

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

SIMULER LA SCIENCE : UN MODÈLE MULTI-AGENTS  
DE L'ÉVOLUTION DES IDÉES SCIENTIFIQUES

THÈSE  
PRÉSENTÉE  
COMME EXIGENCE PARTIELLE  
DU DOCTORAT EN PHILOSOPHIE

PAR  
NICOLAS PAYETTE

JUIN 2018

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.10-2015). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

On dit que c'est le voyage qui compte, plus que la destination. La métaphore est usée, mais je le suis aussi, alors je vais m'en prévaloir. J'adhère aussi à l'idée que ce sont les détours qui comptent le plus, ce qui, vu le nombre et la durée des miens, n'est pas rien. Mais avant d'en venir aux détours, je dois remercier les organismes et les personnes qui se sont assurés que je parviens à destination. Une mise en garde, toutefois : si je nommais tout ceux et celles qui m'ont aidé à un moment ou à un autre, ces remerciements seraient plus longs que la thèse elle-même. Si vous lisez ceci, dites-vous qu'il y a de bonnes chances que je vous sois reconnaissant même si vous n'y trouvez pas votre nom.

D'abord, les organismes. Tout ce cheminement aurait été beaucoup plus difficile sans le soutien du Fonds à l'accessibilité et à la réussite de l'UQAM (FARE), du Fonds québécois de la recherche sur la société et la culture (FQRSC) et du Conseil de recherches en sciences humaines du Canada (CRSH).

Ensuite, les personnes. Pierre Poirier vient au premier rang d'entre elles. En tant que directeur de cette thèse, il a droit au crédit pour ses meilleurs éléments (les erreurs et les confusions ne relevant, bien sûr, que de moi). Sa contribution à mon cheminement intellectuel va toutefois bien plus loin. Depuis le début de mes études universitaires, Pierre a été avec moi d'une générosité exceptionnelle. J'espère seulement, en contribuant à sa valeur sélective conceptuelle inclusive, pouvoir lui rendre un iota de ce qu'il m'a apporté.

Frédéric Bouchard, Daniel Memmi, Serge Robert et Roger Villemare, les autres professeurs ayant siégé aux comités d'évaluation de mon doctorat, ont aussi joué un rôle im-

portant. Ils m'ont parfois empêché de prendre les chemins les plus courts, me forçant à en explorer d'autres qui, bien que plus longs et difficiles, ont mené à une meilleure thèse.

Tous ceux et celles qui m'ont enseigné pendant mon passage à l'UQAM ont droit à ma reconnaissance, mais j'aimerais remercier plus particulièrement les codirecteurs du laboratoire LANCI : Pierre Poirier et Serge Robert (déjà mentionnés), ainsi que Luc Faucher et Jean-Guy Meunier.

J'ai côtoyé au LANCI plusieurs personnes extraordinaires, mais je ne peux passer sous silence la complicité établie avec Maxime Sainte-Marie et Jean-François Chartier. En plus de travailler avec moi sur plusieurs projets de recherche intéressants, ils ont été de précieux compagnons de route. (Ils ont aussi fait de moi un meilleur joueur de billard.)

Dans la catégorie « complices de l'UQAM », Roxanne Campeau, David Chabot, Valérie Couillard et Mélissa Thériault ont aussi droit de cité pour différentes raisons.

Un autre aspect marquant de mes années à l'UQAM a été l'organisation récurrente du colloque *Cognitio* (<http://cognitio.uqam.ca>). Trop de gens ont collaboré avec moi pour tous les nommer, mais je fais une exception pour Benoit Hardy-Vallée, fondateur de l'événement et partenaire des premières années. Merci Benoit, et longue vie à *Cognitio* !

Vous me permettrez maintenant de passer brièvement à l'anglais, puisque je souhaite dire quelques mots sur mes détours hors-Québec.

*My first real foray in the wider academic world was at George Mason University's Center for Social Complexity. I'd like to thank professors Claudio Cioffi-Revilla and Sean Luke for hosting me there, and also for creating the MASON agent-based modelling toolkit, which I used to develop early versions of my model. I can't mention GMU without also thanking Joseph Harrison, who showed me the ropes over there, and Jennifer Durdall, from whom I rented a room during my stay in Fairfax. They became two of my dearest friends in the*

*world. (Merci aussi au FQRSC d'avoir subventionné ce stage et à Martine Foisy, du CIRST, pour toute l'aide qu'elle m'a apportée dans la coordination de ce projet.)*

*Another special thank you goes to Katy Börner et Andrea Scharnhorst, who recruited me to write the "Agent-based models of science" chapter for the Models of science dynamics book they edited with Peter van den Besselaar (Scharnhorst et al., 2012). That work eventually turned into big chunks of chapters 1 and 2 of this thesis, but it was also, in and of itself, an enlightening experience that opened a lot of doors for me. (It is also, to this day, my most cited publication.)*

*Another milestone event for me was the Computer Simulations in the Humanities NEH summer institute, organised by Anthony Beavers, Marvin Croy, Patrick Grim, and Mirsad Hadzikadic, and held at the University of North Carolina in Charlotte. These guys somehow recruited me as a "mentor" for that institute. I ended up learning a tremendous amount from them, from fellow mentors Jason Alexander, Aaron Bramson, Daniel Singer and Chuck Turnista, and from the students at the institute.*

*The next step I took had an impact on my career which could not be overstated : I went to Northwestern University to work in Uri Wilensky's Center for Connected Learning and Computer-Based Modeling, as part of NetLogo's core development team. I am deeply grateful to Uri for giving me a chance to work on NetLogo, and also for being responsible for NetLogo's existence (I do believe NetLogo makes the world a better place.) I had the privilege to interact with many wonderful people at the CCL, but I want to explicitly mention Jason Bertsche, Frank Duncan, Robert Grider, Bryan Head, Arthur Hjorth and Seth Tisue. These guys are among the best programmers I have ever worked with and I think it's fair to say that I learned more about software development from that bunch than from any other. I must add a special word on top of that for Jason, who took me under his roof when I first moved to Chicago and many other times later when I was occasionally coming to visit the CCL. Holger Niemann and Dana Mills, flatmates and friends, also greatly contributed to making my time in Chicago enjoyable.*

*Another great person to whom I am indebted, intellectually and personally, is Nigel Gilbert. Nigel is, in many ways, the founder of the field to which I am trying to contribute with this thesis. His "Simulation of the Structure of Academic Science" (1997) paved the*

*way for ABMs of science. Little did I know when I started on this project that I would also end up working in Nigel's Centre for Research in Social Simulation at the University of Surrey, getting a chance to collaborate with other amazing researchers such as (amongst others) Jennifer Badham, Peter Barbrook-Johnson and Corinna Elsenbroich.*

Plus près de la maison, je souhaite ajouter un petit mot pour le Atomic Café (3606 rue Ontario Est), où de grands pans de cette thèse ont été écrits. C'est important pour la santé mentale d'avoir des endroits publics où il fait bon travailler et le Atomic a joué ce rôle à merveille pour moi.

Dans la catégorie des lieux importants, la palme revient toutefois au Brown Derby : notre petit chalet au bord du lac Diez-d'Aux qui m'a servi de refuge pendant de longues périodes. Je suis reconnaissant à Alain, Karine et Ernst d'avoir accepté que je l'occupe ainsi, mais aussi et surtout pour leur amour, leur soutien, et tous les bons moments que nous avons passé ensemble (et avec Camélie, Siméon, Justin et Maxence !) dans cet endroit et ailleurs.

Un dernier lieu à mentionner : la petite maison de mon pote Fred et ses parents Françoise et Jean-Claude à St-Michel-Chef-Chef, où je suis allé écrire une grande partie du chapitre 4. Merci de me l'avoir prêtée !

Je ne pourrai pas mentionner tous les amis qui ont été importants pour moi pendant cette période, mais parmi ceux et celles dont le nom n'apparaît pas déjà ci-haut, il y a Gab, Jenny, Simon, Corinne, Christian et Hugues que je ne saurais passer sous silence. Chacun à votre façon, vous avez été là pour moi quand ça comptait. Merci !

Je termine maintenant avec quelques petits mots plus personnels.

Claudine, je ne sais pas si je me serais lancé dans cette aventure si ce n'avait été de ton influence. J'ai vécu auprès de toi un éveil intellectuel qui a changé la façon dont je vois le monde. Je t'en serai toujours reconnaissant.

*Paula, thank you for holding my hand during those last few miles, when I was out breath and my legs were tired. You saw me through the end of an adventure and the beginning of another. I hope we'll have many more together.*

Anne-Marie et Kateri, mes sœurs adorées, Alex, mon beau-frère préféré (quoique Marc te fait maintenant compétition...), je ne saurais dire assez à quel point c'est précieux de vous avoir. Il y a un petit clin d'œil pour vous à la page 113.

Papa, Maman, comment vous remercier assez pour tout ce que vous avez fait pour moi ? Le doctorat a été un long chemin parsemé d'embûches, mais j'ai toujours senti votre soutien au fil de celui-ci, toujours pu aller me réfugier auprès de vous pour me ressourcer, me reposer, me recadrer. Mais au-delà de ça, je vous dois aussi le goût de lire, d'apprendre, et de découvrir le monde. Il y a dans cette thèse beaucoup de vous deux. Je vous aime fort.

## DÉDICACE

*À Camélie, Siméon, Justin, Maxence,  
Jules, Adèle et Medan.*

## TABLE DES MATIÈRES

REMERCIEMENTS . . . . .	ii
DÉDICACE. . . . .	vii
LISTE DES FIGURES . . . . .	xii
LISTE DES TABLEAUX. . . . .	xv
LISTE DES EXTRAITS DE CODE SOURCE . . . . .	xvi
RÉSUMÉ. . . . .	xvii
INTRODUCTION . . . . .	1
CHAPITRE I	
INTRODUCTION AUX MODÈLES MULTI-AGENTS . . . . .	15
1.1 Un peu d'histoire . . . . .	15
1.2 Qu'est-ce qu'un agent ? . . . . .	21
1.3 Caractéristiques des MMA . . . . .	24
1.4 La question des niveaux : micro et macro . . . . .	26
CHAPITRE II	
LES MODÈLES MULTI-AGENTS DE LA SCIENCE . . . . .	30
2.1 Le modèle original de Gilbert : articles et kènes. . . . .	30
2.2 Un point de vue inspiré des sciences cognitives. . . . .	35
2.3 Un modèle de la distribution des ressources cognitives . . . . .	37
2.4 La science comme système cognitif distribué. . . . .	41
2.5 La science dans un paysage épistémique. . . . .	43

2.6	Les réseaux épistémiques. . . . .	48
2.7	Conclusion . . . . .	50
CHAPITRE III		
	LE PROCESSUS SCIENTIFIQUE SELON DAVID HULL . . . . .	56
3.1	Contexte philosophique . . . . .	57
	3.1.1 Internalisme et externalisme . . . . .	58
	3.1.2 Épistémologie évolutionniste . . . . .	60
3.2	Une approche évolutionniste de la science . . . . .	63
	3.2.1 Une analyse générale des processus de sélection . . . . .	67
	3.2.2 La science comme processus de sélection . . . . .	69
	3.2.3 Objections à la science comme processus de sélection . . . . .	75
3.3	Les mécanismes sociaux à l'œuvre en science . . . . .	85
	3.3.1 Crédit vs support . . . . .	85
	3.3.2 Valeur sélective conceptuelle inclusive . . . . .	87
	3.3.3 Structure démique de la science. . . . .	89
3.4	Conclusion . . . . .	90
CHAPITRE IV		
	THÉORIES, MODÈLES ET SIMULATIONS . . . . .	92
4.1	L'approche syntaxique des théories . . . . .	93
4.2	L'approche sémantique des théories . . . . .	97
	4.2.1 Les modèles comme structures ensemblistes. . . . .	98
	4.2.2 Les modèles comme trajectoires dans un espace d'états . . . . .	102
	4.2.3 Les modèles comme structures computationnelles . . . . .	105
4.3	Qu'est-ce qui constitue un modèle de la théorie de Hull ? . . . . .	109
4.4	Relation entre modèle et monde . . . . .	113
4.5	Les modèles comme explications. . . . .	120
CHAPITRE V		
	UN MODÈLE DU PROCESSUS SCIENTIFIQUE . . . . .	125
5.1	Description des entités du modèle . . . . .	126

5.2	La notion de paysage épistémique . . . . .	.131
5.2.1	Les paysages épistémiques utilisés dans le modèle . . . . .	.136
5.3	Génération de la population initiale du modèle . . . . .	.139
5.4	Algorithme général de la simulation . . . . .	.143
5.4.1	Idéation . . . . .	.144
5.4.2	Tests . . . . .	.148
5.4.3	Écriture . . . . .	.151
5.4.4	Publication . . . . .	.156
5.4.5	Lecture . . . . .	.158
5.4.6	Retraite . . . . .	.162
5.4.7	Éducation . . . . .	.163
5.5	Conclusion . . . . .	.167
CHAPITRE VI		
RÉSULTATS DE SIMULATION ET ANALYSE . . . . .		
6.1	Expériences . . . . .	.172
6.2	Performance épistémique générale de la communauté . . . . .	.177
6.3	Distributions. . . . .	.183
6.4	Performance sociale en fonction des caractéristiques individuelles . . . . .	.190
6.4.1	Influence de $\alpha$ sur le crédit des scientifiques. . . . .	.190
6.4.2	Influence de $\beta$ sur le crédit des scientifiques. . . . .	.194
6.4.3	Influence de $\gamma$ sur le crédit des scientifiques. . . . .	.201
6.4.4	Influence de $\delta$ sur le crédit des scientifiques. . . . .	.204
6.5	Performance épistémique en fonction des caractéristiques individuelles . . . . .	.206
6.5.1	Influence de $\alpha$ sur la valeur objective des articles. . . . .	.209
6.5.2	Influence de $\beta$ sur la valeur objective des articles. . . . .	.212
6.5.3	Influence de $\gamma$ sur la valeur objective des articles . . . . .	.215
6.5.4	Influence de $\delta$ sur la valeur objective des articles . . . . .	.218
6.5.5	Relation entre crédit et valeur objective . . . . .	.222
6.6	Conclusion . . . . .	.224

CONCLUSION . . . . .	.229
APPENDICE A TABLEAUX DE CORRÉLATIONS. . . . .	.240
BIBLIOGRAPHIE . . . . .	.243

## LISTE DES FIGURES

1.1	Le « constructeur universel » de Von Neumann. . . . .	16
1.2	Exemple de patron complexe dans le jeu de la vie de Conway . . . . .	17
1.3	Modèle de Schelling. . . . .	19
1.4	Une image du <i>Sugarscape</i> . . . . .	20
2.1	Représentation de l'espace dans la simulation de Gilbert . . . . .	33
2.2	Relations numériques dans la simulation de Gilbert . . . . .	34
2.3	Architecture CLARION . . . . .	36
2.4	Publications par auteurs dans <i>Econometrica</i> et <i>Chemical Abstracts</i> . . . . .	37
2.5	Modèle de Edmonds . . . . .	43
2.6	Le paysage épistémique utilisé par Weisberg et Muldoon . . . . .	45
2.7	Voisinage de Moore . . . . .	47
2.8	Position des « Suiveurs » dans le paysage épistémique . . . . .	47
2.9	Progrès épistémique dans les simulations de Weisberg et Muldoon . . . . .	48
2.10	Paysages épistémiques utilisés dans les simulations de Grim . . . . .	50
2.11	Performance selon la structure sociale dans les simulations de Grim . . . . .	51
3.1	<i>The Battle of Ten Naked Men</i> , Pollaiuolo (1465) . . . . .	64
3.2	Héritité pluriparentale vs. biparentale des idées. . . . .	80
3.3	La classification qui, selon Hull, sous-tend l'objection de l'intentionnalité. . . . .	84
3.4	La classification des processus de sélection selon Hull. . . . .	85
5.1	Les entités qui composent le modèle . . . . .	126

5.2	Paysages épistémiques. . . . .	.137
5.3	Paysages épistémiques avec populations initiales . . . . .	.141
5.4	Paysages épistémiques subjectifs. . . . .	.149
5.6	Effet de l'ancienneté sur la probabilité de retraite . . . . .	.164
5.7	Effet de l'ancienneté sur la probabilité de retraite, ajusté par le crédit normalisé . . . . .	.164
6.1	Évolution de la valeur objective moyenne des articles . . . . .	.178
6.2	Distribution spatiale finale des articles sur les différents paysages épistémiques pour des simulations individuelles.. . . .	.179
6.3	Répartition des articles par auteurs. . . . .	.184
6.4	Répartition des citations par articles . . . . .	.187
6.5	Répartition des citations par auteurs . . . . .	.188
6.6	Distribution du crédit par scientifiques . . . . .	.188
6.7	Crédit en fonction de $\alpha$ . . . . .	.192
6.8	Projection de la valeur d'une caractéristique comme compromis . . . . .	.193
6.9	Relation de compromis entre caractéristiques individuelles et crédit ( $E_1$ ). . . . .	.194
6.10	Crédit en fonction de $\beta$ . . . . .	.196
6.11	Relation linéaire entre caractéristiques individuelles et crédit ( $E_1$ ) . . . . .	.197
6.12	Crédit en fonction de $\gamma$ . . . . .	.203
6.13	Crédit en fonction de $\delta$ . . . . .	.205
6.14	Relation de compromis entre caractéristiques individuelle et valeur objective ( $E_2$ ). . . . .	.207
6.15	Relation linéaire entre caractéristiques individuelle et valeur objective ( $E_2$ ). . . . .	.208
6.16	Valeur objective des articles en fonction de $\alpha$ . . . . .	.210
6.17	Valeur objective en fonction du crédit pour $\alpha$ . . . . .	.211
6.18	Valeur objective des articles en fonction de $\beta$ . . . . .	.213
6.19	Valeur objective en fonction du crédit pour $\beta$ . . . . .	.214
6.20	Valeur objective des articles en fonction de $\gamma$ . . . . .	.216
6.21	Valeur objective en fonction du crédit pour $\gamma$ . . . . .	.217

6.22	Valeur objective des articles en fonction de $\delta$ . . . . .	.219
6.23	Valeur objective en fonction du crédit pour $\delta$ . . . . .	.220
6.24	Distribution spatiale finale des articles sur EXP pour différentes valeurs de $\delta$ . . . . .	.221
6.25	Relation linéaire entre crédit et valeur objective ( $E_1$ vs $E_2$ ) . . . . .	.222

## LISTE DES TABLEAUX

4.1	Structure des théories selon l'approche syntaxique . . . . .	94
5.1	Comparaison du modèle et de la théorie de Hull . . . . .	168
6.1	Paramètres constants utilisés dans les simulations. . . . .	173
A.1	Relation de compromis entre caractéristiques individuelles et crédit ( $E_1$ ). .	241
A.2	Relation linéaire entre caractéristiques individuelles et crédit ( $E_1$ ) . . .	241
A.3	Relation de compromis entre caractéristiques individuelle et valeur objective ( $E_2$ ). . . . .	241
A.4	Relation linéaire entre caractéristiques individuelle et valeur objective ( $E_2$ ).241	
A.5	Relation linéaire entre crédit et valeur objective ( $E_1$ vs $E_2$ ) . . . . .	242

## LISTE DES EXTRAITS DE CODE SOURCE

4.1	Le jeu de la vie en NetLogo . . . . .	.106
5.1	Déclaration des espèces ideas, drafts et papers. . . . .	.128
5.2	Déclaration de l'espèce scientists. . . . .	.129
5.3	Déclaration de l'espèce retired-scientists. . . . .	.130
5.4	Déclaration des espèces de liens. . . . .	.130
5.5	Déclaration des variables de cellules. . . . .	.131
5.6	Création de la population initiale de scientifiques. . . . .	.140
5.7	Procédure go. . . . .	.143
5.8	Génération des nouvelles idées. . . . .	.145
5.9	Fonction relative-subjective-value. . . . .	.147
5.10	Sélection des idées à tester et révision des évaluations correspondantes. . . . .	.151
5.11	Procédure principale de rédaction des manuscrits. . . . .	.151
5.12	Sélection des idées pour l'écriture. . . . .	.152
5.13	Implémentation NetLogo de l'équation 5.2. . . . .	.156
5.14	Calcul du pointage par les relecteurs d'un article. . . . .	.157
5.15	Fonction papers-to-read. . . . .	.159
5.16	Procédure read. . . . .	.159
5.17	Procédure adjust-evaluation-for-credit. . . . .	.161
5.18	Procédure attribute-credit-for. . . . .	.162
5.19	Choix des scientifiques qui prendront leur retraite. . . . .	.165
5.20	Éducation des nouveaux scientifiques. . . . .	.166

## RÉSUMÉ

Cette thèse se penche sur les mécanismes épistémiques et sociaux par lesquels s'opère le changement scientifique. Ces questions sont abordées sous l'angle de l'épistémologie évolutionniste, selon laquelle le processus d'acquisition des connaissances et celui de l'évolution biologique relèvent d'un même algorithme général de réplication et de sélection. La théorie développée par le philosophe David L. Hull dans *Science as a Process* (1988), sert d'ancrage à ce travail. Le projet se distingue par l'utilisation d'une méthode encore peu usitée en philosophie et en sciences humaines : la construction d'un modèle multi-agents. Il s'agit d'élaborer une description formelle du comportement des agents qui composent un système et à encoder ce modèle dans un langage de programmation pour le simuler par ordinateur. En explorant l'espace d'états du modèle, on peut examiner l'influence de certains comportements individuels sur le succès épistémique des scientifiques et sur celui de la communauté dans son ensemble. Nous examinons principalement quatre facteurs : le compromis entre crédit et support dans les pratiques de citation scientifique, les comportements de citation des scientifiques envers leurs descendants intellectuels, l'importance accordée au crédit de l'auteur lors de l'évaluation d'une idée contenue dans un article, et la créativité intervenant dans la génération de nouvelles idées scientifiques.

**MOTS CLÉS :** modèles multi-agents, David Hull, épistémologie évolutionniste, paysages épistémiques, philosophie des sciences.

## INTRODUCTION

Quel que soit le statut qu'on est prêt à accorder aux théories que produit la science moderne, on peut difficilement contester à cette dernière une certaine efficacité. De nouvelles théories sont constamment proposées. Certaines sont écartées et d'autres sont retenues. Parfois, les théories retenues ouvrent de nouveaux champs de connaissances. D'autres fois, elles remplacent des théories rendues obsolètes. Il arrive aussi qu'elles cohabitent avec des théories rivales pendant un certain temps. Le paysage scientifique est constamment remodelé. Les théories changent, mais elles ne sont pas les seules. Les problèmes, les méthodes, les instruments, les modes de représentation et même les croyances des scientifiques quant aux buts qu'ils et elles<sup>1</sup> poursuivent changent aussi. Dans la mesure où l'on croit que ces changements augmentent l'adéquation entre les théories et le monde (ou à tout le moins notre emprise pragmatique sur celui-ci), on peut dire que la science « évolue ».

Une fois ce constat posé, reste à expliquer comment s'opèrent ces changements. Plusieurs réponses ont été tentées, certaines visant une reconstruction rationnelle des changements scientifiques, d'autres préférant s'attarder aux contingences sociales et historiques qui les accompagnent. Une autre voie consiste plutôt à tenter de décrire le mécanisme grâce auquel fonctionne la science. Une des thèses sur la nature de ce mécanisme, c'est qu'il faut laisser tomber les guillemets lorsqu'on dit que la science « évolue ». Les philosophies qui adhèrent à des versions plus ou moins fortes de cette thèse sont rangées (depuis Campbell, 1974) sous l'étiquette « épistémologie évolutionniste ».

---

<sup>1</sup>Il serait lourd d'écrire « ils et elles » à chaque fois. Au fil du texte, nous alternerons entre féminin et masculin, mais il faudra considérer que nos propos s'appliquent sans discrimination quant au genre.

Parmi ceux qui ont tenté d'expliquer le changement scientifique à l'aide d'un mécanisme évolutionniste, on trouve le philosophe de la biologie David L. Hull. Dans un ouvrage intitulé *Science as a Process : an Evolutionary Account of the Social and Conceptual Development of Science* (1988c), il expose les détails de ce qu'il croit être le fonctionnement de ce mécanisme.

Notre thèse poursuit trois objectifs :

1. Explorer les conditions sous lesquelles la science fonctionne le mieux. Plus spécifiquement : examiner l'influence de certains comportements individuels sur le succès scientifique de la communauté dans son ensemble.
2. Jeter un nouvel éclairage sur certains aspects de la théorie de Hull et, par extension, sur l'épistémologie évolutionniste en général.
3. Montrer que les modèles multi-agents sont un outil approprié pour remplir les deux premiers objectifs.

Nous sommes donc confrontés à la tâche de construire un modèle multi-agents inspiré du mécanisme d'évolution de la science exposé par Hull dans *Science as a Process*. Notre modèle ne pourra couvrir tous les aspects de la théorie de Hull : nous nous concentrerons sur la dynamique évolutionnaire de la science et le rôle joué dans celle-ci par certains comportements des scientifiques, notamment en ce qui a trait à la recherche de « crédit », notion centrale chez Hull.

Les modèles multi-agents ont plusieurs avantages, mais nous insisterons sur deux d'entre eux. Le premier est peut-être celui qui a l'impact le plus significatif pour la philosophie : ce sont des modèles qui peuvent être conçus de façon à avoir une portée normative. Le second avantage s'applique de façon plus générale à l'étude des systèmes complexes : la modélisation force l'explicitation.

Ce n'est pas d'hier que les philosophes des sciences utilisent des méthodes formelles. La logique est traditionnellement leur outil privilégié. Les empiristes logiques, par exemple, ont tenté de formaliser divers aspects des théories scientifiques : la relation

entre données sensorielles et théories (Carnap, 1928) la réduction interthéorique (Nagel, 1961), etc. Mais la logique (particulièrement la logique des prédicats du premier ordre) se heurte à certaines limites quand vient le temps de formaliser des systèmes complexes. Un système complexe est composé de multiples parties. Les propriétés du système dans son ensemble diffèrent de celles des parties et sont difficiles à prédire. Les modèles multi-agents sont adaptés à la formalisation de tels systèmes<sup>2</sup>, puisqu'ils permettent de se cantonner à décrire les éléments les plus simples du système et les interactions entre eux, puis de laisser ensuite l'ordinateur *simuler* le résultat, ce qui permet de constater l'émergence de nouvelles propriétés sans nécessiter de préconceptions quant à la nature de celles-ci. La science dans son ensemble est un système complexe : le progrès scientifique résulte (entre autres) des interactions entre les idées, les chercheurs et les artéfacts qu'ils produisent. La relation entre ces interactions et les propriétés de la science dans son ensemble serait difficile à exprimer avec la logique du premier ordre, mais les philosophes ont maintenant la possibilité de se tourner vers les modèles multi-agents.

Hull, étant donné les outils qu'il avait à sa disposition, a choisi d'exprimer sa théorie sous forme exclusivement verbale. Nous ne remettons pas en question la valeur d'une telle démarche, mais nous soutenons qu'une démarche supplémentaire de formalisation ajoute de la valeur à la théorie, entre autres parce que, pour être formalisée, la théorie doit être précisée. Une simulation informatique ne tolère aucune ambiguïté dans le système modélisé.

Le processus n'est pas trivial. Lors de la modélisation, il faut tenter de traduire le plus fidèlement possible la théorie prise pour cible, mais celle-ci comporte inévitablement des zones grises, ce qui nous force à prendre des décisions. Certaines de ces décisions sont de nature conceptuelle, par exemple la décision de modéliser une idée comme un point dans un espace à  $n$  dimensions. D'autres décisions devraient, idéalement, reposer sur des données empiriques. Il arrive que ces données ne soient pas dispo-

---

<sup>2</sup>Un modèle multi-agents est un système computationnel, et tout système computationnel peut être exprimé par le calcul lambda (Church, 1936), qui est lui-même une logique formelle. Il serait donc erroné de soutenir qu'une représentation purement logique des systèmes complexes est *impossible*. Si on veut en tirer les conséquences, notre modèle doit toutefois, pour des raisons pragmatiques, être encodé sous forme exécutable, dans un langage de programmation.

nibles et qu'il faille poser des hypothèses, quitte à réviser le modèle si cette situation change. Comment est distribuée, dans la population universitaire, la tendance à citer ses propres étudiants? Une réponse empirique existe à cette question. Nous ne la connaissons pas. Il nous faut statuer arbitrairement. On pourrait, par exemple, supposer que cette tendance suit une distribution uniforme. Si une étude scientométrique venait montrer que ce n'est pas le cas, le modèle pourrait être révisé.

La démarche de modélisation ne comporte pas qu'une exigence d'explicitation : elle comporte aussi une exigence de simplification. Plus un système est complexe, plus il devient difficile d'en identifier les éléments les plus significatifs. C'est une situation analogue au problème imaginé par Lewis Carroll (1893) de l'inutilité d'une carte géographique à l'échelle 1 : 1. C'est un piège dans lequel il est facile de tomber avec les simulations informatiques, qui mettent à notre disposition une puissance de calcul considérable. Cette puissance de calcul n'est toutefois pas infinie et c'est une autre raison de s'en tenir à un modèle simple.

La modélisation impose de nombreuses décisions. Pour les prendre de façon éclairée, il est nécessaire d'avoir un objectif concret en fonction duquel on pourra juger des avantages et des inconvénients des options qui s'offrent à nous. Ce qui nous ramène à la principale motivation pour construire un modèle multi-agents du processus scientifique : sa portée normative.

Dans la conclusion de *Science as a Process*, Hull énumère une série de questions soulevées par sa théorie :

« Apparently, lying is much more prevalent in science than stealing, but exactly how great are the discrepancies? How prevalent can stealing become before the likelihood that one will get credit for one's contribution is so reduced that the system ceases to work? Is it true that scientists violate financial trust much less frequently than other professional or are they simply much better at hiding their transgressions? Does science develop more quickly in areas characterized by competing factions than in areas where scientists work largely alone? What effect does the demic structure of science have on science? Social cohesion is necessary in maintaining research programs. How much cohesion is enough? How much is too much? Can group selection influence selection at the level of the individual scientist? How much competition is too much competition? What happens when science becomes too competitive? Scientists aggressively seek credit. Can scientists become too aggressive in their quest? What effects does

extreme aggression have on the careers of individual scientists? On science itself? »  
(Hull, 1988c, p. 520-521)

Certaines de ces questions exigent une démarche empirique (p. ex., à quel point le mensonge est-il plus fréquent que le vol en science ?), mais d'autres (p. ex., quel est l'effet, sur la science et sur les individus, d'une recherche trop agressive de crédit ?) se prêtent à la simulation. Incidemment, ces questions sont souvent celles qui revêtent un caractère normatif. Une fois qu'on a un modèle adéquat du fonctionnement de la science, on peut aborder l'épistémologie sous l'angle d'un problème d'optimisation : quelles sont les variations de paramètres qui sont susceptibles d'augmenter l'efficacité du processus scientifique ?

Les questions que nous aborderons à l'aide de notre modèle, bien qu'inspirées par celles que pose Hull, ne seront pas exactement les mêmes. Notre analyse sera guidée par la question générale : « Sous quelles conditions le système scientifique fonctionne-t-il le mieux ? » Concrètement, cela implique d'analyser les effets des différentes variations possibles dans les paramètres du système sur sa performance globale. Nous nous attarderons aussi à la question de savoir si l'intérêt individuel des scientifiques coïncide avec celui de la communauté en général. Hull suppose que c'est le cas (2001, p. 135). Nous chercherons à voir si notre modèle est en accord avec cette hypothèse.

Cette façon d'utiliser la modélisation multi-agents en philosophie des sciences est en accord avec la conception que proposait Quine d'une épistémologie naturalisée :

« Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. For me, normative epistemology is a branch of engineering. It is the technology of truth-seeking, or, in more cautiously epistemological term, prediction. Like any technology, it makes free use of whatever scientific findings may suit its purpose. [...] The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed. » (Quine, 1986, p. 664-665)

Cette thèse est constituée de six chapitres. Le premier chapitre présente les modèles multi-agents en général et le second chapitre présente les modèles multi-agents de la science en particulier. Le troisième chapitre, lui, s'attarde à la théorie de Hull (1988c) sur les mécanismes évolutionnaires et sociaux qui sous-tendent le processus scienti-

fique. Le cinquième chapitre met ensuite la table pour le modèle à venir en dressant un portrait du paysage philosophique autour de la question des théories, des modèles et des simulation. Une fois ces questions abordées, nous proposons un modèle multi-agents du processus scientifique qui s'appuie sur la théorie de Hull. Le sixième et dernier chapitre présente et analyse les résultats obtenus grâce à un ensemble de simulations utilisant ce modèle.

Ces différents chapitres présentent de façon exhaustive le domaine, la théorie, le modèle et les résultats, mais il nous semble tout de même approprié de faire ici un bref tour d'horizon qui permettra au lecteur de se faire une idée générale du projet avant de plonger dans les détails de celui-ci.

Ce tour d'horizon s'amorce avec quelques concepts clés de la théorie de Hull : les répliqueurs, les interacteurs, la valeur sélective conceptuelle et le crédit.

Nous avons mentionné d'entrée de jeu que la théorie de Hull propose une vision évolutionniste de la science. Cette vision repose sur l'idée que toute forme d'évolution, qu'elle soit biologique, scientifique, ou autre, dépend de la présence d'entités aptes à jouer le rôle fonctionnel qui permet à cette évolution d'avoir lieu : des *répliqueurs* et des *interacteurs*. Nous analyserons plus longuement ces concepts au chapitre 3 (section 3.2.1), mais il suffit pour l'instant de dire que les répliqueurs sont les entités qui sont transmises d'une génération à l'autre avec des variations et que les interacteurs sont les entités par l'intermédiaire desquelles cette transmission a lieu. En biologie, les répliqueurs sont les gènes et les interacteurs sont les organismes. En science, les répliqueurs sont les idées et les interacteurs sont les scientifiques.

Les interacteurs possèdent une « valeur sélective » (*fitness*) : ils sont plus ou moins susceptibles de laisser des copies de leurs répliqueurs, qui laisseront à leur tour, via une nouvelle génération d'interacteurs, d'autres copies d'eux-mêmes, et ainsi de suite. Dans le cas des scientifiques et de leurs idées, Hull parle de « valeur sélective conceptuelle ».

La valeur sélective conceptuelle d'une scientifique dépend, bien sûr, de la qualité de ses idées : les bonnes idées ont plus de chances d'être répliquées. Mais la valeur sélective

conceptuelle dépend aussi, selon Hull, du « crédit » que la communauté scientifique accorde à une scientifique : plus elle est reconnue par ses pairs, plus ses idées ont de chances d'être répliquées. Et puisque le « crédit » se traduit aussi généralement en succès professionnel (emploi, salaire, subventions, etc.), la recherche de crédit est au cœur de la démarche des scientifiques (voir section 3.3).

La principale façon de conférer du crédit à une scientifique, c'est de citer ses travaux. La scientifique qui en cite une autre reconnaît explicitement l'apport de cette dernière et « partage » ainsi le crédit rattaché à sa propre contribution scientifique<sup>3</sup>. En échange, la scientifique qui cite reçoit le « support » de la scientifique citée : en montrant que ses idées sont appuyées sur d'autres idées déjà reconnues, elle augmente les chances que les siennes soient acceptées. Pour obtenir du support, il faut partager son crédit. Hull fait grand cas de ce compromis (*trade-off*), et cette préoccupation sera reflétée dans notre modèle (voir section 5.4.4), où une variable  $\alpha$  représentera la position de chaque scientifique vis-à-vis de ce compromis. La valeur de  $\alpha$  est située dans l'intervalle  $[0; 1]$  : plus la variable  $\alpha$  d'une scientifique tend vers 1, plus elle accorde d'importance au support qu'elle peut obtenir.

En plus de la question de savoir quelle quantité de crédit ils sont prêts à partager avec leurs compétiteurs en échange de support, les scientifiques doivent aussi décider de la quantité de crédit qu'ils sont prêts à partager avec leurs étudiants. Ces derniers ne leur apportent pas beaucoup de support, mais ils ont autre chose à offrir : la promesse de transmettre à leur tour les idées héritées de leurs professeurs, augmentant par le fait même la valeur sélective conceptuelle de ces derniers. Cet apport de nos « proches parents » scientifiques donne lieu à ce que Hull appelle la « valeur sélective conceptuelle inclusive » (*conceptual inclusive fitness*). L'importance accordée par chaque scientifique à l'aspect inclusif de la valeur sélective conceptuelle sera représentée dans notre modèle par la variable  $\beta$ , aussi située dans l'intervalle  $[0; 1]$  : plus la variable  $\beta$  d'une scientifique tend vers 1, plus elle accorde d'importance à l'aspect « inclusif » de la valeur sélective conceptuelle.

---

<sup>3</sup> Il existe bien sûr des citations « négatives », qui portent un regard critique sur le travail cité, mais Hull suppose que le rôle de celles-ci est périphérique et que, de toute façon, même une citation négative confère une certaine quantité de crédit, puisqu'elle reconnaît par le fait même l'importance de l'œuvre citée.

Avant de décrire les entités qui composent notre modèle et la façon dont elles interagissent entre elles, il nous faudra introduire une notion qui jouera un rôle crucial : celle de paysage épistémique (voir section 5.2). Cette notion n'est pas directement présente dans la théorie de Hull. Son utilisation découle d'une contrainte propre à la modélisation : si on veut mesurer la performance d'un système épistémique artificiel, tel que notre modèle, il nous faut une façon d'attribuer une valeur objective aux « idées » qui sont générées par les agents du modèle.

La notion de paysage épistémique est fortement inspirée de la notion de paysage adaptatif (*fitness landscape*) utilisée en biologie évolutionniste<sup>4</sup> (Wright, 1932). Un paysage adaptatif est un espace à plusieurs dimensions, où une dimension (par convention, la hauteur) est utilisée pour représenter la valeur adaptative et les autres dimensions pour représenter les différents génotypes ou phénotypes possibles dans une population. Deux individus qui se ressemblent se trouveront à proximité l'un de l'autre dans cet espace. L'analogie avec un paysage est riche : elle nous permet de parler de populations qui se « déplacent dans l'espace » et doivent traverser des « vallées » de valeur sélective moindre pour atteindre des « montagnes » ou même des « pics » de valeur sélective.

Parmi les modèles multi-agents de la science que nous présenterons dans le chapitre 2, ceux de Weisberg et Muldoon (2009) et de Grim (2009) font intervenir la notion de paysage épistémique (voir section 2.5). Une épistémologie n'a pas besoin d'être évolutionniste pour que la notion de paysage épistémique puisse être utilisée : peu importe le mécanisme par lequel on suppose que le changement scientifique advient, il suffit de supposer que les objets produits par la science entretiennent entre eux des relations de proximité et qu'on puisse leur attribuer une valeur. C'est le cas des deux modèles mentionnés ici. Les objets situés dans le paysage épistémique sont des « approches » dans le cas de Weisberg et Muldoon et des « hypothèses » dans le cas de Grim, mais ça pourrait aussi être des « idées », des « théories », des « méthodes », ou tout autre objet auquel on peut attribuer une valeur épistémique.

---

<sup>4</sup>La notion de paysage adaptatif est aussi présente dans le domaine de la computation évolutionnaire (voir Luke, 2009), où on utilise ce qu'on appelle une « fonction objectif » pour tester la capacité d'un algorithme à trouver une solution optimale pour un problème en faisant varier un ensemble de paramètres dont les valeurs déterminent la position de la solution dans l'espace.

Le terme « valeur épistémique » laisse place à plusieurs interprétations. Selon différentes conceptions de l'épistémologie, il pourrait désigner une valeur de vérité, une probabilité objective, un potentiel de confirmation empirique, ou toute autre notion qui suppose une forme de correspondance entre l'objet et le monde. Weisberg et Muldoon, de même que Grim, refusent aussi de s'engager plus fortement, parlant respectivement de « *epistemic significance* » et de « *epistemic payoff* ». Dans notre modèle, nous parlerons de la valeur *objective*<sup>5</sup> des idées, qui sera à distinguer de la valeur *subjective* (potentiellement erronée) qu'attribuent les agents à ces différentes idées.

Pour le présent modèle, nous avons choisi de travailler avec des paysages épistémiques en trois dimensions : une pour représenter la valeur objective des idées, et deux pour représenter leur position dans l'espace des idées possibles. Nous examinerons le comportement du modèle dans six paysages différents (voir figure 5.2 à la page 137). Il va sans dire que, pour représenter adéquatement la diversité des idées scientifiques, plus de deux dimensions seraient nécessaires. Ce choix résulte cependant de contraintes bien réelles. Il a, d'abord, une valeur heuristique : travailler en trois dimensions nous permet de visualiser le paysage épistémique et la position des objets qui s'y trouvent. Et cette visualisation n'est pas que statique : elle permet d'observer les dynamiques de mouvement d'une population, ce qui est utile pour développer des intuitions sur le comportement du système, même si ces intuitions doivent ensuite être confirmées par une analyse numérique. Finalement, il y a la question de la lourdeur computationnelle. Plus on ajoute de dimensions, plus le modèle est gourmand en mémoire et en puissance de traitement.

Maintenant que nous avons abordé les concepts principaux de la théorie de Hull et la notion de paysage épistémique, il est temps de se pencher brièvement sur les entités

---

<sup>5</sup>L'idée même d'une valeur *objective* suppose une épistémologie non relativiste. C'est le cas chez Hull et, par extension, dans notre modèle.

qui composent notre modèle (nous le ferons plus longuement à la section 5.1). Il y en a trois types principaux : des idées, des scientifiques et des articles<sup>6</sup>.

Les objets les plus simples, ce sont les idées. Une idée possède un seul attribut interne : sa position dans l'espace épistémique. Cette position détermine la valeur objective de l'idée, mais celle-ci n'est, *a priori*, connue d'aucun agent.

Ceux et celles qui génèrent ces idées, ce sont les scientifiques, qui remplissent ainsi leur rôle d'interacteurs. À l'exception des idées « originelles » introduites au tout début de la simulation, les autres idées sont toutes créées à partir de variations sur une idée existante (leur « parent »), ce qui leur permet aux idées de jouer le rôle de répliqueurs.

Chaque scientifique fait aussi une évaluation de la valeur des idées auxquelles elle est exposée. Bien qu'elles aient la possibilité de « tester » une idée, aucune scientifique n'a un accès direct à sa valeur objective. Cette évaluation sera donc subjective, et on parlera de la « valeur subjective » d'une idée. Nous expliquerons en détail au chapitre 5 comment cette valeur est attribuée.

Les scientifiques entretiennent aussi des relations entre elles. D'abord, elles s'accordent l'une à l'autre une certaine quantité de *crédit*. Quand on parlera, dans le contexte du modèle, de la quantité de crédit que « possède » une scientifique, il ne s'agit en fait que de la somme du crédit qui lui est actuellement attribué par ses pairs. Si l'une d'entre elles quitte la scène, le crédit qu'elle attribuait à toutes les autres disparaît avec elle.

Ensuite, il y a les relations de professeur à étudiant. Les nouveaux scientifiques qui entrent dans le système commencent leur carrière en étudiant auprès d'un superviseur (et les scientifiques qui ont plus de crédit attirent plus d'étudiants). Le superviseur a

---

<sup>6</sup> Nous avons choisi de limiter notre modèle aux entités essentielles à la formalisation des aspects de la théorie de Hull que nous avons choisi d'aborder. Il serait tentant, bien sûr, d'ajouter à celles-ci différentes institutions qui participent à l'entreprise scientifique, mais cela augmenterait significativement la complexité du modèle et rendrait difficile son analyse. Cette entreprise est à ranger dans la catégorie « recherches futures ».

le privilège de transmettre à ses étudiants plusieurs de ses idées, ce qui a ensuite une influence sur sa décision de les citer ou non lorsque viendra le temps d'écrire des articles, comme le veut la théorie de la valeur sélective conceptuelle inclusive.

Les articles complètent notre triumvirat d'entités principales. Les articles sont générés par les scientifiques, qui choisissent pour chaque article une idée à exposer. L'article ainsi produit est situé dans le paysage épistémique à la position de l'idée correspondante. Il a donc lui aussi, par extension, une valeur objective. Et, bien sûr, un article en cite d'autres, le choix des articles à citer étant influencé par les variables  $\alpha$  et  $\beta$  mentionnées plus haut.

La brève description que nous venons de donner des entités qui composent le modèle est plutôt statique. Mais ce qui fait l'intérêt d'un tel modèle, c'est sa dynamique : comment ses composantes interagissent-elles au fil du temps ? Nos simulations se déroulent une année universitaire à la fois, et chaque année comporte les sept étapes suivantes :

1. Les scientifiques génèrent de nouvelles idées (section 5.4.1) ;
2. Les scientifiques procèdent à des tests empiriques et révisent leurs évaluations en fonction des résultats, ce qui leur permet de s'approcher de la valeur objective des idées testées (section 5.4.2) ;
3. Les scientifiques rédigent de nouveaux articles (section 5.4.3) ;
4. Les nouveaux articles sont évalués et les meilleurs sont publiés (section 5.4.4) ;
5. Les scientifiques lisent des articles. Ils évaluent les idées qui s'y trouvent et attribuent du crédit à l'auteur de l'article et à ceux des articles cités (section 5.4.5) ;
6. Certains scientifiques, plus vieux ou moins performants, prennent leur retraite (section 5.4.6) ;
7. Une nouvelle génération d'étudiants termine ses études et fait son entrée dans la communauté scientifique (section 5.4.7).

Au bout du compte, le succès d'une scientifique individuelle se mesure à la quantité de crédit qu'elle réussit à accumuler. Le succès de la communauté en général, lui, se mesure à la somme de la valeur objective des articles qui sont publiés. En utilisant ces deux mesures, notre modèle nous permettra de tester trois hypothèses distinctes, qui se concentrent autour des notions de compromis entre crédit et support et de valeur conceptuelle adaptative inclusive.

Voici ces hypothèses :

- H<sub>1</sub> : Hull suppose que les scientifiques doivent faire un *compromis* entre crédit et support : « Science is so organized that scientists are forced to trade off credit for support. » (1988a, p. 127) La position adoptée vis-à-vis à ce compromis est représentée pour chaque scientifique par la variable  $\alpha$ . Nous faisons l'hypothèse que les valeurs les plus avantageuses de  $\alpha$  pour un scientifique individuel devraient se situer autour de 0,5.
- H<sub>2</sub> : Hull suppose que les scientifiques doivent tenir compte de la valeur sélective conceptuelle *inclusive* lorsqu'ils font le choix des articles à citer : « graduate students [...] are likely to be the chief conduits for one's work to later generations. Their success increases one's own conceptual inclusive fitness. » (1988a, p. 128) Ce choix est représenté dans le modèle par la variable  $\beta$ . Nous faisons l'hypothèse que les valeurs les plus avantageuses de  $\beta$  pour un scientifique individuel devraient se situer autour de 0,5.
- H<sub>3</sub> : Hull suppose que l'intérêt des scientifiques individuels coïncide avec l'intérêt de la communauté scientifique en général : « By and large, what is good for the individual scientist is actually good for the group. » (1988a, p. 129). Nous faisons l'hypothèse que les valeurs de  $\alpha$  et de  $\beta$  qui sont les plus avantageuses pour les scientifiques individuels seront aussi les plus avantageuses pour la communauté en général.

Nous reviendrons en détail sur ces hypothèses au chapitre 6, où nous expliquerons quelles expériences nous avons réalisées pour les mettre à l'épreuve et ce que le modèle permet de conclure par rapport à elles.

Mais avant de clore cette introduction et de passer au vif du sujet, nous souhaitons mentionner trois paires d'oppositions conceptuelles qui traverseront, sous diverses formes, l'ensemble de cette thèse. Sans prétendre « dépasser » ces oppositions (quoique cela puisse vouloir dire exactement), nous croyons que, dans les trois cas, la théorie de Hull et le modèle qui en découle permettent de tenir compte de chacun des pôles sans rencontrer d'incompatibilités majeures.

Il y a d'abord l'opposition entre internalisme et externalisme en philosophie des sciences. Nous reviendrons sur celle-ci plus en détail à la section 3.1, mais le dilemme auquel on fait face est le suivant. Si on ne tient compte que des facteurs « internes » à la science (c'est-à-dire le raisonnement logique à partir de données empiriques), on passe à côté d'une bonne partie de la réalité du monde scientifique. Il devient difficile d'expliquer, en particulier, pourquoi des idées erronées connaissent parfois un certain succès. Mais lorsqu'on fait intervenir les facteurs « externes » (tout le reste, en fait : les facteurs sociologiques, politiques, psychologiques, etc.), c'est le contraire qui se produit : il devient difficile d'expliquer comment la science en vient à progresser malgré ces facteurs externes. Comme nous le verrons au chapitre 3, l'approche évolutionniste de Hull fournit un cadre où il est possible d'expliquer le progrès scientifique (ou son absence) en tenant compte des deux types de facteurs, cadre que nous tenterons de formaliser explicitement au chapitre 5.

Vient ensuite l'opposition entre le subjectif et l'objectif. Hull épouse à tout le moins une certaine forme de réalisme et suppose que : « Eternal and immutable regularities exist out there in nature. » (1988c, p. 476). Mais aucun d'entre nous n'a un accès *direct* à ces vérités objectives. Nous ne pouvons procéder que par approximations successives et raffiner notre compréhension du monde en générant constamment de nouvelles hypothèses, qui survivront ou non lorsque confrontées au monde empirique. La figure 5.4 à la page 149 montre bien comment le décalage entre la représentation subjective du monde chez un individu et la structure du monde objectif s'exprime dans notre modèle.

Finalement, il y a l'opposition entre les aspects individuels et les aspects collectifs de la science. Celle-ci tombe sous la coupe plus générale de l'opposition entre les niveaux « micro » et « macro » d'un système (voir chapitre 1, section 1.4 en particulier). La

science est une entreprise collective : non seulement le progrès scientifique repose-t-il sur l'accumulation de contributions qui ont été faites en se tenant « sur les épaules des géants », mais le paysage scientifique est aujourd'hui si vaste et si complexe que nul individu ne peut prétendre l'embrasser entièrement. La science n'a d'autre choix que de se constituer en système distribué, où des individus spécialisés explorent des sections locales du paysage épistémique. Une des questions qui se pose alors, c'est de savoir comment les interactions entre ces individus permettent à la communauté d'identifier les sommets du paysage épistémique. L'autre grande question, c'est celle qui sous-tend notre hypothèse  $H_3$  : l'intérêt des scientifiques individuels coïncide-t-il avec celui de la communauté scientifique ? Hull soutient que oui. Nous verrons au chapitre 6 ce que le modèle permet d'en dire.

Mais avant d'en arriver là, nous avons beaucoup de chemin à parcourir, et il est temps de s'engager sur celui-ci en nous penchant sur les modèles multi-agents de la science.

## CHAPITRE I

### INTRODUCTION AUX MODÈLES MULTI-AGENTS

La modélisation multi-agents n'est pas tout à fait nouvelle, mais a gagné en popularité dans les dernières années. En tant que méthodologie pour construire des modèles du processus scientifique, par contre, elle est encore marginale. Nous soutiendrons qu'elle possède des caractéristiques qui sont tout particulièrement utiles à cette entreprise.

Ce premier chapitre est une introduction aux MMA en général<sup>7</sup>. Nous verrons d'abord d'où ils viennent et quelles sont leurs caractéristiques principales. Nous mettrons ensuite de l'avant quelques considérations méthodologiques.

#### Un peu d'histoire

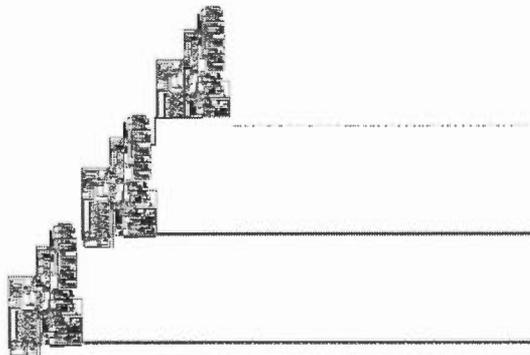
Les modèles multi-agents sont intimement liés à l'informatique et, sans grande surprise, on peut compter John von Neumann comme pionnier de ces deux disciplines. Vers la fin des années 1940, Von Neumann, ayant en tête des applications en intelligence artificielle, s'est intéressé aux systèmes auto-reproducteurs et auto-régulés.

---

<sup>7</sup>Une majeure partie du contenu de ce chapitre et du chapitre 2 est tiré de « Agent-Based Models of Science » (Payette, 2012) publié dans Scharnhorst *et al.* (2012).

Inspiré par les idées de son collègue Stanisław Ulam, il a conçu le premier automate cellulaire.

Son modèle consistait en un système de « cellules » disposées sur une grille orthogonale discrète (décrit plus tard dans von Neumann, 1966). Le temps, dans ce système, est aussi discret, et à chaque intervalle de temps, chaque cellule met à jour son état en fonction d'un ensemble de règles, en se basant sur son état précédent et sur l'état de ses voisins sur la grille. Chaque cellule est une simple machine à états finis, mais le comportement global du système peut être assez complexe. Von Neumann a utilisé ce cadre pour concevoir ce qu'il a appelé un « constructeur universel » : un patron de cellules capable de se reproduire soi-même au fil du temps. Ce faisant, il a montré comment une propriété systémique importante (l'autoreproduction) peut être obtenue grâce à l'interaction de parties individuelles dont le comportement est indépendant de celui du tout.

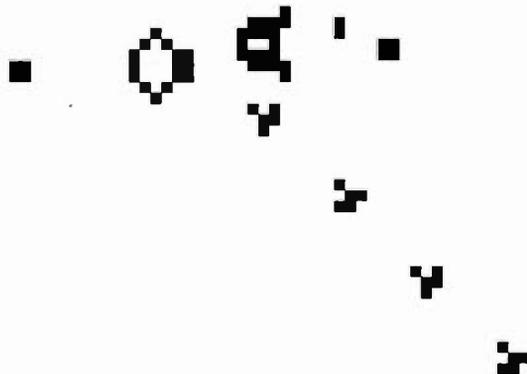


**Figure 1.1** Le « constructeur universel » de Von Neumann. Source : [http://en.wikipedia.org/wiki/File:Nobili\\_Pesavento\\_2reps.png](http://en.wikipedia.org/wiki/File:Nobili_Pesavento_2reps.png).

Ce qui a emmené les automates cellulaires à l'avant-plan, par contre, c'est le « Jeu de la vie » du mathématicien John Conway (Gardner, 1970). Alors que les cellules de von Neumann's pouvaient se trouver dans 29 états différents et que des douzaines de règles étaient nécessaires pour décrire les transitions entre ces états, les cellules de Conway sont soit « vivantes » ou « mortes », selon qu'elles ont ou non un « jeton » sur elles. Seules trois règles sont nécessaires pour décrire leur comportement (traduit de Gardner, 1970, p. 120) :

1. **Survie.** Chaque jeton ayant deux ou trois jetons voisins survit pour la génération suivante.
2. **Mort.** Chaque jeton ayant quatre jetons voisins ou plus meurt de surpopulation. Chaque jeton ayant un seul ou aucun jeton voisin meurt d'isolation.
3. **Naissance.** Chaque cellule vide adjacente à exactement trois jetons—ni plus, ni moins—est le lieu d'une naissance. Un nouveau jeton est placé sur elle lors de la génération suivante.

Ces règles simples, lorsqu'appliquées à différents patrons de cellules initiaux, donnent naissance à une ménagerie impressionnante (et bien documentée) d'objets au comportement complexe : « oscillateurs », « grenouilles », « vaisseaux », « puffeurs », « pulsars », « canons », « mathusalems », etc. Cela démontre à nouveau que des éléments simples peuvent être arrangés de façons qui mènent à des résultats surprenants (c.-à-d., difficiles à prévoir).



**Figure 1.2** Exemple de patron complexe dans le jeu de la vie de Conway. Il s'agit du « canon à navires » de Bill Gosper : celui-ci construit des « navires » qui s'en éloignent. Le canon continue à générer des navires tant qu'il n'est pas perturbé par des interférences causées par d'autres patrons. Source : [http://en.wikipedia.org/wiki/File:Gospers\\_glider\\_gun.gif](http://en.wikipedia.org/wiki/File:Gospers_glider_gun.gif).

Les systèmes que nous avons vus jusqu'à maintenant sont des modèles de phénomènes très généraux (la « vie », l'« autoreproduction »), mais l'idée des automates cellulaires est aussi applicable à plusieurs phénomènes sociaux. Indépendamment des débats autour de l'individualisme méthodologique, plusieurs problèmes des sciences sociales

peuvent être modélisés comme un ensemble d'agents individuels qui interagissent localement les uns avec les autres dans un espace explicite.

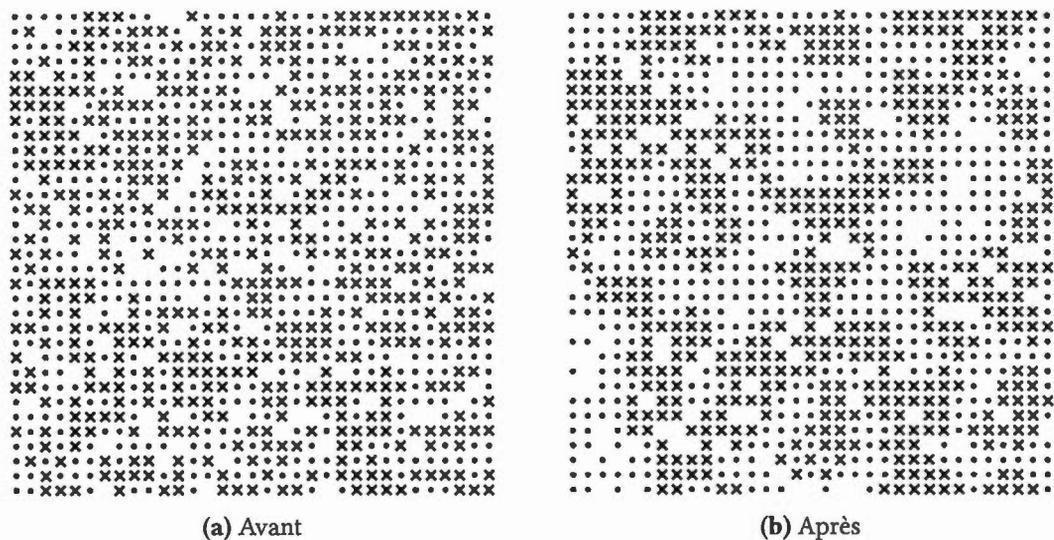
On considère habituellement que les premiers<sup>8</sup> de ces modèles sont les « modèles de ségrégation » de Schelling (1969 ; 1971b ; 1971a). Dans ceux-ci, Schelling explore les mécanismes qui mènent à la formation de regroupements d'agents homogènes dans l'espace géographique (autrement dit, de ghettos). Tout comme dans les automates de von Neumann et de Conway, l'espace y est modélisé comme une grille discrète. Cette fois-ci par contre, chaque cellule individuelle peut contenir un agent « humain » à qui on attribue une de deux (ou plus) « ethnicités » arbitraires. Ces agents ont des préférences quant à l'ethnicité de leurs voisins sur la grille. Si ces préférences ne sont pas satisfaites, ils se déplacent vers l'endroit le plus proche qui convient à leurs critères. Schelling a exploré les dynamiques de ces modèles pour plusieurs configurations initiales et pour plusieurs distributions de préférences.

À la surprise de Schelling lui-même, la proportion de voisins similaires exigée par les agents n'a pas besoin d'être bien élevée pour que la ségrégation apparaisse. Même lorsque le pourcentage désiré est de 30% (et que les agents sont donc prêts à tolérer jusqu'à 7 voisins sur 10 d'une autre ethnicité), une ségrégation assez forte s'installe rapidement. C'est ce que montre la figure 1.3 à la page suivante, où les deux ethnicités sont représentées par des points rouges (•) et des « x » bleus (×). On constate que l'état de mixité qui prévalait au début de la simulation (figure 1.3a à la page suivante) laisse place à la ségrégation (figure 1.3b) une fois le modèle stabilisé. Comme il le fait lui-même remarquer, le résultat particulier d'une simulation dépend des détails des conditions initiales, mais le caractère général de ce résultat, lui, n'en dépend pas.

Faisant un grand bond dans le temps, un autre jalon est le beaucoup plus complexe Sugarscape (Epstein et Axtell, 1996). Alors que la plupart des autres modèles sont conçus pour étudier un phénomène particulier, le Sugarscape se veut être une plate-forme d'expérimentation permettant d'étudier un vaste éventail de phénomènes : l'évolu-

---

<sup>8</sup> Le philosophe Rainer Hegselmann a récemment porté à mon attention l'existence d'un pionnier oublié : James M. Sakoda qui, dans sa thèse de doctorat (1949) anticipe d'une vingtaine d'années les idées de Schelling. Son modèle est publié plus tard (Sakoda, 1971) mais tombe ensuite dans une relative obscurité jusqu'à ce que Hegselmann et Flache (1998) contribuent à lui réattribuer un certain crédit.

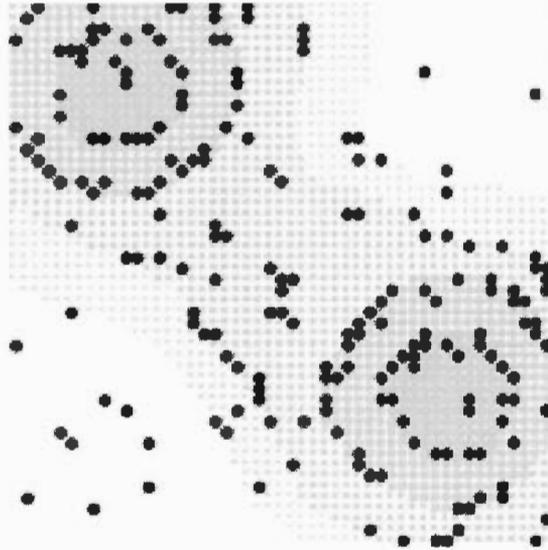


**Figure 1.3** Modèle de Schelling avant (a) et après (b) la simulation. On passe d'une ethnicité mixte à un état de ségrégation.

tion culturelle et biologique, le commerce, la guerre, la transmission de maladies, l'immigration, la pollution, etc. Les agents, dans le Sugarscape, sont aussi situés sur une grille, mais l'environnement n'est pas, cette fois-ci, qu'une coquille vide : il contient du « sucre » et des « épices », ressources génériques dont les agents ont besoin pour survivre. Les agents se déplacent sur la grille et recueillent ces ressources (qui, ensuite, « repoussent »). Les agents diffèrent aussi les uns des autres par d'autres caractéristiques que leur appartenance à un groupe : ils ont différents taux métaboliques, différentes portées visuelles et différentes espérances de vie. Ces différences donnent lieu à des opportunités d'interactions entre agents. Prenons par exemple le taux métabolique : si un agent a besoin de plus de sucre pour survivre, et que l'autre a besoin de plus d'épices, ils peuvent échanger sucre contre épices.

Bien que l'analyse du comportement des agents individuels sur la grille soit intéressante par elle-même, ce sont les phénomènes globaux, se produisant au niveau de la population, qui ont le plus d'intérêt pour les sciences sociales. Par exemple, la richesse individuelle des agents dans le Sugarscape—la quantité de ressources qu'ils ont accumulées—suit le « Principe de Pareto » : une distribution en loi de puissance où seuls quelques agents contrôlent la majeure partie de la richesse dans le système. Quoique le principe de Pareto ait été observé dans plusieurs « vrais » systèmes so-

ciaux (à commencer par la propriété terrienne en Italie au début du vingtième siècle), le Sugarscape est reconnu comme le premier système computationnel à générer ce phénomène grâce à un ensemble de micromécanismes suffisants pour en causer l'émergence.



**Figure 1.4** Une image du *Sugarscape*, tirée de la reproduction en code source ouvert que Tony Bigbee a faite du modèle d'Axtell et Epstein sur la plate-forme MASON (Bigbee *et al.*, 2007).

Nous avons choisi de suivre la voie historique des automates cellulaires pour introduire les modèles multi-agents, mais d'autres influences doivent aussi être reconnues. La théorie des jeux (p. ex., Axelrod et Hamilton, 1981), la vie artificielle (Reynolds, 1987), le connexionnisme (McClelland et Rumelhart, 1987), les algorithmes génétiques (Holland, 1975) et la recherche générale en intelligence artificielle ont tous joué, à divers degrés, des rôles importants.

Nous avons ouvert la présente section en affirmant que les modèles multi-agents sont intimement liés aux ordinateurs. Bien que vraie, cette affirmation peut être légèrement trompeuse : le plan de von Neumann pour son constructeur universel n'a été pleinement implémenté que bien plus tard (Pesavento, 1995), Conway a conçu le jeu de la vie en utilisant une planche de Go, et Schelling a « exécuté » la plupart de ses simulations en utilisant des pièces de monnaie comme jetons.

Dans plusieurs cas, les règles de comportement local sont assez simples pour que leurs résultats puissent être calculés « à la main ». L'ordinateur est requis quand le nombre d'agents et d'étapes dans la simulation devient trop grand pour les ressources computationnelles limitées d'un être humain.

La croissance des ressources computationnelles dans les dernières années a encouragé les chercheurs à concevoir des modèles toujours de plus en plus détaillés, cherchant à capturer les aspects des plus fins des phénomènes sociaux. Un coup d'œil rapide au *Journal of Artificial Societies and Social Simulations*<sup>9</sup>, par exemple, permet de le constater.

Qu'est-ce qu'un agent ?

Dans le survol qui précède, nous avons employé le terme « agent » à quelques reprises, sans toutefois le définir précisément. Cette omission découle d'une difficulté qui accompagne le domaine depuis longtemps, comme en témoignent les propos de Carl Hewitt, rapportés par Wooldridge et Jennings (1995, p. 116) :

« the question *what is an agent?* is embarrassing for the agent-based computing community in just the same way that the question *what is intelligence?* is embarrassing for the mainstream AI community. The problem is that although the term is widely used, by many people working in closely related areas, it defies attempts to produce a single universally accepted definition. »

Wooldridge et Jennings tentent de circonscrire cette difficulté en adoptant une distinction entre agentivité *forte* et agentivité *faible*. L'agentivité forte est conceptualisée en termes d'états mentaux qui s'appliquent aussi aux êtres humains. Ces termes varient selon la philosophie de l'esprit adoptée. Ils font généralement intervenir des notions intentionnelles telles que *croyances*, *désirs*, *intentions*, etc., mais on pourrait tout aussi bien définir l'agentivité forte en termes de réseaux connexionnistes, par exemple. L'agentivité faible, par contraste, se contente de déterminer les propriétés minimales requises pour que les « agents » puissent servir à construire le système cible.

---

<sup>9</sup><http://jasss.soc.surrey.ac.uk>

Pour mettre en place une notion d'agentivité faible relativement à un système cible, il faut identifier celui-ci. Cela nous emmène faire un détour qui nous permettra d'aborder une autre distinction : celle entre ingénierie logicielle et modélisation scientifique.

En ingénierie logicielle, on cherche à atteindre un but de la façon la plus efficiente possible avec les moyens à notre disposition. On peut penser, par exemple, à un système de contrôle aérien où plusieurs agents logiciels communiquent entre eux pour coordonner les vols entre différents aéroports. Dans un tel système, on souhaite simplement que chaque vol parvienne à destination de façon sécuritaire le plus rapidement possible et au moindre coût possible.

La modélisation scientifique, elle, cherche à reproduire un phénomène du monde réel dans le but de mieux le comprendre. On peut ici entendre « reproduire » et « phénomène » dans un sens très large. Pensons au jeu de la vie, abordé ci-dessus, qui modélise de façon très abstraite le phénomène de « la vie ». À l'autre extrémité du spectre, on peut avoir, par exemple, un modèle très concret de l'activité touristique actuelle dans les îles Galapagos (Pizzitutti *et al.*, 2014).

En anglais, la distinction se reflète dans le vocabulaire : on parle de *Multi-Agent Systems* en ingénierie et de *Agent-Based Models* en modélisation. En français, suivant Ferber (1997), on range souvent les deux types de démarche sous l'étiquette des « systèmes multi-agents ». L'utilisation que nous faisons ici de l'expression « modèles multi-agents » (MMA) désigne plus spécifiquement les *Agent-Based Models*.

Cela étant dit, la frontière entre ingénierie logicielle et modélisation scientifique est poreuse. L'ingénierie fait souvent usage de modèles. Leur utilisation est alors subordonnée au but à atteindre, mais n'en est pas moins cruciale : pour mettre à l'épreuve un système de contrôle aérien, par exemple, il est plus sage de le tester d'abord à l'aide d'un modèle de l'activité aérienne dans les aéroports ciblés plutôt que de le mettre directement en service. De même façon, un modèle d'un phénomène social ou naturel, s'il est bien conçu, permet d'examiner des scénarios très concrets (p. ex., que se

passerait-il si on construisait un nouvel aéroport sur l'île d'Isabela<sup>10</sup> ?) C'est d'ailleurs cette capacité à tester des scénarios hypothétiques qui nous permettra d'effectuer un passage du descriptif au normatif sur lequel nous reviendrons au chapitre 6.

Ces considérations nous emmènent à poser la question suivante : une notion d'agentivité forte est-elle nécessaire à la modélisation des phénomènes sociaux ? Dans le cas d'un système de contrôle aérien, le réalisme psychologique des agents logiciels importe peu, quoiqu'on pourrait aussi soutenir que l'adoption d'une notion d'agentivité forte est la façon la plus efficace de construire un tel système.

Mais qu'en est-il dans un contexte de modélisation multi-agents, lorsque les agents modélisés représentent explicitement des êtres humains ?

Plusieurs démarches de modélisation multi-agents reposent sur l'hypothèse implicite que ce sont les *interactions* entre les agents (et entre les agents et leur environnement) qui génèrent les phénomènes d'intérêt dans le système à l'étude. Si ce n'était pas le cas, on pourrait laisser tomber le « multi » et se concentrer sur un seul agent en isolation. Pour les besoins de cette hypothèse, il n'est pas nécessaire que l'architecture cognitive des agents soit un modèle de l'architecture cognitive humaine.

L'insistance des MMA sur les interactions entre agents ne les soustrait pas pour autant à toutes contraintes de réalisme psychologique : il faut à tout le moins que le *comportement* des agents soit un modèle du comportement humain. Mettre ensemble des agents omniscients qui auraient accès une information complète et posséderaient une capacité illimitée à traiter celle-ci, par exemple, ne nous permettrait de faire des inférences légitimes sur le système social étudié.

---

<sup>10</sup>Réponse de Pizzitutti *et al.* (2014, sect. 5.5) : « With a new airport in Isabela and without an upgraded touristic infrastructure on the same island that is able to attract and host an increased number of tourist agents, in the short term, no change in the tourist agent distribution will result among the islands of the archipelago. »

## Caractéristiques des MMA

Les contraintes de réalisme psychologique se retrouvent parmi les caractéristiques des MMA identifiées par Joshua M. Epstein (2006), sur lesquelles nous allons finalement nous pencher : *hétérogénéité, autonomie, espace explicite, interactions locales* et *rationalité limitée*. D'autres définitions sont possibles, bien sûr. Wooldridge et Jennings (1995) en recense plusieurs, et celle de Ferber (1997, p. 13), par exemple, a eu un impact significatif dans le monde francophone. Les principales caractéristiques évoquées se recoupent d'une liste à l'autre, mais il faut nous en tenir à un ensemble limité et la liste d'Epstein a le mérite d'être pensée en fonction des modèles multi-agents en sciences sociales. Elle nous semble aussi être appropriée pour caractériser les MMA des processus scientifiques, et nous nous y référerons au fil de ce chapitre lorsque viendra le temps d'aborder certain de ces modèles.

Quoi qu'il en soit, les caractéristiques évoquées ne devraient pas être considérées comme des conditions nécessaires pour qu'un modèle ait droit au titre de « modèle multi-agents ». Elles devraient plutôt être vues comme établissant une sorte de ressemblance de famille wittgensteinienne. Elles ne sont pas non plus orthogonales. Certaines d'entre elles, comme les interactions locales et l'espace explicite, par exemple, sont intimement liées.

L'*hétérogénéité* exige que les agents ne soient pas seulement regroupés dans quelques catégories homogènes (Epstein, 2006, p. 6). Ils peuvent différer l'un de l'autre d'autant de façons que l'espace des paramètres pour chacune de leurs propriétés individuelles le permettra. Et bien que ça soit quelque chose qui serait très difficile à rendre avec un modèle analytique traditionnel, l'ordinateur permet potentiellement de travailler avec des millions d'agents qui seraient tous différents les uns des autres.

J'ajouterai à cette définition que l'on peut penser à ces différentes propriétés comme étant statiques ou dynamiques, et internes ou externes. Les propriétés statiques sont celles qui ne changent pas au cours de la vie d'un agent. Cela ne veut pas dire qu'elles doivent être considérées comme « innées » : c'est simplement que leur valeur reste constante au fil de la simulation.

Plus intéressantes, peut-être, sont les propriétés dynamiques des agents : celles qui changent et dont la valeur peut-être suivie au fil d'une exécution de la simulation. Une propriété dynamique peut être aussi simple que la quantité d'argent à laquelle un chercheur a actuellement accès, mais elle peut aussi être une propriété plus complexe : la liste des propositions qu'une scientifique considère comme vraies, le but actuel de sa recherche, etc.

Les propriétés que nous avons données en exemple jusqu'à maintenant sont toutes des propriétés internes. Ce que nous appelons les propriétés externes concerne les relations entre un agent et son environnement. À quelle université un chercheur est-il attaché ? Qui sont ses collaborateurs ? Si l'espace est représenté, où se trouve-t-il ? Les propriétés externes sont souvent dynamiques, mais elles peuvent aussi être statiques, selon ce que le modèle cherche à capturer.

L'*autonomie* réfère à l'absence de contrôle centralisé. Dans le contexte de la simulation sociale, cette idée est en lien avec l'individualisme méthodologique : bien que les institutions (et autres macrostructures) puissent établir des politiques (règles, valeurs, etc.) qui influencent le comportement des agents, elles ne sont pas en mesure de contrôler directement les agents. À chaque étape de la simulation, les agents prennent leurs propres décisions en fonction de leurs buts individuels.

L'*espace explicite* exige que les agents soient situés dans un environnement défini. Les comportements possibles pour l'agent sont partiellement déterminés par sa position dans cet environnement. Dans plusieurs MMA, cet environnement est une grille qui représente un espace géographique, mais il n'est pas nécessaire qu'il en soit ainsi. Cela peut être quelque chose de plus abstrait comme (nous le verrons plus tard) la position d'une scientifique dans un espace épistémique ou encore dans un réseau social de collaboration. Pour citer de nouveau Epstein : « The main *desideratum* is that the notion of "local" be well posed » (2006, p. 6). Cette exigence est directement liée à la propriété suivante.

Les *interactions locales* sont typiques des MMA. Quand les agents interagissent avec d'autres agents, c'est habituellement avec leurs voisins—ceux qu'ils côtoient dans l'espace géographique ou dans l'espace social : leurs collaborateurs, leurs collègues, leurs

étudiants, etc. Le fait que ça ne soit pas tout le monde qui interagisse avec tout le monde peut faire une différence importante dans certaines situations. Les simulations de Zollman (2007) et de Grim (2009), par exemple, sur lesquelles nous nous pencherons plus loin, montrent d'importants effets épistémiques de la non-universalité de la communication dans les réseaux scientifiques.

La *rationalité limitée*, finalement, affirme que les agents n'ont pas accès à l'information globale sur le système, et n'ont pas non plus une puissance computationnelle illimitée. Ils agissent normalement selon des règles simples basées sur de l'information locale (Epstein, 2006, p. 6).

La question des niveaux : micro et macro

La plupart des considérations reliées à la modélisation en général s'appliquent aussi aux MMA. Nous reviendrons plus en détail aux questions philosophiques concernant les modèles multi-agents au chapitre 4 (notamment en ce qui a trait à leur fonction explicative), mais nous souhaitons tout de même aborder ici la question des niveaux de modélisation, qui est cruciale pour la compréhension de ces modèles.

Comme nous l'avons laissé entendre plus haut, les MMA sont concernés par les processus de niveau *micro* qui donnent naissance à des phénomènes observables de plus haut niveau. Si un MMA peut générer un macro-phénomène, on doit considérer que le mécanisme qu'il utilise pour ce faire peut à tout le moins être considéré comme un candidat pour l'explication du phénomène en question. Et quand on prend la possibilité de ce type d'explication au sérieux, celle-ci peut même devenir une exigence. C'est ce que Josh Epstein appelle « la devise générativiste » : « *If you didn't grow it, you didn't explain it.* » (2006, p. 51)

Dans l'optique générativiste, une régularité de haut niveau comme, par exemple, la loi de Lotka (Lotka, 1926) doit être expliquée, et une dérivation algébrique de cette loi, comme celle de Herbert Simon (1955, p. 148), ne suffit pas à cette explication. Il est nécessaire de spécifier le *mécanisme* qui aurait pu générer cette distribution. Dans

le cas particulier de la loi de Lotka, un mécanisme possible a été proposé par Nigel Gilbert (1997), comme nous le verrons dans la section 2.1.

Ce qui est en jeu, c'est notre conception de l'explication scientifique. Le modèle classique d'explication déductive-nomologique de Hempel et Oppenheim, par exemple, ne pose pas formellement l'exigence d'une « microthéorie », mais affirme néanmoins que celle-ci est souhaitable pour expliquer un phénomène :

« It is often felt that only the discovery of a micro-theory affords real scientific understanding of any type of phenomenon, because only it gives us insight into the inner mechanism of the phenomenon, so to speak. » (Hempel et Oppenheim, 1948, p. 147)

Au-delà du « sentiment » d'une « réelle » compréhension scientifique d'un phénomène, c'est l'objectif d'en arriver éventuellement à une science unifiée qui se profile ici. Renoncer à la possibilité d'expliquer un phénomène de niveau « macro », c'est en quelque sorte admettre que celui-ci est « émergent » au sens classique du terme, supposant qu'il ne pourrait pas (même en principe) être inféré à partir du comportement de ses composantes de bas niveau. On renonce alors à tout lien formel interniveaux. Dans les mots célèbres de Samuel Alexander :

« The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the “natural piety” of the investigator. It admits no explanation. » (1920, p. 46–47)

Hempel et Oppenheim (p. 152) trouvent cette attitude « étouffante » (*stifling*) pour la recherche scientifique, et c'est à ce diagnostic que fait écho la devise d'Epstein. Cela ne veut pas dire qu'une explication donnée en termes purement « macro » est invalide, ou même inutile. La macroéconomie, par exemple, peut très bien fonctionner sans l'aide de la microéconomie<sup>11</sup>. Ses théories sont à la fois utiles et empiriquement validées. Cela suppose, par contre, que de telles théories sont incomplètes, ce qui est somme toute une critique assez mineure dans la mesure où l'idéal d'une science complète (si tant est qu'il soit atteignable) constitue un horizon vers lequel on souhaite tendre et non pas l'aune à laquelle on mesure toute théorie.

---

<sup>11</sup>Ce qui n'empêche pas la modélisation multi-agents d'y faire quelques avancées sous l'étiquette *Agent-Based Computational Economics* (Amman *et al.*, 2006 ; Hamill et Gilbert, 2016).

Notons que la « science unifiée » que nous évoquons ici n'implique pas un réductionnisme à la Nagel (1961). Le schéma nagélien fait intervenir une relation de déduction entre les lois de la théorie réduisante  $T$  et celles de la théorie réduite  $T'$ . Et pour opérer cette réduction, il faut faire intervenir des « lois-pont » qui relient les termes de  $T$  à ceux de  $T'$ . Dans un générativisme inspiré d'Epstein, il n'est nullement question d'utiliser un modèle basé sur une théorie  $T$  de niveau micro pour déduire *les lois* d'une théorie  $T'$  de niveau macro. La seule contrainte, c'est que les résultats du modèle soient *compatibles*<sup>12</sup> avec les lois de  $T'$ . Si c'est le cas, le modèle aura à tout le moins servi à montrer qu'il est *possible* de générer les phénomènes décrits par  $T'$  en se basant sur les lois de  $T$ . On aura ainsi troqué une vision de la science où les relations entre les théories de niveau micro et les théories de niveau macro sont des relations de nécessité basées sur des lois-ponts pour une vision où nous avons plutôt affaire à des relations de possibilité basées sur ce que nous suggérons d'appeler des « modèles-ponts ». On aurait donc, à terme, une science où chaque théorie serait reliée, par l'intermédiaire d'au moins un modèle-pont, à une théorie de niveau inférieur.

Quoi qu'il en soit, la question de l'échelle à adopter pour un modèle se pose. Le sens des termes « micro » et « macro », dans le contexte du modèle, est relatif à ce choix. Après tout, si nous décidions de modéliser une scientifique cellule par cellule, le comportement de la scientifique serait de niveau macro. Dans les modèles multi-*agents*, il est présumé que ce sont les agents qui occupent le niveau micro. Mais qu'est-ce qu'un agent ? La plupart des MMA de la science adoptent les scientifiques individuels comme agents, mais rien, en principe, ne nous empêche de considérer plutôt les équipes de recherche, les laboratoires, les institutions universitaires, ou même les pays comme des agents.

Dans le modèle de Gilbert (choix contestable, peut-être) ce sont les articles scientifiques qui jouent le rôle des agents. Au bout du compte, il revient au modélisateur d'identifier ce que Claudio Cioffi-Revilla appelle « la distribution des rôles princi-

---

<sup>12</sup>Cette contrainte est similaire à celle posée par le modèle de réduction indirecte de Kemeny et Oppenheim (1956), lui-même un cas particulier du « paradigme général de réduction » proposé plus tard par Schaffner (1967).

paux » : « the main social entities themselves and their main interactions or causal dynamics » (2009, p. 30).

Une fois que le niveau cible a été déterminé et que les entités correspondantes ont été identifiées, il reste à régler la question du degré de détail à adopter pour la modélisation. Les premiers MMA utilisaient habituellement des agents très simples. Dans les modèles de Schelling, les seules propriétés d'un agent étaient sa position et son niveau de tolérance. Les propriétés intéressantes d'un modèle sont celles qui résultent des relations et des interactions entre les objets qui le composent, et non les propriétés des objets eux-mêmes. Il est toutefois important de s'assurer que ce ne sont pas les simplifications adoptées qui sont responsables du comportement du modèle.

Nous verrons, dans le chapitre suivant, comment les principaux pionniers de la modélisation multi-agents du processus scientifique ont abordé ces questions.

## CHAPITRE II

### LES MODÈLES MULTI-AGENTS DE LA SCIENCE

Dans le chapitre précédent, nous avons introduit les modèles multi-agents en général. Le temps est venu de nous pencher sur quelques modèles qui prennent le processus scientifique pour objet. Nous commencerons par le modèle original de Nigel Gilbert (1997), qui est peut-être le MMA de la science le plus connu. Nous allons le décrire en détail et l'utiliser ensuite comme point de référence pour évaluer les autres modèles.

Le modèle original de Gilbert : articles et kènes

Gilbert cherche à expliquer les régularités quantitatives qu'on retrouve traditionnellement en science. Cela inclut la loi de Lotka<sup>13</sup>(1926), mais aussi beaucoup des caractéristiques de la « petite science » identifiées par de Solla Price dans *Little Science, Big Science* (1963) : par exemple, la croissance exponentielle du nombre d'articles scientifiques et le fait que les articles tendent à référer à la littérature plus récente.

Le point de départ de Gilbert est un modèle visant à simuler la loi de Lotka. Dans les mots de Gilbert, la loi de Lotka dit que : « for scientists publishing in journals, the number of authors is inversely proportional to the square of the number of papers

---

<sup>13</sup>Pour une autre simulation de la loi de Lotka, voir Saam et Reiter (1999).

published by those authors » (1997, sect. 4.1). La plupart des auteurs ne publient qu'un ou deux articles, quelques-uns d'entre eux en publient un peu plus et très peu en publient plus de dix. Herbert Simon (1955) décrit la probabilité qu'un article soit publié par un scientifique ayant déjà  $i$  publications comme étant :  $f(i) = a/i^k$  (où  $a$  et  $k$  sont des constantes). Simon a aussi trouvé une probabilité constante,  $\alpha = n/p$ , que le prochain article dans un journal soit écrit par un auteur jamais publié auparavant (où  $p$  est le nombre d'articles dans un journal et  $n$  le nombre d'auteurs<sup>14</sup>). Le mécanisme proposé par Gilbert pour simuler ces régularités est assez simple :

- « 1. Select a random number from a uniform distribution from 0 to 1. If this number is less than  $\alpha$ , give the publication a new (i.e., previously unpublished) author.
2. If the publication is not from a new author, select a paper randomly from those previously published and give the new publication the same author as the one so selected. » (1997, sect. 4.4)

Il est important de noter que Gilbert n'utilise pas directement  $f(i)$ . Les données produites par le modèle de Gilbert sont à la fois très proches des estimations de Simon et des données bibliométriques, même si la simulation est agnostique quant à la distribution attendue des articles par auteurs.

Notons aussi que le modèle est centré sur les articles : ils sont, en quelque sorte, les « agents » dans cette simulation. Cela reste le cas quand Gilbert passe à une simulation plus complexe, où les articles ont, cette fois, un contenu. Chaque article contient un « quantum » de connaissance, représenté par ce que Gilbert appelle un « kène ». Un kène est une séquence de bits qui pourrait, en théorie, être de n'importe quelle longueur. Pour permettre l'affichage des kènes dans un espace à deux dimensions, Gilbert leur donne une longueur de 32 bits, ce qui permet d'encoder deux nombres entiers de 16 bits représentant des coordonnées  $x, y$  sur une grille de  $65\,536 \times 65\,536$ , ce qui permet de parler de la position d'un kène ou de celle d'un article. (La position d'un article est la même que celle du kène qu'il contient.)

Le terme « kène » est choisi pour sonner comme « gène », et la raison pour cela est qu'il y a une composante évolutionnaire dans le modèle de Gilbert. À chaque intervalle

---

<sup>14</sup>Il y a une coquille dans l'article de Gilbert, où il est écrit que  $\alpha = p/n$  alors que c'est plutôt  $\alpha = n/p$ .

de temps, au moins un article se « reproduit » et tous les autres articles<sup>15</sup> ont aussi une petite probabilité  $\omega = 0,0025$  de se reproduire. L'auteur d'un nouvel article est soit un nouvel auteur (avec une probabilité de  $\alpha$ ), soit l'auteur de l'article « parent ». Le nouvel article a, lors de sa création, le même kène que son parent. Il a aussi ses références : il choisit, au hasard, d'autres articles situés dans un rayon de  $\epsilon = 7\,000$ . Chacune de ces références a une influence sur le kène original, de telle façon que le kène du nouvel article est une combinaison du kène original et des kènes de ses références. Si on pense à un kène comme à un point dans l'espace, on peut penser à chacun des kènes des références comme exerçant une attraction gravitationnelle qui « tire » le kène du nouvel article dans sa direction. Plus formellement, étant donné un nombre aléatoire  $m$  entre 0 et 1, qui augmente de façon monotonique avec chaque référence :

$$x'_p = x_p + (x_r - x_p) \frac{1 - m}{2} \quad \text{et} \quad y'_p = y_p + (y_r - y_p) \frac{1 - m}{2} \quad (2.1)$$

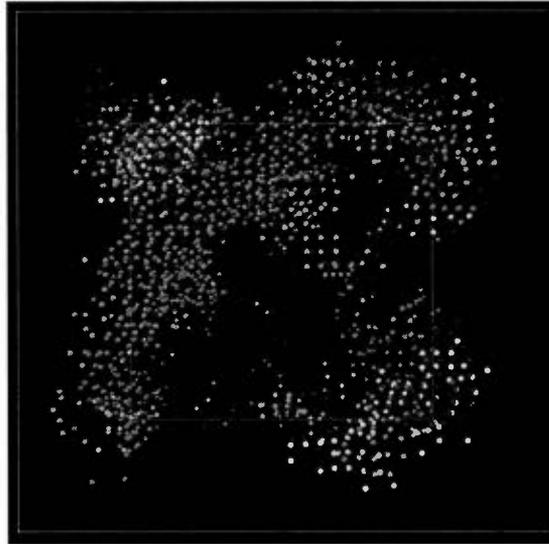
Ce modèle plus détaillé produit aussi une distribution d'articles par auteurs qui suit la loi de Lotka, ce qui n'est guère surprenant puisque le mécanisme qui génère cette distribution est presque identique à celui du modèle précédent<sup>16</sup>. Le modèle produit aussi une distribution asymétrique du nombre de citations par auteurs, qui suit qualitativement les données empiriques. Le taux de croissance du nombre d'articles (qui dépend de la probabilité  $\omega$  de générer un nouvel article) correspond aussi aux observations de de Solla Price.

Finalement, un nouveau résultat de ce modèle plus complexe est qu'on peut maintenant observer différentes « grappes » (*clusters*) d'articles dans l'espace des kènes possibles. C'est une conséquence du mécanisme évolutionnaire choisi par Gilbert, où chaque nouvel article est créé à proximité de son parent. Ces grappes sont interprétées par Gilbert comme représentant différentes spécialités à l'intérieur d'un champ. Cette interprétation pose problème en ce que la position du kène n'est pas prise en

---

<sup>15</sup>Quoique Gilbert ne le mentionne pas explicitement, sa simulation est initialisée avec un certain nombre d'articles séminaux : par exemple, 1 000.

<sup>16</sup>La seule différence est que les auteurs sont maintenant « retirés » après un intervalle de temps aléatoire (où le maximum  $\phi = 480$ )

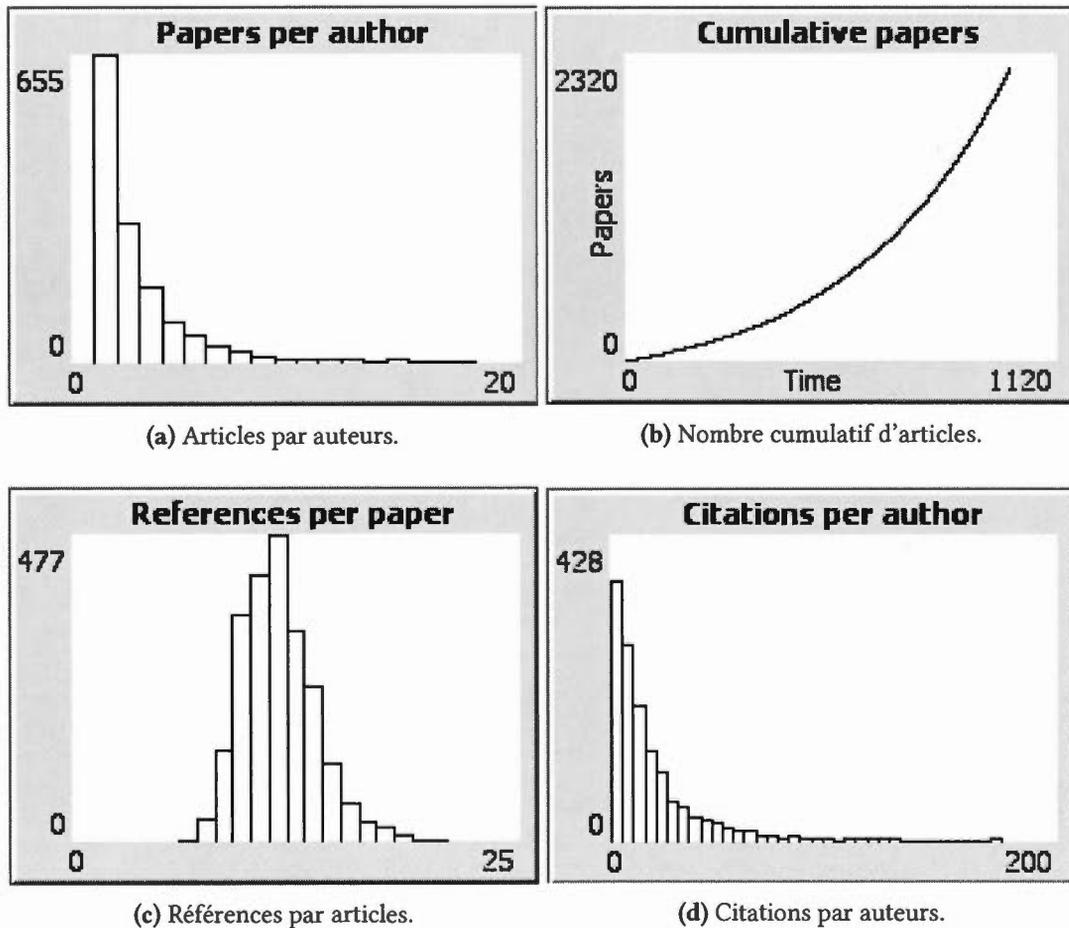


**Figure 2.1** Représentation de l'espace dans la simulation de Gilbert, montrant le regroupement des articles en « spécialités ».

compte lorsque le nouvel article « choisit » son auteur. Il serait assez simple, toutefois, d'intégrer cette considération dans la simulation (par exemple, en augmentant la probabilité qu'un auteur soit choisi s'il a récemment écrit un article se trouvant à proximité du nouvel article).

Quoiqu'assez simple, le modèle de Gilbert est un exemple frappant des possibilités de la modélisation multi-agents en sciences. Gilbert lui-même, avec ses collaborateurs Andreas Pyka et Petra Arweiler, a poussé l'idée plus loin dans une série d'articles sur les réseaux d'innovation (Gilbert *et al.*, 2001, 2007 ; Pyka *et al.*, 2002, 2007, 2009 ; Arweiler *et al.*, 2004). Il a aussi inspiré directement des modèles dans des domaines assez différents (p. ex., Boudourides et Antypas, 2002).

Dans les sections suivantes, nous nous pencherons sur des modèles par Sun et Naveh (2009), Muldoon et Weisberg (2011) et Edmonds (2007). Nous poursuivrons ensuite avec des modèles par Weisberg et Muldoon (2009) et Grim (2009) qui, quoique pas directement préoccupés par les questions bibliométriques explorées par Gilbert, sont liés de près à une idée que celui-ci est venu bien près d'aborder avec sa distribution spatiale des kènes : celle de paysage épistémique.



**Figure 2.2** Relations numériques dans une exécution particulière de la simulation de Gilbert. La figure 2.2a montre une reproduction approximative de la loi de Lotka. Les figures sont tirées de la réplique NetLogo du modèle original de Gilbert, disponible sur <http://www.openabm.org/model/2296/version/2/view>.

## Un point de vue inspiré des sciences cognitives

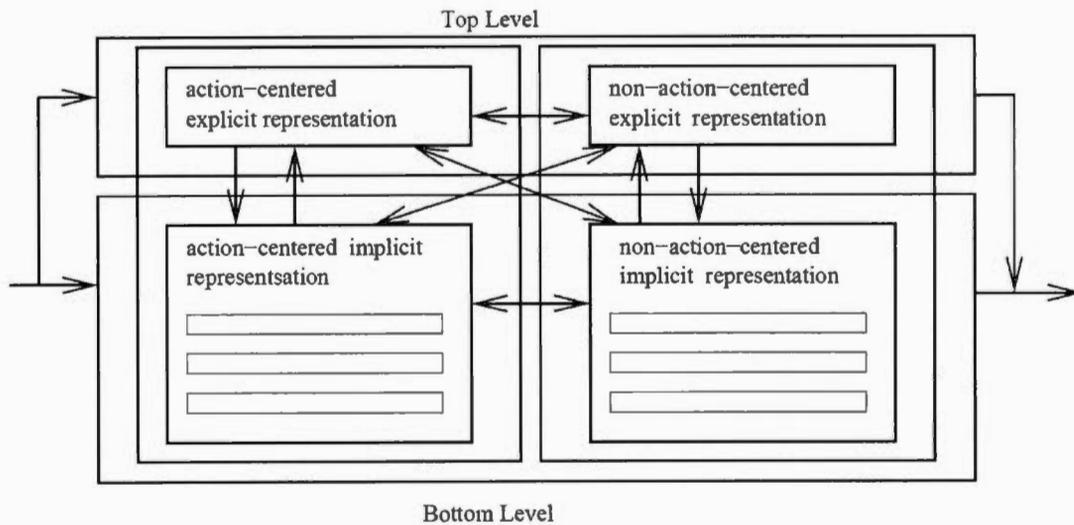
Le fait que les scientifiques ne jouent qu'un rôle mineur dans le modèle de Gilbert donne prise à la critique. Il peut être difficile d'accepter l'idée que la seule différence entre une auteure qui a publié un seul article et une qui en a publié quinze est que plus d'articles ont « sélectionné » cette dernière.

Ron Sun et Isaac Naveh, en particulier, n'ont pas hésité à cibler cette lacune : « Gilbert's model lacks agents capable of meaningful autonomous action » (2007, p. 142). Ils visent plutôt un modèle où : « authors are not merely passive placeholders, but cognitively capable individuals whose success or failure depends on their ability to learn in the scientific world » (2006, p. 321). Pour y arriver, ils utilisent une architecture cognitive qu'ils appellent CLARION, acronyme pour « *Connectionist Learning with Adaptive Rule Induction ON-line* ». Ce nom constitue une assez bonne description de ce que CLARION fait. Il s'agit d'une architecture hybride (figure 2.3 à la page suivante), qui possède un mécanisme d'apprentissage implémenté par un réseau de neurones artificiels, mais qui est en mesure d'extraire des règles symboliques à partir de ce qui a été appris au niveau connexionniste et d'utiliser ces règles pour guider son comportement. Nous n'allons pas entrer dans les détails de CLARION<sup>17</sup>, mais cette architecture vise à être la plus réaliste possible sur le plan cognitif. Sun lui-même a soutenu avec insistance la nécessité de ce type de systèmes cognitifs hybrides (Sun, 2002) et ceux-ci sont de plus en plus courants en sciences cognitives (Evans, 2008).

Dans le modèle de Sun et Naveh, comme on s'y attendrait normalement, ce sont les scientifiques qui choisissent quelles idées seront répliquées, et non pas l'inverse. Les scientifiques choisissent aussi les idées voisines qui seront utilisées pour modifier l'idée originale, mais ils ne s'arrêtent là : ils optimisent aussi l'idée par eux-mêmes, en cherchant l'espace environnant pour y trouver une meilleure position (les idées, ici aussi, sont représentées par des points dans un espace multidimensionnel.) Le fait qu'une telle optimisation ait cours implique que, contrairement à ce qui était le cas dans le modèle de Gilbert, certaines idées sont meilleures que d'autres. Sun et Naveh

---

<sup>17</sup>Voir Sun (2006) pour un aperçu général et Sun (2003) pour une description détaillée

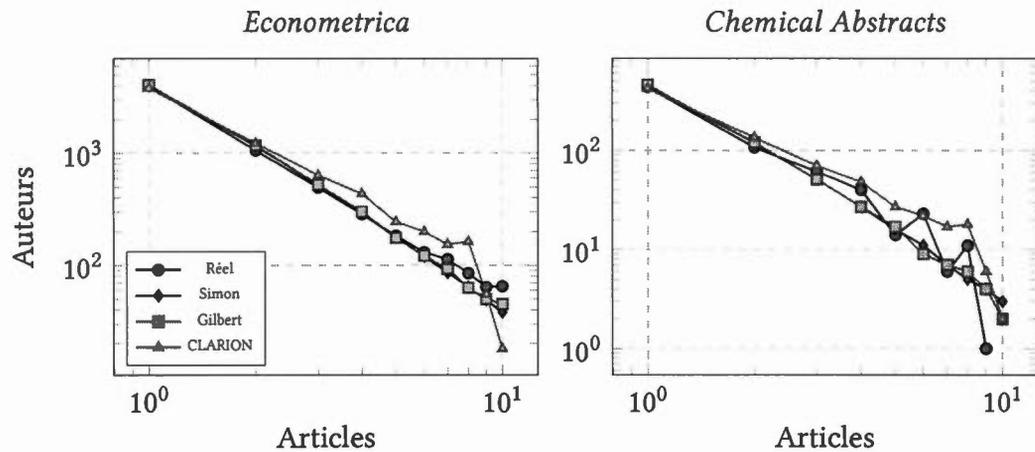


**Figure 2.3** Architecture CLARION (Naveh et Sun, 2006), montrant l'interaction entre représentations explicites (symboliques) et implicites (connexionnistes).

mentionnent quelques propriétés par lesquelles les idées diffèrent : clarté, perspicacité (*insightfulness*), soutien empirique, résultats théoriques et potentiel d'application. Tous les agents possèdent une « fonction subjective » pour évaluer ces différentes propriétés d'une idée, et ils raffinent cette fonction au fil de la simulation, tout en tentant de s'approcher de la fonction « communautaire » qui détermine si un article est publié ou non. Aucun agent n'a accès à la fonction communautaire : tout ce qu'ils ont, c'est la rétroaction qu'ils obtiennent lorsqu'ils soumettent un article, à savoir s'il est publié ou non. Ils utilisent cette rétroaction pour optimiser deux tâches : (1) choisir l'idée cible et (2) choisir les idées qui vont l'influencer. Les agents qui n'arrivent pas à publier assez d'articles sont retirés de la simulation et remplacés par de nouveaux agents. Dans le modèle de Sun et Naveh, c'est ce processus d'apprentissage, plutôt que la chance, qui est responsable de la différence dans le nombre de publications de chaque agent.

Le modèle de Sun et Naveh correspond aux données empiriques, mais pas d'aussi près que celui de Gilbert. Il y a, par contre, une bonne raison pour cela :

« We put more distance between mechanisms and outcomes, which makes it harder to obtain a match with the human data. Thus, the fact that we were able to match



**Figure 2.4** La distribution du nombre d'auteurs ayant publié entre 1 et 10 articles dans les revues *Econometrica* et *Chemical Abstracts* selon la réalité, l'équation de Simon et les modèles multi-agents de Gilbert et de Sun et Naveh.

the human data shows the power of our cognitive agent-based approach compared to traditional methods of simulation. » (Naveh et Sun, 2007, p.200-201)

Le modèle de Sun et Naveh leur permet d'étudier les effets des différences cognitives sur le succès de la communauté dans son ensemble. Ce dernier est mesuré par le nombre total d'articles publiés. Dans le modèle de Gilbert, ce nombre est un résultat direct du paramètre  $\omega$  (la probabilité qu'un article donne naissance à un nouvel article). Ici, c'est plutôt le résultat de la capacité des agents à apprendre les règles de publication de la communauté. Les paramètres cognitifs sont nombreux (par exemple, le taux d'apprentissage des agents, la probabilité d'utiliser l'apprentissage implicite plutôt que l'apprentissage explicite, la part d'aléatoire dans le processus d'optimisation locale) et ils ont tous un effet sur le nombre total d'articles publiés.

#### Un modèle de la distribution des ressources cognitives

Nous poursuivons avec un modèle conçu par Ryan Muldoon et Michael Weisberg (2011), qui se penchent sur la distribution des ressources cognitives entre différents projets scientifiques.

L'intérêt principal de ce modèle, en ce qui nous concerne, c'est qu'il illustre bien les avantages que peuvent avoir la capacité des MMA à intégrer des facteurs comme l'hétérogénéité des agents ou la rationalité limitée de ceux-ci.

Le problème abordé par Muldoon et Weisberg est le suivant. Étant donnés plusieurs projets avec différentes probabilités de succès, il existe une assignation optimale des scientifiques aux projets. Comment s'assurer que la distribution effective des scientifiques s'approche de cet optimum ?

La question avait été précédemment étudiée par Philip Kitcher (1990) et Michael Strevens (2003) en utilisant des modèles analytiques. Le but de Muldoon et Weisberg est de montrer que certaines des idéalizations adoptées par Kitcher et Strevens mènent à des résultats qualitativement différents de ceux qui sont obtenus avec un modèle plus réaliste du comportement des agents.

En d'autres mots, ils considèrent que le modèle de Kitcher et Strevens ne remplit que la première des deux exigences suivantes :

« [M]odels of cognitive labor must be simple enough for us to understand their dynamics, but faithful enough to reality that we can use them to analyze real scientific communities. » (Muldoon et Weisberg, 2011, p. 161)

Kitcher et Strevens ont construit leur modèle en utilisant ce que Muldoon et Weisberg appellent l'approche *contribution marginale/récompense* (CMR), dans laquelle on assigne à chaque projet une *fonction de succès*, qui représente la capacité du projet à utiliser de façon productive les ressources cognitives des scientifiques et à transformer ces ressources en possibilités de succès (Muldoon et Weisberg, 2011, p. 163). Les scientifiques travaillant à un projet couronné de succès sont récompensés, selon un schéma qui peut varier, et chaque scientifique choisit de travailler sur le projet qui maximise ses propres chances d'être récompensé. Le problème consiste à trouver un schéma de récompense qui produit la meilleure distribution de scientifiques par projet.

Muldoon et Weisberg affirment que le modèle de Kitcher et Strevens repose sur au moins deux hypothèses irréalistes :

1. L'hypothèse de la *distribution* : chaque scientifique connaît la distribution des ressources cognitives avant de déterminer sur quel projet travailler.
2. L'hypothèse de la *fonction de succès* : la fonction de succès de chaque projet, qui prend comme *input* une quantité de travail cognitif et qui a comme *output* une probabilité de succès, est connue par tous les scientifiques dans le modèle.

Ces deux hypothèses supposent une connaissance complète de la part des scientifiques. Pour rendre leur propre modèle plus réaliste, Muldoon et Weisberg introduisent des complexifications qui s'accordent avec les caractéristiques que nous avons vues dans la section 1.3 : les agents n'ont pas de connaissances parfaites (rationalité limitée) et ce n'est pas chaque agent qui sait ou croit la même chose (hétérogénéité).

Commençons par l'hypothèse de la distribution. Les scientifiques de Muldoon et Weisberg sont distribués sur une grille (un tore, en fait) de  $35 \times 35$ , où la distance représente la « distance de communication ». Les scientifiques ont un « rayon de vision » : ils « voient » les choix de projet des autres agents à l'intérieur de ce rayon. Pour reproduire le scénario d'information parfaite de Kitcher et Strevens, le rayon de vision doit être d'au moins  $\sqrt{578}$ , la distance à laquelle chaque scientifique voit tous les autres<sup>18</sup>.

Quand Muldoon et Weisberg utilisent cette distance et adoptent le schéma de récompense appelé « *Marge* » par Strevens (2003), où la récompense est divisée également entre tous les agents ayant travaillé au projet qui a réussi, ils obtiennent le même résultat que Strevens :

« As the number of agents was increased, an incentive was created for a minority of scientists to work on the harder project. When the number of agents was increased further, scientists allocated themselves to both projects, and eventually the number of

---

<sup>18</sup>Pour un agent se tenant exactement au milieu d'un tore de  $35 \times 35$  aplani, la distance euclidienne avec chaque coin =  $\sqrt{17^2 + 17^2} = \sqrt{578} \approx 24,042$ .

scientists working on the harder project overtook the number working on the easier project. » (2011, p. 166)

Quand Muldoon et Weisberg diminuent le rayon de vision, par contre—c'est-à-dire, quand ils assouplissent l'hypothèse de la distribution—l'allocation des agents est affectée : quand la vision tombe en dessous de 7, plus personne ne travaille sur le projet le plus difficile. D'un point de vue collectif, cet état de choses est loin d'être optimal.

En ce qui concerne l'hypothèse de la fonction de succès, Muldoon et Weisberg soutiennent qu'il est irréaliste que chaque scientifique connaisse la probabilité objective de succès de chaque projet. Ces probabilités devraient être subjectives et varier d'un scientifique à l'autre. Dans leur modèle, Muldoon et Weisberg utilisent une fonction de probabilité de succès empruntée à Kitcher<sup>19</sup>. Cette fonction a un paramètre de « difficulté », et c'est dans l'évaluation de cette difficulté qu'ils introduisent des différences entre agents.

Muldoon et Weisberg supposent que la distribution des croyances des agents quant à la difficulté d'un projet tombe sous une distribution normale dont la moyenne est la probabilité objective de succès du projet. Une variance de zéro dans cette distribution reproduit le scénario d'information parfaite de Kitcher et Strevens, ce que confirme une simulation avec deux projets (un plus facile, un plus difficile). Mais dès qu'une certaine variance est introduite—dès que certains agents se trompent sur la probabilité de succès—l'allocation qui en résulte est sous-optimale.

Une partie de l'attrait des modèles de la science (et des modèles de la société en général) est qu'un bon modèle peut être utilisé pour guider l'élaboration de politiques. C'est en partie ce que Strevens essaie de faire quand il compare le schéma *Marge* (où chaque scientifique ayant travaillé au projet est récompensé également) au schéma *Priorité* que l'on utilise dans la réalité (où le premier scientifique à compléter un projet avec succès obtient tout le crédit). Dans le modèle de Strevens, le schéma *Priorité* produit une meilleure distribution des ressources cognitives. Dans le modèle, plus réa-

---

<sup>19</sup>Il s'agit d'une fonction de croissance logistique :  $P = \frac{K}{1+e^{-rN}}$ , où  $K$  est la probabilité maximum de succès,  $N$  est le nombre de scientifiques qui travaillent au projet et  $r$  le niveau de difficulté du projet.

liste, de Muldoon et Weisberg, *Priorité* donne de moins bons résultats que *Marge*. La leçon à retenir est que les détails sont importants. Et, comme le montre le travail de Muldoon et Weisberg, les MMA sont une bonne façon de donner aux détails leur juste place.

### La science comme système cognitif distribué

La cognition peut être conçue comme allant au-delà du niveau individuel. Plusieurs philosophes (par exemple, Thagard 1993b ; Giere 2002 ; Cummins *et al.* 2004 ; Magnus 2007) soutiennent qu'on devrait considérer la science dans son ensemble comme un système cognitif distribué. Joshua Epstein, lui, va encore plus loin et généralise cette conclusion à l'ensemble de la société :

« The agent-based approach invites the interpretation of society as a distributed computational device, and in turn the interpretation of social dynamics as a type of computation. » (Epstein, 2006, p. 4)

Bruce Edmonds (2007) prend cette idée au sérieux. Il propose un modèle multi-agents de la science en tant que démonstrateur de théorèmes distribué (*distributed theorem prover*). Par opposition à ce que nous avons vu jusqu'à maintenant, la connaissance acquise par les agents de Edmonds se présente sous forme très structurée : elle est constituée de phrases vraies dans la logique des propositions du premier ordre. Le but des agents est de découvrir de nouveaux théorèmes en combinant des connaissances existantes (les prémisses) pour arriver à de nouvelles connaissances grâce à des inférences logiques.

Dans le modèle de Edmonds, les agents sont limités au *modus ponens* comme règle d'inférence :

$$\frac{p \rightarrow q \quad p}{\therefore q}$$

Chaque agent a une banque de connaissances—des phrases qui peuvent être utilisées comme prémisses pour faire de nouvelles inférences. Ces phrases proviennent des inférences précédentes faites par les agents, mais aussi d'une banque de connaissances

publique : un « journal » dans lequel les agents publient certains des théorèmes qu'ils trouvent. À chaque intervalle de temps, chaque agent :

1. Remplace certaines des phrases dans sa banque de connaissances privée par des phrases en provenance du journal.
2. Essaie de combiner des phrases de sa banque privée et ajoute le résultat des inférences réussies à celle-ci.
3. Soumet certaines des phrases non publiées de sa banque privée au journal.

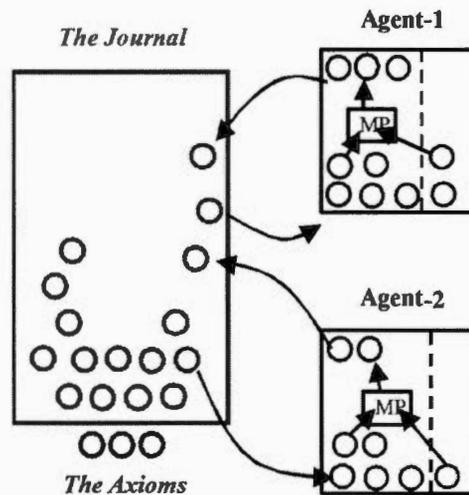
À la fin de chaque intervalle de temps, le journal ordonne les soumissions reçues en fonction de :

« the extent to which a formula had the effect of shortening formula when applied as the major premise in MP ; the shortness of the formula itself ; the past publishing success of the author ; and the fewness of the number of distinct variables in the formula. » (Edmonds, 2007, sect. 2.5).

Le succès de la communauté est évalué selon le nombre de théorèmes utiles qu'elle arrive à trouver en un temps donné. Et par « utiles » on entend *vraiment* utiles : le système est évalué à partir d'une liste de 110 théorèmes tirés de manuels de logique.

Un des résultats intéressants de Edmonds est que le nombre de théorèmes utiles est relativement indépendant du taux de publications du journal (le nombre de publications qu'il accepte à chaque étape). Un autre résultat intéressant est la disparité entre les agents : certains d'entre eux publient plus que d'autres. Et quoique pas tout à fait aussi près de la loi de Lotka que le sont les distributions obtenues par Gilbert ou Sun et Naveh, celle-ci est tout de même assez asymétrique.

Indépendamment des détails du modèle de Edmonds, celui-ci révèle que les MMA de la science peuvent être utilisés pour s'attaquer à de « vrais » problèmes scientifiques. Le calcul propositionnel avec une seule règle d'inférence, tel qu'utilisé par Edmonds, est quelque chose comme un « problème jouet », mais on pourrait aussi imaginer un système qui travaille sur des problèmes plus réalistes, plus difficiles. Ces problèmes



**Figure 2.5** Une représentation des agents et de leurs interactions avec la banque de connaissances publique. (Edmonds, 2007).

resteraient probablement des problèmes formels bien définis plutôt que les problèmes plus ouverts sur lesquels travaillent généralement les scientifiques. Le but ici n'est pas d'utiliser les MMA pour résoudre des problèmes scientifiques, mais plutôt de souligner la similarité entre les systèmes computationnels de résolution de problèmes<sup>20</sup> et le processus scientifique lui-même. Il est concevable que les méthodes d'optimisation développée dans un de ces domaines puissent être transférées à l'autre, et *vice versa*.

### La science dans un paysage épistémique

La plupart des modèles mentionnés ici supposent quelque chose comme un espace (possiblement à haute dimensionnalité) que les agents essaient d'explorer. Certaines positions dans cet espace sont considérées comme meilleures que d'autres et les agents

<sup>20</sup>Certains programmes d'intelligence artificielle sont capables de quelque chose qui s'approche du raisonnement scientifique. L'exemple classique est BACON (Langley *et al.*, 1981), qui a « redécouvert » la loi de réfraction de Snail, la loi de conservation de la quantité de mouvement, la loi de la chaleur spécifique de Black et la formule de Joule sur la conservation de l'énergie. Le programme PI (Thagard et Holyoak, 1985) arrive aussi à des résultats similaires, tout en s'apparentant un peu plus à une approche multi-agents (Thagard, 1993a, ch. 10).

essaient de trouver ces positions. Différents modèles adoptent différentes sémantiques pour cet espace. Dans le cas le plus commun, la proximité dans l'espace représente une forme de similarité conceptuelle, théorique ou pragmatique.

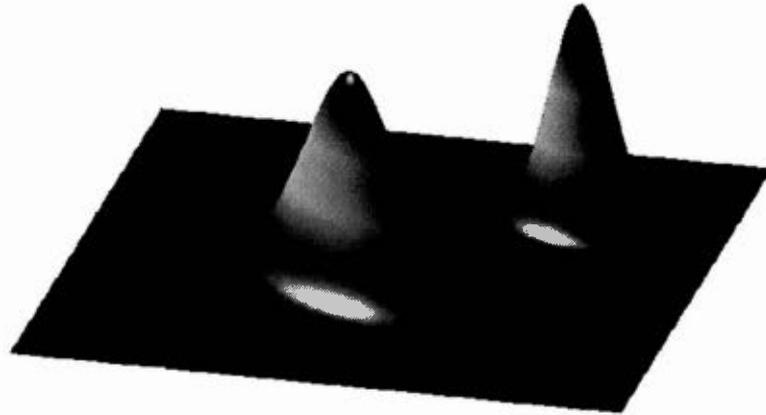
Pour qu'un espace soit interprété comme un « paysage », une de ses dimensions doit représenter la « hauteur ». Les kènes de Gilbert sont situés dans un espace à deux dimensions, mais aucune valeur n'est attachée à leur position : l'espace est plat. Dans le modèle de Sun et Naveh, les idées sont aussi situées dans un plan bi-dimensionnel, mais différentes idées ont différentes valeurs : certaines sont plus claires, mieux supportées empiriquement, etc. Si on fait une somme pondérée de ces différentes valeurs, on obtient une troisième dimension : la hauteur du paysage. On peut aussi, bien sûr, avoir un espace à  $n$  dimensions—l'important est que l'on cherche à maximiser une de ces dimensions<sup>21</sup>. Un facteur sur lequel Sun et Naveh n'ont pas insisté, par contre, c'est la façon dont la forme de cet espace affecte la dynamique scientifique. Pour aborder cette question, nous nous pencherons sur un autre modèle de Weisberg et Muldoon, qui élaborent sur le travail amorcé dans le modèle que nous avons présenté dans la section 2.3.

Les agents dans Muldoon et Weisberg (2011) étaient<sup>22</sup> situés dans l'espace, mais la distance dans cet espace représentait la distance de communication entre les chercheurs et non pas la valeur des projets sur lesquels ils travaillaient. Weisberg et Muldoon sont toujours intéressés par la division du travail cognitif, mais cette fois-ci, plutôt que de comparer deux projets avec différentes probabilités de succès, ils considèrent un vaste éventail d'approches possibles pour les scientifiques dans un champ d'études. Comme

---

<sup>21</sup>On pourrait aussi chercher à maximiser *plusieurs* dimensions. Les problèmes de ce type sont connus sous le nom de « problèmes d'optimisation multiobjectifs » (Steuer, 1986 ; Sawaragi *et al.*, 1985). Dans ces cas-là, on est à la recherche de ce qu'on appelle la « frontière de Pareto » : l'ensemble des positions dans l'espace qui ne sont pas « strictement dominées » par une autre. En réalité, la science est très certainement un problème d'optimisation multiobjectifs, mais à notre connaissance, personne ne s'est encore attaqué au problème sous cet angle.

<sup>22</sup>L'article de 2011 a été publié après celui de 2009, mais traite d'un modèle antérieur.



**Figure 2.6** Le paysage épistémique utilisé dans Weisberg et Muldoon (2009). L'axe vertical représente la « significativité épistémique » (*epistemic significance*).

on pourrait maintenant s'y attendre, ces approches sont représentées par la position des scientifiques dans un espace à deux dimensions<sup>23</sup>.

La troisième dimension est ce qu'ils appellent la « significativité épistémique » (*epistemic significance*) des approches. Le but des agents est de trouver les sommets de plus haute significativité dans le paysage. Le paysage utilisé par Weisberg et Muldoon a deux sommets, générés par des fonctions Gaussiennes (figure 2.6). La façon dont les agents se déplacent dans le paysage dépend de la stratégie (des règles de comportement) qu'ils adoptent. Le but des auteurs consiste à étudier la façon dont des popula-

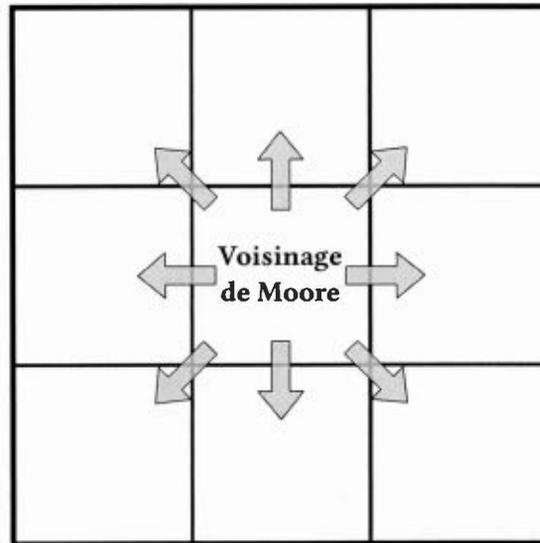
---

<sup>23</sup>Weisberg et Muldoon, comme Gilbert avant eux et Patrick Grim—que nous verrons dans la section 2.6—réservent l'exploration des paysages hautement dimensionnels à des recherches futures. L'avantage principal d'un paysage en trois dimensions, c'est bien sûr qu'il peut être facilement visualisé. Cela simplifie aussi la programmation et garde la charge computationnelle dans des limites raisonnables. Il serait toutefois intéressant de connaître l'impact d'une plus grande dimensionnalité sur certains modèles. La « malédiction de la dimensionnalité » est un problème pour certaines tâches d'optimisation et certains informaticiens tentent de trouver des algorithmes qui soient en mesure de traiter celle-ci (par exemple Powell, 2007), alors il est concevable que cela puisse faire une différence dans les résultats des simulations qui nous préoccupent.

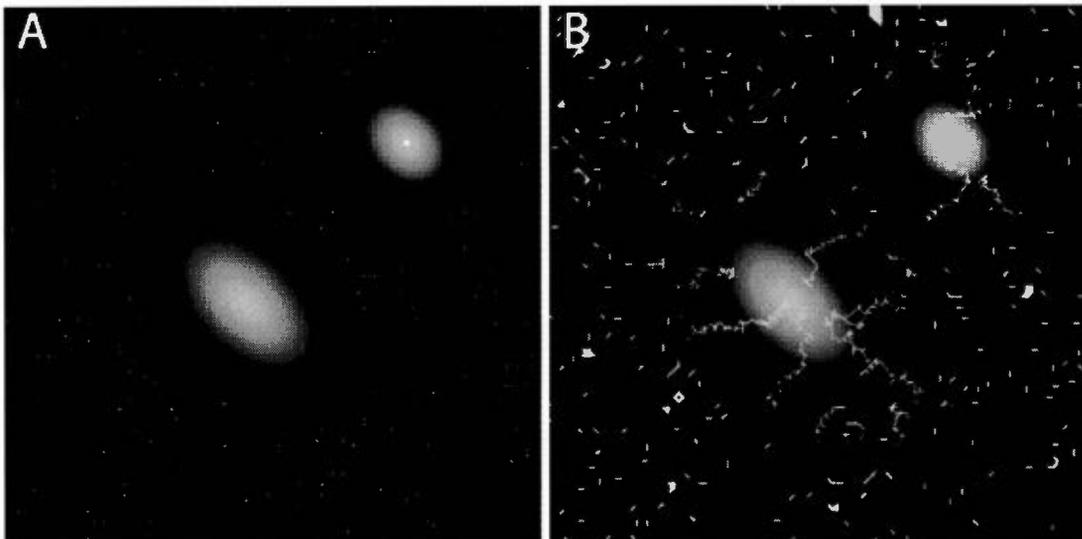
tions avec différentes répartitions de stratégies explorent l'espace. Ils examinent trois stratégies différentes, décrites ici en termes assez généraux :

- Les **Contrôles** sont, à toute fin pratique, des « escaladeurs de collines » (*hill climbers*) : ils adoptent une direction et avancent dans celle-ci tant qu'ils obtiennent de meilleurs résultats. S'ils obtiennent de moins bons résultats, ils reviennent sur leurs pas et adoptent, de façon aléatoire, une nouvelle direction. Ce sont des « contrôles » en ce sens qu'ils ne tiennent pas compte de ce que les autres agents font, et que Weisberg et Muldoon sont surtout intéressés par la dynamique introduite par les interactions entre agents.
- Les **Suiveurs** (*Followers*) commencent par examiner toutes les cases dans leur voisinage de Moore (figure 2.7 à la page suivante) qui ont *déjà* été visitées et qui ont une plus grande significativité que leur approche actuelle. S'il existe de telles cases, ils se déplacent vers la meilleure de celle-ci (choisissant au hasard en cas d'égalité). S'il n'y en a pas, ils choisissent au hasard une case non visitée. (En d'autres mots : ces agents vont seulement innover s'ils n'ont pas le choix.) Finalement, si toutes les cases de leur voisinage ont été visitées et qu'il n'y en a pas de meilleures que leur case actuelle, ils s'arrêtent.
- Les **Mavericks** ressemblent aux contrôles en ce que, lorsque leur position courante est pire que leur position précédente, ils reviennent en arrière et changent de direction. Mais si leur position courante est égale ou meilleure à la précédente, ils choisissent au hasard une case *non visitée* dans leur voisinage. Ce n'est que lorsqu'il n'y a pas de case non visitée qu'ils agissent en Suiveurs et choisissent la meilleure approche connue dans leur voisinage.

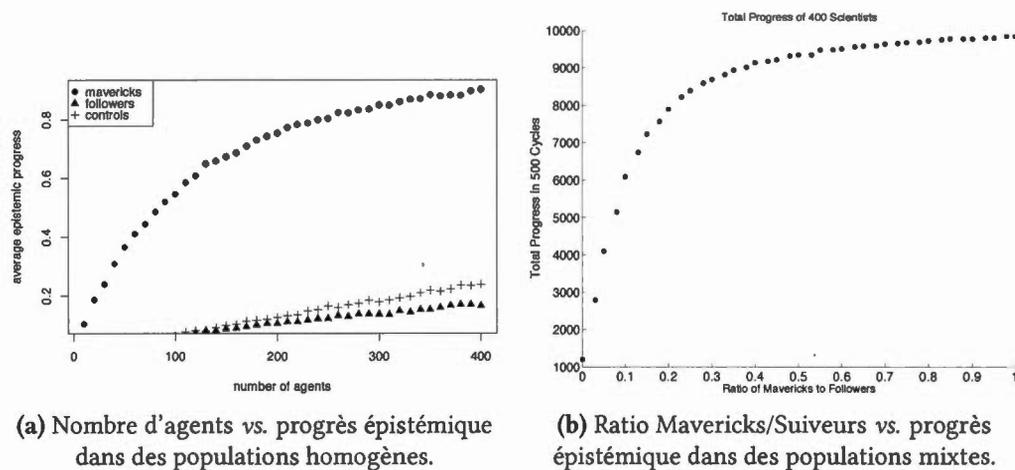
Laissés à eux-mêmes, les Contrôles ne sont pas très efficaces. Ils trouvent éventuellement les sommets, mais puisqu'ils ne peuvent pas apprendre l'un de l'autre, ils ont besoin de plusieurs étapes avant d'y arriver. Les Suiveurs, lorsqu'ils sont seuls, font encore pire : ils restent assez rapidement coincés dans les régions de faible significativité épistémique. À moins qu'ils soient chanceux, ils ne font que se suivre l'un l'autre et « tourner en rond ». Les Mavericks, par contre, sont efficaces : ils trouvent toujours les sommets, et les trouvent plus rapidement que les Contrôles.



**Figure 2.7** Un voisinage de Moore est constitué des neuf cases entourant une position particulière sur la grille.



**Figure 2.8** Position initiale (A) et subséquente (B) des Suiveurs dans le paysage épistémique pendant une simulation. Les traces indiquent les chemins empruntés par les agents.



**Figure 2.9** Progrès épistémique pour des populations homogènes (a) et mixtes (b) dans les simulations de Weisberg et Muldoon (2009).

C'est lorsqu'on commence à examiner des populations mixtes que les choses deviennent plus intéressantes. Même l'ajout d'un seul Maverick à une population de Suiveurs fait une différence significative : le Maverick aide les Suiveurs à débloquer. Plus nombreux sont les Mavericks dans une population, plus la performance du système s'améliore, jusqu'à ce qu'on atteigne 100% de Mavericks dans la population, ce qui est optimal. Dans le monde réel, par contre, un certain équilibre entre Suiveurs et Mavericks est probablement nécessaire. Les Suiveurs semblent bien adaptés à ce que Khun (1962) appelle le *puzzle solving* : trouver des solutions à des problèmes spécifiques avec des méthodes bien définies. Être un Maverick est plus risqué pour un individu : s'écarter des chemins bien balisés, et possiblement échouer, peut être dévastateur pour une carrière scientifique. Quoi qu'il en soit, le travail de Weisberg et Muldoon montre que la façon dont les chercheurs prennent en compte le travail des autres agents influence le succès global de la communauté.

### Les réseaux épistémiques

Kevin Zollman est un des premiers à avoir entrepris de simuler l'effet de la structure sociale de la communauté scientifique sur sa performance épistémique. La question est importante, selon lui, puisque :

« Once one fully articulates a theory of individual epistemic rationality, it is still an open question what the optimal community structure is for these agents—the individualistic question is only part of the answer. » (Zollman, 2007)

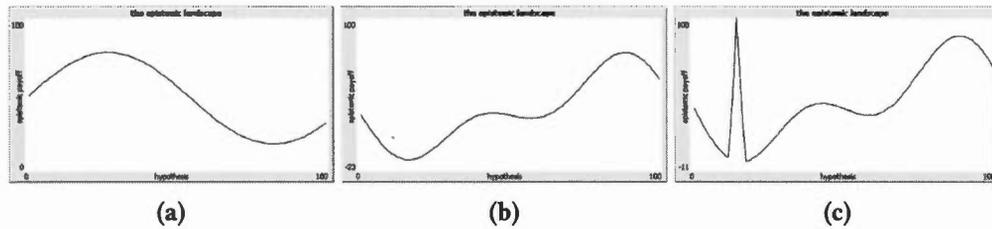
Pour tenter d'identifier la « structure communautaire optimale », Zollman simule le comportement de réseaux de scientifiques tentant de choisir entre deux hypothèses distinctes à partir d'information limitée, où les scientifiques individuels ne peuvent communiquer leurs résultats qu'à leurs voisins immédiats. Nous n'allons pas entrer dans le détail de ses expériences, mais ses conclusions montrent qu'un réseau plus connecté tend à converger plus rapidement vers une hypothèse, mais qu'il est probable que celle-ci soit la *mauvaise* hypothèse : il y a un compromis important entre rapidité et justesse.

En partie inspiré par Zollman, Patrick Grim a aussi travaillé sur la façon dont la structure sociale de la science affecte ses résultats. Lui aussi soutient que cette attention à la communauté est essentielle :

« How does an individual figure out the structure of the world? The truth is that no individual does. It is cultures and communities that plumb the structure of reality; individuals figure out the structure of the world only as they participate in the epistemic networks in which they are embedded. » (Grim, 2009)

La différence principale avec Zollman est que les agents de Grim, tout comme ceux de Weisberg et Muldoon, sont à la recherche de la meilleure hypothèse possible dans un paysage épistémique. Mais plutôt que de voir les résultats obtenus par les agents qui sont près d'eux dans ce paysage (comme les agents de Weisberg et Muldoon), les agents de Grim voient les résultats de ceux avec qui ils sont connectés dans un réseau social (comme les agents de Zollman). À chaque intervalle de temps, un agent a une probabilité de 50% de modifier son hypothèse courante en la déplaçant jusqu'à mi-chemin du lieu où se trouve la meilleure hypothèse entretenue actuellement par un des agents auxquels il est connecté.

Cette approche permet à Grim de tester la performance de différentes structures de réseaux : réseaux en anneau, réseaux en roue, petit monde, réseau aléatoire et réseau pleinement connecté. Résultat : ce sont les réseaux en anneau qui performant le mieux et les réseaux pleinement connectés qui performant le moins bien. En utilisant des



**Figure 2.10** Différentes formes de paysages épistémiques en deux dimensions utilisées dans les simulations de Grim. Remarquez la façon dont le paysage (c) est trompeur pour les agents, ce qui rend difficile la découverte du sommet optimal.

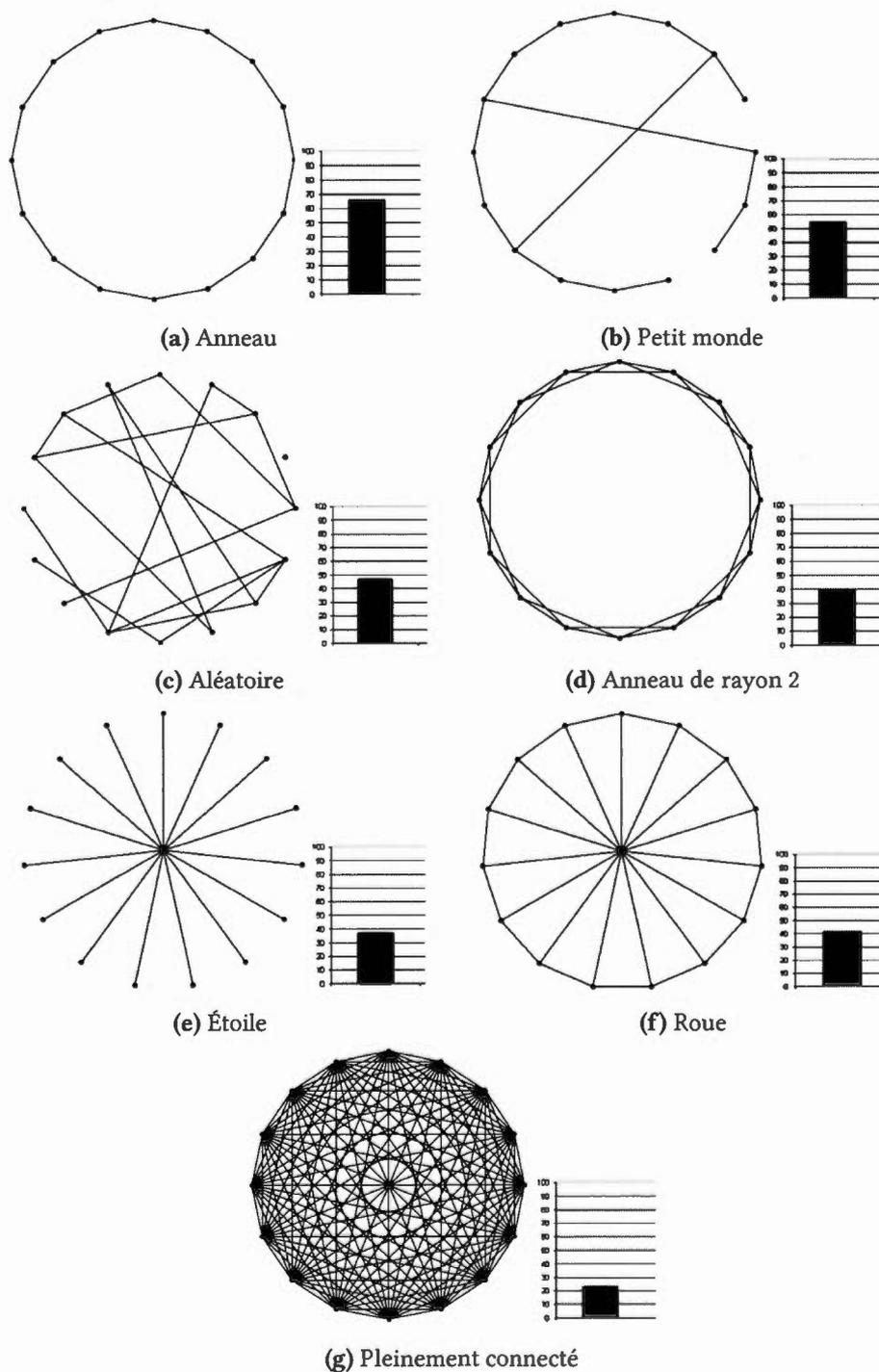
réseaux aléatoires avec différents niveaux de connexion, Grim montre qu'en général, au-delà d'un seuil assez bas, ajouter des liens à un réseau en diminue la performance. Ce résultat est compatible avec ceux de Zollman.

Suite à l'analyse de ses résultats, Grim émet l'hypothèse que pour certains types de problèmes, le réseau scientifique du XVII<sup>e</sup> siècle, où les communications entre chercheurs étaient plus rares et limitées, pourrait avoir été mieux adapté que le réseau pleinement connecté vingt-quatre heures sur vingt-quatre du XXI<sup>e</sup> siècle.

Ce qui se produit avec les réseaux pleinement connectés est similaire à ce qui se produit avec les Suiveurs de Weisberg et Muldoon : les chercheurs restent confinés à des régions déjà explorées du paysage épistémique. Cela rend la communauté susceptible de rester coincée sur des sommets de valeur épistémique non optimale (comme celui qui se trouve à la droite de la figure 2.10c) parce que tous les agents convergent sur la meilleure hypothèse initiale et que personne ne continue à explorer lorsque celle-ci est atteinte. Ce qu'il faudrait à la simulation de Grim, c'est quelques Mavericks : des chercheurs qui évitent délibérément de dupliquer les hypothèses de leurs pairs.

## Conclusion

Après nous être penchés sur cette liste variée (bien que non exhaustive) de modèles multi-agents de la science, nous terminerons ce chapitre en mentionnant brièvement trois aspects du système scientifique sur lesquels les modèles abordés ci-dessus ne se



**Figure 2.11** Pourcentage des simulations où la communauté trouve l'hypothèse optimale pour différentes structures de réseau social sur le paysage 2.10c. En général, le moins connecté le réseau, meilleures sont ses performances.

penchent pas suffisamment, puis en mentionnant tout de même certaines des leçons que nous pouvons en tirer.

Comme nous le verrons en détail dans le chapitre 3, David Hull soutient que la science est un système évolutionnaire. Quoique la simulation de Gilbert fasse intervenir une petite composante évolutionnaire, en ce que chaque « kène » hérite d'un kène « parent », il manque un élément important : les articles qui sont choisis pour se reproduire ne le sont pas en fonction du contenu de leur kène—cette reproduction est entièrement aléatoire. Il suffirait pourtant que la reproduction soit différentielle, basée sur la position du kène dans un paysage épistémique similaire à celui qui est utilisé par Weisberg et Muldoon ou Grim, pour que l'on obtienne une forme d'*adaptation* : que les kènes (ou les articles, idées, théories, propositions, etc.) ayant une position plus élevée dans le paysage épistémique génèrent plus de descendants et que les « meilleurs » kènes se retrouvent ainsi en surnombre. À notre connaissance, cette idée n'a pas encore été pleinement explorée à l'aide de modèles multi-agents<sup>24</sup>.

Une autre composante importante du système scientifique, c'est le processus d'enseignement et d'apprentissage : un professeur transmet ses idées à ses étudiants qui, s'ils ont une carrière réussie, vont enseigner à leur tour ces idées à une nouvelle génération d'étudiants, et ainsi de suite. On peut construire une généalogie des chercheurs tout comme on peut construire une généalogie familiale. Les modèles multi-agents peuvent reproduire ce processus et nous permettre de visualiser cet arbre et d'en analyser la structure. Dans le cadre de la théorie de Hull, un tel modèle nous permettrait d'explorer des concepts comme celui de « valeur sélective conceptuelle inclusive », basé sur le concept biologique de valeur sélective inclusive (Hamilton, 1964) : tout comme un comportement altruiste envers nos proches parents biologiques favorise la survie de nos gènes, un comportement altruiste envers nos étudiants favorise la transmission de nos idées. Nous reviendrons à ce concept dans le chapitre 3 et montrerons dans le chapitre 5 comment celui-ci peut-être implémenté dans le cadre d'un MMA.

---

<sup>24</sup>Notons au passage que dans le domaine de la vie artificielle (Langton, 1997) exploite à fond la possibilité de simuler des systèmes évolutionnaires. Les algorithmes évolutionnaires constituent aussi un exemple important de reproduction computationnelle des principes généraux de l'évolution (Holland, 1975 ; Luke, 2009).

Un dernier élément important laissé de côté par les modèles ci-dessus, c'est la *dynamique* des réseaux de recherche. Nous avons vu avec Grim (2009) et Zollman (2007) que la *structure* de ces réseaux a un impact significatif sur les processus épistémiques sociaux, mais comment ces réseaux se forment-ils ? Comment changent-ils au fil du temps ? Est-ce qu'avoir beaucoup de crédit permet à un chercheur d'attirer de nombreux collaborateurs, ce qui lui permettrait d'obtenir encore plus de crédit ? Peut-on simuler la formation de groupes de recherche distincts ? Qu'en est-il de la rivalité entre ces groupes de recherche ? Toutes ces questions, et plusieurs autres, sont susceptibles d'être abordées de façon fructueuse à l'aide de modèles multi-agents.

La plupart des modèles que nous avons vus dans ce chapitre ont des caractéristiques différentes, mais beaucoup se recoupent. Nous les avons comparés entre eux d'un point de vue qualitatif, d'assez haut niveau. Nous ne doutons pas que des comparaisons numériques plus systématiques (comme le suggère Axtell *et al.* (1996)) seraient bénéfiques, mais dans son état actuel, le domaine des MMA de la science est encore trop fragmenté pour rendre cette entreprise viable. Nous avons aussi soutenu ailleurs (Payette, 2011) que les MMA de la science gagneraient à adopter une vision plus intégrée du système scientifique. La modélisation multi-agents en sciences sociales est encore un domaine relativement immature, et c'est encore plus vrai pour les MMA de la science. Beaucoup de chercheurs se penchent sur des questions méthodologiques (Axtell *et al.*, 1996 ; Cioffi-Revilla, 2009 ; Epstein et Axtell, 1996 ; Bonabeau, 2002 ; Gilbert et Troitzsch, 2005 ; Epstein, 2008), mais nous n'avons pas encore un cadre bien établi qui soit adopté par tous.

Dans leur état actuel, l'impact normatif des MMA de la science reste limité. On peut toutefois, déjà, reconnaître quelques leçons importantes suggérées par les modèles que nous avons vus.

Dans tous les systèmes de recherche de connaissance, il y a un compromis entre exploitation et exploration : un équilibre délicat entre le raffinement des connaissances acquises et la recherche de connaissances complètement nouvelles. Comme Weisberg et Muldoon (2009) l'ont montré, une population de scientifiques a besoin de quelques "Mavericks". De façon plus générale, il s'agit de répartir adéquatement les ressources cognitives du système : nous voulons que les chercheurs soient distribués entre diffé-

rents projets d'une façon qui profite à la communauté dans son ensemble. La façon dont sont récompensés les chercheurs individuels a une influence majeure sur cette répartition, mais comme l'ont montré Muldoon et Weisberg (2011), ce système n'est pas toujours aussi efficace qu'on pourrait le penser. La traditionnelle règle de priorité, par exemple, devrait peut-être être réévaluée.

Le concept de paysage épistémique a aussi une portée normative importante : différentes normes risquent d'avoir différents effets dans différents paysages épistémiques. La forme du paysage devrait donc, théoriquement, être prise en compte au moment d'essayer d'influencer le système scientifique. À l'heure actuelle, par contre, nous ne possédons pas de méthode bien développée qui nous permettrait de déterminer quelle est la forme du paysage épistémique pour un domaine scientifique en particulier. Ce n'est pas une question que nous tenterons de résoudre dans cette thèse mais c'est, selon nous, l'une des questions importantes auxquelles sera confrontée l'étude des systèmes scientifiques dans les années à venir.

Autre leçon cruciale : les scientifiques font partie de communautés, et la structure de ces communautés est importante. Les résultats que nous avons vus jusqu'à présent (Zollman, 2007 ; Grim, 2009) suggèrent qu'une communication trop grande entre scientifiques pourrait mener à l'atteinte d'un consensus trop rapide entre chercheurs, emmenant la communauté à négliger l'exploration de théories alternatives. Si c'est le cas, la pression exercée sur les scientifiques pour qu'ils publient rapidement leurs résultats pourrait contribuer à exacerber ce phénomène.

L'épistémologie classique s'est surtout penchée sur la connaissance individuelle et sur la façon dont celle-ci pouvait être justifiée, négligeant parfois le rôle de la communauté scientifique dans ce processus de justification. L'aspect social de la science en fait un objet particulièrement approprié pour les modèles multi-agents. Dans le mots de Bruce Edmonds :

« I see no reason why messy and intricate social processes should result in knowledge that is less reliable than the process of reason implemented by equally messy and intricate processes between neurons, but it might be so. The only way to examine and tease out the various and apparently contradictory effects of these social processes will be to simulate them letting the results emerge from the interaction. In other words, wi-

thout a prior commitment to such processes regarding the soundness of their ultimate outcomes. » (Edmonds, 2011, § 19.3)

Les quelques leçons abordées ici, même si elles ne constituent qu'un portrait fragmenté de l'impact potentiel des MMA pour l'étude de la science, laissent entrevoir un vaste éventail d'applications possibles, autant en philosophie des sciences que dans les disciplines ayant une approche plus sociologique.

Nous tenterons d'actualiser certaines de ces possibilités dans le chapitre 5, mais nous allons d'abord nous pencher sur la théorie qui nous servira de base pour ce faire : celle de David Hull.

## CHAPITRE III

### LE PROCESSUS SCIENTIFIQUE SELON DAVID HULL

David Hull propose une conception évolutionniste du processus scientifique. L'objectif principal de ce chapitre n'est ni de critiquer, ni de défendre la théorie de Hull, mais plutôt d'expliquer les aspects de celle-ci qui serviront de base au modèle multi-agent présenté au chapitre 5.

Nous commencerons tout de même par situer la théorie de Hull dans le contexte philosophique où elle a été développée. Hull était préoccupé par l'apparent conflit entre internalisme et externalisme en philosophie des sciences, et nous aborderons brièvement cette distinction. Nous verrons aussi comment Hull se situe par rapport à l'épistémologie évolutionniste et comment il construit sa propre théorie de la sélection en science. Celle-ci repose sur une analyse générale de la sélection faisant intervenir les concepts de réplicateur et d'interacteur, et nous verrons ceux-ci en détail. Nous nous pencherons ensuite sur les réponses de Hull à quelques objections possibles envers sa théorie, puis nous conclurons notre survol de sa théorie en examinant trois des mécanismes sociaux à l'œuvre en science selon lui : le compromis entre crédit et support, la valeur conceptuelle sélective inclusive et la structure démique de la science.

## Contexte philosophique

Au moment où Hull est plongé dans la recherche qui mènera à son œuvre maîtresse, *Science as a Process*, la philosophie des sciences, elle, est plongée dans un débat au sujet des facteurs qui doivent être pris en compte pour expliquer le changement scientifique. D'un côté, une philosophie des sciences plus traditionnelle qui se cantonne à considérer les facteurs proprement épistémiques : raisonnement, argumentation et données probantes (*evidence*). C'est ce que Hull appelle la « trinité internaliste », et dont il considère Larry Laudan (1977 ; 1981a ; 1981b ; 1984 ; 1986) comme un représentant exemplaire. D'un autre côté, une approche qui, dans la lignée de Kuhn (1962) et sous l'influence de la sociologie, met de l'avant des facteurs externes à la science telle qu'on la conçoit traditionnellement : culture, politique et, surtout, intérêts individuels. Cette école qui s'incarne entre autres dans ce qu'on a appelé le *strong programme* (Bloor, 1976 ; Barnes, 1977, 1981 ; Collins, 1981a,b,c, 1982) tend parfois vers le relativisme, refusant de considérer la vérité ou la fausseté d'une théorie dans l'explication du sort réservé à celle-ci.

Aucune de ces deux approches n'est entièrement satisfaisante pour Hull. Il est crucial, selon lui, de pouvoir rendre compte du progrès scientifique : les théories scientifiques actuelles, sans offrir le genre de certitude qu'ont cherchée les philosophes au fil de l'histoire, sont tout de même supérieures aux théories qu'elles ont remplacées. Ayant toutefois observé de près le comportement des scientifiques au fil de sa carrière (celui des biologistes, tout particulièrement, lorsqu'il était président de la *Society for Systemic Zoology*) Hull est conscient que l'image du scientifique désintéressé, en quête de la vérité pure, est erronée. Il cherche à intégrer les facteurs qui préoccupent ces deux écoles dans un cadre plus large. Ce cadre sera un cadre *évolutionniste* qui sera non seulement compatible avec l'idée que les facteurs sociaux jouent un rôle crucial dans le développement de la science, mais qui permettra également d'expliquer comment certains de ces facteurs contribuent au progrès de la science.

## Internalisme et externalisme

La distinction entre internalisme et externalisme intervient dans plusieurs branches de la philosophie. La forme générale du problème consiste à se demander si un phénomène peut être expliqué (ou si une propriété peut être définie) en ne faisant intervenir que des éléments internes à un système. À cette question, un internaliste répond « oui » ; un externaliste répond « non ».

En philosophie morale, par exemple, on sera internaliste ou externaliste selon que l'on soutient que le comportement d'un agent peut être dicté seulement par des raisons internes à l'agent (motivations, désirs, buts) ou que l'on admet également des raisons externes à l'agent (loi morale, pressions sociales, etc.) (voir Williams, 1981). En sémantique, on sera internaliste ou externaliste selon que l'on admet que des éléments externes à l'individu interviennent dans le sens d'un énoncé (Putnam, 1973, 1975). En théorie de la connaissance, une question importante est de savoir si les bases d'une croyance justifiée sont internes à l'agent qui entretient cette croyance et si ce dernier a accès à ces bases (voir Pappas, 2012)<sup>25</sup>. Quoiqu'il soit possible d'établir des liens entre ces différentes problématiques, l'adoption d'une position internaliste ou externaliste dans un domaine philosophique n'implique pas nécessairement l'adoption de la position correspondante dans les autres domaines.

C'est souvent la peau du crâne qui sert de frontière entre l'interne et l'externe. En philosophie des sciences, par contre, lorsqu'on cherche à rendre compte de la production des connaissances, cette frontière est plutôt conçue comme une délimitation entre l'activité scientifique et le reste de l'activité humaine. L'endroit où la frontière est tracée peut varier, mais on retrouve traditionnellement à l'intérieur de celle-ci ce que Hull appelle « la trinité internaliste » (1988c, p. 2) : raisonnement, argumentation et données probantes (*evidence*). Du point de vue internaliste, donc, des facteurs comme

---

<sup>25</sup>Cette question concerne spécifiquement la théorie de la connaissance et ne doit pas être confondue avec celle qui nous préoccupera plus loin, qui concerne la philosophie des sciences, de même que la sociologie et l'histoire de celles-ci.

l'appartenance à une classe sociale, la religion, le sexe, les intérêts personnels, etc., ne devraient pas être invoqués en philosophie des sciences<sup>26</sup>.

L'internalisme est souvent associé au positivisme logique de la première moitié du vingtième siècle et à certains de ses critiques, notamment Popper. Il reste considéré comme la position orthodoxe en philosophie des sciences<sup>27</sup>. Les facteurs internes s'articulent autour de deux principes : la confrontation entre les théories scientifiques et le monde et l'utilisation de la raison pour tirer les conséquences de ces théories. Il en résulte généralement une conception progressiste du changement scientifique.

Du point de vue externaliste, l'opposition entre internalisme et externalisme est elle-même un construit de l'école internaliste. Ce que soutiennent les externalistes, ce n'est pas tant qu'il faut tenir compte des facteurs externes, mais plutôt que la frontière elle-même est illusoire et qu'il est impossible d'isoler (ne serait-ce que conceptuellement) la science du reste de l'activité humaine. Au moment où Hull écrit, cette position est fortement associée au *strong programme* des écoles d'Édimbourg et de Bath—p. ex., Bloor (1976) et Barnes (1977, 1981) à Édimbourg et Collins (1981a,b,c, 1982) à Bath.

S'il ne s'agissait que d'un débat sur les facteurs qui doivent être pris en compte en plus des facteurs internes à la science, les enjeux du débat seraient relativement mineurs. Ils prennent toutefois une importance indéniable lorsque certains externalistes affirment

---

<sup>26</sup>Notons ici que la distinction entre facteurs internes et externes à la science diffère de la distinction entre contexte de découverte et contexte de justification—du moins chez les empiristes logiques. Cette dernière distinction, popularisée par Reichenbach, concerne plutôt la différence entre « the system of logical interconnections of thought and the actual way in which thinking processes are performed » (1938, p. 1). Pour Reichenbach, les facteurs externes sont le domaine de la sociologie et le contexte de découverte est le domaine de la psychologie. L'épistémologie doit se préoccuper des facteurs internes en contexte de justification.

<sup>27</sup>Hull invoque à maintes reprises Larry Laudan (1977 ; 1981a ; 1981b ; 1984 ; 1986) pour illustrer la position internaliste.

que la raison, les arguments et les données ne jouent *aucun* rôle (ou encore un rôle négligeable) dans le changement scientifique<sup>28</sup>.

Si on accepte de minimiser (ou même de nier) le rôle des facteurs internes, il devient difficile de soutenir une vision progressiste de la science, du moins si cette progression est conçue en termes d'adéquation croissante entre les théories et le monde. Si on rejette la confrontation entre les théories et la réalité et la comparaison rationnelle entre les théories comme moteurs du changement scientifique, il devient difficile d'affirmer que les théories actuelles sont meilleures que les théories qui les ont précédées. Certains externalistes vont même jusqu'à rejeter toute ontologie réaliste et adopter plutôt différentes formes de relativisme et de constructivisme (voir Bunge, 1991, pour un survol de ces positions).

### Épistémologie évolutionniste

Entre internalisme naïf et externalisme radical, Hull cherche une voie alternative. Il est prêt à admettre le rôle des facteurs sociaux dans le développement de la science, mais pas à rejeter les facteurs internes. Ce qu'il cherche à sauvegarder d'abord et avant tout, c'est une conception progressiste de la science. Donald Campbell, figure de proue de l'épistémologie évolutionniste, décrit ainsi la position dans laquelle se retrouve Hull :

« For some scholars, the success of science is both obvious and not at all puzzling. They have no need for naturalistic-epistemological theories as to how this success might be possible. For other scholars (especially the social constructivist, ontologically relativist, strong programme sociologists and historians of science, plus some philosophers) the fact that there is no apodictic proof of improved competence of reference in temporal sequences of conceptual change in science also removes any puzzle. Since there is no completely demonstrable fit between scientific beliefs and any supposed independent referents of these beliefs, there is no puzzle needing explanation. These two groups do not need selection theory. David Hull is in neither camp. » (1988, p. 172)

---

<sup>28</sup>Ces positions se réclament souvent (tout en allant beaucoup plus loin) des critiques de l'empirisme logique, qui ont soutenu, entre autres, que la logique inductiviste menait nécessairement à une certaine forme de subjectivisme (Popper, 1934), que la distinction entre énoncés analytiques et synthétiques était intenable et qu'il fallait donc renoncer à une épistémologie fondationnaliste (Quine, 1951, 1969) et que les révolutions scientifiques produisent des paradigmes incommensurables (Kuhn, 1962).

Ce que Campbell laisse entendre ici, c'est que la « théorie de la sélection » (*selection theory*<sup>29</sup>) est nécessaire pour expliquer le progrès scientifique dans un cadre qui tient compte des facteurs externes, puisque nous n'avons aucune autre option qui serait « nearly as plausible as selection theory in providing a conjectural solution to that puzzle » (1988, p. 172)<sup>30</sup>. La théorie de la sélection offre un cadre plus large, dans lequel les deux types de facteurs peuvent à la fois être opérants sans que la notion de progrès soit mise en péril : si les facteurs internes jouent un rôle suffisamment important dans le processus de sélection des idées scientifiques, on assistera, éventuellement, à une augmentation de l'adéquation entre les théories et le monde.

De surcroît, pour Campbell, l'adoption d'une épistémologie évolutionniste est inévitable en dernière analyse, puisque toute nouvelle acquisition de connaissance doit reposer sur un processus sélectionniste :

« In going beyond what is already known, one cannot but go blindly. If one can go wisely, this indicates already achieved wisdom of some general sort. » (1974, p. 422)

On peut bien tirer, par déduction, toutes les conséquences possibles de nos connaissances existantes (ce que Campbell appelle « going wisely »), mais on arrive éventuellement à un point où, pour progresser, il faut faire appel à l'induction et générer de nouvelles propositions qui n'offrent pas de certitude épistémique (« going blindly »). C'est le vieux problème de l'induction qui nous pousse vers une forme de sélectionnisme : il ne peut y avoir d'avancées dans la connaissance sans que soient générées des variations incertaines sur ce qui est déjà connu et que ces variations soient confrontées au monde de façon à ce que soient éliminées celles qui sont inadéquates.

C'est de cette façon que fonctionnent les formes d'apprentissage les plus simples, chez les organismes les plus primitifs. Comme, par exemple, lorsque différents comporte-

---

<sup>29</sup>Campbell parle de *selection theory* plutôt que de *evolutionary theory* pour marquer que l'approche de Hull (ni celle de Campbell, d'ailleurs) ne procède pas par analogie avec l'évolution biologique, mais s'appuie plutôt sur un cadre plus général qui regroupe à la fois l'évolution biologique et culturelle, de même que des processus comme le fonctionnement du système immunitaire.

<sup>30</sup>Il ne faut pas confondre la notion de progrès évoquée ici avec celle qui serait impliquée par une conception orthogénétique de l'évolution, qui prétendrait que celle-ci tend de façon unilinéaire vers un but prédéterminé. Le progrès dont parle Campbell consiste simplement en une augmentation de l'adaptation d'un système à son environnement.

ments sont tentés et que ceux qui génèrent du plaisir sont répétés alors que ceux qui génèrent de la douleur ne le sont pas. Dans ce cas-ci, le mécanisme de plaisir/douleur constitue ce que Campbell décrit comme une forme de « sagesse générale préétablie », qui est aussi le produit d'un processus de sélection préalable : la sélection naturelle, biologique.

Une des tâches de l'épistémologie évolutionniste est de décrire cette hiérarchie de mécanismes, en partant de la sélection naturelle et des formes primitives d'apprentissage pour aller jusqu'à la science institutionnalisée. Une autre tâche, plus difficile et plus controversée, est de montrer comment ces mécanismes nous assurent (ou à tout le moins nous permettent) de connaître le monde.

C'est sur le niveau le plus élevé de cette hiérarchie—la science institutionnalisée—que se penche Hull, tentant de décrire le plus finement possible les mécanismes de sélection qui y opèrent. Il doit donc à la fois rendre compte de l'algorithme général de sélection par lequel les idées évoluent et expliquer comment celui-ci est implémenté dans la science telle qu'elle se pratique en Occident depuis 350 ans<sup>31</sup>, avec tout ce que cela suppose de particularités et d'idiosyncrasies. Cette tâche est analogue à celle du biologiste se penchant sur une espèce donnée (ou même sur la vie en général) qui, tout comme la science, est selon Hull un individu spatiotemporellement délimité.

Tout en étant réfractaire à l'attribution d'une valeur normative à sa théorie, Hull soutient que ces mécanismes constituent une bonne explication du progrès scientifique. Et il est en mesure de fournir cette explication sans que se présente de conflit entre les facteurs internes et externes à la science. Cette distinction s'efface : tous ces facteurs sont considérés chez Hull comme faisant partie de l'environnement de sélection des idées scientifiques. Bien sûr, les facteurs internes (particulièrement l'expérimentation) jouent un rôle primordial dans la sélection des idées scientifiques. Ce sont eux qui, au bout du compte, garantissent le caractère progressif de la science. Cela dit, l'analyse de Hull ne cherche pas à caractériser les facteurs externes comme « irrationnels » du point de vue de l'entreprise scientifique. Certains de ces derniers ont aussi un rôle

---

<sup>31</sup>Hull situe le début de cette période en 1665, soit l'année de la publication du premier numéro des *Philosophical Transactions* de la *Royal Society of London*.

positif à jouer dans l'évolution de la science. Hull s'intéresse tout particulièrement au fait que les scientifiques adoptent souvent, dans le but de favoriser leurs carrières individuelles, des comportements qui vont à l'encontre de l'idéal selon lequel ils devraient œuvrer « pour le bien de la science », de façon objective et désintéressée.

Le choix d'une gravure représentant des guerriers engagés dans un violent combat (figure 3.1 à la page suivante) pour illustrer la page couverture de *Science as a Process*, n'est pas innocent. Les rivalités personnelles et professionnelles (et les comportements qui s'ensuivent) ont une influence déterminante sur le processus par lequel les variantes conceptuelles et théoriques sont adoptées ou rejetées par la communauté scientifique. Hull soutient que non seulement la science progresse *malgré* ces luttes intestines, mais aussi qu'elle progresse en partie *grâce* à elles. Cette attention portée à l'intérêt personnel des scientifiques place Hull dans une perspective où la rationalité individuelle et la « raison scientifique » universelle ne coïncident pas nécessairement, mais ne sont pas incompatibles non plus<sup>32</sup>.

### Une approche évolutionniste de la science

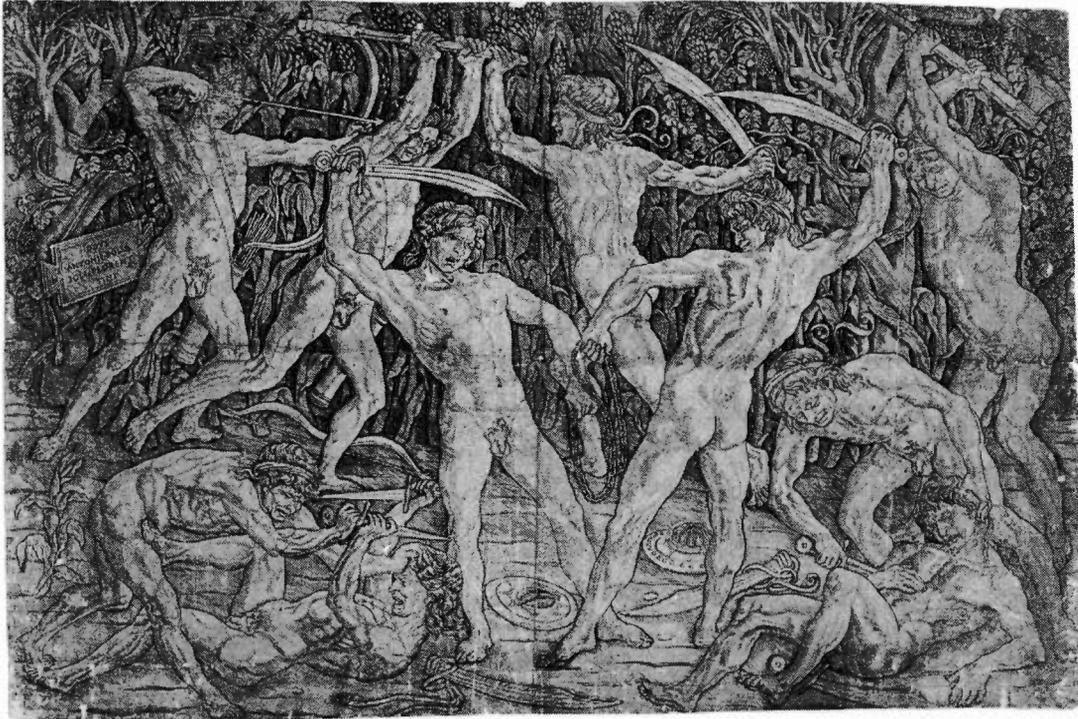
Certains ouvrages classiques de philosophie des sciences mentionnent au passage l'analogie entre l'évolution des théories scientifiques et l'évolution biologique (Popper, 1934 ; Kuhn, 1962), et d'autres avant Hull ont aussi développé la question explicitement (p. ex., Toulmin 1972 ; Campbell 1974 ; voir Bradie 1986 pour une recension exhaustive).

Popper, le principal précurseur de l'épistémologie évolutionniste, affirme que :

« We choose the theory which best holds its own in competition with other theories; the one which, by natural selection, proves itself the fittest to survive. » (1959, p. 108)

---

<sup>32</sup>On peut tracer ici un parallèle avec la situation qui prévaut en économie, où ce sont les comportements non coordonnés d'une multitude d'agents aux intérêts parfois contradictoires qui génèrent des macro-phénomènes comme, par exemple, le mécanisme de fixation des prix. Hull admet d'ailleurs qu'il est légitime de ranger sa théorie avec les explications de type « main invisible ». Voir Hull (2001, p. 139-148) pour une discussion plus approfondie de ce parallèle, qui n'a pas été sans susciter critiques et réactions (p. ex. Solomon, 1994 ; Ylikoski, 1995 ; Wray, 2000).



**Figure 3.1** *The Battle of Ten Naked Men*, Pollaiuolo (1465). Source : [http://en.wikipedia.org/wiki/File:Battle\\_of\\_the\\_Naked\\_Men\\_MetNY.jpg](http://en.wikipedia.org/wiki/File:Battle_of_the_Naked_Men_MetNY.jpg).

Un bon scientifique doit donc : a) s'assurer que ses théories sont falsifiables ; et b) montrer que celles-ci résistent à toute tentative de falsification. Le bon scientifique adhère à ces normes parce qu'il reconnaît que la *logique* de la recherche scientifique exige qu'il en soit ainsi. Et dans la mesure où la science en général fonctionnerait ainsi, ce serait parce que nous aurions établi, rationnellement, que cela doit être le cas. Popper n'est pas naïf : il sait bien que les scientifiques ont leurs motivations individuelles. Mais en tant que philosophe, internaliste de surcroît, son rôle se limite à établir la norme.

Hull est d'accord avec Popper sur le point fondamental, qui est que la science progresse par un processus de sélection « naturelle », mais cela ne lui suffit pas. Il souhaite montrer *comment* les comportements des scientifiques individuels (qui, comme nous l'avons vu plus haut, n'ont rien d'exemplaire) résultent en cette sélection à l'échelle

de la science et, surtout, *pourquoi* les scientifiques individuels adoptent ces comportements.

On pourrait résumer (grossièrement) la différence en disant que la théorie de Popper fait appel à la rationalité épistémique, alors que celle de Hull fait appel à la rationalité économique *tout en montrant que les deux types de rationalité coïncident*.

Thomas Kuhn aussi, même si ce n'est pas ce qu'on a retenu de lui, se réclame d'un certain darwinisme. Dans les dernières pages de *The Structure of Scientific Revolutions*, il affirme que :

« The analogy that relates the evolution of scientific ideas can easily be pushed too far. But with respect to the issues of this closing section it is very nearly perfect. The process described in section XII as the resolution of revolutions is the selection by conflict within the scientific community of the fittest way to practice future science. » (1962, p. 172)

Outre le fait que Kuhn, au contraire de Hull, en reste à l'analogie et ne tente pas de développer un cadre conceptuel général pour analyser les processus de sélection, la principale différence entre les deux est affaire de degré. Pour Kuhn, la question du progrès (et des facteurs qui l'assurent) n'est pas entièrement claire. À l'intérieur d'un paradigme, la science normale procède en résolvant toujours de plus en plus de *puzzles*, et il y a là une certaine forme de progrès, mais qu'en est-il des révolutions par lesquelles la communauté passe d'un paradigme à un autre ? Les scientifiques qui adhèrent au paradigme « vainqueur » y voient certainement une forme de progrès, mais Kuhn doute qu'il existe un point de vue extérieur, objectif, à partir duquel on pourrait poser un tel jugement. Conséquemment, il insiste sur la nature non téléologique du processus d'évolution :

« And the entire process may have occurred, as we now suppose biological evolution did, without benefit of a set goal, a permanent fixed scientific truth, of which each stage in the development of scientific knowledge is a better exemplar. » (1962, p. 172-173)

Pour Hull, par contre, il n'y a pas de doute : le principal facteur de sélection reste l'expérience empirique. Ce sont des facteurs externes qui motivent cette confrontation avec la réalité et qui déterminent en partie l'accueil qu'on fait à ces résultats, mais c'est la trinité internaliste (raison, arguments et preuves) qui, au bout du compte, assure le progrès de la science.

Nous avons jusqu'ici fait écho à certaines préoccupations de Hull, qui se demande comment expliquer le caractère progressif de la science sans négliger les processus historiques et sociaux qui traversent cette dernière. Pour Hull, l'utilisation de la théorie évolutionniste en épistémologie constitue une façon de répondre à cette question.

Ce qui distingue principalement Hull des auteurs qui l'ont précédé, cependant, c'est, dans les mots de Cecilia Heyes, le fait que :

« Many proponents of an evolutionary or selectionist epistemology are beginning to think that they must shortly “put up, or shut up,” but Hull is unusual because he has “put up”, attempted to cash one of evolutionary epistemology’s promissory notes by suggesting how it might apply to conceptual change in science. » (Heyes, 1988, p. 194)

Hull refuse toutefois d'adhérer à l'étiquette « épistémologie évolutionniste », accolée à cette école depuis Campbell (1974). Le problème avec celle-ci, dit-il, n'est pas qu'elle est évolutionniste, mais bien qu'elle cherche à être une *épistémologie* (Hull, 1988c, p. 12). L'épistémologie traditionnelle est à la recherche des critères selon lesquels on pourra dire que la connaissance est *justifiée*. Sans exclure que ceux-ci puissent exister, c'est un engrenage dans lequel Hull ne souhaite pas mettre le pied : même si le fait de considérer le changement scientifique sous l'angle de l'évolution échouait à remplir les critères de l'épistémologie traditionnelle, cela n'en serait pas moins une façon appropriée de rendre compte du progrès scientifique.

La stratégie de Hull consiste à considérer l'évolution comme la forme générale d'un algorithme dont l'évolution biologique n'est qu'une instanciation particulière<sup>33</sup>. Sa théorie de la sélection fait intervenir deux types d'entités : des « répliqueurs » et des « interacteurs » (Hull, 1980). En biologie, les répliqueurs sont les gènes et les interacteurs sont des organismes complets. En science, les répliqueurs sont les « éléments du contenu substantiel de la science » (Hull, 1988a, p. 140), c'est-à-dire les idées scientifiques, les théories, etc., et les interacteurs sont les scientifiques eux-mêmes<sup>34</sup>.

---

<sup>33</sup>D'autres, avant et après lui, emploient aussi cette stratégie (p. ex. Campbell, 1974 ; Dawkins, 1976 ; Dennett, 1995).

<sup>34</sup>La question de savoir si les scientifiques peuvent vraiment être considérés comme des interacteurs au sens de la définition de Hull a été soulevée (p. ex. Heyes, 1988), et Hull lui-même a modifié quelque peu sa position vers la fin de sa carrière.

Dans le cas des gènes comme dans le cas des idées scientifiques, il suffit que l'environnement exerce une pression de sélection suffisante pour que la reproduction de ces réplicateurs soit différenciée. Les réplicateurs qui performant mieux que les autres au jeu de la reproduction laisseront plus de copies d'eux-mêmes. Est-ce que cela veut dire que les idées qui finissent par prévaloir en science sont nécessairement les plus « vraies » ? Non. Mais dans la mesure où le monde empirique intervient dans cette sélection (à tout le moins via les tests expérimentaux que les scientifiques font subir aux idées), les idées les mieux adaptées à la réalité jouissent d'un certain avantage.

L'approche évolutionniste de Hull ne nie pas que la rationalité individuelle joue un rôle important en science, mais elle montre qu'on peut rendre compte du progrès scientifique sans que la possibilité de celui-ci repose entièrement sur les critères de rationalité traditionnellement recherchés par l'épistémologie.

#### Une analyse générale des processus de sélection

Quand Hull soutient que la science « évolue », il n'a pas besoin de guillemets. Ce n'est pas une métaphore : tant l'évolution scientifique que l'évolution biologique sont des processus qui tombent sous une même analyse générale des processus de sélection.

Pour obtenir de l'évolution, il suffit d'avoir des entités qui puissent fonctionner en tant que *réplicateurs* :

« an entity that passes on its structure largely intact in successive replications. » (Hull, 1988a, p. 134)

Soulignons ici l'utilisation de l'expression « *largely intact* », qui laisse implicitement supposer que des variations sont introduites dans le processus de réplication (1988b, p.255), ce qui est une condition essentielle à l'évolution.

Le rôle de réplicateur n'est toutefois pas, à lui seul, suffisant pour expliquer cette évolution. Il faut aussi des entités qui puissent fonctionner en tant qu'*interacteurs* :

« an entity that interacts as a cohesive whole with its environment in such a way that this interaction *causes* replication to be differential. » (1988a, p. 134, italiques originaux)

Ensemble, les réplicateurs et les interacteurs permettent la *sélection* :

« a process in which the differential extinction and proliferation of interactors *causes* the differential perpetuation of the replicators that produced them. » (1988a, p. 134, italiques originaux)

Dont le résultat est des *lignées* :

« an entity that persists indefinitely through time either in the same or an altered state as a result of replication. » (1988a, p. 135)

Une des particularités de l'approche de Hull, c'est que les notions de réplicateur et d'interacteur réfèrent d'abord et avant tout à des rôles fonctionnels : ces notions sont explicitement détachées des entités qui jouent habituellement ces rôles. Cette souplesse conceptuelle permet de les appliquer à un vaste éventail d'entités, y compris celles du monde biologique.

Les réplicateurs traditionnels, ce sont bien sûr les gènes. Mais dans certaines circonstances particulières, il est possible pour des entités de plus haut niveau de fonctionner comme réplicateurs, dans la mesure où ces entités ont une structure qui peut être reproduite d'une génération à l'autre. Hull donne comme exemple la paramécie : un organisme unicellulaire qui se reproduit par fission. Cet organisme a la particularité de pouvoir transmettre, au fil des générations, certaines modifications subies pendant son existence, sans que ces modifications passent par le matériel génétique : si on enlève par chirurgie une partie du cortex cellulaire de la paramécie et qu'on le repose avec les cils cellulaires pointant vers l'intérieur plutôt que vers l'extérieur, cette modification sera transmise aux générations suivantes. Dans un tel cas, c'est la cellule entière qui fonctionne comme réplicateur.

Dans le cas des interacteurs, cette souplesse conceptuelle est encore plus évidente. Les organismes sont les plus communs de ces « tous cohésifs », mais ils ne sont pas les seuls. Tout en bas de l'échelle, les cellules elles-mêmes sont des interacteurs : leurs caractéristiques phénotypiques influencent la reproduction des réplicateurs qui les ont produites. Dans le cas de la paramécie, la cellule est donc à la fois réplicateur et in-

teracteur. En allant vers l'autre extrémité de la hiérarchie, l'analyse de Hull admet que des populations ou même des espèces entières puissent fonctionner comme interacteurs<sup>35</sup>. Outre une certaine cohésion, le critère essentiel pour qu'une entité soit considérée comme un répliqueur, c'est qu'elle joue un rôle *causal* dans la réplique différentielle : pour être considérée comme un interacteur, il ne suffit pas qu'une entité soit le produit de répliqueurs, il faut que ses interactions avec l'environnement aient une influence sur la perpétuation de ceux-ci.

Les répliqueurs et les interacteurs sont manifestement des individus spatiotemporellement déterminés, mais ils ne sont pas les seuls : c'est aussi le cas des lignées, un point sur lequel Hull a insisté beaucoup en tant que philosophe de la biologie (1976 ; 1978) et qui joue aussi un rôle important dans sa philosophie de la science. Une lignée est formée par une séquence de répliqueurs. Selon les répliqueurs que l'on considère, on aura une lignée différente : on peut ainsi parler de lignées de gènes autant que de lignées d'organismes ou bien, comme nous le verrons, de lignées d'idées.

#### La science comme processus de sélection

Si on souhaite analyser la science en tant que processus de sélection, il faut d'abord identifier les éléments qui y jouent le rôle des répliqueurs. Pour Hull, comme nous l'avons mentionné plus, ce sont les « éléments du contenu substantiel de la science » :

« beliefs about the goals of science, proper ways to go about realizing these goals, problems and their possible solutions, modes of representation, accumulated data, and so on. » (1988c, p. 434)

Bref, à peu près tout ce qui occupe l'esprit d'une scientifique dans le cadre de son travail et, surtout, ce qui peut être communiqué d'une personne à l'autre. Notons aussi à quel point cette définition est large : Hull ne s'intéresse pas qu'au contenu logico-formel de la science.

---

<sup>35</sup>Qu'une telle « sélection de groupe » soit effectivement opérante dans l'évolution biologique est une question controversée. Mais que ça soit le cas ou non, la terminologie peut s'en accommoder.

Hull réfère parfois à l'ensemble de ces éléments comme à des *conceptual tokens*. Il refuse toutefois de donner un sens linguistique ou philosophique trop étroit à ce terme : « I use the phrase “conceptual” in its broadest sense to include all the factors that are operative in conceptual change, not just concepts as philosophers use this term. » (Hull, 1988b, p. 253). À d'autres endroits, Hull parle de « mêmes ». Plus souvent, il utilise simplement « idées », et c'est cette terminologie que nous adopterons ici.

Il ne serait pas injuste de reprocher à Hull d'entretenir un certain flou quant à la nature exacte des répliqueurs en science. Pour un philosophe analytique comme Neil Tennant, par exemple, la question de savoir si les « idées » de Hull réfèrent à des  *croyances*  (entités structurées, de niveau sententiel) ou à des  *concepts*  (entités possiblement non structurées, de niveau sub-sententiel) est nécessaire à une analyse plus poussée de sa théorie (Tennant, 1988, pp. 225–226). Hull préfère pourtant rester agnostique sur ce point, laissant volontairement toutes les possibilités ouvertes :

« I have no strong preferences on the precise way that the conceptual world is subdivided into kinds, just so long as these kinds permit the discovery of regularities in conceptual change. » (1988b, p. 254)

Que l'on soit prêt ou non à accorder à Hull cette ambiguïté sémantique, elle pose un problème concret lorsque vient le temps de construire un modèle formel de sa théorie : il nous faudra à tout le moins statuer sur la  *structure*  des répliqueurs, même s'il nous sera possible de garder leurs propriétés sémantiques à un minimum (voir section 5.2).

Un autre élément à ne pas perdre de vue pour bien interpréter (et formaliser) la nature des idées selon Hull, c'est le fait que les répliqueurs sont des individus spatiotemporellement déterminés. Cela a des conséquences importantes pour la question de l'identité des répliqueurs :

« [I]deas in selection processes are individuated in terms of descent. Two tokens of the same type count as different in the competition for credit if they occur in different conceptual lineages. » (1988c, p. 434)

Supposons un instant que nous adoptions les propositions de la logique du premier ordre pour représenter les idées (ça ne sera pas le cas). Le fait que les idées soient individuées en termes de descendance implique que, si deux scientifiques arrivent

indépendamment à la proposition  $((P \rightarrow Q) \wedge P) \rightarrow Q$ , ce ne serait *pas* la même idée, même si le contenu sémantique est identique. Même chose si une idée est représentée par une chaîne de bits (ce qui, dans un modèle computationnel, sera nécessairement le cas au plus bas niveau) : les chaînes 01011000 et 01011000 ne correspondent à la même idée *que* si elles représentent le même individu dans la lignée conceptuelle, peu importe leur contenu sémantique.

Même si elles sont de nature informationnelle, les idées nécessitent un support physique, que ce soit un livre, un ordinateur, des ondes sonores éphémères ou bien le cerveau d'une scientifique. La réplication d'une idée constitue en sa transmission d'un véhicule à l'autre. Certains de ces véhicules sont passifs, mais d'autres jouent un rôle actif dans la transmission des idées : ils fonctionnent comme interacteurs. Dans la conception originale de Hull, c'est clairement le cas des scientifiques individuels :

« Scientists' brains can serve as vehicles for replication sequences, but scientists themselves are anything but passive vehicles for such sequences. They also function as interactors. » (1988a, p. 140)

Il s'agit toutefois d'un aspect de sa théorie qui a été pris pour cible par certaines critiques. Nous nous penchons ici sur deux points importants soulevés par Cecilia Heyes :

« The identity of Hull's conceptual replicators and interactors is incompatible with his general analysis of selection processes in two respects : (i) elements of the substantive content of science (conceptual replicators) do not produce scientists (conceptual interactors); and (ii) it is not the differential extinction and proliferation of scientists (conceptual interactors) which causes the differential perpetuation of elements of the substantive content of science (conceptual replicators). This is the case whether one regards the extinction of scientists to consist in their death or their withdrawal from scientific activity. » (Heyes, 1988, p. 198)

En ce qui concerne le point (ii), sur l'« extinction » des scientifiques, il est pourtant indéniable que celle-ci joue un rôle dans la perpétuation différentielle des idées scien-

tifiques<sup>36</sup>. C'est peut-être William Bechtel (1988, p. 158) qui exprime le mieux ce fait :

« [T]he scientist is extinguished if he or she is no longer able to influence other scientists. This means that if a scientist is not able to teach students or get papers published in appropriate journals, then, even if the scientist continues to carry out investigations and makes major discoveries, he or she is largely extinguished as an interactor. »

On peut toutefois adopter une autre lecture de Heyes, et souligner avec elle que ce n'est pas *que* l'extinction et la prolifération différentielle des interacteurs qui cause la perpétuation différentielle des réplicateurs : c'est aussi, et surtout, leur comportement. Hull est, bien sûr, d'accord avec cet énoncé, et sa définition d'« interacteur » s'accorde parfaitement avec celui-ci. Sa définition de « sélection », par contre, laisse de côté le comportement des interacteurs pour n'insister que sur leur extinction et leur prolifération.

Le point (i) de Heyes, sur la *production* des interacteurs est tout aussi délicat. L'idée de départ de Hull, exprimée dans *Science as a Process*, était que les scientifiques *en tant que* scientifiques, sont effectivement « produits » par leurs idées, en ce que celles-ci influencent fortement leur comportement.

Cette explication n'est toutefois pas entièrement satisfaisante et, surtout, trop peu développée par Hull pour convaincre les sceptiques. Même Bechtel, tout en étant plutôt sympathique à la caractérisation des scientifiques comme interacteurs, se permet de remarquer laconiquement que « It would be a useful enterprise to detail how ideas produce scientists. » (1988, p. 164).

Plutôt que d'approfondir la thèse selon laquelle les idées « produisent » les scientifiques, Hull choisit de relâcher sa définition de sélection. Dans sa réponse à Heyes, il propose de remplacer l'exigence de production par un lien causal plus ténu :

« [S]cientists are the agents in conceptual interaction. They bring about the realization of conceptual phenotypes, but they themselves do not exhibit them, nor are they as entities produced by them. Hence, in my characterization of selection, "produce" must

---

<sup>36</sup>Notons toutefois que, contrairement à ce que Max Planck laisse entendre dans son autobiographie (Planck, 1950), les scientifiques peuvent changer d'idée à tout âge, et il n'est pas nécessaire d'attendre la disparition d'une génération pour que triomphe une nouvelle théorie (Hull *et al.*, 1978).

be replaced by a more general relation such as the “replicators to which they themselves are causally related.” » (2001, p. 110)

Mais quand il publie à nouveau ses définitions en 2001, dans le recueil *Science and Selection*, il révisé plutôt la notion de sélection en ne gardant que le lien causal des interacteurs aux réplicateurs et en omettant toute relation dans l’autre sens :

« *selection* — a process in which the differential extinction and proliferation of interactors *cause* the differential perpetuation of the replicators. » (2001, p. 110, italiques originaux)

Hull ne s’arrête toutefois pas là et va jusqu’à renier complètement la thèse selon laquelle les scientifiques sont des interacteurs :

« Scientists certainly enter into conceptual change in science, but as agents that facilitate interaction, not as interactors themselves. » (2001, p. 110)

Dans ce contexte, pour reprendre une image employée par Tennant (1988, p. 226), les scientifiques individuels se retrouvent dans une position similaire à celle des éleveurs d’animaux dans le processus de sélection artificielle : ce sont eux qui décident quelles sont les caractéristiques désirables qui doivent être perpétuées ou éliminées, mais ils ne sont pas *eux-mêmes*, en tant qu’individus, constitués de ces caractéristiques.

Mais qui sont les interacteurs, alors ?

Heyes (1988, p. 199) suggère de considérer les idées dans le cerveau des scientifiques comme les réplicateurs, et les autres véhicules pour les idées (voix, texte, image, gestes, etc.) comme les interacteurs. Ces éléments ne pourraient fonctionner comme interacteurs que dans le contexte de la « machinerie cognitive des scientifiques », préservant ainsi le rôle crucial de ces derniers dans le processus, mais Hull tient à l’idée que les idées véhiculées par les scientifiques influencent leur comportement, et la suggestion de Heyes occulte quelque peu celle-ci.

Tennant, lui, y va d’une suggestion qui, tout en pouvant sembler contre-intuitive de prime abord, offre une solution élégante aux problèmes abordés ci-dessus :

« [I]f anything to do with scientists themselves is to count, in Hull’s schema, as an interactor (for the concepts and beliefs identified as replicators), then it should be *ideological time-slices* of living scientists ; or *momentary intellectual profiles* ; or *instantaneous me-*

*motypes*—call them what you will. These can be suitably construed, given a materialist theory of mind, as *in the physical world* : they will exist as patterns of neuronal linkages and states of excitation, say (Schilcher and Tennant 1984, p. 117 and p. 124). Scientists themselves—the population of actual living bodies—are then just part of the *selective environment* with which *these* interactors, now properly identified as such, interact. » (Tennant, 1988, p. 227, italiques originaux.)

Sans s'étendre outre mesure sur les conséquences de celle-ci, Hull affirme trouver la suggestion de Tennant « certainement appropriée » (1988b, p. 256). Nous soutiendrons ici qu'elle échappe aux deux problèmes soulevés par Heyes, tout en respectant l'exigence de Hull concernant le lien entre les idées qui sont répliquées et le comportement des scientifiques individuels.

Le premier point de Heyes était que les réplicateurs conceptuels ne « produisent » pas les scientifiques. Et si on considère les scientifiques en tant que personnes à part entière, de leur naissance à leur mort, il faut bien donner raison à Heyes : on ne peut les réduire au produit de leurs idées. Mais si on suit Tennant et que l'on adopte des « tranches temporelles idéologiques » ou des « profils intellectuels momentanés » comme interacteurs, ce n'est plus un problème que de considérer ceux-ci comme le produit des idées. Selon la façon dont on définit ce concept, cela peut même devenir tautologique : une tranche temporelle idéologique pourrait être, *par définition*, le produit d'un ensemble d'idées.

Le second point de Heyes est lui aussi résolu par la suggestion de Tennant. Cette fois-ci, c'est bel et bien l'extinction et la prolifération différentielle des interacteurs qui causent la perpétuation différentielle des éléments du contenu substantiel de la science. Quand une scientifique abandonne une théorie pour une autre, quand elle cesse de promouvoir une idée, il y a une tranche temporelle idéologique qui vient de s'éteindre. Quand une chercheuse s'enthousiasme pour une nouvelle théorie, quand elle se met à propager une idée qu'elle avait ignorée jusque là, il y a un nouveau profil intellectuel momentané qui vient de naître.

Finalement, on en revient à l'exigence de Hull : les idées doivent être, dans une certaine mesure, la source du comportement des interacteurs. C'est par ce comportement que les idées sont confrontées au monde et c'est au travers lui que s'opère le processus

de sélection. Dans le schème de Tennant, le comportement des scientifiques *en tant que* profils intellectuels momentanés est effectivement le produit des idées qui composent ce profil intellectuel. Cela ne veut pas dire que le comportement du scientifique en tant que personne à part entière n'est pas influencé par d'autres caractéristiques personnelles, qui font partie de l'environnement de sélection des idées, mais il y a tout de même un sens où l'activité intellectuelle du scientifique est le résultat des idées qui l'habitent.

Nous ne voulons pas soutenir ici que la suggestion de Tennant règle tous les problèmes conceptuels reliés au statut des scientifiques comme interacteurs dans la théorie de Hull. Pour ce faire, cette suggestion demanderait à être développée bien au-delà de son niveau de détail actuel. De nombreuses questions sont laissées en suspend, comme par exemple problème de l'individuation des tranches temporelles. Quand est-ce qu'elles commencent ? Quand est-ce qu'elles finissent ? Peuvent-elles se recouper entre elles ?

Ce que nous soutenons, par contre, c'est que pour les besoins de notre démarche, le schème de Tennant est amplement suffisant. Comme nous le verrons dans le chapitre 5, les agents qui peuplent notre modèle ne sont, de toute façon, pas grand-chose de plus qu'un « profil intellectuel » idéalisé. Ils ont des caractéristiques personnelles qui influencent leur comportement et ils occupent une position particulière dans un réseau social, mais ils sont d'abord et avant tout constitués par un ensemble d'idées auxquelles ils adhèrent à divers degrés. En ce sens, ils ont pleinement droit au titre d'interacteurs.

### Objections à la science comme processus de sélection

Malgré sa stratégie consistant à établir d'abord une analyse des processus de sélection qui soit suffisamment générale pour couvrir à la fois l'évolution biologique et l'évolution socioculturelle (incluant le processus scientifique), Hull est conscient que ses critiques ne manqueront pas de souligner les différences entre ces deux processus pour tenter de montrer que : « these phenomena are so different that no analysis could

possibly apply to both. » (1988c, p. 440) Il recense donc les principales différences qui pourraient être ainsi évoquées, et tente de montrer qu'elles relèvent de l'erreur ou de la confusion conceptuelle. Ces objections sont au nombre de sept, et nous les aborderons tour à tour. Notre objectif ici n'est pas tant de défendre la théorie que d'utiliser ces points pour en clarifier certains aspects.

Première objection : « l'évolution socioculturelle est beaucoup plus rapide que l'évolution biologique »<sup>37</sup>.

En termes purement temporels, il est vrai que l'évolution socioculturelle semble fonctionner à une échelle plus petite que l'évolution biologique. Les changements génétiques dans l'espèce humaine, par exemple, sont relativement mineurs depuis l'époque de Socrate, par exemple, alors que les changements culturels, eux, ont été majeurs. Et combien de révolutions scientifiques depuis Newton pendant que l'évolution biologique faisait augmenter légèrement la fertilité maternelle à L'Île-aux-Coudres ? (Milot *et al.*, 2011)

Notre tendance à considérer le taux de changement évolutionnaire en termes d'années est toutefois biaisée. Du point de vue de l'évolution, l'unité temporelle pertinente, c'est la génération. Et de ce point de vue, l'évolution des répliqueurs culturels ou scientifiques procède exactement à la même vitesse que celle des répliqueurs biologiques : une génération à la fois. Dans les mots de Hull :

« Each time an old man shows a young boy how to make a slingshot, that is a conceptual generation. Each semester in colleges and universities around the world, the Pythagorean theorem is taught. Each replication forms a conceptual generation. » (1988c, p. 441)

La durée temporelle des générations conceptuelles peut varier grandement, mais c'est aussi le cas de celle des générations biologiques : la durée d'une génération chez l'éléphant africain (*Loxodonta africana*) oscille autour de vingt-cinq ans (Hanks et McIn-

---

<sup>37</sup>Le libellé des objections est traduit directement de Hull (1988c, p. 440).

tosh, 1973), alors que celle du virus de la grippe saisonnière se situe entre trois et quatre jours (Cowling *et al.*, 2009).

Deuxième objection : « les entités qui sont sélectionnées par l'évolution biologique sont atomiques alors que celles qui sont sélectionnées par l'évolution socio-culturelle ne le sont pas ».

Cette objection, comme la plupart de celles qui sont abordées ici, relève selon Hull d'une conception simpliste de la biologie évolutionniste. Oui, il est vrai que, au niveau moléculaire, l'ADN est composé de quatre bases bien distinctes (adénine, thymine, cytosine et guanine, notées A, C, G et T) qu'il est en tentant de considérer comme les particules atomiques sur lesquelles opère la sélection naturelle. L'idée qu'il existe une correspondance « un pour un » entre chaque caractère phénotypique et une séquence d'ADN donnée est toute aussi problématique que celle selon laquelle l'ensemble de nos connaissances peut se réduire à des propositions atomiques portant sur des états de choses dans le monde. Le concept de gène a changé significativement depuis les débuts de la génétique mendélienne (Portin, 2002) et fait encore aujourd'hui l'objet de débats philosophiques (Rheinberger et Müller-Wille, 2010).

Nous avons insisté plus haut sur le rôle fonctionnel que Hull attribue aux entités qui composent selon lui le processus de sélection. Est un réplicateur, selon lui, ce qui *fonctionne comme* un réplicateur dans un processus de sélection, peu importe sa taille, et peu importe s'il est ou non constitué de particules atomiques. Il rejoint en cela le concept de réplicateur chez Dawkins (1976) et celui de gène évolutionnaire chez Williams (1996).

La situation est similaire pour l'évolution socioculturelle. Hull s'appuie sur Rosenberg (1982, p. 22–23) pour affirmer que :

« anyone who proposes to provide a selection model of sociocultural change will not be able to simply adopt the “units” identified and individuated by social scientists, conceptual historians, and philosophers and treat them as units of sociocultural selection. Instead, they will have to individuate the relevant units through attempting to apply their models to real cases. » (1988c, p. 443)

S'il y a bel et bien des unités dont la structure a été transmise « largement intacte » d'une génération à l'autre, et si les véhicules portant ces unités sont entrés dans des interactions qui ont favorisé la reproduction de certaines de ces unités plutôt que d'autres, alors il devrait être possible de les identifier dans un domaine donné, que ce soit des idées scientifiques, des gènes biologiques ou encore des mèmes internet<sup>38</sup>.

Troisième objection : « l'évolution biologique est toujours biparentale alors que l'évolution socioculturelle l'est rarement ».

Considérer l'évolution biologique comme étant typiquement biparentale relève de l'anthropocentrisme. Oui, la reproduction est biparentale chez l'*Homo Sapiens* et beaucoup d'autres animaux avec lesquels nous interagissons, mais : « Throughout most of the history of Earth, the only organisms available for selection to act upon were asexual » (Hull, 1988c, p. 444). Bien que la reproduction asexuelle n'ait plus l'exclusivité aujourd'hui, elle reste la plus commune chez les organismes unicellulaires, et elle est présente chez de nombreux organismes multicellulaires (plantes, champignons, animaux invertébrés, et même certains vertébrés).

Le fait qu'une nouvelle idée puisse être une variante d'une seule idée précédente n'est donc pas un problème pour la théorie de Hull. Qu'en est-il des cas où une nouvelle idée est une combinaison de plus de deux idées ? La question de la plurisexualité est une question que se posent les biologistes depuis au moins Fisher :

« No practical biologist interested in sexual reproduction would be led to work out the detailed consequences experienced by organisms having three or more sexes ; yet what else should he do if he wishes to understand why the sexes are, in fact, always two ? » (Fisher, 1930, p. ix)

Nous savons maintenant que les organismes plurisexuels sont rares, mais qu'ils existent (Iwasa et Sasaki, 1987). Et même s'il n'y a pas encore de consensus quant

---

<sup>38</sup>Ces derniers ne sont qu'une manifestation particulière des mèmes au sens de Dawkins (1976) mais ils offrent aussi un potentiel intéressant pour des études concrètes du processus de répliation (p. ex., Chielens et Heylighen, 2005 ; Shifman et Thelwall, 2009)

à la raison de cette rareté, il reste que sur le plan conceptuel, la reproduction pluriparentale est tout à fait envisageable. Il se peut très bien que, dans le contexte biologique, les circonstances qui lui permettraient d'évoluer et de persister de façon stable soient extrêmement rares, mais ça n'implique pas qu'il en soit de même dans le monde de l'évolution scientifique ou socioculturelle.

Combien de parents ont les idées, alors ? Hull admet toutes les possibilités : « Ideas do not always have two sources. Sometimes they have one, sometimes two, sometimes more. » (1988c, p. 444). Comme pour la question de l'identification des répliqueurs eux-mêmes, l'identification des parents de ceux-ci est une opération qui doit être menée cas par cas. À chaque idée sa ou ses sources.

Dans le modèle multi-agents qui sera décrit au chapitre suivant, nous adoptons la reproduction biparentale pour les idées. Celle-ci nous semble correspondre à l'intuition selon laquelle on génère de nouvelles idées en combinant des idées existantes, tout en évitant les complexifications formelles qui viendraient avec une reproduction pluriparentale<sup>39</sup>.

Dans *Science et méthode*, Henri Poincaré soutient que l'inventivité mathématique repose sur la combinaison d'idées entre elles. Il décrit ainsi sa propre expérience :

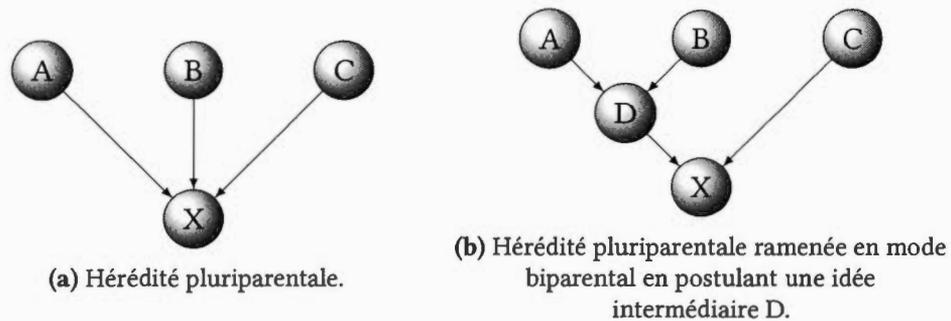
Un soir, je pris du café noir, contrairement à mon habitude, je ne pus m'endormir : les idées surgissaient en foule ; je les sentais comme se heurter, jusqu'à ce que deux d'entre elles s'accrochassent, pour ainsi dire, pour former une combinaison stable. Le matin, j'avais établi l'existence d'une classe de fonctions fuchsiennes, celles qui dérivent de la série hypergéométrique ; (1908, p. 51)

Donald Campbell (1960) fait aussi reposer sa théorie de la créativité humaine sur le fait que la combinaison d'idées existantes est un des principaux mécanismes permettant de générer les variations nécessaires à la création de nouvelles idées. Plus récemment, le psychologue Dean Keith Simonton (1999 ; 2003) soutient que : « scientific creativity constitutes a form of constrained stochastic behavior. That is, it can be accurately modeled as a quasi-random combinatorial process. » (2003, p. 475)

---

<sup>39</sup>De futures versions du modèle pourraient comparer entre eux les différents types de reproduction, mais c'est une complication que nous avons choisi d'écarter pour l'instant.

Comme l'illustre la figure 3.2, une approche biparentale n'empêche pas de combiner plus de deux idées pour donner naissance à une nouvelle idée, mais cette combinaison doit se faire en plusieurs étapes, en générant à chaque fois une idée intermédiaire résultant de deux parents (idée D dans la figure 3.2b). Nous reviendrons plus en détail à la génération de nouvelles idées dans la section 5.4.1.



**Figure 3.2** Hérité pluriparentale vs. biparentale des idées.

Quatrième objection : « les emprunts inter lignées sont plus fréquents dans l'évolution socioculturelle que dans l'évolution biologique ».

On peut faire ici le même genre de réponse que pour les questions précédentes : les emprunts inter lignées sont plus fréquents qu'on le pense en biologie. Mais pour Hull, indépendamment de l'état des choses en biologie, cette objection relève surtout d'une confusion conceptuelle.

Quand on invoque les transferts inter lignées dans le monde conceptuel, ce qu'on a en tête, c'est souvent deux lignées *sociales* distinctes qui s'échangent des idées. Mais pour Hull, c'est la lignée conceptuelle qui importe, et dans un cas comme celui-ci, on n'a pas affaire à un transfert de lignée conceptuelle : c'est une même lignée conceptuelle qui se poursuit, même si celle-ci traverse la frontière entre deux groupes sociaux :

« From [the perspective of conceptual lineages], all sociocultural inheritance is vertical. Claims that one sort of replication is "horizontal" result from viewing one replication sequence from the perspective of another. » (Hull, 1988c, p. 452)

Cela étant dit, un certain chevauchement entre les lignées sociales et les lignées conceptuelles est bien réel. Il résulte de la structure démique de la science et joue un rôle dans le concept de valeur sélective conceptuelle inclusive, deux notions que nous aborderons bientôt dans la section 3.3.

Cinquième objection : « l'évolution socioculturelle est lamarckienne alors que l'évolution biologique est darwinienne ».

Pour pouvoir parler d'évolution lamarckienne, au moins deux conditions doivent être remplies :

1. L'organisme doit pouvoir acquérir des caractéristiques au cours de son existence en réponse à des facteurs environnementaux.
2. Ces caractéristiques doivent pouvoir être transmises à sa progéniture de façon génétique.

L'évolution socioculturelle n'est donc clairement *pas* lamarckienne au sens littéral (et personne ne soutient qu'elle l'est) : les idées acquises au cours de la vie d'une personne ne sont pas transmises à sa progéniture par voie génétique. Ceux qui soutiennent que l'évolution culturelle est lamarckienne l'entendent plutôt dans un sens métaphorique :

« In this *metaphorical* sense, the analogue to genes is ideas, concepts, practices, and so on—in short, memes. Thus, the transmission of ideas or practices from person to person is an instance of the inheritance of acquired *memes*, not acquired *characteristics*. From the perspective of genetic transmission, memes are characteristics ; from the memetic perspective, they are analogue to genes. Social learning can be made to look Lamarckian only by mixing the literal and metaphorical interpretations. » (Hull, 1988c, p. 454, italiques originaux)

Du point de vue de Hull, les idées d'une personne sont en quelque sorte son « génome » conceptuel, et la façon dont ces idées s'expriment dans le comportement constitue le « phénomène ». Les changements au génome (aux idées) causent les changements au phénomène (les comportements), mais pas l'inverse, comme ça serait le cas si l'évolu-

tion conceptuelle était lamarckienne. Cela ne veut pas dire que l'interaction avec le monde extérieur ne peut pas être la source de nouvelles idées. Au contraire, cela se produit constamment. Tout processus d'apprentissage est une instance de ce phénomène. Mais dans ce cas, si on s'en tient à la perspective proposée par Hull, ce sont, justement, de *nouvelles* idées. Celles-ci peuvent être des variations d'idées préalables, mais ce ne sont pas les idées existantes qui sont directement modifiées.

Hull ne considère pas explicitement cette possibilité, mais si on interprétait les changements conceptuels provoqués par l'apprentissage comme étant des changements aux idées existantes plutôt que comme la génération de nouvelles variantes, cela ferait-il de l'évolution conceptuelle un processus lamarckien ? C'est une possibilité que nous ne pouvons pas complètement exclure, et quelqu'un qui souhaiterait livrer bataille à Hull sur ce point trouverait peut-être ici une faille à exploiter. Ce n'est pas notre objectif.

Sixième objection : « l'évolution conceptuelle, à tout le moins en science, est progressive, alors que l'évolution biologique ne l'est pas ».

L'histoire de la biologie est jalonnée de débats sur le caractère progressif ou non de l'évolution, mais le consensus actuel, c'est que l'évolution biologique ne peut être dite progressive que dans un sens *local* : dans un environnement stable, le processus de sélection a tendance à produire des organismes qui sont de mieux en mieux adaptés à cet environnement. Lorsque l'environnement change, ce qui était avantageux ne l'est plus nécessairement, et l'évolution peut entraîner une espèce (si elle survit) dans une direction complètement différente. Qu'en est-il de la science ? Hull pose le problème de la façon suivante :

« Thus, if scientific change is both locally and globally progressive, an important difference exists between the two. One alternative is to argue that science is no more globally progressive than is biological evolution. As far as I can see, the history of science precludes such a conclusion. » (1988c, p. 464)

Bien qu'il soit possible d'identifier quelques régressions locales dans l'histoire de la science, il est difficile de nier que, globalement, elle progresse. C'était d'ailleurs un des

objectifs de Hull que d'identifier les mécanismes par lesquels advient ce progrès. Mais si l'évolution n'est en mesure de produire qu'une adaptation locale, peut-on encore soutenir que c'est ce mécanisme qui est responsable du progrès scientifique ?

Hull soutient que la différence entre l'évolution biologique, localement progressive, et l'évolution scientifique, globalement progressive, n'en est pas une de mécanisme, mais bien d'environnement. L'environnement de sélection des idées scientifiques, c'est le monde dans son ensemble. Comme le disait Aristote : « la science consiste dans la connaissance universelle » (1995, 87b). Les éléments qui changent, dans le monde, n'intéressent la science que dans la mesure où elle peut décrire les mécanismes ou les lois par lesquelles surviennent ces changements. (Aristote : « c'est d'une pluralité de cas particuliers que se dégage l'universel ». (88a)) Mais encore faut-il supposer une certaine stabilité dans le monde :

« It is the successive approximations of scientific theories to [eternal, immutable regularities] which are responsible for the global progress in science. If no such regularities exist, scientists could not approximate them. » (Hull, 1988c, p. 467)

C'est le vieux problème humien de l'induction : il n'y a pas de certitude possible quant à l'immutabilité des lois de la nature. Mais ce n'est pas un problème pour la théorie de Hull, ou pour l'épistémologie évolutionniste en général. S'il s'avérait demain matin que les lois de la nature ont changé, ça ne serait, au fond, qu'un changement d'environnement pour la science. Son progrès apparent n'aurait été que local et il lui faudrait, comme une espèce biologique, s'adapter à une nouvelle réalité (ou disparaître).

Septième objection : « l'évolution socioculturelle est intentionnelle alors que l'évolution biologique ne l'est pas ».

Hull ne nie pas, bien au contraire, qu'il y ait un aspect intentionnel<sup>40</sup> à la science : « If anything counts as intentional, science does. » (1988c, p. 457). Ce qu'il nie, par contre, c'est que ça soit un problème pour sa théorie. L'intentionnalité joue un rôle important dans le processus d'évolution scientifique, tout comme elle peut jouer un

---

<sup>40</sup>On entend ici « intentionnel » au sens de « résultat d'une volonté, dirigé vers un but » plutôt qu'au sens philosophique plus large d'« être à propos de quelque chose ».

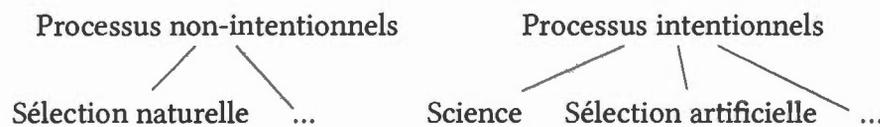
rôle important dans l'évolution biologique : l'agriculture et l'élevage font intervenir une sélection artificielle, intentionnelle, mais n'en sont pas moins des processus d'évolution.

Ce qui sous-tend cette objection c'est la distinction, traditionnellement fondamentale, entre les processus intentionnels, domaine des sciences humaines (*Geisteswissenschaften*), et les processus non intentionnels, domaine des sciences de la nature (*Naturwissenschaften*).

Cette distinction relève d'une forme d'anthropocentrisme pernicieuse :

« *Homo sapiens* is unique. All species are. But this uniqueness is not enough for many, probably most people, philosophers included. For some reason, it is very important that the species to which we belong be uniquely unique. » (1988c, p. 469)

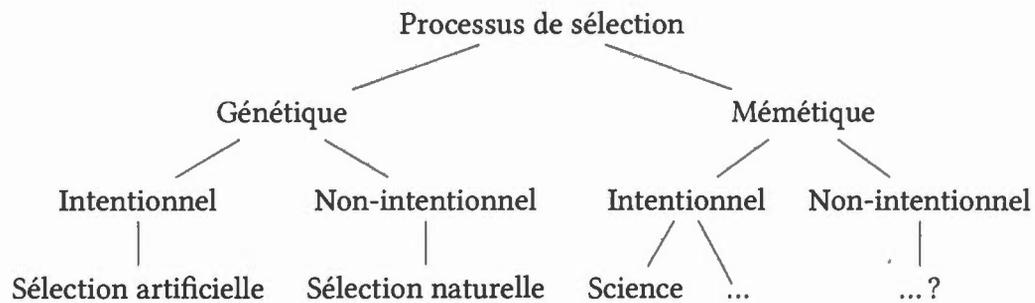
Et si on tente de maintenir cette distinction fondamentale, on se retrouve dans la position inconfortable de devoir ranger la sélection artificielle, avec la science, du côté des processus intentionnels (figure 3.3).



**Figure 3.3** La classification qui, selon Hull, sous-tend l'objection de l'intentionnalité.

Ce que Hull propose plutôt, c'est de ranger tous les processus de sélection sous une même catégorie et de subdiviser d'abord celle-ci selon le type de réplicateurs qui y interviennent : les gènes d'un côté et les réplicateurs culturels de l'autre (les mêmes, en fait, quoique Hull évite d'utiliser le terme). Le caractère intentionnel ou non du processus n'intervient qu'en second lieu (figure 3.4 à la page suivante).

Dans cette optique, l'intentionnalité qui intervient en science n'est pas un obstacle à son étude sous la loupe évolutionnaire : c'est simplement un aspect important dont on doit tenir compte lorsqu'on tente de comprendre l'environnement dans lequel se fait la sélection des réplicateurs impliqués. Et c'est justement sur cet environnement que nous allons nous pencher dans la prochaine section.



**Figure 3.4** La classification des processus de sélection selon Hull.

### Les mécanismes sociaux à l'œuvre en science

L'analyse de Hull jette un éclairage nouveau sur certains des mécanismes sociaux qui sont à l'œuvre dans le monde scientifique. Nous nous pencherons tout particulièrement sur trois d'entre eux : le compromis entre crédit et support, la valeur sélective conceptuelle inclusive et la structure démique de la science.

#### Crédit vs support

Une fois établi le rôle des scientifiques en tant qu'agents dans le processus de changement scientifique, il faut se pencher sur le but vers lequel tendent leurs actions. Hull identifie deux motivations principales dans le comportement des scientifiques : une sincère curiosité envers le monde naturel et le désir de recevoir du crédit pour leurs contributions à la science. Malgré l'importance de la curiosité dans le choix d'une carrière intellectuelle, et la reconnaissance d'un désir sincère de comprendre la Nature chez la plupart des scientifiques, Hull n'accorde pas à ces notions un rôle explicatif particulièrement important dans sa théorie. La curiosité explique en partie *pourquoi* on s'adonne à la science, mais nous éclaire assez peu sur *comment* on s'adonne à la science, et c'est cette dernière question qui l'intéresse le plus. Il considère toutefois le désir d'obtenir du crédit comme un des facteurs principaux guidant le comportement des scientifiques.

La notion de crédit, en tant que reconnaissance professionnelle recherchée par les scientifiques, s'exprime surtout par les citations dans les articles scientifiques : citer le travail d'un autre c'est, la plupart du temps, reconnaître sa valeur<sup>41</sup>. Le crédit accumulé par un chercheur lui donne généralement accès à des postes plus prestigieux, des subventions, des offres de collaboration, des invitations à des conférences, etc. Et tous ces avantages permettent à leur tour d'accumuler plus de crédit<sup>42</sup>.

S'ils veulent obtenir du crédit pour leurs idées, les scientifiques doivent d'abord les faire accepter par leurs pairs. Pour ce faire, ils ont besoin de support. La notion de support découle du fait que la science, telle qu'elle se pratique maintenant, est une entreprise à la fois collective et cumulative. Le mythe du génie solitaire qui révolutionne tout à partir de sa tour d'ivoire n'est que cela : un mythe. Qui souhaite contribuer à l'entreprise scientifique n'a d'autre choix que de s'appuyer sur les travaux de ceux et celles qui l'ont précédé, ne serait-ce que pour se donner une chance d'être pris au sérieux par la communauté scientifique.

Quant au compromis entre les deux :

« One cannot gain support from a particular work unless one cites it, and this citation automatically both confers worth on the work cited and detracts from one's own originality. Scientists would like total credit and massive support, but they cannot. Science is so organized that scientists are forced to trade off credit for support. » (Hull, 1988a, p. 127)

Il y a tout un continuum de stratégies possibles pour une scientifique quant au compromis entre crédit et support, allant des stratégies à haut risque, potentiellement très payantes en terme de crédit à des stratégies beaucoup moins payantes mais bien moins risquées. La distribution de ces stratégies dans la population scientifique aura un im-

---

<sup>41</sup>Rappelons ici que même les citations négatives, où un auteur s'attaque aux idées d'un autre, confèrent généralement une certaine valeur à ces idées, en reconnaissant implicitement qu'elles sont dignes d'attention et méritent d'être critiquées.

<sup>42</sup>C'est une variante de ce que le pionnier de la sociologie des sciences Robert K. Merton a baptisé l'« effet Matthieu », en référence au verset : « Car on donnera à celui qui a, et il sera dans l'abondance, mais à celui qui n'a pas on ôtera même ce qu'il a. » (Matthieu 13:12, trad. Louis Segond). En science, dans les mots de Merton (1968, p. 57) : « eminent scientists get disproportionately great credit for their contributions to science while relatively unknown scientists tend to get disproportionately little credit for comparable contributions. »

pact sur l'entreprise scientifique dans son ensemble, et c'est une des questions que nous étudierons avec notre modèle.

#### Valeur sélective conceptuelle inclusive

Il y a toutefois un phénomène que le compromis entre crédit et support n'explique pas : la tendance des chercheurs à citer leurs propres étudiants. Pourquoi partager le crédit avec ses étudiants alors que le fait de les citer apporte bien peu de support ? L'explication de Hull fait intervenir un concept important en biologie : celui de « valeur sélective inclusive » (*inclusive fitness*—Hamilton 1964), qui devient, dans le contexte de l'évolution scientifique, la « valeur sélective *conceptuelle* inclusive » (*conceptual inclusive fitness*).

En biologie, de façon très informelle, la théorie de la valeur sélective inclusive nous dit qu'un gène codant pour un comportement sera favorisé par la sélection naturelle si l'avantage procuré au bénéficiaire de ce comportement, multiplié par la probabilité que celui-ci partage le gène en question, est plus grand que le coût associé au comportement. C'est ce qu'on appelle parfois « la règle de Hamilton ». La probabilité de partager un gène correspond au « coefficient d'apparentement » de Wright (1922) :  $1/1$  pour des jumeaux identiques,  $1/2$  entre une mère et son enfant,  $1/2$  entre un frère et sa sœur,  $1/4$  entre un oncle et sa nièce, etc. La valeur sélective inclusive n'est pas la seule explication possible pour l'altruisme en biologie, mais elle joue un rôle central dans les relations entre les membres d'une même famille.

En quoi cela peut-il éclairer ce qui se passe en science ? Le système de crédit érige une barrière élevée pour qui veut entrer dans le champ scientifique. Puisqu'un nouveau scientifique n'a pas encore accumulé de crédit, le fait de citer ses travaux n'offre de support à personne. Et il n'obtiendra pas de crédit à moins d'être cité. Pour briser ce cercle vicieux, un comportement altruiste semble être requis de la part d'au moins un scientifique bien établi. C'est souvent le superviseur qui, le premier, cite ses propres étudiants. D'autres gestes altruistes, comme aider l'étudiant à se trouver un poste après ses études, ou même le simple fait d'agir comme mentor, sont non seule-

ment fréquents, mais font aussi partie des normes de comportement communément acceptées en science. Pourquoi ? Puisque Hull a choisi de poser la recherche de crédit comme principale motivation des scientifiques, il doit expliquer pourquoi ceux-ci sont disposés à partager leur crédit sans pour autant en retirer un avantage immédiat.

Pour comprendre le mécanisme qui est à l'œuvre, il suffit toutefois de regarder ce qui se passe à long terme : le fait de partager une partie de son crédit avec ses étudiants augmente les chances que ceux-ci aient une carrière fructueuse, lors de laquelle ils seront des vecteurs de transmission pour les idées de leur ancien superviseur, lequel augmentera ainsi les chances d'obtenir du crédit pour ses idées :

« Just as organisms behave in ways which result in replicates of their own genes or duplicates of these genes in close kin being transmitted to later generations, scientists behave in ways calculated to get their views accepted as their view by other scientists, in particular those scientists working on problems most closely associated with their own. » (Hull, 1988a, p. 128)

Il nous faut toutefois conclure cette section en mentionnant une certaine équivoque entre le concept de crédit et celui de valeur sélective. Il est vrai que la scientifique qui a beaucoup de crédit a plus de chances d'attirer des étudiants plus nombreux et, de façon générale, plus de chances de transmettre ses idées. Mais contrairement à la valeur sélective biologique, qui est *définie* en termes de potentiel reproductif, le crédit est seulement *corrélé* avec le potentiel de reproduction conceptuelle. Il n'est pas défini en ces termes.

Hull n'établit pas cette distinction de façon aussi claire qu'on pourrait le souhaiter, mais cette ambiguïté n'invalide pas le concept de valeur sélective conceptuelle inclusive. Nous profiterons par ailleurs de ce lien étroit entre crédit et valeur sélective conceptuelle inclusive lorsque, au chapitre 6, nous examinerons l'effet des comportements des scientifiques sur le crédit, et supposerons implicitement que toute augmentation de crédit résulte en une augmentation de la valeur sélective conceptuelle inclusive. Nous reviendrons plus concrètement sur cette question lors de l'analyse des résultats de simulation (section 6.4.2).

## Structure démique de la science

Finalement, une notion chère à Hull est que la communauté scientifique est structurée en « dèmes ». En biologie, un dème est une sous-population isolée, dont les membres se reproduisent surtout entre eux :

« Ideally a deme consists of organisms in sufficient proximity to each other that they all have equal probability of mating with each other and producing offspring, provided that they are sexually mature, of the opposite sex, and equivalent with respect to sexual selection. To the extent that these conditions are met, the organisms belonging to a deme share the same gene pool. » (Hull, 1988c, p. 433).

Hull observe que, en science, on retrouve aussi ce genre de communauté, où les chercheurs interagissent et collaborent surtout entre eux. La fonction première de ces groupes est le partage des ressources conceptuelles : différents membres apportent au groupe différentes connaissances et habilités scientifiques, et la mise en commun de ces ressources profite au groupe.

Le dème sert aussi d'incubateur pour les idées scientifiques. Il permet à un chercheur de soumettre de premières ébauches d'idées à des collègues qui aideront à sélectionner les meilleures de celles-ci, sans risquer sa réputation auprès de la communauté scientifique dans son ensemble.

L'échange d'idées à l'intérieur d'un dème bien intégré est assez fréquent pour que la valeur sélective conceptuelle inclusive joue un rôle important dans le comportement intradémique (1988c, p. 435), mais cette intégration a aussi un autre effet : le succès des idées véhiculées par les membres d'un dème scientifique est lié au succès (et à la survie) du dème lui-même. Dans le cas où les idées issues d'un dème sont réfutées par la communauté scientifique, la « sélection interdémique » est « surimposée » sur la sélection individuelle (Hull, 1988a, p. 130). En général, les intérêts de l'individu coïncident avec ceux du dème dont il fait partie, mais il arrive que ça ne soit pas le cas. Selon Hull, les conflits internes à un dème sont souvent causés par une telle divergence, réelle ou perçue (1988a, p. 132).

## Conclusion

La théorie de Hull n'a pas été sans susciter de nombreuses réactions. En 1988, un peu avant la parution de *Science as a Process*, la revue *Biology and Philosophy* publie un numéro spécial autour d'un précis rédigé par Hull (1988a) sur son livre à venir. De nombreux chercheurs s'intéressant à ces problèmes y trouvent l'occasion de prendre position sur la démarche de Hull et de souligner les problèmes qu'ils ne manquent pas d'y découvrir. Les réponses de Hull à ces objections (1988b) sont une précieuse source de précisions sur sa théorie.

De plus, près de vingt-cinq ans après la publication de *Science as a Process* la théorie de Hull exerce encore une influence au-delà des cercles restreints de la philosophie des sciences, comme en témoigne par exemple la publication de *Archaeology as a Process : Processualism and Its Progeny* (O'Brien *et al.*, 2007). Les auteurs de ce dernier ouvrage, des archéologues, tentent de voir comment le mécanisme décrit par Hull s'applique à leur domaine. Ils peuvent s'inspirer en cela de la démarche de Hull lui-même, qui consacre près de la moitié de *Science as a Process* à une étude semblable, centrée plutôt sur l'évolution des systèmes de classification en biologie depuis Darwin (même si on aimerait parfois que Hull soit plus explicite quant aux liens entre son mécanisme général d'évolution scientifique et les événements qu'il décrit dans son étude de cas).

Malgré une certaine postérité, Hull est le premier reconnaître le caractère programmatique (et inachevé) de sa démarche. La troisième partie de *Science and Selection* (2001), intitulée « Testing Our Views about Science », regroupe cinq essais qui constituent ensemble un plaidoyer bien senti pour approche « plus scientifique » de l'étude de la science.

Après avoir détaillé certaines de ses hypothèses sur les interactions entre différents éléments constitutifs du processus scientifique, Hull pose le problème de la façon suivante :

« The preceding observations about the interconnections between use, citation, credit, support and authority may sound plausible enough, but are they true? The first prerequisite for answering such questions is the formulation of explicit models. » (2001, p. 170)

Hull a en tête des modèles formels plus traditionnels, mais les modèles multi-agents tels que ceux qui sont décrits au chapitre précédent recueilleraient aussi son assentiment : le prérequis est qu'ils permettent d'examiner comment sont reliés entre eux les différents comportements que peuvent adopter les scientifiques et de voir comment les conditions initiales (les paramètres du modèle) influencent les résultats obtenus.

Les modèles prennent un intérêt particulier lorsqu'on les utilise pour pallier à la difficulté de réaliser des expériences contrôlées dans certains domaines, et les préoccupations normatives inhérentes à l'épistémologie donnent souvent lieu à de telles difficultés. Hull en est bien conscient :

« Philosophers have also traditionally expressed normative claims about science. In general, such prescriptions are extremely difficult to test. One suggestion is to convince groups of scientists. Have them adopt one's views about how science should be conducted and see what happens. If science in such areas grinds to a halt, then possibly something is wrong with one's normative claims. Conversely, if those scientists who adopt our views are even more successful in attaining their epistemic goals, then possibly there is something to be said for these norms. » (2001, p. 194)

Le modèle proposé au chapitre 5 permet de contourner, du moins dans un premier temps, la difficulté de tester *in vivo* les normes scientifiques : il est plus facile de « convaincre » d'adopter les comportements voulus les agents d'une simulation que de véritables scientifiques...et il sera plus facile de convaincre les scientifiques de les adopter si des modèles en montrent l'efficacité.

## CHAPITRE IV

### THÉORIES, MODÈLES ET SIMULATIONS

Dans le chapitre 3, nous avons parlé de la « théorie » de Hull. Dans le chapitre qui suit, nous allons proposer un « modèle » de cette théorie. Mais qu'entend-on par « théorie » et « modèle » ? Et quelle est la relation entre ces deux concepts ?

Nous avons attendu jusqu'au présent chapitre avant de tenter d'établir ces concepts sur des bases philosophiques plus rigoureuses parce qu'il nous paraissait nécessaire d'avoir bien exposé au préalable les objets auxquels nous souhaitons les appliquer.

Dans les chapitres 1 et 2, nous avons vu plusieurs exemples de modèles. Il s'agissait en l'occurrence de modèles multi-agents, mais ce n'est pas le seul type de modèles possible. Nous en mentionnerons quelques autres ci-dessous. Dans le chapitre 3, nous avons examiné les principaux éléments de la théorie de Hull, mais nous sommes restés assez vagues quant à son statut de « théorie ».

Nous avons besoin d'un ancrage philosophique qui nous permette de caractériser la théorie de Hull en tant que théorie, de caractériser notre modèle en tant que modèle de la théorie de Hull et d'établir le rôle et la valeur épistémique possible d'un tel modèle.

La philosophie des sciences a développé, au vingtième siècle, deux approches principales concernant la structure des théories scientifiques : l'approche syntaxique et

l'approche sémantique. Nous les aborderons tour à tour et verrons si elles peuvent s'appliquer à notre cas. Nous soutiendrons que l'approche sémantique, moyennant quelques ajustements, peut faire le travail nécessaire.

Il nous apparaît toutefois important de mentionner ce que nous ne recherchons *pas* dans cet ancrage philosophique : nous ne souhaitons pas prendre les concepts définis ici comme *objets* de notre modèle. En d'autres termes, nous ne souhaitons pas affirmer que les théories et/ou les modèles sont les répliqueurs dans le processus scientifique, même s'il est concevable que ce soit le cas. Hull reste explicitement agnostique quant aux unités de transmission qui sont opérantes en science, et c'est ce que nous ferons aussi, référant simplement à ces entités comme étant des « idées » (sachant que ces « idées » pourraient fort bien être des modèles, des théories, ou encore autre chose).

Cette précision étant faite, nous pouvons maintenant nous pencher sur la structure des théories scientifiques.

### L'approche syntaxique des théories

L'approche syntaxique a longtemps représenté l'orthodoxie en philosophie des sciences. C'est l'approche développée par les empiristes logiques (p. ex. Carnap, 1966 ; Hempel, 1966 ; Nagel, 1961). Elle reflète l'idée, chère à ceux-ci, que les problèmes philosophiques sont essentiellement des problèmes de langage et que la logique est l'outil privilégié pour aborder ces problèmes. Les éléments qui interviennent selon eux dans la structure d'une théorie sont des termes, des phrases et des langages.

Il y a trois types de termes. Les termes théoriques, les termes d'observation et les termes logiques. Les termes théoriques, ce sont les termes qui réfèrent aux concepts de la théorie : p. ex. « force », « accélération », « température », « molécule », « gène », « crédit », etc. Les termes d'observation, eux, renvoient à des qualités ou des relations qui peuvent être directement perçues par les sens : « rouge », « chaud », « plus haut que », etc. (La question de savoir où s'arrête l'observation et où commence la théorie

est une question délicate à laquelle ne nous nous attarderons pas ici.<sup>43</sup>) Finalement, les termes logiques, ce sont des connecteurs comme « et » ( $\wedge$ ), « ou » ( $\vee$ ), « implique que » ( $\rightarrow$ ) ou des quantificateurs comme « pour tout » ( $\forall$ ) et « il existe » ( $\exists$ ).

En combinant des termes logiques avec un ou les deux autres types de termes, on peut former des phrases (*sentences*). Si la phrase ne contient que des termes théoriques et des termes logiques, ce sera une phrase théorique ( $T_S$ ). Les phrases de ce type appartiennent au « langage théorique » ( $L_T$ ). Le langage théorique contient aussi les règles du système logique utilisé pour l'exprimer. Si, par exemple, une théorie est exprimée en logique des prédicats du premier ordre,  $L_T$  contient les axiomes de cette logique. Un langage théorique en bonne et due forme constitue un système axiomatisé permettant de dériver tous les énoncés de la théorie.

Une phrase qui ne contient que des termes logiques et des termes d'observation est une phrase d'observation ( $O_S$ ), ces phrases appartiennent à leur tour au langage d'observation ( $L_O$ ). Le lien entre  $L_T$  et  $L_O$  est établi par un autre type de phrase : les phrases de coordination ( $C_S$ ), qui en plus des termes logiques contiennent à la fois des termes théoriques et des termes d'observation. Le tableau 4.1, traduit et adapté de Winther (2016), résume la situation.

**Tableau 4.1**  
Structure des théories selon l'approche syntaxique

	<b>Théorie</b>		<b>Observation</b>
Phrases	$T_S$	$C_S$	$O_S$
Vocabulaire	Théorique et logique	Théorique, logique et observationnel	Observationnel et logique
Langage	$L_T$	$L_T$ et $L_O$	$L_O$

<sup>43</sup>Mentionnons seulement que le fait que l'approche syntaxique exige une telle distinction est vu comme une de ses faiblesses par les tenants de l'approche sémantique, p. ex. : « if the distinction can be drawn in a manner satisfactory for the purposes of the Received View, things will be exceedingly complex. The fact that science manages to go about its business without involving itself in such complexities suggests that the distinction is not really required or presupposed by science, and so is extraneous to an adequate analysis of scientific theories. » (Suppe, 1989, p. 62)

Le cœur de la théorie, c'est  $T_S$ . Et tant qu'on s'en tient à  $T_S$ , on est dans le domaine de la pure syntaxe : on peut dériver les conséquences de la théorie sans se soucier de la façon dont celle-ci est liée au monde empirique. Mais puisqu'il s'agit de science et non pas de logique pure, la théorie n'est complète que si elle établit explicitement un lien avec le monde : c'est le rôle de  $C_S$ , qui donne à la théorie sa sémantique.

Notons ici un point : selon l'approche syntaxique, une théorie n'est ni plus ni moins que l'ensemble des phrases théoriques et des phrases de coordination qui la constituent, c'est-à-dire  $\{T_S; C_S\}$ . Si on change une de ces phrases, il ne s'agit plus de la même théorie.

Un autre point à mentionner, et celui-ci nous concerne de plus près, c'est que la notion de modèle n'apparaît pas dans le portrait que nous avons dressé ci-dessus. Cela ne veut pas dire que la notion de modèle soit incompatible avec l'approche syntaxique des théories, mais elle y joue un rôle assez périphérique.

En logique, dans un sens très général, un modèle est une structure qui rend vrais les énoncés d'une théorie : « A possible realization in which all valid sentences of a theory  $T$  are satisfied is called a model of  $T$ . » (Tarsky, 1953, p. 11) Pour prendre un exemple trivial, si on a une théorie  $T$  constituée d'un seul énoncé  $(\forall p, q) (p \rightarrow q)$ , alors l'ensemble de paires  $\{\langle V, V \rangle; \langle F, V \rangle\}$  peut être un modèle de  $T$ , mais pas l'ensemble  $\{\langle V, F \rangle; \langle F, V \rangle\}$ . Notez bien que je dis « *peut* être un modèle de  $T$  » : la relation établie entre  $T$  et son modèle fait intervenir une interprétation particulière, celle selon laquelle les variables  $p$  et  $q$  réfèrent respectivement au premier et au deuxième élément de chacune des paires de valeurs de vérité qui sont les éléments du modèle. Bref, pour être associée à un modèle, une théorie a besoin d'une sémantique. Nous avons vu plus haut que ce sont les énoncés de coordination,  $C_S$ , qui ont pour tâche de donner à une théorie sa sémantique. On peut imaginer que nos paires de valeurs de vérité dans l'exemple ci-dessus correspondent à des états du monde observables : par exemple, à des voyants lumineux sur un appareil de laboratoire. On aurait alors un « modèle » constitué de phrases d'observation, et ce modèle serait ou non un modèle de  $T$  selon que les états observés respectent ou non les contraintes établies par la théorie (si le premier voyant est allumé, le second doit l'être aussi).

Dans ce sens, on peut aussi dire qu'une théorie est vraie si le monde est un modèle de celle-ci. Et l'approche syntaxique suppose que c'est le but recherché par  $C_S$  : faire du monde un modèle de la théorie. Mais qu'en est-il si on souhaite interpréter une théorie à l'aide d'un système autre que le monde qu'elle est originalement censée décrire ? Il faut alors une version alternative de  $C_S$  et dès lors il ne s'agit plus, au sens strict, de la *même* théorie. Une théorie ainsi modifiée pour accommoder différents modèles peut avoir une valeur heuristique ou pédagogique, mais elle s'écarte de son essence. La reconstruction rationnelle des théories prônée par l'approche syntaxique se soucie de la correspondance entre une théorie et le monde, pas de celle entre une théorie et ses modèles.

Ce rôle limité réservé aux modèles dans l'approche syntaxique l'éloigne de la pratique scientifique (Bailer-Jones, 2009). Mais ce n'est pas la seule façon dont l'approche syntaxique s'éloigne de la pratique scientifique. Patrick Suppes, critique de l'approche syntaxique et pionnier de l'approche sémantique, affirme que :

« it is unheard of to find a substantive example of a theory actually worked out as a logical calculus in the writings of most philosophers of science. Much handwaving is indulged in to demonstrate that this working out of the logical calculus is simple in principle and only a matter of tedious detail, but concrete evidence is seldom given. » (Suppes, 1967, p. 56)

Suppes ne mâche pas ses mots, mais il touche à un point important. Bien qu'il existe quelques tentatives d'axiomatisation de théories particulières (p. ex., Reichenbach (1924) pour la théorie de la relativité), celles-ci sont rares. Quand elles existent, elles sont le fait d'armes de quelque philosophe particulièrement motivé. La pratique scientifique courante ne s'embarrasse pas d'axiommatisations logiques.

Et dans le cas des systèmes complexes pour lesquels on tend à avoir recours aux modèles multi-agents et à la simulation, le type d'axiomatisation prôné par l'approche syntaxique n'est pas que laborieux : il est souvent mathématiquement insuffisant :

« In most cases, the equations that form the theoretical basis of systems of interest to simulationists are analytically unsolvable. That is, there is no mathematically expressible function that is the solution to these equations. Therefore, representations of these systems cannot possibly come in the form of linguistic entities derivable from a linguistic

theoretical structure. The syntactic view is clearly not going to do the work that we need here. » (Winsberg, 2001, p. S448)

Ce n'est pas la seule faille de l'approche syntaxique<sup>44</sup>, mais c'est celle qui revêt la plus grande importance pour les questions qui nous concernent. Dans la section suivante, nous allons nous pencher sur la principale alternative à l'approche syntaxique, soit l'approche sémantique, et tenter de voir si elle peut mieux convenir à nos objectifs.

### L'approche sémantique des théories

L'approche sémantique des théories renverse complètement la relation entre théorie et modèles. Selon elle, les modèles sont constitutifs de la théorie : une théorie n'est rien d'autre que l'ensemble des modèles qui la composent. Cela ne veut pas dire qu'une axiomatisation de la théorie n'est ni possible ni souhaitable. Cela veut dire, cependant, qu'une telle axiomatisation a perdu son statut essentiel : une autre axiomatisation qui serait vraie pour tous les modèles de la théorie serait une axiomatisation de la même théorie. Dans les mots de van Fraassen :

« The syntactic picture of a theory identifies it with a body of theorems, stated in one particular language chosen for the expression of that theory. This should be contrasted with the alternative of presenting a theory in the first instance by identifying a class of structures as its models. In this second, semantic, approach the language used to express the theory is neither basic nor unique ; the same class of structures could well be described in radically different ways, each with its own limitations. The models occupy centre stage. » (1980a, p. 44)

Un des effets les plus désirables de ce renversement, c'est qu'il ouvre tout grand la porte à l'utilisation de l'entièreté des mathématiques pour procéder à cette axiomati-

---

<sup>44</sup>Suppe (2000, p. S103) recense les principales critiques faites à l'approche syntaxique : « (1) its observational-theoretical distinction was untenable ; (2) correspondence rules were a heterogeneous confusion of meaning relationships, experimental design, measurement, and causal relationships some of which are not properly parts of theories ; (3) the notion of partial interpretation associated with more liberal correspondence rules was incoherent ; (4) theories are not axiomatic systems ; (5) symbolic logic is an inappropriate formalism ; (6) theories are not linguistic entities and thus theories are individuated incorrectly. »

sation. Van Fraassen (1980b, p. 675) attribue à Suppes le slogan selon lequel : « mathematics, and not metamathematics, is the appropriate tool for philosophers of science ». La logique des prédicats du premier ordre n'est pas un outil facile à manier quand vient le temps de formaliser une théorie scientifique. Plus besoin, maintenant, de s'y limiter.

Un autre des effets de ce renversement, c'est que la relation entre la théorie et le monde n'a plus à être établie à l'aide d'un ensemble d'énoncés de coordination (le  $C_S$  de la section 4.1). Ce sont maintenant les modèles de la théorie qui se tiennent en relation entre la théorie et le monde et donnent à celle-ci son pouvoir représentationnel. (Cela ne veut pas dire, pour autant, que la question de la représentation soit réglée. Nous y reviendrons ci-dessous.)

Cette mise à l'avant-plan des modèles dans l'approche sémantique exige que l'on sache de quoi on parle quand on parle de modèle. Pour les philosophes qui ont développé l'approche sémantique, il s'agit de structures mathématiques. Suppes (1957 ; 1960 ; 1962) privilégie l'utilisation de la théorie des ensembles pour la construction de modèles, mais ce n'est pas la seule possibilité. Van Fraassen et d'autres (p. ex. Beth, 1948, 1949, 1960 ; Lloyd, 1984, 1994) préfèrent caractériser les modèles en termes de trajectoires dans un espace d'état. Nous aborderons brièvement ces deux possibilités avant de nous pencher sur la notion de modèle computationnel proposée par Weisberg (2013).

### Les modèles comme structures ensemblistes

Les modèles ensemblistes mis de l'avant par Suppes respectent la définition de Tarsky (1953) mentionnée plus haut : ils sont des structures qui peuvent constituer la « réalisation » d'une théorie, c'est-à-dire rendre vrais tous ses énoncés. Pour Suppes, un modèle est constitué des éléments suivants :

« an ordered  $(n + 2)$ -tuple  $\mathfrak{M} = \langle \mathfrak{S}, A, m_1, \dots, m_n \rangle$ , where  $\mathfrak{S}$  is the usual system of real numbers under the operations of addition and multiplication and the relation less than [...],  $A$  is a finite, non-empty set, and  $m_1, \dots, m_n$ , are real-valued functions on  $A$ . » (1965, p. 365)

En bref : il s'agit d'un ensemble d'objets (censés correspondre à des objets du monde physique) et diverses fonctions qui associent à ces objets des valeurs numériques en utilisant les ressources de l'arithmétique des nombres réels. Derrière cette relative simplicité se cache un pouvoir expressif considérable. Les fonctions  $m_1, \dots, m_n$  permettent de décrire les différentes propriétés des objets de  $A$  et, supposant qu'elles aient pour argument une variable  $t$ , leur évolution au fil du temps.

Un exemple classique de ce genre de structure est le modèle de la mécanique des particules développé par Suppes (1957, ch. 12) :

« We may axiomatize classical particle mechanics in terms of the five primitive notions of a set  $P$  of particles, an interval  $T$  of real numbers corresponding to elapsed times, a position function  $s$  defined on the Cartesian product of the set of particles, the time interval and the set of positive integers (the set of positive integers enters into the definition of the force function simply in order to provide a method of naming the forces). A possible realization of the axioms of classical particle mechanics, that is, of the theory of classical particle mechanics, is then an ordered quintuple  $\varphi = \langle P, T, s, m, f \rangle$ . A model of classical particle mechanics is such a quintuple. » (Extrait de Suppes, 1960, p. 6)

Dans ce modèle, l'ensemble des particules  $P$  constitue l'ensemble des objets  $A$ , l'intervalle de nombres réels  $T$  est un sous-ensemble de  $\mathbb{S}$  et les fonctions  $s$ ,  $m$  et  $f$  correspondent à  $m_1$ ,  $m_2$  et  $m_3$ .

Peut-on utiliser une telle structure pour décrire un modèle multi-agents ? De prime abord, cela semble être le cas. Prenant par exemple le jeu de la vie de Conway, puisque c'est le plus simple des modèles que nous avons vus au chapitre 1. Rappelons ses règles :

1. **Survie.** Chaque jeton ayant deux ou trois jetons voisins survit pour la génération suivante.
2. **Mort.** Chaque jeton ayant quatre jetons voisins ou plus meurt de surpopulation. Chaque jeton ayant un seul ou aucun jeton voisin meurt d'isolation.
3. **Naissance.** Chaque cellule vide adjacente à exactement trois jetons—ni plus, ni moins—est le lieu d'une naissance. Un nouveau jeton est placé sur elle lors de la génération suivante.

Pour les besoins de la formalisation, nous prendrons comme objets de notre modèle du jeu de la vie les cellules de la grille plutôt que les jetons dont parle la description verbale du modèle<sup>45</sup>. Nous considérerons qu'une cellule est dans l'état 1 s'il y a un jeton qui s'y trouve et qu'elle est dans l'état 0 si elle est vide. Notre modèle comporte les éléments suivants :

- ▷  $G = \langle V, E \rangle$ , un graphe représentant notre grille de jeu.
  - ▷  $V$ , l'ensemble des nœuds (*vertices*) du graphe, représente les cellules de la grille.
  - ▷  $E$ , l'ensemble des liens (*edges*) entre ces nœuds, représente les relations d'adjacence entre les cellules de la grille, où  $E \subseteq \{\{u, v\} \mid u, v \in V\}$ .
- ▷  $i : V \rightarrow \{0; 1\}$ , une fonction, définie en extension, qui retourne l'état initial (0 ou 1) d'une cellule.
- ▷  $p : V \times \mathbb{N} \rightarrow \mathbb{N}$ , une fonction qui pour une cellule  $v$  à un temps  $t$ , retourne la « population » autour de  $v$  au temps  $t - 1$ , c'est-à-dire la somme des zéros et des uns dans les cellules de son voisinage (ces valeurs étant obtenues par la fonction  $s$ , définie plus loin) :

$$p(v, t) = \sum_{u \in \{u \mid \{u, v\} \in E\}} s(u, t - 1) \quad (4.1)$$

- ▷  $s : V \times \mathbb{N} \rightarrow \{0; 1\}$ , une fonction qui pour une cellule  $v$  à un temps  $t$ , retourne l'état de  $v$  : zéro ou un. On peut définir cette fonction ainsi :

$$s(v, t) = \begin{cases} i(v) & \text{si } t = 0, \\ 1 & \text{si } (p(v, t) = 3) \vee ((s(v, t - 1) = 1) \wedge (p(v, t) = 2)), \\ 0 & \text{sinon.} \end{cases} \quad (4.2)$$

---

<sup>45</sup>C'est la règle de la « naissance » qui rend nécessaire la formalisation de la grille elle-même.

Nous avons là une structure  $\langle \mathfrak{S}, G, i, p, s \rangle$  qui constitue une description complète du jeu de la vie dans les termes de la théorie des ensembles. Il semblerait donc que l'approche sémantique telle que développée par Suppes soit en mesure d'accueillir en son sein au moins une certaine classe de modèles multi-agents (en l'occurrence, les automates cellulaires).

Il y a pourtant une remarque à faire avant de se réjouir. Notre description recèle une certaine complexité computationnelle. La fonction  $s$  est récursive : elle fait appel à elle-même. De plus,  $s$  et  $p$  sont co-récessives :  $s$  fait appel à  $p$  et  $p$  fait appel à  $s$ . Ce n'est pas un problème en théorie : la récursion est un des fondements de la computation. C'est toutefois un problème en pratique : si le jeu de la vie était implémenté dans ces termes à l'aide d'un langage informatique qui supporte la récursion, les ressources computationnelles requises seraient telles que seules de très petites grilles pourraient être simulées, et seulement pour un temps très court. Imaginons une grille toroïdale de  $3 \times 3 = 9$  cellules. À  $t = 0$ , la fonction  $s$  va, pour chaque cellule, puiser son résultat directement dans  $i$ , qui est définie en extension. C'est neuf opérations. À  $t = 1$ , pour chaque cellule,  $s$  doit obtenir l'état de cette cellule et l'état de ses huit voisines au temps  $t = 0$ . C'est donc  $9 \times 9 = 81$  opérations. À  $t = 2$ , les neuf cellules doivent obtenir chacune neuf états au temps  $t = 1$ , ce qui demande  $9 \times 9 \times 9 = 729$  opérations. Bref, pour un nombre  $n$  de cellules au temps  $t$ , la complexité computationnelle est au moins<sup>46</sup> de  $O(n^{t+1})$ . À cela s'ajoute le problème de la profondeur de la récursion, susceptible de causer une erreur de dépassement de pile (*stack overflow*).

Du point de vue de l'approche sémantique telle que conçue à l'origine, rien de tout cela n'a d'importance : ce qui constitue le modèle, c'est la structure mathématique mathématique abstraite décrite par  $\langle \mathfrak{S}, G, i, p, s \rangle$ , et non la façon dont elle est décrite. Il serait possible, au prix de quelques contorsions syntaxiques, de reformuler les fonctions ci-dessus d'une façon qui réduise leur coût computationnel, mais sur le plan purement sémantique, il s'agirait toujours du *même* modèle.

---

<sup>46</sup>Nous avons ci-dessus fait abstraction du coût computationnel lié à la construction de l'ensemble  $\{u \mid \{u, v\} \in E\}$  qui constitue le voisinage d'une cellule  $v$ .

Et pourtant, si ce n'était de la possibilité d'implémenter efficacement le jeu de la vie sur un ordinateur, il ne serait qu'une curiosité mathématique obscure dont on ne saurait que faire. Ce qui fait qu'il reste une source de fascination inépuisable depuis maintenant près de 50 ans, c'est qu'on peut expérimenter avec le modèle : changer ses conditions initiales et observer (voir !) les *patterns* qui en émergent.

Paul Humphrey, qui tente de développer une notion de modèle computationnel qui se veut complètement indépendante de l'approche sémantique, affirme que :

« The importance of syntax to applications, and especially to computational tractability, is something that the semantic account of theories, for all its virtues, is essentially incapable of capturing. This is because one of the main points of the semantic view is to abstract from the particular syntactic representation and to identify the theory with an underlying abstract structure. Important as the semantic approach is for baring the underlying abstract structure of a theory, it moves us in exactly the opposite direction from where we should be when we apply computational models. Syntax matters. » (Humphreys, 2002, p. S3)

Nous sommes d'accord avec Humphrey sur un point : la syntaxe est importante. Ou du moins : elle l'est dans la mesure où elle influence les propriétés computationnelles du modèle. Mais contrairement à ce qu'il affirme, il nous semble que l'idée de modèle computationnel n'est pas incompatible avec l'approche sémantique. Nous reviendrons sous peu à cette question.

Pour l'instant, nous allons plutôt nous pencher sur le second type de structure mathématique qui peut traditionnellement constituer un modèle selon l'approche sémantique. Van Fraassen (1970) reprend de Beth (1948, 1949, 1960) l'idée que celle-ci peut être un espace d'états<sup>47</sup>.

### Les modèles comme trajectoires dans un espace d'états

Dans cette conception, chacune des variables d'un modèle correspond à une dimension dans un espace mathématique<sup>48</sup>. L'état d'un modèle à  $n$  variables peut donc être représenté par un vecteur de  $n$  éléments qui correspond à un point dans l'espace d'états à  $n$  dimensions du modèle. Spécifier un modèle, c'est donc à la fois spécifier les dimen-

sions de l'espace d'état du modèle et spécifier quelles sont les positions possibles pour le modèle dans cet espace (Lloyd, 1994, p.19–20).

Pour prendre un exemple des plus simples, imaginons un modèle à deux variables,  $x, y \in \mathbb{N}$ . L'espace d'états d'un tel modèle est un plan cartésien à deux dimensions. Si on spécifie ensuite que les états possibles du modèle sont décrits par l'équation  $y = 2x$ , les points  $\{(0, 0); (1, 1); (2, 4); (3, 8); (4, 16); \dots\}$  représenteront les différents états du modèle dans cet espace.

On peut aussi conceptualiser les états d'un modèle plus complexe comme des points dans un espace à  $n$  dimensions. Reprenons notre exemple du jeu de la vie avec une grille de  $3 \times 3 = 9$  cellules, l'état de chacune d'elle étant représenté par 0 ou 1. On a un espace à neuf dimensions, chacune de ces dimensions n'admettant que 0 ou 1 comme position, et donc  $2^9 = 512$  points dans cet espace, chacun d'eux correspondant à un état possible de notre mini jeu de la vie. Mais un modèle dynamique comme le jeu de la vie ne spécifie pas seulement quels sont les états possibles du système, il spécifie aussi quels états peuvent succéder à quels autres états. Ainsi, selon la règle de la « mort » (voir p. 99), seul l'état  $(0, 0, 0, 0, 0, 0, 0, 0, 0)$  peut succéder à l'état  $(0, 0, 0, 0, 1, 0, 0, 0, 0)$ . Ce qui constitue notre modèle, c'est donc un ensemble de *trajectoires* possibles dans un espace d'états (voir van Fraassen, 1970, p. 330–331).

On remarquera que le nombre de dimensions dans l'espace d'états augmente avec le nombre d'objets représentés par le modèle et le nombre de variables décrivant l'état de chacun de ces objets. Dans un modèle à  $n$  objets ayant chacun  $m$  variables, on aura au moins  $n \times m$  dimensions, en plus des dimensions nécessaires pour représenter les variables globales du modèle. Cette explosion dimensionnelle concerne tout particulièrement les modèles multi-agents, puisque l'intérêt de ceux-ci consiste justement à

---

<sup>46</sup>Quand on a affaire à des variables continues, on tend à parler d'« espace des phases » plutôt que d'« espace d'états », mais le concept est similaire. Puisque les modèles computationnels qui nous intéressent ici sont, par nature, discrets, nous nous en tiendrons à l'expression « espace d'états ».

<sup>48</sup>La notion d'espace d'état est conceptuellement similaire à la notion de paysage épistémique abordée à la section 2.5, mais il faudra se garder de confondre les deux : dans le modèle que nous allons définir au chapitre 5, chaque « idée » instanciée dans le modèle ajoutera chacune des dimensions de l'espace épistémique dans lequel elle se situe à l'espace d'état global du modèle.

représenter des systèmes dont les entités sont nombreuses. De plus (contrairement à notre exemple simplifié d'un jeu de la vie à seulement neuf cellules), les modèles multi-agents tendent à ne pas imposer de limites préalables quant au nombre d'agents qui seront représentés dans le modèle. Dans le modèle de Gilbert (1997) présenté à la section 2.1, par exemple, le nombre d'articles publiés et le nombre d'auteurs actifs dans le système augmentent au fil de la simulation. Ce sera aussi le cas pour le modèle que nous développerons au chapitre 5.

Peut-on concevoir l'espace d'état de tels modèles comme possédant suffisamment de dimensions pour accommoder le nombre maximum d'objets qui peuvent être représentés par le modèle ? Conceptuellement, cela ne pose pas problème : un espace à  $n$  dimensions reste le même genre d'objet mathématique, que  $n$  soit grand ou petit. En pratique, toutefois, notre capacité à explorer l'espace d'états de ces modèles complexes est limitée. Le langage NetLogo, par exemple, supporte techniquement des ensembles pouvant contenir jusqu'à  $2^{31} - 1$  agents (soit un peu plus de deux milliards), mais un ordinateur personnel, limité en mémoire et en rapidité de traitement, peine à représenter plus de quelques milliers d'agents. À part pour les modèles multi-agents les plus triviaux (notre jeu de la vie à neuf cellules, par exemple), l'exploration complète de l'espace d'états est hors de notre portée. Les expériences présentées au chapitre 6, par exemple, ont exigé chacune plusieurs jours de traitement et n'explorent, au bout du compte, qu'une infime partie de l'espace d'états du modèle.

Cette situation n'est pas tant une limitation de l'approche par espaces d'état qu'une conséquence inévitable de la complexité des systèmes étudiés. La gestion de la taille de l'espace d'état est d'ailleurs un des principaux défis auquel fait face toute personne qui tente de concevoir un modèle multi-agents.

En bref : s'il est vrai qu'en définissant un modèle mathématique, on définit par le fait même un ensemble de trajectoires possibles dans un espace d'état, dans le cas de la plupart des modèles computationnels, il y a un prix à payer pour accéder à cet espace.

D'un point de vue pragmatique, ce qui fait la valeur d'un modèle scientifique, ce n'est pas que la structure mathématique abstraite qui est définie par le modèle, mais aus-

si ce que l'on peut *faire* avec cette structure. Nous croyons que cette exigence n'est pas incompatible avec l'esprit de l'approche sémantique. Une des principales motivations de cette approche, à l'origine, était de donner à l'analyse philosophique des outils formels appropriés à la façon dont les modèles sont concrètement utilisés en science.

La démonstration de l'utilité des structures ensemblistes a été faite pour plusieurs sciences (*p. ex.*, la mécanique newtonienne (Suppes, 1957), la psychologie de l'apprentissage (Suppes, 1962)). C'est aussi le cas pour l'approche par espace d'états (*p. ex.*, la physique quantique (van Fraassen, 1970), la génétique des populations (Lloyd, 1984)). Dans la section qui suit, nous allons proposer d'ajouter à l'arsenal de l'approche sémantique un autre type de structure mathématique, élargissant par le fait même l'éventail des modèles dont elle peut rendre compte, que ce soit dans les sciences susmentionnées ou pour toute autre science qui utilise la simulation pour étudier des systèmes complexes.

### Les modèles comme structures computationnelles

Le type de structure que nous souhaitons ajouter, ce sont des structures computationnelles. Pour être plus précis :

« the relevant kind of computational structure is an *algorithm*, which is a set of instructions for carrying out a procedure. Computational models are thus sets of procedures that take a starting state as an input and specify how this state changes, ultimately yielding an output. » (Weisberg, 2013, pp. 29–30)

Weisberg identifie trois caractéristiques des structures computationnelles qui les rendent particulièrement appropriées pour modéliser certains systèmes difficiles à modéliser autrement :

- Elles permettent d'exprimer des règles de transition conditionnelles.
- Elles permettent d'exprimer des règles de transition probabilistes.
- Elles permettent de représenter des processus parallèles.

Rien de tout cela n'est impossible à représenter à l'aide de structures ensemblistes ou de trajectoires dans un espace d'états, mais une approche computationnelle permet d'exprimer ces changements d'état dans un langage relativement direct, et par conséquent facile à comprendre et à manipuler. À titre d'exemple (du moins en ce qui concerne les règles de transition conditionnelles), comparez la description du jeu de la vie en termes ensemblistes donnée à la section 4.2.1 à son équivalent computationnel en NetLogo (code 4.1).

---

```

patches-own [ n ]

to setup [ live-patches ]
  clear-all
  ask patches [ set pcolor white ]
  ask live-patches [ set pcolor black ]
end

to go
  ask patches [ set n count neighbors with [ pcolor = black ] ]
  ask patches [
    if n = 3 [ set pcolor black ]
    if n < 2 or n > 3 [ set pcolor white ]
  ]
end

```

---

**Code 4.1** Le jeu de la vie en NetLogo

Il y a une part de subjectivité qui intervient ici, mais pour quiconque est familier avec le langage, il s'agit là d'une description claire et simple<sup>49</sup>. Et comme l'approche sémantique l'a montré elle-même, il y a des avantages à formaliser nos théories d'une façon qui s'approche de celle qui est utilisée dans le cadre de la pratique scientifique.

De plus, le code 4.1 offre une preuve constructive de l'existence d'un algorithme qui réalise le jeu de la vie avec une complexité computationnelle linéaire, soit  $O(n)$ , où  $n$  est le nombre de cellules. L'existence d'un algorithme efficace est une condition

---

<sup>49</sup>Quand on décrit un modèle en NetLogo (ou dans tout autre langage de programmation), on ne part pas de zéro. Certains éléments du monde de NetLogo sont définis pour nous : on a accès à une grille de cellules (*patches*) qui sont initialisées pour nous. Ces cellules ont des propriétés que nous pouvons utiliser pour implémenter notre modèle : par exemple, nous avons adopté ici la convention selon laquelle une cellule blanche est « vivante » et une cellule noire est « morte ». Cette ontologie préalable contribue à la concision du modèle.

nécessaire à l'utilisation du modèle. Et même lorsqu'un tel algorithme est disponible, on ne peut souvent explorer qu'une infime partie de son espace d'états. Eric Winsberg souligne le fait que cet état de choses rapproche les modèles computationnels des méthodes expérimentales :

« Numerical methods, on the other hand, result in a big pile of numbers. If simulationists want to learn about the general qualitative features of a class of systems, then they must apply all the usual tools of experimental science for analyzing data : visualization, statistics, data mining, etc. If they want to discover functional dependencies, then they must also run a barrage of trials, looking at the results across a wide range of parameters. It is without a doubt this aspect of simulation that carries the most obvious methodological characteristics of experimental work. » (Winsberg, 2003)

Notre chapitre 6 illustrera concrètement ce fait. Pour l'instant, le point à retenir, c'est que les propriétés qui dérivent de la description d'un modèle en termes computationnels ont un impact significatif sur le travail scientifique qui peut être accompli à l'aide d'un modèle.

Paul Humpreys va encore plus loin et soutient que la syntaxe utilisée pour décrire un modèle computationnel joue un rôle crucial :

The dominant logical tradition has emphasized propositional representations in the syntactic approach, or set-theoretical representations in the semantic approach, simply because, in principle, any data can be thus represented. Yet this 'in principle' argument ignores the epistemic impossibility of humans' assimilating the information contained in millions of data points from astrophysical or biological instruments when the data are presented propositionally. (2004, p. 99)

Nous sommes d'accord avec Humphrey sur le fait que la représentation joue un rôle cognitif important. Il nous faut toutefois insister sur le fait que la description d'un modèle ne constitue pas le modèle lui-même. Selon Giere (1988, p. 83), la relation entre un ensemble d'énoncés et un modèle est une relation de « définition » : les énoncés définissent le modèle, qui est une structure abstraite. Cela veut dire qu'un même modèle peut admettre une infinité de descriptions différentes, dans une infinité de langages différents. De même, une description, si elle est ambiguë (comme c'est souvent le cas), peut décrire plusieurs modèles possibles. Dans les termes de Weisberg (2013, p. 35), une description *spécifie* un ou plusieurs modèles et un modèle *réalise* une infinité de descriptions.

Tout modèle computationnel peut théoriquement être spécifié en termes de machines de Turing (1936) ou du calcul lambda de Church (1936)<sup>50</sup>. Nous savons que ces modèles de la computation sont équivalents : tout ce qui peut être calculé par une machine de Turing peut aussi l'être par le calcul lambda, et vice versa (voir Kleene, 1967, p. 232). Les raisons de préférer une description à une autre sont pragmatiques. Dans le cas des modèles computationnels, la description coïncide généralement avec l'implémentation du modèle : décrire le modèle, c'est par le fait même écrire un programme qui peut être exécuté par un ordinateur.

Il faut ici dire quelques mots sur la distinction entre « modèle » et « simulation ». Dans un article de 1990, après avoir cité plusieurs alternatives<sup>51</sup>, Humphrey propose la définition suivante :

« A computer simulation is any computer-implemented method for exploring the properties of mathematical models where analytic methods are unavailable. » (Humphreys, 1990, p. 501)

Stephan Hartmann a deux critiques à formuler envers la définition de Humphrey. Premièrement, celle-ci ne tient pas compte du caractère dynamique de la simulation : une simulation est quelque chose qui se déroule dans le temps. Quand on parle de modèles statiques, même si on utilise un ordinateur pour en dériver les propriétés, on ne parle pas de « simulation ». Deuxièmement, on peut très bien utiliser la simulation même lorsque des méthodes analytiques sont disponibles. Il propose plutôt la définition suivante :

« *a simulation imitates one process by another process*. In this definition, the term “process” refers solely to some object or system whose state changes in time. If the simulation is run on a computer, it is called a *computer simulation*. » (Hartmann, 1996, p. 83, italiques originales)

Notons que la définition de Hartmann laisse ouverte la possibilité que le système simulé soit un système hypothétique (voir aussi Winsberg, 2015, § 1.1). Qu'il s'agisse d'un système réel ou d'un système hypothétique, la simulation suppose l'existence d'un

---

<sup>50</sup>Les machines de Turing et le calcul lambda sont les modèles de la computation les plus connus, mais ce ne sont pas les seuls (voir Savage, 1998 ; Sipser, 2006 ; Fernández, 2009).

<sup>51</sup>Ord-Smith et Stephenson, 1975, p. 3 ; Bennett, 1974, p. 2 ; Neelamkavil, 1987, p. 1 ; Reddy, 1987, p. 162.

modèle computationnel de ce processus : un modèle qui décrit la façon dont l'état du système change au fil du temps. Pour une simulation sur ordinateur, on a nécessairement un programme qui décrit ce modèle dans un langage informatique. Lorsque ce programme est *exécuté*, le processus physique qui a lieu dans l'ordinateur imite le système modélisé : certains changements d'état dans la mémoire de l'ordinateur correspondent aux changements d'état du système.

La principale différence entre les modèles computationnels et les structures ensemblistes ou les trajectoires dans des espaces d'états sur lesquelles l'approche sémantique s'est traditionnellement concentrée, c'est que les modèles computationnels spécifient *comment* les choses sont calculées en plus de spécifier *ce qui* est calculé. En définissant un modèle computationnel, on définit *par le fait même* une structure ensembliste et un ensemble de trajectoires dans l'espace d'état du modèle. Nous ajoutons simplement l'exigence (non triviale, dans le cas des systèmes complexes) de fournir un algorithme pour calculer ces trajectoires de façon efficiente.

Qu'est-ce qui constitue un modèle de la théorie de Hull ?

Selon l'approche sémantique des théories, que nous avons décidé d'adopter, une théorie est constituée d'un ensemble de modèles. Si la théorie de Hull est bien une théorie, il faut donc qu'elle définisse un ou plusieurs modèles.

Hull ne nous offre pas explicitement de modèle formel. Comme nous l'avons vu au chapitre 3, il décrit un mécanisme. Cette description, si tant est qu'il soit possible de la traduire en un modèle formel, constitue ce que Winther (2006) appelle un « modèle verbal » ou un « modèle narratif ». Dans les mots de Michael Weisberg :

« Such models are presented in narrative form, but usually involve verbally sketching some possible mechanism and explaining how the behavior of this mechanism could account for some real world phenomenon. [...] However, no mathematics is used to formulate this model and nothing concrete was constructed. » (Weisberg, 2013, p. 17)

Pour Weisberg, de tels modèles ne constituent pas une catégorie distincte. Une description narrative peut autant spécifier des modèles physiques concrets que des modèles

mathématiques abstraits (incluant les modèles computationnels). Rappelons-nous que la relation entre description et modèle est une relation d'un à plusieurs : une description peut-être réalisée par plusieurs modèles différents. Il y a toute une famille de modèles qui peuvent tomber sous la coupe d'une description narrative comme celle de Hull.

Nous croyons que le modèle de la science proposé par Hull est un modèle computationnel. Le mécanisme qu'il décrit repose sur un algorithme, et les propriétés computationnelles de cet algorithme sont un aspect constitutif de la science. La description donnée par Hull dans *Science as a process* et *A mechanism and its metaphysics* est de haut niveau, mais elle établit tout de même un ensemble de contraintes que doit respecter un modèle qui cherche à la réaliser. Nous croyons que les principales contraintes posées par Hull sont les suivantes :

- Le modèle doit faire intervenir des répliqueurs et des interacteurs possédant les propriétés décrites pour ce type d'entités à la section 3.2.1.
- Les répliqueurs, identifiés comme « éléments du contenu substantiel de la science » (section 3.2.2), doivent être différenciés quant à leur valeur objective.
- Les interacteurs, identifiés comme étant les scientifiques, ont pour objectif premier d'obtenir du « crédit » auprès des autres scientifiques (section 3.3.1).
- Le véhicule principal pour l'attribution du crédit est la citation dans un article scientifique (section 3.3.1).

Nous verrons en détail au chapitre 5 comment nous interprétons ces contraintes en termes de modèle computationnel. Pour juger de la qualité de notre interprétation, il faudra prendre un à un chacun des éléments de notre modèle et voir comment ceux-ci se rattachent à la théorie de Hull. Nous avons, par exemple, choisi de modéliser les répliqueurs comme des points dans un espace multidimensionnel, assignant à une de ces dimensions le rôle de « valeur objective » de l'idée correspondante (section 5.2). Est-ce une interprétation raisonnable de la notion de répliqueur chez Hull ? Nous tenterons de faire valeur que oui, et de montrer comment les « idées », définies ainsi,

ont les propriétés nécessaires pour fonctionner comme répliqueurs dans le système scientifique. Et ainsi de suite pour chacun des éléments du modèle que nous allons développer. La section 5.5 résume les correspondances entre notre modèle et la théorie de Hull.

Et, quel que soit le verdict que l'on porte sur la qualité de notre interprétation, une chose est sûre : elle n'est pas la seule possible. La description de Hull est suffisamment générique pour ouvrir la porte à une infinité de modèles différents.

Cette situation n'est pas que l'apanage des descriptions narratives. La tâche de construire un modèle computationnel n'est jamais triviale. Même dans le cas d'une théorie qui se penche sur un système physique bien défini, qui peut être décrit, par exemple, avec des équations différentielles non linéaires, les contraintes computationnelles et formelles liées à la simulation exigent de nombreuses décisions de la part du modélisateur. Eric Winsberg va jusqu'à dire que la simulation est la « progéniture bâtarde » de la théorie :

« Successful numerical methods, therefore, invariably require of the simulationists that they transform the model suggested by theory in significant ways. Idealizations, approximations, and even self-conscious falsifications are introduced into the model. In the end, the model that is used to run the simulation is an offspring of the theory, but it is a mongrel offspring. It is also substantially shaped by the exigencies of practical computational limitations and by information from a wide range of other sources. » (Winsberg, 2003, p. 108)

Il rejoint en cela Nancy Cartwright, qui critique la tendance (inhérente à l'approche syntaxique, mais à laquelle l'approche sémantique n'échappe pas toujours) à voir les théories comme des « machines distributrices » dont on pourrait tirer sans effort les modèles recherchés (1999, p. 184)<sup>52</sup>. Cartwright souligne que la relation entre une théorie et ses modèles n'est pas une relation de déduction.

Comment savoir, alors, si les décisions prises dans le cadre de la modélisation sont les bonnes ? Pour Winsberg, il n'est pas possible d'établir directement la valeur épistémologique de la simulation. Il soutient (1999 ; 2001 ; 2009) qu'une épistémologie de la simulation (qui reste à construire) doit être à la fois descendante, autonome et dis-

---

<sup>52</sup>Voir aussi, par exemple, Hacking (1983) ou Giere (1999).

parate (*motley*). Elle est descendante dans le sens où la simulation tire en partie sa justification de la théorie dont elle dérive. Elle est autonome dans le sens où elle n'est pas entièrement contrainte par la théorie (qui n'est pas une « machine distributrice »). Elle est disparate dans le sens où elle doit fait appel à de multiples sources différentes pour établir la crédibilité de la simulation. C'est cette dernière caractéristique qui, pour Winsberg, est la plus importante.

La construction de simulations repose sur un ensemble de techniques qui sont développées au fil du temps : des façons de déterminer quels paramètres doivent être inclus dans la simulation, des façons de réduire la complexité computationnelle du modèle, des façons de représenter les entités d'un domaine, des stratégies pour explorer l'espace d'état d'un modèle, des techniques d'abstraction, d'idéalisation, mais aussi « the blood, sweat, and tears of much trial and error » (2009, p. 897). Au bout du compte :

« Whenever these techniques and assumptions are employed successfully, that is, whenever they produce results that fit well into the web of our previously accepted data, our observations, the results of our paper and pencil analyses, and our physical intuitions, whenever they make successful predictions or produce engineering accomplishments, their credibility as reliable techniques or reasonable assumptions grows. » (Winsberg, 2003, p. 122)

Si on en croit Winsberg, la valeur épistémique des techniques de modélisation employées dérive des succès obtenus grâce à elles au fil du temps. Comme l'illustre la citation ci-dessus, Winsberg adopte une approche assez pragmatique quant à ce qui constitue le succès d'une démarche de modélisation. Nous sommes d'accord avec lui, mais nous croyons aussi qu'il est nécessaire d'être plus explicite sur la relation que peut entretenir un modèle avec le monde. À cette fin, nous nous tournerons vers la vision développée par Michael Weisberg dans *Simulation and Similarity: Using Models to Understand the World*.

## Relation entre modèle et monde

Weisberg établit une distinction entre deux types de modélisation, selon que celle-ci soit dirigée vers une cible spécifique (*target-directed modeling*) ou non. Le cas où la modélisation n'a pas de cible spécifique dans le monde réel est à son tour subdivisé en trois catégories : la modélisation généralisée (*generalized modeling*), la modélisation hypothétique (*hypothetical modeling*) et la modélisation sans cible (*targetless modeling*).

La modélisation dirigée vers une cible spécifique est le cas le plus simple. On commence par identifier un phénomène qui correspond à une région particulière de l'espace-temps. Par exemple : un Boeing 747 volant à son altitude de croisière, le comportement sexuel des langoustes en période de migration dans le golfe du Mexique, ou bien les pratiques de citation des scientifiques ayant publié dans *Systematic Zoology* entre 1961 et 1983. Ensuite, il faut définir le système cible, ce qui implique déjà une certaine abstraction. Par « abstraction », ici, on entend la sélection d'un sous-ensemble de propriétés du phénomène étudié. Le choix des propriétés pertinentes dépend de nos intérêts de recherche. On inclura généralement les propriétés qui nous intéressent directement, de même que les propriétés qui sont causalement reliées à celles-ci. Si ce sont les propriétés aérodynamiques du Boeing 747 qui nous intéressent, on inclura la forme de son fuselage dans le système cible, mais on fera probablement abstraction du système d'alimentation électrique de l'avion. Notons qu'il y a déjà un contenu théorique dans le choix de ces propriétés : si on suppose qu'on peut ignorer le système d'alimentation électrique, c'est qu'on a déjà une théorie sur ce qui a (ou non) un impact sur l'aérodynamisme de l'appareil. Notons aussi que, en pratique, le processus est souvent itératif : on ajoute et retranche des propriétés au système cible à mesure que la modélisation nous éclaire sur les relations causales entre celles-ci.

Définir le système cible en terme de ses propriétés d'intérêt c'est déjà, aussi, par le fait même, définir un modèle (Suppes, 1962) : les variables choisies pour représenter ces propriétés peuvent être identifiées aux dimensions d'un espace d'état (Weisberg, 2013, p. 95–96). Cela suggère, déjà, une façon d'établir une relation entre un modèle

et le système modélisé : les trajectoires générées par le modèle dans son espace d'état doivent correspondre aux trajectoires du système cible dans son espace d'état. Cette relation, selon Weisberg, est une relation de *similarité*.

Nous reviendrons sous peu à la relation de similarité, mais il nous faut avant cela nous pencher sur les cas où la modélisation n'a pas de cible spécifique.

Commençons par le plus abstrait de ces cas : celui où la modélisation n'a pas de cible du tout. Nous avons déjà examiné en détail un tel modèle : le jeu de la vie. Ce type de modélisation s'apparente aux mathématiques pures : on étudie les propriétés formelles d'un système, sans égard à sa relation avec le monde ou à ses applications éventuelles. Cela ne veut pas dire que le modèle ne trouvera pas d'applications. L'histoire des mathématiques est parsemée de théories aux applications inattendues (on a qu'à penser aux géométries non euclidiennes). Et même sans applications concrètes, les modèles abstraits peuvent être utilisés. Un modèle comme le jeu de la vie, par exemple, permet de développer des intuitions sur les systèmes complexes en général : il illustre à merveille le fait que des règles simples peuvent générer des *patterns* de haut niveau (Dennett, 1991).

Un autre cas, c'est celui-ci où la modélisation a bien une cible, mais où la cible n'existe pas dans le monde réel. C'est ce que Weisberg appelle la « modélisation hypothétique ». Il nous donne l'exemple classique de Fisher, qui suggère de se pencher sur la reproduction sexuée à plus de deux sexes :

« No practical biologist interested in sexual reproduction would be led to work out the detailed consequences experienced by organisms having three or more sexes ; yet what else should he do if he wishes to understand why sexes are, in fact, always two ? » Fisher (1930, p. ix)

Mise à part la question de savoir s'il vrai que les sexes sont effectivement toujours au nombre de deux, la modélisation nous permet d'établir les conditions qui seraient nécessaires pour que puisse évoluer une espèce à plus de deux sexes. La biologie est un terrain fertile pour de telles questions hypothétiques. Peut-on imaginer un code génétique fonctionnant sur d'autres bases que les nucléotides (adénine, cytosine, guanine ou thymine) de l'ADN ? Peut-on imaginer une chimie organique basée sur le silicone plutôt que le carbone ? C'est le genre de questions qu'explore la discipline

de la vie artificielle (Langton, 1997), qui tente d'étudier « la vie telle qu'elle pourrait être » pour mieux comprendre la vie telle qu'elle est. Les modèles hypothétiques (que l'on pourrait aussi appeler modèles contre-factuels), nous aident à faire la cartographie des mondes possibles et, par conséquent, de situer parmi eux le monde actuel en départageant les propriétés nécessaires des propriétés contingentes.

Les modèles hypothétiques peuvent aussi nous permettre d'explorer les propriétés de systèmes qui n'existent pas dans le monde actuel, mais qui pourraient potentiellement être mis en place. Prenons un exemple relié au financement du système scientifique. Dans le monde actuel, les agences de financement attribuent les fonds de recherche en fonction de demandes de subventions qui sont évaluées par les pairs. Le processus est coûteux et demande une somme de travail colossal : en 2012, la National Science Foundation américaine a fait appel à plus de 17 000 scientifiques pour évaluer 53 556 demandes de subventions (National Science Foundation, 2013). Johan Bollen et ses collègues (2013 ; 2014) proposent un système complètement différent et explorent les conséquences possibles de cette alternative à l'aide d'un modèle multi-agents. Leur modèle est complètement décentralisé :

« Our proposed system would require funding agencies to give all scientists within their remit an unconditional, equal amount of money each year. However, each scientist would then be required to pass on a fixed percentage of their previous year's funding to other scientists whom they think would make best use of the money. Every year, then, scientists would receive a fixed basic grant from their funding agency combined with an elective amount of funding donated by their peers. As a result of each scientist having to distribute a given percentage of their previous year's budget to other scientists, money would flow through the scientific community. Scientists who are generally anticipated to make the best use of funding will accumulate more. » Bollen *et al.* (2014, p. 1)

Les auteurs testent leur suggestion à l'aide d'une simulation où ils supposent que les scientifiques attribuent des fonds à leurs pairs en proportion du nombre de citations faites de ces auteurs dans leurs propres articles au cours de cinq dernières années. Si, par exemple, j'ai cité trois l'article A et deux fois l'article B, j'attribue aux auteurs de l'article A trois cinquièmes des fonds que j'ai à distribuer et deux cinquièmes de ces fonds aux auteurs de l'article B (2013, p. 5).

Le système cible n'existe pas, puisque ce n'est pas comme ça que les fonds de recherche scientifique sont attribués en réalité, mais on peut très bien comparer les résultats du

modèle hypothétique aux résultats du système de financement actuel. Dans ce cas-ci, les résultats du modèle ressemblent à la réalité. Comme dans la réalité, la distribution du financement est fortement asymétrique : quelques chercheurs obtiennent la plus grande part des fonds disponibles. De plus, le financement des chercheurs individuels dans le modèle est corrélé<sup>53</sup> avec leur financement réel : ce sont, à peu de choses près, les mêmes chercheurs qui obtiennent les mêmes budgets. Mise à part la question de savoir si cette distribution des fonds de recherche est optimale, le modèle laisse supposer qu'elle pourrait être obtenue pour une fraction du coût et de l'effort exigés par le système actuel.

Le troisième type de modélisation abordé par Weisberg, sur lequel nous allons maintenant nous pencher, est celui qui nous intéresse le plus, puisque c'est celui qui correspond le mieux à notre propre démarche : la modélisation généralisée. Dans la modélisation généralisée, on aborde un phénomène qui est présent dans plusieurs systèmes différents. Weisberg donne comme exemple la reproduction sexuée :

« A general model of sexual reproduction isn't supposed to be about kangaroo sex or fungi sex, but about sex itself. [...] But nothing in the world looks like "sex in general." There is kangaroo sex, Tasmanian devil sex, and human sex, but not sex in general. Sex in general is an abstraction over these more specific kinds of sex. » (Weisberg, 2013, p. 116)

Chacun de ces différents systèmes cibles a des propriétés différentes. Il faut donc construire, par abstraction, un système cible qui ne comporte que des propriétés communes à tous les systèmes cibles : l'*intersection* des ensembles de propriétés des différents systèmes. Le degré d'abstraction de cette cible généralisée dépend de la similarité entre les différents systèmes cibles.

C'est exactement la situation à laquelle nous faisons face dans notre tentative de modéliser la science. La pratique scientifique varie d'une discipline à l'autre. Les théories et les modèles développés en physique nucléaire diffèrent de ceux développés par les biologistes. Les méthodes expérimentales utilisées en chimie organique ont peu à voir avec le travail de terrain fait en anthropologie culturelle et encore moins avec la preuve d'un théorème en mathématiques pures. Il y a toutefois des propriétés qui sont

---

<sup>53</sup>Pearson  $R = 0,2683$  et Spearman  $\rho = 0,2999$

communes à toutes ces disciplines, et ce sont ces propriétés que Hull (et nous, par le fait même) tente de capturer. Dans tous les cas, des idées sont transmises d'un scientifique à l'autre. Dans tous les cas, un des vecteurs principaux de transmission de ces idées est la publication d'articles scientifiques. Dans tous les cas, la citation joue un rôle important dans l'attribution de crédit aux membres de la communauté. Et ainsi de suite.

Pour construire un modèle computationnel, on ne peut toutefois se contenter de faire abstraction des propriétés qui ne sont pas partagées par les systèmes cibles. Certaines de ces propriétés sont nécessaires au fonctionnement du système, et on se voit forcé d'adopter des valeurs spécifiques pour celles-ci. Par exemple, dans un modèle généralisé de la reproduction sexuée :

« Such a model would not represent the distribution of genotypic fitness in a population, but would have specific genotypes tied to specific organisms that engaged in reproduction. This would require the theorists to make assumptions about the life-cycle, spatial distribution, mating interactions, and fitness of each organism in the model. The model would no longer contain just those types of properties shared with the abstract target ; it would be closer in concreteness to specific-organism targets. » (Weisberg, 2013, p. 116)

Notre modèle généralisé de la science nous force à faire le même type de choix. Pour ne prendre qu'un exemple, pensons à la co-écriture des articles scientifiques : le nombre de coauteurs peut atteindre plusieurs centaines en physique nucléaire alors qu'il est en moyenne inférieur à deux dans les sciences humaines et sociales (Harzing et Alakangas, 2016).

C'est au moment d'interpréter les résultats du modèle que l'impact de cette « concrétisation » de certaines propriétés se fait sentir. Cet impact sera différent selon que la propriété concrétisée a une influence causale sur les résultats du modèle. Si la propriété joue un rôle causal, la portée du modèle doit être restreinte aux systèmes cibles qui possèdent cette propriété. Si la propriété ne joue pas de rôle causal, on peut généraliser l'interprétation à l'ensemble des systèmes cibles, même ceux qui ne possèdent pas la propriété.

La question de savoir si la propriété joue un rôle causal est elle-même une question qui demande une recherche intensive. Si le modèle est suffisamment général pour

représenter différentes valeurs de la propriété, on peut procéder à une analyse de sensibilité. Il existe différentes méthodes pour procéder à cette analyse (Thiele *et al.*, 2014 ; Lee *et al.*, 2015 ; Ten Broeke *et al.*, 2016), mais celles-ci sont limitées par le fait que la taille de l'espace d'états d'un modèle computationnel est souvent trop grande pour être explorée à fond. Il arrive aussi que le modèle ne soit pas suffisamment général pour représenter toutes les différentes façons de concrétiser une propriété. C'est le cas de la propriété « nombre de coauteurs » dans le modèle que nous allons présenter au chapitre 5 : notre modèle ne couvre que le cas où les articles scientifiques ont un seul auteur. Pour savoir si le nombre de coauteurs joue un rôle causal dans les résultats obtenus par le modèle, il faudrait construire un second modèle, différent, où le nombre de coauteurs peut varier, et comparer les résultats obtenus par celui-ci aux résultats du premier modèle. Vu l'ampleur de la tâche, nous n'avons d'autre choix que de la réserver à des recherches futures. Il nous faut donc restreindre l'interprétation des résultats de notre modèle aux disciplines où la norme est d'avoir un seul auteur par article.

Selon Weisberg, un modèle est constitué à la fois de sa structure mathématique et de son interprétation (que Weisberg appelle « *construal* »). Il rejoint en cela Ronald Giere (2010) qui soutient que, dans la mesure où il n'y a pas d'isomorphisme parfait entre un modèle et le monde, il faut faire intervenir l'intentionnalité du modélisateur pour ancrer la sémantique du modèle. Il y a trois composantes à l'interprétation d'un modèle, et celles-ci sont étroitement reliées :

- La portée désirée pour le modèle (*intended scope*) détermine les propriétés du système cible qui sont visées par le modèle.
- L'assignation des éléments structurels du modèle aux propriétés qu'ils représentent.
- Les critères de fidélités, par lesquels on évalue la similarité entre les éléments du modèle et les propriétés du système cible. Weisberg distingue deux types de critères de fidélité : les critères de fidélité dynamiques et les critères de fidélité représentationnels.

Les critères de fidélité dynamiques concernent la relation entre les *outputs* du modèle (ses variables dépendantes) et les propriétés mesurables du système cible. Quand le système cible est un système spécifique, ces critères tendent à être assez serrés. Si, par exemple, on modélise les fluctuations dans la population des renards sur l'île d'Anticosti pendant l'année 2016, on peut exiger que le nombre de renards représentés dans le modèle soit, à chaque temps  $t$  pour lequel on a une mesure de la population sur l'île, à l'intérieur d'un intervalle de  $\pm 10\%$  de celle-ci.

Dans le cas d'une démarche de modélisation généralisée comme la nôtre, les exigences de fidélité sont différentes puisqu'il n'y a pas un système cible en particulier qui puisse nous fournir des données empiriques précises. On peut toutefois se rabattre sur des *patterns* de plus haut niveau.

Dans le cas particulier des modèles du système scientifique, Meyer (2011) établit une liste de « faits stylisés » qui peuvent être utilisés comme critères de fidélité. Parmi ceux-ci, on retrouve la loi de Lotka (Lotka, 1926 ; Simon, 1957) et l'effet Mathieu (Merton, 1968) qui, comme nous le verrons, sont reproduits par notre modèle (figures 6.3 et 6.6). D'autres ne s'appliquent pas à notre modèle, comme la croissance exponentielle du nombre de journaux scientifiques (de Solla Price, 1963), ou la loi de Bradford (Cole, 1962), qui décrit la concentration des articles scientifiques les plus référencés dans un petit nombre de journaux, puisque nous avons décidé de faire abstraction des différents journaux scientifiques (section 5.1).

En plus des critères de fidélité dynamiques, on retrouve les critères de fidélité représentationnels :

« these criteria specify how closely the model's internal structure must match the causal structure of the real world phenomenon to be considered an adequate representation. »  
(Weisberg, 2013, p. 41)

Les modèles computationnels portent généralement une plus grande attention aux critères de fidélité représentationnels que les modèles mathématiques à base d'équations. Ces derniers visent souvent à décrire les relations entre des variables de haut niveau, sans égard aux mécanismes qui causent ces changements de valeur. Dans les équations de Lotka-Volterra, par exemple, la population de proies est représentée

uniquement par son nombre total et fluctue en fonction du nombre de prédateurs et d'autres paramètres comme son taux de croissance ou le taux de prédation. Les mécanismes qui se trouvent derrière un paramètre comme le taux de prédation ne sont pas spécifiés. Dans un modèle multi-agents comme le *Wolf-Sheep Predation* de Wilensky (1997), par contre, chaque loup et chaque mouton individuel est représenté. Quand un mouton meurt, c'est un loup particulier qui est responsable de sa mort : la relation causale entre l'action du loup et la mort du mouton est rendue explicite par le modèle.

Dans le modèle développé au chapitre 5, nous avons accordé une grande importance aux critères de fidélité représentationnels. Les mécanismes causaux identifiés par Hull sont représentés directement dans le modèle. La lecture d'un article par un scientifique, par exemple (voir section 5.4.5), a plusieurs ramifications causales : elle cause la répliation de l'idée présentée dans l'article dans l'esprit du lecteur, elle déclenche l'évaluation de cette idée par le lecteur en fonction de ses idées existantes et du crédit attribué à l'auteur de l'article et aux auteurs des articles cités et elle cause l'attribution de crédit par le lecteur à l'auteur de l'article et aux auteurs cités. Tous ces mécanismes (et plusieurs autres) sont à l'oeuvre pour chacun des agents du modèle. L'hypothèse de Hull (et par extension, la nôtre), c'est que des mécanismes équivalents, où les mêmes facteurs ont les mêmes propriétés causales, existent dans le monde réel.

Comme nous le verrons dans la prochaine section, cette attention portée aux mécanismes causaux dans les modèles multi-agents joue un rôle important dans leur pouvoir explicatif.

### Les modèles comme explications

Pendant une bonne partie du vingtième siècle, la notion dominante d'explication en philosophie des sciences, c'est le modèle nomologico-déductif (Hempel et Oppenheim, 1948). Hull, non sans faire preuve d'un certain sarcasme, décrit ainsi la situation en 1975 :

« Perhaps philosophers have been unable to discover the essence of beauty, goodness, horses, lemons, and games, but they have discovered the essence of scientific explanation, and it is subsumption under a scientific law. There is one and only one way of explaining a particular natural phenomenon or regularity and that is to derive it from one or more laws of nature and, when necessary, relevant particular circumstances. » (Hull, 1975, p. 273)

Expliquer un phénomène, c'est montrer comment celui-ci aurait pu être prédit par les lois de la nature étant donné les conditions qui prévalaient avant l'occurrence de ce phénomène. Malgré le ton adopté, Hull ne rejette pas entièrement cette approche. Il est toutefois préoccupé par le fait qu'elle s'applique mal aux sciences historiques (l'histoire humaine, mais aussi la cosmogonie, la géologie, la paléontologie, etc.) Les phénomènes examinés par celles-ci peuvent difficilement être directement subsumés sous des lois scientifiques.

Les « lois » utilisées par les sciences historiques s'apparentent plus à des heuristiques (*rules of thumb*) qu'aux lois précises de la physique. Il donne l'exemple (1975, p. 264–265) de la « loi de la superposition » utilisée par les géologues pour reconstruire l'histoire d'une région de la croûte terrestre. Celle-ci dit simplement que chaque couche sédimentaire tend à être plus jeune que celle qui se trouve en dessous d'elle, mais plus vieille que celle qui est au-dessus d'elle. Il y a cependant nombre d'exceptions à cette loi, puisque les mouvements de la croûte terrestre peuvent parfois inverser l'ordre des couches. Une géologue confrontée à cette situation fera appel à d'autres principes de reconstruction, lesquels sont aussi susceptibles de souffrir des exceptions. Cela ne veut pas dire (loin de là) que les lois fondamentales de la nature ne sont pas à l'œuvre dans la formation d'une région particulière de la croûte terrestre, mais en l'absence d'information suffisante sur les conditions précises qui prévalaient lors de la formation de la région étudiée, il est impossible de placer le phénomène en relation déductive avec ces lois<sup>54</sup> :

« To put the matter simply, scientific theories tell us what can happen. Within these limits, certain additional rules help historians decide what actually did happen. At present, considerable slippage exists between these two sets of principles. The rele-

---

<sup>54</sup>Dans un article subséquent (1992), Hull insiste sur la compatibilité entre le modèle déductif-nomologique et son modèle d'explication narrative (qu'il appelle alors « modèle des circonstances particulières »).

vant theories do not greatly constrain the possible states of the system, and few of the various rules of thumb used to reconstruct the past are inferential consequences of these theories. » (Hull, 1975, p. 266)

Dans le même article, Hull suggère qu'il faut considérer la possibilité d'admettre ce qu'il appelle des « explications intégratives », où l'action d'expliquer consiste à replacer un événement dans le contexte du « tout organisé » qui en est responsable. Ce « tout organisé », comme ce sera le cas dans *Science as a process*, peut fort bien constituer un mécanisme. La relation entre le mécanisme et l'événement expliqué n'est pas une relation de déduction, mais plutôt une relation de *pattern recognition* (p. 274).

La notion d'explication mécaniste a été développée par plusieurs philosophes (indépendamment de Hull) depuis le début des années 1980<sup>55</sup>. Salmon (1984), pionnier de cette approche, considère que les explications mécanistes sont de nature *ontique*, puisque les mécanismes identifiés sont directement opérants dans le monde réel alors que les explications nomologiques sont *épistémiques*, puisque les lois auxquelles elles font appel sont le produit de notre activité cognitive et ont une valeur descriptive plutôt que causale. Bechtel et Abrahamsen (2005) font toutefois remarquer que les explications mécanistes font intervenir un *modèle* du mécanisme et non pas le mécanisme lui-même et qu'elles sont, en cela, tout aussi épistémiques que les explications déductives-nomologiques.

Voici comment Bechtel et Abrahamsen (2005) définissent un mécanisme :

« A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena. » (p. 423)

La définition de Bechtel et Abrahamsen recoupe d'autres définitions proposées ailleurs (Bechtel et Richardson, 1993 ; Glennan, 1996 ; Machamer *et al.*, 2000) en ce qu'elle fait intervenir les mêmes éléments essentiels (Hedström et Ylikoski, 2010 ; Craver et Tabery, 2017) : des composantes, organisées d'une façon particulière, qui entretiennent

---

<sup>55</sup> Avec comme prédécesseurs, entre autres, Fodor (1968), Simon (1969), Kauffman (1970) et Wimsatt (1974, 1976). On peut faire remonter les explications mécanistes au moins jusqu'à Galilée (Machamer, 1998).

une relation causale avec un phénomène. Expliquer le phénomène, c'est proposer un modèle du mécanisme qui en est la cause :

« A model of a mechanism describes or portrays what are taken to be its relevant component parts and operations, the organization of the parts and operations into a system, and the means by which operations are orchestrated so as to produce the phenomenon. When they are correct, models of mechanisms accurately describe relevant aspects of the mechanism operative in the world. » Bechtel et Abrahamsen (2005, p. 425–426)

Comme nous l'avons soutenu dans les sections précédentes—et comme Bechtel et Abrahamsen (2005, p. 431) le font remarquer—les modèles computationnels sont un outil approprié pour décrire ce genre de structure. Machamer, Darden et Craver (2000) proposent une ontologie des mécanismes en termes d'*entités* et d'*activités*. Les entités ont des propriétés, et s'engagent dans des activités qui modifient ces propriétés. Ce sont les activités des entités qui causent les changements d'état du mécanisme. Cette ontologie s'aligne parfaitement avec celle des modèles multi-agents, où les entités sont des agents (avec des propriétés) et les activités sont les actions de ces agents. Ils insistent aussi sur la nature hiérarchique des mécanismes : une entité opérante dans un mécanisme peut elle-même être constituée de mécanismes de plus bas niveau.

Machamer *et al.* soutiennent que cette façon de découper le monde est essentielle pour rendre intelligibles les phénomènes qu'on y observe :

« The understanding provided by a mechanistic explanation may be correct or incorrect. Either way, the explanation renders a phenomenon intelligible. Mechanism descriptions show *how possibly*, *how plausibly*, or *how actually* things work. Intelligibility arises not from an explanation's correctness, but rather from an elucidative relation between the explanans (the set-up conditions and intermediate entities and activities) and the explanandum (the termination condition or the phenomenon to be explained). » (Machamer *et al.*, 2000, p. 21)

Ils rejoignent en cela la devise générativiste de Epstein (2006, p. 51) évoquée à la section 1.4 : « *If you didn't grow it, you didn't explain it.* » Il s'agit, bien sûr, d'une condition nécessaire, mais non suffisante. Entre « *how possibly* » et « *how actually* », il y a de nombreux pas à franchir.

L'objectif général que poursuit David Hull dans *Science as a process*, c'est d'expliquer comment les motivations individuelles des scientifiques permettent à la science de

progresser. En proposant un mécanisme évolutionnaire qui serait la source du changement conceptuel en science, il franchit un premier pas dans cette direction. La description de Hull est toutefois fort générale et laisse place à l'élaboration de plusieurs modèles pouvant satisfaire les contraintes qu'elle met en place. Au chapitre suivant, nous proposons un modèle précis (quoiqu'encore fort abstrait) du mécanisme décrit par Hull. Avec celui-ci, nous espérons franchir un pas supplémentaire vers une explication *possible* de la façon dont la science progresse.

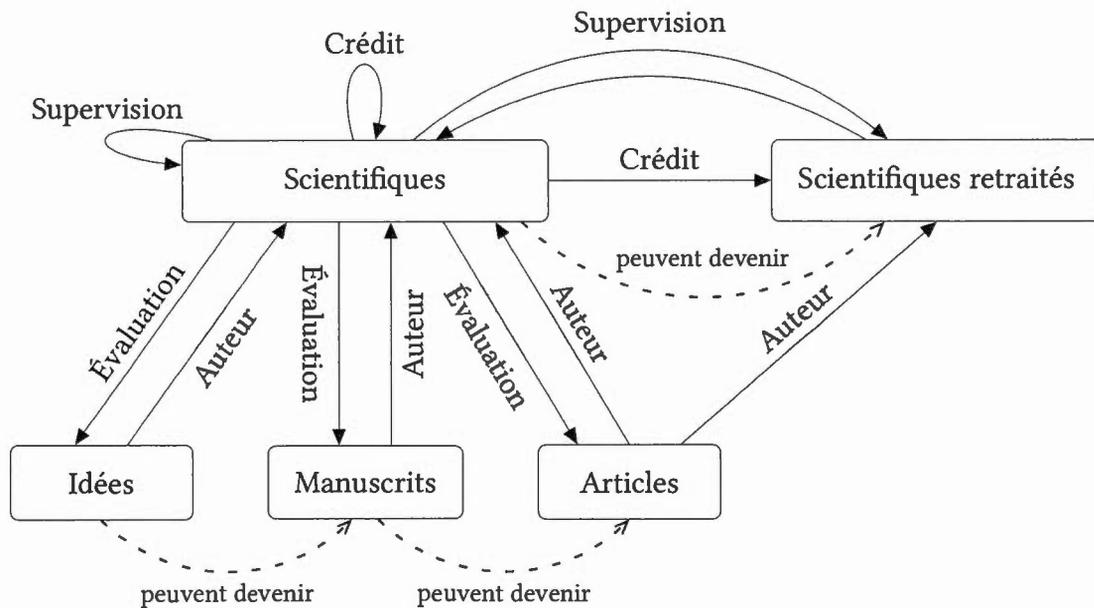
## CHAPITRE V

### UN MODÈLE DU PROCESSUS SCIENTIFIQUE

*In short, all of the prerequisites necessary for engaging in the process commonly termed “science” are to some extent programmed into us.*

– David Hull (1988a, p. 124)

Dans le chapitre 2, nous avons vu plusieurs modèles multi-agents de la science qui traitaient de différents aspects du système scientifique, sans toutefois embrasser pleinement l’aspect évolutionnaire de la science, ni aborder la notion de crédit qui sous-tend plusieurs des aspects sociaux du système scientifique. Dans le chapitre 3, nous nous sommes penchés sur la théorie de David Hull qui, elle, met les notions d’évolution et de crédit à l’avant-plan, tout en étant exprimée sous une forme verbale qui ne permet pas d’en explorer systématiquement les conséquences. Dans le présent chapitre, nous allons tenter de procéder à la synthèse de ces deux visions et construire un modèle multi-agents inspiré de la théorie de Hull.



**Figure 5.1** Les entités qui composent le modèle

### Description des entités du modèle

La figure 5.1 donne un aperçu de haut niveau des entités qui composent le modèle et des relations qu'elles entretiennent entre elles, mais la façon la plus précise de les spécifier est sans doute de reproduire ici les déclarations qui se trouvent dans le code source du modèle<sup>56</sup>. Celui-ci a été développé à l'aide du langage NetLogo (Wilensky, 1999), une variante du langage pédagogique Logo (Feurzeig et Papert, 1967) spécialisée dans la création de MMA.

NetLogo permet de travailler avec trois types d'agents distincts :

- Des agents immobiles, qui forment l'environnement du modèle. Le monde de NetLogo est une grille bidimensionnelle composée de cellules discrètes. Dans le

<sup>56</sup> Ce n'est pas tout à fait de la « programmation littéraire » (Knuth, 1984), mais notre approche s'en inspire. Le présent document est composé avec un module  $\LaTeX$  qui permet d'extraire des parties de code du modèle et de les insérer dans le texte, nous assurant ainsi de la synchronisation entre les deux.

langage de NetLogo, ces cellules sont appelées *patches*. Même si leur position est fixe, elles n'en constituent pas moins des agents à part entière, pour lesquels on peut définir des variables et des comportements.

- ▶ Des agents mobiles, qui se déplacent sur la grille de cellules<sup>57</sup>. NetLogo hérite du Logo original le terme *turtles* pour désigner ces agents. Ces « tortues » ont aussi leurs variables et leurs comportements.
- ▶ Des liens (*links*) entre les tortues. Ces liens permettent de définir les relations entre les tortues et de les organiser en réseaux. Tout comme les cellules et les tortues, ils peuvent avoir leurs variables et leurs comportements, ce qui en fait aussi des agents à part entière.

Le langage permet de définir différentes « espèces » (*breeds*<sup>58</sup>) de tortues et de liens. Ces espèces correspondent approximativement à la notion de « classes » dans les langages orientés objet, en ce sens qu'elles permettent de définir des variables qui sont possédées par chaque membre de l'espèce. Une même tortue peut toutefois changer d'espèce au fil de la simulation tout en conservant les valeurs des variables qui sont communes à son ancienne et sa nouvelle espèce.

Notre modèle fait un usage important de cette dernière possibilité, notamment pour les espèces représentant les trois stades de la vie des idées (*ideas*), dont certaines peuvent devenir des manuscrits (*drafts*) qui, s'ils sont acceptés pour publication, deviendront finalement des articles (*papers*). Le code 5.1 à la page suivante montre la déclaration de ces trois espèces. Le mot clé *breed* permet de définir une nouvelle espèce en spécifiant son nom au pluriel et au singulier. On peut ensuite déclarer les

---

<sup>57</sup> Bien que la grille soit constituée de cellules discrètes, les coordonnées des agents dans NetLogo sont exprimées par des nombres à virgule flottante à double précision. Ainsi, déterminer la position d'un agent ne se limite pas à identifier la cellule dans laquelle il se trouve : la position de celui-ci à l'intérieur de la cellule peut aussi être prise en compte.

<sup>58</sup> Le terme « races » serait une traduction plus littérale de « *breeds* », mais « espèces » nous semble néanmoins plus approprié.

variables propres à l'espèce en ajoutant le suffixe « -own » au pluriel de l'espèce et en énumérant ensuite les variables désirées<sup>59</sup>.

---

```
breed [ ideas idea ]
ideas-own [ author ]

breed [ drafts draft ]
drafts-own [ author ]

breed [ papers paper ]
papers-own [ author year ]
```

---

**Code 5.1** Déclaration des espèces ideas, drafts et papers.

Les trois espèces possèdent la variable `author`<sup>60</sup>, dont la valeur sera conservée lors du passage d'une espèce à l'autre, mais seuls les articles possèdent la variable `year`, qui désigne l'année de publication de l'article.

Notons aussi que chaque type d'agent dans NetLogo (cellules, tortues et liens) possède aussi un certain nombre de variables prédéfinies, disponibles peu importe l'espèce. Dans le cas des idées, manuscrits et articles, les variables prédéfinies `xcor` et `ycor` prendront une importance particulière, puisque ce sont elles qui spécifient leurs positions dans le paysage épistémique, auquel nous reviendrons plus loin.

Il y a aussi une entité qui n'est pas représentée explicitement à la figure 5.1, ni dans le code NetLogo : le *Journal*. Chaque fois qu'un manuscrit est publié et devient un article, il faut supposer qu'il est publié dans un journal scientifique particulier auquel

---

<sup>59</sup>NetLogo étant un langage à typage dynamique, il n'est pas nécessaire de spécifier le type des variables.

<sup>60</sup>Le lecteur attentif aura remarqué qu'il s'agit de `author`, au singulier, et non de `authors`, au pluriel. Nous avons travaillé aussi sur une version du modèle où les articles pouvaient être coécrits par plusieurs auteurs, mais cet ajout soulevait un nombre étonnant de questions : Comment les scientifiques choisissent-ils leurs coauteurs ? Comment négocient-ils l'ordre des noms sur l'article publié ? Comment choisissent-ils l'idée à présenter dans l'article ? Cette idée est-elle l'idée d'un coauteur en particulier ou bien une idée commune qui émerge en combinant des idées des coauteurs ? Comment choisissent-ils les articles à citer (question particulièrement cruciale vu le rôle des citations dans le modèle) ? À cela s'ajoutent, entre autres, des questions sur la lecture et l'évaluation d'articles coécrits. Dans l'intérêt de préserver la simplicité relative du modèle, nous avons choisi de nous concentrer sur la version à auteur unique.

ont accès tous les scientifiques. Ce choix nous donne une idée de la « taille » du domaine scientifique qui peut être représenté par notre modèle : il s'agit, grosso modo, du champ d'investigation couvert par un journal spécialisé. On aurait pu tenter de modéliser un domaine plus vaste, où compétitionnent plusieurs journaux scientifiques (avec un lectorat et des facteurs d'impacts différents), mais nous avons jugé cette complexification peu utile. Hull lui-même, dans *Science as a Process*, se concentre sur la communauté qui gravite autour d'un seul journal : *Systematic Biology*.

Bien qu'idées, manuscrits et articles soient techniquement des agents au sens de Net-Logo, ce sont des entités passives pour lesquelles aucun comportement n'est défini. Les agents actifs dans le modèle, ce sont les « scientifiques » (`scientists`), définis dans le code 5.2. Notez la liste de variables, plus longue dans leur cas. Nous reviendrons sur certaines d'entre elles lorsque nous discuterons, plus loin, des comportements des scientifiques.

---

```
breed [ scientists scientist ]
scientists-own [
  start-year
  alpha
  beta
  gamma
  delta
  my-credit
  my-normalized-credit
]
```

---

**Code 5.2** Déclaration de l'espèce `scientists`.

Les « scientifiques retraités » (`retired-scientists`) constituent une dernière espèce de tortues (voir code 5.3 à la page suivante)<sup>61</sup>. Les scientifiques passent à cette espèce lorsqu'ils cessent d'être actifs dans le système. Ils continuent, toutefois, à entretenir des liens avec d'autres entités du modèle : les articles qu'ils ont écrits et les autres scientifiques qui, potentiellement, leur attribuent du crédit.

---

<sup>61</sup>Nous sommes conscients que, prise hors contexte, la présente phrase serait pour le moins étrange.

---

```
breed [ retired-scientists retired-scientist ]
retired-scientists-own [ start-year alpha beta gamma delta ]
```

---

**Code 5.3** Déclaration de l'espèce `retired-scientists`.

Le moment est venu, justement, de nous pencher sur les différents liens que peuvent entretenir entre elles les espèces du modèle. Le code 5.4 contient les déclarations pour toutes les espèces de liens. Notez les remarques qui apparaissent après le caractère « ; » (qui indique le début d'un commentaire dans le langage NetLogo). Celles-ci indiquent de façon informelle les espèces de tortues qu'une espèce de liens est censée relier. Les liens de l'espèce `citations`, par exemple, relient des articles entre eux (`paper -> paper`) : ce sont des liens orientés (`directed-link-breed`) qui vont de l'article qui cite à l'article cité.

---

```
directed-link-breed [ citations citation ] ; paper -> paper
directed-link-breed [ supervisions supervision ] ; scientist -> scientist
directed-link-breed [ credits credit ] ; scientist -> scientist
credits-own [ amount normalized-amount ]
directed-link-breed [ evaluations evaluation ] ; scientist -> {idea, draft, paper}
evaluations-own [ subjective-value tested? ]
```

---

**Code 5.4** Déclaration des espèces de liens.

Les liens de type `supervisions` représentent la relation entre un scientifique et ses étudiants. Ce sont des liens orientés (`directed-link-breed`). La distance géodésique entre deux scientifiques sert à établir le degré d'apparentement entre deux scientifiques utilisé dans le calcul de la valeur sélective conceptuelle inclusive (à laquelle nous reviendrons sous peu—voir équation 5.2, figure 5.5 et code 5.13, pages 154–156).

Les liens orientés de type `credits` représentent la quantité de crédit accordé à une scientifique par une autre, cette quantité étant stockée dans la variable `amount`. Si deux scientifiques s'accordent mutuellement du crédit, deux liens seront créés entre eux : un dans chaque direction.

Finalement, les liens de type `evaluations` relient une scientifique à une idée, un manuscrit ou un article, et représentent l'opinion qu'elle s'en fait. Cette valeur, comprise dans l'intervalle  $[0; 1]$ , est stockée dans la variable `subjective-value` (voir sec-

tion 5.4.1). Ces liens possèdent aussi la variable booléenne `tested?`, qui indique si l'idée évaluée a été testée par la scientifique (voir section 5.4.2). La valeur *subjective* diffère généralement de la valeur *objective*, à laquelle les scientifiques n'ont pas un accès direct. Cette valeur objective est définie pour le dernier type d'« agents » dans la simulation : les cellules (`patches`).

---

```
patches-own [ objective-value ]
```

---

**Code 5.5** Déclaration des variables de cellules.

Comme nous l'avons expliqué plus haut, les cellules forment l'environnement dans lequel sont situés tous les autres agents, et la position de ces derniers est exprimée par les coordonnées `xcor` et `ycor`.

Comme le montre le code 5.5, chaque cellule possède une variable `objective-value`, qui correspond à sa valeur objective. Cet environnement constitue le « paysage épistémique » dont les scientifiques essaient d'identifier les sommets. Ces valeurs sont déterminées à partir de différentes « fonctions objectif » et sont aussi comprises dans l'intervalle  $[0; 1]$ . La figure 5.2 à la page 137 montre les différents paysages disponibles. La difficulté de trouver les sommets varie d'un paysage à l'autre, et nous reviendrons très bientôt sur chacun des paysages choisis (section 5.2.1).

### La notion de paysage épistémique

Nous avons déjà abordé la notion de paysage épistémique dans la section 2.5, quand nous avons examiné le modèle de Weisberg et Muldoon (2009). Dans celui-ci, les agents devaient trouver les « approches » ayant la plus grande « significativité épistémique » dans un paysage en trois dimensions (figure 2.6 à la page 45). Nous avons aussi vu, à la section 2.6, le modèle de Grim (2009), où des communautés scientifiques organisées en réseaux sont à la recherche de l'« hypothèse » ayant la meilleure « valeur épistémique » dans différents paysages en deux dimensions (figure 2.10 à la page 50).

Dans ces deux modèles (tout comme dans le nôtre, où les agents seront à la recherche de la « valeur objective » la plus élevée), le paysage établit une sémantique toute simple pour les objets qu'on y place, qu'il s'agisse d'« approches » (Weisberg et Muldoon), d'« hypothèses » (Grim) ou d'« idées » (notre modèle). Cette sémantique se limite à deux aspects :

- ▶ Chaque objet se voit attribuer une valeur plus ou moins élevée (« significativité épistémique », « valeur épistémique », « valeur objective », etc.).
- ▶ Chaque objet entretient avec les autres objets des relations de plus ou moins grande proximité.

C'est tout.

Est-il légitime de réduire la sémantique des idées scientifiques à ces deux aspects ? Nous soutenons que oui, tout en étant conscients qu'il s'agit d'un modèle limité et que le petit nombre de dimensions des paysages utilisés, en particulier, en limite le réalisme.

Pour aider à comprendre comment nous pouvons soutenir une telle chose, nous évoquerons la « Bibliothèque de Babel », une idée popularisée par une nouvelle de Borges (1944), qui s'est lui-même inspiré de Laßwitz (1907)<sup>62</sup>. Dans l'histoire imaginée par Borges, l'univers tout entier est constitué par une bibliothèque où se succèdent, apparemment sans fin, des salles hexagonales aux murs garnis de livres. Chaque livre contient 410 pages, chacune d'elle comportant quarante lignes de quatre-vingts caractères, pour  $410 \times 40 \times 80 = 1\,312\,000$  caractères par livre. Les caractères utilisés sont au nombre de vingt-cinq : espace, point, virgule et vingt-deux lettres différentes. La bibliothèque est donc immense, mais pas infinie : elle contient les  $25^{1\,312\,000} \approx 1,956 \times 10^{1\,834\,097}$  livres possibles. Les humains qui peuplent cet univers consacrent leur vie à arpenter la bibliothèque à la recherche de livres sensés. La quasi-totalité des

---

<sup>62</sup> J'ai aussi évoqué cette métaphore dans un article précédent (Payette, 2011). Je ne suis pas non plus le premier à en faire usage dans un tel contexte. Dennett (1995, p. 107–108), entre autres, l'utilise pour parler du code génétique et de la notion, pas très éloignée de celle qui nous concerne ici, de paysage adaptatif.

livres ne sont qu'une suite de caractères pêle-mêle, mais parmi eux se trouvent aussi nécessairement tous les livres jamais écrits, y compris la Bible, les œuvres complètes de Molières, *On the Origin of Species* et *Science as a Process* (quitte à ce que ça soit en plusieurs volumes lorsqu'un ouvrage dépasse 410 pages.) Il s'y trouve aussi tous les numéros, publiés ou à venir, de *Nature*, *Science*, *PLOS ONE* et tous les autres journaux scientifiques.

Et qu'est-ce, au fond, que la bibliothèque de Babel, sinon un espace à 1 312 000 dimensions, où chaque dimension peut prendre vingt-cinq valeurs différentes ? Et si on y ajoute une un-million-trois-cent-douze-mille-et-unième dimension qui indique la valeur de chaque livre (peu importe la façon dont cette valeur est déterminée), on se retrouve avec une structure similaire à celle de nos paysages épistémiques. La différence principale (outre le nombre de dimensions) c'est qu'il faut supposer que les relations de proximité dans nos paysages sont des relations de proximité *sémantique*, et non pas *syntactique*, comme dans l'espace des livres possibles dans la bibliothèque de Babel. Ainsi, dans le paysage épistémique de la biologie, par exemple, *De l'origine des espèces* occupe une position très proche de *On the Origin of Species*, même si les caractères qui composent les deux livres sont très différents.

La notion de paysage épistémique n'est pas présente chez Hull. C'est un ajout qu'il nous a fallu faire à sa théorie pour rendre compte d'un des aspects importants de celle-ci : le fait que les idées scientifiques ont une valeur objective et que le processus scientifique (avec son algorithme évolutionniste et sa combinaison de facteurs internes et externes à la science) permet à la communauté de découvrir et de trouver les idées de plus grande valeur. Dans la théorie de Hull, c'est le monde *lui-même* qui joue le rôle de paysage épistémique. C'est sa structure à lui que les scientifiques essaient de découvrir par leurs expériences et leurs raisonnements. Alors quand Hull parle de ce qui se passe avec les scientifiques dans « la vraie vie », nul n'est besoin de postuler une structure formelle qui serait différente du monde lui-même. Dans notre modèle, nous n'avons pas ce luxe : il faut bien donner à nos agents quelque chose à découvrir.

Nous ne connaissons pas le véritable paysage épistémique de la science. C'est le rôle de l'entreprise scientifique, justement, d'en faire la cartographie, et cette tâche est bien loin d'être complétée. À la limite, nous pourrions tenter de nous pencher sur un sous-

domaine scientifique particulièrement bien exploré et tenter de simuler le comportement de la communauté scientifique dans le paysage formé par ce sous-domaine. C'est un peu ce qu'a fait Edmonds (2007) en utilisant le calcul des prédicats du premier ordre avec une seule règle d'inférence pour son modèle de la science. Quoique Edmonds ne formule pas explicitement le problème en ces termes, celui-ci pourrait théoriquement être ramené à la forme d'un paysage épistémique où chaque théorème se voit attribuer une valeur selon certains critères (voir section 2.4) et où les relations de proximité sont établies par la possibilité de dériver un théorème à partir d'un autre.

Nous n'avons toutefois pas choisi cette voie, puisque c'est plutôt la façon dont la communauté scientifique se comporte par rapport à la *structure* d'un paysage qui nous intéresse, et non pas le contenu de celui-ci. Par « structure », nous voulons surtout désigner le fait qu'un paysage peut comporter un ou plusieurs sommets de valeur épistémique, et que le chemin pour parvenir à ces sommets peut être plus ou moins parsemé d'embûches : minimum locaux, où la communauté peut rester coincée, et vallées de valeur épistémique basse, qui n'offrent aucun indice quant à la présence d'un sommet à proximité. Nous avons aussi choisi de comparer la performance de la communauté sur plusieurs paysages épistémiques différents, ce qui aurait été plus difficile s'il nous avait fallu faire l'exercice de traduire plusieurs domaines scientifiques différents sous forme de paysages. Cette décision nous a aussi permis de bénéficier du fait qu'il soit possible de trouver dans la littérature sur les algorithmes génétiques plusieurs « problèmes d'optimisation continus » dont la structure est appropriée à servir de paysage épistémique.

Aussi, il ne faut pas supposer que les paysages épistémiques que nous utilisons dans la simulation représentent la science dans son ensemble. Il faut, d'abord, supposer que nos agents tentent d'explorer un champ scientifique restreint, dont l'étendue serait à peu près celle qui est couverte par un journal scientifique spécialisé. Il faut aussi supposer que les « idées » produites dans le modèle font intervenir un ensemble de présupposés implicites qui ne sont pas remis en question. On peut évoquer ici l'image, proposée par Quine, selon laquelle la science est un réseau dont le noyau (quoique jamais à l'abri d'une révision éventuelle) est stable :

La totalité de ce qu'il est convenu d'appeler notre savoir ou nos croyances, des faits les plus anecdotiques aux lois les plus profondes de la physique atomique ou même des mathématiques pures et de la logique, est une étoffe tissée par l'homme, et dont le contact avec l'expérience ne se fait qu'aux contours. (Quine, 1980, p. 108)

Pour traduire cette image en termes de paysage épistémique, on peut supposer que la science dans son ensemble est un paysage à  $m$  dimensions, mais que nos agents n'explorent qu'un nombre limité de dimensions,  $n$ , où  $n \ll m$ . Une idée « complète » serait alors représentée par le vecteur  $(a_1, \dots, a_n, a_{n+1}, \dots, a_m)$  où les composantes  $a_1, \dots, a_n$  varient d'une idée à l'autre et les composantes  $a_{n+1}, \dots, a_m$  représentent le noyau dur de la science et sont fixes pour toutes les idées du modèle. On peut donc se contenter, dans le modèle, de représenter les idées par des vecteurs à  $n$  dimensions, tout en gardant en tête que ces dimensions ne représentent que la périphérie de la toile.

Si on souhaitait représenter la possibilité de changement scientifique radical, on pourrait inclure dans le modèle certaines dimensions plus fondamentales et trouver une façon de limiter les possibilités de mouvement dans celles-ci. Ce pourrait être simplement de placer sur ces dimensions quelques sommets très étroits et éloignés les uns des autres, de façon à ce que, pour passer d'un sommet à l'autre, il faille faire un grand saut conceptuel. On pourrait par exemple imaginer qu'une des dimensions du paysage épistémique de l'astronomie représente les possibles conceptions du système solaire et qu'il y ait sur cette dimension deux sommets très étroits, correspondant respectivement au géocentrisme (moins élevé) et à l'héliocentrisme (plus élevé). À l'Antiquité, la communauté scientifique occidentale adopte, pour cette dimension, une valeur fixe qui correspond au sommet du géocentrisme. Elle se contente pendant longtemps d'explorer d'autres dimensions : postuler des épicycles ou des orbites elliptiques, par exemple. Un saut conceptuel vers le sommet correspondant à l'héliocentrisme est fait par Nicolas Copernic, qui publie son *De revolutionibus orbium coelestium* en 1543. Il faut quelques siècles pour que l'ensemble de la communauté scientifique l'y rejoigne, mais c'est à nouveau, aujourd'hui, une valeur fixe : les astronomes contemporains ne se déplacent plus dans cette dimension particulière ; ils en explorent d'autres.

## Les paysages épistémiques utilisés dans le modèle

Les paysages épistémiques utilisés dans le modèle ont été générés par l'extension NetLogo Landscapes<sup>63</sup>. Cette extension permet d'assigner une valeur entre 0,0 et 1,0 à la variable *objective-value* de chacune des cellules de l'environnement NetLogo. À la figure 5.2 à la page suivante, la valeur 0,0 est représentée par la couleur noire alors que les valeurs qui tendent vers 1,0 sont représentées par un vert de plus en plus clair. Nous avons utilisé pour nos simulations un environnement de  $601 \times 601$ , pour un total de 361 201 cellules.

L'extension permet de générer une soixantaine de paysages différents, mais pour des raisons de puissance computationnelle, il n'aurait pas été réaliste de tester le comportement du modèle sur un si grand nombre de paysages. Nous avons donc choisi d'en garder six parmi ceux qui nous semblaient les plus pertinents :

- (a) Le paysage DE JONG F1 (De Jong, 1975) est généré par une fonction classique généralement utilisée dans le domaine de la computation évolutionnaire pour faire une première évaluation d'un algorithme d'optimisation : elle comporte un seul sommet, central, vers lequel monte de toute part une pente douce et continue. C'est dans ce paysage que notre communauté scientifique simulée devrait obtenir les meilleurs résultats collectifs : c'est-à-dire parvenir rapidement à publier des articles qui se concentrent autour du sommet du paysage.
- (b) Le paysage EXP (Garrett, 2004) offre une contrepartie plus difficile à DE JONG F1 : il comporte aussi un seul sommet, central, mais la pente qui y mène est beaucoup plus « à pic ». La majeure partie de l'espace est de valeur épistémique nulle et ne comporte aucun indice (aucune « pente ») qui mène au sommet du paysage.

---

<sup>63</sup> Nous avons développé cette extension dans le cadre d'un stage prédoctoral effectué au Center for Connected Learning and Computer-Based Modelling de l'université Northwestern. Elle permet d'initialiser les cellules de NetLogo à l'aide de valeurs générées par les fonctions disponibles dans le *Optimization Algorithm Toolkit* de Jason Brownlee (2007). L'extension s'assure que toutes les fonctions retournent des valeurs normalisées entre 0,0 et 1,0, où 1,0 est la valeur à atteindre. Elle est disponible, avec son code source, sur <https://github.com/NetLogo/Landscapes-Extension>.

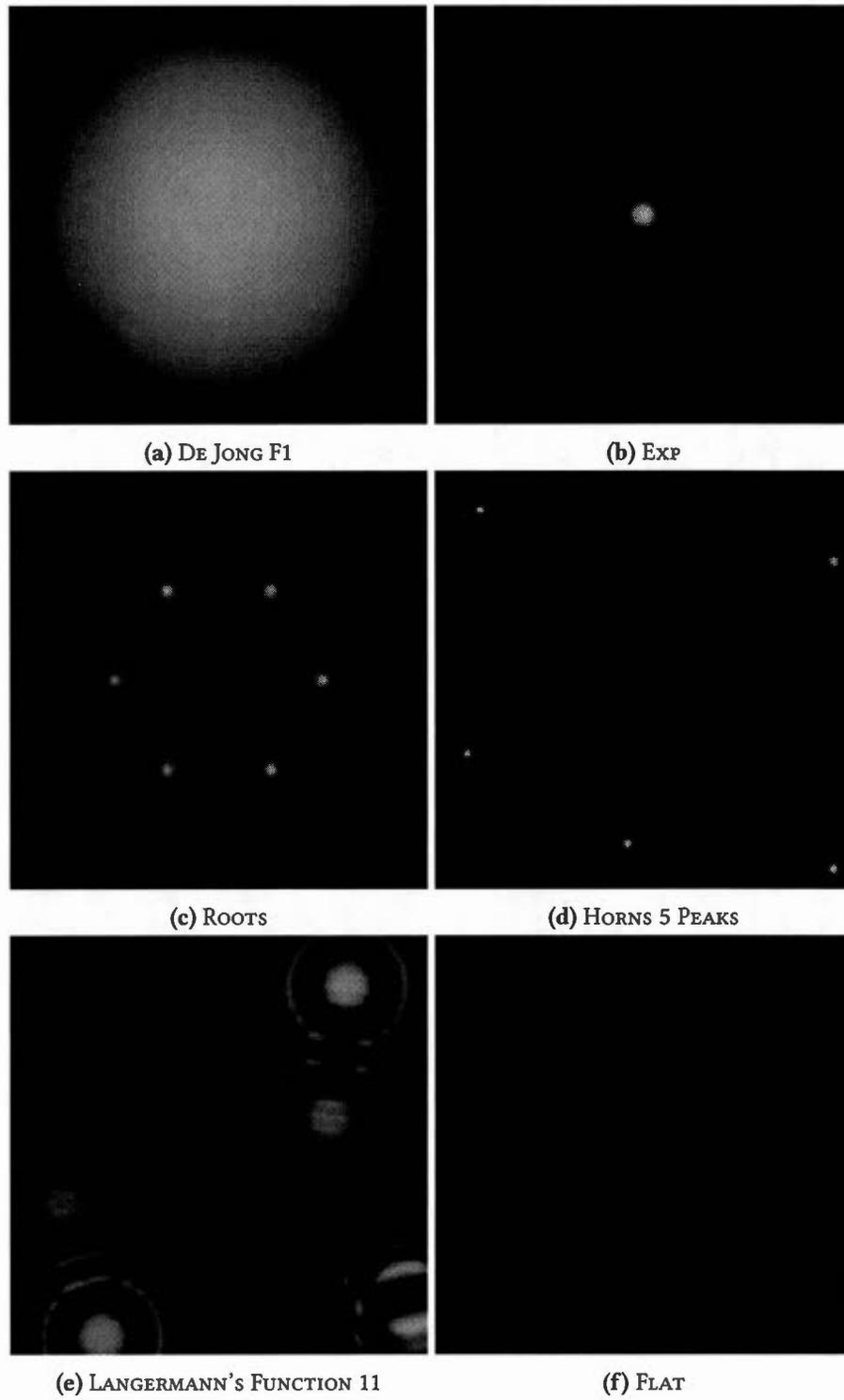


Figure 5.2 Paysages épistémiques

La performance de la communauté dans EXP devrait être bien inférieure à sa performance dans DE JONG F1.

- (c) Le paysage ROOTS (de Castro et Timmis, 2002) comporte six sommets distincts. Seuls deux de ces sommets ont la valeur objective maximale (= 1,0), mais les quatre autres ne sont pas beaucoup moins hauts ( $\approx 0,98$ ). Une pente continue monte vers chacun des sommets et ils devraient pouvoir être atteints facilement par la communauté. À la différence de DE JONG F1 et EXP, ROOTS devrait nous permettre d'observer le comportement d'une communauté fragmentée, où différents chercheurs concentrent leur activité autour de différents sommets.
- (d) Le paysage HORNS 5 PEAKS (Horn et Goldberg, 1994) comporte lui aussi plusieurs sommets (cinq, comme son nom l'indique ; un à 1,0 et les autres juste un peu plus bas), mais ceux-ci ont la particularité d'être séparés les uns des autres par des vallées de basse valeur épistémique, que la communauté devra traverser pour atteindre ces sommets, rendant l'entreprise beaucoup plus difficile que dans le cas de ROOTS, et peut-être même plus difficile que dans le cas de EXP, puisque la structure du paysage tend à induire la communauté en erreur.
- (e) Le paysage LANGERMANN'S FUNCTION 11 (Bersini *et al.*, 1996) est le troisième et dernier de nos paysages à plusieurs sommets. Il y a six, mais seulement trois de valeur objective plus grande que 0,7. Les sommets dans ce paysage sont d'ailleurs moins clairs pour la communauté : c'est comme si chacun d'eux résultait de l'effet d'ondulation créé par un caillou que l'on jette à la surface de l'eau. Ces ondes créent, autour de chaque sommet, de petites vallées successives de valeur épistémique moindre. La capacité de la communauté à performer dans ce paysage dépendra de son aptitude à traverser ces vallées.
- (f) Finalement, le paysage FLAT, comme son nom l'indique, est un paysage complètement plat, où toutes les idées possibles ont la même valeur objective : 0,5. Les agents de la simulation ne sont, bien sûr, pas au courant de ce fait, et ils attribuent quand même des valeurs subjectives différentes à chaque idée. Ce paysage nous permettra de voir quels sont les comportements les plus avantageux pour les individus dans un domaine où la valeur des idées est complètement in-

différenciée (une pseudoscience, par exemple) et seule la dynamique du crédit détermine le succès individuel.

Au bout du compte, la sélection des paysages à utiliser est relativement arbitraire : nous avons simplement essayé d'avoir une certaine variété dans les caractéristiques de ceux-ci, tout en maintenant une certaine plausibilité quant à la possibilité qu'un sous-domaine scientifique soit structuré ainsi<sup>64</sup>. Nous verrons au chapitre 6 que, bien que certains aspects du comportement du système soient robustes d'un paysage à l'autre, d'autres aspects (dont la performance générale de la communauté) sont fortement influencés par la structure du paysage.

#### Génération de la population initiale du modèle

La tâche de l'algorithme de génération de la population initiale (code 5.6 à la page suivante) est de créer les scientifiques et d'attribuer à chacun une idée originale. Ces idées constitueront les ancêtres de toutes les autres idées de la simulation, qui seront créées à partir d'elles et de leurs descendantes. Ce sont, en quelque sorte, les idées séminales du domaine scientifique représenté par le modèle, domaine dont les membres de la population initiale sont les « fondateurs ».

Pour créer une situation initiale qui permet à la population de progresser sur le plan épistémique, l'algorithme doit veiller à ce que les idées initiales ne soient pas créées trop près des sommets de valeur objective : la performance de la communauté doit dépendre de sa capacité à explorer le paysage et à trouver ces sommets.

Les sommets du paysage épistémique (identifiés par la procédure `find-peaks`), sont définis comme étant les cellules de valeur objective plus grande que 0,7 qui ont la valeur objective la plus élevée dans un rayon de cent cellules. Les cellules de départ sont

---

<sup>64</sup>Beaucoup des paysages disponibles dans l'extension `Landscapes` présentent des structures très régulières qui pourraient être plausibles pour des sciences formelles bien particulières, mais qui (intuitivement, du moins) semblent moins aptes à représenter la structure d'une science empirique.

---

```

to create-initial-scientists
  let peaks find-peaks
  let starting-patches rnd:weighted-n-of-with-repeats population patches [
    (distance-to-closest peaks) ^ 2
  ]
  foreach starting-patches [ p ->
    ask p [
      sprout-scientists 1 [
        set color color-of-scientists
        init-scientist
        create-initial-idea
      ]
    ]
  ]
end

```

---

**Code 5.6** Création de la population initiale de scientifiques.

choisies de façon aléatoire<sup>65</sup>, pondérée par le carré de la distance entre chaque cellule et le sommet le plus proche : en bref, plus une cellule est éloignée des sommets, plus elle a de chance de donner naissance à une scientifique et son idée initiale. Le résultat est une population d'idées généralement éloignées des sommets, tout en préservant la possibilité d'avoir quelques idées plus rapprochées de ceux-ci. La figure 5.3 à la page suivante montre des exemples de paysages épistémiques avec les populations initiales de scientifiques. Notez que les scientifiques sont placés dans le paysage à la position de leur idée initiale à des fins d'illustration, mais c'est la position de l'idée qui compte dans le modèle, pas celle du scientifique.

Chaque scientifique est donc doté, au départ, d'une seule idée, dont il doit faire une évaluation préliminaire. À ce stade de la simulation, cependant, la population ne sait rien sur le paysage épistémique et n'a aucune information sur la valeur objective des idées (cela viendra plus tard, lorsqu'ils testeront leurs idées). La première évaluation est donc faite de façon entièrement aléatoire : chaque scientifique attribue à son idée initiale une valeur subjective entre 0 et 1 tirée d'une distribution uniforme.

---

<sup>65</sup>Le générateur de nombres pseudo-aléatoires utilisé par NetLogo est une implémentation Java de l'algorithme Mersenne Twister (Matsumoto et Nishimura, 1998) tirée de la plate-forme MASON (Luke *et al.*, 2005).

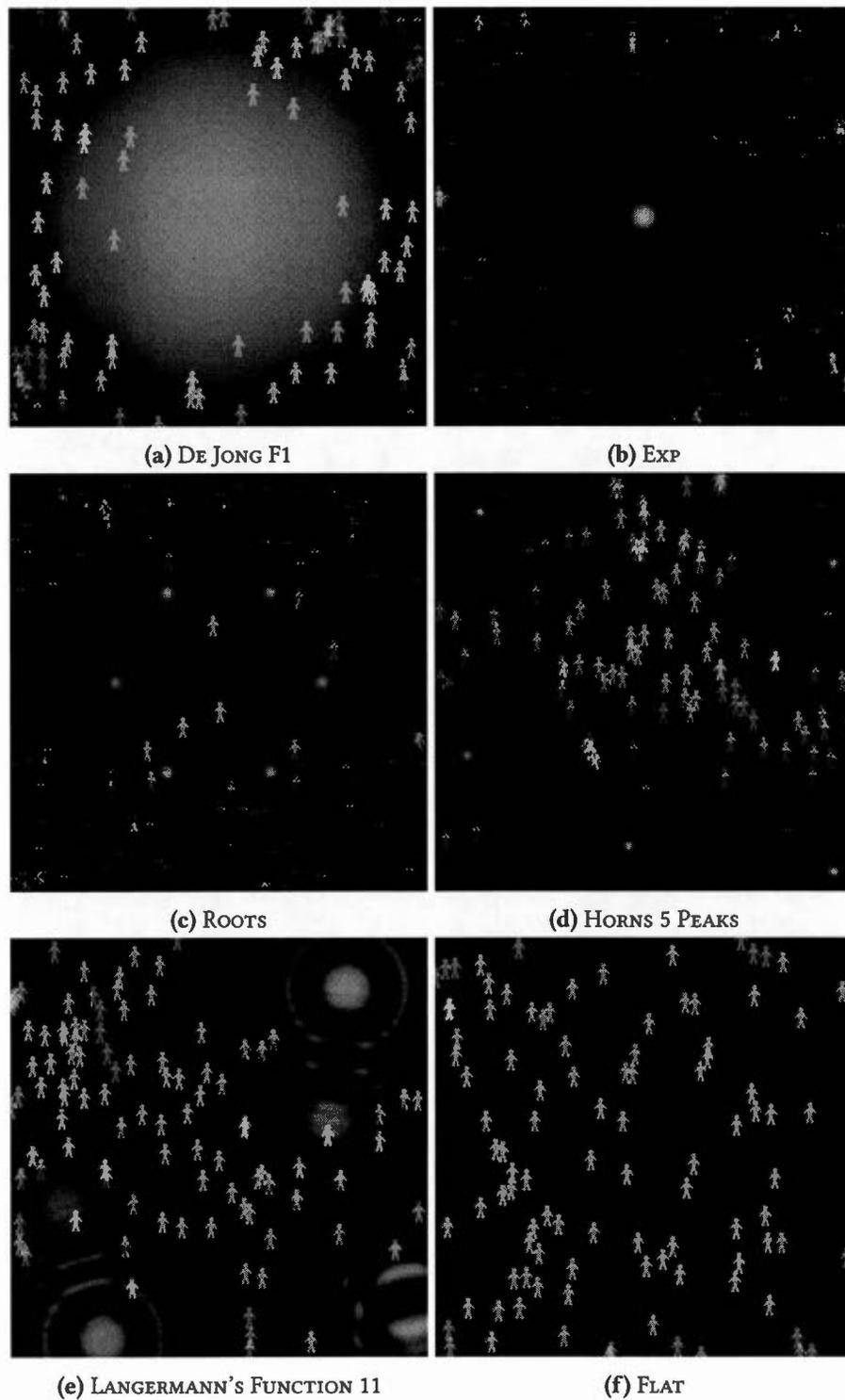


Figure 5.3 Paysages épistémiques avec populations initiales

Lorsqu'un nouveau scientifique est créé, ses caractéristiques personnelles sont aussi initialisées. Celles-ci sont au nombre de quatre et sont identifiées par des lettres grecques :  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  (ou alpha, beta, gamma et delta dans le code NetLogo), et chacune d'elle reçoit, pour chaque scientifique, une valeur tirée de l'ensemble  $\{0; 0,1; 0,2; \dots; 1\}$ . Nous reviendrons plus en détail sur leur rôle dans les sections pertinentes, mais, en bref :

- ▶  $\alpha$  représente le comportement d'une scientifique face au compromis décrit par Hull entre crédit et support lorsque vient le temps de choisir quels articles citer. Plus  $\alpha$  est élevé, plus la scientifique accorde d'importance au support. Moins  $\alpha$  est élevé, plus la scientifique accorde d'importance au crédit. (Voir section 5.4.3.)
- ▶  $\beta$  représente le comportement d'une scientifique par rapport à l'aspect « inclusif » de la valeur sélective conceptuelle inclusive. Plus  $\beta$  est élevé, plus la scientifique aura tendance à citer ses étudiants.
- ▶  $\gamma$  représente l'importance accordée au crédit de l'auteur lors de l'évaluation d'une idée contenue dans un article. Plus  $\gamma$  est élevé, plus la scientifique y accordera d'importance.
- ▶  $\delta$  représente en quelque sorte la « créativité » d'une scientifique : sa tendance à dévier des idées précédentes lorsqu'elle génère de nouvelles idées. Plus  $\delta$  est élevé, plus cette déviation sera grande.

Les valeurs de  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  sont généralement déterminées au hasard, mais dans certaines des simulations dont nous examinerons les résultats au chapitre 6, nous attribuerons une valeur fixe à une de ces caractéristiques et examinerons l'influence que cela aura sur la performance de la communauté en général. Que se passe-t-il, par exemple, lorsque  $\alpha = 0,0$  pour toute la population et que, conséquemment, les scientifiques ne tiennent compte que du support au moment de décider qui citer ? Réponse figure 6.16 à la page 210.

## Algorithme général de la simulation

Ayant identifié les entités qui composent le modèle, les relations qu'elles entretiennent entre elles et la façon dont le modèle est initialisé, il nous faut maintenant nous pencher sur les algorithmes qui règlent le déroulement de la simulation.

---

```

to go
  ask scientists [ generate-new-ideas ]
  display
  ask scientists [ perform-tests ]
  ask scientists [ write-drafts ]
  publish-papers
  ask ideas [ die ]
  display
  ask scientists [ read-papers ]
  update-credit
  retire-scientists
  breed-scientists
  tick
end

```

---

### Code 5.7 Procédure go.

Avant de plonger dans ces procédures, nous allons toutefois en rappeler les grandes étapes et dire quelques mots sur chacune d'elles. Le code 5.7 montre la boucle principale, appelée go, comme le veut la convention NetLogo. Chaque itération de cette boucle représente une « année » dans la simulation, et à chaque année :

1. Les scientifiques génèrent de nouvelles idées. Ces idées sont générées à partir des articles qu'ils ont lus ou écrits.
2. Les scientifiques testent certaines de leurs évaluations.
3. Les scientifiques écrivent de nouveaux manuscrits qui, s'ils sont acceptés pour publication, deviendront des articles.
4. Les idées qui n'ont pas été sélectionnées pour devenir des manuscrits « meurent ».

5. Des évaluateurs sont affectés aux manuscrits et choisissent ceux qui seront publiés par le journal.
6. Les scientifiques lisent certains articles parmi ceux qui ont été publiés et attribuent du crédit aux auteurs de ces articles et des articles qui sont cités dans ceux qu'ils ont lus.
7. Un certain nombre de scientifiques prennent leur « retraite ».
8. Une nouvelle génération de scientifiques arrive dans la communauté.
9. Parmi les scientifiques qui restent actifs, un certain nombre sont choisis pour être les « superviseurs » des nouveaux arrivants et procèdent à leur éducation.
10. Finalement, la commande `tick` augmente le compteur d'une année.

Dans les sections qui suivent, nous reprendrons ces étapes une à une.

## Idéation

Chaque scientifique génère un certain nombre de nouvelles idées chaque année. Ce nombre, donné par le paramètre `nb-new-ideas-per-year`, est le même pour tous les scientifiques. Le modèle ne différencie pas les scientifiques selon leur productivité : chacun d'eux a le même nombre d'idées, effectue le même nombre de tests, rédige le même nombre de manuscrits et lit le même nombre d'articles chaque année. Il s'agit, bien sûr, d'une simplification par rapport au monde réel, mais elle nous permet de tester la performance des scientifiques en fonction d'un nombre de paramètres limités (les caractéristiques  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$ ), toutes choses étant égales par ailleurs.

Pour chaque nouvelle idée générée, il y a trois étapes (voir code 5.8) :

1. Choisir les idées qui seront les « parents » de la nouvelle idée<sup>66</sup> ;
2. Combiner ces parents pour donner naissance à une nouvelle idée ;
3. Évaluer la nouvelle idée et lui attribuer une valeur subjective.

---

```

to generate-new-ideas
  if any? my-out-evaluations [
    repeat nb-new-ideas-per-year [
      let parents n-values 2 [ choose-evaluation-from my-out-evaluations ]
      let new-idea combine-and-mutate parents
      generate-new-evaluation-for new-idea
    ]
  ]
end

to-report choose-evaluation-from [ candidate-evaluations ]
  report rnd:weighted-one-of candidate-evaluations [ subjective-value ^ 2 ]
end

```

---

**Code 5.8** Génération des nouvelles idées.

Les idées parentes sont choisies de façon aléatoire, pondérée par le carré de la valeur subjective qui leur est accordée, parmi toutes les idées pour lesquelles la scientifique a une évaluation (fonction `choose-evaluation-from` dans le code 5.8). Plus la scientifique accorde une valeur subjective élevée à une idée, plus celle-ci a de chances de devenir un parent d'une autre idée.

À l'étape de la combinaison, la nouvelle idée hérite, pour chacune de ses coordonnées ( $x$  et  $y$ ) dans l'espace épistémique, de la coordonnée d'un de ses deux parents. Chaque coordonnée est sélectionnée aléatoirement en fonction du carré de la valeur subjective du parent (fonction `choose-evaluation-from`, ici aussi). Il est donc possible qu'un

---

<sup>66</sup>Comme nous l'avions annoncé à la section 3.2.3, les idées dans le modèle ont généralement deux parents. Il serait possible de faire varier le nombre de parents d'une idée à l'autre, mais pour des raisons de simplicité, nous avons choisi de nous en tenir à la biparentalité. Notons toutefois qu'il est possible qu'une idée soit choisie deux fois comme parent d'une même nouvelle idée ; dans ces cas (plus rares), la nouvelle idée n'a qu'un seul parent.

même parent transmette ses deux coordonnées à la nouvelle idée ; l'autre parent n'apportant dans ce cas aucune contribution. (Ce qui reflète la possibilité qu'une nouvelle idée ne soit pas le résultat de la combinaison de deux idées distinctes.)

Les coordonnées des nouvelles idées ne sont toutefois pas exactement celles de leurs parents : chaque coordonnée est modifiée par une variation tirée d'une distribution normale. L'écart-type de cette distribution est déterminé à la fois par la valeur  $\delta$  attribuée au scientifique et par le paramètre général *max-new-idea-std-dev* (exprimé en pourcentage de la taille du paysage épistémique) selon la formule :  $\delta \times \text{max-new-idea-std-dev}$ . Si, par exemple,  $\delta = 0,5$  et *max-new-idea-std-dev* = 10 %, l'écart-type de la distribution sera de 5%. C'est cette mutation qui permet d'introduire dans le système la variation nécessaire au processus d'évolution. Sans elle, la communauté serait condamnée aux combinaisons de coordonnées déjà présentes lors de l'initialisation du modèle.

Ce processus de sélection, combinaison et mutation des idées permet à ces dernières de jouer le rôle de *réplicateur* exigé par la théorie de Hull, puisqu'il fait de chacune d'elles « an entity that passes on its structure largely intact in successive replications. » (Hull, 1988a, p. 134) C'est aussi une des façons dont les scientifiques jouent leur rôle d'interacteurs puisque, en choisissant les idées qui seront répliquées, ils sont « an entity that interacts as a cohesive whole with its environment in such a way that this interaction *causes* replication to be differential. » (1988a, p. 134)

Notons aussi que le processus de génération de nouvelles idées à partir d'idées existantes donne naissance à des lignées conceptuelles, entités dont Hull fait grand cas puisqu'elles sont, comme les espèces biologiques, des individus spatiotemporellement délimités. Outre le fait de renforcer la caractérisation de la science comme processus évolutionnaire—où cette individualité joue un rôle important selon Hull (1976, 1978, 1980)—cette notion affecte surtout la façon d'identifier les idées qui appartiennent une même théorie : « Like species, theories must be individuated in terms of some sort of descent and cohesiveness, not similarity. » (1978, p. 356). Puisque les idées générées dans le modèle constituent de telles lignées, et il serait possible de mettre en place cette forme d'identification des théories, mais celle-ci n'aurait pour l'instant aucun rôle fonctionnel à jouer.

Nous venons d'expliquer comment une nouvelle idée est créée, mais la tâche du scientifique n'est pas encore tout à fait accomplie : il doit attribuer une valeur subjective à sa nouvelle idée. Cette valeur est calculée en fonction de la position de l'idée dans le paysage épistémique, relativement à celles des autres idées pour lesquelles la scientifique a déjà une évaluation. Pour ce faire, on prend la moyenne des autres évaluations faites par la scientifique, pondérée par l'inverse du carré de la distance normalisée entre les objets de ces autres évaluations et l'idée qu'elle doit évaluer. Il s'agit, en quelque sorte, d'un « modèle gravitationnel » : plus une idée ou un article existant est proche dans l'espace épistémique, plus son influence est grande sur l'évaluation.

Pour trouver la valeur  $v_i$  d'une idée  $i$  à évaluer, étant donné  $I$  l'ensemble des idées déjà évaluées et  $d(i, j)$  la distance entre deux idées, on a :

$$v_i = \sum_{j \in I} \left( v_j \times \frac{\frac{1}{d(i,j)^2}}{\sum_{k \in I} \frac{1}{d(i,k)^2}} \right) \quad (5.1)$$

Le code 5.9 montre l'implémentation NetLogo de cette équation.

```
to-report relative-subjective-value [ location ]
  let total-value 0
  let total-isd 0
  ask my-out-evaluations [
    let d [ distance location ] of end2
    if d > 0 [
      let isd 1 / (d ^ 2) ; inverse square distance
      set total-isd total-isd + isd
      set total-value total-value + (isd * subjective-value)
    ]
  ]
  report ifelse-value (total-value = 0) [ 0 ] [ total-value / total-isd ]
end
```

**Code 5.9** Fonction relative-subjective-value.

Cette valeur subjective relative représente l'évaluation préliminaire que le scientifique fait d'une nouvelle idée. Il se pose simplement la question : « ressemble-t-elle plus à des idées que je trouve bonnes ou à des idées que je trouve mauvaises ? » Le résultat de

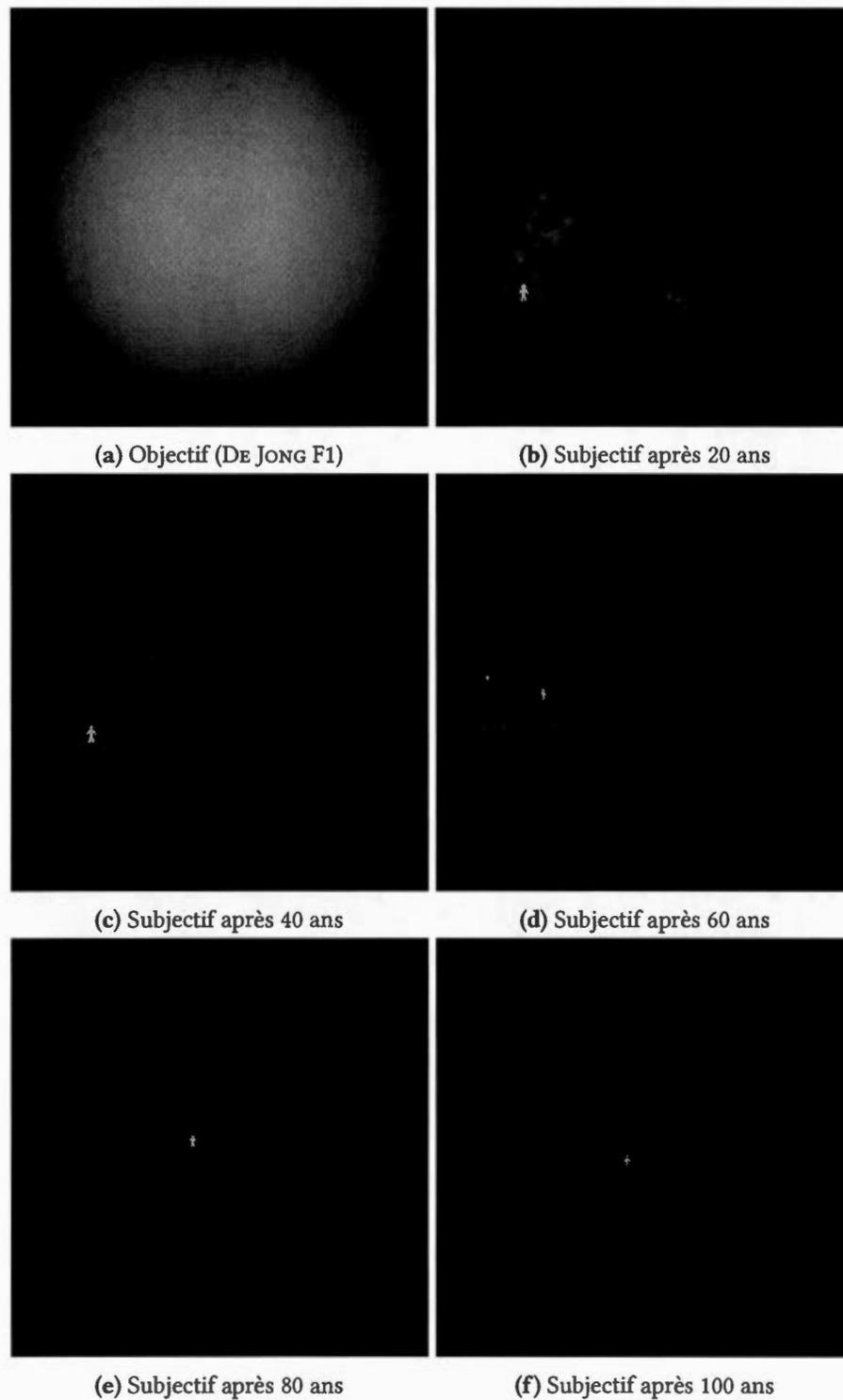
notre modèle gravitationnel, au fond, c'est que chaque scientifique a, « dans sa tête », un paysage épistémique *subjectif*, ou chaque position possible de l'espace possède une valeur (figure 5.4 à la page suivante).

De cette façon, les idées existantes exercent une influence sur le comportement futur du scientifique : elles déterminent la valeur subjective accordée aux nouvelles idées, et cette valeur subjective déterminera à son tour quelles idées seront testées puis, éventuellement, mises par écrit. Nous verrons aussi à la section 5.4.5 que la valeur subjective accordée aux idées présentées dans les articles qui seront lus détermine le crédit qui est accordé aux auteurs de ces articles et que le crédit intervient à son tour dans plusieurs des choix qui sont faits par les scientifiques. En ce sens, on pourra dire qu'on a satisfait à l'exigence de Hull selon laquelle il doit y avoir un lien causal entre les répliqueurs et le comportement des interacteurs (voir section 3.2.2).

À mesure que les évaluations que fait la scientifique changent—parce qu'elle fait des tests (voir section 5.4.2) ou qu'elle lit des articles écrits par des auteurs à qui elle accorde plus ou moins de crédit (voir section 5.4.5)—la forme de ce paysage subjectif change aussi. Pour une scientifique qui aurait une vision parfaite du monde, la forme de ce paysage épistémique subjectif serait identique à celle du paysage objectif. Comme on peut le constater à la figure 5.4, ce n'est pas nécessairement le cas : même la scientifique ayant accumulé le plus de crédit à différentes étapes de la simulation a un paysage subjectif très différent du paysage objectif. Et ceux-ci différencieraient encore plus si ce n'était pas de la possibilité de procéder à des tests empiriques, qui sont le sujet de notre prochaine section.

## Tests

On a parlé dans la section précédente de la façon dont une valeur subjective est attribuée aux idées, mais celles-ci ont aussi, il est bon de le rappeler, une valeur *objective*. Et le but des scientifiques, au bout du compte, c'est de découvrir celle-ci : de s'assurer que la valeur subjective qu'ils attribuent à une idée est le plus près possible de sa valeur objective. Malheureusement, dans le modèle comme dans la réalité, aucun



**Figure 5.4** Paysages épistémiques subjectifs. La figure 5.4a montre un paysage épistémique objectif (DE JONG F1, en l'occurrence). Les figures 5.4b à 5.4f montrent ce même paysage tel que perçu subjectivement par le scientifique qui a couramment le plus de crédit après 20 ans, 40 ans, 60 ans, 80 ans et 100 ans, lors d'une simulation individuelle.

scientifique n'a un accès direct à la valeur objective des idées. Il peut toutefois s'en approcher, et la façon de le faire c'est de procéder à des tests empiriques<sup>67</sup>.

Dans un monde où les ressources de la communauté scientifique seraient illimitées, on pourrait tester toutes les idées. Ce n'est toutefois pas le cas : procéder à des tests prend du temps (et, dans la vraie vie, de l'argent). On ne peut donc en faire qu'un nombre limité. Le choix des idées à tester est donc crucial.

Dans le modèle, le nombre de tests que peut faire chaque scientifique chaque année est déterminé par le paramètre général *nb-tests-per-year*. Le scientifique doit d'abord choisir sur quelles évaluations il souhaite se pencher à lumière du monde empirique. Dans le modèle, nous limitons celles-ci à celles qui n'ont pas déjà été testées, reflétant le fait qu'une scientifique teste rarement deux fois la même idée. Les idées à tester sont choisies au hasard, mais ce choix est pondéré par la valeur subjective de l'idée : plus une scientifique pense qu'une idée est bonne, plus elle a de chances de souhaiter s'en assurer empiriquement. C'est aussi un prérequis à la publication : les scientifiques n'écrivent des articles que sur des idées qui ont été testées.

Notons aussi que les scientifiques ne testent que les idées dont ils sont les auteurs. Selon Hull :

« Some commentators on science urge scientists to test any and all work before they utilize it. They insist that a scientist should personally check everything that goes into any paper that bears his or her name (Broad and Wade 1982). I cannot conceive of worse advice. The whole point of scientists working together is to pool conceptual resources. [...] The best advice for a scientist who begins to doubt the reliability of the work of a colleague is to sever professional ties. One cannot waste one's time checking the result of one's co-workers. » (1988a, p. 131)

Le seul cas où une scientifique devrait consacrer du temps à tester les résultats obtenus par quelqu'un d'autre, c'est lorsqu'il s'agit d'un rival et que les résultats en question menacent ses propres théories. Ces cas sont toutefois suffisamment rares et nous avons choisi de ne pas en tenir compte dans notre modèle.

---

<sup>67</sup>Il est bon de se rappeler ici que nous ne modélisons que l'exploration d'un nombre limité de dimensions du paysage épistémique et que, pour reprendre à nouveau l'image de Quine, le savoir « est une étoffe tissée par l'homme, et dont le contact avec l'expérience ne se fait qu'aux contours » (1980, p. 108) : les dimensions de notre paysage représentent les contours de l'étoffe.

Ce qui se passe dans le modèle lorsqu'un test est effectué est relativement simple (voir code 5.10) : la scientifique obtient un résultat égal à la moyenne entre la valeur subjective attribuée à l'idée et la valeur objective de celle-ci. Ainsi, elle s'approche de la valeur objective, sans toutefois y parvenir entièrement. Cette différence représente l'incertitude reliée au test, de même que le biais relié à l'évaluation subjective préalable de la scientifique.

---

```

to perform-tests
  let candidates my-out-evaluations with [
    not tested? and [ author ] of end2 = myself
  ]
  let n min list count candidates nb-tests-per-year
  ask rnd:weighted-n-of n candidates [ subjective-value ] [
    set subjective-value (subjective-value + [ objective-value ] of end2) / 2
    set tested? true
  ]
end

```

---

**Code 5.10** Sélection des idées à tester et révision des évaluations correspondantes.

## Écriture

Nous allons maintenant nous pencher de plus près sur l'étape d'écriture des articles. La procédure principale de rédaction des manuscrits (code 5.11), exécutée pour chaque scientifique actif dans la simulation, nous montre que cette rédaction se déroule en deux étapes : sélectionner les idées à propos desquelles on veut écrire et procéder ensuite à l'écriture des manuscrits.

---

```

to write-drafts
  let selected-ideas select-ideas-for-drafts
  foreach selected-ideas create-draft
end

```

---

**Code 5.11** Procédure principale de rédaction des manuscrits.

La sélection des idées est relativement simple (code 5.12 à la page suivante). Le scientifique choisit aléatoirement parmi celles qu'il a testées et dont il est l'auteur, et ce choix est pondéré par la valeur subjective qu'il attribue à ces idées. Ici aussi, le scientifique

fait figure d'interacteur : son choix des idées à répliquer par l'écriture a une influence directe sur la reproduction différentielle de celles-ci.

---

```

to-report select-ideas-for-drafts
  let candidates my-out-evaluations with [
    tested? and
    [ breed ] of end2 = ideas and
    [ author ] of end2 = myself
  ]
  let n min list nb-drafts-per-year count candidates
  let selected-evaluations rnd:weighted-n-of n candidates [ subjective-value ]
  report [ end2 ] of selected-evaluations
end

```

---

**Code 5.12** Sélection des idées pour l'écriture.

Chaque manuscrit présente une et une seule idée. L'idée choisie est transformée en manuscrit : c'est-à-dire qu'elle change d'« espèce » NetLogo et devient un agent de type draft qui occupe la même position que l'idée dans le paysage épistémique. Il reste ensuite à choisir des articles à citer et à créer les liens de citation entre le manuscrit et les articles cités.

C'est la première de ces étapes, le choix des références, qui constitue le cœur du processus. L'auteur établit une liste de candidats parmi les articles pour lesquels il a une évaluation (et donc, qu'il a lus) et sélectionne aléatoirement un certain nombre<sup>68</sup> de ceux-ci (nb-references-per-paper).

La sélection des articles est pondérée en fonction de leur « citabilité », concept qui demande quelques explications. C'est ici que nous revenons à deux éléments centraux de la théorie de Hull : le compromis entre crédit et support, et la valeur sélective conceptuelle inclusive. Nous avons abordé ces deux questions plus en détail aux sec-

---

<sup>68</sup> Que le nombre de références soit le même pour tous les articles est une des simplifications du modèle. Une étude de Abt et Garfield montre toutefois que le nombre de références dans les articles scientifiques varie de façon linéaire avec la taille de l'article : « In each of 41 research journals in the physical, life, and social sciences there is a linear relationship between the average number of references and the normalized paper lengths. For most of the journals in a given field, the relationship is the same within statistical errors. » (2002, p. 1106) Dans la mesure où on suppose que notre journal fictif exige des articles d'une longueur particulière, cette simplification nous paraît justifiée.

tions 3.3.1 et 3.3.2, mais je me permets de redonner brièvement la parole à Hull sur ce sujet. D'abord, en ce qui concerne le compromis entre crédit et support :

« One cannot gain support from a particular work unless one cites it, and this citation automatically both confers worth on the work cited and detracts from one's own originality. Scientists would like total credit and massive support, but they cannot. Science is so organized that scientists are forced to trade off credit for support. » (1988a, p. 127)

Puis, au sujet de la valeur sélective conceptuelle inclusive :

« When one fails to cite a well-known author, one gains neither credit nor support. The same observation do not hold for those lower in the scientific hierarchy, in particular one's graduate students. In the short term, their support is worth very little. However, graduate students are not entirely powerless because they are likely to be the chief conduits for one's work to later generations. Their success increases one's own conceptual inclusive fitness. » (2001, p. 102)

Les scientifiques sont donc confrontés, dans un premier temps, au choix de privilégier crédit ou support dans la sélection des articles à citer. Puis, dans la mesure où ils souhaitent obtenir du crédit, ils doivent choisir les articles à privilégier parmi les leurs et ceux de leurs étudiants (et des étudiants de leurs étudiants, et ainsi de suite). Dans le modèle, ces relations sont exprimées par les liens de type *supervisions*.

La position des agents relativement à ces deux axes de choix s'incarne dans deux variables dont la valeur, située dans l'intervalle  $[0; 1]$  peut être différente d'une scientifique à l'autre :

- La variable  $\alpha$  (alpha, dans le code NetLogo). Plus la valeur de  $\alpha$  est élevée, plus une scientifique accordera d'importance au *support* que peuvent lui apporter les articles qu'elle cite. Dans un cas où  $\alpha = 1$ , ce serait son seul critère. Plus  $\alpha$  s'approche de zéro, par contre, plus c'est du *crédit* qu'elle souhaite obtenir.
- La variable  $\beta$  (beta, dans le code NetLogo) qui exprime la stratégie adoptée par la scientifique pour tenter d'augmenter sa valeur sélective conceptuelle inclusive. Plus la valeur de  $\beta$  est élevée, plus une scientifique aura tendance à citer ses étudiants. Dans le cas où  $\beta = 0$ , la scientifique considère que seuls les articles dont elle est elle-même l'auteur valent la peine d'être cités. (Une telle scientifique néglige l'aspect *inclusif* de la valeur sélective conceptuelle inclusive.)

## Évaluation de la valeur sélective conceptuelle inclusive (VSCI)

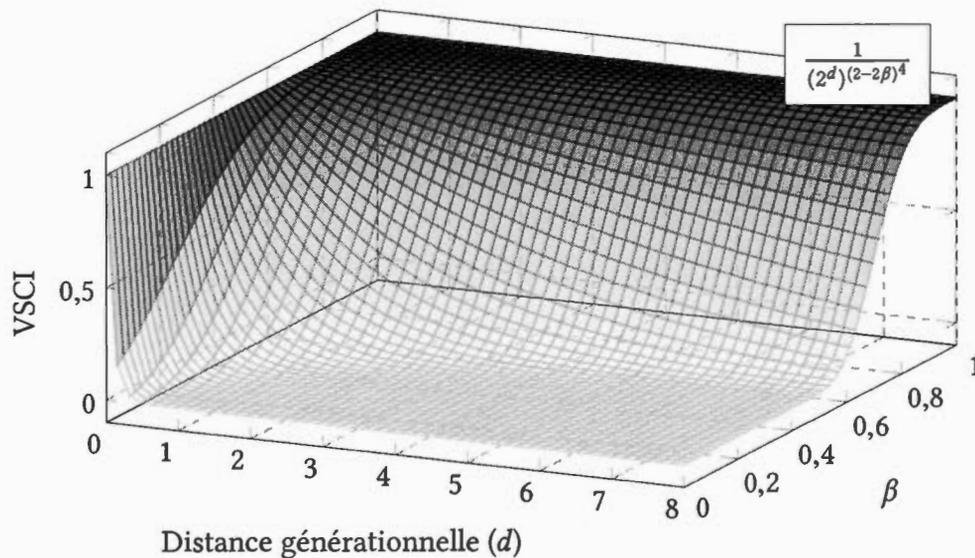


Figure 5.5

Dans le chapitre 6, nous examinerons l'influence des variables  $\alpha$  et  $\beta$  sur le crédit des scientifiques individuels et sur la performance de la communauté scientifique dans son ensemble. Pour l'instant, nous nous contenterons de voir comment elles influencent la « citabilité » d'un article. Celle-ci est définie par l'équation 5.2, où  $c$  représente le crédit normalisé<sup>69</sup> attribué par le lecteur à l'auteur et  $d$ , la distance entre le scientifique et l'auteur de l'article dans la « généalogie » superviseurs/étudiants. Notons que la distance  $d$  n'est définie que s'il existe un chemin entre le scientifique et l'auteur (ce qui explique la clause « si  $d \in \mathbb{R}$  » dans l'équation).

$$\alpha c + (1 - \alpha)f, \quad \text{où } f = \begin{cases} \frac{1}{(2^d)(2-2\beta)^4} & \text{si } d \in \mathbb{R} \\ 0 & \text{sinon} \end{cases} \quad (5.2)$$

<sup>69</sup> Par « crédit normalisé », nous entendons ici la quantité de crédit attribué à l'auteur par le lecteur, divisée par la quantité totale de crédit attribué par le lecteur à tous les scientifiques.

Aucun des deux termes de l'équation ne représente une valeur absolue : les deux sont des indices qui permettent de comparer entre elles les références potentielles en fonction de leur citabilité. Ce calcul n'est pas non plus objectif : il dépend des préférences du scientifique ( $\alpha$  et  $\beta$ ) et du crédit qu'il attribue aux auteurs des articles considérés.

Le premier terme de l'équation représente le degré de support que le scientifique croit pouvoir lui être fourni par l'article concerné. Ce support augmente avec le crédit  $c$  attribué à l'auteur de l'article. Le second terme, lui, représente le degré de valeur sélective conceptuelle inclusive qu'il croit pouvoir lui être fourni en citant l'article concerné. De façon générale, plus l'auteur de l'article à citer est lié de près au scientifique, plus cette valeur est élevée, mais plus  $\beta$  s'approche de zéro, plus elle diminue rapidement avec la distance  $d$  (voir figure 5.5 à la page 154). Pour l'auteur lui-même,  $d = 0$  : ses propres articles sont donc considérés comme lui apportant le maximum de valeur sélective conceptuelle inclusive. Et pour un auteur qui aurait  $\beta = 0$ , ce serait même les seuls articles considérés pour cela : un tel auteur ne citerait ses étudiants *que* si ceux-ci lui apportent un support immédiat. Un auteur pour qui  $\beta = 1$ , au contraire, accorderait autant de poids aux articles de ses descendants intellectuels qu'aux siens propres dans le calcul de la valeur sélective conceptuelle inclusive.

Le code 5.13 à la page suivante montre l'implémentation NetLogo de l'équation 5.2, où l'expression `normalized-credit-for [ author ] of paper` et la fonction `inclusive-fitness-provided-by` correspondent aux deux termes de l'équation.

En ce qui concerne le premier terme l'équation 5.2, il serait légitime de se demander pourquoi c'est le crédit de l'auteur de l'article qui est pris en compte, et non pas le nombre de citations reçues par l'article. Pour comprendre ce choix, il faut garder en tête que nous modélisons des agents qui n'ont pas accès à une information parfaite et illimitée. Il n'est pas inutile non plus de se replacer dans le contexte historique où Hull a élaboré sa théorie : dans les années 1980, le nombre exact de citations reçues par un article donné n'était pas une information facilement disponible : les bases de données bibliométriques comme Google Scholar, PubMed, Scopus et Web of Science n'existaient pas encore. Une façon d'évaluer le support potentiellement fourni par

---

```

to-report citability [  $\alpha$  ] ; paper reporter
  report  $\alpha$  * normalized-credit-for author +
    (1 -  $\alpha$ ) * inclusive-fitness-provided
end

to-report inclusive-fitness-provided ; paper reporter
  let d nw:distance-to author
  report ifelse-value is-number? d
    [ 1 / ((2 ^ d) ^ ((2 - 2 * beta) ^ 4)) ]
    [ 0 ]
end

```

---

**Code 5.13** Implémentation NetLogo de l'équation 5.2.

un article, dans ce contexte, c'est de prendre en compte les citations qu'a reçues cet article dans les articles qu'on a lu. Et puisque les lecteurs attribuent du crédit aux auteurs cités dans les articles qu'ils lisent, le crédit total de l'auteur reflète en partie le nombre de citations qu'il a reçues. De fonctionner avec le crédit de l'auteur a aussi l'avantage de tenir compte du fait qu'un article récent par un auteur respecté peut être une source significative de support même si cet article n'a pas encore recueilli de citations. Ce dernier point nous permet de croire que, encore aujourd'hui, les auteurs ne se contentent pas de regarder le nombre de citations pour identifier les articles qui leur offriront du support<sup>70</sup>.

## Publication

Beaucoup d'articles sont écrits, mais ils ne sont pas tous publiés. Comme le montre l'équation 5.3 à la page suivante, le taux d'acceptation dépend des paramètres généraux du modèle. Si, par exemple, une population de 100 scientifiques, que chacun produit deux manuscrits par année, et que le *Journal* en publie 20, on aura un taux d'acceptation de 10 %, ce qui représente un domaine très compétitif.

---

<sup>70</sup>Cela étant dit, il pourrait être intéressant d'intégrer cette question dans une future version du modèle : en supposant que tous les agents ont accès au nombre de citations recueillies par chaque article, comment devraient-ils intégrer cette information dans leur stratégie de citation ?

$$\text{Taux d'acceptation} = \frac{\text{population} \times \text{nb-drafts-per-year}}{\text{nb-papers-published-per-year}} \quad (5.3)$$

Comment le *Journal* fait-il pour déterminer quels articles seront publiés ? Un premier critère est l'originalité de l'idée : un manuscrit doit présenter une idée qui occupe une cellule du paysage épistémique où aucune autre idée n'a été publiée. Le second critère, c'est la valeur de l'idée. Le *Journal* n'a pas plus accès à la valeur objective de l'idée que les scientifiques individuels : il doit baser sa décision sur des évaluations subjectives. Pour ce faire, il procède comme dans la réalité : il assigne des évaluateurs aux articles. Dans le modèle, un nombre prédéterminé d'évaluateurs (*nb-reviewers-per-paper*) est assigné de façon aléatoire à chaque manuscrit. Cette sélection est pondérée par le crédit que possèdent les évaluateurs<sup>71</sup> : les scientifiques qui ont accumulé beaucoup de crédit ont donc l'occasion d'évaluer plus d'articles et, par conséquent, d'influencer le choix des articles publiés par le journal. Au bout du compte, le *Journal* publie les *nb-papers-published-per-year* qui obtiennent la meilleure évaluation moyenne et les autres manuscrits sont éliminés du modèle.

---

```

to-report review-score [ the-draft ]
  let candidate-reviewers [ other scientists ] of [ author ] of the-draft
  let reviewers rnd:weighted-n-of nb-reviewers-per-paper candidate-reviewers [ my-credit ]
  ask reviewers [
    generate-new-evaluation-for the-draft
    adjust-evaluation-for-credit the-draft [ cited-authors ] of the-draft
  ]
  report mean [ [ subjective-value ] of out-evaluation-to the-draft ] of reviewers
end

```

---

**Code 5.14** Calcul du pointage par les relecteurs d'un article.

---

<sup>71</sup>Dans *Science as a Process*, Hull souligne l'importance du choix des évaluateurs par l'éditeur d'un journal : « [Editors] can influence the results of the refereeing process by their choice of referees. An editor soon learns that certain referees have very stringent standards. They rarely think a paper is good enough to publish. Others are very lenient. Almost anything strikes them as publishable. Also editors are well aware of the preferences of their referees. Some like quantitative papers ; others do not. Some insist that even the most theoretical papers contain masses of data ; others are not so particular. And so on. » (1988c, p. 164). Nous avons choisi de ne pas modéliser le rôle de l'éditeur dans le système.

Le code 5.14 à la page 157 montre le choix des évaluateurs et calcul du pointage pour chaque article. Le calcul du pointage se fait en deux temps. Dans un premier temps, l'évaluateur génère une évaluation subjective basée uniquement sur la position de l'idée présentée dans le manuscrit par rapport aux idées existantes de l'évaluateur : c'est à nouveau l'équation 5.1 à la page 147 qui détermine la valeur subjective attribuée. Dans un second temps, l'évaluation est ajustée en fonction du crédit des auteurs cités dans l'article. Cet ajustement reflète l'aspect « support » offert par les citations : un auteur présente ses idées comme reposant sur les idées d'auteurs réputés a plus de chance de voir ses propres idées acceptées.

Cet ajustement en fonction du crédit procède de la même façon que celui qui a lieu lors de la lecture d'un article (voir la prochaine section et plus particulièrement le code 5.17 à la page 161), à une différence près : le crédit de l'auteur de l'article lui-même n'est pas pris en compte lors de l'ajustement. Cela reflète le fait que le processus d'évaluation se fait à l'aveugle : les évaluateurs ne savent pas qui est l'auteur de l'article et ne peuvent donc pas tenir compte du crédit qu'ils lui accordent dans leur évaluation<sup>72</sup>.

Notons que les évaluateurs jouent un rôle important dans la reproduction différentielle des répliqueurs, même s'il ne s'agit pas ici de leurs propres idées. En choisissant les articles qui seront publiés, ils déterminent quelles idées survivent et pourront être répliquées à nouveau suite à la lecture de l'article par d'autres scientifiques. La lecture est, d'ailleurs, la prochaine étape de notre simulation.

## Lecture

Lors de l'étape de lecture, chaque scientifique doit, dans un premier temps, sélectionner les articles à lire (code 5.15 à la page suivante) puis, dans un second temps, les lire (code 5.16 à la page suivante).

---

<sup>72</sup>Nous avons choisi de modéliser un processus à l'aveugle, mais ce ne sont pas tous les journaux scientifiques qui suivent cette procédure. Dans la réalité, il arrive aussi que les évaluateurs soient en mesure d'identifier l'auteur d'un article même quand l'évaluation est censée se faire à l'aveugle. Il est concevable qu'ils laissent alors leur opinion de l'auteur influencer leur évaluation, mais nous avons choisi de ne pas en tenir compte dans notre modèle.

---

```

to-report papers-to-read
  let candidates papers with [ not in-evaluation-neighbor? myself ]
  let n min list nb-papers-read-per-year count candidates
  report rnd:weighted-n-of n candidates [ citability [ alpha ] of myself ]
end

```

---

**Code 5.15** Fonction papers-to-read.

La sélection des articles à lire fonctionne sur le principe familier consistant à établir une liste de candidats, puis en choisissant un certain nombre d'entre eux aléatoirement en fonction d'un facteur de pondération. Dans ce cas-ci, les candidats sont les articles que la scientifique n'a pas encore lus (et donc pour lesquels elle n'a pas encore fait d'évaluation). Le facteur de pondération, lui, c'est la « citabilité » de l'article (voir équation 5.2 à la page 154). Pourquoi la citabilité, puisqu'il s'agit ici de choisir les articles à lire et non pas à citer ? Parce qu'il est plus utile de lire les articles qu'on pourra éventuellement citer. Et parce qu'une scientifique qui accorde de l'importance à son crédit futur et tient compte de l'aspect inclusif de la valeur sélective conceptuelle (c'est-à-dire, pour qui  $\beta$  est élevé et  $\alpha$  ne l'est pas trop) doit lire les articles de ses étudiants si elle souhaite pouvoir les citer éventuellement. L'équation 5.2 reflète donc les priorités de la scientifique (déterminées par  $\alpha$  et  $\beta$ ) d'une façon qui s'applique également à la lecture et l'écriture.

Encore ici, les scientifiques jouent un rôle d'interacteurs : en choisissant quels articles lire, ils choisissent les idées qui pourront donner naissance à d'autres. Remarquons aussi que, à cette étape, le choix n'a *rien* à voir avec la *valeur* de l'idée : c'est uniquement le crédit de l'auteur et sa proximité dans le réseau de parenté scientifique qui déterminent le choix des articles à lire.

---

```

to read [ the-paper ]
  generate-new-evaluation-for the-paper
  adjust-evaluation-for-credit the-paper [ author-and-cited-authors ] of the-paper
  attribute-credit-for the-paper
end

```

---

**Code 5.16** Procédure read.

Une fois les articles sélectionnés, la lecture de chacun se déroule en trois étapes : la scientifique produit une évaluation « de base » de celui-ci (generate-new-

evaluation-for), ajuste cette évaluation en fonction du crédit qu'elle accorde déjà à l'auteur de l'article *et* aux auteurs des articles qu'il cite (adjust-evaluation-for-credit) et, finalement, attribue du nouveau crédit à l'auteur de l'article qu'elle vient de lire *et* aux auteurs qu'il cite. Cette procédure tient compte du compromis entre crédit et support : le crédit des auteurs cités est pris en compte dans l'évaluation de l'article, mais le nouveau crédit reçu suite à celui-ci doit être partagé avec eux. Remarquez aussi la boucle de rétroaction positive mise en place autour du crédit : plus un auteur possède de crédit, plus ses articles seront évalués positivement, et plus il recevra de nouveau crédit pour ceux-ci. Cette rétroaction positive intervient aussi à un autre niveau, puisque le crédit augmente la probabilité d'être lu, et que c'est lorsqu'on est lu qu'on reçoit du crédit. Comme nous le verrons au chapitre 6 (figure 6.6), cette boucle de rétroaction génère une distribution du crédit fortement asymétrique. Et la seule chose qui limite, au bout du compte, la quantité de crédit qui peut être accumulé par un auteur, c'est le fait qu'on est dans un système fermé : il y a un nombre fini de scientifiques qui liront un nombre fini d'articles chaque année et attribueront à ceux-ci une quantité finie de crédit. Notons aussi que nos simulations sont d'une durée relativement courte : cent ans, dans les expériences présentées au chapitre 6. Dans certains domaines du savoir humain, les durées sont plus longues. Platon, par exemple, continue à recevoir du crédit pour ses œuvres plus de 2 300 ans après leur publication.

Prenons maintenant tour à tour les trois étapes de la lecture d'un article. Celle-ci commence par une procédure generate-new-evaluation-for, qu'il serait peu utile d'inclure ici puisqu'elle se contente de créer le lien d'évaluation entre la scientifique et l'article et d'appeler la fonction relative-subjective-value (code 5.9 à la page 147) qui, elle, fait le travail crucial.

La fonction relative-subjective-value (expliquée plus haut à la section 5.4.1) ne tient compte que de la position des idées dans l'espace épistémique, mais le crédit déjà accordé à l'auteur de l'article et aux auteurs des articles cités doit aussi intervenir dans l'évaluation. La procédure adjust-evaluation-for-credit (code 5.17 à la page suivante) se charge de cette tâche. Elle commence par faire la moyenne du « crédit normalisé » des auteurs concernés. Par « crédit normalisé », on entend le crédit accor-

dé à une scientifique par une autre, divisé par la somme du crédit que cette dernière accorde à *tous* les scientifiques. Cette moyenne résulte en une valeur dans l'intervalle  $[0; 1]$ , qui établit, en quelque sorte, la « crédibilité » de l'article. On ajuste ensuite la valeur subjective de l'article en combinant cette valeur de crédibilité et la valeur subjective relative établie précédemment<sup>73</sup>. C'est la valeur de la caractéristique individuelle  $\gamma$  qui détermine le poids accordé à chacune de ces valeurs. Soit  $v$  la valeur subjective relative de l'article et  $c$  la crédibilité calculée pour lui, la valeur subjective finale qui lui est attribuée est égale à  $(1 - \gamma)v + \gamma c$ . En d'autres termes, plus  $\gamma$  est élevé, plus le lecteur tient compte du crédit dans son évaluation.

---

```

to adjust-evaluation-for-credit [ the-paper authors ]
  let g gamma
  let credit-amount ifelse-value empty? authors [ 0 ] [
    mean map normalized-credit-for authors
  ]
  ask out-evaluation-to the-paper [
    set subjective-value ((1 - g) * subjective-value + (g * credit-amount))
  ]
end

```

---

**Code 5.17** Procédure `adjust-evaluation-for-credit`.

La dernière étape de la lecture consiste à attribuer aux auteurs une quantité de crédit totale égale à la valeur subjective qui vient d'être attribuée à l'article lu (code 5.18 à la page suivante). Cette quantité de crédit est répartie au prorata du nombre d'articles entre l'auteur de l'article lu et les auteurs des articles cités : c'est-à-dire que l'auteur dont deux articles sont cités obtiendra deux fois plus de crédit que l'auteur dont un seul article est cité. Ainsi un auteur qui ne citerait que ses propres articles (c.-à-d., un auteur pour lequel  $\alpha = 0$  et  $\beta = 0$ ) obtiendrait tout le crédit pour son article. Il n'obtiendrait par contre aucun support externe et ne ferait aucun investissement dans sa valeur sélective conceptuelle inclusive. La question de savoir si une telle stratégie est viable est de celles auxquelles nous tenterons de répondre suite à l'analyse des résultats du modèle au chapitre 6.

---

<sup>73</sup>C'est aussi comme cela que fonctionne l'ajustement pour le crédit par les évaluateurs à l'étape de la publication (section 5.4.4), à la différence que le crédit de l'auteur est exclu du calcul lors de celle-ci.

---

```

to attribute-credit-for [ the-paper ]
  let authors [ author-and-cited-authors ] of the-paper
  let the-credit [ subjective-value ] of out-evaluation-to the-paper / length authors
  foreach authors [ a -> add-credit a the-credit ]
end

```

---

**Code 5.18** Procédure `attribute-credit-for`.

## Retraite

Lorsqu'il est question du changement scientifique, on entend souvent évoquer le « principe de Planck » :

« A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it. » (Planck, 1950, p. 33)

Hull lui-même s'est penché sur la question (Hull *et al.*, 1978) en examinant l'âge des scientifiques au moment où ils ont accepté la théorie de l'évolution dans les années qui ont suivi la publication de *On the Origin of Species* (1859) et il n'a trouvé aucune corrélation entre l'âge et le moment de l'acceptation pour ceux qui ont accepté l'évolution darwinienne dans les dix ans qui ont suivi la publication de la théorie, puis une très faible corrélation pour ceux qui l'ont acceptée plus tard (voir aussi Hull, 1988c, p. 379–382, 533). Quoi qu'il en soit, le fait demeure : il faut bien, éventuellement, que les scientifiques prennent leur retraite, et notre modèle doit en tenir compte.

La retraite (et l'éducation, que nous aborderons à la section suivante) est aussi un cas de sélection, au sens où l'extinction et la prolifération différentielle des interacteurs causent la perpétuation différentielle des répliqueurs. Il s'agit, bien sûr, d'extinction *en tant que scientifique*, pas d'extinction littérale. Comme le dit William Bechtel à propos de la théorie de Hull :

« the scientist is extinguished if he or she is no longer able to influence other scientists. This means that if a scientist is not able to teach students or get papers published in appropriate journals, then, even if the scientist continues to carry out investigations and make major discoveries, he or she is largely extinguished as an interactor. » (1988, p. 158)

Le modèle choisit donc, chaque année, un certain nombre (déterminé par le paramètre population-turnaround) de scientifiques qui quitteront la communauté pour faire de la place aux nouveaux arrivants (code 5.19 à la page 165). Ce ne sont pas nécessairement les plus vieux : les retraités sont choisis de façon aléatoire, pondérée par l'équation 5.4, où  $a$  représente l'ancienneté du scientifique (c'est-à-dire la durée de sa carrière en années) et  $c$ , la quantité de crédit qu'il a accumulée, normalisée dans l'intervalle  $[0; 1]$ . Le premier terme de l'équation 5.4 produit une « courbe en baignoire » (voir figure 5.6 à la page suivante) où la probabilité de quitter le monde académique est forte en début de carrière, mais diminue rapidement ensuite, pour finalement remonter un peu avant 30 ans de carrière et approcher 1,0 après 35 ans. Selon une étude du Gouvernement du Canada (2005), l'âge moyen d'obtention d'un doctorat dans les universités canadiennes en 2003–2004 était d'environ 36 ans. Trente-cinq ans de carrière plus tard, ça nous mène à 71 ans. Nous ne soutenons toutefois pas que cette courbe représente exactement la probabilité de retraite dans une communauté scientifique donnée : seulement qu'elle est suffisamment plausible pour les besoins de notre modèle.

L'effet de l'ancienneté sur la probabilité de retraite est contrebalancé par le crédit accumulé (second terme de l'équation 5.4) : plus une scientifique a de crédit, moins elle a de chance de se retirer. Cela vaut en fin de carrière, bien sûr, où les professeurs émérites continuent à contribuer à la recherche, mais aussi en début de carrière, où les jeunes scientifiques qui ont réussi à publier ont plus de chances de s'accrocher. La figure 5.7 à la page suivante montre comment le crédit vient « pencher » la courbe générée par le premier terme de l'équation 5.4.

$$\max\left(\frac{1}{2a}, \frac{1}{1 + e^{-(a-30)}}\right) + \frac{1-c}{2} \quad (5.4)$$

## Éducation

En plus du rôle crucial que le crédit joue dans la dynamique de publication (c'est-à-dire augmenter les chances d'être lu et d'être cité), il joue un rôle non moins crucial

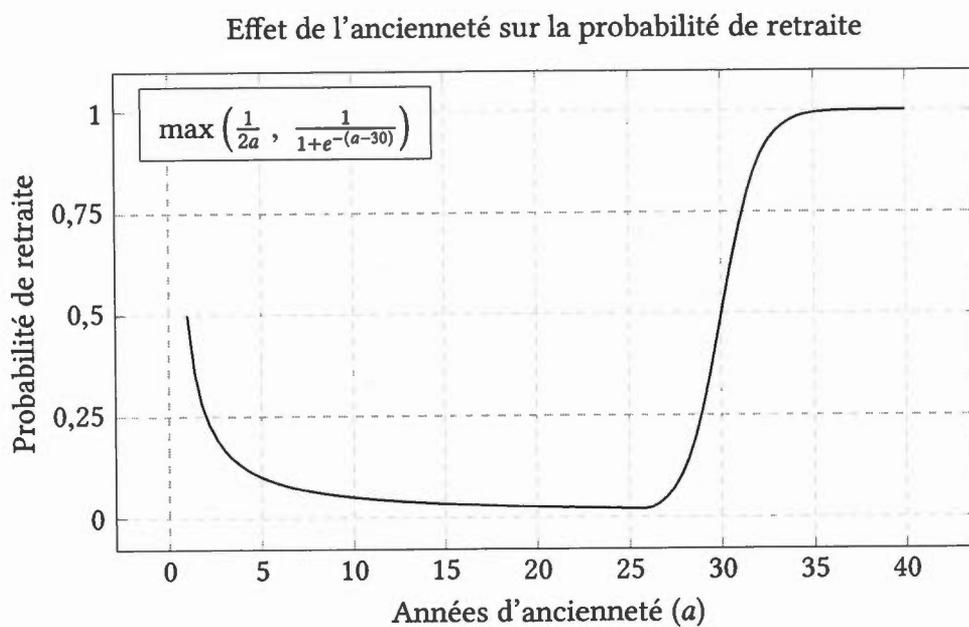


Figure 5.6

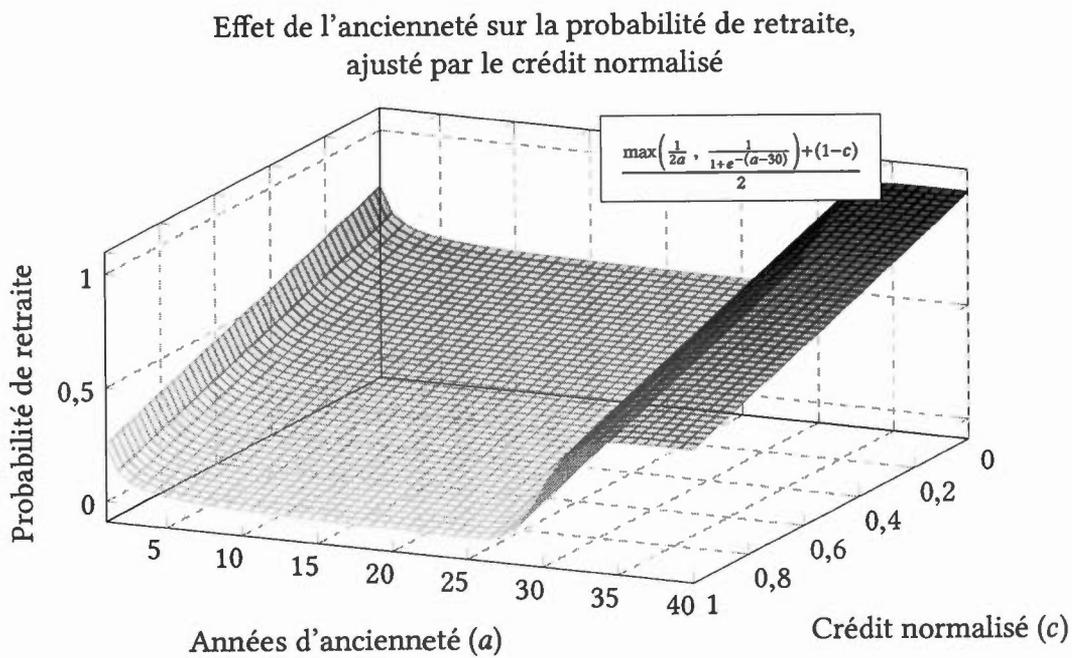


Figure 5.7

---

```

to retire-scientists ; observer procedure
  let n max list 0 (count scientists - (population - population-turnaround))
  ask rnd:weighted-n-of n scientists [ retirement-probability ] [
    retire
  ]
end

```

---

**Code 5.19** Choix des scientifiques qui prendront leur retraite.

dans la dynamique de l'éducation : attirer des étudiants et s'assurer que ceux-ci obtiennent éventuellement un emploi. Notez l'analogie avec le monde biologique où un organisme doit non seulement se reproduire, mais aussi s'assurer de la survie de ses rejetons.

Chaque année, on a autant de nouveaux scientifiques qui font leur entrée dans la communauté qu'on a eu de scientifiques qui l'ont quittée (le paramètre *population-turnaround*). On peut considérer qu'il s'agit des chercheurs qui viennent d'obtenir un doctorat et qui se dénichent un stage postdoctoral ou un premier emploi de professeur. Il peut aussi supposer, par ailleurs, qu'il y a grand nombre d'étudiants (pas modélisés) qui ne terminent pas leur doctorat ou quittent le monde académique après l'obtention de celui-ci. Le rôle du crédit est donc double : attirer les bons étudiants, qui ont plus de chances d'être diplômés, et, par le prestige qui rejaillit sur eux, augmenter les chances que ceux-ci se trouvent un emploi.

Une fois établi le rôle du crédit, le processus d'éducation est relativement simple (voir code 5.20 à la page suivante). On choisit les superviseurs de façon aléatoire, pondérée par le crédit accumulé :

« Single scientists exist through time, sometimes engendering professional descendants, sometimes not. » (Hull, 1988c, p. 508)

Un même scientifique peut être choisi plusieurs fois (et donc voir plusieurs de ses étudiants se joindre à la communauté scientifique la même année).

Chacun de ces nouveaux scientifiques est initialisé de la même façon que les scientifiques initiaux et se voit attribuer une valeur pour chacune de ses variables  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  (aléatoire ou fixe, selon les expériences—voir chapitre 6). Reste à leur donner accès à des idées, qu'ils pourront ensuite utiliser pour générer leurs propres idées.

```

to breed-scientists
  let supervisors rnd:weighted-n-of-with-repeats population-turnaround scientists [ my-credit ]
  foreach supervisors [ supervisor ->
    create-scientists 1 [
      init-scientist
      move-to supervisor
      create-supervision-from supervisor [
        show-or-hide-link show-supervision-links?
      ]
      add-credit supervisor 1.0
      foreach [ self ] of papers with [ author = supervisor ] read
    ]
  ]
end

```

**Code 5.20** Éducation des nouveaux scientifiques.

Contrairement aux scientifiques initiaux, qui produisent *ex nihilo* des idées entièrement originales, les nouveaux scientifiques sont d'abord exposés aux idées de leur superviseur.

Cette exposition se fait, dans le modèle, à travers la lecture de tous les articles écrits par le superviseur. Est-ce tous les étudiants qui, dans le monde réel, lisent *tous* les articles écrits par leur superviseur? Certainement pas. Est-il réaliste de supposer qu'ils ne lisent *que* ceux-ci? Non plus. Mais au bout du compte, cette simplification<sup>74</sup>

---

<sup>74</sup>On pourrait, il est vrai, ajouter à la simulation un paramètre qui déterminerait le pourcentage des articles de leurs superviseurs qui sont lus par les étudiants, mais cet ajout soulèverait plusieurs autres questions : Quelle doit être la valeur de ce paramètre? Cette valeur varie-t-elle d'un étudiant à l'autre? Si oui, comment ces variations sont-elles distribuées dans la population? Et si ce ne sont pas tous les articles de leurs superviseurs qui sont lus par les étudiants, comment choisissent-ils ceux qu'ils lisent? De plus, la situation de l'étudiant d'un jeune professeur, qui n'a encore que peu de publications, est différente de celui qui rejoint un superviseur plus expérimenté, avec des tonnes d'articles à son actif. Aucune de ces questions ne pose d'obstacle insurmontable, mais les réponses qu'on pourrait leur donner exigeraient de nouvelles études empiriques ou, faute de mieux, d'autres décisions basées sur nos intuitions. Le processus de modélisation multi-agents a le mérite de souligner à grands traits l'ampleur des trous dans notre connaissance du domaine abordé.

a des conséquences qui nous semblent, elles, représenter adéquatement la situation qui prévaut au terme du processus d'éducation réel :

- Les étudiants sont exposés aux idées défendues par leur superviseur et peuvent les utiliser pour créer ensuite leurs propres idées (causant ainsi la reproduction différentielle de ces répliqueurs) ;
- Les étudiants, n'ayant lu à leur arrivée que ces articles, n'auront d'autre choix que de les citer dans leurs premières publications, ce qui est cohérent avec la thèse de la valeur sélective conceptuelle inclusive de Hull ;
- Les étudiants, lors de la lecture de ces articles, attribuent du crédit à leur superviseur *et aux auteurs cités par celui-ci*, ce qui inclut fort probablement le superviseur de leur superviseur, et ainsi de suite. Ce sont donc ces auteurs qu'ils liront (et citeront) en priorité par la suite.

C'est avec cette étape que prend fin chaque « année » de notre simulation. Suite à elle, la boucle recommence : idéation, tests, écriture, publication, lecture, retraite et éducation adviennent de nouveau, et ainsi de suite jusqu'à ce qu'on atteigne un nombre d'itérations arbitrairement fixé.

## Conclusion

Notre description du modèle est maintenant complète. La tâche de modéliser la théorie de Hull n'était pas triviale : *Science as a Process* est une œuvre massive et, quoiqu'on y trouve bel et bien une théorie, celle-ci n'est pas présentée de façon formelle<sup>75</sup>. Cela étant dit, pour les besoins de l'exercice, on peut se référer à l'article « A Mechanism and Its Metaphysics : An Evolutionary Account of the Social and Conceptual Development of Science » (Hull, 1988a) qui résume les principaux éléments de celle-ci. Le tableau 5.1 à la page suivante tente d'identifier ces différents éléments, donnant la ré-

---

<sup>75</sup>C'est d'ailleurs un des avantages de l'approche par modélisation que de *forcer* une présentation formelle

férence chez Hull (1988a) lorsqu'applicable. Pour la théorie de Hull autant que pour notre modèle, les éléments centraux sont indiqués par un grand crochet (✓) et les éléments périphériques par un petit (✓).

**Tableau 5.1**  
Comparaison du modèle et de la théorie de Hull

Élément	Théorie	Modèle	Réf. (Hull, 1988a)	Réf. modèle
Compromis entre crédit et support	✓	✓	p. 127	Sections 5.4.3 et 5.4.5
Valeur adaptative conceptuelle inclusive	✓	✓	p. 128	Sections 5.4.3 et 5.4.5
Norme contre le mensonge	✓	-	p. 129-130	
Norme contre le vol de propriété intellectuelle	✓	-	p. 130	
Structure démique de la science	✓	✓	p. 130-132	
Tests	✓	✓	p. 131, 140-141, 148	Section 5.4.2
Réplicateurs	✓	✓	p. 134-136, 140-149	Section 5.4.1
Interacteurs	✓	✓	p. 134-136, 140-149	Sections 5.4.1 à 5.4.7
Sélection	✓	✓	p. 134-136, 140-149	Section 5.4.1
Lignées	✓	✓	p. 135-138, 150-153	Section 5.4.1
Paysage épistémique	-	✓	-	Section 5.2

La plupart de ces éléments ont déjà été traités précédemment, au fil de la description du modèle. Le tableau 5.1 donne les références appropriées. Les seules questions qui n'ont pas été abordées directement, ce sont celle du vol, du mensonge et la structure démique de la science.

Hull considère que sa théorie explique la nature autorégulatrice (*self-policing*) de la science et, en particulier, les normes contre le « mensonge » (c.-à-d., la publication de résultats frauduleux) et le « vol » (c.-à-d., l'appropriation du crédit pour les recherches d'un tiers). La norme contre le mensonge est plus forte que celle contre le vol puisque l'utilisation de résultats erronés peut nuire à tous ceux qui les utilisent alors que l'appropriation illégitime du crédit n'a pour victime que la personne dépouillée. Modéliser

l'émergence de ces normes aurait été un projet intéressant, mais fort différent de celui auquel nous nous sommes attaqués. Le vol et le mensonge ne sont pas représentés dans le modèle actuel.

Finalement, Hull emprunte à la biologie la notion de « dème », qui désigne une sous-population à l'intérieur de laquelle les échanges de gènes sont fréquents. En science, ce sont les groupes de recherches, où les échanges d'idées sont fréquents, qui constituent des dèmes. Selon Hull, ces groupes sont suffisamment cohérents pour agir comme interacteurs dans le processus scientifique et même faire l'objet de sélection différenciée. Le modèle reproduit partiellement cette structure en générant des liens de parenté entre les superviseurs et leurs étudiants (ce qui a pour effet de lier entre les étudiants d'un même professeur, et leurs étudiants à eux, etc.) Dans la mesure où les scientifiques accordent de l'importance à la valeur adaptative conceptuelle inclusive (c.-à-d.,  $\beta > 0$ ), les membres d'un même groupe se citent entre eux, favorisant ainsi l'accumulation de crédit pour tous les membres du groupe et, par le fait même la survie du groupe. Aurions-nous gagné à avoir une structure de groupe plus explicite dans le modèle, avec des réseaux de collaboration (distinct des liens de parenté) dont la structure aurait pu évoluer au fil du temps? Sans doute. De premières ébauches du modèle faisaient intervenir des liens de collaboration explicite, mais comme ce fut le cas lorsque nous avons tenté d'intégrer les articles coécrits dans le modèle, leur présence venait grandement complexifier le modèle. Qu'est-ce qui fait qu'une scientifique offre à une autre de collaborer avec elle? Qu'est-ce qui fait que la seconde accepte ou refuse? Qu'est-ce qui fait qu'un lien de collaboration peut être brisé? À quelle fréquence les collaborateurs s'échangent-ils des idées? Comment évaluent-ils les idées de leurs collaborateurs? Toutes ces questions peuvent, bien sûr, trouver des réponses, et nous aurions pu choisir de concentrer nos efforts sur celles-ci. Mais dans la mesure où nous étions surtout intéressés par le rôle du compromis entre crédit et support et de la valeur sélective conceptuelle inclusive, nous avons tenté de ne garder dans le modèle que les éléments essentiels à l'étude de ceux-ci. Nous sommes toutefois bien conscients qu'une étude plus poussée des réseaux de collaboration scientifique est à la fois nécessaire à notre compréhension des aspects sociaux de la science et une cible tout à fait appropriée pour les modèles multi-agents.

## CHAPITRE VI

### RÉSULTATS DE SIMULATION ET ANALYSE

Dans le chapitre 5, nous avons présenté un modèle multi-agents du processus scientifique inspiré de la théorie de Hull. Le modèle lui-même possède selon nous une valeur intrinsèque : il nous a permis de donner une description précise, formelle, de plusieurs concepts abordés par Hull dans sa théorie. Tant les facteurs internes à la science (la relation entre les idées et le monde et l'exploration de celle-ci par l'intermédiaire des tests empiriques) que les facteurs externes (la diffusion des idées dans la communauté scientifique et l'influence du crédit sur celle-ci) ont trouvé leur place dans un cadre évolutionnaire fidèle à celui suggéré par Hull.

Mais un modèle multi-agents ne se limite pas à formaliser un cadre théorique : il nous permet d'explorer les conséquences de cette théorie par l'intermédiaire de la simulation. Comme nous l'avons annoncé dans l'introduction, nous testerons trois hypothèses relatives à la théorie de Hull. Nous les répétons ici :

$H_1$  : Hull suppose que les scientifiques doivent faire un *compromis* entre crédit et support : « Science is so organized that scientists are forced to trade off credit for support. » (1988a, p. 127) La position adoptée vis-à-vis à ce compromis est représentée pour chaque scientifique par la variable  $\alpha$ . Nous faisons l'hypothèse que les valeurs les plus avantageuses de  $\alpha$  pour un scientifique individuel devraient se situer autour de 0,5.

H<sub>2</sub> : Hull suppose que les scientifiques doivent tenir compte de la valeur sélective conceptuelle *inclusive* lorsqu'ils font le choix des articles à citer : « graduate students [...] are likely to be the chief conduits for one's work to later generations. Their success increases one's own conceptual inclusive fitness. » (1988a, p. 128) Ce choix est représenté dans le modèle par la variable  $\beta$ . Nous faisons l'hypothèse que les valeurs les plus avantageuses de  $\beta$  pour un scientifique individuel devraient se situer autour de 0,5.

H<sub>3</sub> : Hull suppose que l'intérêt des scientifiques individuels coïncide avec l'intérêt de la communauté scientifique en général : « By and large, what is good for the individual scientist is actually good for the group. » (1988a, p. 129). Nous faisons l'hypothèse que les valeurs de  $\alpha$  et de  $\beta$  qui sont les plus avantageuses pour les scientifiques individuels seront aussi les plus avantageuses pour la communauté en général.

Les résultats de notre modèle ne se limitent pas à voir si ces hypothèses sont confirmées ou non, mais celles-ci nous ont servi de guide dans l'élaboration du modèle et dans le choix des paramètres à tester. Dans la section qui suit (6.1), nous décrivons les expériences réalisées pour tester ces hypothèses. Les résultats de ces expériences seront ensuite exposés en quatre temps, passant des caractéristiques plus générales du système aux effets des variations individuelles sur le comportement de celui-ci :

- ▶ Nous nous pencherons d'abord sur la performance épistémique générale de la communauté : réussit-elle, dans les différents paysages épistémiques, à trouver les idées qui ont la plus haute valeur objective ou, à tout le moins, à trouver des sommets locaux ? (Section 6.2.)
- ▶ Nous analyserons ensuite quelques distributions qui résultent de la simulation. Cela inclut des distributions bibliométriques comme la répartition des articles par auteurs, la répartition des citations par articles et la répartition des citations par auteurs, mais aussi la distribution du crédit par scientifiques (Section 6.3.)
- ▶ Nous examinerons ensuite l'effet des différentes caractéristiques individuelles sur la performance « sociale » des scientifiques, c'est-à-dire sur la quantité de

crédit qu'ils arrivent à accumuler. L'examen de  $\alpha$  et  $\beta$  nous permettra de tester les hypothèses  $H_1$  et  $H_2$ , mais nous regardons aussi l'effet de  $\gamma$  et  $\delta$  sur l'accumulation de crédit. (Section 6.4.)

- Nous nous tournerons ensuite vers la performance épistémique de la communauté scientifique en fonction des caractéristiques  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$ . Nous verrons l'effet qu'aurait de fixer celles-ci à différentes valeurs pour la communauté dans son ensemble sur la valeur objective moyenne des articles produits pendant une simulation. En parallèle, nous aborderons  $H_3$  : est-il vrai que ce qui est « bon pour le scientifique » est « bon pour la science » ? La communauté performe-t-elle mieux avec les valeurs qui permettent d'accumuler le plus de crédit ? (Section 6.5.)

Nous compléterons le chapitre par un résumé des conclusions auxquelles nous serons parvenues et de ce que le modèle peut nous apprendre.

## Expériences

La première condition pour pouvoir tester nos hypothèses est de définir ce qu'on entend par « avantageuses pour une scientifique individuelle » et « avantageuses pour la communauté en général ».

Pour la communauté en général, la mesure toute désignée est la moyenne de la valeur objective des articles publiés. Plus la communauté trouve rapidement les sommets du paysage épistémique et mieux elle arrive à concentrer son activité scientifique autour de ces sommets une fois qu'elle les a trouvés, plus la valeur objective moyenne des articles publiés sera élevée.

Pour les scientifiques individuels, il faut plutôt mesurer le crédit qu'ils arrivent à accumuler. Contrairement à la valeur objective des articles, qui est fixe, la quantité de crédit attribuée à une scientifique varie au fil de sa carrière et même après sa retraite, puisque ses articles peuvent continuer d'être lus et cités. Il serait « injuste » de considérer simplement le total du crédit accumulé au fil du temps, puisque les scientifiques

arrivés en début de simulation ont plus de temps pour profiter de ce crédit « post-hume ». Pour tenir compte de cette situation, la mesure utilisée sera la quantité totale de crédit attribuée à la scientifique au fil du temps, divisée par le nombre d'années écoulées entre le début de sa carrière et la fin de la simulation. Quand nous parlerons du crédit d'une scientifique dans les sections qui vont suivre, c'est toujours de cette mesure qu'il sera question.

Finalement, il faut voir comment nous allons déterminer les valeurs des différents paramètres du modèle pour pouvoir tester nos hypothèses.

Il y a, d'abord, un ensemble de paramètres que nous ne ferons pas varier, mais pour lesquels il faut tout de même établir une valeur. Le tableau 6.1 montre les valeurs choisies pour ces paramètres et donne une brève description de chacun d'eux.

**Tableau 6.1**  
Paramètres constants utilisés dans les simulations

Paramètre	Valeur	Description
max-new-idea-std-dev	5 %	Écart type maximum pour la création de nouvelles idées, exprimé en pourcentage de la taille du paysage
nb-drafts-per-year	2	Nombre de brouillons par scientifique par année
nb-new-ideas-per-year	10	Nombre de nouvelles idées produites par scientifique par année
nb-papers-published-per-year	20	Nombre d'articles publiés par le journal par année
nb-references-per-paper	15	Nombre de références par article
nb-reviewers-per-paper	3	Nombre de relecteurs assigné par le journal par article
nb-tests-per-year	5	Nombre de tests effectués par scientifique par année
population	100	Population totale des scientifiques
population-turnaround	10	Nombre de scientifiques quittant et arrivant par année

Plusieurs des paramètres du modèle pourraient théoriquement être calibrés en fonction d'un domaine scientifique et d'un journal en particulier. Certaines valeurs

pour ces paramètres sont disponibles sur des bases de données bibliographiques<sup>76</sup> : nb-papers-published-per-year, nb-references-per-paper. Certaines demanderaient plus de recherches, mais pourraient tout de même être obtenues : population, population-turnaround, nb-new-scientists-per-year, nb-drafts-per-year, nb-reviewers-per-paper. Dans d'autres cas, par contre, il faudrait se contenter d'estimations plausibles, ou alors procéder à des recherches sociologiques originales : nb-tests-per-year, nb-papers-read-per-year, nb-new-ideas-per-year, new-idea-std-dev. Ces deux derniers paramètres, en particulier, sont difficilement opérationnalisables : Qu'est-ce qui compte comme une nouvelle idée ? Et comment mesure-t-on la distance entre une idée et une autre ? Nous n'avons pas, pour l'instant, de réponses à ces questions.

Il y a deux raisons principales pour lesquelles nous n'avons pas choisi de calibrer notre modèle en fonction d'un domaine particulier. La première, c'est que malgré les liens apparents qu'elle entretient avec ces disciplines, notre démarche n'est pas une étude de scientométrie ou même de sociologie des sciences. Ce que nous cherchions à faire, c'est de traduire les concepts principaux de la théorie de Hull dans un modèle multi-agents et de voir comment un tel modèle peut être utilisé pour explorer l'influence de différents facteurs sur la performance individuelle et collective de la communauté scientifique. Des concepts comme la nature évolutionnaire du processus scientifique, le rôle du crédit et celui de la valeur sélective conceptuelle inclusive sont censés être suffisamment généraux pour s'appliquer à toutes les communautés scientifiques et ne devraient pas dépendre, en général, des idiosyncrasies de l'une d'entre elles. Si c'est bien le cas, la petite communauté scientifique fictive décrite par les paramètres ci-dessus est tout à fait appropriée pour explorer les dynamiques identifiées par Hull.

Il est possible, par ailleurs, que le monde réel recèle quelques exceptions à la théorie de Hull. On peut penser, par exemple, au collectif de mathématiciens français anonymes qui publient depuis 1939 sous le pseudonyme de Nicolas Bourbaki, renonçant ainsi au crédit individuel normalement recherché lors de la publication. Hull lui-même remarque que :

---

<sup>76</sup>Par exemple, sur <http://www.scopus.com>.

« If gaining credit from one's peers is as central to the effective functioning of science as I claim it is, Bourbaki is more than a curiosity. It is an anomaly that must eventually be converted into a confirming instance. » (1988c, p. 292–293)

Il faudra bien, un jour, si on prend la théorie de Hull au sérieux, actualiser ce « *eventually* » dans le cas de Bourbaki, mais aussi tenter de voir comment Hull peut nous aider à comprendre l'évolution de domaines scientifiques particuliers. C'est un peu ce qu'ont essayé de faire O'Brien, Lyman et Schiffer dans un ouvrage intitulé *Archaeology as a Process* (2007) où ils retracent, d'une façon explicitement inspirée par Hull, la généalogie des idées archéologiques et les relations entre les différents scientifiques œuvrant dans le domaine depuis les années 1960. Ce n'est pas, toutefois, notre objectif à nous.

L'autre raison pour laquelle nous avons choisi de modéliser une communauté scientifique fictive est plutôt pragmatique : nous sommes limités par nos ressources computationnelles. Les paramètres de la simulation ont été choisis de façon à être plausibles tout en limitant le plus possible la durée d'exécution d'une simulation. Notre échelle est donc plus petite que celle de la plupart des journaux scientifiques. Le journal *Systematic Biology*, pour prendre un point de repère cher à Hull, a publié 79 articles en 2012. C'est près de quatre fois plus que dans notre modèle, où notre *Journal* fictif en publie 20 par année. La durée d'exécution de nos simulations varie de façon linéaire en fonction du nombre d'articles publiés et les expériences que nous nous apprêtons à décrire prennent déjà plus d'une semaine à exécuter. En quadruplant cette durée, on arrive à plus d'un mois, ce qui est incompatible avec les contraintes de temps imposées au projet. Notons toutefois que ces contraintes sont contingentes et ne constituent pas un obstacle fondamental pour qui souhaiterait effectuer des simulations à plus grande échelle : avec suffisamment de temps et de ressources computationnelles, c'est une entreprise réalisable.

Nos résultats seraient-ils qualitativement différents avec des milliers d'agents et un journal qui publie des centaines d'articles ? Nos intuitions nous disent que non, mais s'il y a une chose dont les modèles multi-agents nous ont appris à nous méfier, c'est bien de nos intuitions. Le fait est que nous ne pouvons pas le savoir à moins de l'essayer, ce qui est impossible dans le cadre de ce projet. Et cette remarque ne vaut pas

seulement pour les paramètres dont la valeur est liée à un coût computationnel significatif. Il ne serait pas beaucoup plus coûteux d'affecter à chaque article dix évaluateurs plutôt que trois. Mais pour essayer toutes les valeurs possibles pour le paramètre *nb-reviewers-per-paper* dans l'intervalle [1; 10], il faut faire dix fois plus de simulations. Et ainsi de suite pour chaque dimension de l'espace des paramètres que l'on souhaite explorer. C'est pourquoi il nous a fallu limiter le nombre de questions posées et concevoir nos expériences en fonction des hypothèses énoncées à la section précédente. Les autres questions devront être rangées dans la catégorie des recherches futures.

Pour les recherches présentes, nous avons exécuté les expériences suivantes :

$E_1$  : Dans une première expérience, les valeurs des caractéristiques  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  varient d'une scientifique à l'autre et sont tirées d'une distribution aléatoire uniforme parmi les valeurs  $\{0,0; 0,1; 0,2; \dots; 1,0\}$ . Le seul paramètre qui varie d'une simulation à l'autre, c'est le paysage épistémique utilisé : nous avons exécuté 1 100 simulations pour chacun des six paysages épistémiques pour un total de 6 600 simulations. Chaque simulation dure cent « ans » (c'est-à-dire cent itérations de l'algorithme). Cette expérience nous a permis d'examiner le lien entre les valeurs de  $\alpha$  et  $\beta$  et le crédit accumulé par les scientifiques individuels et de tester ainsi les hypothèses  $H_1$  (concernant le compromis entre crédit et support) et  $H_2$  (concernant la valeur sélective conceptuelle inclusive). Elle nous a aussi permis de faire le même exercice avec les valeurs de  $\gamma$  et  $\delta$ , bien que nous n'ayons pas formulé au préalable d'hypothèses à leur égard. C'est aussi à partir des résultats de cette expérience que nous examinerons les caractéristiques globales du modèle, de même que la performance épistémique générale de la communauté sur chacun des paysages.

$E_2$  : Dans une seconde expérience, nous fixons tour à tour la valeur des caractéristiques  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  pour toute la communauté. Ainsi,  $\alpha$  sera d'abord fixée à 0,0 pour tous les scientifiques et  $\beta$ ,  $\gamma$  et  $\delta$  choisies aléatoirement pour chaque scientifique. Une fois qu'on a passé toutes les valeurs  $\alpha \in \{0,0; 0,1; 0,2; \dots; 1,0\}$ , c'est au tour de  $\beta$  d'être fixe alors que  $\alpha$ ,  $\gamma$  et  $\delta$  sont aléatoires, et ainsi de suite. On a donc toutes les combinaisons (*caractéristique, valeur, paysage*)  $\in \{\alpha; \beta; \gamma; \delta\} \times \{0,0; 0,1; 0,2; \dots; 1,0\} \times \{\text{DE JONG F1; EXP; ROOTS; HORNS 5 PEAKS;}$

LANGERMANN F11; FLAT}), pour un total de 264 combinaisons. Exécutées 100 fois chacune, cela fait un total de 26 400 simulations. Cette expérience nous sert à mettre à l'épreuve l'hypothèse  $H_3$  : elle nous permet de voir quelles sont les valeurs de  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  qui, lorsqu'elles sont adoptées par toute la communauté, produisent la valeur objective moyenne la plus élevée pour les articles publiés. Si ces valeurs étaient les mêmes que celles qui sont les plus avantageuses pour les scientifiques individuels,  $H_3$  serait vérifiée.

### Performance épistémique générale de la communauté

La première question qu'on pourrait se poser à propos du modèle est : fonctionne-t-il ? En d'autres termes : réussit-il à tout le moins à rendre compte d'une certaine forme de progrès scientifique ? Ou, encore plus concrètement : les agents du modèle réussissent-ils à identifier les sommets du paysage épistémique ? La réponse générale à cette question, c'est que dans tous les paysages sauf FLAT (où c'est impossible), la communauté scientifique arrive à progresser. Mais l'ampleur de ce progrès est fortement déterminée par la structure du paysage. La figure 6.1 à la page suivante montre l'évolution de la valeur objective moyenne des articles publiés au fil des années dans chaque paysage. Puisque nous avons tiré ces données de l'expérience  $E_1$ , il s'agit, pour chaque paysage, de la moyenne<sup>77</sup> de 1 100 simulations. La figure 6.2 à la page 179 montre des exemples de distributions finales des articles dans l'espace épistémique pour chaque paysage, pour des simulations choisies au hasard.

Sans surprise, c'est dans le paysage DE JONG F1 que la communauté performe le mieux. La pente régulière et graduelle permet à la communauté de grimper rapidement vers le sommet du paysage et de s'installer à proximité. Comme on peut le voir à la figure 6.2a, la majeure partie des articles ne se retrouve pas au sommet du paysage. Les premiers articles publiés dans la zone de valeur objective plus élevée bénéficient d'un effet d'attachement préférentiel : ils sont lus et cités tôt dans le processus, ce qui permet à leurs

---

<sup>77</sup>En fait, pour chaque année de chaque simulation, on prend la moyenne de la valeur objective des articles publiés, puis on fait (encore pour chaque année) la moyenne de ces moyennes. Les autres résultats que nous présenterons appliquent aussi cette méthode.

## Évolution de la valeur objective moyenne des articles

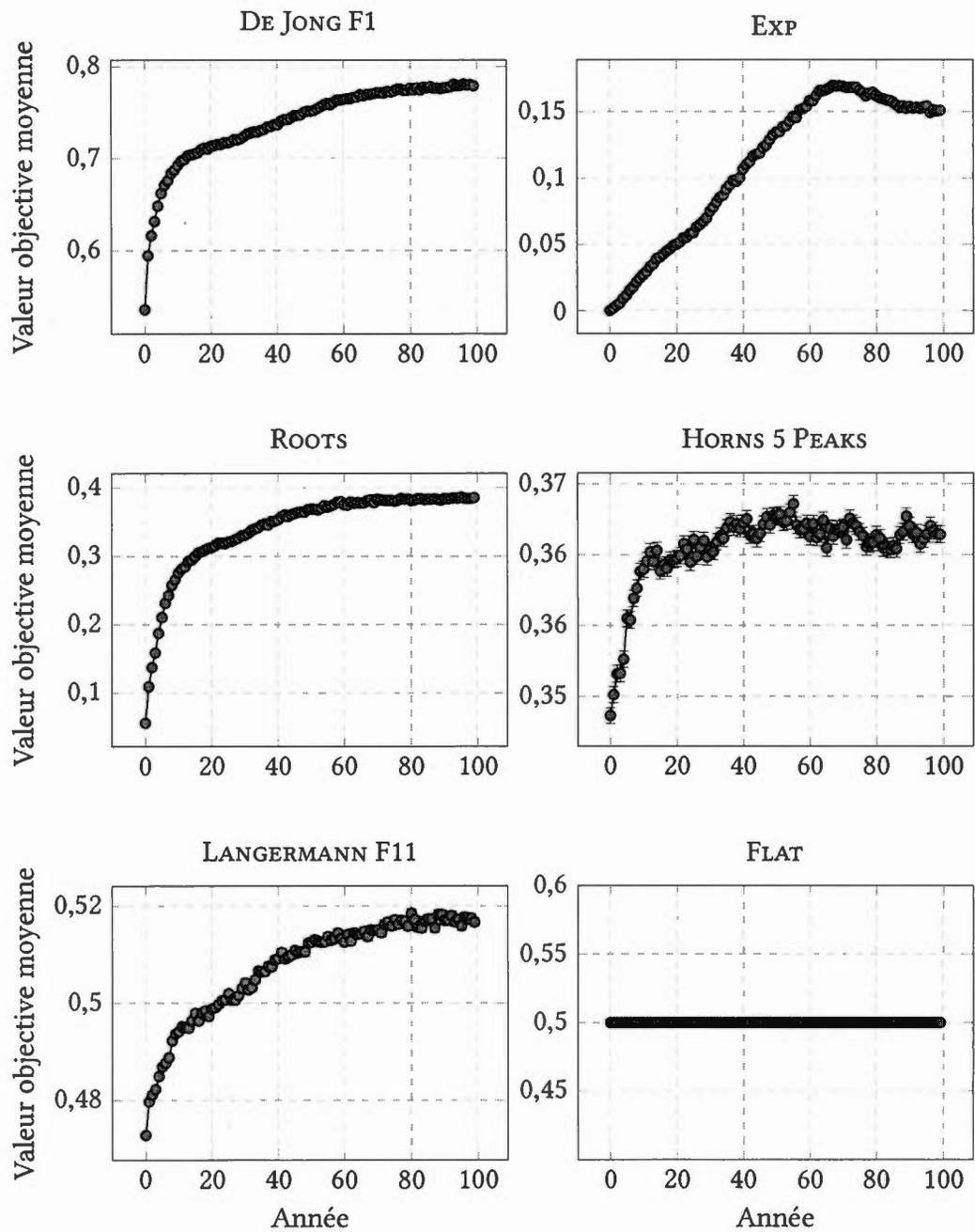
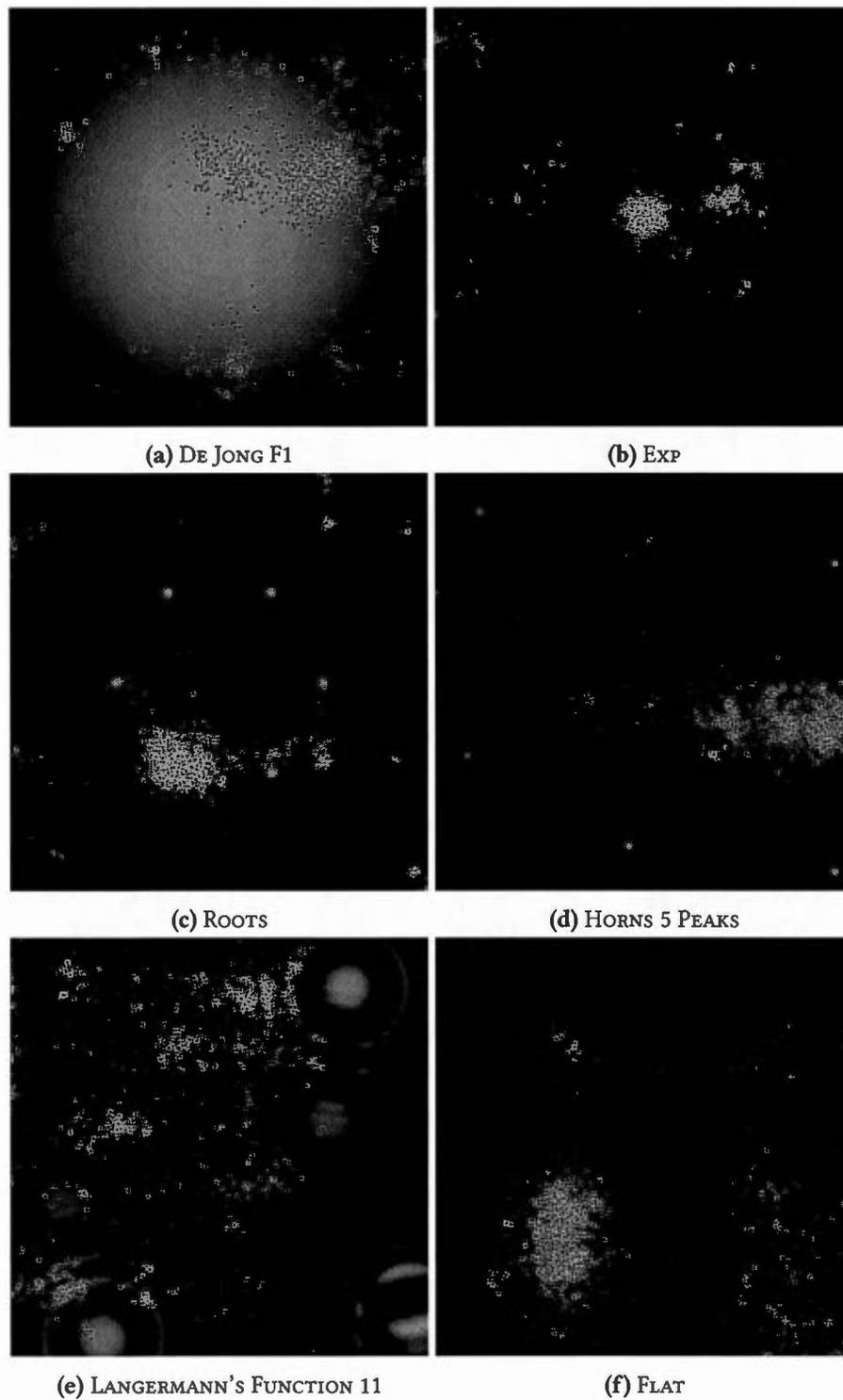


Figure 6.1



**Figure 6.2** Distribution spatiale finale des articles sur les différents paysages épistémiques pour des simulations individuelles.

auteurs d'accumuler un crédit qui fait que leurs articles continuent à être lus et cités, et donc à générer de nouvelles idées. Cela génère des « grappes » d'articles, semblable à ce qu'on a pu observer avec les kènes dans la simulation de Gilbert (figure 2.1 à la page 33). Nous verrons plus loin (figure 6.4 à la page 187) que la distribution des citations par articles résultant de ce processus est fortement asymétrique. Mais malgré ce biais favorable envers les articles séminaux et leurs auteurs, l'information sur la valeur objective des idées, acquise par l'intermédiaire des tests empiriques, se diffuse lentement dans le système et la communauté tend à se stabiliser dans la zone de valeur 0,8.

On observe un comportement différent sur le paysage EXP. La valeur objective étant nulle sur la majeure partie de sa superficie, le phénomène d'attachement préférentiel domine en début de simulation. Ce sont les facteurs externes (c'est-à-dire le crédit et les liens de parenté entre les scientifiques) qui ont le dessus pendant cette période. Mais dès que quelques articles sont publiés dans la zone de valeur objective élevée, la différence de valeur entre les résultats empiriques obtenus pour ceux-ci et pour les autres est suffisamment grande pour leur permettre d'exercer rapidement une influence. Les facteurs internes (c'est-à-dire la valeur objective des idées telle qu'estimée par les tests) retrouvent une certaine place pendant cette période, mais des articles continuent à être publiés à l'écart du sommet, ce qui explique que la valeur moyenne maximum atteinte soit autour de 0,15. La croissance linéaire de la valeur objective moyenne pour EXP à la figure 6.1 reflète le fait que la découverte du sommet se produit à des moments différents d'une simulation à l'autre. Vers la soixante-dixième année, par contre, on observe un phénomène unique à ce paysage : la valeur objective moyenne des articles publiés se met à baisser. Ce qui se passe, c'est que la communauté scientifique se retrouve à l'étroit sur ce sommet trop petit : il devient difficile d'y publier des articles originaux, et les facteurs externes sont à nouveau seuls à dominer. Dans la réalité, un tel phénomène serait le signe qu'il est temps pour la communauté scientifique de s'atteler à résoudre des problèmes différents.

Le paysage ROOTS donne lieu à une dynamique similaire à celle qu'on observe dans EXP, mais du fait qu'il y a six sommets plutôt qu'un seul, et un plateau de valeur objective 0,5 au centre de ceux-ci, la communauté tend à trouver plus rapidement un

sommet et à maintenir une valeur objective plus élevée par la suite, grâce aux articles publiés sur le plateau plutôt que dans la zone de valeur zéro.

Le paysage HORNS 5 PEAKS présente une difficulté supplémentaire pour la communauté : ses sommets sont isolés, séparés du reste du paysage par des vallées de valeur épistémique nulle. Et cela s'avère une réelle difficulté pour les agents du modèle : dans la plupart des simulations, ils ne réussissent pas à trouver les sommets. La communauté reste coincée sur les « crêtes » de valeur objective moyenne qui serpentent entre ceux-ci. Et même lorsque quelques articles sont publiés sur les sommets, la communauté ne tend pas à s'installer autour de ceux-ci, puisque les idées qui sont générées autour d'eux ont tendance à tomber dans la zone de valeur épistémique nulle et sont rejetées suite aux tests empiriques. Et puisque, au départ d'une simulation, les idées initiales sont générées sur cette crête (voir figure 5.3d à la page 141), on n'observe pas une grande variation dans la valeur objective moyenne au fil de la simulation pour HORNS 5 PEAKS à la figure 6.1.

La communauté fait face à des difficultés similaires sur LANGERMANN F11. Même si les vallées qui séparent les sommets sont moins abruptes, la communauté tend à rester sur le plateau de valeur objective moyenne où elle est placée au départ de la simulation. Les rares simulations où elle s'installe autour d'un sommet (quand même plus vastes que dans HORNS 5 PEAKS) ne sont pas suffisantes pour faire une grande différence sur la moyenne des simulations.

Finalement, la figure 6.1 ne nous apprend pas grand-chose sur le paysage FLAT puisque la valeur objective moyenne des articles est nécessairement la même tout au long de la simulation. Les articles tendent à se concentrer rapidement autour d'une ou deux grappes très fortes, et les articles qui s'éloignent de celles-ci n'ont que très peu de chance d'avoir un impact. C'est une situation similaire à celle qu'on observe au début d'une simulation dans EXP, sauf qu'ici, il n'y a pas de porte de sortie : les seuls facteurs qui peuvent différencier un article d'un autre sont les facteurs externes ; la valeur objective est la même pour tous, et les tests empiriques ne font que confirmer cet état de choses.

Que doit-on retenir de tout ça ? Que la performance épistémique générale de la communauté est surtout déterminée par la structure du paysage épistémique qu'elle cherche à explorer. Cela étant dit, dans la mesure où chaque paysage le permet, l'algorithme fonctionne. Dans les bonnes circonstances, la communauté arrive à trouver les sommets, même si elle a tendance à rester coincée sur des sommets locaux (ce qui est aussi le cas dans la science réelle). C'est une bonne nouvelle pour notre modèle, mais c'est aussi une bonne nouvelle pour la théorie de Hull : le processus évolutionnaire postulé pour la répliation des idées scientifiques est suffisant pour rendre compte du progrès scientifique, et ce même lorsqu'on fait intervenir des facteurs externes comme le crédit ou les liens entre scientifiques. Cela ne constitue pas une validation absolue de sa théorie, mais cela montre à tout le moins qu'il est plausible qu'un tel mécanisme soit à l'œuvre en science. Il nous faut toutefois ajouter ici un bémol : le comportement de notre modèle est un comportement de science « normale », au sens de Kuhn (1962). Rien, dans ce que nous avons observé de nos simulations, ne permet de rendre compte de ce qui se produit lors des révolutions scientifiques. Cela ne veut pas dire qu'il est impossible de construire un modèle évolutionnaire de la science qui permette de générer des révolutions scientifiques, mais ce n'était pas spécifiquement notre objectif : il faut simplement garder en tête que les conclusions que nous nous apprêtons à tirer du modèle ne s'appliquent qu'à la science normale.

Notons aussi, finalement, que l'influence déterminante du paysage épistémique ne nous empêche pas d'explorer l'effet d'autres facteurs sur la performance de la communauté, puisque nous étudierons ces effets à l'intérieur de chaque paysage. Le fait que la communauté performe moins bien sur EXP que sur DE JONG F1 ne nous empêche pas de voir comment la performance sur EXP est influencée par  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$ . Et il se peut très bien que les facteurs qui ont une forte influence sur la performance dans un paysage n'aient qu'une faible influence dans un autre. Nous nous pencherons sur cette question à la section 6.5.

## Distributions

Notre modèle permet de produire, par simulation, des données bibliométriques qui, bien que fictives, sont de même nature que celles que l'on retrouve dans la réalité. En théorie, toutes les méthodes d'analyse bibliométrique qui font intervenir des motifs de citations et des réseaux de coauteurs seraient applicables aux résultats de nos simulations.

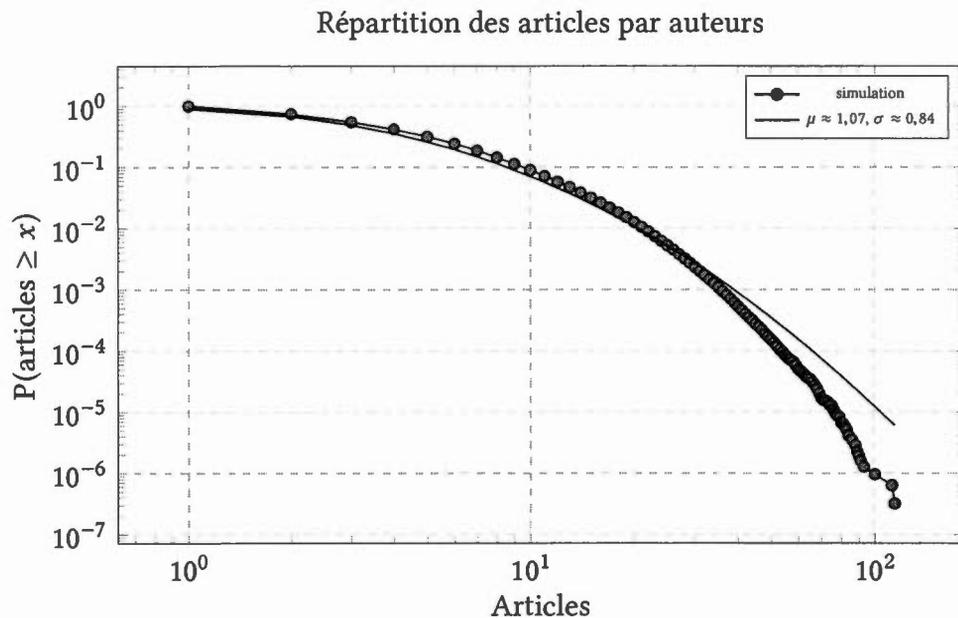
Pousser à fond l'analyse bibliométrique s'écarterait de notre objectif principal, qui est d'explorer les aspects du modèle qui ont un impact normatif, mais celle-ci aura néanmoins une valeur heuristique. Il est connu depuis longtemps que la distribution du nombre d'articles par scientifiques (Lotka, 1926) et la distribution du nombre de citations par articles (de Solla Price, 1965) suivent des lois de puissance. Si notre modèle générerait des distributions correspondant à ces lois, cela montrerait, sans en constituer une validation définitive, qu'il représente un élément d'un mécanisme plausible pour la production d'articles scientifiques. Le modèle ne génère pas *tout à fait* des lois de puissance : il génère des distributions qui, comme les lois de puissances, sont fortement asymétriques, mais qui, statistiquement, n'en sont pas : lorsqu'on trace la réciproque de la fonction de répartition<sup>78</sup> dans un graphique aux axes logarithmiques, on n'obtient pas une ligne entièrement droite. Les figures 6.3 à 6.6 montrent quatre de ces distributions : le nombre d'articles par auteurs, le nombre de citations par articles, le nombre citations par auteurs et, finalement, la quantité de crédit par scientifique.

Dans le reste de cette section, nous allons prendre une à une ces distributions et dire quelques mots sur chacune d'elles, tentant d'identifier, lorsque c'est possible, les éléments du modèle qui sont responsables de leurs formes particulières.

La figure 6.3 à la page suivante, montre la distribution du nombre d'articles par auteur. C'est l'équivalent, pour notre modèle, des distributions générées par les modèles de Gilbert (1997) et Sun et Naveh (2009) (voir figure 2.2 à la page 34). Le modèle de

---

<sup>78</sup> La « complementary cumulative distribution function » (CCDF), en anglais.



**Figure 6.3**

Gilbert et celui de Sun et Naveh génèrent des lois de puissance. Dans notre modèle, le nombre d’auteurs ayant publié un nombre élevé d’articles décline plus rapidement que s’il s’agissait d’une véritable loi de puissance. Nous avons trouvé que la distribution générée s’approche plus d’une loi log-normale (tracée en rouge à la figure 6.3). Comment cela peut-il s’expliquer ?

La première chose à mentionner, c’est la proximité générale entre les lois de puissances et les lois log-normales. Dans les distributions empiriques, il peut être difficile de les distinguer statistiquement l’une de l’autre. En se penchant sur l’historique des modèles générant ces deux types de lois, Mitzenmacher (2003, p. 227) constate que :

« power law and log-normal distributions are intrinsically connected. Very similar basic generative models can lead to either power law or lognormal distributions, depending on seemingly trivial variations. »

Pour générer une loi de puissance, la façon la plus directe est d’utiliser un modèle d’attachement préférentiel. C’est ce que fait Gilbert. Dans son modèle, lorsqu’un article est produit par un auteur existant, la probabilité qu’il soit écrit par un auteur

plutôt qu'un autre est directement proportionnelle au nombre d'articles déjà publiés par l'auteur. Un tel mécanisme ne peut faire autrement que de générer une loi de puissance.

Dans le modèle de Sun et Naveh, ce qui détermine le succès des auteurs dans le processus de publication, c'est leur capacité à apprendre la « fonction d'évaluation communautaire » des articles (qui joue, dans leur modèle, le même rôle que nos paysages épistémiques). Pour différencier les agents entre eux, ils font varier différents paramètres de l'architecture cognitive CLARION (par exemple, le taux d'apprentissage). Ce sont des facteurs qui ont une influence directe sur la capacité des agents à produire des articles qui seront publiés, et les agents qui démarrent la simulation avec les bonnes combinaisons de paramètres bénéficient d'un avantage qui leur permet de publier beaucoup plus que les autres.

Dans notre modèle, afin d'accentuer le rôle des facteurs sociaux, nous avons fait le choix de ne pas différencier les scientifiques par rapport à leur productivité : chaque année, chacun a le même nombre d'idées, lit le même nombre d'articles, effectue le même nombre de tests et soumet le même nombre de manuscrits. Et à part peut-être pour la créativité ( $\delta$ ), les caractéristiques où ils diffèrent entre eux ( $\alpha$ ,  $\beta$  et  $\gamma$ ) n'ont pas une influence directe sur la qualité des idées produites. En ce sens, il est remarquable que notre modèle produise une distribution à ce point asymétrique.

Nous croyons qu'il y a deux facteurs principaux qui expliquent cette asymétrie. Le premier, c'est la longévité des scientifiques qui ont accumulé beaucoup de crédit. Les scientifiques qui réussissent à publier tôt dans leur carrière accumulent un crédit qui diminue leurs chances de quitter le système scientifique. Cela leur permet de publier d'autres articles, ce qui leur permet d'obtenir plus de crédit, ce qui diminue d'autant leurs chances de retraite, et ainsi de suite. L'autre facteur est lié à l'évaluation des articles avant publication. Les évaluateurs choisis par le journal ne savent pas qui est l'auteur de l'article, mais ils basent quand même partiellement leur évaluation sur le crédit des auteurs cités dans l'article. Et à part quand  $\alpha = 1$  et que l'auteur mise tout sur le support (ce qui, nous le verrons plus loin, n'est pas la meilleure stratégie), les

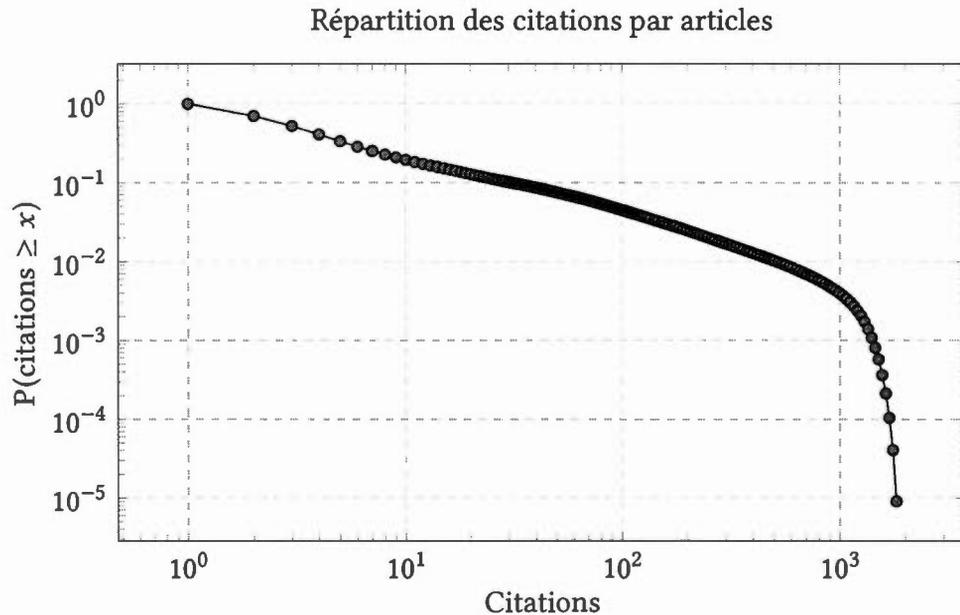
citations incluent des autocitations, et les auteurs qui ont déjà beaucoup de crédit<sup>79</sup> sont avantagés.

Dans le monde réel, il y a aussi d'autres facteurs, qui ne sont pas pris en compte par notre modèle. Il y a, d'abord, les articles coécrits : les scientifiques qui ont beaucoup de crédit bénéficient de nombreuses occasions de collaborer avec d'autres et de coécrire des articles, ce qui augmente d'autant leur nombre d'articles publiés, et donc leur crédit, et ainsi de suite. Ces coauteurs sont parfois les propres étudiants du scientifique, ce qui nous amène à un autre facteur de productivité : les scientifiques les mieux établis ont souvent une petite armée d'étudiants et d'assistants de recherche qui travaillent pour eux (et les fonds pour les payer). Finalement, on peut supposer que (au moins chez certains auteurs) la productivité augmente avec l'expérience de publication : que le dixième article est moins difficile à écrire que le premier, et que le centième est moins difficile que le dixième. Si on intégrait tous ces facteurs dans le modèle (peut-être en plus de facteurs cognitifs à la Sun et Naveh), on aurait probablement plus de scientifiques qui produisent un nombre d'articles plus élevé et, conséquemment, une distribution qui s'approche un peu plus d'une loi de puissance.

La figure 6.4 à la page suivante montre la distribution du nombre de citations par articles générée par le modèle. Ce qu'on observe ici, c'est une distribution qui ressemble à une loi de puissance jusqu'à environ 1 000 citations, puis tombe rapidement par la suite. Cette chute finale dans la probabilité de trouver des articles avec un grand nombre de citations est cohérente avec ce qu'on observe dans les distributions empiriques de citations par articles (voir Wallace *et al.*, 2009, fig. 5). Elle s'explique par le fait que nous nous trouvons dans un système fini : il y a une limite au nombre de citations qu'un article peut accumuler. Dans la simulation, un maximum de vingt articles est publié chaque année. Un article qui serait publié à l'an un de la simulation et qui serait ensuite cité par *tous* les articles subséquents pourrait donc accumuler un maximum de 1980 citations dans les quatre-vingt-dix-neuf années qui suivent. Nos ré-

---

<sup>79</sup>En fait, les auteurs qui ont déjà beaucoup de crédit *auprès de leurs évaluateurs* sont avantagés, mais les auteurs qui ont déjà beaucoup de crédit en général ont plus de chances d'avoir beaucoup de crédit auprès de leurs évaluateurs. C'est d'autant plus vrai que les évaluateurs sont eux-mêmes choisis en fonction de leur crédit et sont donc des scientifiques avec une certaine ancienneté qui ont fort probablement eu l'occasion de lire les autres articles de l'auteur.



**Figure 6.4**

sultats s'approchent de cela : sur toutes les simulations de  $E_1$ , le maximum de citations obtenues par un article est 1891.

Étant données la distribution du nombre d'articles par auteurs et la distribution du nombre de citations par articles que nous venons de décrire, nous ne serons pas surpris de trouver une distribution similaire pour le nombre de citations par auteurs (figure 6.5 à la page suivante). L'auteur le plus cité, sur toutes les simulations de  $E_1$ , obtient 27141 citations.

Finalement, la figure 6.6 à la page suivante montre la distribution du crédit par scientifiques. La forme de celle-ci découle partiellement de la distribution des citations par scientifiques, puisque c'est, entre autres, par l'intermédiaire des citations qu'on acquiert du crédit dans le modèle. Toutes les citations ne sont pas également profitables, par contre : on peut très bien être cité dans un article qui ne sera jamais lu par qui que ce soit. À l'autre extrémité du spectre, il peut suffire d'une seule citation dans un article très populaire pour accumuler une grande quantité de crédit. L'autre façon d'acquérir du crédit c'est, bien sûr, lorsque nos propres articles sont lus. Et les

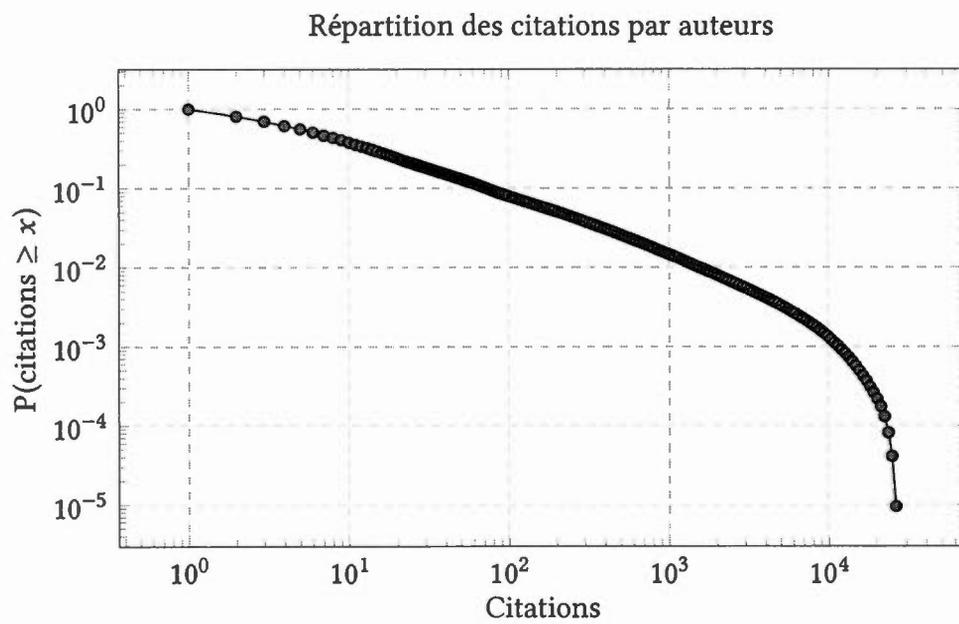


Figure 6.5

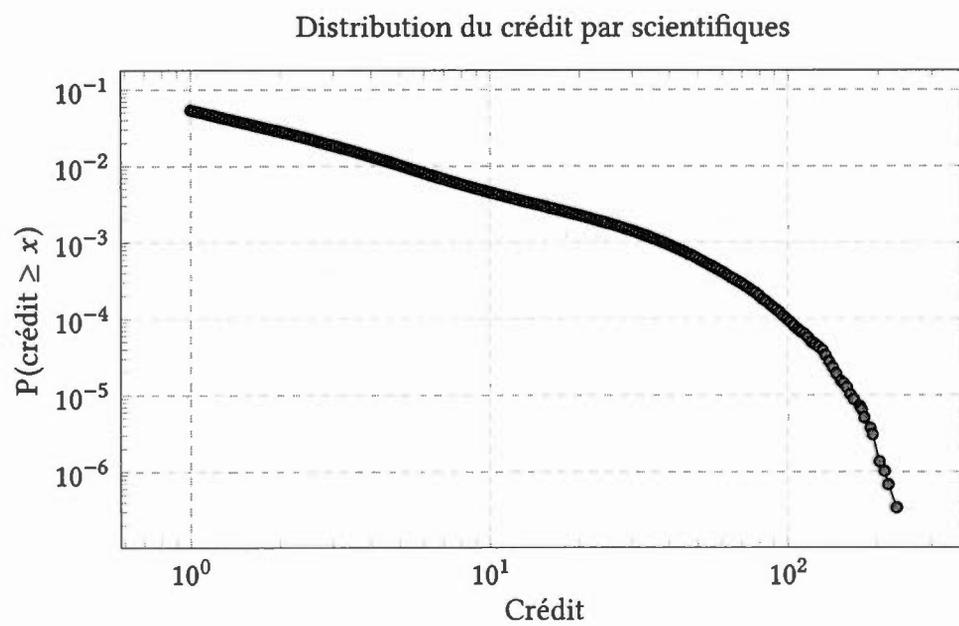


Figure 6.6

remarques précédentes s'appliquent à cette situation aussi : un article peut très bien ne jamais être lu, alors qu'un autre sera lu par toute la communauté.

Notons aussi qu'il y a plusieurs boucles de rétroaction qui sont à l'œuvre en ce qui concerne le crédit, ce qui permet à la queue de la distribution de tomber un peu moins rapidement que pour le nombre d'articles et le nombre de citations. La première et la plus évidente de ces boucles de rétroaction, c'est simplement le fait que d'avoir plus de crédit augmente les chances que les articles d'une scientifique soient lus et qu'il obtienne d'autre crédit lors de cette lecture. La seconde boucle de rétroaction, c'est le fait que le crédit lui permet d'attirer des étudiants à qui elle fera lire ses articles (ce qui lui donne du crédit) et qui la citeront éventuellement (encore plus de crédit). La troisième boucle de rétroaction est un peu plus subtile : lorsque  $\gamma > 0$ , une partie de l'évaluation qui est faite lors de la lecture d'un article est fonction du crédit attribué : plus le lecteur accorde de crédit préalable à l'auteur, plus son idée sera évaluée positivement. Et puisque le crédit attribué à l'auteur suite à la lecture d'un article est fonction de l'évaluation qui est faite de l'idée proposée dans celui-ci, ce crédit sera d'autant plus élevé. Finalement, la dernière boucle de rétroaction est une version indirecte de la précédente : les idées sont jugées en partie sur le crédit, mais elles sont aussi jugées (quand  $\gamma < 1$ ) sur la base de leur similarité avec des idées existantes évaluées plus ou moins positivement. Si le crédit a aidé à faire accepter les idées passées d'un auteur, il aide aussi à faire accepter ses idées futures lorsque celles-ci sont similaires aux idées passées (ce qui risque fort d'être le cas puisque les idées passées servent de parents aux idées futures).

Dans cette section, nous avons examiné la façon dont les auteurs, leurs articles et les citations entre les articles sont répartis dans la simulation. Nous avons constaté que ces distributions, comme dans le système scientifique du monde réel, sont fortement asymétriques : peu de scientifiques sont productifs, beaucoup le sont peu ; peu d'articles sont cités, beaucoup le sont peu ; et par conséquent, peu d'auteurs sont cités, beaucoup le sont peu. Sans surprise, on a trouvé le même genre d'asymétrie pour le crédit. Nous n'avons pas de données empiriques en ce qui concerne le crédit dans le monde réel, mais nous pouvons supposer que cette asymétrie y existe aussi.

Pour conclure cette section, nous ferons simplement remarquer que toutes les distributions présentées ici émergent des mécanismes mis en place pour la simulation : elles sont le résultat des interactions entre les agents et non pas d'un modèle mathématique conçu en fonction de produire ces distributions particulières.

### Performance sociale en fonction des caractéristiques individuelles

Le temps est venu d'aborder une des questions qui nous préoccupent plus particulièrement : l'influence des caractéristiques individuelles des scientifiques sur leur « performance sociale » : leur capacité à accumuler du crédit. Nous accorderons un peu plus d'attention à  $\alpha$  et  $\beta$ , puisque c'est à leur propos que nous avons formulé, respectivement, les hypothèses  $H_1$  et  $H_2$ , mais nous examinerons aussi l'effet de  $\gamma$  et  $\delta$ .

Nous utiliserons pour ce faire les résultats de l'expérience  $E_1$  où, rappelons-le, les valeurs de  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  sont fixées de façon aléatoire pour chaque scientifique dans la simulation. Nous calculons, pour chaque valeur possible de chaque caractéristique, la quantité moyenne de crédit accumulée par les scientifiques qui possèdent cette valeur pour cette caractéristique. La figure 6.7 à la page 192 montre ces résultats pour  $\alpha$  dans chaque paysage épistémique<sup>80</sup>. La figure 6.10 à la page 196, de même que les figures 6.12 et 6.13, en annexe, font de même pour  $\beta$ ,  $\gamma$  et  $\delta$ , respectivement.

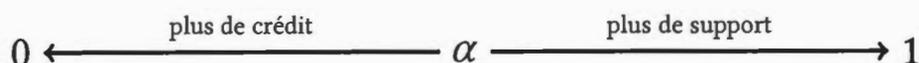
### Influence de $\alpha$ sur le crédit des scientifiques

Un des éléments de la théorie de Hull (que nous avons traduit par  $H_1$ ), c'est qu'un scientifique, s'il veut que ses idées aient une chance d'être acceptées par la communauté, n'a d'autre choix que de citer les travaux de ceux qui l'ont précédé, leur conférant par le fait même une partie du crédit qui sera attribué lors de la lecture de l'article. Si ce n'était de ce besoin de support, il lui serait plus avantageux de consacrer toutes ses citations à recherche de crédit, pour lui-même ou pour ses étudiants (ce dernier choix

---

<sup>80</sup>Dans ces graphiques, comme dans les semblables qui suivront, les barres d'erreur représentent l'erreur type de la moyenne.

étant déterminé par  $\beta$ ). La caractéristique  $\alpha$ , dont nous avons examiné le rôle en détail à la section 5.4.3, représente l'attitude d'un scientifique par rapport à ce compromis. Plus  $\alpha$  est élevé, plus le scientifique aura tendance à citer les articles qui peuvent lui apporter du support, au détriment de ses propres articles et de ceux de ses étudiants (et donc de son crédit). En résumé :



Qu'en est-il dans la simulation ? Au vu de la figure 6.7, il apparaît clairement qu'une certaine dose de support est nécessaire : les scientifiques pour qui  $\alpha = 0$  (et qui n'accordent donc aucune importance au support) sont ceux qui, de loin, accumulent le moins de crédit. Les valeurs de  $\alpha$  les plus payantes se situent, selon les paysages, entre 0,1 et 0,4, puis tendent à être de moins en moins avantageuses jusqu'à l'autre extrême, 1,0, où la scientifique ne considère *que* le support que peut lui apporter un article sans se privilégier elle-même ou ses étudiants.

On a donc trouvé, à l'œil, quelque chose comme la relation de compromis recherchée : il faut un peu de support, mais pas trop. Nous aimerions maintenant nous assurer de façon plus rigoureuse que ce résultat est significatif. Si c'était une relation linéaire que nous avons constatée, il suffirait de calculer le coefficient de corrélation entre les deux variables. Mais puisque nous recherchons un autre type de relation, il nous faut franchir quelques étapes supplémentaires. La première, c'est de définir ce que nous entendons exactement par une relation « de compromis ». Dans la formulation de  $H_2$ , « les valeurs les plus avantageuses de  $\alpha$  pour une scientifique individuelle devraient se situer autour de 0,5 », mais ce n'est pas tout à fait suffisant : il nous faut, pour chaque valeur possible de  $\alpha$ , assigner une valeur correspondante représentant son statut en tant que compromis. Selon cette projection, les extrêmes  $\alpha = 0$  et  $\alpha = 1$  auraient une valeur de compromis de 0 (pas du tout un compromis) alors que  $\alpha = 0,5$  aurait une valeur de compromis de 1 (le maximum). On recherche donc quelque chose comme

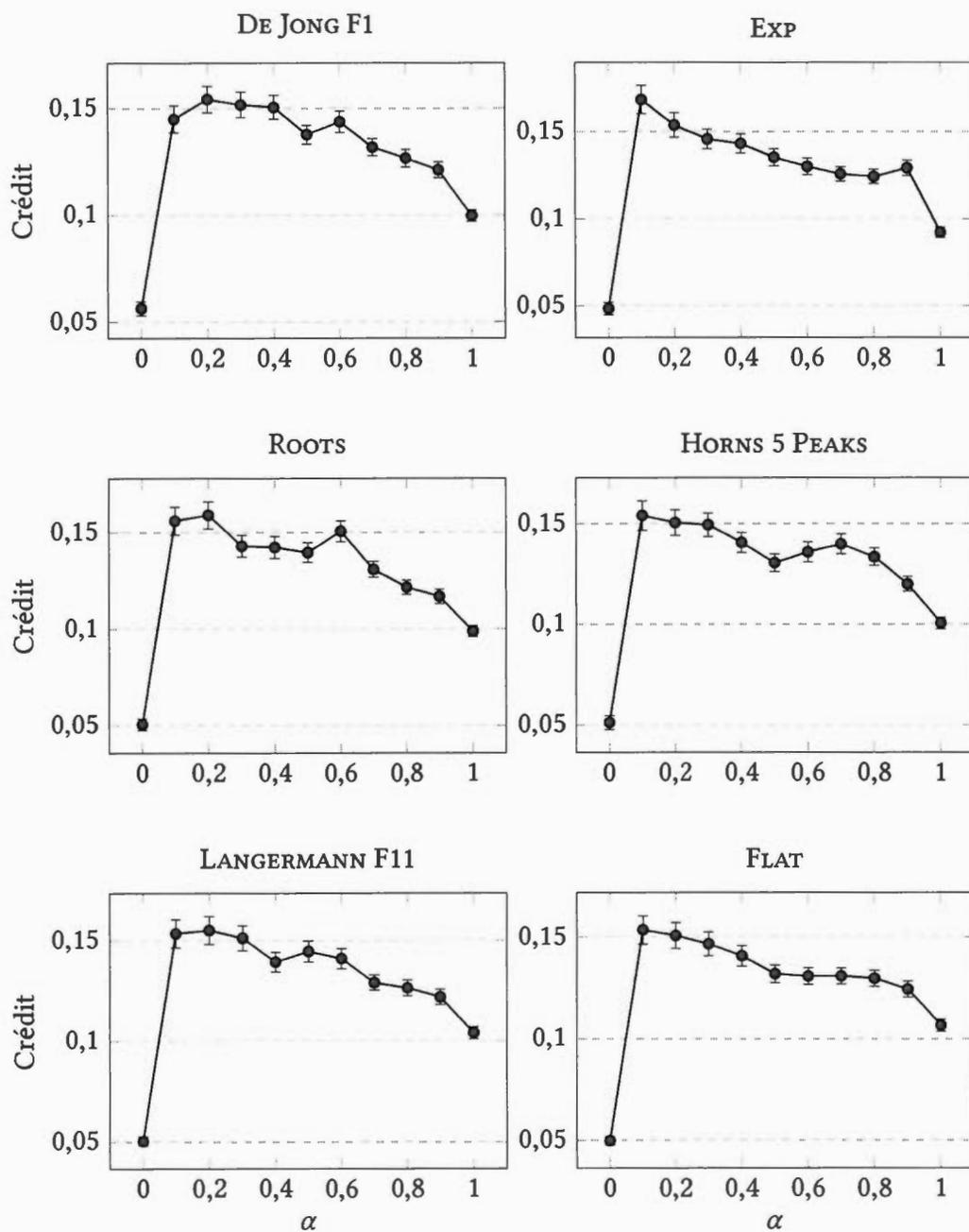
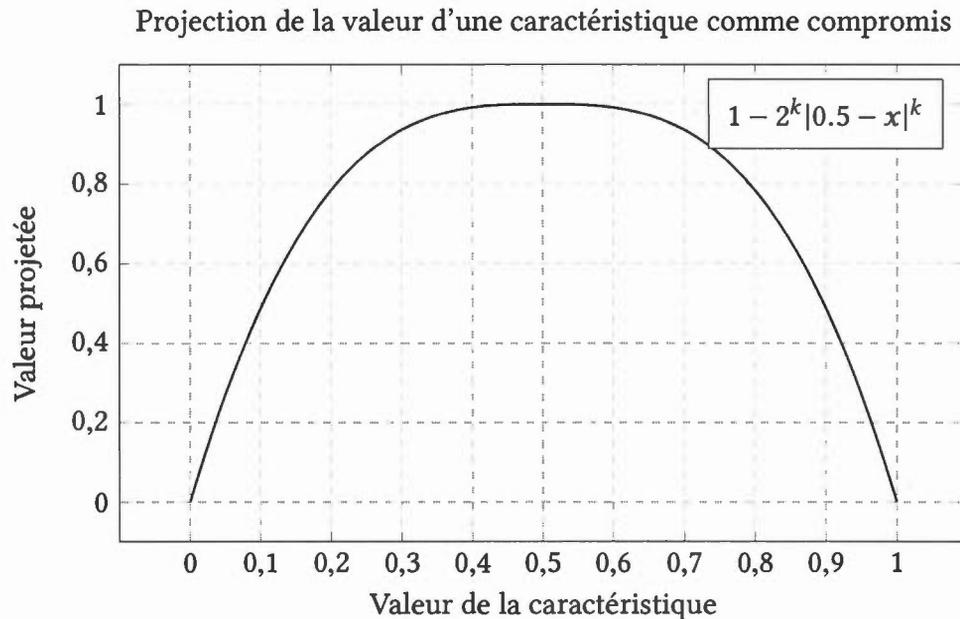
Crédit en fonction de  $\alpha$ 

Figure 6.7



**Figure 6.8**

une « cloche », qu'il est relativement simple de modéliser avec l'équation :

$$1 - 2^k |0.5 - x|^k, \quad (6.1)$$

où  $k$  est une constante qui permet d'ajuster la forme de la courbe pour la rendre plus ou moins à pic. La figure 6.8 représente cette équation de façon visuelle. Nous avons choisi  $k = 3$  pour représenter un compromis incitant surtout à éviter les valeurs extrêmes (0 et 1).

On peut maintenant calculer notre corrélation. Si le crédit obtenu pour chacune des valeurs possibles de la caractéristique  $\alpha$  suivait exactement cette courbe, on aurait une corrélation parfaite. On pourrait alors accepter l'hypothèse  $H_1$ , selon laquelle les scientifiques font face à un compromis entre crédit et support. Les résultats que nous avons vus à la figure 6.7 ne suivent pas tout à fait la courbe de la figure 6.8 : les valeurs  $0,1 \leq \alpha \leq 0,4$  sont celles qui ont tendance à apporter le plus de crédit. Néanmoins, lorsqu'on calcule le coefficient de corrélation de Pearson entre les valeurs de  $\alpha$  projetée à l'aide de l'équation 6.1, on trouve une relation significative ( $p \leq 0,05$ )

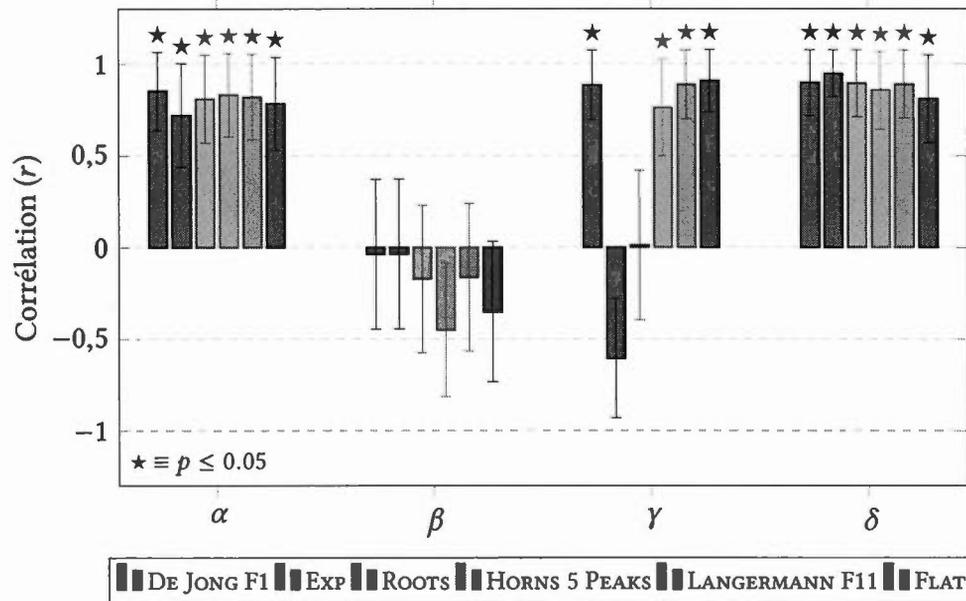
Relation de compromis entre caractéristiques individuelles et crédit ( $E_1$ )

Figure 6.9

dans tous les paysages. La figure 6.9 présente ces résultats sous forme de diagramme à barres verticales<sup>81</sup>. Cela nous semble suffisant pour accepter l'hypothèse  $H_1$ , avec le bémol que les valeurs plus basses de  $\alpha$  sont légèrement favorisées : l'idéal, c'est d'avoir juste assez de support.

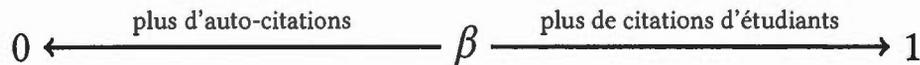
Influence de  $\beta$  sur le crédit des scientifiques

La figure 6.9 ne présente pas que la relation de compromis entre  $\alpha$  et le crédit accumulé : elle fait la même chose pour  $\beta$ ,  $\gamma$  et  $\delta$ . Nous reviendrons plus loin sur  $\gamma$  et  $\delta$ , mais le moment est venu de nous attarder à  $\beta$ , dont il faut d'abord rappeler le rôle. Comme pour  $\alpha$ , c'est surtout au moment de l'écriture des articles<sup>82</sup> que  $\beta$  intervient. La valeur de  $\beta$  détermine la tendance du scientifique à citer ses « descendants intellectuels » : ses étudiants, les étudiants de ceux-ci, et ainsi de suite. Quand  $\beta = 0$ , les étudiants

<sup>81</sup>Voir tableau A.1 à la page 241, en annexe, pour les chiffres exacts.

<sup>82</sup>Mais aussi, dans une moindre mesure, lors du choix des articles à lire (voir section 5.4.5).

n'ont aucun poids : toutes les citations qui ne sont pas consacrées à l'obtention de support sont des autocitations. Quand  $\beta = 1$ , au contraire, les articles des descendants ont autant de chances d'être cités que ceux du scientifique lui-même, peu importe la distance générationnelle. Et pour les valeurs intermédiaires  $0,1 \leq \beta \leq 0,9$  la probabilité d'être cité descend plus ou moins vite avec la distance générationnelle, selon que  $\beta$  est près de zéro ou de un. En bref :



L'hypothèse de Hull quant au fait de citer ses étudiants, c'est que même s'ils ne sont pas en mesure de fournir du support à court terme (parce qu'ils n'ont pas encore accumulé de crédit et ne sont donc pas en mesure d'influencer positivement les lecteurs de l'article dans lequel ils sont cités), il peut être payant à long terme de les citer puisqu'ils ont le potentiel d'être les principaux vecteurs de transmission des idées de leur superviseur. C'est l'hypothèse  $H_2$  : on s'attend à ce qu'il y ait un compromis à faire quant à la valeur de  $\beta$  la plus avantageuse. Concrètement, dans la simulation, l'avantage qu'il y a à citer ses étudiants se traduit de la façon suivante :

- Lors du processus d'éducation (voir section 5.4.7), l'étudiant lit tous les articles de son superviseur ;
- Ce faisant, l'étudiant ajoute ces articles à sa banque d'articles lus et attribue du crédit à son superviseur ;
- Lorsque l'étudiant écrit ses premiers articles, il cite, parmi les articles qu'il a lus, ceux des auteurs à qui il accorde du crédit : en l'occurrence, son superviseur.

Et pourtant, si on retourne à la figure 6.9 à la page 194, on constate qu'il n'y a pas de relation de compromis significative entre  $\beta$  et le crédit accumulé. Au contraire, si on examine la situation dans le détail (figure 6.10 à la page suivante), ce qui semble se dessiner, c'est que les valeurs les plus avantageuses se trouvent, selon les paysages, dans l'intervalle  $0,0 \leq \beta \leq 0,3$  et que, par la suite, plus  $\beta$  est élevé, moins on accumule

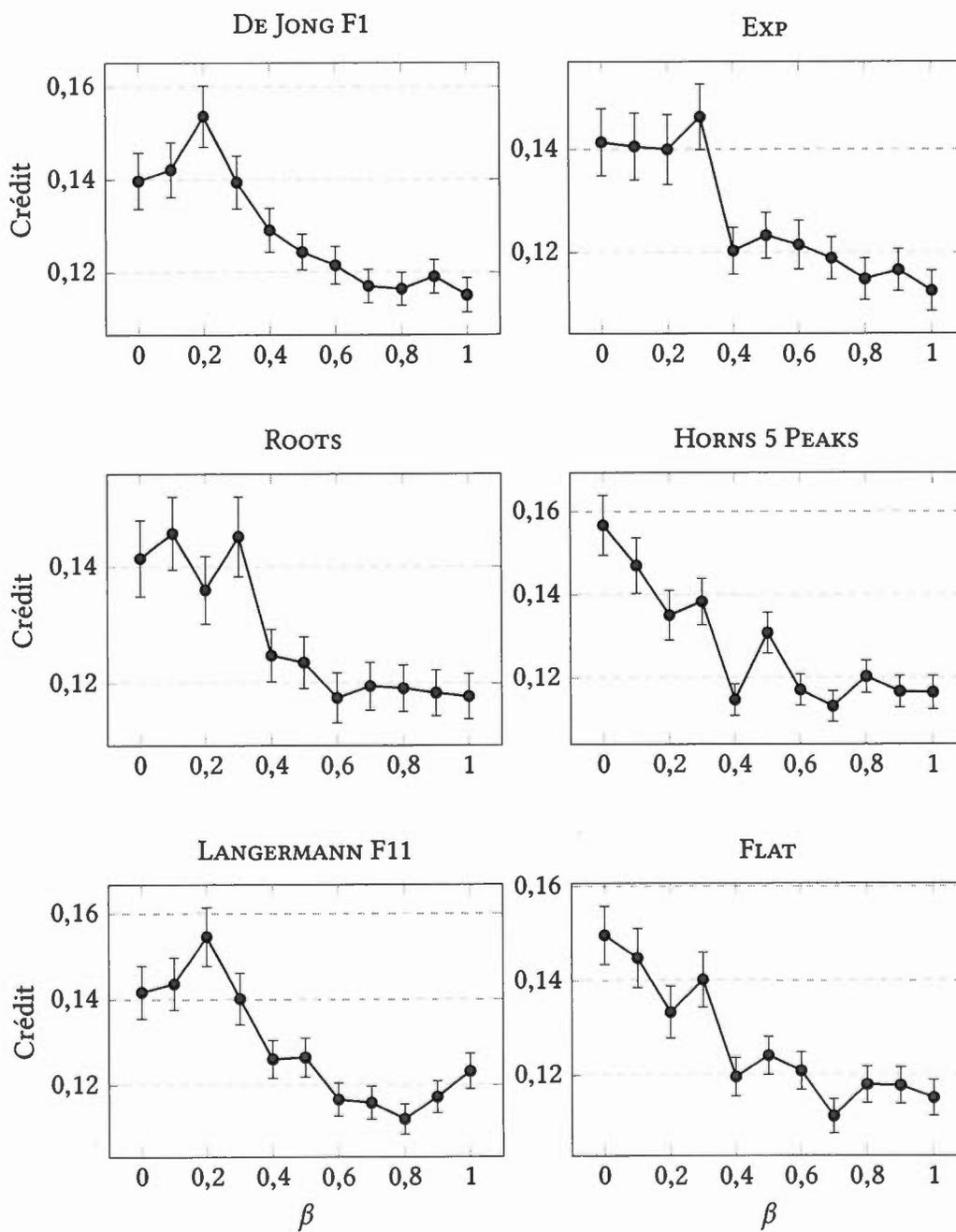
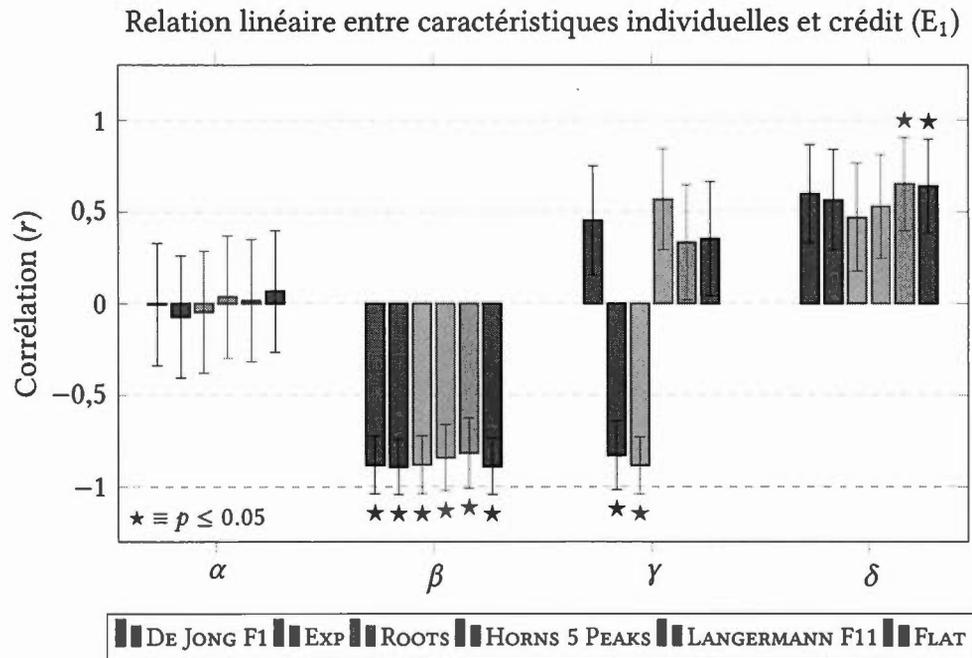
Crédit en fonction de  $\beta$ 

Figure 6.10



**Figure 6.11**

de crédit. Pour confirmer cela statistiquement, il suffit cette fois d'établir une simple corrélation linéaire entre  $\beta$  et le crédit accumulé. C'est ce que montre (entre autres choses<sup>83</sup>) la figure 6.11 : pour tous les paysages, on a une corrélation négative inverse très forte entre  $\beta$  et le crédit accumulé. Dans le contexte de notre modèle, il est désavantageux de citer ses étudiants plutôt que de se citer soi-même. Sur la base de ces résultats, il nous est donc impossible d'accepter  $H_2$ .

La principale leçon à tirer de ce résultat c'est qu'il ne faut pas se fier à nos intuitions lorsqu'on a affaire à un système complexe. Nous avons déterminé, dans les moindres détails, les comportements individuels des agents du système, mais, au bout du compte, le comportement global du modèle n'est pas celui auquel nous nous attendions. Cela montre bien, selon nous, le danger qu'il y a à réfléchir sans le support d'un modèle formel ou de données empiriques.

<sup>83</sup>Nous avons calculé les corrélations linéaires pour toutes les caractéristiques. Sans surprise, la figure 6.11 ne montre pas de relation linéaire pour  $\alpha$ , pour laquelle nous avons déjà trouvé une relation de compromis. Nous reviendrons plus loin à  $\gamma$  et  $\delta$ . Voir tableau A.2 à la page 241, en annexe, pour les chiffres exacts.

Quoi qu'il en soit, la question se pose : qu'est-ce qui explique cette corrélation négative entre le fait de citer ses étudiants et le crédit obtenu ? Nous voyons deux facteurs principaux (ce qui n'exclut pas qu'il en existe d'autres) :

- ▶ Favoriser ses étudiants en citant leurs articles, c'est quelque chose qui ne peut être fait qu'à partir du moment où ils ont déjà publié un article. Et dans la mesure où la publication d'un premier article est le principal obstacle au lancement d'une nouvelle carrière, les citations altruistes du superviseur arrivent trop tard pour faire une différence significative : au moment où il peut citer l'étudiant, celui-ci est déjà bien lancé, et ce n'est pas quelques citations qui feront une grande différence sur son succès futur...
- ▶ ...à moins que ces citations ne proviennent d'une scientifique qui a elle-même, déjà, beaucoup de crédit, et dont les articles sont lus par la majeure partie de la population. Mais, comme nous l'avons vu, les scientifiques qui ont tendance à citer leurs étudiants ont moins de chances d'obtenir beaucoup de crédit et donc de se retrouver en position de faire une différence significative pour leurs étudiants en les citant. Et de plus, s'ils se retrouvent dans cette position, c'est qu'ils ont déjà réussi, et que le crédit futur qu'ils pourraient obtenir grâce à leurs étudiants ne pèse pas très lourd dans la balance. (Rappelons que la distribution du crédit dans la simulation est fortement asymétrique : ceux qui en ont en ont beaucoup, et les autres en ont très peu.)

Étant donné ces facteurs, il est plus avantageux pour les scientifiques, dans la simulation, de consacrer la partie « crédit » de leurs citations (l'autre partie étant consacrée au support) à se citer eux-mêmes plutôt qu'à citer leurs étudiants. Cela augmente leurs chances d'accumuler beaucoup de crédit en début de carrière, à un moment où, de toute façon, ils ne peuvent pas faire grand-chose pour leurs étudiants.

On pourrait être tenté de penser que cette situation découle du fait que le modèle ne permet qu'un nombre fixe de références par article, et que chaque citation d'un étudiant se fait donc au détriment d'une potentielle autocitation. Ce n'est toutefois pas exactement le cas, puisque le crédit attribué aux auteurs cités, suite à la lecture

d'un article, est attribué en fonction de la *proportion* de citations de chaque auteur, sans égard au nombre absolu de celles-ci (voir code 5.18 à la page 162).

Peut-être qu'une stratégie payante pour les scientifiques serait de moduler  $\beta$  en fonction du crédit accumulé jusqu'ici : une scientifique qui n'a pas beaucoup de crédit agirait de façon plus égoïste (avec  $\beta$  près de zéro) au départ, mais deviendrait plus altruiste avec ses étudiants (en augmentant  $\beta$ ) si elle parvenait à accumuler une quantité suffisante de crédit. Il serait possible d'intégrer une telle stratégie dans notre modèle, mais c'est une tâche à ranger dans la catégorie des recherches futures.

Notons aussi que, dans la vraie vie, il y a d'autres façons, à part les citations, de favoriser la carrière de ses étudiants. D'abord, le crédit du superviseur rejaillit « gratuitement » sur ses étudiants. Avoir fait son doctorat avec Chomsky, par exemple, a longtemps procuré un certain avantage dans le cadre d'une carrière en linguistique, au-delà des efforts particuliers que Chomsky a pu faire pour cela. Et dans tous les cas, particulièrement pour les superviseurs qui bénéficient de moins de crédit, il existe de nombreux moyens d'aider ses étudiants : réseautage, recommandations, mentorat, etc. Il se peut très bien que ces facteurs aient, au bout du compte, une influence positive sur le crédit du superviseur et qu'il vaille la peine pour lui d'y consacrer des ressources. Notre hypothèse ne concernait toutefois que les citations, et le modèle ne permet pas de la confirmer.

Il nous faut aussi revenir sur l'ambiguïté, mentionnée à la section 3.3.2, entre le crédit et la valeur sélective conceptuelle adaptative. Quand Hull parle de l'avantage qu'il y a à citer ses étudiants, il en parle à la fois en termes de propagation des *idées* et en termes de crédit obtenu grâce à cette propagation. Ce que nous avons mesuré avec notre modèle, c'est uniquement l'avantage en termes de *crédit* qu'il y a à citer ses étudiants. On suppose que propagation des idées et obtention de crédit sont fortement corrélées, mais il est logiquement possible que le fait de citer ses étudiants contribue à la propagation des idées d'une scientifique sans qu'elle récolte sous forme de crédit les fruits de cette propagation. Notre modèle, sous sa forme actuelle, ne nous permet pas de vérifier si c'est le cas : il faudrait, pour cela, garder trace des relations de descendance entre les idées qui sont générées par les scientifiques, ce qui serait coûteux sur le plan computationnel. Reste que le lien entre crédit et valeur sélective conceptuelle

adaptative mériterait d'être exploré plus à fond, et qu'une future version du modèle devrait mesurer indépendamment ces deux éléments pour s'assurer que la corrélation supposée entre eux tient dans toutes les circonstances. Mais quoi qu'il en soit, notre hypothèse  $H_2$  portait spécifiquement sur les effets de  $\beta$  sur l'obtention de crédit, et nos résultats ne nous laissent d'autre choix que de rejeter celle-ci.

Et c'est potentiellement un problème pour la théorie de Hull : dans la mesure où les comportements des scientifiques sont principalement motivés par le désir d'obtenir du crédit, pourquoi ceux-ci citeraient-ils leurs étudiants si cela ne leur apportait pas de crédit ? Nous voyons plusieurs possibilités :

- Peut-être que notre modèle omet un élément important, et que dans la réalité, citer ses étudiants apporte bel et bien du crédit ;
- Peut-être que les scientifiques *pensent* que le fait de citer leurs étudiants leur apportera du crédit, mais qu'ils ont tort ;
- Peut-être que la tendance à citer ses étudiants n'est pas motivée par le crédit, mais par une autre forme d'altruisme ;
- Peut-être que la tendance à citer ses étudiants est motivée par des raisons non liées à l'altruisme : par exemple, le fait que le superviseur connaît bien les travaux de ses étudiants, que ces travaux sont habituellement proches de ses propres intérêts de recherche, etc. ;
- Peut-être que, dans la réalité, les superviseurs ne citent pas leurs étudiants. Il nous semble, comme il semblait à Hull, que ce genre de citation est relativement fréquent, mais pour s'en assurer, il faudrait faire une étude empirique rigoureuse.

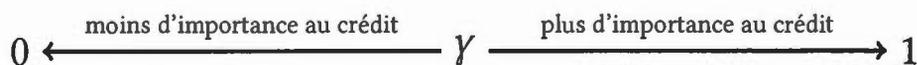
Notre conclusion, c'est qu'on en est à un point où il serait souhaitable d'entreprendre une démarche empirique. Il y a au moins un domaine où nous avons une certaine quantité d'information sur les liens de « parenté » entre scientifiques : les mathématiques.

Le « Mathematics Genealogy Project »<sup>84</sup> répertorie les liens superviseurs/étudiants chez les docteurs en mathématique de Leibniz à nos jours. En recoupant ces liens avec une base de données bibliographique, on pourrait vérifier si les superviseurs citent bel et bien leurs étudiants et si on peut établir des liens avec le succès futur du superviseur et de l'étudiant. Il s'agit toutefois d'un projet de recherche à part entière qui dépasse largement le cadre de cette thèse.

### Influence de $\gamma$ sur le crédit des scientifiques

Avant de passer à la performance épistémique en fonction des caractéristiques individuelles, nous allons encore dire quelques mots sur la performance sociale, en nous penchant (un peu plus brièvement) sur l'effet de  $\gamma$  et  $\delta$ .

La caractéristique  $\gamma$  intervient lors de la lecture des articles (section 5.4.5). Elle détermine la proportion de l'évaluation d'un article qui est basée sur le crédit de l'auteur de l'article (et de ceux des articles cités dans celui-ci) par rapport à l'évaluation « de base » qui est faite en fonction des idées existantes. Lorsque  $\gamma = 0$ , le crédit ne joue aucun rôle : tout ce qui compte, c'est la proximité de l'idée présentée dans l'article avec des idées déjà évaluées de façon positive ou négative. Quand  $\gamma = 1$ , au contraire, ça n'est que le crédit qui compte, et la position de l'idée dans l'espace épistémique n'est pas du tout considérée. Visuellement :



Nous n'avions pas formulé d'hypothèse explicite quant à l'effet de  $\gamma$  sur le crédit. Nous nous doutons que, pour ce qui est de la performance épistémique, il se pourrait qu'il soit nuisible d'accorder trop d'importance au crédit, mais nous verrons cela plus loin, à la section 6.5. Mais peut-il être bénéfique, pour son propre crédit, d'accorder une

---

<sup>84</sup> La base de données, administrée par Harry Coonce est disponible sur <http://genealogy.math.ndsu.nodak.edu>. Voir Jackson (2007) pour un bref historique du projet.

certain importance au crédit des autres lorsqu'on évalue des articles ? Si on se réfère à la figure 6.12 à la page suivante, on constate que la réponse dépend du paysage épistémique sur lequel on se situe : on a une relation de compromis pour DE JONG F1, HORNS 5 PEAKS, LANGERMANN F11 et FLAT, mais une relation linéaire négative pour EXP et ROOTS. Nos diagrammes 6.9 (p. 194) et 6.11 (p. 197) montrent que toutes ces corrélations sont significatives.

Qu'est-ce qui explique ces relations ? Notre tentative de réponse, c'est que sur les paysages où on observe une relation de compromis, la raison pour laquelle il est important d'accorder au moins une certaine importance au crédit, c'est que cela nous amène à valoriser les idées qui sont déjà, probablement, valorisées par la communauté. Cela repose sur le fait que le crédit qu'une scientifique accorde à un auteur a tendance à être corrélé avec le crédit que tous les autres scientifiques accordent à ce même auteur puisque ce sont, de façon générale, les auteurs les plus publiés et les plus cités qui accumulent le plus de crédit. Et si on base nos propres idées sur des idées qui sont déjà valorisées par les autres, on a plus de chances de voir nos propres idées acceptées, ce qui augmente notre crédit. On pourrait dire que le crédit, en général, est une force conservatrice : plus on lui accorde d'importance, moins on s'écarte des idées reçues, ce qui, bien souvent, est une voie plus sûre.

Ce conservatisme n'est toutefois pas avantageux dans tous les paysages. Pourquoi ? Dans EXP, en particulier, cela saute aux yeux : il faut savoir reconnaître les bonnes idées lorsque, finalement, elles se présentent. Si la communauté a passé la majeure partie de son temps à errer dans la zone de valeur objective nulle, le crédit est un bien piètre indicateur de la qualité des idées. Et le sommet de valeur objective dans EXP est si étroit que les scientifiques qui ne saisissent pas rapidement l'occasion d'explorer cette zone passent à côté de la chance d'accumuler beaucoup de crédit. C'est aussi ce qui se produit, dans une moindre mesure, sur ROOTS, avec ses six sommets. On aurait pu s'attendre à observer un phénomène similaire sur HORNS 5 PEAKS, mais puisque la communauté n'arrive généralement pas à trouver les sommets (qui sont « cachés » dans des vallées de basse valeur objective), la situation ne tend pas à se présenter.

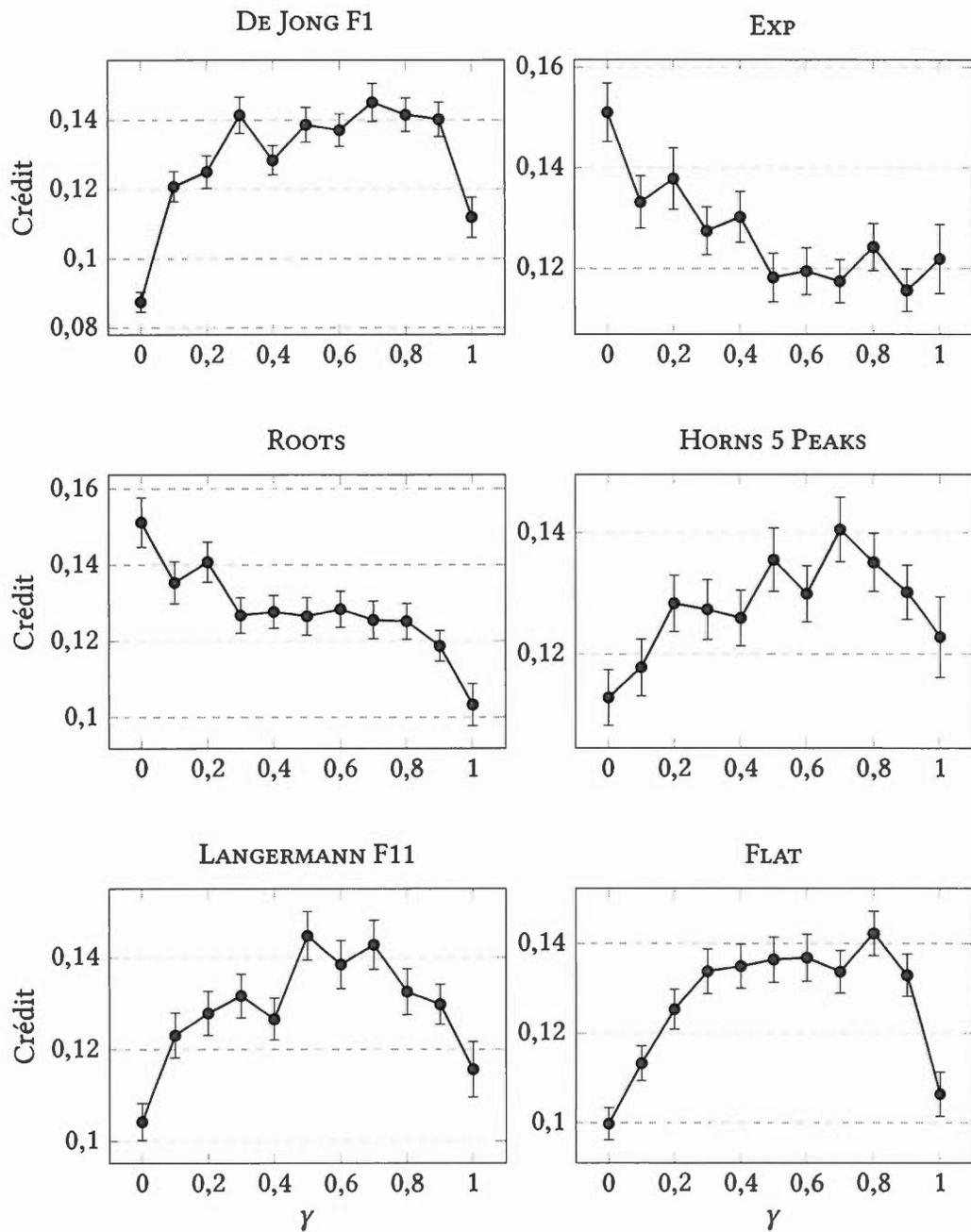
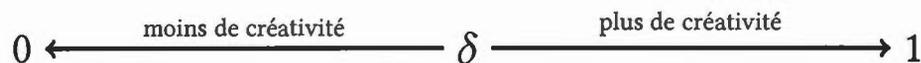
Crédit en fonction de  $\gamma$ 

Figure 6.12

En conclusion : l'importance qu'une scientifique doit apporter au crédit des autres pour favoriser son crédit personnel varie selon la structure du paysage épistémique dans lequel elle se trouve.

### Influence de $\delta$ sur le crédit des scientifiques

On en vient à la dernière des caractéristiques personnelles de nos scientifiques :  $\delta$ . Celle-ci représente, d'une certaine façon, le « taux de mutation » des idées : après qu'une ou deux idées « parentes » aient été combinées pour donner naissance à une nouvelle idée, on applique une mutation gaussienne à chacune de ses coordonnées, et  $\delta$  détermine l'écart-type de cette variation. Lorsque  $\delta = 0$ , la seule source de variation est la recombinaison des coordonnées des parents. En d'autres termes,  $\delta$  représente la créativité d'une scientifique :



Alors, est-il bon d'être créatif ? Pour LANGERMANN F11 et FLAT, on a aussi une corrélation linéaire positive, mais celle-ci repose surtout sur le fait que lorsque  $\delta = 0$ , il est très difficile pour une scientifique de publier des articles originaux, et son crédit s'en ressent dramatiquement. Les résultats (figure 6.13 à la page suivante) nous montrent que l'important, c'est d'être au moins un peu créatif. La figure 6.9 à la page 194 nous montre une relation de compromis significative pour  $\delta$ , puisque les courbes obtenues s'alignent fort bien sur la « cloche » de la figure 6.8, à part pour  $\delta = 1$ .

La situation s'apparente tout de même au compromis entre « exploration » et « exploitation » que l'on retrouve dans plusieurs problèmes d'optimisation et d'apprentissage. Celui-ci est présent notamment dans les algorithmes génétiques (Holland, 1975) et les algorithmes d'apprentissage par renforcement (Barto et Sutton, 1998), mais a été d'abord porté formellement à l'attention de la communauté scientifique par Robbins (1952) dans le cadre de ce qu'on appelle maintenant le problème des bandits manchots.

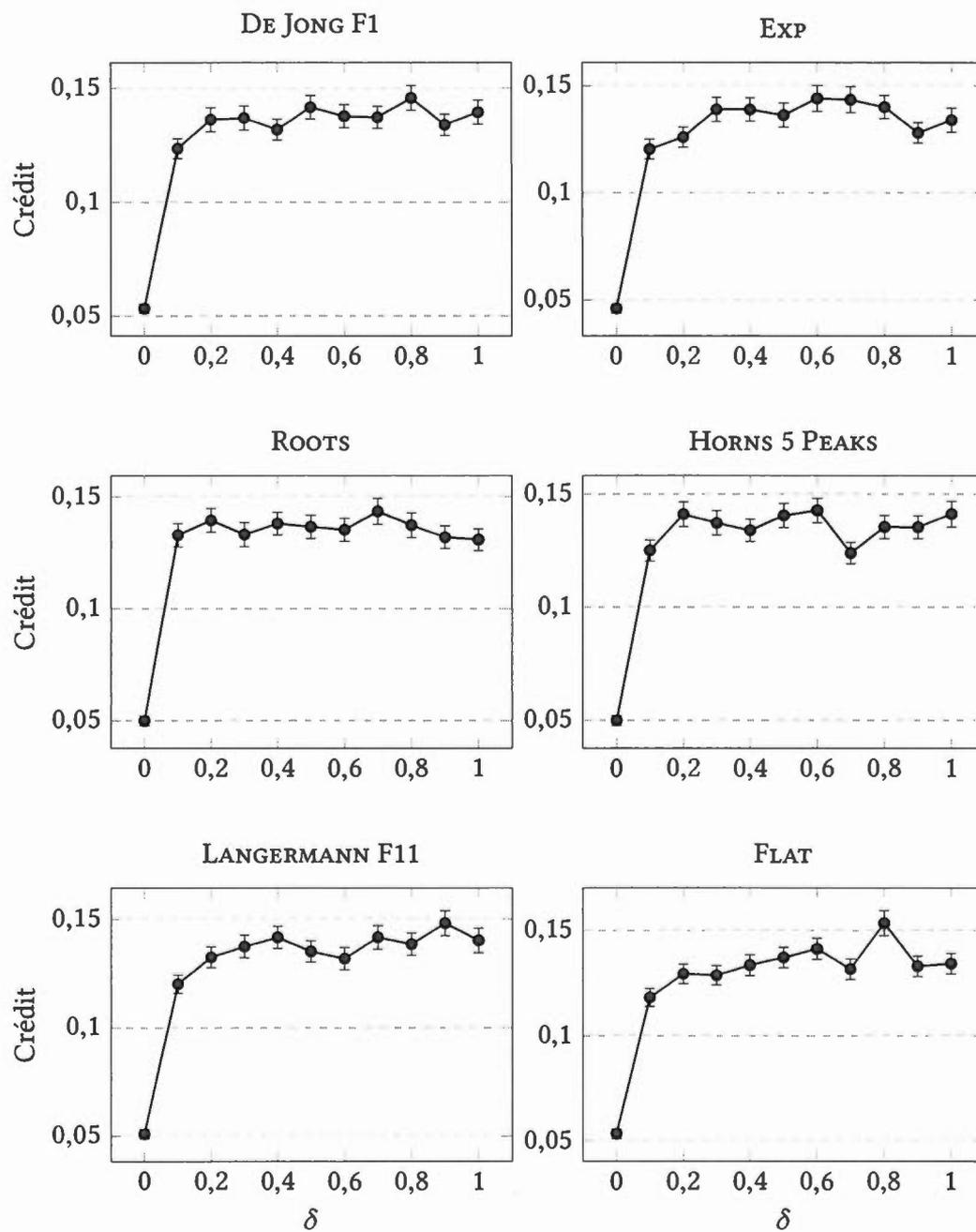
Crédit en fonction de  $\delta$ 

Figure 6.13

Dans celui-ci, on est confronté à un nombre  $n$  de bandits manchots (plus connus au Québec sous le nom de « machines à sous »). La probabilité de gagner, inconnue, est différente pour chaque machine. Le problème consiste à déterminer le moment optimal pour cesser d'« explorer » (essayer les différentes machines pour tenter de déterminer la probabilité de gagner sur chacune d'elle) et se mettre à « exploiter » (se concentrer sur la machine la plus payante). Les valeurs moindres de  $\delta$  correspondent ici à l'exploitation : on se tient près des idées que l'on sait déjà être bonnes. Les valeurs élevées correspondent à l'exploration : on va vers l'inconnu dans l'espoir de trouver des idées meilleures que celles que l'on possède déjà.

Comme nous en avons fait l'hypothèse, plus haut, pour  $\beta$ , une stratégie viable serait peut-être de varier  $\delta$  en fonction du crédit accumulé : avoir un  $\delta$  élevé en début de carrière pour explorer plus de possibilités, puis limiter notre créativité une fois qu'on a trouvé un bon filon à exploiter. Mais peut-être aussi que la meilleure stratégie serait, au contraire, de commencer de façon conservatrice, en s'alignant sur les idées des autres et de ne laisser libre cours à notre créativité qu'une fois acquise une certaine sécurité. Pour le savoir, il faudrait ajouter ces stratégies au modèle et les comparer entre elles.

Mais pour l'instant, nous en sommes plutôt arrivés à une autre étape de notre analyse : nous pencher sur la performance épistémique de la communauté.

### Performance épistémique en fonction des caractéristiques individuelles

Jusqu'à maintenant, nous avons examiné la façon dont les valeurs des caractéristiques individuelles influençaient l'accumulation de crédit par les individus. La question à laquelle nous allons tenter de répondre maintenant est tout autre : que se passe-t-il lorsqu'on impose une norme commune et qu'on force tous les scientifiques à adopter la même valeur pour la même caractéristique ? C'est ce que nous avons fait dans l'expérience  $E_2$  (décrite à la section 6.1) : dans chaque simulation, il y a une caractéristique qui est fixée à une valeur donnée et les autres caractéristiques sont attribuées de façon aléatoire. Notre critère de mesure pour la performance épistémique de la communau-

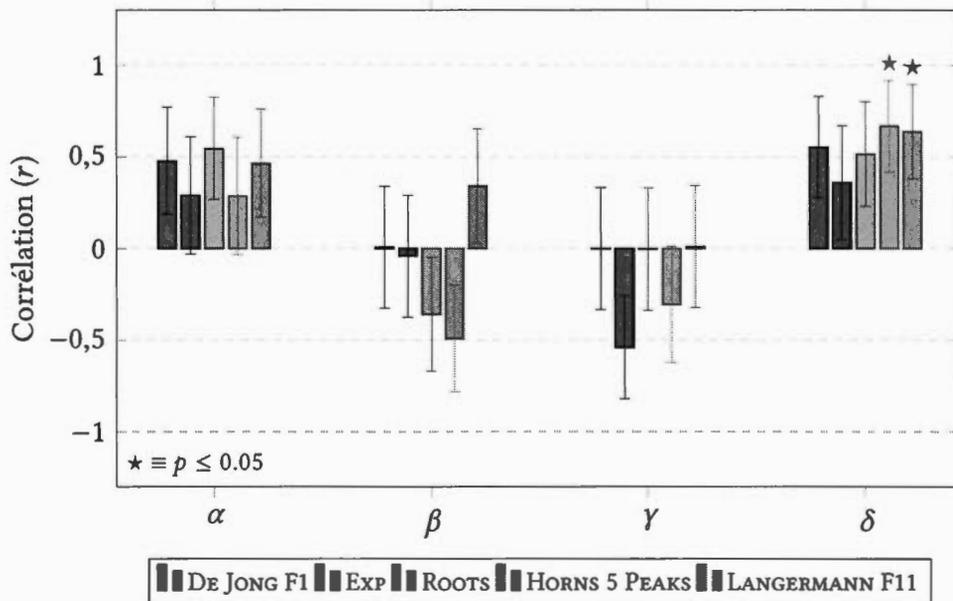
Relation de compromis entre caractéristiques individuelle et valeur objective ( $E_2$ )

Figure 6.14

té, c'est simplement la valeur objective moyenne des articles qui sont publiés pendant chaque simulation.

La présente section constitue un prérequis pour la section suivante, où nous tenterons de voir si ce qui est bon pour les scientifiques individuels est aussi bon pour la science (l'hypothèse  $H_3$ ). Nous avons vu, à la section précédente, ce qui était bon pour les scientifiques, alors il nous faut maintenant voir ce qui est bon pour la science.

Comme nous l'avons fait pour le crédit à la section précédente, nous verrons pour chacune des caractéristiques s'il y a une relation de compromis (figure 6.14) ou bien une relation linéaire (figure 6.15) entre la valeur de celle-ci et, cette fois, la valeur objective moyenne des articles.

Les graphiques de la figure 6.16 à la page 210 montrent ce qui arrive à la valeur objective moyenne des articles lorsqu'on fixe la valeur de  $\alpha$  sur différents paysages. Les figures 6.18, 6.20 et 6.22 font la même chose pour  $\beta$ ,  $\gamma$  et  $\delta$ , respectivement.

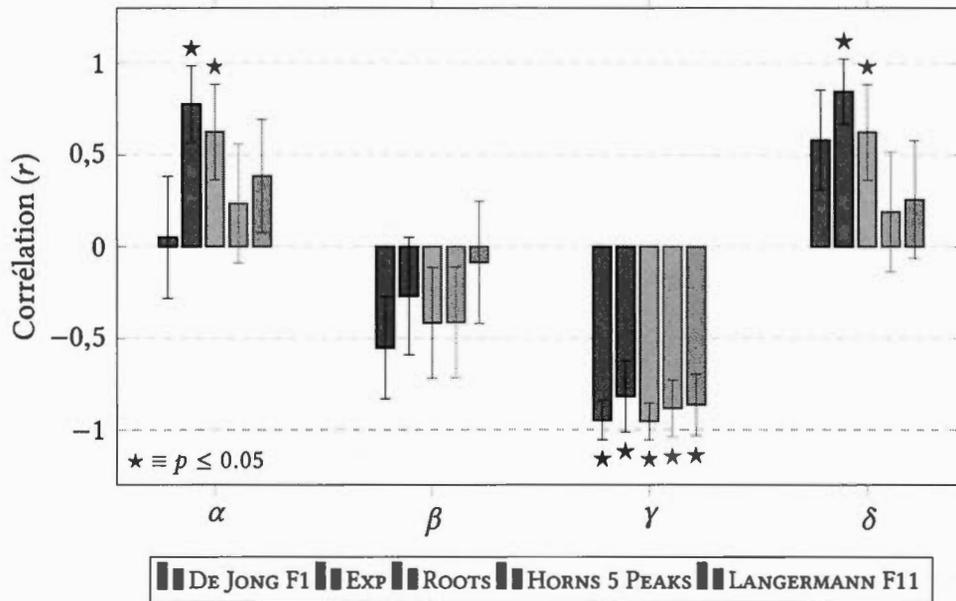
Relation linéaire entre caractéristiques individuelle et valeur objective ( $E_2$ )

Figure 6.15

Notons d'abord que sur tous ces graphiques, nous avons ajouté une ligne horizontale rouge qui montre la valeur objective des articles sur le paysage correspondant dans  $E_1$ , alors que toutes les caractéristiques étaient attribuées de façon aléatoire. Cela nous permet de nous assurer que le simple fait d'assigner la même valeur à tous les individus n'a pas un effet démesuré par rapport à la performance obtenue dans une population diversifiée. Pour  $\alpha$ , comme pour les autres caractéristiques, les résultats de la population diversifiée sont semblable à ceux des populations fixes.

Notons aussi que, même si le paysage FLAT est inclus sur les figures 6.16 à 6.20, il n'y a pas grand-chose à dire sur celui-ci : dans tous les cas, la valeur objective moyenne des articles sur celui-ci est 0,5, puisque c'est la seule possible sur ce paysage. Lorsque nous ferons, dans les paragraphes qui suivent, des énoncés concernant « tous les paysages », il faudra bien entendu exclure FLAT.

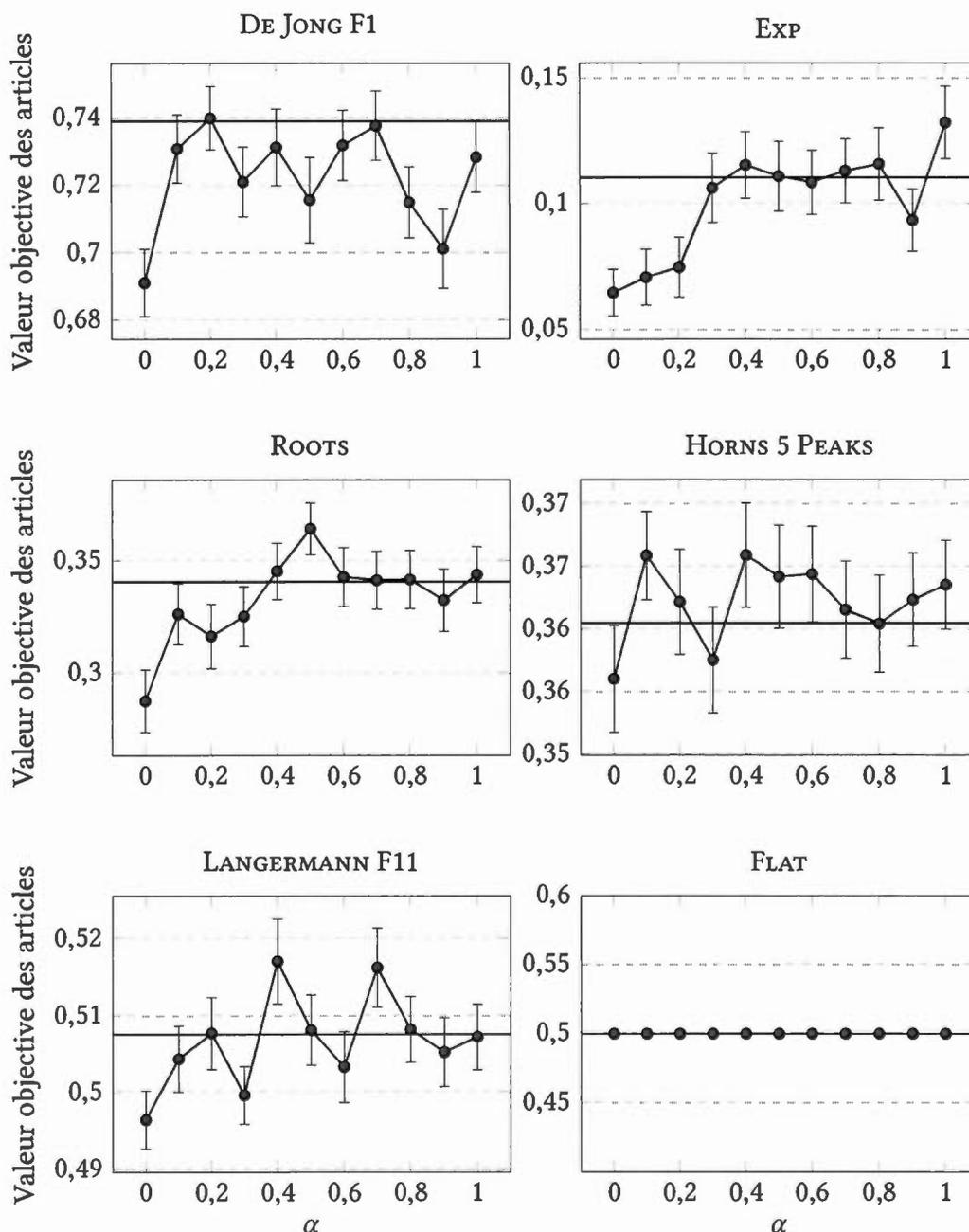
### Influence de $\alpha$ sur la valeur objective des articles

Alors, à quoi ressemble la relation entre  $\alpha$  et la valeur objective des articles ? Il y a beaucoup de variation d'une simulation à l'autre, comme le montre la taille des barres d'erreur. À deux exceptions près, on ne trouve pas de relation significative entre les deux. Cela s'explique, selon nous, par le fait que  $\alpha$  influence surtout à qui sera attribué le crédit dans la simulation, mais pas tant quelle zone du paysage épistémique sera explorée par la communauté. Du point de vue de la performance épistémique générale, peu importe qui reçoit le crédit pour une idée : l'important, c'est que l'idée soit diffusée dans la communauté. On remarque toutefois que, dans tous les paysages,  $\alpha = 0$  mène à une moins bonne performance. Cela s'explique, selon nous, par le fait que, lorsque tous les scientifiques se contentent de citer leurs propres articles et ceux de leurs étudiants, il n'y a aucune place pour la circulation des idées entre les différentes familles intellectuelles. Les résultats de la simulation montrent que cette circulation n'a pas besoin d'être bien grande, mais qu'elle doit avoir lieu. Le paysage EXP est celui dans lequel on retrouve ce phénomène de la façon la plus marquée : dès que quelques articles sont publiés sur le sommet de valeur épistémique, il est crucial que ces idées soient diffusées (c'est-à-dire citées) le plus rapidement possible. C'est aussi, dans une moindre mesure, le cas pour ROOTS, mais dans ce dernier paysage, ce sont les valeurs de  $\alpha$  autour de 0,5 qui semblent optimales. De plus amples recherches seraient nécessaires pour comprendre pourquoi on observe, dans ce paysage en particulier, une légère baisse de performance pour les valeurs de  $\alpha > 0,5$ .

Si on devait tirer une conclusion normative de ces résultats, ce serait quelque chose comme : accordez une certaine attention au support lorsque vous choisissez quels articles citer. Rien de révolutionnaire ici : c'est ce que les scientifiques font dans la vraie vie, de toute façon. Le modèle ne fait que confirmer que c'est avantageux à la fois pour eux et pour la science. Mais au-delà du fait qu'un peu de support est nécessaire, il est difficile de tirer une conclusion générale en ce qui concerne l'alignement de l'intérêt individuel avec l'intérêt collectif pour  $\alpha$ .

Pour évaluer notre hypothèse  $H_3$  en ce qui concerne  $\alpha$  (appelons-la  $H_{3,\alpha}$ ), il faut faire le lien entre les résultats présentés à figure 6.7 à la page 192, qui montre le crédit

### Valeur objective des articles en fonction de $\alpha$



**Figure 6.16** Les points du graphique représentent la valeur objective moyenne des articles pour chaque simulation lorsque la valeur de  $\alpha$  est fixée pour tous les scientifiques dans  $E_2$ . La ligne horizontale supplémentaire rouge représente, à fin de comparaison, la valeur objective obtenue dans  $E_1$  lorsque la valeur de  $\alpha$  varie d'un scientifique à l'autre.

### Valeur objective en fonction du crédit pour $\alpha$

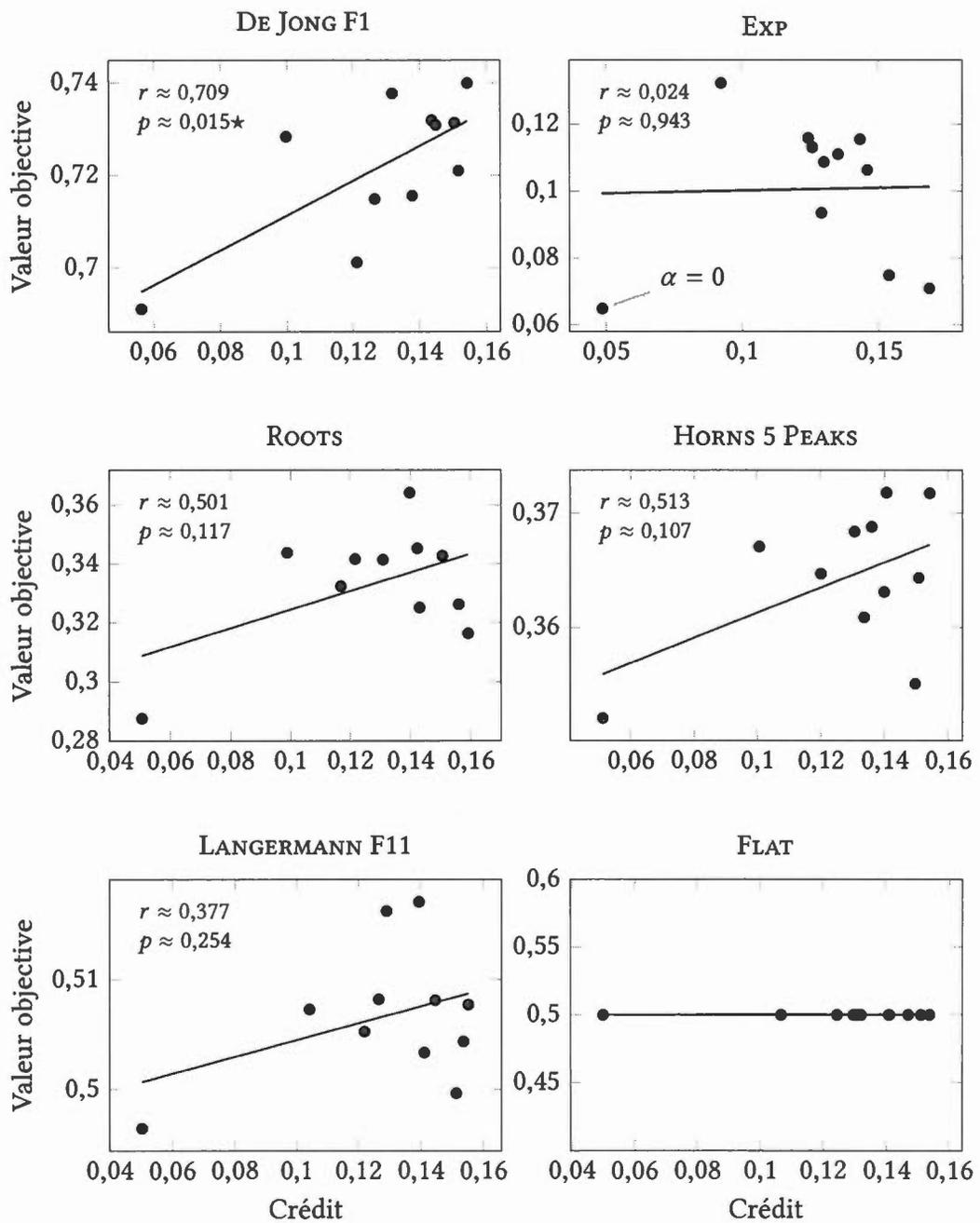


Figure 6.17

accumulé en fonction de  $\alpha$ , et ceux que nous venons de voir à la figure 6.16, qui montre la valeur objective en fonction de  $\alpha$ . Il suffit pour cela de construire, pour chaque valeur de  $\alpha$ , le point (crédit, valeur objective) correspondant. C'est ce que nous avons fait à la figure 6.17 à la page 211. Et puisque c'est des corrélations linéaires qu'on recherche ici, nous avons tracé la ligne de régression et indiqué les valeurs de  $r$  et de  $p$  pour chaque paysage. Une étoile (★) indique les corrélations significatives.

On n'en retrouve toutefois qu'une seule, pour DE JONG F1. Sur les autres paysages, le lien entre  $\alpha$  et la valeur objective n'était pas assez fort pour qu'on puisse s'attendre à quoi que ce soit lorsqu'on tente de corréler celle-ci avec le crédit.

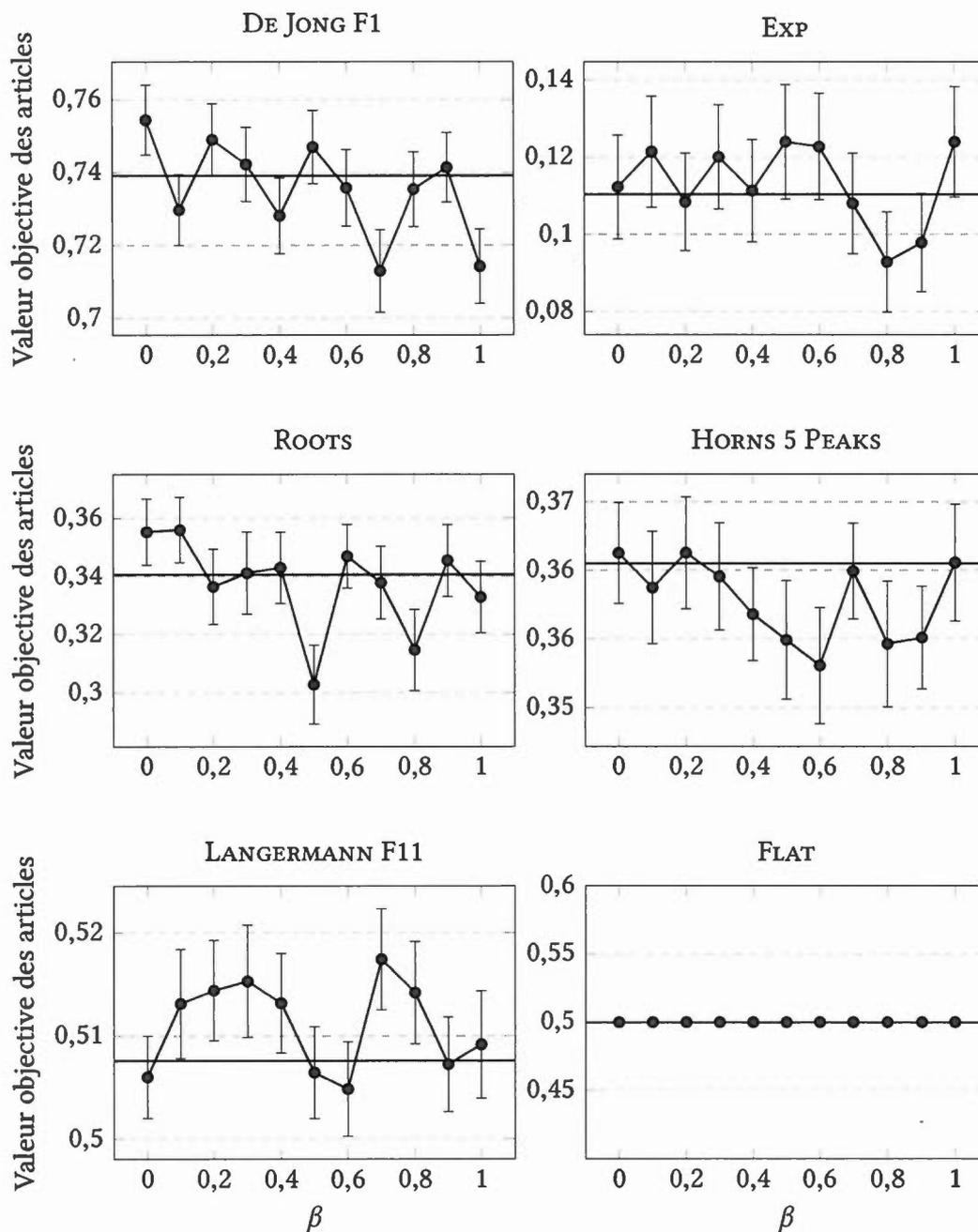
La seule exception, c'est EXP, où si ce n'était du point  $\alpha = 0$ , pour lequel le crédit et la valeur objective sont tous deux très bas, on aurait une corrélation négative. Il y a, sur ce paysage, une corrélation négative entre  $\alpha$  et la valeur objective. Mais nous avons constaté plus haut qu'il y avait une relation de compromis entre  $\alpha$  et le crédit, avec un biais vers les plus petites valeurs de  $\alpha$ . Bref : à part pour  $\alpha = 0$ , cela vaut mieux pour l'individu lorsque  $\alpha$  est petit, mais mieux pour la science lorsque  $\alpha$  est élevé.

### Influence de $\beta$ sur la valeur objective des articles

En ce qui concerne le lien entre la valeur de  $\beta$  et la performance épistémique de la communauté, on ne trouve pas une seule corrélation significative. Au mieux, on trouve de très faibles corrélations linéaires négatives (voir figure 6.15 à la page 208), mais la seule qui s'approche un tant soit peu d'être significative, c'est DE JONG F1 avec  $p \approx 0,078$ . Ce que cela voudrait dire, c'est que la communauté performe un peu mieux lorsque les superviseurs citent moins leurs étudiants. On pourrait alors supposer que c'est parce que, sans cette aide, ils doivent être jugés au mérite : ils accumulent du crédit si leurs articles recueillent des évaluations positives et sont ensuite cités pour cette raison, plutôt que parce qu'ils bénéficient des citations de leurs superviseurs qui sont préoccupés par leur propre postérité intellectuelle.

On se rappellera qu'on avait aussi trouvé des corrélations linéaires négatives (significatives, celles-là) entre  $\beta$  et le crédit accumulé. Se pourrait-il que nous soyons en

### Valeur objective des articles en fonction de $\beta$



**Figure 6.18** Les points du graphique représentent la valeur objective moyenne des articles pour chaque simulation lorsque la valeur de  $\beta$  est fixée pour tous les scientifiques dans  $E_2$ . La ligne horizontale supplémentaire rouge représente, à fin de comparaison, la valeur objective obtenue dans  $E_1$  lorsque la valeur de  $\beta$  varie d'un scientifique à l'autre.

### Valeur objective en fonction du crédit pour $\beta$

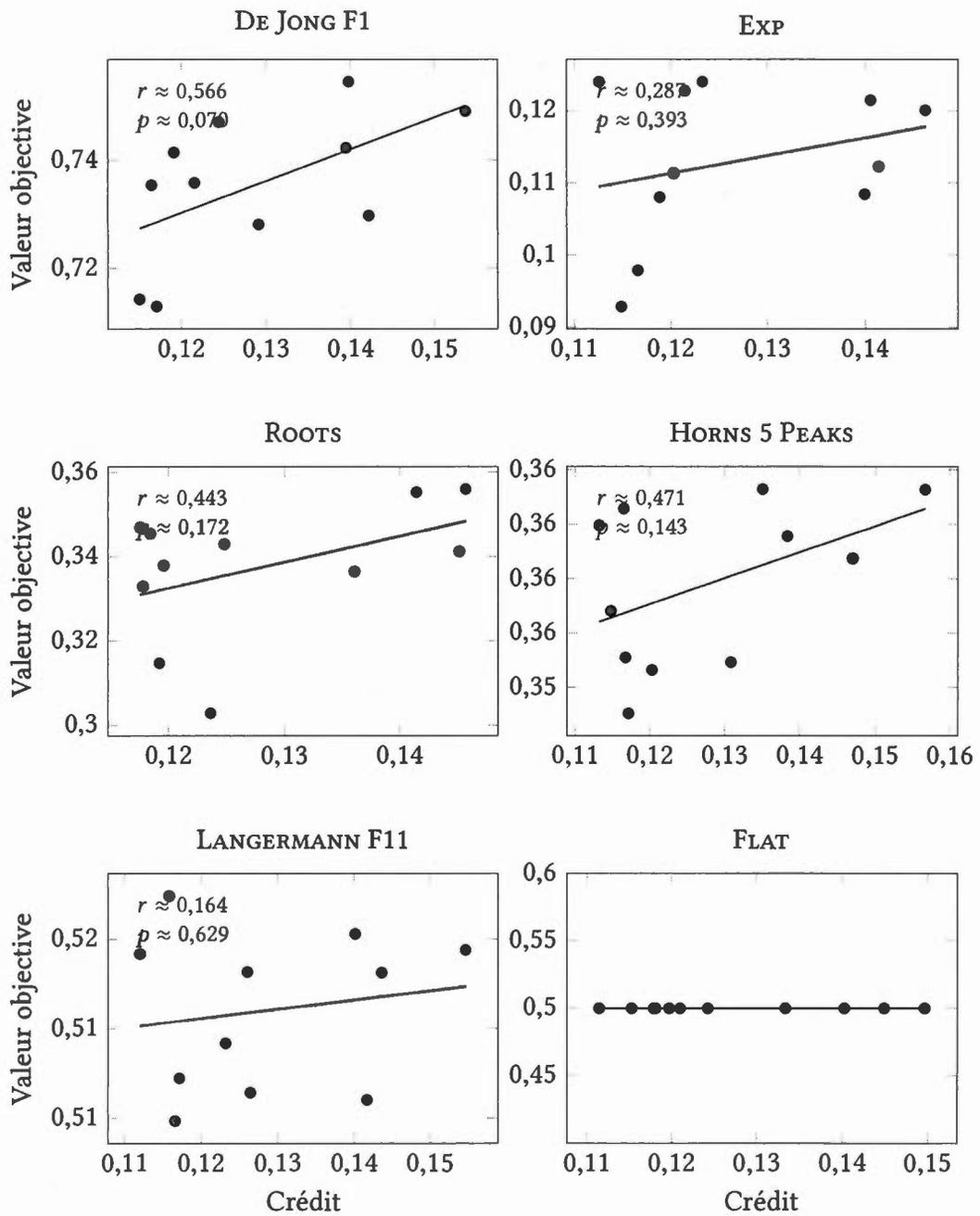


Figure 6.19

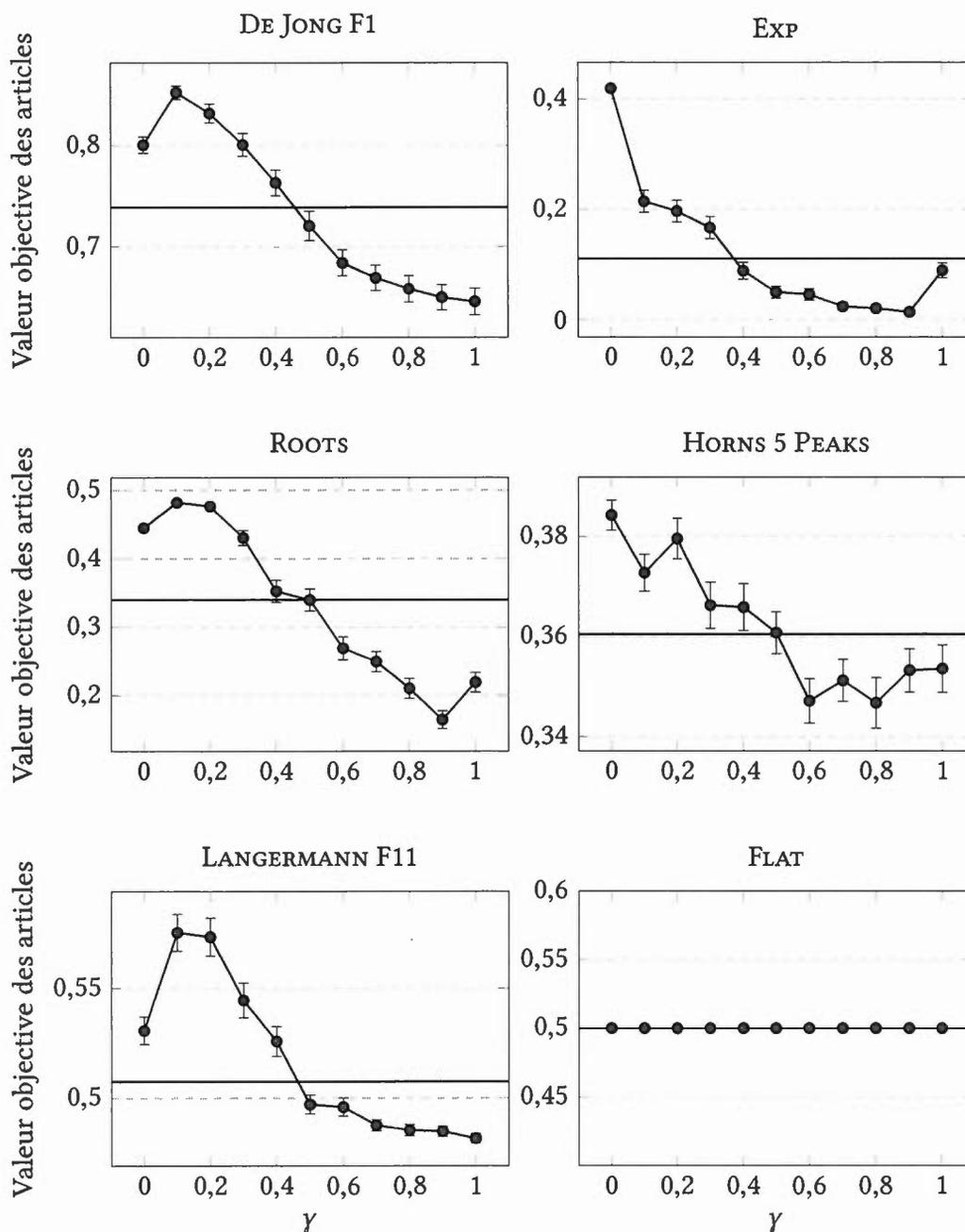
mesure d'accepter  $H_{3\beta}$ , et qu'on puisse dire avec confiance que le « népotisme intellectuel » n'est bon ni pour l'individu ni pour la communauté? Pas tout à fait, non. Comme on peut le voir à la figure 6.19 à la page 214, on ne trouve aucune relation significative. Il faut mettre l'absence de confirmation sur la faiblesse de la relation entre  $\beta$  et la valeur objective.

### Influence de $\gamma$ sur la valeur objective des articles

Le cas de  $\gamma$  est beaucoup plus tranché : comme on peut le voir clairement à la figure 6.20 à la page suivante, on a une corrélation linéaire négative significative sur tous les paysages. Moins les scientifiques accordent de l'importance au crédit lors de l'évaluation, mieux c'est pour la science.

On touche ici à un point crucial : il se pourrait que le rôle joué par le crédit dans le processus scientifique ne soit pas entièrement positif. Il constitue, bien sûr, une source de motivation importante et présente, en ce sens, des avantages indéniables. Mais lorsqu'il vient teinter l'évaluation des idées, comme c'est le cas dans le modèle lorsque  $\gamma$  est élevé, la science s'en ressent : les idées qui ont le plus de chances d'être acceptées ne sont pas nécessairement les meilleures, mais bien celles qui proviennent de la source la plus respectée. Cela n'est pas que négatif, puisqu'on peut supposer, dans le monde réel comme dans le modèle, que les scientifiques qui ont produit de bonnes idées par le passé produisent encore de bonnes idées dans le futur. Mais ce n'est pas nécessairement le cas non plus. Dans un paysage comme *Exp*, par exemple, il peut s'écouler plusieurs années avant que qui que ce soit trouve une idée dont la valeur objective est plus grande que zéro. Et pourtant, pendant ces années, certains scientifiques publient des articles, accumulent du crédit, attirent des étudiants, etc. Pourtant, ce succès individuel n'est nullement garant de leur capacité à trouver le sommet de valeur objective avant les autres. Il se peut très bien que ce soit un scientifique en début de carrière qui trouve ce sommet en premier, mais dans les simulations où  $\gamma$  est élevé, la communauté perd un précieux temps à attendre qu'un chercheur reconnu publie un article dans cette région du paysage épistémique.

## Valeur objective des articles en fonction de $\gamma$



**Figure 6.20** Les points du graphique représentent la valeur objective moyenne des articles pour chaque simulation lorsque la valeur de  $\gamma$  est fixée pour tous les scientifiques dans  $E_2$ . La ligne horizontale supplémentaire rouge représente, à fin de comparaison, la valeur objective obtenue dans  $E_1$  lorsque la valeur de  $\gamma$  varie d'un scientifique à l'autre.

### Valeur objective en fonction du crédit pour $\gamma$

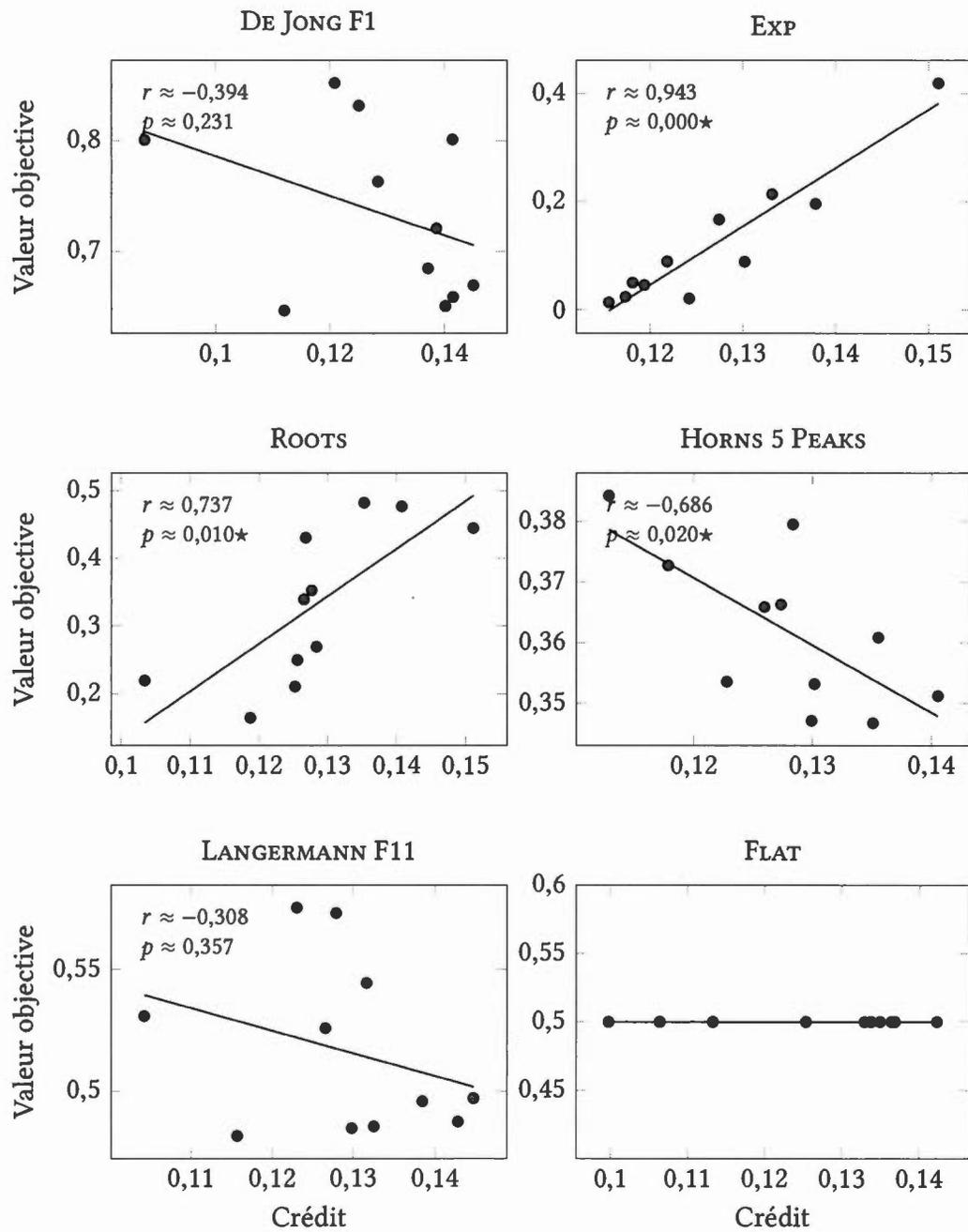


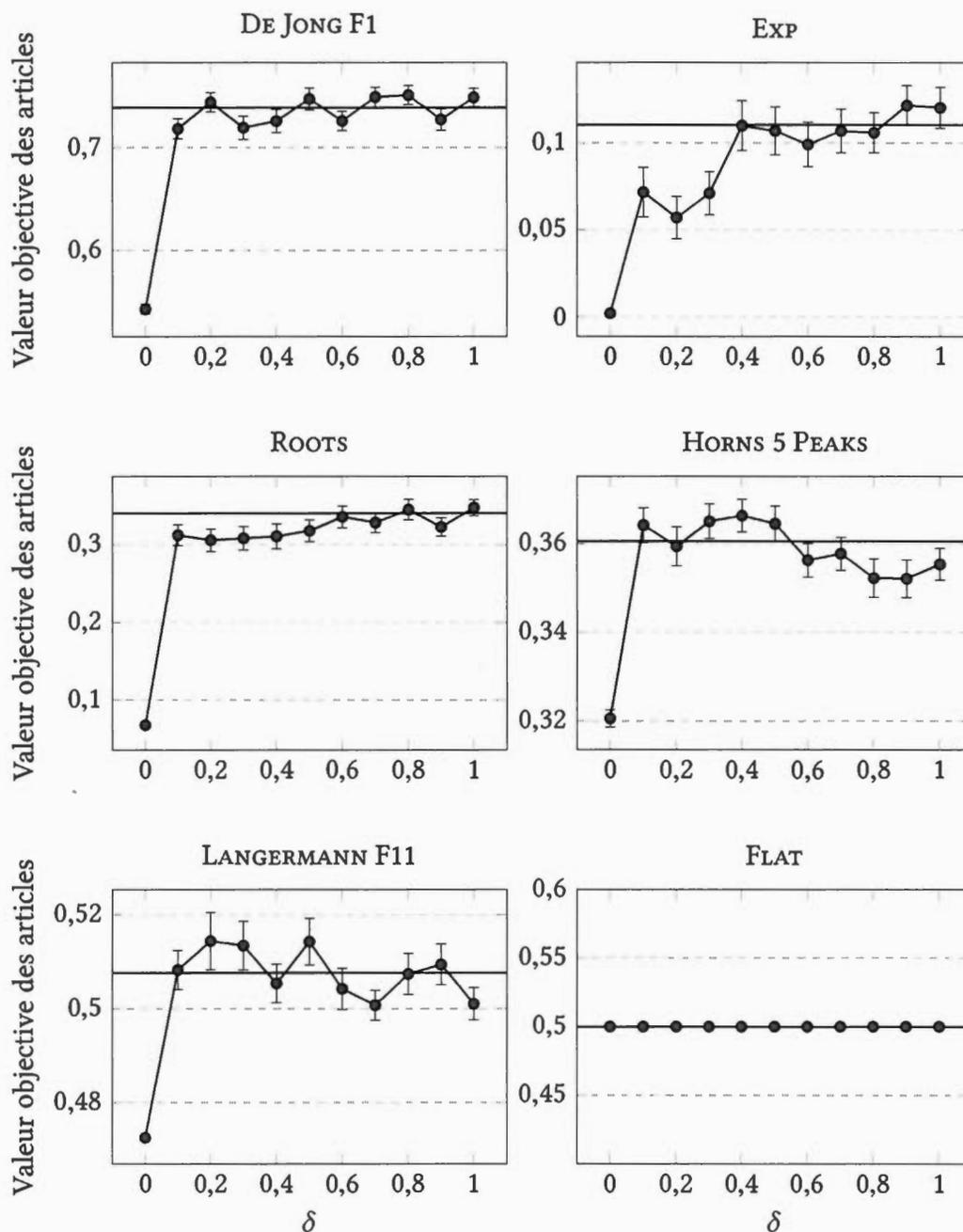
Figure 6.21

Cet état de choses ne serait toutefois une mauvaise nouvelle pour la science que si l'intérêt des scientifiques individuels les poussait à accorder une importance démesurée au crédit lors de l'évaluation des idées. Heureusement, nous avons vu à la figure 6.12 à la page 203 que ce n'est pas le cas. Sur EXP et ROOTS, à tout le moins, on avait le même genre de relation entre  $\gamma$  et le crédit qu'entre  $\gamma$  et la valeur objective : une corrélation linéaire négative. Dans ces deux paysages,  $H_{3\gamma}$  se confirme : il vaut mieux, à la fois pour les individus et pour la communauté, ne pas accorder trop d'importance au crédit. Sur les autres paysages, par contre, c'est moins clair : c'est une relation de compromis qu'on a trouvée entre  $\gamma$  et le crédit. Il est important, dans ceux-ci, d'accorder une certaine importance au crédit dans le processus d'évaluation. On constate même, à la figure 6.20 à la page 216, que la relation entre l'intérêt individuel et l'intérêt collectif  $\gamma$  est négative (quoique pas significative, à part pour HORNS 5 PEAKS).

On a donc, une fois de plus, un portrait mitigé pour  $H_3$  : ce qui est bon pour la science n'est pas nécessairement bon pour les individus.

#### Influence de $\delta$ sur la valeur objective des articles

On se rappellera que, lorsqu'on a examiné la relation entre  $\delta$  et le crédit individuel, on a trouvé que le plus important pour une scientifique, c'est de faire preuve d'au moins un peu de créativité. On ne sera pas surpris de retrouver la même chose en ce qui concerne la valeur objective à la figure 6.22 à la page suivante : lorsque  $\delta = 0$ , la communauté peine à produire de bonnes idées. Si on se rappelle du processus de génération des idées décrit à la section 5.4.1, ce n'est pas surprenant : sans les mutations permises par  $\delta > 0$ , les coordonnées des nouvelles idées sont limitées aux coordonnées  $x$  et  $y$  de leurs parents. Faute de variation, les idées initiales de la simulation limitent radicalement la proportion du paysage épistémique qu'il est possible d'explorer. La figure 6.24a à la page 221 montre un exemple visuel frappant de cette condition dans le paysage EXP, où elle est particulièrement néfaste. Les autres images de la figure 6.24 montrent que la communauté arrive à explorer une proportion de plus en plus grande du paysage à mesure que  $\delta$  augmente (ce qui n'est toutefois pas une garantie de trouver le sommet, comme le montre 6.24e).

Valeur objective des articles en fonction de  $\delta$ 

**Figure 6.22** Les points du graphique représentent la valeur objective moyenne des articles pour chaque simulation lorsque la valeur de  $\delta$  est fixée pour tous les scientifiques dans  $E_2$ . La ligne horizontale supplémentaire rouge représente, à fin de comparaison, la valeur objective obtenue dans  $E_1$  lorsque la valeur de  $\delta$  varie d'un scientifique à l'autre.

### Valeur objective en fonction du crédit pour $\delta$

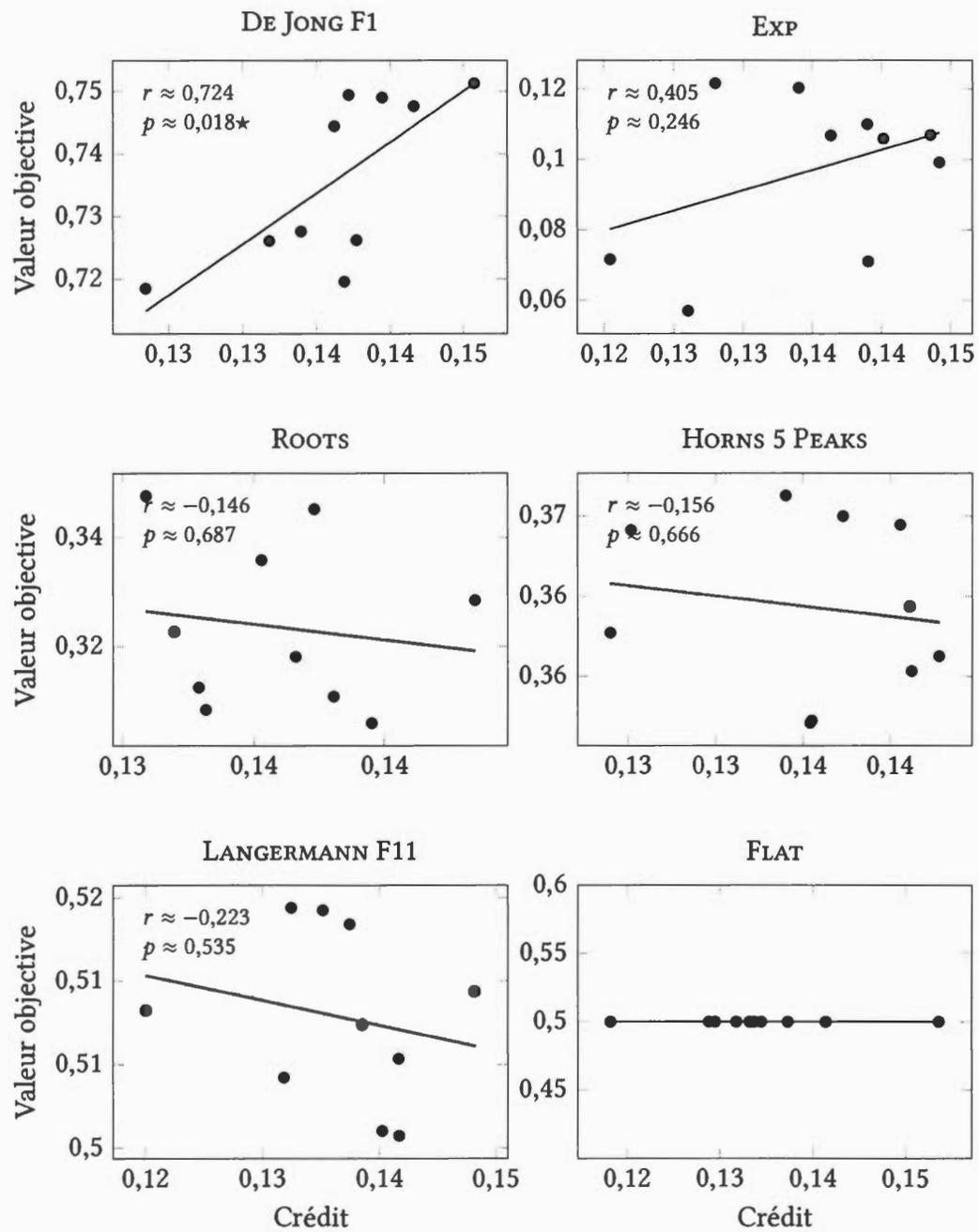
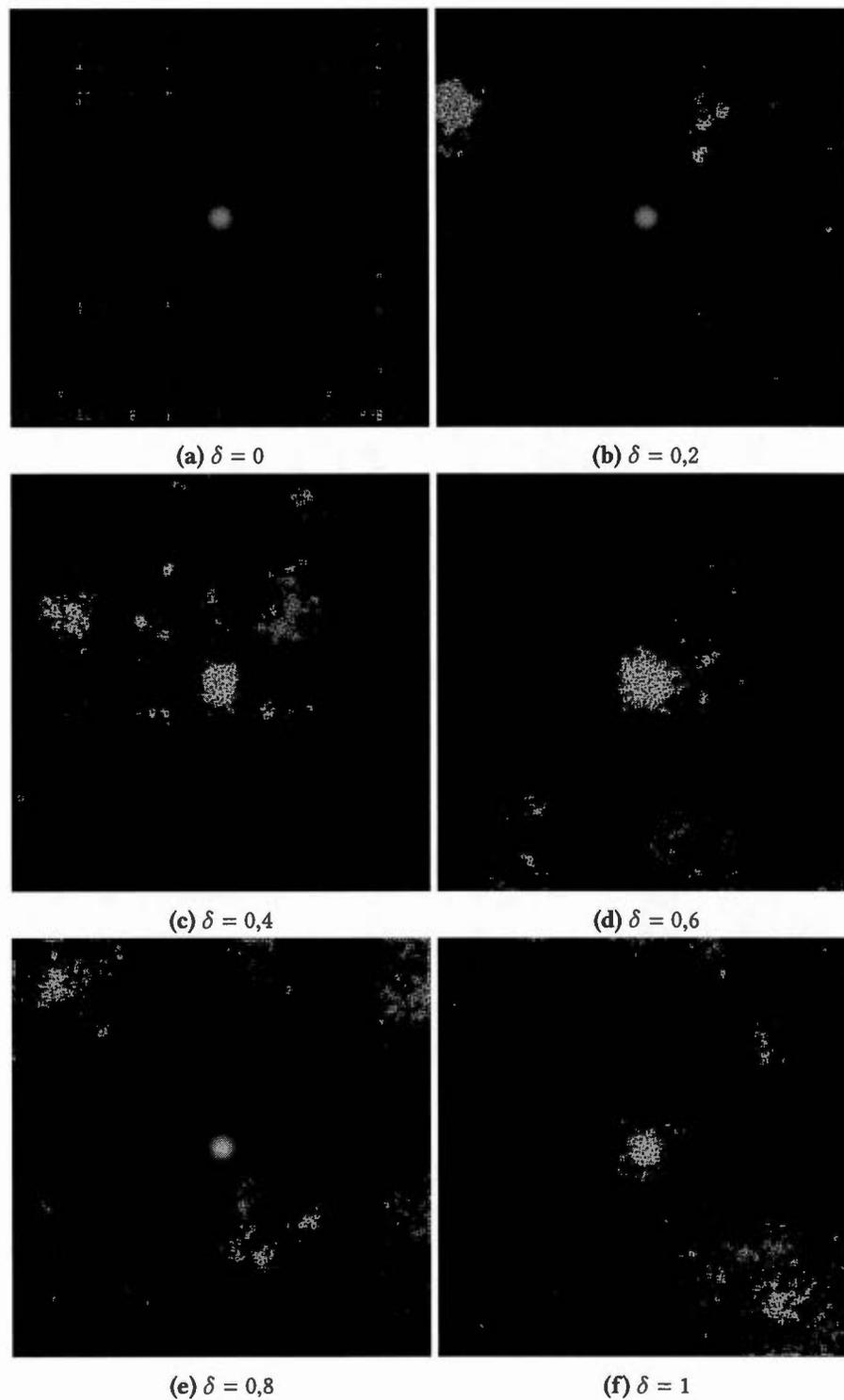


Figure 6.23



**Figure 6.24** Distribution spatiale finale des articles sur Exp pour différentes valeurs de  $\delta$ .

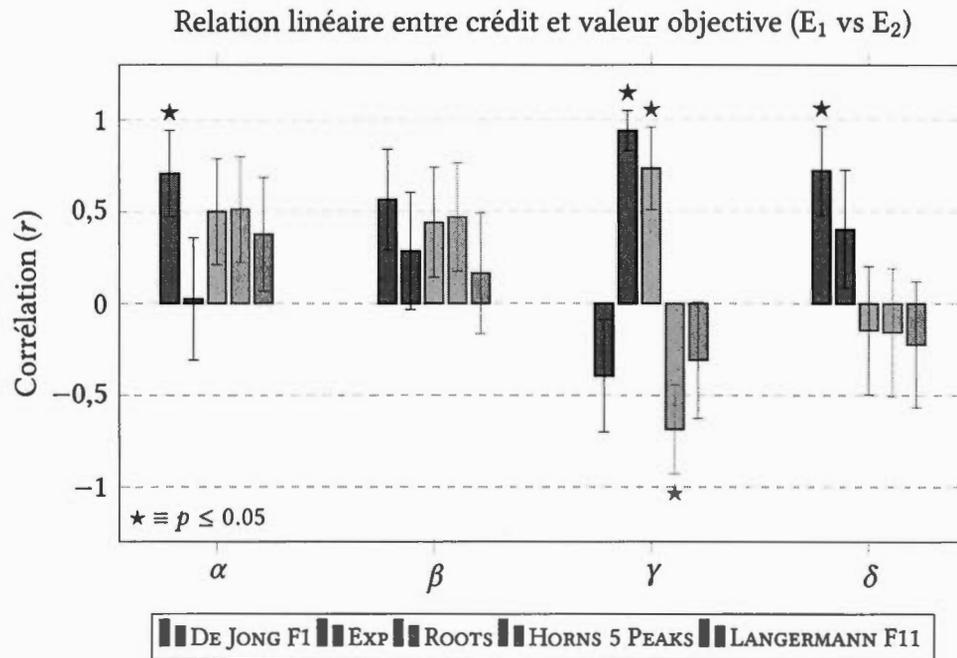


Figure 6.25

Si on établit nos corrélations entre le crédit et la valeur objective pour  $\delta$  en tenant compte de  $\delta = 0$ , on obtient des corrélations positives significatives. Mais il nous a paru pertinent de nous pencher plutôt sur ce qu'il en était lorsqu'on exclut ce point aberrant (cet *outlier*). Les corrélations présentées à la figure 6.23 à la page 220 n'incluent donc pas  $\delta = 0$ .

Résultat ? Une seule corrélation positive significative (sur DE JONG F1) entre le crédit et la valeur objective pour  $\delta$ . Dans ces circonstances, nous n'avons d'autre choix que de rejeter  $H_{3_\delta}$ .

#### Relation entre crédit et valeur objective

Dans la mesure où nous n'avons pu accepter pleinement aucune des sous-hypothèses  $H_{3_{\alpha.. \delta}}$ , il nous est également impossible d'accepter  $H_3$  en général. La figure 6.25 résume la situation.

On trouve des relations positives significatives çà et là ( $\alpha$  et  $\delta$  sur DE JONG F1,  $\gamma$  sur EXP et ROOTS), mais, de façon générale, nous ne pouvons pas conclure que l'intérêt individuel et l'intérêt collectif coïncident.

« Scientists need very little encouragement to adhere to the institutional norms of science. Scientists do not band together in weekly support groups to urge each other to mend their sinful ways. They need not because by and large their individual goals coincide with the institutional goals of science. Scientists adhere to the norms of science because it is in their own self-interest to do so. » (Hull, 1988c, p. 394)

Hull pense que le système de crédit est généralement bon pour la science. Et quand il fait ce genre d'énoncé, il fait aussi référence à des types de normes que notre modèle ne nous permet pas d'aborder. Il pense entre autres à la fraude (la fabrication de données, par exemple), qui est sévèrement punie, pas seulement parce qu'elle nuit à la science, mais aussi parce qu'elle nuit aux scientifiques qui ont basé leurs propres travaux sur ces résultats frauduleux et qui voient alors leurs propres contributions mises en péril. Il y a d'autres normes qui sont respectées de façon moins stricte. Le « vol » (c'est-à-dire l'appropriation de résultats d'autres scientifiques) par exemple, est plus fréquent que la fraude. Mais le vol ne fait de victime que celui qui s'est fait voler : si cela reste à l'intérieur de certaines limites, la communauté scientifique dans son ensemble n'en souffre pas trop.

Ce qu'il cherche à faire, entre autres, c'est de montrer qu'une vision de la science comme un système social où règnent la compétition et les intérêts personnels n'est pas incompatible avec le progrès scientifique. On n'a pas besoin de supposer que le comportement des scientifiques est exemplaire, sur le plan moral comme sur le plan épistémique, pour expliquer comment la science fonctionne :

« Many commentators find one or more features of science as it has existed for the past couple hundred years to be less than palatable. They find scientists too polemical, aggressive, arrogant, and elitist. Scientists are too anxious to publish so that they can scoop their competitors. They seem more intent on enhancing their own reputations than in helping humanity. According to these commentators, scientists should turn their attention from the problems that they find most interesting to those that are currently most relevant to our survival. The mechanism which I propose explains why scientists do not behave in the way that these critics think that they should behave. » (1988a, p. 154)

Cela étant dit, nous avons construit un modèle de la science basé sur les principaux éléments du mécanisme proposé par Hull et nous avons trouvé plusieurs situations où l'intérêt collectif et l'intérêt individuel ne coïncident pas. Cela ne veut pas dire qu'il faut remettre en question la théorie de Hull (ou le fonctionnement du système scientifique), mais cela laisse supposer qu'il y a des zones de tension qui subsistent. Cela semble être particulièrement le cas lorsque l'intérêt individuel pousse à un certain conservatisme (pour ce qui est de l'importance accordée au crédit et de la créativité, notamment) alors que la science dans son ensemble pourrait bénéficier d'un peu plus d'indépendance d'esprit et d'audace.

## Conclusion

Dans les sections qui précèdent, nous avons exploré le comportement du modèle sous différentes conditions. Nous avons d'abord fait varier de façon aléatoire toutes les caractéristiques individuelles des scientifiques et examiné l'effet de cette variation sur leur capacité à accumuler du crédit dans la simulation. Nous avons ensuite pris ces caractéristiques une à une et nous en avons fixé la valeur pour tous les scientifiques dans une simulation. Cela nous a permis de voir quelles valeurs des caractéristiques individuelles permettaient la meilleure performance épistémique pour la science en général. Finalement, nous avons comparé les résultats de ces deux expériences pour tenter de voir si les valeurs des caractéristiques qui sont les plus avantageuses pour les individus étaient les mêmes que celles qui sont avantageuses pour la communauté. Tout cela nous a permis de tirer certaines conclusions quant aux trois hypothèses formulées en introduction.

Nous pouvons accepter l'hypothèse  $H_1$  : dans tous les paysages épistémiques, nous avons trouvé une relation de compromis statistiquement significative entre la valeur de  $\alpha$ , qui représente le compromis entre crédit et support, et la quantité de crédit obtenue par les scientifiques (figure 6.9 à la page 194).

Notre modèle ne nous permet pas d'accepter l'hypothèse  $H_2$ , selon laquelle il y aurait aussi une relation de compromis par rapport à l'aspect inclusif de la valeur sélective

conceptuelle (figure 6.9 à la page 194). Par contre, on trouve dans tous les paysages une relation linéaire inverse significative : plus un scientifique a tendance à citer ses étudiants, moins il obtient de crédit pour lui-même (figure 6.11 à la page 197).

Les résultats sont mitigés en ce qui concerne  $H_3$  : est-ce que ce qui est bon pour la scientifique est bon pour la science ? Cela varie. Dans certains paysages et pour certaines variables, on a une corrélation positive significative entre le crédit obtenu par les scientifiques individuels pour différentes valeurs des caractéristiques individuelles et la valeur moyenne des articles lorsque la population dans son ensemble adopte ces valeurs (figure 6.25 à la page 222). Le résultat le plus clair concerne  $\gamma$  : dans les paysages où les sommets sont à pic et isolés (EXP et ROOTS), les scientifiques *et* la communauté ont avantage à minimiser l'importance du crédit dans l'évaluation des idées (ce qui s'accorde bien avec nos intuitions concernant le monde réel). Les autres cas ne nous permettent pas de tirer des conclusions bien fortes. Bref, on ne peut pas affirmer, de façon générale, que l'intérêt individuel coïncide avec l'intérêt collectif.

Nous ne nous sommes toutefois pas contentés d'évaluer les hypothèses formulées au préalable. Nous avons examiné d'autres aspects du système et tiré, au fil de notre analyse, quelques autres conclusions secondaires.

Premièrement, nous avons constaté que, malgré ses difficultés à s'échapper des minimums locaux sur certains paysages épistémiques, le système fonctionne : la valeur objective des idées produites par les scientifiques tend à augmenter avec le temps. Cela montre à la fois qu'un processus évolutionnaire tel que décrit par Hull est suffisant pour rendre compte d'un certain progrès scientifique, et que ce progrès peut avoir lieu malgré l'importance accordée au crédit par la communauté scientifique.

Nous avons aussi constaté que le modèle produisait des distributions fortement asymétriques pour ce qui est du nombre d'articles par auteurs, du nombre de citations par articles, du nombre de citations par auteur et du crédit par auteur. Nous croyons que ces distributions ressemblent suffisamment à celle qu'on retrouve dans la réalité pour ne pas mettre en péril la plausibilité de notre modèle. Ce qui est remarquable, par ailleurs, c'est que ces distributions émergent naturellement du comportement des

agents dans le système, sans que nous ayons fait d'effort particulier pour les générer.

Sur le plan de la performance sociale, notre analyse ne s'est pas limitée à examiner l'influence de  $\alpha$  et  $\beta$  : elle s'est aussi portée sur  $\gamma$  et  $\delta$ . On observe, dans la plupart des paysages, une relation de compromis entre l'importance accordée au crédit ( $\gamma$ ) et l'importance accordée à la similarité avec les idées déjà évaluées lors de la lecture des articles. Lorsqu'on a affaire à un paysage où se trouvent un ou plusieurs sommets très étroits, par contre, mieux vaut accorder le moins d'importance possible au crédit. Bref, l'importance qu'une scientifique doit apporter au crédit des autres pour favoriser son crédit personnel varie selon la structure du paysage épistémique dans lequel elle se trouve. Pour ce qui est de la créativité ( $\delta$ ), il semblerait que le plus important pour une scientifique soit de faire preuve d'un minimum de celle-ci. Mais ensuite, dans la plupart des paysages, un certain conservatisme est de mise : mieux vaut ne pas être trop créatif, sous peine d'avoir un peu plus de difficulté à faire accepter ses idées.

Nous nous sommes finalement penchés sur la performance épistémique de la communauté en fonction des caractéristiques individuelles. Comme nous l'avons mentionné ci-dessus, en abordant  $H_3$ , les valeurs qui sont optimales pour la communauté ne sont pas toujours les mêmes que celles qui le sont pour les individus. Alors, quelles sont-elles ?

Si les caractéristiques  $\alpha$  et  $\beta$  ont une certaine importance pour le crédit des scientifiques, elles ne jouent pas un rôle énorme dans la performance épistémique de la communauté. Le compromis entre crédit et support, bien réel pour les individus, ne se retrouve pas ici. Sur les paysages EXP et ROOTS, la communauté performe généralement mieux lorsque les scientifiques accordent plus d'importance au support. Et dans tous les paysages, il semblerait qu'il faille accorder au moins un peu d'importance au support. Il est difficile de dériver des normes scientifiques bien fortes de ces résultats. Le rôle de  $\beta$  dans la performance de la communauté est encore moins significatif : il semblerait que le fait de citer ses étudiants n'ait pas grand impact sur celle-ci. Ici encore, il n'y a pas de conclusion normative à tirer.

La situation est différente pour  $\gamma$  : il y a une corrélation négative assez forte entre l'importance accordée au crédit lors de l'évaluation des idées et la performance épistémique de la communauté scientifique. Lorsque les scientifiques accordent trop d'importance au crédit, la performance se dégrade. On a ici une conséquence normative : pour le bien de la science, les scientifiques ne devraient pas accorder trop d'importance au crédit. Dans certains paysages (EXP et ROOTS), cette norme coïncide avec l'intérêt individuel et devrait donc être suivie sans coercition par les scientifiques. Dans les autres paysages, par contre, ceux-ci ont plutôt avantage à baser au moins en partie leur évaluation des idées sur le crédit de l'auteur d'un article et des auteurs cités.

Finalement, nous avons découvert pour  $\delta$  qu'il est crucial que les scientifiques fassent preuve d'une certaine créativité. Mais une fois posé ce constat, il est difficile d'en dire plus : l'importance d'être créatif varie d'un paysage à l'autre. Elle se fait surtout sentir lorsque le paysage comporte des sommets de valeur objective plus limités et qu'il est important pour la communauté d'explorer l'espace épistémique le plus rapidement possible.

Voilà qui termine le résumé des conclusions que le modèle nous a permis de tirer. Notre analyse nous a permis de découvrir, parmi les caractéristiques individuelles des scientifiques, celles qui sont bonnes pour les individus et celles qui sont bonnes pour le système scientifique. Ces caractéristiques ne coïncident pas toujours complètement, mais nous n'avons pas non plus observé de conflits flagrants. Le constat initial de Hull demeure : la science fonctionne, à tout le moins dans une certaine mesure. Le système de crédit, malgré ses aspects parfois injustes, ne semble pas être un obstacle au progrès.

Cela ne veut pas dire que tout est parfait. Des interventions ciblées, des micro-optimisations, sont possibles, et notre modèle peut en suggérer. En termes de politiques scientifiques concrètes, cela se traduit surtout par la valorisation de certains comportements au moment d'attribuer des subventions, des chaires de recherche, etc. Nous avons constaté l'importance d'avoir des chercheurs qui font preuve de créativité : cela suggère qu'on aurait peut-être avantage à subventionner les projets de recherche plus originaux. Nous avons constaté que la science fonctionne mieux quand la communauté accorde un peu moins d'importance au

crédit. Il peut être difficile d'encourager les chercheurs individuels à changer leurs comportements lorsque cela va à l'encontre de leurs intérêts, mais il est possible de tenir compte de cette leçon à d'autres niveaux : peut-être faudrait-il accorder moins d'importance au CV d'un chercheur lorsque vient le temps des embauches et des subventions, et plus d'importance à la qualité de ses idées ?

Cela étant dit, notre analyse reste rudimentaire. Peut-être que ces facteurs ne sont pas indépendants les uns des autres. Peut-être que ce qui est optimal, c'est d'avoir une population hétérogène où les différents comportements suivent une distribution particulière. Comment utiliser la simulation multi-agents pour répondre à ces questions plus complexes ? La question est ouverte, et dans nombre de disciplines (économie, sociologie, biologie, etc.) les chercheurs s'attaquent au problème. C'est tout un champ de recherche qui se déploie devant nous, et nous avons la conviction que la philosophie des sciences peut à la fois en bénéficier et y jouer un rôle important.

## CONCLUSION

*Essentially, all models are wrong, but some are useful.*

– George E. P. Box (1987, p. 424)

On n'échappe pas au verdict de l'incontournable citation de Box : *all models are wrong*, le nôtre y compris. Reste à voir s'il peut prétendre être inclus dans ce sous-ensemble de modèles qui sont à tout le moins utiles. Afin de soutenir que c'est bien le cas, nous reviendrons brièvement sur chacun des trois objectifs énoncés dans l'introduction.

Notre premier objectif était d'explorer les conditions sous lesquelles la science fonctionne le mieux et, plus spécifiquement, d'examiner l'influence de certains comportements individuels sur le succès scientifique de la communauté dans son ensemble.

Notre démarche correspondait à ce que Michael Weisberg appelle la modélisation généralisée (voir section 4.4). Nous ne cherchions pas à modéliser une communauté scientifique en particulier, mais bien à examiner les mécanismes de haut niveau qui opèrent sur le processus scientifique dans son ensemble. Est-il légitime de considérer la science comme un processus généralisable ? La biologie n'est pas plus identique à la physique que la reproduction chez les kangourous n'est identique à la reproduction chez les humains. Ces deux branches de la science, comme ces deux systèmes de re-

production, ont néanmoins des propriétés en commun, et ce sont ces propriétés que nous avons tenté de modéliser.

Quand Hull parle de la science, il ne parle pas d'un système entièrement abstrait non plus. Il réfère à une entité historique : la science telle qu'elle se pratique en occident depuis la fin du XVII<sup>e</sup> siècle, avec son système de publications et de citations. C'est une des contributions importantes de Hull que d'explicitier le rôle de ce système (et de la notion de « crédit » qui relève de celui-ci) dans l'évolution des idées scientifiques. Il est possible que les différences dans le fonctionnement de ce système entre les différentes disciplines scientifiques aient un impact significatif sur l'évolution des idées, mais c'est justement en commençant par modéliser le système au niveau le plus général possible que nous aurons un point de référence à partir duquel nous pourrions examiner (lors de recherches futures) l'impact de ces différences.

Dans le chapitre 6, nous avons montré comment, dans le contexte du modèle, la distribution des caractéristiques individuelles dans la population influence la performance du système scientifique. Nous avons trouvé que, de façon générale, la science fonctionne mieux quand les scientifiques accordent une certaine importance au support dans le choix des références de leurs articles, quand ils font preuve d'une certaine dose de créativité dans la génération de nouvelles idées et quand ils accordent le moins d'importance possible au crédit lors de l'évaluation des idées.

La validité de ces conclusions dépend de la qualité du modèle lui-même. Nous croyons avoir réussi à capturer plusieurs des éléments essentiels du processus scientifique tel que décrit par Hull, mais ce n'est pas une garantie de validité. Quoi qu'il en soit, on peut toujours considérer le projet dans son ensemble comme une expérience de pensée : « Si la science fonctionne ainsi, que peut-on conclure ? » D'autre part, construire un modèle de la science qui saisisse toutes les caractéristiques du processus scientifique de façon à ce que le comportement du modèle corresponde entièrement à la réalité n'est pas une mince affaire. Un tel modèle, fut-il possible en théorie, n'existe pas à ce jour, et notre modèle n'a pas la prétention de remédier à cet état de choses. Ce qu'il fournit, par contre, c'est un cadre de travail à l'intérieur duquel on pourra espérer s'approcher de cet idéal. On pourrait dire, en ce sens, qu'il a une valeur heuristique. Il soulève, par exemple, plusieurs questions empiriques : « Comment une scientifique

décide-t-elle des articles qu'elle lira ? », « Comment une scientifique choisit-elle les idées qui valent la peine d'être transformées en articles ? », « Comment choisit-elle la liste de références de l'article qu'elle écrit ? » Les réponses données à ces questions par notre modèle ne sont, au mieux, qu'approximatives. Derrière chacune d'elles se cache un projet de recherche à part entière. Si de tels projets étaient éventuellement menés à terme, par nous ou par d'autres, il suffirait d'intégrer leurs résultats au modèle pour que la qualité de celui-ci (et des conclusions qu'on en tire) soit augmentée d'autant. Et s'il s'avérait que des réponses concurrentes soient proposées à ces questions, tant mieux : le modèle permettrait de comparer entre elles les conséquences de ces réponses.

L'autre facteur qui influence la validité de nos conclusions, c'est la qualité de l'analyse que nous avons présentée au chapitre 6. Dans un monde idéal, la taille de l'espace des paramètres exploré par nos simulations aurait été bien plus grande. Peut-être aurions-nous trouvé alors qu'un paramètre dont la valeur est restée fixe durant nos expériences a une influence déterminante sur la performance de la communauté scientifique. Étant données les ressources computationnelles à notre disposition, nous nous sommes contentés de faire varier les paramètres qui nous semblaient les plus directement reliés aux considérations évoquées par Hull ( $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$ ) et d'examiner l'effet de ces variations sur deux mesures de performance assez simples : le crédit accumulé par les individus et la valeur objective moyenne des articles publiés par la communauté. Nous croyons que cela s'est avéré suffisant pour atteindre notre objectif, mais il y aurait encore de la place pour une analyse plus poussée.

Notre second objectif visait à apporter un éclairage nouveau sur la théorie de Hull. Cette théorie a déjà fait l'objet d'une attention considérable, mais c'est la première fois qu'elle est formalisée ainsi.

Dans l'article où il met en place les bases de l'approche sémantique des théories, Patrick Suppes souligne le rôle que la modélisation peut jouer dans l'évaluation d'une théorie présentée de façon verbale :

« The emptiness and shallowness of many classical theories in the social sciences is well brought out by the attempt to formulate in any exact fashion what constitutes a model of the theory. The kind of theory which mainly consists of insightful remarks

and heuristic slogans will not be amenable to this treatment. The effort to make it exact will at the same time reveal the weakness of the theory. » (Suppes, 1960, p. 13)

Nous ne croyons pas que notre démarche ait exposé la théorie de Hull comme étant « vide » ou « superficielle », mais il nous faut reconnaître que nous avons sous-estimé la difficulté qu'il y aurait à modéliser celle-ci. Comme nous l'avons souligné à la section 4.3, la description de Hull établit un ensemble de contraintes qui définissent une famille de modèles possibles pour sa théorie. En tentant d'extraire ces contraintes et de construire un modèle qui les respecte, nous avons toutefois pu constater à quel point elles laissaient place à l'interprétation. Hull cherche à décrire « un mécanisme et sa métaphysique », mais la connexion entre le mécanisme et la métaphysique est parfois ténue.

Prenons, par exemple, les concepts de « réplicateurs » et d'« interacteurs ». Ceux-ci sont bien développés sur le plan abstrait (voir section 3.2.1), mais leur application au processus d'évolution qui a cours en science est laissée pour le moins vague. Hull évite de prendre position sur la nature exacte des réplicateurs dans le processus scientifique. Il nous dit simplement que ceux-ci forment « le contenu substantiel de la science » (voir section 3.2.2). L'applicabilité de la notion d'interacteur n'est pas mieux établie : la question à savoir ce qu'ils sont s'ils ne sont pas les scientifiques eux-mêmes n'est jamais clairement résolue (section 3.2.2 aussi). Un tel flou peut être acceptable pour une première ébauche du mécanisme scientifique, mais il se prête mal à l'élaboration d'un modèle précis.

Cette difficulté à faire le pont entre les considérations théoriques et la façon dont celles-ci s'appliquent concrètement est inhérente à l'œuvre de Hull. Pourtant, il est lui-même conscient de l'importance des exemples concrets (voir Hull, 1997) et tente d'en faire un élément central de sa démarche. *Science as a process* est un ouvrage qui comprend deux parties bien distinctes. Dans les sept premiers chapitres, Hull fait l'historique détaillé de la biologie systématique au vingtième siècle, portant à la fois attention aux relations personnelles entre les scientifiques travaillant dans ce domaine et aux idées défendues par ceux-ci. Dans les sept chapitres suivants, il développe sa théorie évolutionnaire du changement scientifique. La seconde partie est censée, a

*posteriori*, éclairer la première. Le soin de voir exactement comment elle s'y applique est toutefois laissé au lecteur, et le gouffre entre les deux est parfois vertigineux.

Avec notre modèle, nous avons tenté de combler une partie de ce gouffre : nous avons pris la description de Hull, et nous l'avons rendue un peu plus concrète. Pas suffisamment concrète pour retracer l'évolution d'une discipline scientifique particulière comme la biologie systématique, mais à tout le moins assez concrète pour reproduire certains éléments mesurables du système scientifique, comme la distribution des citations ou des articles par auteurs. C'est encore peu de choses, mais c'est un pas dans la bonne direction.

Si nous avons choisi Hull pour ce projet, ce n'est pas parce qu'il y avait urgence à se pencher sur sa théorie plutôt que sur une autre. C'est parce que nous croyions que l'attention qu'il porte à la fois aux interactions sociales et aux questions plus proprement épistémiques en faisait une théorie particulièrement adaptée à la modélisation multi-agents. La description offerte par Hull s'est toutefois révélée plus fragmentaire que nous l'avions cru de prime abord.

Hull était-il tout de même le bon choix pour ce projet ? C'est une question à laquelle il est difficile de répondre, puisque les obstacles auxquels nous avons fait face ne sont devenus apparents qu'au fil de notre démarche. Pour évaluer l'applicabilité de notre méthode à une théorie autre que celle de Hull, il faudrait refaire avec cette autre théorie le même exercice. Nous soupçonnons toutefois que les difficultés rencontrées seraient semblables, à moins que la théorie n'ait été conçue dès le départ avec l'idée de servir de base à un ou plusieurs modèles multi-agents. Si Hull avait eu cet outil à sa disposition, sa théorie aurait probablement été bien différente.

Notre dernier objectif consistait justement à démontrer l'utilité des modèles multi-agents pour la philosophie des sciences. Si on considère que nos deux premiers objectifs sont remplis, on peut dire que le troisième l'est aussi, puisque c'est un modèle multi-agents qui a permis de réaliser les deux autres.

Serait-il néanmoins possible de considérer que nous avons échoué aux deux premiers objectifs, tout en atteignant le troisième ? Oui. Il se pourrait que nous ayons pris

de mauvaises décisions pendant la modélisation ou que notre analyse des résultats du modèle comporte des failles, mais que notre démarche n'en illustre pas moins la valeur d'un tel modèle. Le simple fait d'avoir mené le projet à terme montre à tout le moins que c'est possible. Pour quelqu'un qui s'opposerait à l'utilisation des MMA en philosophie, il ne suffirait pas de critiquer les lacunes de notre modèle. Il faudrait montrer que celles-ci ne sont pas idiosyncrasiques, mais qu'elles témoignent plutôt de limitations fondamentales de la méthode.

Même si notre dernier objectif est subordonné aux deux autres, il revêt une importance particulière à nos yeux. Nous aurions pu, bien sûr, le mettre explicitement à l'avant-plan et écrire une thèse purement méthodologique : essayer de montrer pourquoi, *en théorie*, les modèles multi-agents sont une bonne façon d'étudier la science. Mais comme Cecilia Heyes le faisait remarquer à propos de l'épistémologie évolutionniste, il y a un point où la théorie n'est plus suffisante : *put up, or shut up*. Comme Hull, nous avons préféré la première alternative. Cela n'enlève rien à la valeur d'un travail théorique autour des MMA en philosophie : nous pensons que les deux entreprises doivent être menées de front et qu'elles peuvent se nourrir l'une de l'autre.

L'utilisation de la modélisation multi-agents en philosophie des sciences reste une partie relativement inexplorée du paysage épistémique. Celui qui s'y aventure ne dispose pas de beaucoup de recherches existantes sur lesquelles s'appuyer. La démarche comporte donc une part de risque. Peut-être s'agit-il d'un cul-de-sac évolutionnaire, mais nous croyons que les idées exprimées ici s'inscrivent dans une branche théorique qui portera fruit. Au chapitre zéro de leur livre sur la modélisation multi-agents, Wilensky et Rand (2015) introduisent la notion de « restructuration » : un changement dans la façon d'encoder la connaissance dans un domaine qui modifie radicalement la façon de résoudre des problèmes à l'intérieur de ce domaine. Pour illustrer ce concept, ils reprennent un exemple développé par Andrea diSessa à propos de la puissance des représentations algébriques et de leur absence dans l'œuvre de Galilée.

Dans son *Discours concernant deux sciences nouvelles*, Galilée consacre des pages et des pages *de texte* à la dérivation du théorème suivant :

Si deux mobiles sont animés d'un mouvement uniforme, le rapport de leurs vitesses sera composé du rapport des espaces parcourus et du rapport inverse des temps. (Galilei, 1638, thm. IV)

Exprimé en notation algébrique moderne, où pour un mobile  $i$ ,  $d_i$  est la distance,  $r_i$  la vitesse et  $t_i$  l'intervalle de temps, on a :

$$\frac{r_1}{r_2} = \frac{d_1 t_2}{d_2 t_1}$$

Pour montrer la puissance de la notation algébrique, diSessa refait *en quelques lignes* la démonstration qui avait pris des pages à Galilée. Il commente :

« So now we've redone a significant piece of work by one of the great geniuses of Western Science, with amazing ease. Solving problems is almost always easier after the first time around, but the difference here is almost mindboggling. [...] The fault is not with Galileo or with the education provided by his parents or with the schooling of the times. Algebra simply did not exist at that time. » (diSessa, 2001, p. 14)

Wilensky et Rand soutiennent que les modèles multi-agents peuvent avoir le même effet de « restructuration » pour penser aux systèmes complexes. Réfléchir verbalement à l'effet global des interactions locales entre multiples agents est une tâche longue, ardue, et fortement sujette à l'erreur. Ce qui exige des pages et des pages de description peut souvent être ramené à quelques lignes de code NetLogo. Cela permet non seulement une économie descriptive, mais aussi l'exploration automatique des conséquences des interactions programmées.

La science est un système complexe. Beaucoup plus complexe, sans rien vouloir enlever au génie de Galilée, que la relation entre la vitesse de deux mobiles. Et sans rien vouloir enlever non plus au génie de Hull (ni à celui de tous les philosophes qui ont réfléchi à la science), nous croyons que sa théorie se heurte aux limites du mode de représentation utilisé. Nous ne prétendons surtout pas que notre modèle de la théorie de Hull rende obsolète la formulation originale de cette dernière, puisqu'elle est beaucoup plus riche et complète que ce que nous sommes arrivés à traduire formellement. Nous prétendons toutefois avoir contribué à une restructuration de celle-ci, dans le sens entendu par Wilensky et Rand. Nous espérons avoir montré que le fait de définir formellement toutes les entités impliquées dans le processus (les scientifiques et

leurs idées), l'environnement dans lequel elles se trouvent (le paysage épistémique), et les interactions qui se produisent entre elles, puis de laisser l'ordinateur procéder à « l'expérience de pensée » qui résulte de ce modèle nous permet de dépasser les limites imposées par nos intuitions et nos capacités de raisonnement limitées. Nous n'en sommes qu'aux balbutiements de cette méthode, mais nous prédisons qu'il sera un jour aussi impensable de réfléchir aux interactions entre scientifiques sans le support de modèles multi-agents qu'il le serait maintenant de réfléchir au mouvement de particules physiques sans faire appel à la notation algébrique.

Les avancées dans le domaine des MMA (de même que dans l'analyse des réseaux sociaux, dans l'analyse statistique des bases de données bibliographiques, dans l'analyse de texte automatisée, etc.) sont en train de faire naître une « science de la science » interdisciplinaire qui vient bonifier les méthodes (utiles, mais parfois cloisonnées) de la philosophie et la sociologie des sciences traditionnelles. Mais pour qu'advienne pleinement cet état de choses, il y a encore bien du travail à accomplir. Dans les derniers paragraphes de cette thèse, nous indiquons quelques directions qui nous semblent pouvoir être fructueuses.

Il y a, d'abord, un travail conceptuel qui doit se poursuivre. Le modèle présenté ici fait intervenir des « idées » situées dans un « paysage épistémique ». Nous sommes persuadés que ces deux concepts ont un rôle important à jouer dans les simulations de la science puisqu'ils capturent l'idée que la science vise à produire des unités d'information dont elle cherche à maximiser la valeur.

La nature de ces unités peut être envisagée de façon très pluraliste : elles peuvent tout aussi bien être des propositions empiriques concrètes («  $x$  est corrélé avec  $y$  dans  $z$  ») que des propositions de nature plus philosophique (« il faut repenser le concept  $X$  »). Dans notre modèle, elles sont représentées par des vecteurs à  $n$  dimensions (avec  $n = 2$ ) et correspondent à quelque chose comme « l'idée principale contenue dans un article scientifique ». C'est suffisant pour les besoins du modèle, mais cela reste une grossière approximation. Admettant qu'il soit légitime de découper le contenu de la science en unités distinctes (pouvant servir de répliqueurs dans une théorie comme celle de Hull),

reste à voir quel est le bon<sup>85</sup> niveau de découpage et comment identifier concrètement ces unités. Si on voulait construire des modèles de la science dont les résultats sont plus directement comparables avec les données empiriques que nous possédons sur le monde scientifique, il serait bon d'opérationnaliser le concept. Idéalement, il devrait être possible de forer les bases de données d'articles scientifiques pour en extraire automatiquement les « idées ». Nos modèles multi-agents pourraient ensuite tenter de générer des patrons de propagation des idées qui correspondent à ceux que l'on aurait observés dans la littérature scientifique.

La question des paysages épistémiques est étroitement reliée à celle des unités d'information : il suffit de pouvoir situer ces unités dans un espace conceptuel<sup>86</sup> et de pouvoir attribuer à chacune d'elle une valeur épistémique pour qu'on ait, par le fait même, un paysage épistémique. Dans le modèle, nous avons utilisé des paysages épistémiques arbitraires, puisés dans la littérature sur les algorithmes d'optimisation. Mais en supposant qu'on puisse identifier les « idées » dans un domaine scientifique, il serait envisageable de reconstruire *a posteriori* le paysage épistémique de ce domaine. Cela suppose toutefois qu'on puisse attribuer à ces idées une valeur plus ou moins objective, ce qui soulève son lot de questions—questions dont les réponses risquent de varier d'un domaine à l'autre.

Ce dernier constat, selon lequel il faudra éventuellement tenir compte des différences entre les domaines scientifiques, en suggère un autre : il faudra peut-être aussi se concentrer sur des aspects plus spécifiques du processus scientifique. Influencés par Hull et par la tradition philosophique, nous avons abordé la science de façon très générale et très abstraite. Ce n'est pas, en soi, une mauvaise chose, mais ce n'est pas la seule approche possible. L'expérience acquise au fil du projet nous a permis de constater qu'il y a ici quelque chose comme un cercle herméneutique, où la compréhension du tout repose sur la compréhension des parties et la compréhension des parties présume la compréhension du tout. Il n'y a pas de voie royale vers la compréhension

---

<sup>85</sup>Nous ne pensons pas qu'il y a une seule bonne façon de découper la science en idées : le découpage choisi devra dépendre des objectifs spécifiques poursuivis par un modèle.

<sup>86</sup>Nous avons travaillé ici avec un espace où les relations entre les idées étaient des relations de proximité sémantique, mais cela n'exclut pas d'autres types de relations : conceptuelles, inférentielles, voire sociologiques.

de tels systèmes : celle-ci doit se bâtir peu à peu, en portant tour à tour notre attention sur l'ensemble du système, puis sur ses détails (et ainsi de suite). Nous décrivons à la section 5.4.4 le processus d'évaluation par les pairs utilisé dans le modèle. Notre approche est relativement simple et (nous l'espérons) cohérente avec le reste du modèle. Il s'agit pourtant d'un sujet d'étude suffisamment riche pour faire l'objet de modèles complets (par exemple : Thurner et Hanel, 2010 ; Allesina, 2012 ; Squazzoni et Gandelli, 2012 ; Paolucci et Grimaldo, 2014). Et ce n'est qu'un exemple : chacune des parties de la section 5.4 aurait pu faire l'objet d'une thèse à part entière. Dans un monde idéal, ces modèles locaux seraient conçus de façon à pouvoir être intégrés dans un modèle plus large, similaire à celui que nous avons proposé ici. Cela présente toutefois des difficultés techniques puisque les plates-formes de simulation actuelles ne sont pas conçues de façon à favoriser une telle modularité. Cela présente aussi des difficultés conceptuelles puisque les différents modèles locaux doivent à tout le moins s'entendre sur la structure de certaines entités de base. Cela n'en reste pas moins, selon nous, l'horizon vers lequel on doit tendre.

Il y a un autre avantage à porter attention aux modèles locaux : la possibilité accrue de valider ceux-ci à partir de données empiriques. Suite à notre expérience, nous serions même portés à utiliser les données comme point de départ de la modélisation : choisir un domaine scientifique en fonction de la quantité d'information qu'on possède déjà sur lui et construire un modèle qui utilise une partie de cette information comme *input* (pour calibrer les paramètres du modèle) et l'autre partie comme *output* (pour valider les résultats du modèle). Ici aussi, nous croyons qu'une alternance entre les approches peut être bénéfique : tout comme la compréhension des parties peut aider à la compréhension du tout (et *vice versa*), une approche plus empirique peut nourrir ensuite une approche plus théorique (et *vice versa*).

Le point final que nous souhaitons soulever concerne le passage du descriptif au normatif. Pour reprendre à nouveau le mot de Quine (1986, p. 664) : « normative epistemology is a branch of engineering ». Et comme les modèles des ingénieurs ont pour but de les aider à construire des choses dans le monde, ultimement, le but d'un modèle de la science devrait être de nous aider à mieux *faire la science*. Nous avons mentionné quelques interventions suggérées par notre modèle à la conclusion du chapitre 6, mais

on peut pousser plus loin cette démarche. Pour cela, il serait bon d'identifier en amont des questions de recherche dont les réponses peuvent mener à des politiques concrètes. Pour reprendre ici un exemple cité au chapitre 4, Bollen *et al.* (2014) utilisent un modèle multi-agents pour examiner les effets potentiels d'un système d'attribution de fonds de recherche où ce seraient l'ensemble des scientifiques d'un domaine qui décideraient comment répartir l'argent dans la communauté plutôt qu'un comité assigné à cette tâche par un organisme subventionnaire. Leurs résultats suggèrent qu'on obtiendrait, pour une fraction du coût et de l'effort, une répartition des fonds similaire à celle qui a cours aujourd'hui. Que Bollen *et al.* aient raison ou non quant à ce scénario particulier, la leçon demeure : les modèles multi-agents nous permettent d'explorer les conséquences de possibles changements à notre façon de mener la science. Il nous semble sage de tenter d'exploiter à fond cette capacité.

La philosophie a-t-elle encore un rôle à jouer dans de telles entreprises ? On parle, après tout, de programmer des simulations visant à reconstruire un processus social en se basant sur des données quantitatives. Ne seraient-ce pas là, plutôt, des tâches qui relèvent respectivement de l'informatique, la sociologie des sciences et la scientométrie ? C'est le cas dans une certaine mesure, mais toute démarche de modélisation exige un travail conceptuel dont la nature sied bien aux philosophes. Surtout que les concepts qui interviennent dans un domaine comme la science ont tendance à avoir des ramifications philosophiques importantes. Nous avons insisté sur le travail qui reste à faire autour de la notion d'idée ou de celle de paysage épistémique, mais tout l'appareillage conceptuel de la philosophie des sciences a potentiellement un rôle à jouer : théorie, explication, justification, observation, preuve et, surtout, vérité, risquent fort d'intervenir dans nos modèles de la science. Les philosophes, de Platon à Hull, équipés de leur seule raison et (plus récemment) d'un peu de logique formelle, ont donné à ces concepts de riches caractérisations. Nous avons maintenant à notre disposition, aux côtés du λόγος traditionnel, des langages de modélisation comme NetLogo. Ceux-ci ajoutent au paysage épistémique de la philosophie des sciences des dimensions dont nous ignorions l'existence. Reste à les explorer.

## APPENDICE A

### TABLEAUX DE CORRÉLATIONS

Les tableaux A.1 à A.4 résument les relations établies entre les différentes caractéristiques individuelles et le crédit (dans l'expérience  $E_1$ ), puis entre les caractéristiques individuelles et la valeur objective des articles publiés (expérience  $E_2$ ). Finalement, le tableau A.5 montre la relation entre le crédit moyen obtenu pour une valeur de la caractéristique (ce qui est « bon pour l'individu ») et la valeur objective moyenne des articles lorsque la caractéristique est fixée à la même valeur dans toute la population (ce qui est « bon pour la science »).

Pour chaque paysage épistémique et pour chaque caractéristique individuelle, on donne le coefficient de corrélation de Pearson ( $r$ ), l'erreur standard (SE) et la valeur- $p$  ( $p$ ). Ces valeurs sont arrondies à deux décimales pour  $r$  et SE, et à trois décimales pour  $p$ . Lorsque les résultats sont statistiquement significatifs, les valeurs sont en gras.

Toutes ces valeurs sont aussi présentées sous forme d'histogrammes : figure 6.9 à la page 194 pour A.1, figure 6.11 à la page 197 pour A.2, figure 6.14 à la page 207 pour A.3, figure 6.15 à la page 208 pour A.4, et figure 6.25 à la page 222 pour A.5.

**Tableau A.1**  
Relation de compromis entre caractéristiques individuelles et crédit ( $E_1$ )

	$\alpha$			$\beta$			$\gamma$			$\delta$		
	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$
DE JONG F1	<b>0,85</b>	<b>0,21</b>	<b>0,007</b>	-0,04	0,41	0,934	<b>0,89</b>	<b>0,19</b>	<b>0,003</b>	<b>0,90</b>	<b>0,18</b>	<b>0,002</b>
EXP	<b>0,72</b>	<b>0,28</b>	<b>0,044</b>	-0,03	0,41	0,937	-0,60	0,33	0,114	<b>0,95</b>	<b>0,13</b>	<b>0,000</b>
ROOTS	<b>0,81</b>	<b>0,24</b>	<b>0,015</b>	-0,17	0,40	0,685	0,01	0,41	0,976	<b>0,90</b>	<b>0,18</b>	<b>0,003</b>
HORNS 5 PEAKS	<b>0,83</b>	<b>0,23</b>	<b>0,010</b>	-0,45	0,36	0,263	<b>0,76</b>	<b>0,26</b>	<b>0,028</b>	<b>0,86</b>	<b>0,21</b>	<b>0,006</b>
LANGERMANN F11	<b>0,82</b>	<b>0,23</b>	<b>0,013</b>	-0,16	0,40	0,700	<b>0,89</b>	<b>0,19</b>	<b>0,003</b>	<b>0,89</b>	<b>0,19</b>	<b>0,003</b>
FLAT	<b>0,78</b>	<b>0,25</b>	<b>0,021</b>	-0,35	0,38	0,396	<b>0,91</b>	<b>0,17</b>	<b>0,002</b>	<b>0,81</b>	<b>0,24</b>	<b>0,015</b>

**Tableau A.2**  
Relation linéaire entre caractéristiques individuelles et crédit ( $E_1$ )

	$\alpha$			$\beta$			$\gamma$			$\delta$		
	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$
DE JONG F1	0,00	0,33	0,992	<b>-0,88</b>	<b>0,16</b>	<b>0,000</b>	0,45	0,30	0,163	0,60	0,27	0,052
EXP	-0,07	0,33	0,833	<b>-0,89</b>	<b>0,15</b>	<b>0,000</b>	<b>-0,83</b>	<b>0,19</b>	<b>0,002</b>	0,56	0,28	0,071
ROOTS	-0,05	0,33	0,893	<b>-0,88</b>	<b>0,16</b>	<b>0,000</b>	<b>-0,88</b>	<b>0,16</b>	<b>0,000</b>	0,47	0,29	0,146
HORNS 5 PEAKS	0,04	0,33	0,917	<b>-0,84</b>	<b>0,18</b>	<b>0,001</b>	0,57	0,27	0,068	0,53	0,28	0,095
LANGERMANN F11	0,02	0,33	0,962	<b>-0,82</b>	<b>0,19</b>	<b>0,002</b>	0,33	0,31	0,316	<b>0,65</b>	<b>0,25</b>	<b>0,030</b>
FLAT	0,07	0,33	0,847	<b>-0,89</b>	<b>0,15</b>	<b>0,000</b>	0,35	0,31	0,289	<b>0,64</b>	<b>0,26</b>	<b>0,034</b>

**Tableau A.3**  
Relation de compromis entre caractéristiques individuelle et valeur objective ( $E_2$ )

	$\alpha$			$\beta$			$\gamma$			$\delta$		
	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$
DE JONG F1	<b>0,48</b>	0,29	0,137	0,01	0,33	0,983	0,00	0,33	1,000	0,55	0,28	0,078
EXP	0,29	0,32	0,388	-0,04	0,33	0,904	-0,54	0,28	0,087	0,36	0,31	0,277
ROOTS	<b>0,55</b>	<b>0,28</b>	<b>0,082</b>	-0,36	0,31	0,280	0,00	0,33	0,995	0,52	0,29	0,105
HORNS 5 PEAKS	0,29	0,32	0,391	-0,49	0,29	0,126	-0,31	0,32	0,361	<b>0,67</b>	<b>0,25</b>	<b>0,025</b>
LANGERMANN F11	0,46	0,30	0,150	0,34	0,31	0,304	0,01	0,33	0,973	<b>0,64</b>	<b>0,26</b>	<b>0,034</b>

**Tableau A.4**  
Relation linéaire entre caractéristiques individuelle et valeur objective ( $E_2$ )

	$\alpha$			$\beta$			$\gamma$			$\delta$		
	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$
DE JONG F1	0,05	0,33	0,879	-0,55	0,28	0,078	<b>-0,95</b>	<b>0,11</b>	<b>0,000</b>	0,58	0,27	0,061
EXP	<b>0,78</b>	<b>0,21</b>	<b>0,005</b>	-0,27	0,32	0,425	<b>-0,82</b>	<b>0,19</b>	<b>0,002</b>	<b>0,84</b>	<b>0,18</b>	<b>0,001</b>
ROOTS	<b>0,63</b>	<b>0,26</b>	<b>0,039</b>	-0,42	0,30	0,204	<b>-0,95</b>	<b>0,10</b>	<b>0,000</b>	<b>0,62</b>	<b>0,26</b>	<b>0,040</b>
HORNS 5 PEAKS	0,23	0,32	0,487	-0,41	0,30	0,206	<b>-0,88</b>	<b>0,16</b>	<b>0,000</b>	0,19	0,33	0,578
LANGERMANN F11	0,39	0,31	0,241	-0,09	0,33	0,803	<b>-0,86</b>	<b>0,17</b>	<b>0,001</b>	0,26	0,32	0,446

**Tableau A.5**  
Relation linéaire entre crédit et valeur objective ( $E_1$  vs  $E_2$ )

	$\alpha$			$\beta$			$\gamma$			$\delta$		
	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$	$r$	SE	$p$
DE JONG F1	<b>0,71</b>	<b>0,24</b>	<b>0,015</b>	0,57	0,27	0,070	-0,39	0,31	0,231	<b>0,72</b>	<b>0,24</b>	<b>0,018</b>
EXP	0,02	0,33	0,943	0,29	0,32	0,393	<b>0,94</b>	<b>0,11</b>	<b>0,000</b>	0,40	0,32	0,246
ROOTS	0,50	0,29	0,117	0,44	0,30	0,172	<b>0,74</b>	<b>0,23</b>	<b>0,010</b>	-0,15	0,35	0,687
HORNS 5 PEAKS	0,51	0,29	0,107	0,47	0,29	0,143	<b>-0,69</b>	<b>0,24</b>	<b>0,020</b>	-0,16	0,35	0,666
LANGERMANN F11	0,38	0,31	0,254	0,16	0,33	0,629	-0,31	0,32	0,357	-0,22	0,34	0,535

## BIBLIOGRAPHIE

- ABT, H. A. et GARFIELD, E. (2002). Is the relationship between numbers of references and paper lengths the same for all sciences? *J. Am. Soc. Inf. Sci. Technol.*, 53(13) : 1106–1112.
- AHRWEILER, P., PYKA, A. et GILBERT, N. (2004). Simulating knowledge dynamics in innovation networks (SKIN). Rapport technique, Universitaet Augsburg, Institute for Economics.
- ALEXANDER, S. (1920). *Space, Time, and Deity : The Gifford Lectures at Glasgow, 1916-1918*. London : Macmillan and co., limited.
- ALLESINA, S. (2012). Modeling peer review : An agent-based approach. *Ideas in Ecology and Evolution*, 5(2).
- AMMAN, H. M., TESFATSION, L., JUDD, K. L., KENDRICK, D. A. et RUST, J. (2006). *Handbook of Computational Economics, Volume 2 : Agent-Based Computational Economics*. North-Holland.
- ARISTOTE (1995). *Organon : les seconds analytiques. IV. Vrin*, Paris.
- AXELROD, R. et HAMILTON, W. D. (1981). The evolution of cooperation. *Science*, 211 : 1390–96.
- AXTELL, R., AXELROD, R., EPSTEIN, J. M. et COHEN, M. D. (1996). Aligning simulation models : A case study and results. *Computational & Mathematical Organization Theory*, 1(2) : 123–141.
- BAILER-JONES, D. M. (2009). *Scientific Models in Philosophy of Science*. University of Pittsburgh Pre.
- BARNES, B. (1977). *Interests and the Growth of Knowledge*. Routledge and K. Paul.
- BARNES, B. (1981). On the 'hows' and 'whys' of cultural change (response to Woolgar). *Social Studies of Science*, 11(4) : 481–498.

- BARTO, A. G. et SUTTON, R. S. (1998). *Reinforcement Learning : An Introduction*. MIT Press.
- BECHTEL, W. (1988). New insights into the nature of science : What does Hull's evolutionary epistemology teach us? *Biology and Philosophy*, 3(2) : 157–164.
- BECHTEL, W. et ABRAHAMSEN, A. (2005). Explanation : A mechanist alternative. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2) : 421–441.
- BECHTEL, W. et RICHARDSON, R. C. (1993). *Discovering Complexity : Decomposition and Localization as Strategies in Scientific Research*. Princeton University Press.
- BENNETT, A. W. (1974). *Introduction to Computer Simulation*. West Pub. Co.
- BERSINI, H., DORIGO, M., LANGERMAN, S., SERONT, G. et GAMBARDELLA, L. (1996). Results of the first international contest on evolutionary optimisation (1st ICEO). *In Proceedings of IEEE International Conference on Evolutionary Computation, 1996*, pages 611–615.
- BETH, E. W. (1948). Analyse sémantique des théories physiques. *Synthese*, 7 : 206–207.
- BETH, E. W. (1949). Towards an Up-to-date Philosophy of the Natural Sciences. *Methodos*, 1 : 178–185.
- BETH, E. W. (1960). Semantics of Physical Theories. *Synthese*, 12 : 172–175.
- BIGBEE, A., CIOFFI-REVILLA, C. et LUKE, S. (2007). Replication of Sugarscape using MASON. *Agent-Based Approaches in Economic and Social Complex Systems IV*, 3(6) : 183–190.
- BLOOR, D. (1976). *Knowledge and Social Imagery*. Routledge & K. Paul.
- BOLLEN, J., CRANDALL, D., JUNK, D., DING, Y. et BOERNER, K. (2013). Collective allocation of science funding : From funding agencies to scientific agency. *arXiv :1304.1067 [physics]*.
- BOLLEN, J., CRANDALL, D., JUNK, D., DING, Y. et BÖRNER, K. (2014). From funding agencies to scientific agency. *EMBO reports*, 15(2) : 131–133.
- BONABEAU, E. (2002). Agent-based modeling : Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3) : 7280–7287.
- BORGES, J. L. (1944). *Ficciones*. Editorial Sur, Argentina.

- BOUDOURIDES, M. et ANTYPAS, G. (2002). A simulation of the structure of the World-Wide Web. *Sociological Research Online*, 7(1).
- BOX, G. E. P. et DRAPER, N. R. (1987). *Empirical Model-Building and Response Surfaces*. Wiley.
- BRADIE, M. (1986). Assessing evolutionary epistemology. *Biology and Philosophy*, 1(4) : 401–459.
- BROAD, W. et WADE, N. (1982). *Betrayers of the Truth*. Century.
- BROWNLEE, J. (2007). OAT : The optimization algorithm toolkit. Technical Report 20071220A, Swinburne University of Technology, Melbourne, Australia.
- BUNGE, M. (1991). A critical examination of the new sociology of science, part 1. *Philosophy of the Social Sciences*, 21(4) : 524–560.
- CAMPBELL, D. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, 67(6) : 380–400.
- CAMPBELL, D. T. (1974). Evolutionary epistemology. In SCHILPP, P. A., éditeur : *The Philosophy of Karl Popper*. Open Court Publishing.
- CAMPBELL, D. T. (1988). A general ‘selection theory’, as implemented in biological evolution and in social belief-transmission-with-modification in science. *Biology and Philosophy*, 3(2).
- CARNAP, R. (1928). *Der logische Aufbau der Welt*. Weltkreis-verlag, Berlin-Schlachtensee.
- CARNAP, R. (1966). *Philosophical Foundations of Physics : An Introduction to the Philosophy of Science*. Basic Books, inc.
- CARROLL, L. (1893). *Sylvie and Bruno Concluded*. Macmillan.
- CARTWRIGHT, N. (1999). *The Dappled World : A Study of the Boundaries of Science*. Cambridge University Press.
- CHIELENS, K. et HEYLIGHEN, F. (2005). Operationalization of meme selection criteria : Methodologies to empirically test memetic predictions. In *AISB’05 Convention : Proceedings of the Socially Inspired Computing Joint Symposium : Memetic Theory in Artificial Systems and Societies Emerging Artificial Societies Engineering with Social Metaphors*, pages 14–20.
- CHURCH, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2) : 345–363.

- CIOFFI-REVILLA, C. (2009). Simplicity and reality in computational modeling of politics. *Computational and Mathematical Organization Theory*, 15(1) : 26–46.
- COLE, P. F. (1962). A new look at reference scattering. *Journal of Documentation*, 18(2) : 58–64.
- COLLINS, H. M. (1981a). Son of seven sexes : The social destruction of a physical phenomenon. *Social Studies of Science*, 11(1) : 33–62.
- COLLINS, H. M. (1981b). Stages in the empirical programme of relativism. *Social Studies of Science*, 11(1) : 3–10.
- COLLINS, H. M. (1981c). What is TRASP ? The radical programme as a methodological imperative. *Philosophy of the Social Sciences*, 11 : 215–224.
- COLLINS, H. M. (1982). Knowledge, norms and rules in the sociology of science. *Social Studies of Science*, 12(2) : 299–309.
- COPERNICUS, N. (1543). *De Revolutionibus Orbium Coelestium*. Johannes Petreius, Nuremberg.
- COWLING, B. J., FANG, V. J., RILEY, S., MALIK PEIRIS, J. S. et LEUNG, G. M. (2009). Estimation of the serial interval of influenza. *Epidemiology (Cambridge, Mass.)*, 20(3) : 344–347.
- CRAVER, C. et TABERY, J. (2017). Mechanisms in Science. In ZALTA, E. N., éditeur : *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 édition.
- CUMMINS, R., POIRIER, P. et ROTH, M. (2004). Epistemological strata and the rules of right reason. *Synthese*, 141(3) : 287–331.
- DARWIN, C. (1859). *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. J. Murray.
- DAWKINS, R. (1976). *The Selfish Gene*. Oxford University Press.
- DE CASTRO, L. et TIMMIS, J. (2002). An artificial immune network for multimodal function optimization. In *Proceedings of the 2002 Congress on Evolutionary Computation, 2002. CEC '02*, volume 1, pages 699–704.
- DE JONG, K. A. (1975). *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. Thèse de doctorat, University of Michigan, Ann Arbor, MI, USA.
- DE SOLLA PRICE, D. J. (1963). *Little Science, Big Science*. Columbia University Press.

- DE SOLLA PRICE, D. J. (1965). Networks of scientific papers. *Science*, 149(3683) : 510–515.
- DENNETT, D. C. (1991). Real patterns. *The Journal of Philosophy*, 88(1) : 27–51.
- DENNETT, D. C. (1995). *Darwin's Dangerous Idea : Evolution and the Meanings of Life*. Simon & Schuster.
- DI SESSA, A. A. (2001). *Changing Minds : Computers, Learning, and Literacy*. MIT Press.
- EDMONDS, B. (2007). Artificial science : A simulation to study the social processes of science. In EDMONDS, B., TROITZSCH, K. G. et HERNÁNDEZ IGLESIAS, C., éditeurs : *Social Simulation : Technologies, Advances and New Discoveries*, pages 61–67. IGI Global.
- EDMONDS, B. (2011). A brief survey of some relevant philosophy of science. *Journal of Artificial Societies and Social Simulation*, 14(4) : 7.
- EPSTEIN, J. M. (2006). *Generative Social Science : Studies in Agent-Based Computational Modeling*. Princeton University Press.
- EPSTEIN, J. M. (2008). Why model ? *Journal of Artificial Societies and Social Simulation*, 11(4) : 12.
- EPSTEIN, J. M. et AXTELL, R. (1996). *Growing Artificial Societies : Social Science from the Bottom Up*. Brookings Institution Press.
- EVANS, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59(1) : 255–278.
- FERBER, J. (1997). *Les Systèmes Multi-Agents. Vers Une Intelligence Collective*. InterEditions, Paris.
- FERNÁNDEZ, M. (2009). *Models of Computation : An Introduction to Computability Theory*. Springer Science & Business Media.
- FEURZEIG, W. et PAPERT, S. (1967). Logo.
- FISHER, S. R. A. (1930). *The Genetical Theory of Natural Selection*. The Clarendon Press.
- FODOR, J. A. (1968). *Psychological Explanation : An Introduction to the Philosophy of Psychology*. Random House.
- GALILEI, G. (1638). *Discorsi e Dimostrazioni Matematiche Intorno a Due Nuove Scienze*. Elsevier.

- GARDNER, M. (1970). Mathematical games : The fantastic combinations of John Conway's new solitaire game 'Life'. *Scientific American*, 223(4) : 120–123.
- GARRETT, S. (2004). Parameter-free, adaptive clonal selection. *In Congress on Evolutionary Computation, 2004. CEC2004*, volume 1, pages 1052–1058 Vol.1.
- GIERE, R. N. (1988). *Explaining Science : A Cognitive Approach*. University of Chicago Press.
- GIERE, R. N. (1999). *Science Without Laws*. University of Chicago Press.
- GIERE, R. N. (2002). Scientific cognition as distributed cognition. *The cognitive basis of science*.
- GIERE, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, 172(2) : 269.
- GILBERT, N. (1997). A simulation of the structure of academic science. *Sociological Research Online*, 2(2).
- GILBERT, N., AHRWEILER, P. et PYKA, A. (2007). Learning in innovation networks : Some simulation experiments. *Physica A : Statistical Mechanics and its Applications*, 378(1) : 100–109.
- GILBERT, N., PYKA, A. et AHRWEILER, P. (2001). Innovation networks-a simulation approach. *Journal of Artificial Societies and Social Simulation*, 4(3) : 1–13.
- GILBERT, N. et TROITZSCH, K. G. (2005). *Simulation for the Social Scientist*. McGraw-Hill International.
- GLENNAN, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44(1) : 49–71.
- GLUSZYNSKI, T. et PETERS, V. (2005). Survey of earned doctorates : A profile of doctoral degree recipients. 2003–2004. Rapport technique 2005032, Statistiques Canada.
- GRIM, P. (2009). Threshold phenomena in epistemic networks. *In Complex Adaptive Systems and the Threshold Effect : Views from the Natural and Social Sciences*, Washington, DC.
- HACKING, I. (1983). *Representing and Intervening : Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press.
- HAMILL, L. et GILBERT, N. (2016). *Agent-Based Modelling in Economics*. John Wiley & Sons.

- HAMILTON, W. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, 7(1) : 1–16.
- HANKS, J. et McINTOSH, J. E. A. (1973). Population dynamics of the african elephant (*loxodonta africana*). *Journal of zoology*, 169(1) : 29–38.
- HARTMANN, S. (1996). The World as a Process : Simulations in the Natural and Social Sciences. In HEGSELMANN, R., éditeur : *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, pages 77–100. Kluwer, Dordrecht.
- HARZING, A.-W. et ALAKANGAS, S. (2016). Google Scholar, Scopus and the Web of Science : A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2) : 787–804.
- HEDSTRÖM, P. et YLIKOSKI, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, 36 : 49–67.
- HEGSELMANN, R. et FLACHE, A. (1998). Understanding complex social dynamics. <http://jasss.soc.surrey.ac.uk/1/3/1.html>.
- HEMPEL, C. G. (1966). *Philosophy of Natural Science*. Englewood Cliffs, N.J., Prentice-Hall.
- HEMPEL, C. G. et OPPENHEIM, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, 15(2) : 135–175.
- HEYES, C. (1988). Are scientists agents in scientific change ? *Biology and Philosophy*, 3(2) : 194–199.
- HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems : An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press.
- HORN, J. et GOLDBERG, D. E. (1994). Genetic algorithm difficulty and the modality of fitness landscapes. In *Foundations of Genetic Algorithms 3*. Citeseer.
- HULL, D. L. (1975). Central Subjects and Historical Narratives. *History and Theory*, 14(3) : 253–274.
- HULL, D. L. (1976). Are species really individuals ? *Systematic Zoology*, 25(2) : 174–191.
- HULL, D. L. (1978). A matter of individuality. *Philosophy of Science*, 45(3) : 335–360.
- HULL, D. L. (1980). Individuality and selection. *Annual Review of Ecology and Systematics*, 11(1) : 311–332.

- HULL, D. L. (1988a). A mechanism and its metaphysics : An evolutionary account of the social and conceptual development of science. *Biology and Philosophy*, 3(2) : 123–155.
- HULL, D. L. (1988b). A period of development : A response. *Biology and Philosophy*, 3(2) : 241–263.
- HULL, D. L. (1988c). *Science as a Process : An Evolutionary Account of the Social and Conceptual Development of Science*. University of Chicago Press.
- HULL, D. L. (1992). The Particular-Circumstance Model of Scientific Explanation. In NITECKI, M. H. et NITECKI, D. V., éditeurs : *History and Evolution*, pages 69–79. SUNY Press.
- HULL, D. L. (1997). That Just Don't Sound Right : A Plea for Real Examples'. In EARMAN, J. et NORTON, J., éditeurs : *The Cosmos of Science*. University of Pittsburgh Press.
- HULL, D. L. (2001). *Science and Selection : Essays on Biological Evolution and the Philosophy of Science*. Cambridge Univ Pr.
- HULL, D. L., TESSNER, P. D. et DIAMOND, A. M. (1978). Planck's principle. *Science (New York, N.Y.)*, 202(4369) : 717–723.
- HUMPHREYS, P. (1990). Computer Simulations. *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1990 : 497–506.
- HUMPHREYS, P. (2002). Computational Models. *Philosophy of Science*, 69(S3) : S1–S11.
- HUMPHREYS, P. (2004). *Extending Ourselves : Computational Science, Empiricism, and Scientific Method*. Oxford University Press.
- IWASA, Y. et SASAKI, A. (1987). Evolution of the number of sexes. *Evolution*, 41(1) : 49–65.
- JACKSON, A. (2007). A labor of love : The mathematics genealogy project. *Notices of the AMS*, 54(8) : 1002–1003.
- KAUFFMAN, S. A. (1970). Articulation of Parts Explanation in Biology and the Rational Search for Them. *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1970 : 257–272.
- KEMENY, J. G. et OPPENHEIM, P. (1956). On reduction. *Philosophical Studies*, 7(1) : 6–19.
- KITCHER, P. (1990). The division of cognitive labor. *The Journal of Philosophy*, 87(1) : 5–22.

- KLEENE, S. C. (1967). *Mathematical Logic*. Dover Publications.
- KNUTH, D. E. (1984). Literate programming. *The Computer Journal*, 27(2) : 97–111.
- KUHN, T. S. (1962). *The Structure of Scientific Revolutions*. Books LLC.
- LANGLEY, P., BRADSHAW, G. L. et SIMON, H. A. (1981). BACON.5 : The discovery of conservation laws. In *IJCAI'81*, pages 121–126.
- LANGTON, C. G. (1997). *Artificial Life : An Overview*. MIT Press.
- LABWITZ, K. (1907). Die Universalbibliothek. In *Traumkristalle : Neue Märchen*. Felber.
- LAUDAN, L. (1977). *Progress and Its Problems : Towards a Theory of Scientific Growth*. Routledge & K. Paul.
- LAUDAN, L. (1981a). The pseudo-science of science? *Philosophy of the Social Sciences*, 11(2) : 173–198.
- LAUDAN, L. (1981b). *Science and Hypothesis : Historical Essays on Scientific Methodology*. Taylor & Francis US.
- LAUDAN, L. (1984). *Science and Values : The Aims of Science and Their Role in Scientific Debate*. University of California Press.
- LAUDAN, L. (1986). Some problems facing intuitionist meta-methodologies. *Synthese*, 67(1) : 115–129.
- LEE, J.-S., FILATOVA, T., LIGMANN-ZIELINSKA, A., HASSANI-MAHMOOEI, B., STONEDAHL, F., LORSCHIED, I., VOINOV, A., POLHILL, J. G., SUN, Z. et PARKER, D. C. (2015). The Complexities of Agent-Based Modeling Output Analysis. *Journal of Artificial Societies and Social Simulation*, 18(4) : 4.
- LLOYD, E. A. (1984). A semantic approach to the structure of population genetics. *Philosophy of Science*, pages 242–264.
- LLOYD, E. A. (1994). *The Structure and Confirmation of Evolutionary Theory*. Princeton University Press.
- LOTKA, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12) : 317–323.
- LUKE, S. (2009). *Essentials of Metaheuristics*. Lulu.
- LUKE, S., CIOFFI-REVILLA, C., PANAIT, L., SULLIVAN, K. et BALAN, G. (2005). MASON : A multiagent simulation environment. *Simulation*, 81(7) : 517–527.

- MACHAMER, P. (1998). Galileo's machines, his mathematics, and his experiments. In MACHAMER, P., éditeur : *The Cambridge Companion to Galileo*, pages 27–52. Cambridge University Press, New York.
- MACHAMER, P., DARDEN, L. et CRAVER, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67(1) : 1–25.
- MAGNUS, P. D. (2007). Distributed cognition and the task of science. *Social Studies of Science*, 37(2) : 297–310.
- MATSUMOTO, M. et NISHIMURA, T. (1998). Mersenne twister : A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1) : 3–30.
- MCCLELLAND, J. L. et RUMELHART, D. E. (1987). *Parallel Distributed Processing : Psychological and Biological Models*. MIT Press.
- MERTON, R. K. (1968). The Matthew effect in science : The reward and communication systems of science are considered. *Science*, 159(3810) : 56–63.
- MEYER, M. (2011). Bibliometrics, stylized facts and the way ahead : How to build good social simulation models of science ? *Journal of Artificial Societies and Social Simulation*, 14(4) : 4.
- MILOT, E., MAYER, F. M., NUSSEY, D. H., BOISVERT, M., PELLETIER, F. et RÉALE, D. (2011). Evidence for evolution in response to natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences*.
- MITZENMACHER, M. (2003). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2) : 226–251.
- MULDOON, R. et WEISBERG, M. (2011). Robustness and idealization in models of cognitive labor. *Synthese*, 183(2) : 161–174.
- NAGEL, E. (1961). *The Structure of Science : Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World.
- NATIONAL SCIENCE FOUNDATION (2013). Report to the National Science Board on the National Science Foundation's merit review process, Fiscal Year 2012. Rapport technique NSF 13-33, Washington, DC, USA.
- NAVEH, I. et SUN, R. (2006). A cognitively based simulation of academic science. *Computational & Mathematical Organization Theory*, 12(4) : 313–337.

- NAVEH, I. et SUN, R. (2007). Cognitively based modeling of scientific productivity. In RENNARD, J.-P., éditeur : *Handbook of Research on Nature-Inspired Computing for Economics and Management*. Idea Group Inc (IGI), 141-152.
- NEELAMKAVIL, F. (1987). *Computer Simulation and Modelling*. Wiley.
- O'BRIEN, M. J., LYMAN, R. L. et SCHIFFER, M. B. (2007). *Archaeology as a Process : Processualism and Its Progeny*. University of Utah Press.
- ORD-SMITH, R. J. et STEPHENSON, J. (1975). *Computer Simulation of Continuous Systems*. CUP Archive.
- PAOLUCCI, M. et GRIMALDO, F. (2014). Mechanism change in a simulation of peer review : From junk support to elitism. *Scientometrics*, 99(3) : 663–688.
- PAPPAS, G. (2012). Internalist vs. externalist conceptions of epistemic justification. In ZALTA, E. N., éditeur : *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, summer 2012 édition.
- PAYETTE, N. (2011). For an integrated approach to agent-based modeling of science. *Journal of Artificial Societies and Social Simulation*, 14(4) : 9.
- PAYETTE, N. (2012). Agent-based models of science. In SCHARNHORST, A., BÖRNER, K. et BESSELAAR, P., éditeurs : *Models of Science Dynamics : Encounters Between Complexity Theory and Information Sciences*, Understanding Complex Systems, pages 127–157. Springer.
- PESAVENTO, U. (1995). An implementation of von Neumann's self-reproducing machine. *Artificial Life*, 2(4) : 337–354.
- PIZZITUTTI, F., MENA, C. F. et WALSH, S. J. (2014). Modelling tourism in the Galapagos islands : An agent-based model approach. *Journal of Artificial Societies and Social Simulation*, 17(1) : 14.
- PLANCK, M. (1950). *Scientific Autobiography, and Other Papers*. Williams & Norgate.
- POINCARÉ, H. (1908). *Science et méthode*. Paris, E. Flammarion.
- POLLAIUOLO, A. (1465). The battle of ten naked men.
- POPPER, K. R. (1934). *Logik Der Forschung*. Julius Springer, Vienne.
- POPPER, K. R. (1959). *The Logic of Scientific Discovery*. Routledge.

- PORTIN, P. (2002). Historical development of the concept of the gene. *The Journal of medicine and philosophy*, 27(3) : 257–286.
- POWELL, W. B. (2007). *Approximate Dynamic Programming : Solving the Curses of Dimensionality*. Wiley-Interscience.
- PUTNAM, H. (1973). Meaning and reference. *The Journal of Philosophy*, 70(19) : 699–711.
- PUTNAM, H. (1975). The meaning of “meaning”. In GUNDERSON, K., éditeur : *Language, Mind, and Knowledge*. U of Minnesota Press.
- PYKA, A., GILBERT, N. et AHRWEILER, P. (2007). Simulating knowledge-generation and distribution processes in innovation collaborations and networks. *Cybernetics and Systems : An International Journal*, 38(7) : 667.
- PYKA, A., GILBERT, N. et AHRWEILER, P. (2009). Agent-based modelling of innovation networks—the fairytale of spillover. *Innovation Networks*, pages 101–126.
- PYKA, A., KÜPPERS, G. et AHRWEILER, P. (2002). Simulating innovation networks. In PYKA, A. et KÜPPERS, G., éditeurs : *Innovation Networks : Theory and Practice*. Edward Elgar Publishing.
- QUINE, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, 60(1) : 20–43.
- QUINE, W. V. O. (1969). *Ontological Relativity and Other Essays*. Columbia University Press.
- QUINE, W. V. O. (1980). Les deux dogmes de l’empirisme. In JACOB, P., éditeur : *De Vienne à Cambridge : l’héritage du positivisme logique de 1950 à nos jours : essais de philosophie des sciences*, pages 87–112. Gallimard.
- QUINE, W. V. O. (1986). Reply to Morton White. In HAHN, L. E. et SCHILPP, P. A., éditeurs : *The Philosophy of W.V. Quine*, pages 663–665. Open Court.
- REDDY, R. (1987). Epistemology of knowledge based simulation. *Simulation*, 48(4) : 162–166.
- REICHENBACH, H. (1924). *Axiomatik Der Relativistischen Raum-Zeit-Lehre*. F. Vieweg & Sohn, Braunschweig.
- REICHENBACH, H. (1938). *Experience and Prediction : An Analysis of the Foundations and the Structure of Knowledge*. University of Chicago Press.

- REYNOLDS, C. (1987). Flocks, herds and schools : A distributed behavioral model. *SIGGRAPH '87 : Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 25–34.
- RHEINBERGER, H.-J. et MÜLLER-WILLE, S. (2010). Gene. In ZALTA, E. N., éditeur : *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information (CSLI), Stanford University, spring 2010 édition.
- ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5) : 527–535.
- ROSENBERG, A. (1982). Are there culturgens ? *Behavioral and Brain Sciences*, 5(01) : 22–24.
- SAAM, N. J. et REITER, L. (1999). Lotka's law reconsidered : The evolution of publication and citation distributions in scientific fields. *Scientometrics*, 44(2) : 135–155.
- SAKODA, J. M. (1949). *Minidoka ; an Analysis of Changing Patterns of Social Interaction*. Thèse de doctorat.
- SAKODA, J. M. (1971). The checkerboard model of social interaction. *The Journal of Mathematical Sociology*, 1(1) : 119–132.
- SALMON, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.
- SAVAGE, J. E. (1998). *Models of Computation : Exploring the Power of Computing*. Addison Wesley.
- SAWARAGI, Y., NAKAYAMA, H. et TANINO, T. (Orlando : Academic Press, 1985.). *Theory of Multiobjective Optimization*, volume 176 de *Mathematics in science and engineering*. Academic Press, Orlando, FL.
- SCHAFFNER, K. F. (1967). Approaches to reduction. *Philosophy of science*, pages 137–147.
- SCHARNHORST, A., BORNER, K. et BESSELAAR, P. (2012). *Models of Science Dynamics : Encounters between Complexity Theory and Information Sciences*. Springer.
- SHELLING, T. C. (1969). Models of segregation. *The American Economic Review*, pages 488–493.
- SHELLING, T. C. (1971a). Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2) : 143–186.

- SCHELLING, T. C. (1971b). On the ecology of micromotives. *The public interest*, 25(1971) : 59–98.
- SHIFMAN, L. et THELWALL, M. (2009). Assessing global diffusion with Web memetics : The spread and evolution of a popular joke. *Journal of the American Society for Information Science and Technology*, 60(12) : 2567–2576.
- SIMON, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4) : 425–440.
- SIMON, H. A. (1957). *Models of Man : Social and Rational; Mathematical Essays on Rational Human Behavior in Society Setting*. Wiley.
- SIMON, H. A. (1969). *The Sciences of the Artificial*. M.I.T. Press.
- SIMONTON, D. (1999). Creativity as blind variation and selective retention : Is the creative process Darwinian? *Psychological Inquiry*, 10(4) : 309–328.
- SIMONTON, D. K. (2003). Scientific creativity as constrained stochastic behavior : The integration of product, person, and process perspectives. *Psychological bulletin*, 129(4) : 475.
- SIPSER, M. (2006). *Introduction to the Theory of Computation*. Thomson Course Technology.
- SOLOMON, M. (1994). Multivariate models of scientific change. *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1994 : 287–297.
- SQUAZZONI, F. et GANDELLI, C. (2012). Opening the Black-Box of Peer Review : An Agent-Based Model of Scientist Behaviour. *Journal of Artificial Societies and Social Simulation*, 16(2) : 3.
- STEUER, E. (1986). *Multiple Criteria Optimization : Theory, Computations, and Application*. John Wiley & Sons, Inc, New York.
- STREVVENS, M. (2003). The role of the priority rule in science. *The Journal of Philosophy*, 100(2) : 55–79.
- SUN, R. (2002). *Duality of the Mind : A Bottom-up Approach toward Cognition*. Routledge.
- SUN, R. (2003). A tutorial on CLARION 5.0. Technical report, Cognitive Science Department, Rensselaer Polytechnic Institute.

- SUN, R. (2006). The CLARION cognitive architecture : Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction : From cognitive modeling to social simulation*, pages 79–99.
- SUN, R. et NAVEH, I. (2009). Cognitive simulation of academic science. *In Proceedings of the International Joint Conference on Neural Networks*, pages 3011–3017.
- SUPPE, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. University of Illinois Press.
- SUPPE, F. (2000). Understanding Scientific Theories : An Assessment of Developments, 1969-1998. *Philosophy of Science*, 67(3) : 115.
- SUPPES, P. (1957). *Introduction to Logic*. Van Nostrand, Princeton.
- SUPPES, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. Technical Report 33, Office Of Naval Research.
- SUPPES, P. (1962). Models of data. *In NAGEL, E., SUPPES, P. et TARSKY, A., éditeurs : Logic, Methodology, and Philosophy of Science : Proceedings of the 1960 International Congress*, pages 252–261. Stanford University Press, Stanford.
- SUPPES, P. (1965). Logics appropriate to empirical theories. *In ADDISON, J., HENKIN, L. et TARSKY, A., éditeurs : The Theory of Models*, Amsterdam. North Holland Publishing Company.
- SUPPES, P. (1967). What is a scientific theory? *In MORGENBESSER, S., éditeur : Philosophy of Science Today*, pages 55–67. Basic Books, New York.
- TARSKY, A. (1953). A general method in proofs of undecidability. *In TARSKY, A., MOSTOWSKI, A. et ROBINSON, R., éditeurs : Undecidable Theories*. North Holland Publishing Co., Amsterdam.
- TEN BROEKE, G., VAN VOORN, G. et LIGTENBERG, A. (2016). Which sensitivity analysis method should i use for my agent-based model? *Journal of Artificial Societies & Social Simulation*, 19(1).
- TENNANT, N. (1988). Theories, concepts and rationality in an evolutionary account of science. *Biology and Philosophy*, 3(2) : 224–231.
- THAGARD, P. (1993a). *Computational Philosophy of Science*. MIT Press.
- THAGARD, P. (1993b). Societies of minds : Science as distributed computing. *Studies in History and Philosophy of Science*, 24 : 49–67.

- THAGARD, P. et HOLYOAK, K. (1985). Discovering the wave theory of sound : Inductive inference in the context of problem solving. *In Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 610–612.
- THIELE, J. C., KURTH, W. et GRIMM, V. (2014). Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models : A Cookbook Using NetLogo and R. *Journal of Artificial Societies and Social Simulation*, 17(3) : 11.
- THURNER, S. et HANEL, R. (2010). Peer-review in a world with rational scientists : Toward selection of the average. *arXiv :1008.4324 [physics]*.
- TOULMIN, S. (1972). *Human Understanding : The Collective Use and Evolution of Concepts*. Princeton University Press, Princeton N.J.
- TURING, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London mathematical society*, 2(1) : 230–265.
- VAN FRAASSEN, B. C. (1970). On the extension of Beth's semantics of physical theories. *Philosophy of Science*.
- VAN FRAASSEN, B. C. (1980a). *The Scientific Image*. Oxford University Press, New York.
- VAN FRAASSEN, B. C. (1980b). Theory Construction and Experiment : An Empiricist View. *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1980 : 663–678.
- VON NEUMANN, J. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press.
- VON SCHILCHER, F. et TENNANT, N. (1984). *Philosophy, Evolution and Nature*. Routledge & Kegan Paul.
- WALLACE, M. L., LARIVIÈRE, V. et GINGRAS, Y. (2009). Modeling a century of citation distributions. *Journal of Informetrics*, 3(4) : 296–303.
- WEISBERG, M. (2013). *Simulation and Similarity : Using Models to Understand the World*. Oxford University Press, New York.
- WEISBERG, M. et MULDOON, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2).
- WILENSKY, U. (1997). Wolf Sheep Predation model.
- WILENSKY, U. (1999). NetLogo.

- WILENSKY, U. et RAND, W. (2015). *An Introduction to Agent-Based Modeling : Modeling Natural, Social and Engineered Complex Systems with NetLogo*. MIT Press, Cambridge, MA.
- WILLIAMS, B. (1981). Internal and external reasons. *Moral luck*, pages 101–113.
- WILLIAMS, G. C. (1996). *Adaptation and Natural Selection : A Critique of Some Current Evolutionary Thought*. Princeton University Press.
- WIMSATT, W. C. (1974). Complexity and organization. *In PSA 1972*, pages 67–86. Springer.
- WIMSATT, W. C. (1976). Reductionism, levels of organization, and the mind-body problem. *In Consciousness and the Brain*, pages 205–267. Springer.
- WINSBERG, E. (1999). Sanctioning models : The epistemology of simulation. *Science in context*, 12(02) : 275–292.
- WINSBERG, E. (2001). Simulations, models, and theories : Complex physical systems and their representations. *Philosophy of Science*, pages S442–S454.
- WINSBERG, E. (2003). Simulated Experiments : Methodology for a Virtual World. *Philosophy of Science*, 70 : 105–125.
- WINSBERG, E. (2009). Computer Simulation and the Philosophy of Science. *Philosophy Compass*, 4(5) : 835–845.
- WINSBERG, E. (2015). Computer Simulations in Science. *In ZALTA, E. N., éditeur : The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2015 édition.
- WINTHER, R. (2006). Parts and theories in compositional biology. *Biology and Philosophy*, 21 : 471–499.
- WINTHER, R. G. (2016). The Structure of Scientific Theories. *In ZALTA, E. N., éditeur : The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 édition.
- WOOLDRIDGE, M. et JENNINGS, N. R. (1995). Intelligent agents : Theory and practice. *The knowledge engineering review*, 10(02) : 115–152.
- WRAY, K. B. (2000). Invisible hands and the success of science. *Philosophy of Science*, 67(1) : 163–175.
- WRIGHT, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*, 56(645) : 330–338.

WRIGHT, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics*.

YLIKOSKI, P. (1995). The invisible hand and science. *Science Studies*, 8(2) : 32–43.

ZOLLMAN, K. J. S. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5) : 574–587.