875.
- REEVE, ECR & ROBERTSON, FW (1954), Studies in quantitative inheritance. VI. Sternite chaeta number in *Drosophila*; a metamerie quantitative character, *Z. Indukt. Abstamm. Vererbungsl.* 86, 269-288.

- REICHENBACH, H (1949), *The Theory of Probability*, San Francisco, University of California Press.
- RIDLEY, M (2004), *Evolution*, 3rd edition, Malden MA, Blackwell Scientific Publishing.
- RIGNEY, DR (1979), Stochastic model of constitutive protein levels in growing and dividing bacterial cells, *Journal of Theoretical Biology* 76, 453-480.
- RIGNEY, DR & SCHIEVE, WC (1977), Stochastic model of linear, continuous protein synthesis in bacterial populations, *Journal of Theoretical Biology* 69, 761-766.
- ROBERTS, JT (2006), Determinism, In S Sarkar & J Pfeifer (eds.), *The Philosophy of Science. An Encyclopedia* 1, New York, Routledge.
- ROSCHÉ, WA & FOSTER, PL (2000), Determining Mutation Rates in Bacterial Populations, *Methods* 10, 4-17.
- ROSENBERG, A (1994), *Instrumental Biology or the Disunity of Science*, Chicago, University of Chicago Press.
- ROSENBERG, A (1988), Is the Theory of Natural Selection Really a Statistical Theory?, *Canadian Journal of Philosophy*, suppl. V, 14, 187-206.
- ROSENBERG, A (2006), *Darwinian Reductionism, or How to Stop Worrying and Love Molecular Biology*, Chicago, University of Chicago Press.
- ROTH, JR & ANDERSSON, DI (2004), Adaptive Mutation: How Growth under Selection Stimulates Lac⁺ Reversion by Increasing Target Copy Number, *Journal of Bacteriology* 186, 4855-4860.
- RUELLE, D (1991), *Hasard & chaos*, Paris, Editions Odile Jacob, Collection Opus 89.
- RUFFIE, J (1976), *De la biologie à la culture*, Paris, Flammarion.
- RUFFIE, J (1982), *Traité du vivant*, Paris, Fayard.
- RYAN, FJ, NAKADA, D & SCHNEIDER, MJ (1961), Is DNA replication a necessary condition for spontaneous mutation?, *Z. Vererbungslehre* 92, 38-41.
- RYAN, FJ, OKADA, T & NAGATA, T (1963), Spontaneous mutation in spheroplasts of *Escherichia coli* », *Journal of Genetics and Microbiology* 30, 193-199.
- SAINT-SERNIN, B (1998) *Cournot. Le réalisme*, Paris, Vrin.
- SAINT-SERNIN, B (2008), Hasard, *Encyclopedia Universalis* 11, Paris.
- SALMON, WC (1977) Objectively Homogeneous Reference Classes, *Synthese* 36, 399-414.
- SALMON, WC (1988), *Causality and Explanation*, New York, Oxford University Press.

- SAMOILOV, MS, PLYASUNOV, S & ARKIN, AP (2005), Stochastic and signaling in enzymatic futile cycles through noise-induced bistability with oscillations, *Proceedings of the National Academy of Sciences USA* 102, 2310-2315.
- SAMOILOV, MS, PRICE, G & ARKIN, AP (2006), From Fluctuations to Phenotypes: The Physiology of Noise, *Science STKE* 366, re 17, 1-9, www.stke.org/cgi/content/full/2006/366/re17.
- SARKAR, S (1991a). Lamarck *contre* Darwin, Reduction *versus* Statistics: Conceptual Issues in the Controversy Over Directed Mutagenesis in Bacteria, *In* AI Tauber (ed.), *Organism and the Origins of Self*, Netherlands, Kluwer Academic Publishers.
- SARKAR, S (1991b), Haldanes's Solution of the Luria-Delbrück Distribution, *Genetics* 27, 257-261.
- SARKAR, S (2007), *Molecular Models of Life: Philosophical Papers On Molecular Biology*, Boston, The MIT Press.
- SCHAAPER, RM & DUNN, RL (1987), Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: The nature of *in vivo* DNA replication errors, *Proceedings of the National Academy of Sciences USA* 84, 6220-6224.
- SCHAAPER, RM & DUNN, RL (1991), Spontaneous Mutation in the *Escherichia coli lacI* Gene, *Genetics* 129, 317-326.
- SCHAFFER, J (2007), Deterministic Chance?, *British Journal for the Philosophy of Science* 58, 113-140.
- SCHNORR, CP (1971), A Unified Approach to the Definition of Random Sequence, *Mathematical Systems Theory* 5, 246-258.
- SCHNORR, CP (1973), Process Complexity and Effective Random Tests, *Journal of Computer and System Science* 7, 376-378.
- SCHNORR, CP (1977), A Survey of the Theory of Random Sequences, *In* B Hintikka, *Basic Problems in Methodology and Linguistics*, Dordrecht, 193-210.
- SCHRÖDINGER, E (1986[1944]), *Qu'est-ce que la vie ?*, Paris, Seuil.
- SHANAHAN, T (1991), Chance as an Explanatory Factor in Evolutionary Biology, *History and Philosophy of the Life Sciences* 13, 249-269.
- SHANAHAN, T (1992), Selection, Drift, and the Aims of Evolutionary Theory, *In* P Griffiths (ed.), *Trees of Life: Essays in Philosophy of Biology*, Dordrecht, Kluwer, 133-161.
- SHANAHAN, T (2004), *The Evolution of Darwinism*, Cambridge, Cambridge University Press.
- SHANNON, CE (1948), A Mathematical Theory of Communication, *The Bell System Technical Journal* 27, 379-423, 623-656.

- SHANNON, CE (1998), *Communication in the Presence of Noise*, Proceedings of the IEEE 86, 447-457.
- SHAPIRO, JA (1984), Observations on the formation of clones containing araB-lacZ cistron fusion, *Molecular Genetics and Genomics* 194, 79-90.
- SHAPIRO, JA (1995a), Adaptive Mutation: Who's Really in the Garden?, *Science* 268, 373-374.
- SHAPIRO, JA (1995b), 'Adaptative Mutation': The Debate Goes On, *Science* 269, 286-287.
- SHAPIRO, JA (1999), Genome System Architecture and Natural Genetic Engineering in Evolution, In LH Caporale (ed.), *Molecular Strategies in Biological Evolution, Annals of the New York Academy of Science* 870, 23-35.
- SHAPIRO, JA (2005), A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering, *Gene* 345, 91-100.
- SIMPSON, GG (1953), *The Major Features of Evolution*, New York, Columbia University Press.
- SIMPSON, GG (1984[1944]), *Tempo and Mode in Evolution*, New York, Columbia University Press.
- SIMPSON, ML, COX, CD & SAYLER, GS (2003), Frequency domain analysis of noise in autoregulated gene circuits, *Proceedings of the National Academy of Sciences USA* 100, 4551-4556.
- SKIPPER, RA (2002), The persistence of the R.A. Fisher-Sewall Wright controversy, *Biology and Philosophy* 17, 341-367.
- SMART, JJC (1959), Can biology be an exact science?, *Synthese* 11, 359-368.
- SNIEGOWSKI, PD & LENSKI, RE (1995), Mutation and Adaptation: The Directed Mutation Controversy in Evolutionary Perspective, *Annual Review of Ecology and Systematics* 26, 553-578.
- SNIEGOWSKI, PD, GERRISH, PJ & LENSKI, RE (1997). Evolution of high mutation rates in experimental populations of *E. coli*, *Nature* 387, 703-705.
- SNIEGOWSKI, PD, GERRISH, PJ, JOHNSON, T & SHAVER, A (2000), The evolution of mutation rates: separating causes from consequences, *Bioessays* 22, 1057-1066.
- SOBER, E (1984), *The Nature of Selection. Evolutionary Theory in Philosophical Focus*, Chicago, The University of Chicago Press.
- SOBER, E (2006) Parsimony, In S Sarkar & J Pfeifer (eds.), *The Philosophy of Science. An Encyclopedia* 1, New York, Routledge.

- SOLOMONOV, RJ (1964), A Formal Theory of Inductive Inference, *Information and Control* 7, 1-22.
- SOULE, ME (1982), Allomeric Variation. 1. The Theory and Some Consequences, *American Naturalist* 120, 751-764.
- SOULE, ME & BAKER, B (1968), Phenetics of natural populations. IV. The population asymmetry parameter in the butterfly *Coenonympha tullia*, *Heredity* 23, 611-614.
- SPOERING, AL, VULIC, M & LEWIS, K (2006), GlpD and PlsB participate in persister cell formation in *Escherichia coli*, *Journal of Bacteriology* 188, 5136-5144.
- STAHL, FW (1988), A unicorn in the garden, *Nature* 335, 112-113.
- STAHL, FW (1992), Unicorns Revisited, *Genetics* 132, 865-867.
- STAMOS, DN (2001), Quantum Indeterminism and Evolutionary Biology, *Philosophy of Science* 68, 164-184.
- STEARNS, SC & HOEKSTRA, R (2005), *Evolution: An Introduction*, Oxford, Oxford University Press.
- STEBBINS, GL (1971), *Processes of Organic Evolution*, Englewood Cliffs, New Jersey Prentice-Hall Inc.
- STEBBINS, GL (1982), *Darwin to DNA, Molecules to Humanity*, San Francisco, WH Freeman & Company.
- STEELE, EJ (1979), *Somatic Selection and Adaptive Evolution: On the Inheritance of Acquired Characters*, Toronto, Williams-Wallace.
- STELLING, J, SAUER, U, SZALLASI, Z, DOYLE, FJ III & DOYLE, J (2004), Robustness of Cellular Functions, *Cell* 118, 675-685.
- STERELNY, K & GRIFFITHS, PE (eds.), 1999, *Sex and Death. An Introduction to Philosophy of Biology*, Chicago, The University of Chicago Press.
- STERNBERG, RV (2002), On the Roles of DNA Repetitive Elements in the Context of a Unified Genomic-Epigenetic System, *Annals of the New York Academy of Science* 981, 154-188.
- STEWART, FM, GORDON, DM & LEVIN, BR (1990), Fluctuation Analysis: The Probability Distribution of the Number of Mutants Under Different Conditions, *Genetics* 124, 175-185.
- STEWART, FM (1994), Fluctuation Tests: How Reliable Are the Estimates of Mutation Rates?, *Genetics* 137, 1139-1146.
- STREVVENS, M (1998), Inferring Probabilities From Symmetries, *Noûs* 32, 231-246.

- STUBBE, H (1972[1963]), *History of Genetics. From Prehistoric Times to the Rediscovery of Mendel's Laws*, Boston, The MIT Press.
- SUPPES, P (1993), The Transcendental Character of Determinism, In PA French, TE Uehling & HK Wettstein (eds.), *Midwest Studies in Philosophy* 18, Notre Dame Indiana, University of Notre Dame Press, 242-257.
- SZATHMARY, E (2005), Path dependence and historical contingency in biology, In A Wimmer & R Kössler (eds.), *Understanding Change. Models, Methodologies and Metaphors*, New York, Palgrave Macmillan.
- TADDEI, F (1995), *Environnement et contrôles de la variabilité génétique chez Escherichia coli*, Thèse de doctorat, Université Paris 11 Orsay.
- TADDEI, F, VULIC, M, RADMAN, M & MATIC, I (1997), Genetic variability and adaptation to stress, In R Bijlsma & V Loeschcke (eds.), *Environmental Stress Adaptation and Evolution*, Basel-New York: Birkhäuser Verlag.
- TEBB, G & THODAY, JM (1954), Stability in development and relational balance of X chromosomes in *Drosophila melanogaster*, *Nature* 174, 1109-1110.
- TEMIN, HM & MIZUTANI, S (1970), RNA-dependent DNA polymerase in virions of Rous sarcoma virus, *Nature* 226, 1211-1213.
- TENAILLON, O, LE NAGARD, H, GODELLE, B & TADDEI, F (2000). Mutators and sex in bacteria: Conflict between adaptive strategies, *Proceedings of the National Academy of Science* 97, 19, 10465-10470.
- TENAILLON, O, TADDEI, F, RADMAN, M & MATIC, I (2001), Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation, *Research in Microbiology* 152, 11-16.
- TENAILLON, O, DENAMUR, E & MATIC, I (2004), Evolutionary significance of stress induced mutagenesis, *Trends in Microbiology* 12, 264-270.
- THATTAI, M & VAN OUDENAARDEN, A (2001), Intrinsic noise in gene regulatory networks, *Proceedings of the National Academy of Sciences USA* 98, 8614-8619.
- THATTAI, M & VAN OUDENAARDEN, A (2004), Stochastic Gene Expression in Fluctuating Environments, *Genetics* 167, 523-530.
- THODAY, JM (1956), Balance, heterozygosity, and developmental stability, *Cold Spring Harbor Symposia on Quantitative Biology* 21, 318-326.
- THODAY, JM (1958), Homeostasis in a selection experiment, *Heredity* 12, 401-415.
- THOMPSON, D'ARCY W (1917), *On Growth and Form*, Cambridge, Cambridge University Press.

- TIERCELIN, C (2003), Le projet peircien d'une métaphysique scientifique réaliste, *In* JP Cometti & C Tiercelin (eds.), *Cent ans de philosophie américaine, Actes du colloque de Cerisy 1995*, Publications de l'université de Pau, 157-182.
- TURNER, TE, SCHNELL, S & BURRAGE, K (2004), Stochastic approaches for modelling in vivo reactions, *Computational Biology and Chemistry* 28, 165-178.
- VAN KAMPEN, NG (1992), *Stochastic Processes in Physics and Chemistry*, North Holland Publishing.
- VAN LAMBALGEN, M (1987), Von Mises' Definition of Random Sequences Reconsidered, *The Journal of Symbolic Logic* 52, 725-755.
- VAZQUES-LASLOP, N, LEE, H & NEYFAKH, AA (2006), Increased persistence in *Escherichia coli* caused by controlled expression of toxins or other unrelated proteins, *Journal of Bacteriology* 188, 3494-3497.
- VILLE, J (1939), *Etude critique sur la notion de collectif*, Paris, Gauthier-Villars.
- VOLCHAN, SB (2002), What Is a Random Sequence?, *The American Mathematical Monthly* 109, 46-63.
- VON KRIES, J (1886), *Die Principien der Wahrscheinlichkeitsrechnung. Eine logische Untersuchung*, Freiburg, Mohr.
- VON MISES, R (1961[1928]), *Probability, Statistics and Truth*, London, George Allen & Unwin.
- VON PLATO, J (1982), Probability and Determinism, *Philosophy of Science*, 49, 51-66.
- VON PLATO, J (1983), The Method of Arbitrary Functions, *The British Journal for the Philosophy of Science* 34, 37-47.
- WADDINGTON, C (1957) *The Strategy of the Genes*, London, George Allen & Unwin.
- WANG, J, GONZALEZ, KD, SCARINGE, WA., TSAI, K, LIU, N, GU, D, LI, W, HILL, KA & SOMMER, SS (2007), Evidence for mutation showers, *Proceedings of the National Academy of Sciences USA* 104, 8403-8408.
- WATSON, JD et CRICK, FHC (1953), A Structure for Deoxyribose Nucleic Acid, *Nature* 171, 737-738.
- WEBER, M (2001), Determinism, Realism, and Probability in Evolutionary Theory, *Philosophy of Science* 68, S213-S224.
- WEISBERG, M (2006a), Robustness Analysis, *Philosophy of Science* 73, 730-742.
- WEISBERG, M (2006b), Forty Years of 'The Strategy'. Levins on Model Building and Idealization, *Biology and Philosophy* 21, 623-645.

- WEISBERG, M & REISMAN, K (2008), The Robust Volterra Principle, *Philosophy of Science* 75, 106-131.
- WEISMANN, A (1892[1883]), *Ueben Vererbung*, trad. fr., *De l'Hérédité*, réédité In C Lenay (dir.) *La découverte des lois de l'hérédité (1862-1900) : une anthologie*, Presse Pocket, 1990.
- WEISMANN, A (1893[1892]), *Das Keimplasma. Eine Theorie der Vererbung*, Jena, Gustav Fischer, traduction de l'allemand en anglais par WN Parker & H Rönnefeldt, *The Germ Plasm. A Theory of Heredity*, London, The Contemporary Science Series.
- WEST-EBERHARD, MJ (2003), *Developmental Plasticity and Evolution*, New York, Oxford University Press.
- WEYL, H (1952), *Symmetry*, Princeton, Princeton University Press.
- WILLEY, A (1894) *Amphioxus and the ancestry of the vertebrates*, New York, MacMillan.
- WIMSATT, WC (1980), Randomness and perceived-randomness in evolutionary biology, *Synthese* 43, 287-329.
- WIMSATT, WC (1981), Robustness, reliability, and overdetermination, In MB Brewer & BE Collins (eds.), *Scientific inquiry and the social sciences*, San Francisco, Jossey Bass.
- WIMSATT, WC & SCHANK, JC (1988), Two constraints on the evolution of complex adaptations and the means of their avoidance, In MH Nitecki (ed.), *Evolutionary Progress*, Chicago, University of Chicago Press.
- WIMSATT, WC (2000), Emergence as Non-Aggregativity and the Biases of Reductionisms, *Foundations of Science* 5, 269-297.
- WOESE, CR (1967), *The Genetic Code. The Molecular Basis for Genetic Expression*, New York, Harper and Row.
- WRIGHT, BE (2000), A Biochemical Mechanism for Nonrandom Mutations and Evolution, *Journal of Bacteriology* 182, 2993-3001.
- WRIGHT, BE, LONGACRE, A & REIMERS, JM (1999), Hypermutation in derepressed operons of *Escherichia coli* K12, *Proceedings of the National Academy of Science* 96, 5098-5094.
- WRIGHT, S (1922), The Effects of Inbreeding and Crossbreeding on Guinea Pigs. I. Decline of Vigor. II. Differentiation among Inbred Families. III. Crosses between highly Inbred Families, *U.S. Department of Agriculture, Bull.* 1090: 1-63, Bull. 1121: 1-49.
- WRIGHT, S (1930), The Genetical Theory of Natural Selection: A Review, *Journal of Heredity* 21, 349-356.
- WRIGHT, S (1931), Evolution in Mendelian Populations, *Genetics* 16, 97-159.

- WRIGHT, S (1948), On the Roles of Directed and Random Changes in Gene Frequency in the Genetics of Populations, *Evolution* 2, 279-294.
- WRIGHT, S (1952), The genetics of quantitative variability, In ECR Reeve & CH Waddington (eds.), *Quantitative inheritance*, London, Stationery Office, 5-41.
- YI, TM, HUANG, Y, SIMON, MI & DOYLE, J (2000), Robust perfect adaptation in bacterial chemotaxis through integral feedback control, *Proceedings of the National Academy of Sciences USA* 97, 4649-4653.
- ZIRKLE, C (1946), The early history of the idea of the inheritance of acquired characters and of pangenesis, *Transactions of the American Philosophical Society* 35, 91-151.

Table des figures

(p. 181) Figure 1 : Le test de fluctuation de Luria et Delbrück (1943). Distribution des mutants (représentés par des cellules pleines) dans 4 populations distinctes, fondées par 4 bactéries. a) Hypothèse de l'« immunité acquise et héréditaire » : la distribution du nombre de mutants suit une loi de Poisson ; b) Hypothèse de la « mutation » : la distribution du nombre de mutants est caractérisée par une variance beaucoup plus grande que celle prévue par la loi de Poisson. Figure tirée de Taddei, 1995, *Environnement et contrôle de la variabilité génétique chez Escherichia coli*, p. 23.

(p. 204) Figure 2 : Représentation schématique de la réponse SOS. Figure tirée de Taddei *et al*, 1997, « Genetic variability and adaptation to stress », p. 279.

(p. 207) Figure 3 : Les étapes principales du mécanisme moléculaire à l'origine de l'hypermutableté locale, au niveau de l'opéron *leu*, que Wright et ses collègues ont observée chez des populations d'*Escherichia coli* K12. Figure tirée de Wright, 2000, « A Biochemical Mechanism for Nonrandom Mutations and Evolution ».

(p. 225) Figure 4 : La distinction entre mutations « aléatoires » et mutations « non aléatoires » selon Jablonka et Lamb (2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*).

(p. 230) Figure 5 : Les mécanismes mutateurs et les biais de mutation.

*Mécanismes et biais de mutation « NON évolués », c'est-à-dire dus à des contraintes physico-chimiques ou à un état pathologique.

**Mécanismes et biais de mutation « Evolués (par SN) », c'est-à-dire qu'ils sont des sous-produits du processus évolutif ou des résultats adaptatifs de l'évolution par sélection naturelle.

(p. 234) Figure 6 : Notre analyse des mutations génétiques en comparaison avec la vision de Jablonka et Lamb (2005, *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*).

(p. 296) Figure 7 : Trois séries temporelles de la concentration de protéines au sein d'une cellule au cours du temps, l'une générée par simulation déterministe (courbe bleue), l'autre par simulation stochastique (courbe rouge). Les histogrammes montrent la probabilité qu'une cellule ait en son sein une certaine concentration intracellulaire de protéines. Figure tirée de Koern *et al*, 2005, « Stochasticity in gene expression: from theories to phenotypes ».

(p. 298) Figure 8 : La méthode « à deux rapporteurs ». Figure tirée de Maheshri & O'Shea, 2007, « Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression ».

(p. 298) Figure 9 : La méthode « à un rapporteur ». Figure tirée de Maheshri & O'Shea, 2007, *op. cit.*

(p. 300) Figure 10 : Représentation graphique des définitions opérationnelles et des caractéristiques du bruit intrinsèque (A) et du bruit extrinsèque (B). Dans la situation représentée ci-dessus, les deux rapporteurs identiques (GENE1 et GENE1) se trouvent dans la même cellule. Figure tirée de Raser & O'Shea, 2005, « Noise in Gene Expression: Origins, Consequences, and Control ».

(p. 326) Figure 11 : Dessin décrivant l'activation génétique, la transcription, la traduction, la dégradation des ARN messagers et des protéines. Figure tirée de Paulsson, 2005, « Models of stochastic gene expression ».

(p. 366) Figure 12 : Densité de probabilité. Figure tirée de Dress, 2004, *op. cit.*, p. 57.

(p. 369) Figure 13 : Fonction de répartition. Figure tirée de Dress, 2004, *op. cit.*, p. 154.

(p. 371) Figure 14 : Loi normale. Figure tirée de Cichitelli, 2001, *op. cit.*, p. 84.

(p. 373) Figure 15 : Loi uniforme discrète. Figure tirée de Cichitelli, *op. cit.*, p. 58.

(p. 373) Figure 16 : Loi uniforme continue. Figure tirée de Cichitelli, 2001, *op. cit.*, p. 82.

(p. 375) Figure 17 : Processus de Poisson. Figure tirée de Dress, 2004, *op. cit.*, p. 142.

Index des noms

Alberts, B 294,314
Allen, JA 136
Alphey, L 149
Andersson, DI 191,192
Aristote,15,43,57-61,63,64
Arkin, A 287,288,314
Astaurov, BL 284,286
Asteris, G 148
Atmanspacher, H 47
Ayala, FJ 171-172
Balaban, NQ 317
Balibar, E 47
Barberousse, A 47
Bar-Even, A 338
Bartholomay, AF 303
Bateson, W 127,158
Beardmore, JA 284
Beatty, J 16,17,19,66,155,156,158,260,
261,263,269
Becksei, A 312,313,332
Bennett, CH 97,98,105
Benson, SA 190
Benzer, S 165,252
Berg, OG 287
Bergson, H 60
Bernard, C 48,51,53,55
Bernoulli, J 81,111,369,371
Bertrand, J 82
Bishop, R 16,47,56
Blake, WJ 290,298,315,329
Bloch, O 43,48
Boe, L 190
Bohr, N 130
Borel, E 81
Bowler, PJ 25,134-137
Brading, K 340
Bradshaw, AD 284
Brailard, P-A 315
Brandon, R 16,40,71,143,260,261
Brown, R 281
Buettner, VL 255
Cairns, J 30,138,147,179,186-194
Calder, N 264
Calude, CS 97
Caporale, LH 220,239
Carnap, R 108,109
Carrier, M 16
Carson, S 40, 71,260
Castellani, E 340
Cavalli-Sforza, LL182
Chaitin, GJ 79, 90, 95-97, 103-106
Charpentier, E 66
Chou, WH 289
Church, A 79,91-96, 99, 102
Cichitelli, G 365,371,373
Cohen, JS 131
Cohen, L 281
Conche, M 44
Cook, DL 314
Cope, ED 135
Copeland, JB 94
Correns, C 19,158

Coulson, CA
 147,248,249,251,253,256,355
 Cournot, A-A 50,57,60-63,65, 68,107,116
 Crick, FHC 121,131,215
 Crow, JF 249,264
 Culp, S 356-358
 Dalmedico, AD 66
 Darden, L 189,190
 Darwin, C 10,13,14,16-18,20,24-27,29,
 30,40,54,58,66,122,123,128,129,132-
 134,137-139,147,152,155-
 160,163,164,174,175,181,183,189,214,
 244,251,262,359,362
 Daston, L 79,107
 David, P 259
 Davis, BD 190,237
 Dawkins, R 230,237
 Debré, P 292
 De Broglie, L 45,72
 Debs, TA 342
 de Finetti, B 108,110
 Delahaye, J-P 91,92,96-98,101,103-105
 Delbrück, M 30,130,138,147,148,179-
 189,193,248
 Delsol, M 17,63
 Desplan, C 288,310,313
 de Vries, H 19,127,128,158
 Dhar, N 316-318
 Dietrich, M 264
 Dijkmans, R 190
 Dobzhansky, T 17,26,29,160-164,166,169-
 174,251,260
 Drake, JW 146,253-255
 Dress, F 365,366,369,375
 Drouet, I 74
 Dubinin, NP 260
 Dunn, LC 127-129,134,241
 Dunn, RL 146,149,199
 Earman, J 16,44,46,47,49,55,56,89,92
 Eble, GJ 17,155,174
 Eimer, GHT 136
 Einstein, A 281
 Ekeland, I 66
 Eldredge, N 19
 Ellis, RE 107
 Elowitz, MB 297,315,329
 El Samad, H 304,331,349
 Elston, TC 314
 Faye, J 72
 Feller, W 79
 Fetzer, JH 108,113
 Fisher, RA
 26,158,162,164,166,251,262,263
 Fischer, EP 130
 Ford, EB 263
 Foster, PL 148,189-193,248
 Fox Keller, E 31,124,129-
 131,151,195,211,257
 Franceschelli, S 66
 Fraser, HB 310,316,332
 Fréchet, MR 81
 Freese, E 165,252
 Friedberg, EC 139,144,147,203
 Frigg, R 302
 Futuyma, DJ 24,26,139,143,147
 Galavotti, MC 92, 108
 Galton, F 134,158
 Gauvrit, N 88

Gayon, J 16,17,46-
 49,51,55,56,62,132,262,263
 Gerdes, K 318
 Giere, RN 343,352-354
 Gigerenzer, G 16, 158
 Gillespie, DT 303,304,306,307
 Gillespie, J 19,259
 Gillies, D 107,108,113,114
 Giraud, A 151,199
 Gleick, J 66
 Glennan, S 93,103
 Glickman, BW 199
 Goldschmidt, R 129
 Goldstein, S 72
 Goodwin, B 315
 Gould, SJ 16,17,19,66,155
 Graves, L 40
 Gray, RD 23,318
 Griffith, F 289
 Griffiths, PE 23,131,257,318
 Hacking, I 45,79,84,107,108,356
 Hadamard, J 66
 Hagedoorn, AL & AC 262
 Haig, D 217,218,222,223
 Hajek, A 46,107,108,112
 Haldane, JSB 26,164,189,251
 Hall, BG 190,191
 Hall, BK 27,143,284,285,287,288,294
 Hallgrimsson, B
 27,143,284,285,287,288,294
 Hammock, EAD 289
 Hampel, FR 356
 Hänggi, P 312
 Harris, H 264
 Hartmann, S 302
 Hartwell, LH 314
 Hasty, J 312,313
 Heisenberg, W 44,72
 Hendrickson, H 192
 Henslow, G 135
 Hilgevoord, J 72
 Hill, KA 255
 Hodge, MJS 260
 Hoefler, C 16,47
 Hoekstra, R 139
 Huber, PJ 356
 Hull, D 16
 Hume, D 43
 Humphreys, P 99
 Huxley, J 17,26,160,161,164,251
 Hyatt, A 135
 Isaacs, FJ 313
 Ismael, J 71
 Jablonka, E 23,31,123,125,151-
 153,157,182,190,196,197,205,208,209,
 211-229,231,234-239,360
 Jacob, F 27,131,205
 Jaeger, G 147
 Jeffreys, H 109
 Jennings, HS 135
 Johnson, T 200
 Jones, ME 148
 Jordanova, LJ 134,135
 Judson, HF 131
 Jukes, TH 264
 Kærn, M 296,299-301,310,311
 Kammerer, P 135
 Kauffman, S 315

Kay, LE 131
 Kellis, M 314
 Kepler, TB 314
 Keren, I 318
 Keynes, JM 81,108-110
 Kierzek, AM 314
 Kimura, M 146,197,249,260,264
 King, JL 264
 Kistler, M 43,49
 Klingenberg, CP 285,287
 Klug, W 139
 Kojève, A 56
 Kolmogorov, AN 79,89,90,95-
 98,102,104,106,107
 Kramers, HA 291
 Krüger, L 16
 Küppers, B-O 104
 Kussel, E 310,316,317,318
 Laland, KN 23
 Lalande, A 61
 Lamarck, J-B 132 -
 136,147,157,162,175,176,181,183,189,
 218,220,221,225,237,244,360
 Lamb, M 23,31,123,125,151-
 153,157,182,190,196,197,205,208,209,
 211-229,231,234-239
 Lamotte, M 19,261
 Lankester, ER 134
 Laplace, P-S
 48,49,51,55,56,61,62,81,82,86,91,108,1
 09,115
 Laplacette, J de la 61
 Latter, BDH 285
 Lea, DE 147,248,249,251,253,256,355
 L'Ecuyer, P 101
 Lederberg, J 182
 Lederberg, E 182
 Leibler, S 316,317,319
 Leibniz, GW 43,44,48,51,55,97
 Leisner, M 313
 Lenay, C 17,25
 Lennox, J 155
 Lenski, RE 176,179,182,189,190-194
 Lestienne, R 44
 Levin, LA 79, 96, 103
 Levins, R 250,327,340,341,353,354
 Lewin, B 139,294,314
 Lewis, D 16
 Lewis, K 318
 Lewontin, R 17,23,121,133,264,286
 Li, M 96
 Lim, MM 289
 Lipson, C 130
 Lloyd, S 104
 Loewer, B 16
 Lopez-Maury, L 311,316,319
 Losick, R 288,310,313
 Lucrèce, 43,44
 Ludwig, W 284
 Luria, SE 30,138,147,148,179-
 187,189,193,248
 Maamar, H 290, 313
 Machamer, P 183,293
 Macherey, P 47
 Maheshri, N
 287,288,294,297,298,310,311,313,329,3
 32,333
 Malaterre, C 40

Markow, TA 284,285
 Martin, T 62-63
 Martin-Löf, P 79,92,100-103
 Mason, LG 285
 Mather, K 284
 Maupertius, P-LM de 127
 Maynard-Smith, J 315
 Mayr, E 17,23,24,26,29,76,129,161-
 164,166,251,261,263
 McAdams, HH 287,314
 McKinney, JD 316-318
 Mellor, DH 108
 Mendel, G 17,19,127,137
 Méré, chevalier de 79
 Merlin, F 23,40,234,318,344
 Mettetal, JT 317,338
 Mikeladze-Dvali, T 289
 Mill, JS 61
 Miller, D 108,113
 Miller, S 186, 191
 Millstein, RL 17,46,174-
 176,190,225,244,260,261,264
 Mittler, JE 176,190
 Mizutani, S 188
 Møller, AP 285
 Mombaerts, P 289
 Monod, J
 17,60,61,64,65,105,106,131,205,292
 Montague, R 47
 Moore, GE 109
 Morange, M 121,130,131,289,291
 Morgan, TH 18,128,129,147,241
 Moxon, RE 201
 Muller, H 129,130
 Nagel, E 46,76
 Nathans, J 289
 Newcombe, H 182
 Newman, JRS 316,338
 Ninio, J 253,254,256
 Novick, A 182
 Nozick, R 342,358
 Ohta, T 264
 Olby, R 131
 Oller, AR 146,149
 Oppenheim, AB 317
 Orrel, AR 312
 Ortolli, S 72
 Osborn, HF 135
 O'Shea, E 32,287,288,290,294,296-
 300,309-311,313-315,329,332,333
 Overbaugh, J 186
 Oyama, S 16,23
 Ozbudak, EM 297,315,317
 Packard, AS 134,136
 Pagels, H 104
 Palmer, AR 285,288
 Pascal, B 79,370
 Pasteur, L 220,239
 Paulsson, J 31,287,326,327
 Pearson, K 284,286
 Peirce, CS 45
 Perrin, J 281
 Perutz, M 130
 Pharabod, J-P 72
 Poincaré, H 46,47,57,66-69,81,111
 Polak, M 283-286
 Pomian, K 16,47,56

Popper, K 16,47,50-
 53,56,73,74,108,112,113
 Portugal, FH 131
 Provine, WB 263
 Psillos, S 43
 Ptashne, M 290
 Radman, M 144,199,202,204,205
 Raidl, E 87
 Raj, A 332
 Ramsey, FP 108,110
 Rao, CV 311,313,320
 Raser, JM 32,290,296,297,299,300,309-
 311,314,315,329,333
 Redhead, MLG 342
 Reeve, ECR 284,285,287
 Reichenbach, H 90,108,111,112
 Reisman, K 342,343,346
 Ridley, M 26,139,147,151,177
 Rigney, DR 287
 Roberts, JT 47
 Robertson, FW 284,285,287
 Romanes, GJ 134
 Romaschoff, DD 260
 Rosche, WA 148,192,248
 Rosenberg, A 16,76,260
 Roth, JR 191,192
 Ruelle, D 66
 Ruffié, J 264
 Russel, B 109
 Ryan, FJ 193
 Saint-Sernin, B 46,62
 Salmon, WC 43,44,47,91,92,94
 Samadi, S 259
 Samoilov, MS 282,287,293,310,313,338
 Sarkar, S 76,147,148,175,181-
 184,189,190,244,
 Schaaper, RM 146,149,199
 Schaffer, J 16
 Schank, JC 315
 Schieve, WC 287
 Schnorr, CP 79,91,92,101,103
 Schottky, W 281
 Schrödinger, E 73,130
 Scott, WB 135
 Shanahan, T 17,134,260
 Shannon, CE 96,281,282
 Shapiro, JA
 31,124,151,186,187,194,195,206,211
 Simard, R 101
 Simpson, GG
 16,17,26,29,155,160,162,164,168,172-
 174,251
 Simpson, ML 312
 Skipper, RA 261,263
 Smart, JJC 16
 Sniegowski, PD
 179,182,190,191,193,194,199
 Sober, E 17,21,71,149,260
 Solomonov, R 79,90,95-97,106
 Soulé, ME 284,285
 Spoering, AL 318
 Stahl, FW 190
 Stamos, DN 71,75
 Stearns, SC 139
 Stebbins, GL 29,160,164,171,172,251
 Steele, EJ 137
 Stelling, J 311,313
 Sterelny, K 131

Sternberg, RV 31,124,151,196,211
 Stewart, FM 147,148,189
 Stockard, CR 136
 Stotz, K 257
 Strevens, M 81-82
 Stubbe, H 127
 Suppes, P 47,76
 Swaddle, JP 285
 Szathmary, E 16
 Szilard, L 182
 Taddei, F 181,182,193,194,199,202-204
 Tebb, G 284
 Temin, HM 188
 Tenaillon, O 200,202,205
 Thattai, M 310,312,314,318,319
 Thoday, JM 284,285
 Thompson, D'Arcy W 315
 Tiercelin, C 45
 Tschermak-Seysenegg, von E 19
 Turing, A 92-101,104,269,274
 Turner, TE 291,296,303,304,331
 Uffink, J 72
 Valentine, JW 171,172
 Van Kampen, NG 304,306,307
 Van Lambalgen, M 92
 Van Oudenaarden, A 310,312,314,318,338
 Vazques-Laslop, N 318
 Ville, J 91,102
 Vitanyi, P 96
 Volchan, SB 92
 Von Kries, NG 81
 Von Mises, R 79,89-
 92,94,95,102,103,108,111,112
 Von Plato, J 16,81
 Von Wartburg, W 43,48
 Waagen, W 127
 Waddington, C 284
 Wagner, R 144,199
 Wang, J 255
 Watson, JD 121,131
 Weber, M 17
 Weinert, TA 314
 Weisberg, M 250,340-343,346
 Weismann, A 17,25,134,217,218,222
 West-Eberhard, MJ 23,283,286
 Weyl, H 340,342
 Wilkins, MHF 130
 Willey, A 284
 Wimsatt, WC 17,76,315,340,341,344
 Wittgenstein, L 109
 Woese, CR 131
 Wright, BE 31,195,205-
 209,211,219,227,228,230-232,238
 Wright, S 161,162,164,251,262,263,285
 Yi, TM 314
 Young, L 289
 Zenil, H 97
 Zirkle, C 132

Le hasard et les sources de la variation biologique : analyse critique d'une notion multiple

Résumé : La notion de hasard en biologie fait l'objet de débats philosophiques et scientifiques depuis la deuxième moitié du XIX^e siècle, notamment depuis la publication de *L'Origine des Espèces* de Darwin (1859). Dans l'état actuel de la recherche, elle est encore un objet de controverse : le fait que cette notion puisse prendre des significations et des rôles multiples la rend difficile à cerner, même dans un contexte très spécifique. Notre thèse consiste en une analyse épistémologique de la notion de hasard lorsqu'elle est utilisée par les biologistes dans la caractérisation des phénomènes à l'origine de la variation au sein des populations naturelles. Plus exactement, nous abordons la question de savoir quelle notion de hasard est conceptuellement et empiriquement appropriée en ce qui concerne deux sources de variation biologique : les mutations génétiques et le bruit dans l'expression des gènes. Nous apportons une clarification conceptuelle de cette notion, selon une perspective évolutionnaire et d'un point de vue moléculaire, sur la base des avancées récentes au sujet de ces deux causes de la variation. En vertu de la relation apparemment privilégiée entre la notion de hasard et les probabilités, nous traitons aussi la question de l'interprétation des probabilités dans les descriptions formelles de ces phénomènes biologiques. L'objectif principal de ce travail est de fournir un cadre conceptuel précis à l'utilisation de la notion de hasard en biologie.

Chance and the sources of biological variation: a critical analysis of a multiple notion

Summary: The notion of chance in biology is the subject of philosophical and scientific debates since the second half of the XIX century, in particular since the publication of Darwin's *On The Origin of Species* (1859). In the present state of research, it is still an object of controversy: the fact that this notion can take multiple meanings and roles makes it difficult to define, even in a very specific context. My thesis is an epistemological analysis of the notion of chance as it is used by biologists to characterize the phenomena at the origin of variation in natural populations. More exactly, I treat the question of knowing which notion of chance is conceptually and empirically appropriate in the study of two sources of variation: genetic mutations and noise in gene expression. On the basis of the recent studies on these two causes of variation, I provide a conceptual clarification of this notion, both from an evolutionary perspective and from a molecular point of view. Given the apparently special relationship between the notion of chance and the notion of probability, I am also interested in the question of the interpretation of probability in the formal descriptions of these biological phenomena. The main aim of this work is to provide a precise conceptual framework for the use of the notion of chance in biology.

Discipline: Philosophie

Mots-clés: Philosophie des sciences ; Philosophie de la biologie ; Épistémologie ; Hasard ; Interprétation des probabilités ; Variation biologique ; Mutations génétiques ; Bruit dans l'expression des gènes.

Équipe d'accueil : Institut d'Histoire et de Philosophie des Sciences et des Techniques (UMR 8590), 13 rue du Four 75006, Paris.