UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

ÉCOLE DOCTORALE DE PHILOSOPHIE

IHPST

Thèse pour l'obtention du grade de
Docteur en philosophie de l'Université Paris 1 Panthéon-Sorbonne
présentée et soutenue publiquement par

David WASZEK

le 16 décembre 2018

LES REPRÉSENTATIONS EN MATHÉMATIQUES

Directeur de thèse :

M. Marco PANZA	Directeur de recherches au CNRS (IHPST) Université Paris 1 Panthéon-Sorbonne

Composition du jury :

M. Jeremy AVIGAD	Professor, Carnegie Mellon University
Mme Hourya BENIS-SINACEUR	Directrice de recherches émérite au CNRS (IHPST)
M. Jeremy GRAY	Emeritus Professor, Open University *(rapporteur)*
M. Brice HALIMI	Maître de conf. hab., Univ. Paris Ouest *(rapporteur)*
M. David RABOUIN	Chargé de recherches au CNRS (SPHERE)
M. Jean-Jacques SZCZECINIARZ	Professeur, Université Paris Diderot

Table des chapitres

Remerciements

En quatre ans de doctorat, j'ai reçu des impulsions et des idées de toutes parts, certaines fortes et manifestes, d'autres à peine perceptibles et pourtant cruciales, d'autres encore, à n'en pas douter, dont je n'ai même pas conscience. Sans prétendre pouvoir rendre son dû à chacun, j'aimerais exprimer toute ma gratitude à ceux qui suivent.

Les encouragements de Marco Panza et son soutien permanent m'ont été indispensables pour entamer cette thèse, et plus encore pour la terminer alors qu'il me semblait en être encore au travail préliminaire. Avec Annalisa Coliva, il m'a aussi accueilli pour un séjour fructueux de quelques mois au département de philosophie de l'université de Californie à Irvine.

David Rabouin m'a consacré beaucoup de temps, et le travail qui suit est profondément marqué par nos discussions. Jeremy Avigad et Ken Manders m'ont accueilli à Pittsburgh pour quelques jours ; mes échanges avec eux m'ont beaucoup apporté, là-bas comme ici ou encore à Toulouse lors de leurs visites. Jeremy Heis a pris le temps de plusieurs discussions approfondies lors de mon séjour à Irvine. Gianni (Juan-Luis) Gastaldi a été généreux de son temps et sa perspective éloignée de ma formation m'a été précieuse. M'ont également été précieuses mes discussions, sur un continent ou sur un autre, avec Valeria Giardino, Penelope Maddy, Yacin Hamami, Sébastien Gandon, Brice Halimi, Dirk Schlimm, John Mumma, Sun-Joo Shin, Alberto Naibo et Maël Pégny. Je voudrais également remercier Andy Arana et Sébastien Maronne pour leurs commentaires lors de ma soutenance de mi-parcours. Stéphane Dugowson a eu l'amabilité de me faire parvenir un exemplaire de sa thèse de doctorat. Isidora Stojanovic m'a aidé à m'orienter dans les labyrinthes quelque peu oubliés de la sémantique des situations, et Gabriel Scherer m'a fourni quelques indications utiles sur l'informatique théorique d'aujourd'hui.

Un simple commentaire à l'issue d'une intervention, et même un unique conseil de lecture, peut s'avérer décisif. C'est ainsi que lors de ma toute première présentation, quelques mois à peine après le début de ma thèse, Jean Gayon m'a recommandé une lecture marquante qui a joué un rôle crucial dans la genèse de ma seconde partie. Pour leurs commentaires ou leurs conseils, je voudrais également remercier Gerhard Heinzmann, Jessica Carter, Silvia de Toffoli, Danielle Macbeth, Jean-Michel Salanskis, Hourya Benis-Sinaceur, Olivier Rey, Roi Wagner, Gabriel Sandu et Atsushi Shimojima.

Ce travail doit beaucoup aux écoles d'été, journées d'étude et colloques auxquels j'ai eu la chance de participer. Jean-Michel Salanskis m'a permis d'assister à une école de philosophie des sciences organisée à Nancy par Léna Soler, où j'ai beaucoup appris. Sébastien Maronne (souvent avec l'aide de David Rabouin et Brice Halimi) a organisé d'innombrables événements à Toulouse et à Saint-Flour dont j'ai énormément profité. Valeria Giardino m'a permis de me rendre à plusieurs conférences très intéressantes pour moi qu'elle a organisées à Nancy. Les éditions successives du *French PhilMath Workshop* ont aussi été formatrices ; je voudrais remercier les organisateurs de celle de Marseille où j'ai eu l'occasion d'intervenir, en particulier Paola Cantù. L'équipe des rencontres franco-mexicaines d'histoire et de philosophie des sciences, dont Carlos Alvarez et Vincent Jullien, m'ont offert un cadre sympathique et fructueux pour présenter mon travail (David Rabouin y est aussi pour beaucoup). Je garde également un excellent souvenir de la *Novembertagung* de 2016 à Sønderborg, et remercie Line Andersen pour ses efforts d'organisation. Dans deux domaines très différents, l'école d'été de la SoPhA de 2014 (organisée par Filipe Drapeau Vieira Contim, Pascal Ludwig et Max Kistler) et celle de méthodes quantitatives en histoire des mathématiques de 2018 (organisée par Catherine Goldstein, Alexandre Guilbaud et Irène Passeron) m'ont permis de compléter ma formation ; au début de ma thèse, la première rencontre, à Helsinki, du projet avorté de réseau doctoral européen en philosophie a aussi été formatrice (qu'Alberto Naibo soit ici remercié pour y avoir conduit notre petite délégation parisienne).

Pour ces divers déplacements, le soutien financier de l'école doctorale de philosophie de Paris 1 m'a souvent été précieux, ainsi que celui de l'IHPST et, à l'occasion, des projets de recherche de Marco Panza. Une aide substantielle du collège des écoles doctorales m'a aussi permis un séjour plus long à l'étranger ; je voudrais ici remercier Pierre-Marie Morel d'avoir soutenu ma candidature.

Plusieurs collègues et amis ont pris le temps de relire et de discuter avec moi des passages plus ou moins longs de ma thèse, dont encore une fois Gianni Gastaldi et David Rabouin, ainsi que Henri Salha, Nicolas Michel et Julien Gusthiot (sans parler des relectures extrêmement minutieuses de Marco Panza). Toutes les coquilles et aberrations qui subsistent sont évidemment de mon fait, comme l'étaient les nombreuses erreurs qu'ils m'ont permis d'éliminer.

Les discussions avec les jeunes chercheurs et chercheuses de l'IHPST et de SPHERE m'ont aussi été très profitables. Je voudrais en particulier remercier João Cortese pour avoir organisé un groupe de travail d'histoire des mathématiques à SPHERE où j'ai énormément appris, Marina Imocrante qui a rassemblé plus récemment un groupe de travail très productif à l'IHPST, ainsi que Pascal Bertin, Nicolas Michel, Mourtaza Chopra, Julien Gusthiot, Eleonora Sammarchi, Matias Osta et, dans les dernières semaines de ma thèse, CHANG Chun Hao. Emmylou Haffner finissait lorsque j'ai commencé, mais heureusement, cela ne m'a pas empêché de bénéficier de son soutien amical tout au long de mon doctorat et de sa compagnie rassurante dans les séminaires parisiens comme aux colloques les plus exotiques. À Paris 1, la bonne humeur de Marie Michon, Delphine Olivier, Vincent Legeay, Louis Guerpillon, Alexis Anne-Braun, Antoine Naïk, Ludmilla Lorrain, Victor Lefèvre, Sophia Rousseau-Mermans, Nicola Bertoldi et bien d'autres m'a rendu la thèse plus heureuse. À Bordeaux, la compagnie de Benjamin Le Roux et de Marcin Krasnodębski m'a été d'un soutien inestimable, de même que celle d'un ami d'études qui se reconnaîtra ; et, dans la triste ville d'Irvine, celle de Francesca Biagioli.

Ce doctorat a aussi été, pour moi, une période d'enseignement. Pendant mes trois ans de monitorat à Paris 1, la bienveillance de Pierre Wagner m'a permis d'organiser mon service de manière à réserver du temps pour la recherche. Je suis reconnaissant à Pascal Duris et Jérôme Pierrel, de l'université de Bordeaux, de m'avoir fait confiance et accordé un poste d'ATER pour ma quatrième année. Les derniers mois de ce travail, enfin, n'auraient pas été possibles sans le poste que m'a accordé le département SPH de l'université Paris Diderot : je tiens à remercier Jean-Jacques Szczeciniarz et Nadine de Courtenay, ainsi que mes collègues ATERs et moniteurs (Nicolas Michel, Youna Tonnerre, Jean Sanchez, Laura Barbier et Smadar Bustan), pour m'avoir permis de consacrer le début de l'automne à achever ce manuscrit.

Cette thèse n'est guère que le produit de l'éducation que j'ai reçue. Je ne peux finir sans dire ma reconnaissance aux institutions qui m'ont formé. Y étudier a été un immense privilège, dans tous les sens du terme. Les classes préparatoires du lycée Louis-le-Grand m'ont offert une formation scientifique d'une qualité exceptionnelle, à laquelle cette thèse doit beaucoup. L'École Normale Supérieure m'a offert un environnement intellectuel unique et une liberté d'étude qui est peut-être sans équivalent. Je suis particulièrement reconnaissant à Marwan Rashed, qui m'a encouragé tout au long de ma scolarité, et à tous ceux qui ont regardé avec indulgence mon permanent vagabondage disciplinaire, parmi lesquels Pierre Pansu et François Loeser au département de mathématiques, Mathias Girel, Jean-Pascal Anfray, Francis Wolff et Sophie Roux au département de philosophie, ainsi que François

Recanati et Daniel Andler. L'équipe de la licence de lettres classiques de Paris 4 m'a accueilli sans hésiter lorsque j'ai frappé à leur porte. L'UFR de philosophie de Paris 1 a bien voulu m'admettre en master, et Jocelyn Benoist y a dirigé mon premier mémoire en consacrant des heures de son temps à discuter mes idées pourtant très confuses. Je voudrais aussi remercier ceux qui ont dirigé mes mémoires de mathématiques : Olivier Benoist à l'ENS, Ian Grojnowski à l'université de Cambridge et Benoit Fresse à l'université de Lille. Un autre privilège crucial que m'a accordé l'ENS a été la possibilité de séjours à l'étranger, à Cambridge d'abord puis à Boston College, dont le département de langues romanes m'a accueilli pour une année très productive pendant laquelle j'ai conçu le projet de cette thèse. Pour remonter plus loin encore dans le temps, je voudrais remercier l'association France-IOI, qui m'a enseigné à programmer avant que je connaisse quoi que ce soit aux mathématiques et à laquelle je n'ai jamais convenablement exprimé ma gratitude. Leur enseignement s'est avéré utile à cette thèse de la façon la plus inattendue.

Le soutien que mes proches ont apporté à ce travail est bien peu de choses par rapport à tout ce que je leur dois, mais il m'a été indispensable. Merci à celui qui, pendant que j'écrivais des jours sans m'arrêter, est venu sur mon balcon discuter de pays lointains ; à celle qui est venue, chargée d'un plat de fèves, me parler de pays proches ; et à celui qui m'a accueilli pour de nombreux festins en me racontant les merveilles de pays imaginaires. Merci à celle dont je vis entouré des belles lettres et cartes postables. Merci à celui qui, en soutenant l'an dernier, m'a montré le chemin. Merci à celle qui, malgré ses pérégrinations permanentes entre quatre ou cinq pays, a toujours trouvé du temps pour me réjouir de sa bonne humeur inépuisable. Merci à tous ceux avec qui j'ai voyagé pendant ces années. Merci à celle dont les messages et souvent la compagnie, presque tous les jours de ces quatre années, m'ont été le bonheur le plus sûr. Merci à ma famille de Nouvelle-Angleterre, dont la maison est un port chaleureux ; à ma grand-mère, chez qui j'ai écrit bien des pages de cette thèse ; et à ma sœur, qui comprend souvent mes inquiétudes. Merci, enfin, à mes parents.

Introduction

Pour résoudre un problème de mathématiques ou comprendre une démonstration, une figure bien choisie est parfois d'un grand secours. Ce fait souvent remarqué peut être vu comme un cas particulier d'un phénomène plus général. Utiliser une figure plutôt que des phrases, reformuler un problème sous la forme d'une équation, employer telles notations plutôt que telles autres : dans tous ces cas, en un sens, on ne fait que *représenter sous une nouvelle forme* ce qu'on sait déjà, et pourtant, cela peut permettre d'avancer. Plus largement, on peut lire de ce point de vue certaines des transformations les plus importantes de l'histoire des mathématiques, et y voir des évolutions non seulement conceptuelles, mais aussi *représentationnelles.* Pour prendre des exemples célèbres, songeons au développement de l'algèbre, du calcul différentiel ou même plus récemment de la théorie des catégories : il est tentant d'y voir (entre autres !) l'élaboration de modes de représentation permettant de traiter efficacement et uniformément toute une famille de problèmes.

Ce travail part d'une perplexité vis-à-vis de ces phénomènes. Comment est-il possible que l'on puisse progresser par le simple fait de représenter autrement ce que l'on sait déjà ? Comment comprendre les concepts de représentation et d'information, souvent invoqués de manière vague pour parler de ce problème ? Ces concepts sont-ils même adaptés ou faut-il les abandonner ? Voilà, en première approximation, les questions qui orientent cette thèse.

Ces questions sont largement orthogonales aux problèmes centraux de la philosophie des mathématiques du XX^{e} siècle. Pour caricaturer, prenons des théorèmes mathématiques, par exemple des résultats élémentaires de géométrie et d'arithmétique. Les questions classiques sont de savoir de quoi ils parlent (d'objets abstraits ? de propriétés du monde physique ? de rien du tout ? etc.) et d'où viennent leur validité. Ce travail, au contraire, concerne non pas *ce dont* on parle mais *comment* on en parle. Dans un manuel, peut-être notre théorème de géométrie est-il présenté à l'aide d'une figure et notre théorème d'arithmétique à l'aide d'une formule algébrique. Qu'apportent la figure et la formule ? Est-il plus facile de comprendre le théorème, de le démontrer ou de découvrir sa preuve avec elles que sans elles, et si oui pourquoi ? C'est ce genre de questions qui m'intéressent ici ; si elles sont indéniablement importantes pour bien décrire les mathématiques comme activité humaine, elles n'en concernent pourtant pas directement l'ontologie ou les fondements. De ce point de vue, mon travail, comme beaucoup des travaux sur lesquels je m'appuierai, peut être rattaché à la « philosophie de la pratique mathématique » telle que la définit Mancosu (2008a).

Les problèmes que j'ai présentés dépassent bien sûr le cadre des mathématiques : toutes les sciences, sans doute, utilisent aujourd'hui avec profit des graphes, des diagrammes, des illustrations diverses ou encore des formules. Je m'appuierai d'ailleurs régulièrement sur des travaux de philosophie des sciences et de philosophie générale. Les mathématiques sont néanmoins l'objet principal de ma réflexion et ma source essentielle d'exemples. Au-delà du fait contingent que l'histoire et la philosophie des mathématiques sont mon ancrage disciplinaire, ce choix, je crois, se justifie. Nombre des situations scientifiques où, par exemple, un diagramme bien choisi s'avère précieux peuvent être transposées dans un contexte purement mathématique qui en préserve les traits essentiels ; et la rigueur et la précision des mathématiques permettent d'étudier les contributions de différentes représentation de plus près qu'il ne serait aisé de le faire ailleurs. Si les mathématiques sont loin d'épuiser le problème, il me semble donc qu'elles sont un bon laboratoire pour l'aborder.

État des lieux

Le point de départ de cette thèse est un ensemble de travaux récents qui soulignent que les mathématiques ne sont pas qu'affaire de raisonnements pouvant être conduits dans la langue naturelle, mais impliquent de manière essentielle des supports de raisonnement : écritures symboliques, figures, diagrammes [1], etc. À vrai dire, l'existence même de cette littérature mérite une clarification préalable. Ne semblerait-t-il pas évident à tout collégien que les mathématiques sont une affaire de triangles, de cercles, et d'équations avec x et y ? En fait, cette littérature ne peut se comprendre que relativement à une vision des mathématiques bien particulière, issue des bouleversements subis par la discipline autour du tournant du vingtième siècle, selon laquelle toutes les mathématiques peuvent être re-

1. Dans ce travail, j'ai choisi d'employer le terme de « figure » dans le cas de la géométrie élémentaire et de réserver celui de « diagramme » pour des représentations plus abstraites (diagrammes de Venn, diagrammes commutatifs, etc.) ; c'est, d'expérience, l'usage qui semble le plus naturel à ceux qui ne sont pas familiers de la littérature récente de langue anglaise, où le terme « *diagram* » tend à prendre un sens très large. Cela me conduit à traduire l'anglais « *diagram* » tantôt par « figure », tantôt par « diagramme ».

construites formellement d'une manière qui efface toute différence représentationnelle. Je reviendrai plus bas sur ce contexte mathématique et philosophique très particulier.

Contre cette vision des mathématiques se sont développés, depuis un quart de siècle, divers travaux autour du « raisonnement diagrammatique » ou de la « visualisation ». Ces deux expressions correspondent à deux courants de recherche différents issus d'un point de départ commun : l'un comme l'autre critiquent la mise à l'écart des figures ou diagrammes dans l'analyse des mathématiques.

Le premier courant, fait d'études logiques, a été initié autour de 1990 par les logiciens Jon Barwise et John Etchemendy [2]. Leur but était de montrer que l'on peut raisonner de manière parfaitement rigoureuse avec des diagrammes. Pour ce faire, ils ont développé de premiers exemples de systèmes formels diagrammatiques ainsi qu'hétérogènes, c'est-à-dire employant à la fois des diagrammes et des formules logiques. Leur travail a été poursuivi par leurs élèves, entre autres Sun-Joo Shin, qui a développé des systèmes formels pour plusieurs types de diagrammes logiques, et Atsushi Shimojima, qui prolonge leur réflexion sur la distinction entre diagrammes et formules et sur l'apport spécifique des diagrammes au raisonnement. Ces études, quoiqu'elles soient très stimulantes et aient révélé des spécificités logiques précises de certains raisonnements diagrammatiques, sont toutefois limitées à des exemples très simples.

Le second courant, au contraire, étudie des mathématiques plus avancées. Il est pour l'essentiel fondé sur des études de cas historiques ou portant sur la pratique de mathématiciens contemporains [3]. Les philosophes de ce courant rejettent volontiers les investigations logiques précédentes : selon les termes de Paolo Mancosu, celles-ci seraient « au mieux insuffisantes, au pire malavisées [4] », parce qu'en reproduisant les ambitions uniquement justificationnelles de la tradition logique, elles seraient à leur tour incapables de rendre compte de l'usage réel des diagrammes à des fins de découverte ou de compréhension. Les études de cas que proposent ces auteurs, qui sont souvent très descriptives, ont révélé de nombreux phénomènes intéressants, par exemple sur les différents usages heuristiques des diagrammes, sur les cas dans lesquels ils peuvent induire en erreur ou sur les mécanismes cognitifs de visualisation. On peut rattacher à cette seconde tendance une ligne d'investigation distincte, venue de l'histoire des mathématiques, qui met l'accent sur l'importance des figures comme outils de preuve, en particulier dans des traditions mathématiques non

2. Le collectif Allwein et Barwise (1996) rassemble l'essentiel de leurs premiers travaux et de ceux de leurs élèves. Je reviendrai longuement sur ces travaux dans la troisième partie de cette thèse.

3. Pour un aperçu récent de ces travaux, voir Giaquinto 2015 ou Giardino et Hamami 2019.

4. Mancosu 2005, p. 25.

occidentales [5]. Peu de travaux ont su combiner la rigueur du premier courant et l'attention soigneuse aux mathématiques du second. Une exception, qui aura pour cette raison une importance toute particulière dans ce travail, est le travail de Ken Manders sur la géométrie d'Euclide [6].

L'enjeu principal de tous ces travaux est de pointer la distance qui existe entre ce qui compte comme une preuve ou une justification dans diverses pratiques mathématiques (celles d'aujourd'hui comme à travers l'histoire) et l'idéal philosophique de démonstration formelle, et au fond de remettre cet idéal en cause. L'important, pour moi, est ailleurs. Ces recherches soulèvent, quoique parfois de manière seulement implicite, la question de savoir comment comprendre l'idée de différences représentationnelles.

Certains auteurs, au premier rang desquels Ken Manders [7], ont tenté d'inscrire la question des figures et diagrammes en mathématiques dans le cadre d'une réflexion plus générale sur les formes de représentation, qui couvre aussi, par exemple, la différence, en géométrie, entre l'emploi d'équations à la manière de Descartes et l'emploi de figures à la manière d'Euclide. Cette idée est reprise, et élargie, par Brown (2008, chap. 6) ou encore Colyvan (2012, chap. 8), mais ces textes ne sont que programmatiques. De manière plus récente, Dirk Schlimm (2018 ; 2008) s'est intéressé à la question des différences notationnelles, mais le sujet reste peu exploré.

Délimitation du problème

Ce qui m'intéresse, ce sont les situations où les représentations particulières que l'on utilise font une différence. C'est un problème large et vague. Pour commencer, j'ai donc dû faire des choix. En fait de « représentations », je me limite principalement à ce qu'on pourrait appeler les supports matériels utilisés au cours de l'activité mathématique : figures, graphes, diagrammes en tous genres, formules, etc.

Ce choix, comme toute délimitation préalable d'un objet d'étude, reflète au fond des hypothèses de travail et des décisions méthodologiques, que l'on peut contester. Par exemple, on peut vouloir traiter aussi les images mentales du mathématicien ; j'ai toutefois préféré éviter le détour vers la psychologie et m'en tenir à des représentations plus concrètes, que l'on peut étudier directement. Dans une autre direction, il peut être tentant d'élargir la problématique à ce qu'on appelle « représentations » en un sens technique en mathéma-

5. Voir en particulier Chemla 2012 ; Høyrup 1990, 2002.

6. Écrit en 1995, l'article en question a longtemps circulé informellement avant d'être publié (Manders 2008b). J'y reviendrai à plusieurs reprises dans ce travail, en particulier aux chapitres 8, 11 et 12.

7. Voir en particulier Manders 1999, 2012, 2017.

tiques [8], dans l'espoir d'utiliser des outils strictement mathématiques (par exemple issus de la théorie des catégories) pour éclairer le problème dans son ensemble. C'est un projet légitime, mais là encore, j'ai préféré m'en tenir à des objets concrets plutôt que de monter en abstraction.

La langue naturelle pose un problème spécifique. Beaucoup des auteurs que j'étudierai comparent par exemple des figures et des « représentations sous forme de phrases » (souvent d'ailleurs sans distinguer entre une phrase en langue naturelle et sa formalisation dans quelque langage logique), et je les suivrai sur ce terrain. En revanche, j'ai laissé de côté les problèmes qui relèvent plus directement de la philosophie du langage, par exemple celui du changement conceptuel. Cela ne va pas de soi : changer de terminologie ou reconceptualiser, n'est-ce pas aussi une manière qui peut être fructueuse de re-représenter ? Je n'ai pas non plus voulu prendre comme *point de départ* les distinctions philosophiques usuelles pour la langue naturelle, comme celle qui est souvent faite, à la suite de Frege, entre signification ou contenu et référence [9]. C'est pourtant aussi une façon de penser la possibilité de parler de la même chose de différentes manières, et c'est peut-être la bonne façon d'aborder certaines des différences représentationnelles que l'on rencontre en mathématiques. Nous aurons amplement l'occasion d'y revenir dans la troisième partie. Mais quitte à revenir à la philosophie du langage au cours de mon investigation, j'ai préféré partir de travaux plus spécifiques sur les figures ou sur les notations plutôt que d'un cadre général de philosophie du langage dans lequel j'aurais tenté de faire entrer aussi, à toute force, des phénomènes non linguistiques.

---

8. La théorie des représentations de groupes, par exemple, peut être vue comme une manière d'étudier des objets très abstraits (par exemple des groupes définis uniquement par générateurs et relations) à travers leurs « représentations », qui sont ici des sous-groupes de transformations linéaires d'un espace vectoriel. Selon cette acception du terme, les « représentations » d'un groupe sont elles-mêmes des objets mathématiques. En ce qui me concerne, j'emploie ici le terme de représentation en un sens plus matériel : l'écriture d'un groupe par générateurs et relations, par exemple $G = \langle a \mid a^3 \rangle$ est déjà une représentation en ce sens ; l'écriture

$$\begin{array}{rccc} \rho : & G & \to & GL_2(\mathbf{C}) \\ & a & \mapsto & \begin{pmatrix} 1 & 0 \\ 0 & j \end{pmatrix} \end{array}$$

est encore une représentation, toujours en ce sens matériel ; en revanche, le morphisme ρ lui-même en tant qu'objet abstrait, qui est précisément ce qu'en mathématiques on appelle une représentation du groupe G, ne fait pas partie des représentations qui m'intéressent ici. Dans le même esprit, on pourrait également penser à des cas apparentés, mais pour lesquels le terme technique de « représentation » n'est pas utilisé en mathématiques, par exemple la relation entre une variété abstraite et ses plongements dans $\mathbf{R}^n$. Mais là encore, du point de vue auquel je me place, tant la variété que ses plongements sont des objets abstraits et tombent en-dehors du champ de ce travail.

9. L'emploi de ces termes soulève deux problèmes de traduction superposés : il y a la traduction des termes allemands de Frege et la traduction des termes anglais couramment utilisés pour des notions apparentées, quoique pas identiques. Sur tous ces points, cf. *infra*, section 9.1.a), en part. notes 6 et 9 p. 228.

Mise en perspective du problème

J'ai dit plus haut que les questions qui guident ce travail concernaient non pas ce dont on parle en mathématiques, mais comment on en parle. J'ai ensuite rattaché à cette problématique des questions diverses : l'emploi d'écritures symboliques plutôt que phrases, le choix des notations, l'utilisation de figures. Or poser le problème ainsi risque de soulever une objection immédiate. J'ai fait comme s'il était évident qu'utiliser des figures ou des symboles plutôt que des phrases était une simple affaire de commodité ou d'efficacité. Mais peut-on vraiment mettre ainsi entre parenthèses toute question ontologique ? Ne pourrait-on pas défendre que la construction de figures est indispensable au raisonnement géométrique, à l'instar de Kant ; ou encore que les écritures symboliques sont consubstantielles à l'analyse ?

a) Mathématiques modernes et représentations

En fait, la forme que prend le problème, ici comme dans une bonne part de la littérature récente, ne peut se comprendre que par référence aux bouleversements subis par les mathématiques dans les décennies autour de 1900, que Jeremy Gray a appelé leur « transformation moderniste [10] ». Il n'est pas question de rendre ici justice à cette histoire complexe. Concernant la place des représentations, il faudrait sans doute étudier la convergence de plusieurs facteurs. Il y a d'abord des développements, tout au long du XIX^{e} siècle (entre autre l'étude systématique des séries de Fourier et l'émergence de la géométrie projective et des géométries non-euclidiennes) qui suggèrent l'inadéquation ou l'insuffisance de l'intuition géométrique [11]. Ceux-ci conduisent à reformuler l'analyse puis la géométrie sous une forme qui permet de se passer de figures géométriques, et débouchent sur une conception normative de ce qu'est une démonstration qui exclut celles-ci [12]. D'un autre côté, il faudrait tenir compte de la diffusion progressive d'un idéal méthodologique visant à éliminer autant que possible, au profit d'une approche « conceptuelle », les calculs dépendant de notations particulières [13].

10. Voir Gray 2008 ou, pour une introduction plus rapide, Gray 2009. D'un point de vue philosophique, Benis-Sinaceur 2002 offre une synthèse de certaines de transformations générales les plus frappantes de la période.

11. Pour ce qui est du lien entre évolutions de l'analyse et statut de l'intuition géométrique, cf. Volkert 1986. Voir Grattan-Guinness [1980] 2000 pour une introduction à l'histoire de l'analyse au XIX^{e} siècle et Gray 2007 pour une introduction à celle de la géométrie à la même période.

12. Pour quelques références, voir Mancosu 2005, p. 13-15 ou Giaquinto 2015, section 2.

13. Pour un panorama des multiples manières dont cette histoire peut être racontée, voir Haffner 2014, p. 39-77.

Ce qui est sûr, c'est qu'au milieu du XX^{e} siècle, le statut des représentations (figures géométriques d'abord, mais aussi écritures algébriques) a profondément changé. Il est assez largement admis qu'*en principe*, toutes les mathématiques ou presque pourraient être reformulées dans un cadre uniforme, par exemple dans le langage de la théorie des ensembles, ce qui éliminerait toutes les différences représentationnelles. De manière associée quoique plus diffuse, il devient fréquent de considérer que la discipline porte sur des objets abstraits définissables axiomatiquement et n'ayant de lien intrinsèque avec aucune forme particulière de représentation. Le collectif Bourbaki, qui a une profonde influence sur les mathématiques pures dans les décennies d'après-guerre, est emblématique de ce type de vision des mathématiques [14]. Non seulement leurs ouvrages évitent largement le recours aux figures, mais les calculs aussi y sont réduits à un rôle subalterne et si possible éliminés : Claude Chevalley, l'un des fondateurs du groupe, dira rétrospectivement que « tout ce qui est purement le résultat d'un calcul n'était pas considéré par nous comme une bonne démonstration [15] ».

Dans ce contexte, il est clair que l'emploi de telle représentation plutôt que telle autre n'est plus dicté par la nature du sujet : la géométrie n'est pas *essentiellement* affaire de figures, ni l'analyse d'expressions algébriques. Le choix des représentations en est alors réduit à une question d'efficacité voire de goût ; en témoigne le fait qu'André Weil, lui aussi membre fondateur de Bourbaki, puisse écrire comme une évidence que « [l]e contenu d'un théorème ne change guère, qu'on l'exprime en mots ou en formules ; le choix, comme nous le savons tous, est pour l'essentiel une affaire de goût et de style [16] ».

b) Évolutions plus récentes

La discipline a changé, cependant, depuis que le collectif Bourbaki était à son apogée, et sa vision des mathématiques est aujourd'hui moins dominante [17]. Certes, pour l'essentiel, le cadre précédent persiste, en particulier l'idée que les mathématiques pourraient être formulées dans un langage uniforme. Mais on reconnaît plus volontiers l'importance pratique cruciale des figures ou des calculs concrets conduits dans des notations particulières, et plus seulement des présentations axiomatiques. Il me semble que si le problème des représentations suscite aujourd'hui tant d'intérêt, ce n'est pas seulement en raison d'un tournant plus

14. Sur le collectif de mathématiciens Bourbaki et sa vision des mathématiques, voir par exemple Corry 1997.

15. Guedj 1981, p. 24.

16. « The content of a theorem does not change greatly, whether it is expressed in words or in formulas; the choice, as we all know, is mostly a matter of taste and of style. » (Weil 1978, p. 92).

17. Cf. par exemple Dahan-Dalmedico 2001.

général vers la pratique ou d'une certaine lassitude, de la part des philosophes, quant aux problèmes de fondements ; c'est aussi en réaction à ces évolutions récentes de la discipline.

Le cas des figures est particulièrement net : Paolo Mancosu a même pu parler d'un « retour du visuel [18] ». Certes, rien n'interdisait à un mathématicien « moderniste », pour qui les figures n'ont pas leur place dans une démonstration rigoureuse, d'en faire un large usage à des fins heuristiques ou pédagogiques, et certains l'ont beaucoup fait, à commencer par David Hilbert [19]. Il est cependant indéniable que les nouvelles mathématiques, en particulier chez Bourbaki [20], ont été associées à un style d'exposition qui vise avant tout la perfection de l'organisation logique du matériau et ne laisse que peu de place à d'éventuelles figures. Les choses ont toutefois changé. Le souci pédagogique a certainement joué un rôle dans le reflux progressif du style bourbachique dans les manuels [21], de même que la grande utilité des visualisations dans certains domaines (comme l'étude des équations différentielles ainsi que la géométrie et la topologie, en particulier en basse dimension). D'autre part, de nouveaux types de visualisation permis par les ordinateurs ont émergé et pris de l'importance [22].

Dans le même registre, on pourrait parler d'un « retour du calculatoire », qui remet

18. Mancosu 2005, p. 17.

19. Les figures sont omniprésentes dans son ouvrage sur les fondements de la géométrie (Hilbert 1899), dont Jeremy Gray, se faisant l'écho d'un sentiment fréquemment partagé, dit que « s'il y a une seule œuvre exemplaire qui a fait entrer le modernisme, c'est sans doute [celle-là] » (Gray 2008, p. 5). La position de Hilbert transparaît clairement de notes prises lors d'un de ses cours de 1902 sur les fondements de la géométrie : « Man könnte auch von *Figuren* absehen, wir werden dies aber nicht thun, sondern häufig Figuren gebrauchen, uns aber *niemals auf sie verlassen.* » (« On pourrait aussi renoncer aux figures, mais nous ne le ferons pas ; au contraire, nous utiliserons souvent des figures, sans cependant *jamais nous reposer sur elles.* », Hilbert 2004, p. 541.) Il co-écrit plus tard un ouvrage entier de *Géométrie intuitive* (Hilbert et Cohn-Vossen 1932), richement illustré.

20. L'emploi d'un style d'exposition qui évite les figures dépasse cependant le cas spécifique de Bourbaki. Johansen, Misfeldt et Pallavicini ont par exemple procédé à un inventaire des diagrammes et figures publiés dans un grand nombre d'articles représentatifs de la revue *Annals of Mathematics* entre 1885 à 2005, et ont constaté une diminution drastique du nombre de figures dans la première moitié du XX^e^ siècle (conférence donnée le 19 juin 2018 dans le cadre du colloque *Diagrams 2018* à Édimbourg ; la version publiée de leur article ne contient pas le détail des chiffres : Johansen, Misfeldt et Pallavicini 2018). Il faut néanmoins distinguer le style d'exposition écrit des autres pratiques. Même parmi les membres de Bourbaki, il est possible que les pratiques d'enseignement ou de communication informelle de la recherche aient été très éloignées de ce que suggèrent leurs manuels. Plus largement, il serait utile de mener une étude fine des attitudes envers les figures en fonction des différents contextes locaux ou nationaux, des différentes sous-disciplines et des différents types de pratiques (enseignement, conférences, publication d'articles ou de manuels). Je n'ai pas connaissance d'un tel travail et m'en tiens donc ici à des remarques générales et sans doute superficielles.

21. Claude Chevalley lui-même regrettera rétrospectivement la méthode d'exposition choisie par Bourbaki. « Il y avait quelque chose que nous refoulions tous, » déclare-t-il en 1981, « c'était que tout ce que nous écrivions était inutilisable dans l'enseignement » ; il ajoute que « [c]e à quoi je n'adhère certainement plus c'est à la méthode d'exposition. Elle est inutilisable pour un débutant par exemple. » (Guedj 1981, p. 24).

22. Les travaux présentés (de manière accessible et richement illustrée) par Mumford, Series et Wright 2002 en sont un bon exemple.

sur le devant de la scène les questions de notations et d'organisation des calculs [23]. Là encore, même au milieu du XXe siècle, le dédain de Bourbaki pour le calcul était certainement loin d'être largement partagé, en particulier dans les branches plus appliquées des mathématiques ; avec la perte d'influence de leur idée de la hiérarchie des sous-disciplines, le paysage devait nécessairement changer [24]. Mais il y a plus. L'aide des outils informatiques de calcul symbolique [25] rend aujourd'hui accessibles computationnellement des questions qui, auparavant, n'étaient guère abordables que *via* des théorèmes généraux, ce qui remet en cause le primat des méthodes abstraites même en mathématiques pures.

Objectifs et plan

La première ambition de ce travail est de compléter et de mettre en perspective la littérature existante en étudiant un exemple de changement notationnel. Comme nous l'avons vu, de nombreuses études récentes se sont penchées sur le rôle des diagrammes ou des figures dans diverses pratiques mathématiques, mais très peu, en revanche, a encore été écrit sur les notations. Ma première partie vise à combler cette lacune. Elle est consacrée à l'étude d'une innovation notationnelle apparemment minime introduite par Leibniz et de son impact sur quelques investigations ultérieures de Leibniz et Jean Bernoulli.

Armé de cette étude de cas et de la littérature existante, je me tourne ensuite vers le problème général posé plus haut : comment est-il possible que l'on puisse progresser par le simple fait de représenter autrement ce que l'on sait déjà ?

Ce que je propose ici n'est pas un travail systématique qui chercherait à proposer une théorie générale des représentations à partir de premiers principes. Mon approche est exploratoire. Je suis parti de différentes manières dont les différences représentationnelles sont pensées dans la littérature récente de philosophie des mathématiques. J'ai tenté de les comprendre, de les clarifier, de les confronter à mon étude de cas et à d'autres, et finalement de voir comment elles pouvaient permettre de répondre au problème de départ.

Le problème difficile est de comprendre l'idée assez obscure que l'on puisse représenter la même chose sous différentes formes. Pour avancer, je me tourne, dans ma deuxième partie, vers le travail de Herbert Simon, une figure centrale des sciences cognitives naissantes dans la seconde moitié du XXe siècle. En se fondant sur la notion informatique de *structure*

23. Sur ce point, je suis redevable à une journée d'étude sur le calcul organisée à l'université Paris 7 par Jean-Jacques Szczeciniarz le 12 octobre 2015, et en particulier à l'intervention de Joël Merker.

24. Cf. encore une fois Dahan-Dalmedico 2001.

25. Pour un exemple célèbre de succès précoce de ces nouvelles méthodes, voir Petkovšek, Wilf et Zeilberger 1996 (pour une introduction rapide, cf. par ex. Tefera 2010).

de données, celui-ci propose une définition claire de ce que c'est, pour deux représentations, que d'être « informationnellement équivalentes » et explique qu'elles peuvent néanmoins être « computationnellement différentes ». Nous verrons que cette formulation – qui jouit ces dernières années d'une certaine popularité en philosophie des sciences – est assez adaptée à l'étude des notations et éclaire certains des aspects de mon étude de cas. Elle a cependant des limites. D'une part, le modèle informatique sous-jacent, à y bien regarder, n'est pas très clair et ne s'applique qu'imparfaitement aux représentations concrètes qui font l'objet de ce travail. D'autre part, il s'avère que beaucoup des phénomènes présentés plus haut ne relèvent pas d'une authentique équivalence informationnelle au sens de Simon : texte et figure, dans la géométrie d'Euclide, ne sont pas équivalents mais portent plutôt des informations de types différents ; chez Descartes, la configuration géométrique analysée n'est pas équivalente à l'équation qu'il en tire. Les avantages centraux du changement notationnel de Leibniz de la première partie y échappe également. Faire cette distinction est déjà un progrès. Mais comment analyser ces cas-là ? Il faudrait pouvoir comparer des représentations qui parlent de la même chose mais portent des informations différentes, un peu comme ce que permet la notion de contenu en philosophie du langage. Le cadre de Simon ne le permet pas, parce qu'il s'en tient à l'*équivalence* informationnelle sans donner d'outils pour décrire les informations qu'une représentation donnée porte.

Pour ce faire, je consacre ma troisième partie au travail des logiciens Jon Barwise, John Etchemendy et Sun-Joo Shin. Ceux-ci ont, dans les années 1990, tenté d'étendre aux diagrammes les concepts logiques de syntaxe et de sémantique. Cette généralisation donne le moyen d'appliquer les outils usuels des sémantiques en termes de modèles à des représentations de types différents, par exemple des diagrammes et des formules algébriques. Cette approche en termes de modèles permet de progresser sur de nombreux cas, mais en un sens, elle ne permet pas de rendre compte de certains des avantages les plus fréquents de l'usage de diagrammes. Pour prendre un exemple très simple, considérons trois nombres x, y et z tels que $x < y$ et $z < y$. Pour conclure que $x < z$, j'ai besoin de faire une inférence, certes simple, fondée sur la transitivité de la relation $<$. Si au contraire je représente x, y et z sur une droite horizontale (selon la convention que si $a < b$, alors a est à gauche de b), dès lors que je représente $x < y$ et $z < y$, je ne peux pas ne pas représenter aussi $x < z$: en quelque sorte, cette représentation par une figure me donne cette conclusion « gratuitement ». Pourtant, les modèles du diagramme et de la conjonction $x < y \wedge y < z$ sont les mêmes et font disparaître cette différence. Barwise et Etchemendy étaient conscients de cette difficulté : dans leurs termes, une sémantique en termes de modèles ne distingue pas bien les informations « explicitement » représentées, comme l'est $x < z$ sur la figure, et celles qui ne sont qu'implicites, comme l'est $x < z$ dans le cas des formules. Un aspect peu connu de leurs travaux est qu'ils ont tenté de développer une nouvelle sémantique que l'on pourrait dire *informationnelle* pour remédier à ce défaut ; des idées apparentées ont ensuite été appliquées par leur élève Atsushi Shimojima pour rendre compte de manière plus fine du raisonnement diagrammatique. Toujours dans la troisième partie, je consacre deux chapitres à cet aspect moins connu de leurs travaux. J'y confronte pour finir mon étude de cas, qui est apparemment difficile à traiter dans ce cadre.

Les deux parties précédentes sont fondées sur un présupposé commun : ce que j'ai appelé jusqu'ici représentations (les figures, formules, etc. utilisées dans la pratique des mathématiques) portent des informations bien délimitées, en un sens à analyser. Peut-être ce présupposé lui-même est-il une fausse piste. Si la plupart des travaux que j'ai discutés jusqu'ici emploient bien, par défaut, le terme de « représentations », il faut toutefois signaler que certains des auteurs cités essaient d'éliminer le terme, lui préférant par exemples ceux d'« artefacts » (Manders 1999), ou, à la suite de Hutchins (2005), d'« ancrages matériels » (Johansen 2010, 2014 ; Rabouin 2017). La quatrième partie explore quelques raisons que l'on peut avoir d'abandonner ainsi le terme.

Première partie

Étude de cas : l'« analogie des puissances et des différences »

Sommaire de la première partie

Introduction : une étude de cas historique, pourquoi, comment ?

En 1695, en l'espace de quelques mois, Leibniz et Jean Bernoulli font une série de découvertes remarquables. Les formules et les notations y jouent un rôle crucial, paradigmatique de certains des phénomènes que je voudrais approfondir dans ce travail : c'est sans doute en inspectant visuellement des formules complexes que Leibniz est conduit à l'une de ses idées les plus fructueuses ; et c'est en manipulant des formules de manière apparemment insensée – en torturant véritablement les notations admises – que Bernoulli découvre des régularités inattendues. Pour mon propos, cet épisode est donc exemplaire. J'y puiserai tout au long de ce travail, pour y confronter les positions philosophiques que je discuterai.

Cette partie est donc une *étude de cas* : quoiqu'elle puisse, par ailleurs, avoir un intérêt historique propre, son but ici est d'enrichir une discussion philosophique plus générale. Méthodologiquement, mon travail s'inscrit ainsi dans la démarche mise en avant par Paolo Mancosu dans son volume collectif *The Philosophy of Mathematical Practice* (2008b) [26], qui s'est considérablement développée en philosophie des mathématiques depuis une quinzaine d'années.

Cependant, même dans le cadre de cette littérature-là, mener une étude historique détaillée ne va pas de soi : si l'on feuillette le collectif dirigé par Paolo Mancosu, on se rend vite compte que les exemples discutés sont le plus souvent contemporains ou du moins traités d'un point de vue contemporain, et sont discutés de manière bien moins approfondie. Pourquoi avoir choisi cette démarche-là, et pourquoi cet épisode historique en particulier ?

Certaines de mes motivations sont contingentes. Tout d'abord, cette étude de cas vient d'un travail historique plus large que j'ai entrepris, mais qui dépasse le cadre de cette thèse (il vise à réévaluer ce qu'on appelle parfois le « calcul des opérations » au XVIIIe siècle

26. Voir en particulier son introduction : Mancosu 2008a.

et son rôle dans le développement de l'algèbre [27]). Une autre raison de me tourner vers l'histoire, qui a un goût de paradoxe, est que je tenais à étudier des mathématiques en train de se faire. Les études que l'on trouve par exemple dans Mancosu (2008b) tendent à discuter des mathématiques bien établies, venues de manuels. Cela se comprend. Pour les mathématiques contemporaines, il faut en effet choisir : ou bien l'on s'en tient à des questions relativement élémentaires, ou bien l'on rend les détails incompréhensibles à la plupart des lecteurs. L'épisode que j'ai choisi permet d'éviter ce dilemme. D'un côté, il concerne d'authentiques découvertes en train de se faire : comme Leibniz et Bernoulli explorent ensemble, il est facile de suivre pas à pas leurs progrès au fil des lettres qu'ils échangent ; on peut y observer leurs idées se développer comme au ralenti, voir comment l'une conduit à l'autre. D'un autre côté, le sujet reste plutôt facile d'accès, la plupart des spécialistes en histoire et philosophie des mathématiques ayant une certaine familiarité avec le XVIIᵉ siècle.

Mon choix de l'histoire, et de cet épisode, répond toutefois aussi à des motivations plus profondes. Si nous voulons parler des mathématiques plus généralement et pas seulement des mathématiques contemporaines, il me semble essentiel d'élargir notre champ de vision ; d'ailleurs, notre regard sur les mathématiques contemporaines elles-mêmes ne peut qu'être biaisé si l'on ne se donne pas les moyens de voir ce qui leur est spécifique [28]. Les articles de Ken Manders – les seuls du volume de Mancosu qui aient une authentique dimension historique – sont à cet égard exemplaires pour moi ; plus largement, un récent volume collectif sur la notion de preuve dirigé par Karine Chemla (2012) montre bien ce qu'un tel élargissement peut apporter à notre réflexion. Pour mon travail, qui cherche à aborder les mathématiques au niveau le plus concret de la manipulation de représentations, il m'a semblé utile de choisir une étude de cas qui comporte des usages complexes de formules, mais offre toutefois un peu de recul vis-à-vis de la distinction qui est faite aujourd'hui entre celles-ci et les diagrammes.

Dans ce contexte, se tourner vers l'âge classique et plus spécifiquement vers Leibniz peut sembler naturel. L'exemple précis que j'ai choisi est assez célèbre, et est régulièrement cité comme un cas de découverte guidée par les notations : il apparaît souvent dans la littérature française sur Leibniz depuis Couturat [29], et il a même récemment fait son apparition

27. Pour se faire une idée du sujet, voir par exemple Koppelman 1971 ou Lubet 2010.

28. Une telle démarche historique n'a pas besoin de présupposer qu'il existe quelque chose d'universel et d'atemporel comme « les mathématiques », dont la nature et les frontières n'auraient jamais varié. Bien sûr, des jugements initiaux de similarité nous guident nécessairement quand il s'agit de choisir quels cas étudier pour éclairer nos mathématiques, mais on peut laisser de côté toute question normative.

29. Celui-ci y consacre en effet quelques paragraphes en appendice de sa *Logique de Leibniz* : Couturat 1901, p. 497-498.

dans la littérature anglo-saxonne de philosophie des mathématiques [30].

Travailler sur Leibniz pose cependant un problème historiographique considérable. Dans le cas de Leibniz, la question des notations en mathématiques (ou, comme il le dirait lui-même, celle de la « caractéristique ») est loin d'être mineure ou négligée comme elle l'est parfois chez d'autres auteurs. Tout au contraire, les commentateurs y attachent souvent une importance exorbitante et abordent le sujet chargés d'un bagage philosophique très lourd, marqué entre autres par les discussions explicites qu'y consacre Leibniz lui-même ; ils s'intéressent alors aux échos, dans les mathématiques de Leibniz, de sa philosophie et des principes généraux de sa méthode [31]. Dans le cas de l'épisode que j'ai choisi, l'effet de ce parti-pris de lecture est clair. Cela conduit souvent à y voir une illustration triomphale de la Caractéristique, de l'*ars combinatoria* ou de l'*ars inveniendi* leibniziens.

J'ai tenté d'aborder mes textes indépendamment de tout ce bagage, de m'en tenir aux mathématiques de l'épisode, et de mettre en avant, plutôt que le récit univoque d'une découverte permise par une notation bien choisie, la multiplicité des pistes de réflexion mal ajustées avec lesquelles Leibniz compose. Cela me permet, je l'espère, d'arriver à une évaluation un peu plus nuancée du rôle que jouent les notations leibniziennes.

Cet objectif – analyser le rôle exact que jouent les notations dans mon épisode – m'impose une étude plus détaillée qu'il est usuel dans la littérature récente. Il est trop facile, en effet, d'attribuer aux notations un rôle important voire moteur dans une découverte si l'on omet de regarder de près pourquoi elles ont été introduites. Colyvan (2012), par exemple, écrit que de bonnes notations peuvent « faciliter » ou « suggérer » de « nouvelles mathématiques », et même qu'elles peuvent « inciter l'utilisateur à garder en tête des distinctions que l'inventeur de la notation n'avait peut-être même pas remarquées [32] ». Pourtant, il est légitime de soupçonner que les notations qu'il prend comme exemples ont été introduites précisément en raison des bénéfices qu'il leur attribue [33] ; faut-il plutôt dire qu'une bonne

30. Cf. Larvor 2010, p. 199, Johansen 2010, p. 151 et Johansen 2014, p. 93. Larvor le tient probablement de Serfati 2005.

31. Cette attitude est nette chez des auteurs aussi différents que Couturat 1901, Serres 1968 ou Granger 1981, par exemple. Le cas de Serfati 2008 est particulièrement frappant : quoiqu'il s'intéresse avant tout à l'histoire des mathématiques et non à la philosophie de Leibniz, il présente toute l'invention du calcul différentiel comme le résultat d'une pure démarche d'exploration combinatoire.

32. « [G]ood notation can facilitate new mathematics […]. [N]ote how once again good notation might suggest new mathematics. […] Good notation, it seems, prompts the user to keep track of distinctions the inventor of the notation may not have even noticed. » (Colyvan 2012, p. 158-159).

33. Il discute d'abord notre écriture usuelle des nombres (c'est-à-dire l'écriture positionnelle en chiffres arabes), et suggère un scénario fantaisiste dans lequel une société dépourvue des « concepts d'infinité ou de récursion » aurait néanmoins introduit cette écriture et aurait compris l'infinité des nombres entiers grâce à elle (p. 156–157). Il se tourne ensuite vers les notations de Leibniz et de Newton, mais sous une forme anachronique qui rend difficile d'évaluer historiquement ses jugements (p. 158–159). Il discute enfin la représentation de

notation peut permettre à l'étudiant de redécouvrir certaines des raisons pour lesquelles elle a été adoptée ? Pour comprendre de près si et comment les nouvelles notations de Leibniz aident Leibniz et Bernoulli dans leurs explorations, un examen plus minutieux est nécessaire. J'essaierai en particulier de comprendre comment les notations de Leibniz lui permettent de progresser au-delà des motivations qu'il pouvait avoir à les introduire. En d'autres termes, j'essaie de tracer une voie moyenne entre d'un côté la glorification excessive des notations que l'on trouve parfois dans la littérature, de l'autre l'idée qu'elles sont une affaire entièrement superficielle. Comme l'écrit Jeremy Gray :

> Historians of mathematics sometimes used to claim that without this or that piece of notation some idea was unthinkable. This does not seem entirely satisfactory, especially to mathematicians who know very well that new symbols are easy to devise, but the claim has some merit. [34]

Les deux premiers chapitres contiennent respectivement une présentation du contexte et une étude de détail de mes textes. Ils sont strictement historiques. Je ne veux pas dire par là que je m'interdis absolument toute reconstruction ou toute utilisation de mathématiques plus récentes lorsqu'elles me semblent authentiquement éclairantes ; je veux dire plutôt que je cherche à comprendre ce que Leibniz et Bernoulli pensaient faire, dans leurs propres termes. Le passage au troisième chapitre marque en revanche une rupture méthodologique nette. J'y mène une analyse destinée à préparer mon usage de l'étude de cas dans la suite de la thèse. Quoique j'y fasse peut-être autant d'histoire que dans les deux précédents, mon objectif y est différent : je confronte l'étude de cas à une problématique venue d'ailleurs, celle du rôle des notations.

---

surfaces compactes par des polygones dont certains côtés sont supposés identifiés (cf. p. 160–163 ; pour des introductions à cette forme de représentation, voir par exemple de Toffoli et Giardino 2015, p. 317-318 ou Hatcher 2002, p. 5-8). Peut-être a-t-il raison d'écrire qu'elle pourrait conduire à l'idée de surfaces comme la bouteille de Klein qui ne peuvent être plongées dans l'espace à trois dimensions. Il faut cependant souligner qu'elle a été introduite à un moment où ce genre d'objet était déjà parfaitement compris (sur ce point, voir Eckes et Giardino 2018, p. 125-128).

34. « Il fut un temps où les historiens des mathématiques affirmaient que sans telle ou telle notation, telle idée était impensable. Cela ne semble pas entièrement satisfaisant, surtout pour les mathématiciens, qui savent très bien qu'il est facile d'introduire de nouveaux symboles ; pourtant, l'affirmation n'est pas sans mérite. » (Gray 2001, p. 59)

Note bibliographique

a) Sources primaires

Mon étude s'appuie avant tout sur des lettres, pour la plupart publiées par Carl Immanuel Gerhardt dans son édition classique des œuvres mathématiques de Leibniz, que j'abrégerai désormais LMS (une liste des abréviations avec références détaillées se trouve en tête de la bibliographie). La publication récente des volumes pertinents de la nouvelle et plus rigoureuse édition de référence de Leibniz, dite édition des académies (= LAA), m'a permis de vérifier les transcriptions de LMS. Lorsque les deux éditions diffèrent (du fait de coquilles de Gerhardt ou de sa tendance à moderniser les notations) j'ai systématiquement suivi l'édition des académies. Je continue néanmoins à donner aussi les références à LMS, qui est aisément disponible dans la plupart des bibliothèques universitaires.

Pour ce qui est des autres sources primaires, les vastes efforts de numérisation entrepris depuis quelques années m'ont le plus souvent permis de les citer d'après leur édition d'origine, et donc d'éviter tout risque de modernisation ultérieure des notations. Les articles mathématiques les plus importants de Leibniz sur le thème du calcul infinitésimal ont été traduits en français par Marc Parmentier, et accompagnés de notices utiles, dans un volume intitulé *La naissance du calcul différentiel* (= LNC) ; je m'en suis parfois aidé pour les traductions [35].

Dans tous les cas, j'ai pris soin de reproduire les notations originales aussi exactement que possible.

b) Littérature secondaire

Les textes dont je traite sont décrits sommairement par Knobloch (2002) dans le cadre d'une étude générale de la correspondance entre Leibniz et Jean Bernoulli. Pour le reste, ils ont été analysés par les historiens des mathématiques dans trois contextes principaux.

Tout d'abord, à tort ou à raison, presque toutes les études d'ensemble sur l'histoire du « calcul des opérations », « calcul des opérateurs » ou « calcul symbolique » voient dans ces textes la naissance de ce domaine des mathématiques. La plupart ne leur consacrent cependant que quelques lignes et aucune n'en fait une analyse poussée. Les traitements les moins réduits se trouvent chez Davis (1936) et Koppelman (1971).

Ensuite, ces textes sont également évoqués dans le cadre d'histoires générales de la dérivation d'ordre non entier. L'étude de Ross (1977) est fréquemment citée, mais truffée d'er-

35. Une récente traduction allemande des articles mathématiques de Leibniz, par Heinz-Jürgen Hess and Malte-Ludolf Babin (Leibniz 2011), m'a aussi été utile. Elle est plus complète et souvent plus rigoureuse.

reurs de lecture et d'approximations. La thèse de Dugowson (1994), non publiée, contient en revanche une analyse soigneuse (la seule que je connaisse) des brèves remarques de Leibniz sur les différences d'ordre fractionnaire.

Le dernier contexte dans lequel les historiens se sont intéressés à nos textes concerne une formule publiée par Jean Bernoulli en 1694[36], qui est souvent rapprochée de notre « formule de Taylor » (elle a été identifiée dès le XVIIIe siècle à plusieurs résultats de Taylor[37] et a donné lieu à une querelle de priorité souvent étudiée[38]). C'est cette formule de Bernoulli qui sert de point de départ aux réflexions de Leibniz sur l'« analogie des puissances et des différences », et, surtout, Leibniz et Bernoulli en donnent plusieurs démonstrations différentes dans nos lettres. L'importance capitale que prendront, au XVIIIe siècle, les développements en série a donc conduit de nombreux historiens à s'y intéresser, souvent en parallèle avec une étude des travaux de Taylor. C'est dans ce cadre que l'on trouve les études les plus solides et approfondies de notre corpus, en particulier celles de Feigenbaum (1985)[39] et de Panza (1992, p. 386-418), ou encore celle de Ferraro (2008, p. 45-51).

36. Jean Bernoulli 1694 ; voir *infra*.

37. B. Taylor 1715, proposition 11 ou théorème 4 ; corollaire 2 de la proposition 7 ou théorème 3.

38. Voir Pringsheim 1900 et surtout Auchter 1937, p. 28–36 et Fleckenstein 1946.

39. Voir aussi Feigenbaum 1986, un article moins complet mais aussi moins touffu qui se concentre sur la comparaison entre Taylor et Leibniz.

Chapitre 1

Éléments de contexte : le calcul différentiel leibnizien en 1694

Avant de plonger dans la correspondance de Leibniz et de Jean Bernoulli, il faut comprendre le contexte de leur travail : qu'est-ce que le calcul différentiel pour eux en 1694 ? Quels sont les problèmes qu'ils cherchent à résoudre ?

Outre ces questions très générales, je voudrais examiner ici un point spécifique, mais essentiel pour nous, des réflexions de Leibniz sur ses nouvelles méthodes. Il remarque en effet une analogie entre son calcul différentiel et l'« analyse ordinaire », qui traite de la résolution d'équations et de l'extraction de racines. Cette analogie le guide, et bien la comprendre nous aidera beaucoup à interpréter son vocabulaire et à comprendre ses intentions ; cet arrière-plan est également crucial pour distinguer ce qu'il y a de réellement neuf dans les découvertes qui nous occuperont au chapitre 2.

1.1 Le cadre : la géométrie cartésienne des courbes

Tout d'abord, qu'est-ce que le calcul différentiel de Leibniz ? Il s'agit d'une méthode nouvelle, mais qui s'inscrit pleinement dans le cadre de la géométrie de Descartes : c'est avant tout un outil pour l'étude de certains objets géométriques, les courbes (cercles, ellipses, paraboles, spirales, etc.), au moyen de *quantités variables* dénotées par des lettres [1]. Les plus usuelles de ces quantités variables sont illustrées à la figure 1.1 : l'ordonnée y (distance à une droite des ordonnées parallèlement à un axe des abscisses), l'abscisse x,

1. Je m'appuie tout au long de cette section sur l'excellente étude de Bos 1974, à laquelle je renvoie pour davantage de détails.

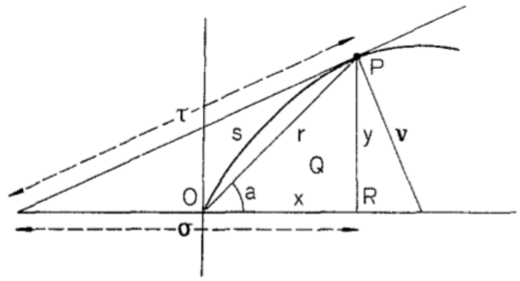


FIGURE 1.1 – Quantités géométriques variables associées à une courbe, d'après Bos (1974) : l'abscisse x, l'ordonnée y, l'arc s, le rayon r, l'arc polaire a, la sous-tangente σ, la tangente τ, la normale ν, la région OPR sous la courbe, etc.

l'arc s (portion de courbe à partir d'un point de départ fixé et jusqu'au point considéré), etc. Ces quantités ne sont définies que dans un contexte géométrique, sur la base d'une figure : pour définir l'ordonnée y d'une courbe, il faut choisir des droites qui pourront servir d'axes ; pour définir l'arc s, il faut choisir un point de départ (le point O pour la fig. 1.1). Elles sont variables au sens où elles prennent des valeurs différentes suivant le point considéré : à chaque point de la courbe, on peut associer une valeur de x, de y, de s, etc. Toutes ces quantités varient simultanément lorsque le point se déplace sur la courbe et ne sont pas considérées (comme on le fait parfois aujourd'hui) comme des *fonctions* d'une quantité privilégiée x, s ou autre. Bien sûr, leurs variations ne sont pas indépendantes les unes des autres : la courbe incarne certaines relations entre les variables. Par exemple, pour la courbe de la fig. 1.1, quand x augmente, y et s aussi. Dans certains cas, ces relations peuvent être exprimées exactement par une équation algébrique ; ainsi, pour une définition convenable des axes, les quantités x et y associées à une parabole vérifient l'équation $xx = ay$ (où a est une certaine longueur constante).

1.2 Les différentielles de Leibniz

Dans ce cadre général, Leibniz innove en introduisant de nouvelles quantités variables, les *différentielles*, dont les valeurs ne sont plus finies mais *infiniment petites*. Pour cela, il se fonde sur une analogie avec les suites de nombres. D'une manière générale, à toute suite de nombres

$$a \quad b \quad c \quad e \quad f \quad \text{etc.}$$

Leibniz associe la suite

$$a-b \qquad b-c \qquad c-e \qquad e-f \qquad \text{etc.}$$

de leurs *différences*[2]. Si l'on approxime une courbe par un polygone, et qu'on ne prend en considération que les sommets de ce polygone, chaque quantité variable associée à notre courbe conduit à une suite de nombres dont on peut alors prendre les différences (puis éventuellement les différences des différences, etc.). Par exemple, si l'on considère une parabole d'équation $ay = xx$ et que l'on place des points de notre polygone à intervalles réguliers selon l'axe des x, on trouve[3] (en posant pour simplifier $a = 1$) :

$x =$	0		1		2		3		4		5	etc.
$dx =$		1		1		1		1		1		etc.
$ddx =$			0		0		0		0		0	etc.
$y =$	0		1		4		9		16		25	etc.
$dy =$		1		3		5		7		9		etc.

Par extrapolation, Leibniz étend cette idée au cas d'une approximation polygonale ayant une infinité de côtés infiniment petits : il identifie la courbe à un tel polygone à une infinité de côtés, et traite les quantités variables x, y comme des suites composées d'une infinité de termes dont les différences dx, dy (appelées dans ce cas *différentielles*) entre termes successifs prennent des valeurs infiniment petites. Ces différentielles sont bien à leur tour des quantités géométriques variables, c'est-à-dire que leur valeur dépend du point de la courbe considéré : dans l'exemple précédent, en supposant que le polygone infini est obtenu en rapprochant les points tout en conservant leur disposition régulière selon l'axe des x, on trouve que dx est constante (ce que l'on peut écrire $ddx = 0$, c'est-à-dire que les différences des différences des x sont nulles), mais que dy ne l'est pas et prend des valeurs différentes en chaque point de la courbe.

Il y a toutefois une difficulté : les différentielles ainsi définies ne sont pas uniques. Pour la parabole précédente, on aurait aussi pu partir d'une approximation polygonale finie dont les points sont régulièrement disposés selon l'axe des y, obtenant quelque chose comme

$y =$	0	1	2	3	4	5	etc.
$x =$	0	1	$\sqrt{2}$	$\sqrt{3}$	2	$\sqrt{5}$	etc.

2. Il y a deux conventions de signe pour les différences : $a-b$, $b-c$ etc. ou $b-a$, $c-b$ etc. Dans la pratique, Leibniz travaille habituellement avec des suites monotones (croissantes ou décroissantes) et adopte la convention qui conduit à des différences positives.

3. Je précise que cet exemple ne figure pas tel quel dans les textes de Leibniz : ici encore, je m'appuie sur la reconstruction de Bos (1974).

d'où une extrapolation au cas infini conduisant à $ddy = 0$ et $ddx \neq 0$. Leibniz est parfaitement conscient de cette sous-détermination des différentielles, et écrit qu'on peut toujours choisir de supposer constantes les différences d'une des variables, selon ce qui nous est le plus commode. Henk Bos a étudié ce problème en détail et a bien montré que ce choix joue dans la théorie de Leibniz un rôle analogue à notre choix d'une variable indépendante (c'est-à-dire à notre choix de considérer toutes les autres variables comme fonctions d'une variable particulière [4]). Dans tous les cas, ce choix n'a pas d'impact sur les relations qui ne font intervenir que les différentielles premières (c'est-à-dire où ne figurent pas de double différentiation comme ddx) ; on peut donc l'ignorer pour les problèmes les plus simples [5].

L'intérêt d'introduire les nouvelles variables dx, dy, etc. tient au fait suivant. Les règles de l'« algorithme du calcul différentiel », publiées par Leibniz en 1684 [6], permettent de passer de l'équation d'une courbe [7], par exemple $ay = xx$, à une équation *différentielle*, c'est-à-dire une équation où apparaissent des quantités infiniment petites comme dx, dy en plus des quantités usuelles : l'équation précédente conduit ainsi à $ady = 2xdx$. Cette équation différentielle permet, entre autres, de déterminer les tangentes [8] et les points extrémaux de la courbe étudiée.

1.3 Le problème inverse : propriétés des tangentes, quadratures et équations différentielles

Inversement, si l'on connaît les tangentes à une certaine courbe, se pose alors la question de déterminer la courbe en question : en trouver une équation algébrique finie, dans la mesure du possible, ou en donner des approximations (par exemple sous la forme d'une expression par série infinie [9]). C'est le « problème inverse des tangentes », plus difficile [10].

4. Bos 1974, en part. §2.16–18, §3.1, §3.2.

5. Voir Bos 1974, §2.21.

6. Leibniz 1684b.

7. Dans ce contexte, une équation algébrique peut contenir des radicaux, c'est-à-dire des puissances fractionnaires ; il n'est pas nécessaire de se ramener d'abord à une équation ne contenant que des puissances entières positives, comme l'exigeaient les méthodes des tangentes de Descartes et de Fermat.

8. À l'époque, il est habituel de repérer la position de la tangente par la *sous-tangente* (souvent notée σ, voir fig. 1.1 ; nous l'exprimerions par la formule $y\frac{dx}{dy}$). Leibniz la calcule en passant par le rapport $\frac{dx}{dy}$, égal à notre dérivée, que l'on obtient facilement à partir de l'équation différentielle, mais ce rapport n'est qu'un intermédiaire et n'a pas le statut privilégié que nous lui donnons depuis le XVIIIe siècle ; voir Bos 1974.

9. Peut-on démontrer l'impossibilité, pour certaines courbes, d'une expression algébrique finie ? Cette question fonde la distinction entre courbes « algébriques » et courbes « transcendantes » et est d'une grande importance pour Leibniz, qui tente d'y répondre dès Leibniz 1684a.

10. Le prototype des problèmes de ce type, qui conduit à une courbe apparentée au graphe de notre fonction logarithme, a été posé par Florimond de Beaune à Descartes. La solution de Descartes, dans une lettre du 20

Or comme le calcul leibnizien permet de traduire une propriété des tangentes à une courbe en équation différentielle de cette courbe [11], le problème inverse des tangentes s'avère être un cas particulier d'un problème inverse plus général : étant donnée une équation différentielle – issue ou non d'une propriété des tangentes – déterminer la courbe correspondante, sous la forme d'une équation algébrique finie ou d'une série infinie.

Une découverte importante de Leibniz est que la recherche de quadratures, c'est-à-dire le calcul de l'aire d'un trapézoïde délimité par une courbe donnée, se ramène aussi à ce problème inverse : puisque la différence entre deux valeurs de l'aire sous la courbe pour des points infiniment proches sera le petit rectangle ydx, il s'agit de trouver S vérifiant l'équation différentielle $dS = ydx$. Cette équation différentielle est d'un type particulièrement simple, qui peut se résoudre en trouvant une quantité (ici S) dont on connaît la différentielle. Par exemple, si $y = 2x$, il faut trouver S tel que $dS = ydx = 2xdx$; or il est facile de voir, d'après les règles du calcul différentiel, qu'on peut prendre $S = xx$. Les frères Bernoulli, au début des années 1690, appellent cette quantité S « intégrale » de ydx (en écriture abrégée, I. ydx). Ainsi, Jean Bernoulli définit en 1691–92, dans un cours destiné au marquis de l'Hôpital, les intégrales des différentielles comme « les quantités dont elles sont les différentielles [12] ». Leibniz, quant à lui, avait de cette opération inverse de la différentiation une conception un peu différente, issue là encore de l'analogie avec les suites de nombres : pour lui, il s'agit d'une opération de *sommation*. Par exemple [13], à partir d'une suite décroissant vers zéro

$$a \qquad b \qquad c \qquad d \qquad \text{etc.}$$

février 1639 qui sera publiée dans sa correspondance en 1667 (Descartes 1657-1667, III, p. 409–416 ; cf. Maronne 2007) rend ce problème célèbre. Leibniz le traite explicitement et nommément dès la fin de son premier article sur le calcul différentiel (Leibniz 1684b), quoique sans utiliser, à ce moment-là, de méthode inverse générale.

11. En effet, la tangente est entièrement connue dès lors que l'on connaît les variables x, y et leurs différentielles dx, dy ou même simplement le quotient différentiel $\frac{dx}{dy}$; il est donc facile d'exprimer une propriété des tangentes sous la forme d'une équation contenant x, y, dx, dy. À titre d'exemple, voir la traduction différentielle que donne Leibniz du problème de Florimond de Beaune à la fin de la *Nova methodus* (Leibniz 1684b, p. 473 = LMS, V, p. 226).

12. « […] quomodo differentialium *Integrales*, id est, eæ quantitates quarum sunt differentiales, inveniantur, monstrabimus. » (« […] nous montrerons maintenant comment l'on trouve les *Intégrales* des différentielles, c'est-à-dire les quantités dont elles sont les différentielles. », *Lectiones mathematicæ de methodo integralium, aliisque*, BOO, III, p. 387. Ces cours n'ont été publiés qu'en 1742 dans les œuvres complètes de Jean Bernoulli, mais leur datation en 1691–92, plausible d'après leur contenu, semble hors de doute depuis la découverte d'un manuscrit datant vraisemblablement de 1705 les contenant exactement ; voir Schafheitlin 1922, p. 1–3.) Pour une autre occurence précoce du terme, cette fois chez son frère aîné Jacques, voir Jacques Bernoulli 1690.

13. Je reprends ici dans ses grandes lignes la présentation que fait Leibniz de la réciprocité entre sommes et différences dans sa lettre à Bernoulli du 28 février 1695 (LAA, III.6A, p. 313 = (LMS, V, p. 168)).

on obtient, par somme, la suite

$$a + b + c + d + \text{etc.} \qquad b + c + d + \text{etc.} \qquad c + d + \text{etc.} \qquad \text{etc.}$$

et par différence la suite

$$a - b \qquad b - c \qquad c - d \qquad \text{etc.}$$

Les sommes des différences sont alors

$$a - b + b - c + c - d + \text{etc.} = a \qquad b - c + c - d + \text{etc.} = b \qquad c \qquad d \qquad \text{etc.}$$

c'est-à-dire qu'on retrouve la suite de départ ; les différences des sommes sont, de même, les éléments de la suite de départ. Le passage aux différences et le passage aux sommes sont donc deux opérations inverses sur les suites. Sur ce modèle, Leibniz conçoit l'opération inverse de la différentiation d'une quantité comme une opération de sommation : il écrit ainsi $\int ydx$, le symbole $\int$ étant un S (pour somme) stylisé, plutôt que I. ydx, et interprète cette quantité comme la somme infinie des valeurs infiniment petites successives de la quantité variable ydx. Les quadratures sont alors intrinsèquement conçues comme des sommes infinies de rectangles infinitésimaux. Jean Bernoulli adoptera la notation de Leibniz, mais non sa terminologie, préférant continuer à parler d'intégrales ; c'est l'usage qui nous est resté.

En bref, le problème inverse général, sur lequel se concentrent Leibniz et Bernoulli à l'époque qui nous occupe, consiste à retrouver une quantité variable à partir d'une équation différentielle quelconque ; un cas particulier important est celui des équations différentielles comme $dS = ydx$ qui peuvent se résoudre par intégration ou sommation (l'opération réciproque à la différentiation).

1.4 Une analogie : différences et sommes, puissances et racines

Une remarque importante pour la suite est que Leibniz met en rapport cette configuration problématique avec celle de l'« analyse ordinaire », où il y a également deux problèmes : la résolution d'équations algébriques générales et l'extraction de racines (qui en est un cas particulier).

En 1694, quelques mois avant nos textes, Leibniz compare ainsi les couples d'opérations réciproques différences-sommes et puissances-racines :

> Car cette méthode, ou ce *calculus différentialis*, sert non seulement aux différences, mais encore aux sommes, qui sont le réciproque des différences, à peu près comme le calcul ordinaire ne sert pas seulement aux puissances, mais encore aux racines, qui sont le réciproque des puissances [14].

Pour expliciter, la réciprocité entre puissances et racines permet de transformer une équation de la forme

$$y^n = \ldots \quad \text{en} \quad y = \sqrt[n]{\ldots}$$

et inversement ; de la même manière, on peut passer de

$$dy = \ldots \quad \text{à} \quad y = \int \ldots$$

et inversement. L'analogie, remarque Leibniz, ne s'arrête pas là. Comme il l'avait déjà signalé plus haut dans le même texte, on ne sait pas, en général, résoudre une équation algébrique par radicaux : « le public n'a pas encore le moyen de trouver les racines du cinquième degré et au dela [15] ». En d'autres termes, on ne sait pas toujours éliminer les puissances (c'est-à-dire résoudre en y) dans une équation complexe qui contient plusieurs puissances différentes d'une même quantité y ; en revanche, on peut toujours éliminer les racines présentes dans une expression en élevant à la puissance (plusieurs fois le cas échéant). De la même manière, alors qu'on peut toujours éliminer les sommes dans une équation (en différentiant autant de fois que nécessaire), on ne peut pas toujours éliminer les différentielles dans une équation où figurent plusieurs ordres de différentielles d'une même quantité (dy et ddy, ou tout simplement dy et y, par exemple) :

> Et l'analogie va plus loin qu'on ne pense. Dans l'analyse ordinaire on peut toujours délivrer le calcul *a vinculo* et des racines par le moyen des puissances : mais le public n'a pas encore la méthode de le délivrer des puissances impliquées par le moyen des racines pures. De même dans notre Analyse des transcendantes, on peut toujours délivrer le calcul *a vinculo* et des sommes par le moyen des différences : mais le public n'a pas encore la méthode de le délivrer des différences impliquées par le moyen des sommes pures ou quadratures [...] [16].

J'interprète ici l'expression de « puissances impliquées » comme désignant les puissances y^e données par une expression contenant d'autres puissances de y : « délivrer le calcul des

14. Leibniz 1694a (LMS, V, p. 308).
15. Leibniz 1694a (LMS, V, p. 308).
16. Leibniz 1694a (LMS, V, p. 308).

puissances impliquées », c'est donc résoudre une équation algébrique par radicaux. De la même manière, j'interprète « différences impliquées » comme désignant les différences dy (ou ddy, etc.) données par une expression en y, dy, etc. : délivrer le calcul de ces différences impliquées, c'est donc résoudre une équation différentielle. Cette interprétation semble confirmée par des passages analogues mais plus explicites. Ainsi, en 1702, dans un contexte très similaire Leibniz parle (cette fois en latin) de racines « affectées », terme dont le sens est bien établi par ailleurs [17] :

> Ut in Algebra reciprocae sibi sunt *Potentiae* et *Radices*, ita in calculo infinitesimali *Differentiae* et *Summae* : et uti in Algebra seu scientia generali finitae magnitudinis potissimus scopus est *extrahere radices* formularum, ita in scientia infiniti *invenire summas* serierum [...]. Et quemadmodum aliae radices *purae* sunt, cum valores ex solis cognitis habentur, aliae *affectae*, cum ipsae earum potentiae valorem ipsarum ingrediuntur : ita quae summanda sunt, aut pure et plante sunt cognita, aut rursus implicant summam quasitam, ut si sit $dy = ayxdx :, ax + yy$, ubi y summa quaesita ingreditur valorem summandi dy [18].

La résolution d'une équation différentielle par sommes, qui n'est pas toujours possible, est donc analogue à la résolution d'une équation algébrique par radicaux.

Cette analogie a une dernière facette, qui jouera un rôle central par la suite. L'évaluation des sommes présente une difficulté analogue à l'extraction de racines : de même que les racines conduisent parfois à des nombres irrationnels, qu'on ne sait pas exprimer sous

17. Par exemple, dans sa lettre à Leibniz dite *Epistola prior* de 1676 (cf. infra, n. 20 p. 39), Newton, après avoir montré comment utiliser sa formule du binôme pour trouver des développements en série de diverses fractions et racines, présente une méthode pour obtenir des développements en série pour les racines d'« équations affectées », c'est-à-dire de polynômes. Pour d'autres passages parallèles où figure aussi ce terme, voir par exemple un brouillon de lettre à Huygens d'octobre 1690 (LAA, III.4, p. 595–596) ou encore une lettre ultérieure à Reyneau : Costabel 1949, p. 314.

18. « Il y a dans le calcul infinitésimal entre les *Différences* et les *Sommes* la même réciprocité qu'en Algèbre entre les *Puissances* et les *Racines*, et tout comme en Algèbre, science générale de la quantité finie, l'objectif principal est d'*extraire les racines* des expressions, il est pour la science de l'infini, de *trouver les sommes* des séries [...] Si la valeur des racines ne fait apparaître que des termes connus, ces racines sont *pures*, mais lorsqu'elles interviennent elles-mêmes avec une certaine puissance dans l'expression qui en donne la valeur, elles sont *affectées* ; de la même manière, les expressions à sommer ou bien sont purement et simplement données, ou bien renferment elles-mêmes la somme cherchée, par exemple $dy = \frac{ayxdx}{ax+yy}$ faisant intervenir la somme inconnue y dans le dy qu'il faut sommer. » (Leibniz 1702, p. 210 = LMS, V, p. 350–351, trad. fr. de LNC.) L'objet de cet article est d'introduire une technique analogue à ce que nous appelons la décomposition en éléments simples d'une fraction rationnelle, afin de ramener la quadrature de courbes données par des équations compliquées à des transcendantes connues, par exemple à la quadrature de l'hyperbole. L'introduction de cet article vise à situer cette contribution dans le programme d'ensemble de la science de l'infini : il faut d'abord savoir ramener la solution d'une équation différentielle à une quadrature, puis – contribution spécifique de cet article – savoir ramener les quadratures à des quadratures plus simples.

forme rationnelle finie, de même les sommes conduisent parfois à des grandeurs qu'on ne sait pas exprimer sous la forme d'une expression algébrique finie. Leibniz écrit ainsi, à la suite du passage de 1694 cité plus haut :

> [...] comme il n'est pas toujours possible de tirer les racines effectivement pour parvenir aux grandeurs rationnelles de l'Arithmétique commune, il n'est pas toujours possible non plus de donner effectivement les sommes ou quadratures, pour parvenir aux grandeurs ordinaires ou algébriques de l'analyse commune. Cependant par le moyen des séries infinies on peut toujours exprimer des grandeurs rompues comme en entiers, et des incommensurables en rationnelles, et des transcendantes en ordinaires [19].

Ce qui importe ici est que pour les racines, il existe une méthode générale d'approximation en séries infinies, qui servira de modèle à Leibniz pour les différentielles. Leibniz tient cette méthode générale d'une célèbre lettre que lui envoie Newton en 1676, dite *Epistola Prior* [20]. Newton y propose une généralisation de la notation exponentielle pour les puissances [21] :

> [S]icut Analystæ, pro aa, aaa, &c. scribere solent a^2, a^3, &c. sic ego, pro $\sqrt{}a$, $\sqrt{}a^3$, $\sqrt{}c.a^5$ [22], &c. scribo $a^{\frac{1}{2}}$, $a^{\frac{3}{2}}$, $a^{\frac{5}{3}}$; & pro $\frac{1}{a}$, $\frac{1}{aa}$, $\frac{1}{aaa}$, scribo a^{-1}, a^{-2}, a^{-3} [23].

À l'aide de cette notation, Newton donne ensuite une formule générale, habituellement connue sous le nom de « formule du binôme », pour développer en série infinie n'importe quelle somme de deux termes élevée à une puissance quelconque (entière ou fractionnaire, positive ou négative). Dans sa lettre, il ne donne les termes successifs qu'au moyen de formules récursives, mais Leibniz en écrit immédiatement une version non récursive en marge [24], et en utilise dès 1677–78 la version suivante : en posant « x aequ. $A + B$ », il

19. Leibniz 1694a (LMS, V, p. 308).

20. Lettre de Newton à Oldenburg du 13 juin 1676, recopiée par Oldenburg et envoyée à Leibniz le 26 juillet 1676 ; Collins 1712, p. 49-57 ou LMS, I, p. 100–113. Newton a adressé en tout et pour tout deux lettres à Leibniz, toutes deux d'une grande importance mathématique ; celle-ci est la première, d'où son appellation.

21. Cette notation vient de la *Géométrie* de Descartes : « [j'écris] aa, ou a^2, pour multiplier a par soy mesme ; Et a^3, pour le multiplier encore une fois par a, & ainsi a l'infini. » (Descartes 1637, p. 299). Quarante ans plus tard, elle est largement utilisée.

22. $\sqrt{}c$ désigne ici la racine cubique. C'est une notation répandue, qui se trouve aussi dans la *Géométrie* de Descartes.

23. « De la même manière que les analystes, pour aa, aaa, etc. ont l'habitude d'écrire a^2, a^3, etc., de même moi, pour $\sqrt{}a$, $\sqrt{}a^3$, $\sqrt{}c.a^5$, etc. j'écris $a^{\frac{1}{2}}$, $a^{\frac{3}{2}}$, $a^{\frac{5}{3}}$; et pour $\frac{1}{a}$, $\frac{1}{aa}$, $\frac{1}{aaa}$, j'écris a^{-1}, a^{-2}, a^{-3}. » (Collins 1712, p. 49 ou LMS, I, p. 101.)

24. Cf. LMS, I, p. 100, note.

obtient pour $\boxed{z}\, x$ (c'est-à-dire x^z) l'expression

$$A^z + \frac{z}{1}A^{z-1}B^1 + \frac{z, z-1}{1, 2}A^{z-2}B^2 + \frac{z, z-1, z-2}{1, 2, 3}A^{z-3}B^3 \text{ etc.}$$

et précise que si z n'est pas entier, on obtient une série infinie [25]. Plus près de la période qui nous occupera, Leibniz continue à employer cette formule. Par exemple [26], dans un article d'août 1694 consacré à la courbe élastique, il en écrit un cas particulier comme ceci [27] :

$$\boxed{e}\,\overline{1+b} = 1 + \frac{e}{1}b + \frac{e.e-1}{1.2}bb + \frac{e.e-1.e-2}{1.2.3.}b^3 \;\&c.$$

Leibniz est impressionné par le fait qu'une notation bien choisie permette d'écrire une formule très générale, subsumant les diverses méthodes connues de développement en séries infinies pour les fractions et les racines [28]. Lorsqu'il s'agira de développer en série non plus des racines, mais des sommes, Leibniz gardera en tête cette série générale, et nous la verrons réapparaître au prochain chapitre.

1.5 Stratégies de résolution du problème inverse

Pour résumer, le problème inverse général – remonter d'une équation différentielle à la courbe correspondante – présente plusieurs difficultés, qui forment le contexte problématique essentiel des textes de Leibniz et Bernoulli qui nous occupent. Tout d'abord, il n'est pas toujours possible de se ramener à des sommes ou intégrales pour résoudre une équation différentielle ; ensuite, même quand on peut le faire (par exemple pour les quadratures et rectifications), on ne peut pas forcément exprimer de manière algébrique finie la somme obtenue.

Que faire face à une somme que l'on ne sait pas calculer ? On peut chercher à la transformer de manière à se ramener à des sommes les plus simples possibles. Cela revient à introduire de nouvelles courbes apparemment non exprimables au moyen d'une équation algébrique finie, dites pour cette raison « transcendantes », par exemple les quadratures du cercle ou de l'hyperbole (logarithme), et à exprimer la quantité recherchée au moyen

25. Manuscrit publié par Eberhard Knobloch (dans Leibniz 1976, p. 74–77), qui le date de 1677–78. Pour un commentaire de ce manuscrit, voir Knobloch 1973, section 2.2, en part. p. 98–99.

26. Pour une autre occurence datée de juillet 1695, donc déjà au même moment que les recherches qui font l'objet du prochain chapitre, voir Leibniz 1695, p. 312 = LMS, V, p. 323.

27. Leibniz 1694b, p. 369 = LMS, V, p. 313 (sur cet article, voir *infra*, note 30).

28. Serfati (2005, p. 390-391) va même jusqu'à soutenir que l'*Epistola Prior* est la source première de toutes les préoccupations ultérieures de Leibniz pour la symbolique.

de ces transcendantes élémentaires [29] (déjà bien étudiées et pour lesquelles on connaît des procédés d'approximation).

Une autre stratégie, qui est la seule possible dans le cas d'équations différentielles qu'on ne sait pas résoudre par sommes, consiste à exprimer directement la quantité recherchée au moyen d'une série infinie. La technique principale [30] dont dispose Leibniz pour obtenir de telles séries est la *méthode des coefficients indéterminés*, qu'il utilise dans le cas d'expressions finies (à la suite de Descartes) déjà dans les années 1680 [31] mais n'étend au cas des séries infinies qu'en 1693 [32]. En voici un exemple. Partons de l'équation différentielle

$$dy = adx :, a + x \quad \text{(c'est-à-dire } dy = \frac{adx}{a+x}\text{)}$$

que Leibniz réécrit

$$ady : dx + xdy : dx - a = 0 \quad \text{(c'est-à-dire } \frac{ady}{dx} + \frac{xdy}{dx} - a = 0\text{).}$$

Le principe est alors de supposer que la quantité recherchée, ici y, est donnée par une série $y = bx + cx^2 + ex^3 + fx^4$ etc. dont les coefficients sont laissés indéterminés et notés par des lettres. Ici, en remplaçant y par cette série, Leibniz obtient :

$$\begin{array}{lclllll} ady : dx & = & ab & +2acx & +3aex^2 & +4afx^3 & \text{etc.} \\ +xdy : dx & = & & +bx & +2cx^2 & +3ex^3 & \text{etc.} \\ -a & = & -a & & & & \end{array}$$

Comme la somme de la colonne de gauche est nulle (c'est notre équation différentielle), la somme de la colonne de droite l'est aussi : on obtient donc une série infinie égale à zéro. La méthode repose alors sur l'idée que tous les coefficients d'une série nulle doivent être nuls : ainsi, Leibniz conclut que $ab - a = 0$ d'où $b = 1$ (terme constant) ; puis $2ac + b = 0$ d'où $c = -1 : 2a$ (coefficient de x) ; puis $3ae + 2c = 0$ d'où $e = 1 : 3a^2$ (coefficient de x^2), et ainsi de suite. Finalement, il en déduit que y peut s'écrire

$$y = \frac{x}{1} - \frac{x^2}{2a} + \frac{x^3}{3a^2} - \frac{x^4}{4a^3} \text{ etc.}$$

Cette méthode est très générale : elle s'applique à toutes sortes d'équations différentielles,

29. C'est le programme qu'expose Leibniz 1686 en l'illustrant brièvement dans le cas de la cycloïde.

30. On rencontre à l'occasion une autre méthode : Leibniz utilise l'intégration terme à terme d'une série infinie obtenue par la formule générale du binôme (cf. section 1.4) pour calculer $\int, xxdx : \sqrt{1 - x4}$, c'est-à-dire $\int \frac{xx}{1-x^4}$ (Leibniz 1694b, p. 369).

31. Voir Leibniz 1684a, 1686.

32. Leibniz 1693.

même contenant des différentielles secondes ou supérieures, et même si elles ne peuvent pas se résoudre par simples intégrations ou sommations.

L'inconvénient de cette méthode, écrit Jean Bernoulli quelques mois plus tard lorsqu'il en propose une autre, est qu'elle exige un calcul différent dans chaque cas particulier [33]. Sa méthode à lui est au contraire directe et uniforme, quoiqu'elle soit moins générale : elle ne s'applique qu'aux quantités que l'on peut exprimer sous la forme d'une intégrale de ndz (ce qui est certes déjà beaucoup puisque, comme le souligne Bernoulli, cela permet d'exprimer tous les problèmes de quadrature et de rectification en une seule formule). Elle repose sur la série infinie suivante [34], que Bernoulli qualifie d'« universelle » et de « très générale » :

$$\text{integr. } ndz = +nz - \frac{1}{1.2} zz \frac{dn}{dz} + \frac{1}{1.2.3} \frac{z^3 ddn}{dz^2} - \frac{1}{1.2.3.4} \frac{z^4 dddn}{dz^3} \text{ \&c.}$$

Cette formule exhibe la forme générale de la série solution en termes de différentielles d'ordre supérieur de n ; il suffit donc pour l'utiliser de différentier n de manière répétée, comme on en a aujourd'hui l'habitude quand on utilise le théorème de Taylor. Par exemple, la quantité y donnée par $dy = \frac{adx}{a+x}$ (que Leibniz avait exprimée par une série infinie *via* la méthode des coefficients indéterminés) peut se ramener à la forme Integr. ndz en prenant $z = x$ et $n = \frac{a}{r}$ avec $r = a + x$. La formule de Bernoulli conduit alors facilement à la série

$$y = \frac{ax}{r^1} + \frac{axx}{2r^2} + \frac{ax^3}{3r^3} + \frac{ax^4}{4r^4} \text{ \&c.}$$

qui se ramène à celle de Leibniz [35].

Pour Bernoulli, cette série apporte une contribution majeure au calcul différentiel, et nous pouvons maintenant comprendre pourquoi : non seulement elle résout un cas central du problème inverse, mais elle le fait *via* une formule générale. La démonstration qu'en donne Bernoulli frappe tout particulièrement Leibniz, parce qu'elle est généralisable et ouvre donc la porte de tout un monde de séries générales de sommes et de différences. C'est là que commence notre histoire.

33. « [...] in quovis exemplo singularis calculus instituendus [est], quo non nisi singularis series provenit quæsito respondens [...] » (« [...] dans n'importe quel exemple, il faut faire un calcul particulier, dont ne provient rien d'autre que la série particulière répondant à la question posée [...] », Jean Bernoulli 1694, p. 437).

34. Jean Bernoulli 1694, p. 438.

35. On passe facilement de la série de Bernoulli à celle de Leibniz en remplaçant a par $r - x$ dans chaque numérateur.

Chapitre 2

Récit : la découverte de l'« analogie »

En 1694, Jean Bernoulli découvre la formule remarquable suivante, déjà mentionnée [1] :

$$\text{integr. } ndz = +nz - \frac{1}{1.2}zz\frac{dn}{dz} + \frac{1}{1.2.3}\frac{z^3ddn}{dz^2} - \frac{1}{1.2.3.4}\frac{z^4dddn}{dz^3} \ \&\text{c.}$$

Stimulés par cette découverte qui les impressionne, Leibniz et Jean Bernoulli sont conduits, d'investigation en investigation, à ce qu'ils appellent l'« analogie des puissances et des différences ». C'est leur chemin que je veux ici retracer pas à pas.

2.1 L'apparition de la notation exponentielle des différences

Dans une lettre du 2 septembre 1694, Bernoulli annonce son résultat à Leibniz en quelques lignes [2] ; pour la démonstration, il renvoie à un article à paraître dans les *Acta Eruditorum* [3]. Le 6–16 décembre, Leibniz – qui n'a pas encore reçu l'article publié et n'a donc pas vu la démonstration – lui répond qu'il avait déjà découvert un résultat similaire longtemps auparavant, et lui envoie sa preuve de l'époque, fondée sur la considération de sommes et différences finies [4]. Je ne l'examinerai pas en détail [5], mais un point mérite une

1. Jean Bernoulli 1694, p. 438. Cette formule est valable « pour dz constant » ; pour quelques explications, voir section 1.2.
2. LAA, III.6A, p. 172 = LMS, III/1, p. 150.
3. Jean Bernoulli 1694, publié dans le volume de novembre.
4. LAA, III.6A, p. 243–244 = LMS, III/1, p. 155–156.
5. Voir Feigenbaum 1985, p. 84–87, Feigenbaum 1986 ou Panza 1992, vol. 1, p. 395–397.

attention toute particulière : Leibniz y utilise une notation nouvelle [6].

En effet, Leibniz emploie ici une notation « exponentielle » pour les différentielles d'ordre supérieur : il écrit d^3y là où Bernoulli écrivait $dddn$. C'est sans doute l'une des toutes premières occurences de cette notation [7]. Avant 1694, je ne l'ai trouvée nulle part chez Leibniz, ni dans les premiers manuscrits (aujourd'hui publiés [8]) où il introduit les symboles d et $\int$, ni dans sa correspondance savante, ni dans ses articles mathématiques. Elle est absente même dans les passages où on l'attendrait : quand Leibniz rédige des présentations de son calcul pour Tschirnhaus [9] ou pour Huygens [10], il énumère dy, ddy, $dddy$ sans abréviation ; de même l'Hôpital, dans son manuel de calcul différentiel tiré des cours que lui donne Jean Bernoulli au début des années 1690, écrit : « on marquera par dd la différence seconde ou du second genre ; par ddd, la différence troisiéme ou du troisiéme genre ; par $dddd$, la différence quatriéme ou du quatriéme genre ; & de même des autres [11] ». Cette absence est frappante parce qu'à partir de 1695, Leibniz utilise très systématiquement la notation d^3 dans ce genre de contexte [12]. En 1694 même, j'en ai trouvé en tout et pour tout deux apparitions discrètes avant notre lettre de décembre [13].

Ce qui est sûr, c'est que dans les premiers temps du calcul infinitésimal, il est rarement nécessaire d'employer des différentielles au-delà du second ordre, et jamais (à ma connaissance) au-delà du troisième ; or tant qu'on en reste au second ordre, il n'est pas gênant d'écrire ddx, et ce d'autant moins que l'usage d'alors préfère aa à a^2, voire parfois aaa à a^3. Si Leibniz, l'Hôpital et les frères Bernoulli sont parfaitement conscients du fait qu'on peut itérer la différentiation autant autant de fois qu'on le souhaite, cette possibilité

6. Leibniz dit avoir découvert la preuve en question des années plus tôt, mais c'était avant d'avoir introduit la notation d, exponentielle ou non : si la preuve est ancienne, la notation est neuve. C'est du moins ce qu'il affirme dans un manuscrit rédigé autour de 1715, l'*Historia et origo calculi differentialis*, qu'il meurt avant de diffuser et qui ne sera publié qu'au XIXe siècle : il y reprend la même preuve que dans cette lettre et écrit : « [...] loquendo stylo a se postea introducto, et terminum seriei vocando y [...] licebit differentiam primam vocare dy, secundam ddy, tertiam d^3y, quartam d^4y [...] » (« en parlant dans un style introduit plus tard, et en appelant y le terme de la série [...] on pourra appeler dy la différentielle première, ddy la seconde, d^3y la troisième, d^4y la quatrième [...] », Leibniz 1846, p. 6).

7. Cajori 1928-1929, II, p. 204–205 et Serfati 2001, p. 209 en situent l'apparition en 1695, ce qui est au mieux légèrement inexact (il y en a quelques occurences en 1694, dont la lettre qui nous occupe ; voir *infra*) mais suggère tout de même qu'elle n'est guère présente auparavant.

8. Voir LAA, VII.3, VII.5.

9. « $\iint y$, $\iiint y$, ddy, $dddy$ », lettre de Leibniz à Tschirnhaus (non datée, sans doute 17 octobre 1684, peut-être non envoyée ou seulement indirectement *via* Mencke), LAA, III.4, p. 164 = LMS, IV, p. 507 = Leibniz 1899, p. 459.

10. « $d\overline{y}$, $d\overline{dy}$, $d\overline{ddy}$ » (lettre de Leibniz à Huygens du 15/25 juillet 1690, LAA, III.4, p. 534 = LMS, II, p. 43).

11. L'Hôpital 1696, p. 55.

12. Pour quelques exemples, voir *infra*, note 89 p. 69.

13. Lettre de Leibniz à Huygens du 12/22 juin 1694 (LAA, III.6A, p. 127 = LMS, II, p. 181) ; lettre de l'Hôpital à Jean Bernoulli du 7 juin 1694 et réponse du 27 juin (BBW, I, p. 224 et 229).

reste donc théorique : même s'ils avaient connu cette notation, il n'en auraient guère eu l'usage. C'est précisément avec la formule intégrale générale de Bernoulli que la situation change. Pour la première fois, Leibniz est confronté à une formule qui fait apparaître des différentielles d'ordre quelconque, et ce n'est qu'à partir de là que la notation exponentielle devient réellement utile. Les découvertes que nous étudierons dans les pages qui suivent convaincront ensuite Leibniz de la pertinence de cette notation, et à partir de l'été 1695, il ne manquera plus une occasion de la mettre en avant [14].

Quoi qu'il en soit, Leibniz avait-il, déjà à ce stade, des raisons de choisir cette notation-là plutôt qu'une autre ? Elle reprend le principe de répétition des puissances usuelles, au sens où d^3 signifie ddd comme a^3 signifie aaa (quoique ici, la juxtaposition ne vaille plus multiplication mais application répétée de la différentiation), et ce parallèle la rend commode et facile à comprendre ; peut-être n'est-il pas nécessaire de chercher plus loin. Il se peut cependant que Leibniz ait eu d'autres motivations. Pour commencer, il n'est pas complètement impossible qu'il ait eu en tête son idée d'une analogie entre les paires puissances-racines et différences-sommes [15], quoique le lien avec son choix notationnel me semble très vague.

Surtout, Leibniz pourrait avoir pensé à ce qu'il appellera plus tard la « loi transcendantale des homogènes ». Voici l'idée : de même que dx est infiniment petit par rapport à x, ddx et $dxdx$ sont infiniment petits par rapport à dx, etc. de sorte qu'il n'y a guère de sens à ajouter entre eux dx et ddx, par exemple, puisque le second est négligeable par rapport au premier. De la même manière que classiquement, tous les termes d'une expression doivent être homogènes en dimension (c'est-à-dire qu'on ne doit pas ajouter de lignes à des surfaces, etc.), ils doivent donc aussi être homogènes en « ordre d'infinité », pour reprendre une expression de Henk Bos [16]. Leibniz fait allusion à ce principe dès 1690, dans une lettre à Huygens ; il y défend son choix de désigner les différences successives de x par dx, ddx etc. plutôt que par des lettres arbitraires, qui ne feraient pas voir quelle quantité est différentiée et combien de fois elle l'est, et ajoute :

> [C]ela paroist sur tout convenable quand il y a plusieurs lettres, et plusieurs degrés de differences à combiner, comme il m'est arrivé quelques fois car il y a alors à observer *une certaine loy d'homogenes toute particuliere*, et la seule vüe decouvre ce qu'on ne deméleroit pas si aisement par des notes vagues, comme sont des simples lettres [17].

14. Pour quelques exemples, voir *infra*, note 89 p. 69.
15. Voir section 1.4.
16. Sur ces deux lois d'homogénéité, cf. par exemple Bos 1974, p. 6, 33-34.
17. Lettre à Huygens du 3/13 octobre 1690, LAA, III.4, p. 620 = LMS, II, p. 49 (je souligne). Pour une discussion plus approfondie de cette lettre, voir *infra*, p. 76 *sq.*.

On peut expliciter le propos de Leibniz ainsi : si x, y, z désignent des lignes, on peut observer que xxz et xyy sont homogènes en dimension en comptant leurs nombres respectifs de facteurs ; de la même manière, on peut observer que $zdxdy$ et $xzddy$ sont homogènes en ordre d'infinité (ou comme dira Leibniz, « transcendantalement ») en comptant leurs nombres respectifs de lettres d. Or à partir de là, un parallèle direct, que Leibniz ne rend explicite que vingt ans plus tard [18], justifie une notation exponentielle pour les différences : si l'on écrit $xxz = x^2z^1$ et $xyy = x^1y^2$, on peut constater l'homogénéité de ces deux termes en additionnant leurs exposants, c'est-à-dire $2+1 = 3 = 1+2$; et si l'on écrit $zdxdy = zd^1xd^1y$ ou $d^0zd^1xd^1y$, et $xzddy = xzd^2y$ ou $d^0xd^0zd^2y$, on peut procéder de la même manière pour la loi transcendantale des homogènes. Il est tout à fait possible qu'en 1694, Leibniz ait déjà remarqué ce parallélisme.

2.2 De la formule de Bernoulli à l'écriture $d^{-1} = \int$

Face à la formule de Bernoulli, Leibniz pense à une démonstration par les différences finies. C'est, pour lui, une démarche très naturelle : comme nous l'avons vu [19], son calcul différentiel est à l'origine une extrapolation à partir des sommes et différences finies ; et c'est, comme il l'écrit à de multiples reprises [20], *via* la réciprocité des sommes et différences finies qu'il a découvert que le problème des tangentes et celui des quadratures sont inverses l'un de l'autre. Revenir aux sommes ou différences finies pour démontrer la formule de Bernoulli, c'est donc revenir à la nature même des sommes et des différences et au fondement de leur réciprocité.

La démonstration que Bernoulli donne de sa formule est toute différente : comme il l'écrit dans sa réponse à Leibniz, « concernant mon autre série [...] tu verras peut-être sans tarder dans les *Acta* que je n'en fais pas venir l'origine d'aussi loin [21] ». Sa démonstration repose sur un usage habile, apparenté à notre « intégration par parties », de la règle de

18. Cf. Leibniz 1710b, p. 165 = LMS, V, p. 381–382, où il introduit l'expression de « loi transcendantale des homogènes » ; voir aussi *infra*, section 2.3.

19. Voir section 1.2.

20. Par exemple, au cours de la même correspondance : « Ego certe in totam hanc methodum me fateor, ex hac consideratione reciprocationis inter summas differentiasque, incidisse, et a Seriebus numerorum ad linearum seu ordinatarum considerationes processisse. » (« Quant à moi, je le proclame, c'est la considération de la réciprocité entre les sommes et les différences qui m'a conduit à toute cette méthode ; je suis allé des séries de nombres à l'examen des lignes ou ordonnées. » Lettre du 28 février 1695, LAA, III.6A, p. 314 = LMS, III/1, p. 168.) Voir aussi Bos 1974, p. 13–14.

21. « Quod concernit alteram meam seriem [...] jam forsan in Actis videbis, me ejus originem non ita longe accersere [...] » (Lettre du 2–12 février 1695, LAA, III.6A, p. 288 = LMS, III/1, p. 160.)

différentiation d'un produit. Il écrit d'abord [22]

$$ndz = +ndz \underbrace{+zdn - zdn}\underbrace{-\frac{zzddn}{1.2\,dz} + \frac{zzddn}{1.2dz}} + \underbrace{\frac{z^3dddn}{1.2.3dz^2}}\ \&c.$$

où tous les termes du membre de droite sauf le premier s'annulent deux à deux (comme Feigenbaum 1985, j'ai ajouté des accolades pour clarifier). Si l'on regroupe les termes autrement, comme ceci (les accolades étant à nouveau de moi),

$$ndz = \underbrace{+ndz + zdn}\underbrace{-zdn - \frac{zzddn}{1.2\,dz}} + \underbrace{\frac{zzddn}{1.2dz} + \frac{z^3dddn}{1.2.3dz^2}}\ \&c.$$

on peut reconnaître dans chaque groupe de deux termes consécutifs la différentielle d'un produit [23], d'où en intégrant :

$$\text{integr. } ndz = +nz - \frac{1}{1.2}zz\frac{dn}{dz} + \frac{1}{1.2.3}\frac{z^3ddn}{dz^2} - \frac{1}{1.2.3.4}\frac{z^4dddn}{dz^3}\ \&c.$$

Notons qu'on passe d'un terme au suivant en intégrant les expressions en z et en différentiant les expressions en n.

Dès sa lettre suivante, Leibniz (qui a entre temps reçu les *Acta* et lu l'article de Bernoulli) qualifiera cette preuve de « très belle et inattendue [24] ». Inattendue, sans doute, parce que contrairement à la preuve de Leibniz qui requiert un retour un peu complexe à la définition des différences successives et de la somme, celle de Bernoulli conduit au résultat presque sans rien, uniquement au moyen de la règle formelle de dérivation d'un produit. Elle ouvre ainsi un domaine de relations générales entre sommes et différences explorable au seul moyen de l'algorithme du calcul différentiel.

En tout cas, cette « très belle » méthode retient grandement l'attention de Leibniz, qui y revient avec enthousiasme dès le post-scriptum de la même lettre [25] : il la généralise en calculant $\int \overline{\overset{e}{z}\cdot\overset{m}{d}\cdot n}$ pour e quelconque et m entier positif [26]. À la différence de Bernoulli

22. Jean Bernoulli 1694, p. 438.

23. Pour expliciter, nous pouvons écrire :

$$ndz + zdn = d(nz), \qquad -zdn - \frac{zzddn}{1\cdot 2\cdot dz} = d\left(\frac{zzdn}{1\cdot 2\cdot dz}\right), \qquad \frac{zzddn}{1\cdot 2\cdot dz} + \frac{z^3dddn}{1\cdot 2\cdot 3\cdot dz^2} = d\left(\frac{z^3ddn}{1\cdot 2\cdot 3\cdot dz^2}\right)$$

et ainsi de suite (en se rappelant qu'on prend dz constant).

24. « Pulcherrima mihi profecto Tua seriei obtinendae ratio visa est, quam in Actis explicuisti, et inexpectata. » (Lettre du 28 février 1695, LAA, III.6A, p. 310 = LMS, III/1, p. 164.)

25. Lettre du 28 février 1695, LAA, III.6A, p. 312–315 = LMS, III/1, p. 166-169.

26. À une légère différence notationnelle près – l'interversion de z et n – l'intégrale de ndz calculée par Bernoulli est le cas particulier $e = 1$, $m = 1$; mais Bernoulli avait supposé dz constant (donc dans la notation de Leibniz, dn constant), alors que Leibniz suppose dz constant, ce qui explique que la formule dérivée par Bernoulli ne s'obtienne pas comme cas particulier de celle de Leibniz.

dans les *Acta*, Leibniz rédige la démonstration directement sous forme intégrale et avance de proche en proche, écrivant successivement les termes qu'il faut ajouter ou retrancher et le terme correspondant de la série finale. Il semble cependant avoir écrit ce post-scriptum en hâte, et fait de multiples erreurs de calcul[27] :

$$\begin{aligned}\int \overline{z^{\overset{e}{\cdot}} d^{\overset{m}{\cdot}} n} + e \int \overline{z^{\overset{\overline{e-1}}{\cdot}} d^{\overset{\overline{m-1}}{\cdot}} n \cdot dz} &= z^{\overset{e}{\cdot}} d^{\overset{m}{\cdot}} n \\ - edz \int \overline{z^{\overset{\overline{e-1}}{\cdot}} d^{\overset{\overline{m-1}}{\cdot}} n} - eedz \int \overline{z^{\overset{\overline{e-2}}{\cdot}} d^{\overset{\overline{m-2}}{\cdot}} n \cdot dz} &= -e \cdot z^{\overset{\overline{e-1}}{\cdot}} d^{\overset{\overline{m-1}}{\cdot}} n \cdot dz \\ + ee\overline{dz}^2 \int \overline{z^{\overset{\overline{e-2}}{\cdot}} d^{\overset{\overline{m-2}}{\cdot}} n} + e^3\overline{dz}^2 \int \overline{z^{\overset{\overline{e-3}}{\cdot}} d^{\overline{m-3}} ndz} &= +ee \cdot z^{\overset{\overline{e-2}}{\cdot}} d^{\overline{m-2}} n \cdot \overline{dz}^2\end{aligned}$$

Si l'on additionne ces lignes, la somme des membres de gauche sera égale à la somme des membres de droite. Or, selon le même principe que ci-dessus, la somme des termes à gauche se simplifie en $\int \overline{z^{\overset{e}{\cdot}} d^{\overset{m}{\cdot}} n}$ (le second terme de chaque ligne s'annule avec le premier terme de la suivante) d'où la série :

$$\int \overline{z^{\overset{e}{\cdot}} d^{\overset{m}{\cdot}} n} = z^e d^{\overset{m}{\cdot}} n - e \cdot z^{\overset{\overline{e-1}}{\cdot}} d^{\overline{m-1}} n \cdot dz + e^2 \cdot z^{\overset{\overline{e-2}}{\cdot}} d^{\overset{\overline{m-2}}{\cdot}} n\overline{dz}^2 - e^3 z^{\overset{\overline{e-3}}{\cdot}} d^{\overline{m-3}} n \cdot \overline{dz}^3 \text{ etc.}$$

Il y a deux erreurs : d'une part une simple erreur d'exposants (le membre de droite de la première ligne devrait être $z^e d^{m-1} n$ et non $z^e d^m n$, par exemple ; la même erreur est reproduite à chaque ligne, donc les exposants du d devant n sont décalés à chaque étape) ; d'autre part, Leibniz semble considérer que $d^2(z^e) = e^2 z^{e-2} dz^2$ et non $e(e-1)z^{e-2}dz^2$, et ainsi de suite pour les puissances plus élevées. Les calculs corrects sont donnés par Bernoulli dans sa réponse[28], et approuvés par Leibniz[29] :

$$\begin{aligned}\int \overline{z^{\overset{e}{\cdot}} d^{\overset{m}{\cdot}} n} + e \cdot \int \overline{z^{\overset{\overline{e-1}}{\cdot}} d^{\overset{\overline{m-1}}{\cdot}} n \cdot dz} &= z^{\overset{e}{\cdot}} d^{\overset{\overline{m-1}}{\cdot}} n \\ - edz \cdot \int \overline{z^{\overset{\overline{e-1}}{\cdot}} d^{\overset{\overline{m-1}}{\cdot}} n} - e \cdot e - 1 \cdot dz \cdot \int \overline{z^{\overset{\overline{e-2}}{\cdot}} d^{\overset{\overline{m-2}}{\cdot}} n \cdot dz} &= -e.z^{\overset{\overline{e-1}}{\cdot}} d^{\overset{\overline{m-2}}{\cdot}} n \cdot dz \\ + e \cdot e - 1 \cdot \overline{dz}^2 \int \overline{z^{\overset{\overline{e-2}}{\cdot}} d^{\overset{\overline{m-2}}{\cdot}} n} + e \cdot e - 1 \cdot e - 2 \cdot \overline{dz}^2 \int \overline{z^{\overset{\overline{e-3}}{\cdot}} d^{\overset{\overline{m-3}}{\cdot}} n \cdot dz} & \\ &= +e \cdot e - 1 \cdot z^{\overset{\overline{e-2}}{\cdot}} d^{\overset{\overline{m-3}}{\cdot}} n\overline{dz}^2\end{aligned}$$

d'où la série :

$$\int \overline{z^{\overset{e}{\cdot}} d^{\overset{m}{\cdot}} n} = z^{\overset{e}{\cdot}} d^{\overset{\overline{m-1}}{\cdot}} n - e \cdot z^{\overset{\overline{e-1}}{\cdot}} d^{\overset{\overline{m-2}}{\cdot}} n \cdot dz + e \cdot e - 1 \cdot z^{\overset{\overline{e-2}}{\cdot}} d^{\overset{\overline{m-3}}{\cdot}} n \cdot \overline{dz}^2$$

27. De surcroît, l'édition de Gerhardt comporte une coquille de plus ; cf. LAA, III.6A, p. 313, LMS, III/1, p. 167.
28. Lettre du 20–30 avril 1695, LAA, III.6A, p. 346–347 = LMS, III/1, p. 171. Ici aussi, il y a des coquilles sur les exposants dans l'édition de Gerhardt absentes de l'édition des académies.
29. Lettre du 6–16 mai 1695, LAA, III.6A, p. 353 = LMS, III/1, p. 175.

$$- e \cdot e - 1 \cdot e - 2 \cdot z^{\frac{e-3}{\cdot}}\, d^{\frac{m-4}{\cdot}}\, n \cdot \overline{dz}^{3} + \text{ etc.}$$

Là encore, on passe d'un terme au suivant en sommant les n et en différentiant les z. Tant que n est affecté d'une différentielle, l'intégrer revient à abaisser son ordre de différentiation : on passe de $d^m n$ à $d^{m-1} n$ puis $d^{m-2} n$, etc. ; mais lorsqu'on atteint dn, la sommation doit conduire à n, puis à $\int n$, $\iint n$, et ainsi de suite. Pour pouvoir traiter ce passage des différentielles aux sommes dans le cadre d'une formule unique, Leibniz introduit une innovation notationnelle :

> Etsi autem sic exhauriri m videatur, posito esse integrum affirmativum, non tamen hoc fit, nam $d^0 n = \int^0 n = n$ et $d^{\frac{-1}{\cdot}} n = \int^{\frac{1}{\cdot}} n$ et d^{-2} est $\iint$ seu $\int^2$ [30].

Leibniz illustre cette convention en donnant les premiers termes pour $m = 1$ et $m = 2$, cas où les intégrales apparaissent rapidement [31]. Sans cette innovation notationnelle, Leibniz ne pourrait pas aussi simplement écrire de formule générale, puisque la règle de formation des termes devrait changer après m étapes.

Cette notation, qui repose sur la réciprocité $\int dx = x$, est donc très naturelle dans ce contexte précis, puisqu'elle permet de traiter tous les cas de manière uniforme et ainsi à écrire une formule simple et générale. Cette innovation s'inscrit parfaitement dans une préoccupation méthodologique constante de Leibniz : la recherche de notations générales permettant de traiter des cas en apparence différents de manière uniforme [32].

Toutefois, à première vue, cette nouvelle notation n'est vraiment utile que dans certains cas particuliers, à savoir les situations où l'on manipule les opérateurs d et $\int$ en leur appliquant d'autres opérateurs d ou $\int$ et où l'on se sert de la réciprocité entre d et $\int$. Pourtant, Leibniz semble voir dès cette première occurence la promesse d'une unité plus profonde et plus substantielle entre sommes et différences, qui permettrait d'obtenir des formules générales contenant des d^n, valables tant pour des différences (n positif) que pour des sommes (n négatif). En ce sens, Leibniz écrit [33] :

---

30. « Quoiqu'il semble qu'ainsi m s'épuise – on a posé m entier positif – ce n'est en fait pas le cas : car $d^0 n = n$ et $d^{-1} n = \int^1 n$ et d^{-2} est $\iint$ ou $\int^2$ » (Lettre du 28 février 1695, LAA, III.6A, p. 313 = LMS, III/1, p. 167).

31. Les formules de Leibniz sont évidemment fausses, puisqu'il les déduit de sa formule générale. Pour $m = 1$, la formule correcte est par exemple

$$\int \overline{z^e dn} = z^e n - e \cdot z^{\frac{e-1}{\cdot}} \int \overline{n} + e \cdot e - 1 \cdot z^{\frac{e-2}{\cdot}} \iint n - e \cdot e - 1 \cdot e - 2 \cdot z^{\frac{e-3}{\cdot}} \iiint n \text{ etc.}$$

qui – contrairement à celle que donne Leibniz – donne lieu à une série *finie* si e est un entier positif, parce que le produit $e(e-1)(e-2)\ldots$ finit par s'annuler. Dans le cas $e = 1$ correspondant à l'intégrale de Bernoulli, on obtient par exemple $\int z dn = zn - \int n$, et non une série infinie comme Bernoulli. La raison en est que Leibniz calcule l'intégrale pour dz constant, et non dn constant comme il faudrait le faire pour trouver la série de Bernoulli (voir aussi note 26 ci-dessus). Cette différence a peut-être échappé à Leibniz, parce que sa formule originale (fausse) contient e^2 au lieu de $e(e-1)$ et ainsi de suite, et conduit donc toujours à une série infinie.

32. Voir *infra*, section 3.4.

33. Encore une fois, la formule que donne Leibniz est fausse ; pour la corriger, il faut décaler les exposants

> Unum addam quod etiam hanc reciprocationem confirmat, si ponamus in aequ. 1 [34] m esse numerum negativum seu $m = -r$ fore $d^m = \int^r$, unde ex aequ. 1 fiet, (posita $dz = 1$) $\int \overline{z^e \int^{\frac{r}{\cdot}} \overline{n}} = z^{\frac{e}{\cdot}} \int^{\frac{r}{\cdot}} n - e\cdot z^{\frac{e-1}{\cdot}} \int^{\frac{r+1}{\cdot}} n + ee\cdot z^{\frac{e-2}{\cdot}} \int^{\frac{r+2}{\cdot}} n - e^3\cdot z^{\frac{e-3}{\cdot}} \int^{\frac{r+3}{\cdot}} n$, etc. [35]

Il essaie donc immédiatement de remplacer les exposants positifs de d par des exposants négatifs dans sa formule précédente. (Il ne justifie pas, mais la méthode de démonstration suivie reste valable : on passe toujours d'un terme au suivant en intégrant les expressions en n et en différentiant les expressions en z.) Dans le même esprit, il envoie peu après sa formule (fausse [36]) pour $\int z^e d^m n$ au marquis de l'Hôpital et lui écrit :

> En donnant la methode des Differences dans vostre écrit [37], vous donnerés Monsieur la Methode des sommes virtuellement ; et en effect je ne distingue pas ces deux calculs. […] Il ne sera point necessaire aussi, que vous vous borniés aux seules differences puisque leur calcul est le meme avec celuy des sommes, l'un estant seulement reciproque de l'autre. Par exemple j'ay trouvé comme x^{-1} est $= 1 : x$ que de même $d^{\frac{-1}{\cdot}} x = \int x$. Par exemple, ayant trouvé cette equation generale
>
> $$\int \overline{z^{\frac{e}{\cdot}} d^m n} = z^e d^m n - ez^{\frac{e-1}{\cdot}} d^{\frac{m-1}{\cdot}} n + eez^{\frac{e-2}{\cdot}} d^{\frac{m-2}{\cdot}} n - e^3 z^{\frac{e-3}{\cdot}} d^{\frac{m-3}{\cdot}}, \text{ etc.}$$
>
> (supposant que dz est l'unité) et faisant specialement $m = 1$, il en proviendra cette equation
>
> $$\int \overline{z^e dn} = z^e dn - ez^{\frac{e-1}{\cdot}} n + eez^{\frac{e-2}{\cdot}} \int \overline{n} - e^3 z^{\frac{e-3}{\cdot}} \iint n, \text{ etc.}$$
>
> Car $d^0 n = n$ et $d^{-1} n = \int n$ et $d^{-2} n = \iint n$ ou $\int^2 n$, c'est-à-dire $\int \overline{\int \overline{ndz} dz}$. Si m estoit 2, $d^m n$ seroit ddn, $d^{\frac{m-1}{\cdot}} n$ seroit dn, $d^{\frac{m-2}{\cdot}} n$ seroit n, $d^{\frac{m-3}{\cdot}} n$ seroit $\int n$, et $d^{\frac{m-4}{\cdot}} n$ seroit $\iint n$, etc. [38]

Les affirmations de Leibniz semblent ici aller au-delà de ce que ses découvertes lui per-

---

des intégrales et remplacer ee par $e \cdot e - 1$, etc.

34. Il s'agit de l'expression générale de $\int \overline{z^e d^m n}$ ci-dessus.

35. « J'ajouterai une chose qui confirme encore cette réciprocité : si, dans l'équation (1) nous posons m négatif, soit $m = -r$, nous aurons $d^m = \int^r$, d'où l'équation (1) fera (en posant $dz = 1$) […] » (Lettre du 28 février 1695, LAA, III.6A, p. 314 = LMS, III/1, p. 168.)

36. Il rectifiera selon la correction faite par Bernoulli dans une lettre ultérieure du 12–22 juillet 1695 (préservée en deux morceaux ; voir LAA, III.6B, p. 450 et LMS, II, p. 294).

37. Leibniz fait référence au futur manuel de calcul infinitésimal de l'Hôpital, qui sera publié en 1696 sous le titre *Analyse des infiniment petits pour l'intelligence des lignes courbes* (L'Hôpital 1696).

38. Lettre de Leibniz à l'Hôpital non datée, mais probablement du 18 mars 1695 (LAA, III.6A, p. 315–316 = LMS, II, p. 274–275).

mettent : si la réciprocité entre sommes et différences permet avec profit d'écrire d^{-1} pour $\int$, les seules formules dans lesquelles la substitution est possible sont la série infinie de Leibniz et ses variantes.

Pour Leibniz, le succès partiel de la notation $d^{-1} = \int$ soulève donc avant tout le problème de l'unification des calculs des différences et des sommes : la question qui dominera la suite de notre épisode, à savoir celle de l'analogie entre le comportement des différences et celui des puissances, est à ce stade absente. Il est vrai qu'il peut être tentant de la voir déjà ici, parce que la notation $d^{-1} = \int$, en mettant l'accent sur la *réciprocité* entre différences et sommes, peut rappeler l'analogie remarquée par Leibniz entre les paires puissances-sommes et différences-racines[39]. Mais il faut se garder d'aller trop vite. En réalité, le d^{-1} s'aligne assez mal avec cette idée de départ, puisque en algèbre, la puissance -1 correspond à l'inverse et non à la racine.

2.3 L'analogie des puissances et des différences de Leibniz

Comme nous l'avons vu, Leibniz envoie le 28 février 1695 la formule générale fausse qu'il tire de la méthode de Bernoulli, et celui-ci lui répond le 20–30 avril 1695 en corrigeant les calculs. C'est dans la lettre suivante de Leibniz, datée du 6–16 mai 1695, qu'apparaît l'« analogie des puissances et des différences » :

> Recte correxisti calculum meum. Nam dum festinabundus in chartam conjicio, quod literas scribenti calculus suggerit, errorem admisi, seriemque male expressi. Multa adhuc in istis summarum et differentiarum progressionibus latent, quae paulatim prodibunt. Ita notabilis est consensus inter numeros potestatum a binomio, et differentiarum a rectangulo[40] et puto nescio quid arcani subesse. Ex. gr.

$$
\begin{array}{ll}
\boxed{1}, x+y = 1x+1y & d^1, xy = 1ydx + 1xdy \\
\quad \text{ou} \quad 1x^1y^0 + 1x^0y^1 & \quad \text{ou} \quad 1d^1xd^0y + 1d^0xd^1y \\
\boxed{2}, x+y = 1x^2 + 2xy + 1y^2 & d^2, xy = 1yddx + 2dydx + 1xddy \\
\boxed{3}, x+y = 1x^3 + 3x^2y + 3xy^2 + 1y^3 & d^3, xy = 1yd^3x + 3dyd^2x + 3d^2ydx + 1xd^3y \\
\boxed{4}, x+y = 1x^4 + 4x^3y + 6x^2y^2 & d^4, xy = 1yd^4x + 4dyd^3x + 6d^2xd^2y \\
\qquad +4xy^3 + 1y^4 & \qquad +4d^3ydx + 1xd^4y
\end{array}
$$

> Et ita porro. Ubi perfectissimus est consensus. Nempe ubi ab una parte ponitur

39. Cf. 1.4.
40. Les différences d'un rectangle, c'est-à-dire d'un produit de deux facteurs.

> $x^m y^n$, ab altera ponitur $d^m x d^n y$. Ita respondent sibi x^2 et $yddx$. Nam x^2 est $x^2 y^0$ et $ddx\ y$ est $d^2 x d^0 y$. Nam $d^0 y = y$. Atque ita realis quidam consensus inter potentiarum indices seu Logarithmos, et nostros differentialium quasi Logarithmos reperitur, qui etiam ad polynomia et multi-rectangula, seu rectangula solida et supersolida [41] extenditur [...] [42].

D'une simple notation commune – l'emploi dans les deux cas d'exposants – on passe ici, d'après Leibniz, à un « accord réel ».

Quel est le lien entre les réflexions qui précèdent et cette analogie ? Quelles sont les « progressions de sommes et de différences » dont parle Leibniz ?

Une première hypothèse [43] est que les « progressions » en question, ce seraient par exemple

$$\iint x = d^{-2}x, \quad \int x = d^{-1}x, \quad x = d^0 x, \quad d^1 x, \quad d^2 x, \quad d^3 x, \quad \text{etc.}$$

ce qui dans cette écriture exponentielle est analogue à une suite (ou « progression ») géométrique comme

$$a^{-2}, \quad a^{-1}, \quad 1 = a^0, \quad a = a^1, \quad a^2, \quad a^3, \quad \text{etc.}$$

La réciprocité différences-sommes, ainsi revenue sur le devant de la scène *via* la notation $d^{-1} = \int$, aurait donc rappelé à Leibniz son idée d'une analogie entre les paires d'opérations réciproques différences-sommes et puissances-racines, et l'aurait conduit à l'approfondir. Je trouve, toutefois, que c'est aller un peu vite. Comme je l'ai déjà signalé, l'analogie de Leibniz entre les deux réciprocités ne s'aligne pas si bien avec sa nouvelle notation d^{-1} : prendre la racine carrée, dans une progression géométrique comme la précédente, c'est remonter à a de a^2, à a^3 de a^6, etc. donc prendre la puissance $\frac{1}{2}$; sommer, en revanche,

41. Comme Leibniz l'explique ensuite, il s'agit de comparer par exemple $\boxed{m}\,\overline{x+y+z}$ et $d^m xyz$, et de même en dimensions supérieures.

42. « Tu as bien corrigé mon calcul. Pendant que je jetais en hâte sur le papier ce que me suggérait le calcul en écrivant la lettre, j'ai effectivement commis une erreur et mal exprimé la série. Il reste encore beaucoup de choses cachées dans ces progressions de sommes et de différences et elles apparaîtront peu à peu. Il faut noter par exemple l'accord entre les puissances d'un binôme et les différences d'un rectangle ; je pense qu'il y a là-dessous je ne sais quoi de secret. Par exemple : [...] et ainsi de suite, où l'accord est total. De fait lorsque d'un côté on pose $x^m y^n$ on a de l'autre $d^m x d^n y$. C'est ainsi que x^2 et $yddx$ se correspondent, puisque x^2 est $x^2 y^0$ et que $yddx$ est $d^0 y d^2 x$. Car $d^0 y = y$. On découvre ainsi un certain accord réel entre les indices des puissances, c'est-à-dire les logarithmes, et nos quasi-logarithmes différentiels ; accord qui s'étend même aux polynômes et aux produits de termes plus nombreux, c'est-à-dire solides et supra-solides [...]. » (Lettre du 6–16 mai 1695, LAA, III.6A, p. 353–354 = LMS, III/1, p. 175. Traduction en partie reprise de LNC, p. 410–411.)

43. C'est plus ou moins, si je le comprends bien, ce que suggère Marc Parmentier dans ses notes sur le mémoire Leibniz 1710b (cf. LNC, p. 410).

c'est se décaler d'un cran vers la gauche, c'est-à-dire prendre la puissance -1. Un second problème avec cette hypothèse est qu'elle n'explique pas comment Leibniz serait passé de l'idée très générale d'une analogie différences-puissances à une comparaison entre les différentielles d'un *produit de deux termes* et les puissances d'une *somme de deux termes*. Il faudrait supposer qu'il y serait arrivé en expérimentant au hasard, ou peut-être qu'il avait déjà remarqué l'accord entre les coefficients (ou le nombre de termes) de $dd(xy)$ et $(x+y)^2$ etc., et que la notation nouvellement acquise $x = d^0 x$ lui a seulement permis de le préciser.

Je voudrais proposer une autre hypothèse, qui, si elle n'exclut pas nécessairement la précédente, du moins la complète et la précise. Il est naturel, dans le contexte de la lettre de Leibniz, de supposer que ces « progressions de sommes et de différences » désignent les successions des termes des séries infinies précédemment obtenues par Leibniz et Bernoulli, en particulier la dernière en date :

$$\int \overline{z^{\overset{e}{\cdot}} d^{\overset{m}{\cdot}} n} = z^{\overset{e}{\cdot}} d^{\frac{m-1}{\cdot}} n - e \cdot z^{\frac{e-1}{\cdot}} d^{\frac{m-2}{\cdot}} n \cdot dz + e \cdot e - 1 \cdot z^{\frac{e-2}{\cdot}} d^{\frac{m-3}{\cdot}} n \cdot \overline{dz}^2$$
$$- e \cdot e - 1 \cdot e - 2 \cdot z^{\frac{e-3}{\cdot}} d^{\frac{m-4}{\cdot}} n \cdot \overline{dz}^3 + \text{ etc.}$$

Quand on passe d'un terme au suivant dans cette formule, l'exposant de différentiation de n diminue, tandis que celui de z^e augmente (on passe de $d^m n$ et z^e à $d^{m-1} n$ et $d(z^e)$ – d'où un dz – puis à $d^{m-2} n$ et $d^2(z^e)$ – d'où un dz^2 – puis encore à $d^{m-3} n$ et $d^3(z^e)$ – d'où un dz^3 – et ainsi de suite). Ce type de « progression », où un exposant diminue et l'autre augmente, est caractéristique de la formule du binôme de Newton, que Leibniz connaît bien et utilise régulièrement. Pour rappel [44], il l'écrit par exemple comme ceci dans un manuscrit : pour ce que nous noterions $(A+B)^z$, il donne

$$A^z + \frac{z}{1} A^{z-1} B^1 + \frac{z, z-1}{1, 2} A^{z-2} B^2 + \frac{z, z-1, z-2}{1, 2, 3} A^{z-3} B^3 \text{ etc.}$$

Il ne fait pas de doute que Leibniz avait cette formule en tête. En effet, comme nous l'avons vu, il avait déjà repéré l'analogie suivante entre les problèmes posés par les racines et par les sommes : dans les deux cas, on ne peut pas toujours les exprimer en quantités « ordinaires » (nombres rationnels dans un cas, expressions algébriques finies dans l'autre), mais il faut se tourner vers des séries infinies. Pour ce qui est des racines, Newton, en s'appuyant sur une notation bien choisie, avait pu exprimer ces séries infinies par une formule générale. Leibniz, qui vient d'introduire une innovation notationnelle similaire, se demande très certainement s'il n'est pas possible d'écrire une formule analogue pour les

44. Cf. section 1.4, en part. p. 40.

sommes. La similarité de structure entre la formule du binôme et les développements en série obtenus avec Bernoulli ne pouvait que le renforcer dans cette idée. Or la formule de Bernoulli se présente, précisément, comme la somme (ou d^{-1}) d'un produit de deux termes. Le chemin qui conduit à l'analogie entre les puissances d'une somme (formule du binôme) et les différentielles d'un produit est alors clair.

Selon mon hypothèse, Leibniz est donc guidé par l'analogie de structure entre la formule du binôme et les développements d'intégrale en séries infinies découverts avec Bernoulli (et pas seulement par l'idée générale d'une analogie entre les réciprocités différences-sommes et puissances-racines). Sur cette base, il aurait conçu dès cette lettre le projet d'une version différentielle et intégrale de la formule du binôme, valable tant pour les exposants positifs (cas des différentielles d'ordre supérieur, qu'il traite ici) que pour les entiers négatifs, correspondant aux intégrales : son but aurait été de déduire les développements de $\int ndz$ et $\int z^e d^m n$ tout juste découverts d'une loi générale analogue à celle de Newton pour les puissances. Ce qu'il écrit après son analogie semble le confirmer :

> Imo videndum an non in summationibus concipere aliquid liceat respondens radicibus irrationalibus, imo affectis [45].

On voit bien que Leibniz a en tête le parallèle avec les racines. Voici, je crois, son idée : de même qu'on peut obtenir par la formule du binôme un développement en série pour les racines irrationnelles, de même est-il peut-être possible de donner une formule pour développer toute intégrale en série, dont les formules obtenues dans les lettres précédentes ne seraient que des cas particuliers. On peut même, suggère-t-il, imaginer aller encore plus loin, jusqu'aux « racines affectées ». Que veut-il dire ? J'ai mentionné plus haut [46] que dans l'*Epistola prior* de 1676, Newton avait envoyé à Leibniz non seulement la formule du binôme, mais aussi une méthode pour développer en série les racines d'« équations affectées », c'est-à-dire de polynômes. Sans doute Leibniz rêve-t-il ici à une méthode permettant de développer en série non seulement les sommes, c'est-à-dire les quantités y données par $dy = \ldots$, mais aussi les solutions d'équations différentielles générales, c'est-à-dire les quantités y données par une équation plus complexe en y, dy, ddy etc., qui seraient l'analogue différentiel des « racines affectées ». Cela donnerait une nouvelle méthode de résolution de ce que j'ai appelé plus haut le problème inverse général [47], que Leibniz ne sait jusqu'alors traiter que par la méthode des coefficients indéterminés.

45. « Et même, il faut voir s'il ne se laisse pas concevoir pour les sommes quelque chose de correspondant aux racines irrationnelles, ou même affectées. » (Lettre du 6–16 mai 1695, LAA, III.6A, p. 354 = LMS, III/1, p. 175.)

46. Voir toujours section 1.4, en part. note 17.

47. Cf. chapitre précédent, en part. sections 1.3 et 1.5.

Leibniz, donc, aurait d'emblée espéré une formule générale pour les sommes et différences récapitulant les divers procédés trouvés jusqu'alors ; mais dans un premier temps, il n'aurait réussi à dégager une analogie que pour les différences. Ce n'est que quelques mois plus tard, en septembre 1695, qu'il parvient à ce qui est, d'après moi, son but initial. Au marquis de l'Hôpital, qui voulait en apprendre davantage sur l'analogie dont Bernoulli lui avait apparemment parlé [48], Leibniz envoie ce qui devait apparemment être une courte introduction, commencée dans une marge, mais qui va finalement plus loin et comprend des résultats nouveaux [49]. Tout d'abord, il présente la même chose qu'à Bernoulli, quoique avec une notation nouvelle pour les puissances ($p^e x$ pour x^e) certainement destinée à rendre l'analogie plus visible. Il étend ensuite l'analogie au cas de la puissance -1, qui le conduit à la « série universelle » de Bernoulli, puis passe finalement au cas général : de même que

$$\begin{aligned} p^e\overline{x+y} &= x^e + \frac{e}{1}x^{e-1}y^1 + \frac{e\cdot e-1}{1\cdot 2}x^{e-2}y^2 + \frac{e\cdot e-1\cdot e-2}{1.2.3}x^{e-3}y^3 \text{ etc.} \\ &= p^e x\cdot p^0 y + \frac{e}{1}p^{e-1}x\cdot p^1 y + \frac{e\cdot e-1}{1\cdot 2}p^{e-2}x\cdot p^2 y + \frac{e\cdot e-1\cdot e-2}{1\cdot 2\cdot 3}p^{e-3}x\cdot p^3 y \text{ etc.} \end{aligned}$$

on a

$$\begin{aligned} d^e\overline{xy} &= d^e x d^0 y + \frac{e}{1}d^{e-1}x\cdot d^1 y + \frac{e\cdot e-1}{1\cdot 2}d^{e-2}x\cdot d^2 y + \frac{e\cdot e-1\cdot e-2}{1\cdot 2\cdot 3}d^{e-3}x\cdot d^3 y \text{ etc.} \\ &= d^e x\cdot y + \frac{e}{1}d^{e-1}x\cdot dy + \frac{e\cdot e-1}{1\cdot 2}d^{e-2}x\cdot ddy \text{ etc.} \end{aligned}$$

On peut supposer que cette lettre à l'Hôpital a été l'occasion pour Leibniz de reprendre le fil de ses réflexions, et de finalement réussir à obtenir une formule générale, valable aussi pour les intégrales, sur le modèle de celle du binôme ; en tout cas, dès sa lettre suivante à Bernoulli, Leibniz lui envoie cette version généralisée de l'analogie, ce qui semble confirmer qu'il s'agit d'un résultat nouveau [50].

Cette lettre à l'Hôpital (et celle du 20–30 octobre à Bernoulli contenant le même résultat) réalise l'ambition de départ de Leibniz et semble clore pour lui cette ligne d'investigation. À ma connaissance, il ne développe plus l'analogie ; il se contente de la publier en 1710, sous une forme restreinte aux exposants entiers positifs et soigneusement justifiée [51]. Ce texte paraît dans le premier numéro des *Miscellanea Berolinensia*, le journal nouvellement créé de l'académie des sciences de Berlin ; il est probable que Leibniz ait souhaité contribuer à son lancement et ait eu l'idée d'y envoyer des résultats plus anciens, mais encore jamais

48. Peut-être dans la lettre de Jean Bernoulli à l'Hôpital du 21 août 1695, qui est perdue (voir BBW, I, p. 303, 306).
49. Lettre du 30 septembre 1695 (nouveau style), LAA, III.6B, p. 508–510 = LMS, II, p. 300–302.
50. Lettre du 20–30 octobre 1695, LAA, III.6B, p. 528 = LMS, III/1, p. 221.
51. Leibniz 1710b.

publiés [52]. Toutefois, malgré l'absence de résultats nouveaux, cet article apporte quelques éléments intéressants.

Tout d'abord, Leibniz le commence en rappelant le parallèle, déjà longuement discuté, entre les réciprocités différences-sommes et puissances-racines, mais souligne bien que l'analogie qu'il veut présenter est différente et « plus cachée » (« *arcanior* ») :

> Ut cujuslibet quantitatis facile est invenire potentiam, ita cujuslibet certa lege variantis possumus invenire differentiam seu Elementum. Sed regressus a potentia ad radicem per extractionem, et regressus a differentia ad terminum per summationem non semper in potestate est. Et uti impossibilitas extractionis in numeris rationalibus quasitae producit quantitates surdas, ita impossibilitas summationis in quantitatibus Algebraicis quasitae producit quantitaes transcendentes, quarum considerationem in Analysin jam olim induximus. […] Sed arcanior quaedam subest inter Potentias et Differentias Analogia, quam hoc loco exponere operae pretium erit. Et primum potentias binomii (seu summae nominum duorum) comparabimus cum differentiis rectanguli (seu facti ex Factoribus binis) […] [53].

En second lieu, Leibniz reprend l'innovation notationnelle ($p^e x$ pour x^e) qu'il avait introduite dans sa lettre à l'Hôpital, en la justifiant explicitement par le souhait de rendre l'analogie plus claire (« *analogiæ clarioris causa* [54] ») ; il précise d'ailleurs plus loin que « par cette nouvelle manière d'écrire, l'analogie entre les puissances et les différences apparaît clairement, alors que par la manière commune ce n'est pas le cas [55] », faisant référence tant à sa notation $p^e x$ pour x^e qu'au fait d'écrire soigneusement $d^0 x$ pour x, $d^1 x$ pour dx et $d^2 x$ pour ddx.

52. Il publie d'ailleurs dans le même numéro un article mineur qui passe en revue diverses formes de notation en algèbre : Leibniz 1710a.

53. « De même qu'il est facile de prendre une puissance de n'importe qu'elle quantité, de même nous pouvons trouver la différence, c'est-à-dire l'Élément, de toute quantité variant selon une loi déterminée. Mais le retour d'une puissance à sa racine par extraction, et le retour d'une différence à son terme par sommation, ne sont pas toujours en notre pouvoir. Et de même que l'impossibilité de l'extraction [de racine], si elle est cherchée en nombres rationnels, produit les quantités sourdes [irrationnelles], de même l'impossibilité de la sommation, si elle est cherchée en quantités algébriques, produit les quantités transcendantes, dont nous avons il y a déjà longtemps introduit la considération en Analyse. […] Mais il existe une certaine Analogie plus cachée entre les Puissances et les Différences, qui devrait valoir la peine d'être exposée ici. Tout d'abord, nous comparerons les puissances d'un binôme (ou somme de deux noms) avec les différences d'un rectangle (ou produit de deux facteurs) […]. » (Leibniz 1710b, p. 160-161 = LMS, V, p. 724–725 ; trad. adaptée de LNC, p. 414-415.)

54. Leibniz 1710b, p. 161.

55. « […] patet, novo scribendi more apparere analogiam inter potentias et differentias, vulgato […] non apparere. » Leibniz 1710b, p. 165.

Troisièmement, Leibniz discute en quelques lignes la « loi transcendantale des homogènes » (« *lex homogeneorum transcendentalis* ») mentionnée plus haut, que nous n'avons pas rencontrée dans ses lettres avec Bernoulli. Elle s'intègre sans mal dans son article, puisqu'elle aussi bénéficie des notations p^e et d^e : comme l'écrit Leibniz, « la notation courante ne permet pas de saisir [cette loi] aussi bien [56] ». Cependant, il faut bien voir que Leibniz la connaissait déjà en 1690 ; peut-être est-ce même elle qui l'a initialement motivé à écrire d^3 pour ddd [57]. Quoi qu'il en soit, elle est largement indépendante de ses découvertes ultérieures sur les puissances et les différences, et en fait ne s'harmonise qu'imparfaitement avec elles : la dissymétrie entre sommes (notées par des puissances négatives) et racines (notées par des exposants fractionnaires) fait qu'elle ne se généralise pas de manière transparente.

Enfin, le fait de se limiter aux exposants positifs permet à Leibniz de justifier l'analogie *via* un parallélisme entre le passage d'une puissance de $x + y$ à la suivante d'une part, le passage d'une différentielle de xy à la suivante d'autre part :

> Quae analogia inter differentiationem & potentiationem servatur perpetuò, continuata potentiatione (seu Potentiæ excitatione) & differentiatione. Nempe ut in nova potentiatione Binomii totum præcedens multiplicatur tam per y quam per x, & priore casu p ipsius y, posteriore p ipsius x, augetur unitate ; ita in differentiando totum præcedens differentiatur tam secundum y quam secundum x, & priore casu d ipsius y, posteriore autem d ipsius x augetur unitate [...] [58].

Leibniz peut ensuite ramener les coefficients de la formule commune pour $d^e(xy)$ et $p^e(x+y)$ à des considérations combinatoires sur le nombre de termes de chaque sorte qui sont produits par un développement exhaustif. C'est cela qui permet à Couturat d'écrire :

> [À] quoi tient cette analogie frappante, qui paraît d'abord à Leibniz et à Bernoulli quelque chose de secret et de mystérieux ? Simplement à ce fait que les règles formelles des deux opérations (élévation aux puissances et différentiation) sont les mêmes, si différentes que soient en elles-mêmes ces opérations.

56. « Eâdem etiam opera apparet quænam sit *Lex homogeneorum transcendentalis*, quam vulgari modo scribendi differentias, non æque agnoscas. » Leibniz 1710b, p. 165.

57. Voir section 2.1, p. 1.

58. « Cette analogie entre la différentiation et l'élévation à une puissance se perpétue si nous poursuivons cette dernière (c'est-à-dire si nous augmentons la puissance), ainsi que la différentiation. En effet, de même que pour obtenir une nouvelle puissance d'un binôme, on multiplie tout ce qui précède tant par y que par x, et que le p de x dans le premier cas, le p de y dans le second cas sont augmentés d'une unité ; de même, dans la différentiation tout ce qui précède est différentié tant selon y que selon x, et le d de y dans le premier cas, le d de x dans le second cas sont augmentés d'une unité. » Leibniz 1710b, p. 162-163, trad. LNC, p. 417 (modifiée).

> Et cette analogie formelle vient, en dernière analyse, de ce que l'une et l'autre obéissent aux mêmes lois générales de combinaison [...]. Cela montre bien que la Combinatoire est plus générale que l'Algèbre, puisqu'on voit les mêmes formules de combinaison s'appliquer à divers calculs hétérogènes [59].

Que Leibniz identifie un principe combinatoire commun aux deux calculs, c'est indéniable. C'est souvent cet aspect, il est vrai remarquable, qui a le plus frappé les commentateurs. Lorsque Hourya Benis-Sinaceur cherche à comprendre comment Gödel a pu voir en la théorie des modèles robinsonienne une réalisation de l'*ars inveniendi* leibnizien, notre exemple est l'un des premiers qu'elle met en avant :

> Ainsi les opérations d'élévation à une puissance et de différentiation, en elles-mêmes distinctes, peuvent s'écrire de même façon. Calcul algébrique et calcul différentiel ont des langages, et donc des structures parallèles. Or, on a là le germe de la démarche de la théorie des modèles : constituer une métathéorie de théories mathématiques par l'analyse de leur caractéristique, c'est-à-dire de leur langage [60].

Sur un point, toutefois, Couturat va trop loin : la combinatoire ne lève pas le mystère et n'apporte aucune clôture à l'investigation. Sans doute se laisse-t-il tromper par la date tardive du mémoire de 1710. Leibniz, pourtant, a pleinement compris le contenu de ce mémoire dès sa lettre du 28 février 1695 : il y expose déjà la correspondance entre les puissances d'un binôme et les différences d'un rectangle, sa généralisation aux trinômes ou davantage, et le lien avec la combinatoire. Or c'est précisément dans cette lettre qu'il écrit : « je pense qu'il y a là-dessous je ne sais quoi de secret [61] »... En 1710, ce n'est plus une hypothèse : il sait que l'analogie se généralise aux sommes et déborde donc le cadre de la petite justification combinatoire qu'il en donne ici. Le cas des puissances positives n'est qu'une manifestation élémentaire d'un phénomène plus profond, pas encore tout à fait clair ; ce n'est qu'un point de départ, pas un aboutissement, et certainement pas l'accomplissement d'un programme de réduction de l'algèbre à la combinatoire. Leibniz peut bien avoir des rêves plus généraux ; comme l'écrit Benis-Sinaceur, l'essentiel reste pour lui « de bâtir le plus possible sans attendre d'avoir les moyens de parachever l'analyse ou de garantir définitivement les fondements [62] ».

59. Couturat 1901, p. 498

60. Benis-Sinaceur 1989, p. 209 (cf. aussi Benis-Sinaceur 1988a,b qui contiennent d'autres versions du même matériau).

61. « puto nescio quid arcani subesse ». Voir ci-dessus, p. 51.

62. Benis-Sinaceur 1989, p. 209.

2.4 Le « genre singulier de calcul » de Bernoulli

Si Leibniz semble satisfait des résultats obtenus et ne cherche pas à les approfondir davantage, Bernoulli, en revanche, tente d'aller plus loin : au cours de l'été 1695, avec les encouragements de Leibniz, il tire de l'analogie une méthode d'intégration suprenante.

L'idée de départ est que les différentielles successives dx, d^2x, d^3x etc. d'une quantité x sont analogues à ses puissances x, x^2, x^3 etc. À ce stade, cette analogie est toutefois très limitée. Elle a été introduite comme simple notation, pour permettre d'exprimer des différentielles de n'importe quel ordre ; elle a ensuite été étendue aux puissances négatives pour des raisons locales, pour écrire de manière uniforme une certaine formule particulière ; mais elle n'a encore qu'une seule confirmation substantielle : les différentielles d'un produit ont la même forme que les puissances d'une somme. Bernoulli tente cependant de la prendre au sérieux, et traite d^n exactement comme si c'était la puissance d'une quantité usuelle.

Plus précisément, dans sa lettre du 8–18 juin 1695, Bernoulli introduit les règles de calcul suivantes. Le principe est de traiter d^n comme une puissance, en ignorant la variable différentiée (sachant qu'il faut systématiquement ajouter d^0 quand la variable apparaît seule : par exemple, on remplace x par d^0x) :

> Nota, quod in hoc scrutinio literae ipsae, quae alias quantitatem denotant x, y, non considerandae sunt ut tales, sed dumtaxat quatenus determinant d, d^2, d^3 etc. Hoc modo quadratum ipsius d^3y non est d^3yy, sed d^6y ; cubus ipsius d^3y non d^3y^3, sed d^9y : Idem puta de multiplicatione, divisione et extractione radicum $d^2y \times d^3y = d^5y$, $\sqrt[3]{d^6y} = d^2y$; item $\frac{d^2y}{d^2y} = d^0y = y$, et hac ratione $\frac{x}{x}$ non est $= 1$, sed $= \frac{d^0x}{d^0x} = d^0x = x$; quoniam autem $d^{-m} = \int^{+m}$, erit ex. gr. $\frac{d^4y}{d^5y} = d^{-1}y = \int y$ [...] [63].

Malgré tout, on ne peut pas omettre complètement les variables x, y, etc. : il faut par exemple pouvoir distinguer d^2x de d^2y. En fait – pour reprendre un artifice notationnel proposé par Lenore Feigenbaum [64] – tout se passe comme si on calculait algébriquement avec plusieurs variables distinctes d_x, d_y, etc.

63. « Remarque que dans cette investigation les lettres x, y, qui dénotent habituellement une quantité, ne doivent pas être considérées comme telles, mais uniquement dans la mesure où elles déterminent d, d^2, d^3 etc. Ainsi, le carré de d^3y n'est pas d^3yy mais d^6y ; le cube de d^3y n'est pas d^3y^3 mais d^9y. De même par exemple pour la multiplication, la division et l'extraction de racines : $d^2y \times d^3y = d^5y$, $\sqrt[3]{d^6y} = d^2y$; pareillement $\frac{d^2y}{d^2y} = d^0y = y$, et pour cette raison $\frac{x}{x}$ n'est pas égal à 1, mais à $\frac{d^0x}{d^0x} = x$; d'autre part puisque $d^{-m} = \int^{+m}$, on aura par exemple $\frac{d^4y}{d^5y} = d^{-1}y = \int y$ [...] » (Lettre du 8–18 juin 1695, LAA, III.6A, p. 399 = LMS, III/1, p. 180.)

64. Feigenbaum 1985, p. 88. On trouve en fait la même suggestion chez Couturat (1901, p. 497).

Ces règles de calcul, Bernoulli les emploie dans le cadre d'une méthode d'intégration. Celle-ci repose à nouveau sur un usage créatif de l'analogie de Leibniz. Dans le cas algébrique, x est la « troisième proportionnelle » de x^2 et x^3, c'est-à-dire que

$$\frac{x}{x^2} = \frac{x^2}{x^3},$$

ce qui permet de calculer x ainsi :

$$x = \frac{(x^2)^2}{x^3}.$$

De la même manière, raisonne Bernoulli, « la troisième proportionnelle de d^3 et dd sera d, celle de d^4 et d^3 sera dd, et ainsi des autres [65]. » Étant donnée une quantité différentielle, disons d^3z, on peut alors l'*intégrer*, c'est-à-dire retrouver d^2z, de la manière suivante. Pour commencer, on calcule d^4z, ce qui est facile : il suffit de différentier d^3z. Puis on prend la troisième proportionnelle de d^3z et d^4z, par exemple en écrivant

$$d^2z = \frac{(d^3z)^2}{d^4z}$$

le calcul devant être fait d'après les règles ci-dessus. Dans les termes de Bernoulli :

> [V]idetur tamen quantitatem propositam differentialem cujusvis gradus summari posse, eam primo differentiando, et dein sumendo tertiam proportionalem hujus novae quantitatis differentialis ad differentialem propositam, consideratis interim d, d², d³, d⁴ etc. tanquam quantitatibus algebraicis et non ut literis tantummodo caracteristicis [66].

Reconstruisons en détail le premier exemple donné par Bernoulli, celui de l'intégration de $xd^3y+dxddy = d^0xd^3y+dxd^2y$. Tout d'abord, on différentie : on trouve $2dxd^3y+d^0xd^4y+d^2xd^2y$. La troisième proportionnelle voulue s'écrit donc

$$\begin{aligned}\frac{(d^0xd^3y+dxd^2y)^2}{2dxd^3y+d^0xd^4y+d^2xd^2y} &= \frac{2dxd^5y+d^0xd^6y+d^2xd^4y}{2dxd^3y+d^0xd^4y+d^2xd^2y}\\ &= \frac{d^0xd^2y(2dxd^3y+d^0xd^4y+d^2xd^2y)}{2dxd^3y+d^0xd^4y+d^2xd^2y}\\ &= d^0xd^2y = xddy,\end{aligned}$$

65. « tertia proportionalis d³ ad dd erit d, et d⁴ ad d³ erit dd, et ita de aliis » (Lettre du 8–18 juin 1695, LAA, III.6A, p. 398 = LMS, III/1, p. 180.)

66. « Il semble que l'on puisse sommer une quantité différentielle proposée d'ordre quelconque en la différentiant d'abord puis en prenant la troisième proportionnelle de cette nouvelle quantité différentielle et de la proposée, si d, d^2, d^3 etc. sont pendant ce temps-là considérées comme des quantités algébriques et pas simplement comme des lettres caractéristiques. » (Lettre du 8–18 juin 1695, LAA, III.6A, p. 398 = LMS, III/1, p. 179–80.)

ce qui est bien l'intégrale recherchée puisque $d(xddy) = xd^3y + dxddy$. Cependant, Bernoulli ne donne pas les détails du calcul et il s'y cache une difficulté : la simplification qui conduit à la dernière ligne n'est pas très claire. Au numérateur, pourquoi factoriser par d^0xd^2y plutôt que par d^2y seul, ce qui conduirait à l'intégrale fausse ddy ? La réponse, sans doute, est que la fraction $\frac{2dxd^3y+d^0xd^4y+d^2xd^2y}{2dxd^3y+d^0xd^4y+d^2xd^2y}$ ne doit pas être simplifiée en 1, comme dans un calcul algébrique usuel, mais en d^0xd^0y. Cette simplification exotique est dans l'esprit des règles données par Bernoulli : pour mémoire, « $\frac{d^2y}{d^2y} = d^0y = y$, et pour cette raison $\frac{x}{x}$ n'est pas égal à 1, mais à $\frac{d^0x}{d^0x} = x$ ». Cependant, elle ne s'en déduit pas. Faut-il comprendre que lorsqu'on simplifie une expression, le résultat doit faire apparaître un d^0z pour chaque variable z qui y apparaît ? En fait, cette difficulté en cache de plus profondes, auxquelles je reviendrai.

Bernoulli applique ensuite sa méthode à un cas plus intéressant : une expression non intégrable de manière finie, à savoir $xd^3y+2dxddy$ (c'est-à-dire $d^0xd^3y+2d^1xd^2y$). Comme précédemment, il commence par différentier l'expression : on trouve $d^0xd^4y + 3dxd^3y + 2d^2xd^2y$. Puis il cherche à calculer

$$\frac{(d^0xd^3y + 2d^1xd^2y)^2}{d^0xd^4y + 3dxd^3y + 2d^2xd^2y} = \frac{d^0xd^6y + 4d^1xd^5y + 4d^2xd^4y}{d^0xd^4y + 3dxd^3y + 2d^2xd^2y}.$$

À ce stade, il ne peut plus continuer comme ci-dessus, parce que cette fraction n'est pas simplifiable. Bernoulli procède alors à une « division continue », c'est-à-dire qu'il applique ce que nous appelons l'algorithme de division polynômiale[67]. Il n'explicite pas le calcul, mais pour clarifier le procédé, voici les premières étapes :

$$\begin{array}{l|l}
d^0xd^6y + 4d^1xd^5y + 4d^2xd^4y & d^0xd^4y + 3d^1xd^3y + 2d^2xd^2y \\
\cline{2-2}
\underline{-d^0xd^6y - 3d^1xd^5y - 2d^2xd^4y} & d^0xd^2y + dxdy - d^2xd^0y + d^3xd^{-1}y \ldots \\
0 \quad + dxd^5y + 2d^2xd^4y & \\
\quad \underline{- dxd^5y - 3d^2xd^4y - 2d^3xd^3y} & \\
\quad 0 \quad - d^2xd^4y - 2d^3xd^3y & \\
\qquad \underline{+2d^2xd^4y + 3d^3xd^3y + 2d^4xd^2y} & \\
\qquad 0 \quad + d^3xd^3y + 2d^4xd^2y & \\
\qquad \qquad \ldots &
\end{array}$$

(Par souci de clarté, j'ai ici choisi la disposition scolaire française pour les divisions : le

67. Au XVIIe siècle, la méthode consistant à appliquer indéfiniment l'algorithme de division à des expressions symboliques pour obtenir des séries infinies est souvent appelée « méthode de Mercator » (en référence à Nicholas Mercator 1668, qui l'a introduite). Elle est bien connue et souvent utilisée. La formule du binôme de Newton (dans le cas d'exposants négatifs) la subsume ; cf. note 30.

dividende est en haut à gauche, le diviseur en haut à droite et le quotient juste en-dessous du diviseur [68].) Le principe est le même que pour la division polynômiale. On divise le premier terme du dividende, d^0xd^6y, par le premier terme du diviseur, d^0xd^4y : on obtient le premier terme du quotient, d^0xd^2y. On multiplie ce premier terme par le diviseur tout entier, et on soustrait le résultat au dividende. Le reste obtenu remplace le diviseur de départ, et on recommence. Bernoulli trouve

$$\begin{aligned} &d^0xddy + dydx - d^0yddx + d^{-1}yd^3x - d^{-2}yd^4x + d^{-3}yd^5x \text{ etc.} \\ &\qquad = xddy + dydx - yddx + \int yd^3x - \iint yd^4x + \int^3 yd^5x \text{ etc.} \end{aligned}$$

Pour finir, il présente une variante de ce calcul. Remarquons que dans la division précédente, les termes du dividende et du diviseur sont ordonnés selon les puissances décroissantes de y ; en conséquence, les termes de la série obtenue le sont aussi. Si l'on ordonne au contraire selon les puissances décroissantes de x, on trouve une série différente, et c'est ce que fait ensuite Bernoulli [69], qui obtient

$$2xddy - \int xd^3y + \iint xd^4y - \int^3 xd^5y + \int^4 xd^6y \text{ etc.}$$

Cette méthode est très surprenante, ce que Leibniz et Bernoulli ne cessent de souligner dans leurs lettres. Pourquoi ? Avant tout parce que Bernoulli manipule les formules d'une manière contraire à leur syntaxe apparente : comme nous l'avons vu, il écrit que dans sa méthode « d, d^2, d^3 etc. sont considérées comme des quantités algébriques et pas simplement comme des lettres caractéristiques ». C'est aussi ce qui frappe Leibniz, qui répond :

> Non sine admiratione vidi quam facile et quam alte penetraveris in ea quae proposueram de singulari calculi genere quo rectangulorum differentiales cum polynomiorum potentiis conferuntur, tantum pro literae x exponentibus, substituendo exponentes ipsius d ipsam x afficientis. Et pulchre notasti, hoc modo ipsas d tractari quasi literas non considerando ipsas x vel y nisi tanquam afficientes literam d, versa rerum vice ; cum alias d sit tantum nota quaedam syncategorematica, x autem et y sint quantitates [70].

68. Avec une organisation verticale des calculs et un quotient à droite, cette disposition n'est pas si éloignée de celle que l'on trouve habituellement chez les auteurs de l'époque ; la différence principale est qu'en général, ils mettent plutôt le diviseur à gauche du dividende. Voir par exemple Mercator 1668, p. 29 ; Wallis 1693, p. 362-364, 367 ; ou le manuel de Guisnée 1705, p. xvii-xxii.

69. « En commençant la division par le dernier membre, on obtient une autre série […] » (« Alia invenitur series incipiendo divisionem ab ultimo membro […] », lettre du 8–18 juin 1695, LAA, III.6A, p. 399 = LMS, III/1, p. 181.)

70. « J'ai vu non sans admiration combien facilement et profondément tu es entré dans les études que je

Pourtant, la méthode fonctionne, au sens où les résultats sont vérifiables par ailleurs. En effet, quoique Bernoulli ne le fasse pas explicitement, on peut différentier terme à terme les séries obtenues. Différentions par exemple jusqu'au cinquième terme la seconde série que Bernoulli donne pour $\int 2dxddy + xd^3y$:

$$\begin{aligned}&2dxddy + xd^3y \cancel{-\int xd^4y} \cancel{+\int xd^4y} \cancel{+\iint xd^5y}\\ &\qquad \cancel{-\iint xd^5y} \cancel{-\int^3 xd^6y} \cancel{+\int^3 xd^6y} + \int^4 xd^7y \text{ etc.}\end{aligned}$$

Dans sa lettre suivante, Bernoulli fournit une « confirmation » supplémentaire de sa méthode : il parvient à l'utiliser pour retrouver sa « série très universelle » – celle-là même qui avait servi de point de départ à toute cette investigation [71]. Sa démarche est toujours la même : pour calculer $\int \overline{ndz}$, il différentie ndz en $dndz + nddz$ puis calcule, toujours en traitant les d algébriquement, la fraction

$$\frac{(ndz)^2}{dndz + nddz} = \frac{d^0nd^2z}{d^1nd^1z + d^0nd^2z} = \frac{d^0nd^1z}{d^1nd^0z + d^0nd^1z}$$

(où il simplifie par d^1z). Il effectue alors la « division continue » en commençant par le terme d^0nd^1z. Encore une fois, Bernoulli omet les calculs, que j'explicite ici à nouveau selon la disposition scolaire française :

$$\begin{array}{llll|l}
d^0ndz & & & & d^0nd^1z + d^1nd^0z \\ \cline{5-5}
-d^0ndz & -d^1nd^0z & & & d^0nd^0z - d^1nd^{-1}z + d^2nd^{-2}z - d^3nd^{-3}z \ldots \\ \cline{1-2}
0 & -d^1nd^0z & & & \\
 & +d^1nd^0z & +d^2nd^{-1}z & & \\ \cline{2-3}
 & 0 & +d^2nd^{-1}z & & \\
 & & -d^2nd^{-1}z & -d^3nd^{-2}z & \\ \cline{3-4}
 & & 0 & -d^3nd^{-2}z & \\
 & & & \ldots &
\end{array}$$

t'avais proposées, concernant ce genre singulier de calcul où les différentielles des produits sont rapprochées des puissances des polynômes en remplaçant les exposants de la lettre x par ceux du d qui affecte ce x. Et tu as joliment remarqué que de cette manière les d eux-mêmes sont traités comme si c'étaient des lettres, en négligeant x ou y sinon dans la mesure où ils affectent d : on inverse ainsi les choses, puisque en d'autres circonstances d n'est qu'une sorte de signe syncatégorématique alors que x et y sont des quantités. » (Lettre du 24 juin 1695, LAA, III.6B, p. 424 = LMS, III/1, p. 191.)

71. Lettre du 17 juillet 1695 (ancien style), LAA, III.6B, p. 453–454 = LMS, III/1, p. 199–200.

Cette division conduit ainsi à la série

$$\int \overline{ndz} = d^0 nd^0 z - d^1 nd^{-1} z + d^2 nd^{-2} z - d^3 nd^{-3} z \text{ etc.}$$
$$= nz - dn \int z + d^2 n \int^2 z - d^3 n \int^3 z \text{ etc.}$$

Pour dz constant, écrit ensuite Bernoulli, on a les égalités $\int z = \frac{zz}{1\cdot 2dz}$, $\int^2 z = \frac{z^3}{1\cdot 2\cdot 3dz^2}$, $\int^3 z = \frac{z^4}{1\cdot 2\cdot 3\cdot 4dz^3}$, $\int^4 z = \frac{z^5}{1\cdot 2\cdot 3\cdot 4\cdot 5dz^4}$ d'où en remplaçant dans la série précédente

$$\int \overline{ndz} = nz - dn\frac{zz}{1\cdot 2dz} + d^2 n\frac{z^3}{1\cdot 2\cdot 3dz^2} - d^3 n\frac{z^4}{1\cdot 2\cdot 3\cdot 4dz^3} \text{ etc.}$$

qui est exactement la série de Bernoulli discutée ci-dessus. Bernoulli s'en félicite :

> hunc enim eventum cum haec inciperem scribere non sperabam, putans longe aliam seriem hac methodo proventuram : Elegans iste consensus mirifice methodorum probitatem praeprimis hujus posterioris ubi tam mirabiliter et contra omnem consuetudinem cum literis d, proceditur, confirmat [72].

Leibniz, lui aussi, y voit une « confirmation merveilleuse de notre méthode [73] ».

Malgré ses succès remarquables, cette méthode cache des difficultés profondes, et c'est à mon avis pour cela que les recherches de Leibniz et de Bernoulli s'arrêteront là. Bernoulli est beaucoup trop optimiste quand il écrit, juste après avoir montré comment intégrer en séries infinies :

> Video me hic inter scribendum et quidem ex insperato incidisse in methodum universalem summandi vel per vel citra seriem, quantitaem differentialem cujuscunque gradus ; video etiam infinita alia adhucdum abscondita hic latere, ea autem eruere, et studiosius excolere nunc non vacat […] [74].

En fait, sa méthode n'est valable que pour une classe très particulière d'expressions, à savoir les sommes de termes de la forme $d^k x d^l y$ (ou leurs analogues en davantage de variables, par exemple $d^k x d^l y d^m z$). Autrement dit, il faut que toutes les variables qui figurent dans l'expression apparaissent dans *chaque* terme, que ce soit différentiées (affectées de d^n pour

72. « Je n'espérais pas ce résultat quand j'ai commencé à écrire : je pensais que cette méthode conduirait à une série bien différente. Cet accord élégant confirme merveilleusement la fiabilité des méthodes, surtout de la seconde, où l'on emploie la lettre de d de manière si extraordinaire et si contraire à toutes les habitudes. » (Lettre à Leibniz du 17 juillet 1695, nouveau style, LAA, III.6B, p. 454 = LMS, III/1, p. 200.)

73. « methodi hujus nostrae mirabilis confirmatio » (Lettre du 29 juillet 1695, LMS, III/1, p. 206.)

74. « Je me rends compte qu'en écrivant cette lettre, j'ai découvert contre toute attente une méthode universelle pour sommer soit avec soit sans série une quantité différentielle de degré quelconque ; je vois aussi qu'une infinité d'autres choses jusqu'à présent dissimulées s'y cachent ; mais je n'ai pas en ce moment le loisir de les tirer au grand jour ni de les cultiver avec davantage d'application […] » (Lettre à Leibniz du 8–18 juin 1695, LAA, III.6A, p. 400 = LMS, III/1, p. 168.)

$n > 0$), intégrées (affectées de d^n pour $n < 0$) ou sous forme non modifiée (c'est-à-dire affectées de d^0) ; de plus, l'expression à intégrer ne peut contenir de puissances, puisque les règles de calcul utilisées rendent indiscernables, par exemple, d^2x et $(dx)^2$.

Dans ce cas très particulier, voici pourquoi la méthode fonctionne. Considérons pour simplifier une somme de termes du type $d^k x d^l y$ (c'est-à-dire limitons-nous à deux variables) et notons-la z. Comme l'écrit Feigenbaum (1985, p. 88), différentier une telle expression revient à multiplier par $d^1xd^0y + d^0xd^1y$ au sens des règles de Bernoulli, puisque

$$d(d^k x d^l y) = d^{k+1} x d^l y + d^k x d^{l+1} y = (d^1 x d^0 y + d^0 x d^1 y) d^k x d^l y.$$

Donc z, dz, d^2z, d^3z etc. forment bien une progression géométrique, dont la raison est $d^1xd^0y + d^0xd^1y$. Cela éclaire la méthode de Bernoulli et permet même de la simplifier : inutile de calculer une troisième proportionnelle, il suffit de diviser directement la quantité à intégrer par $d^1xd^0y + d^0xd^1y$ (c'est d'ailleurs exactement ce que Bernoulli se retrouve à faire, après simplification, dans le calcul qui conduit à sa « série très universelle »).

En définitive, Bernoulli a eu de la chance ; en général, il ne suffit pas de traiter algébriquement les opérateurs différentiels pour obtenir des résultats corrects. Après avoir reçu la lettre de Bernoulli, Leibniz s'imagine tout de suite que « d'une manière générale, les opérations qui ont lieu dans le cas d'une progression géométrique et de ses logarithmes peuvent être imitées ici [75] ». Ce n'est pas aussi simple, et c'est pour cela, sans doute, que Bernoulli et lui n'iront pas plus loin.

2.5 Les différentielles fractionnaires

Peut-on avoir des différentielles d'ordre fractionnaire ? À première vue, si l'on part de l'analogie des suites de nombres pour définir les différentielles, cela n'a aucun sens. En effet, étant donnée une suite de nombres, on peut former la suite de ses différences ; on peut prendre les différences des différences (les différences d'ordre 2) ; les différences des différences des différences (différences d'ordre 3) ; et ainsi de suite. On peut aussi former la suite de ses sommes, et puisque sommer et différentier sont des opérations inverses, il est raisonnable de considérer conventionnellement les sommes comme des différences d'ordre -1. Mais prendre les différences d'ordre $\frac{1}{2}$, qu'est-ce que cela pourrait bien vouloir dire ? Prendre les différences à moitié ? Quand on conçoit les différentielles ainsi, il n'y a pas de réponse claire.

75. « Et omnino, quae in geometrica progressione et logarithmis operationes locum habent, eas hic imitari licet […] » (Lettre du 24 juin 1695, LAA, III.6B, p. 424–425 = LMS, III/1, p. 191.)

Dans la correspondance que nous avons étudiée, la situation change : Leibniz dispose maintenant de formules générales portant sur d^n avec n quelconque. Il se demande alors immédiatement ce que l'on obtiendrait en remplaçant n par une fraction. Cette démarche est certainement très naturelle pour lui, puisque la formule du binôme elle-même est une généralisation de ce genre, valable pour des exposants tant entiers que fractionnaires. Donc, dès qu'il obtient sa première formule générale, à savoir (sans corriger les erreurs de calcul)

$$\int \overline{z^{\dot{e}} d^{\dot{m}} n} = z^e d^{\dot{m}} n - e \cdot z^{\frac{e-1}{\cdot}} d^{\overline{m-1}} n \cdot dz + e^2 \cdot z^{\frac{e-2}{\cdot}} d^{\frac{m-2}{\cdot}} n\overline{dz}^2 - e^3 z^{\frac{e-3}{\cdot}} d^{\overline{m-3}} n \cdot \overline{dz}^3 \text{ etc.}$$

Leibniz se pose la question de ses limites de validité, et écrit à Bernoulli :

> posse quidem e esse numerum non integrum, *sed m semper integrum esse*, nisi quis ad instar Metaphysicarum potentiarum (seu Logarithmorum) etiam metaphysicas nescio quas differentias (vel summas) fingere vellet [76].

Cette première formule ne fait toutefois apparaître $d^m n$ qu'au milieu de beaucoup d'autres termes, et n'est donc pas très adaptée pour caractériser d'éventuelles différentielles d'ordre fractionnaire. La version différentielle de la formule du binôme est un bien meilleur point de départ :

$$d^e\overline{xy} = d^e x d^0 y + \frac{e}{1} d^{e-1} x \cdot d^1 y + \frac{e \cdot e-1}{1 \cdot 2} d^{e-2} x \cdot d^2 y + \frac{e \cdot e-1 \cdot e-2}{1 \cdot 2 \cdot 3} d^{e-3} x \cdot d^3 y \text{ etc.}$$

Comme nous l'avons vu plus haut, Leibniz découvre apparemment cette formule en écrivant à l'Hôpital. Dans la même lettre, juste après la formule, il ajoute :

> Vous voyés par là Monsieur, qu'on peut exprimer par une serie infinie une grandeur comme $d^{\frac{1}{2}}\overline{xy}$ ou $d^{1:2}\overline{xy}$, quoyque cela paroisse eloigné de la Geometrie, qui ne connoist ordinairement que les differences à exposans entiers affirmatifs, ou les negatifs à l'egard des sommes, et pas encor celles, dont les exposans sont rompus. Il est vray, qu'il s'agit encor de donner $d^{1:2}x$ *pro illa serie* [77] ; mais encor cela se peut expliquer en quelque façon [78].

Leibniz traite ensuite un cas particulier [79], celui d'une variable x dont les valeurs ou ordon-

76. « si e peut être un nombre non entier, *m en revanche est toujours entier*, à moins que quelqu'un ne veuille créer, à l'instar des puissances (ou logarithmes) métaphysiques, je ne sais quelles différences (ou sommes) également métaphysiques. » (Je souligne. Lettre du 28 février 1695, LAA, III.6A, p. 313 = LMS, III/1, p. 167.)

77. « au lieu de cette série ».

78. Lettre à l'Hôpital du 30 septembre 1695 (nouveau style), LAA, III.6B, p. 510 = LMS, II, p. 301–302.

79. Il avait utilisé le même exemple quelques mois plus tôt pour justifier l'existence des différentielles d'ordre 2 contre les objections de Nieuwentijt (Leibniz 1695), et avait expliqué cet argument à l'Hôpital dans une lettre précédente (datée du 14–24 juin 1695, LAA, III.6A, p. 416 = LMS, II, p. 288.)

nées sont en progression géométrique (Leibniz choisit une raison unité pour simplifier les calculs) eu égard à des abscisses β de différentielle $d\beta$ constante, de sorte que $dx = xd\beta$. En termes modernes, cela revient à choisir une fonction dont la dérivée est proportionnelle à la valeur, c'est-à-dire une exponentielle. On trouve alors sans surprise

$$dx = xd\beta, \quad ddx = x \cdot \overline{d\beta}^2, \quad d^3x = x \cdot \overline{d\beta}^3, \quad \text{et} \quad d^e x = x \cdot \overline{d\beta}^e,$$

formule qui s'interpole tout naturellement pour des exposants non entiers :

> par cette adresse l'exposant differentiel est changé en exposant potentiel et remettant $d\overline{x} : x$ pour $d\beta$, il y aura $d^e x = \overline{d\overline{x} : x}^e \cdot x$. Ainsi il s'en suit que $d^{1:2}x$ sera egal à $x \cdot \sqrt[2]{dx : x}$. Il y a de l'apparence qu'on tirera un jour des consequences bien utiles de ces paradoxes, car il n'y a gueres de paradoxes sans utilité [80].

Cela ne satisfait pas le marquis de l'Hôpital. Après avoir répondu à Leibniz avec perplexité [81], il écrit à Jean Bernoulli que « [M^{r}. Leibniz] me fait part aussi d'une suitte par laquelle on peut exprimer les differences qui ont pour exposans des nombres rompus. cependant cela ne me contente pas encore, car je ne vois point comment on pourroit se former une idée nette de ces sortes de differences en geometrie [82] ». Dès réception de cette lettre, et avant de répondre à l'Hôpital, Jean Bernoulli – qui n'avait pas encore entendu parler de différences fractionnaires – demande des explications à Leibniz [83]. Ce dernier lui envoie presque la même chose qu'à l'Hôpital, à ceci près que Leibniz traite cette fois le cas d'une variable x en progression géométrique de raison a quelconque (plus seulement $a = 1$), ce qui le conduit à $d^{\frac{1:2}{\cdot}} x = x\sqrt{dh : a}$, où h remplace le β de la lettre à l'Hôpital [84]. Leibniz conclut ainsi :

> Unde vides talium differentiarum valores hoc modo haberi posse per radicem vel potentiam ordinariae differentiae. Quod cum memorabile sit, Tibi non ingratum fore puto. Easdem extraordinarias differentias per seriem infinitam ex ordinariis conflatam exprimi posse, me non monente vides, adeoque suo modo reales esse, etiam hinc patet [85].

80. Lettre à l'Hôpital du 30 septembre 1695 (nouveau style), LAA, III.6B, p. 510 = LMS, II, p. 302.
81. « Je trouve bien difficile de se former une idée nette de ces differences qui ont pour exposans des nombres rompus et que vous avez trouvé le moyen d'exprimer par des suittes infinies. » (Lettre du 1er décembre 1695, LAA, III.6B, p. 556 = LMS, II, p. 304.)
82. Lettre de l'Hôpital à Jean Bernoulli du 24 décembre 1695, Jean Bernoulli (BBW, I, p. 308).
83. Lettre du 17–27 décembre 1695, LAA, III.6B, p. 589 = LMS, III/1, p. 225.
84. En termes modernes, nous dirions qu'il traite $x = e^{ah}$.
85. « On voit qu'ainsi, les valeurs de telles différences peuvent être données par des racines ou des puis-

Leibniz veut sans doute dire que (par exemple) la racine de dh qui apparaît dans l'expression $d^{\frac{1:2}{\cdot}}x = x\sqrt{dh : a}$, dont le sens n'est pas très clair, peut être remplacée par une série infinie de puissances ordinaires de dh (ce qui est effectivement facile en utilisant la formule du binôme). On peut donc exprimer $d^{\frac{1:2}{\cdot}}x$ en n'utilisant que des objets parfaitement admis jusqu'alors.

Mais pourquoi, dans les deux lettres, Leibniz passe-t-il si abruptement de la formule très générale du binôme au cas très particulier d'un x en progression géométrique ? Leibniz semble raisonner ainsi : la formule du binôme conduit facilement à des relations générales vérifiées par d'éventuelles différentielles fractionnaires ; ces relations générales semblent cohérentes, ce qui indique qu'on peut sans doute donner un sens à ces nouveaux objets ; essayons donc de les définir autrement, au moins dans un cas particulier. Je voudrais cependant proposer une autre interprétation de la démarche de Leibniz, plus conjecturale. Peut-être Leibniz songe-t-il à utiliser sa formule du binôme pour donner une authentique *définition* des différentielles fractionnaires, selon la démarche suivante. Écrivons tout d'abord la formule du binôme pour $d^{\frac{1}{2}}(xy)$ (ce qu'aucun de nos auteurs ne fait explicitement) :

$$d^{\frac{1}{2}}(xy) = d^{\frac{1}{2}}xd^0y + \frac{1}{2}d^{-\frac{1}{2}}x.d^1y - \frac{1}{2}d^{-\frac{3}{2}}x.d^2y + \frac{1}{4}d^{-\frac{5}{2}}x.d^3y \text{ etc.}$$

Le problème avec cette formule est qu'il y a des différentielles fractionnaires non seulement à gauche, mais aussi à droite. *A priori*, elle ne permet donc pas de définir $d^{\frac{1}{2}}$ de manière non circulaire. Toutefois, dans le membre de droite c'est seulement la quantité x qui est affectée de différentielles fractionnaires : si l'on sait définir les différentielles d'ordre quelconque d^ex de x, on pourra aussi exprimer $d^{\frac{1}{2}}\overline{xy}$ par une série infinie ; en posant $y = \frac{z}{x}$, on pourra même définir $d^{\frac{1}{2}}z$ pour une quantité z quelconque. En d'autres termes, le cas particulier que traite Leibniz peut ouvrir la porte à une définition générale, et c'est peut-être ce qu'il a en tête lorsqu'il écrit à Bernoulli que « ces différences extraordinaires peuvent être exprimées par une série infinie composée d'ordinaires »… Ce n'est toutefois qu'une conjecture.

Leibniz ne reviendra pas en détail sur les différentielles d'ordre fractionnaire. En réponse aux perplexités du marquis de l'Hôpital, il se contentera d'admettre :

> Quant aux différences dont les exposans sont des nombres rompus, j'avoue qu'on ne les sçauroit comprendre, mais ces sortes de grandeurs quand elles ne seroient qu'imaginaires peuvent servir à trouver des verités réelles. Et il est

sances de différences ordinaires. Voilà qui est mémorable : tu seras, je pense, heureux de l'admettre. Tu constates sans que j'aie besoin de le dire que ces différences extraordinaires peuvent être exprimées par une série infinie composée d'ordinaires ; cela aussi montre qu'à leur manière, elles sont réelles. » (Lettre du 23 décembre 1695, LAA, III.6B, p. 601–602 = LMS, III/1, p. 228.)

tousjours vray qu'elles ont un *fundamentum in re*[86].

Il établit donc une comparaison avec les nombres imaginaires (que Bernoulli, dans sa propre réponse à l'Hôpital, rend plus explicite[87]).

Jusqu'à sa mort, Leibniz mentionnera occasionnellement les différentielles d'ordre fractionnaire, en général comme argument pour souligner les avantages de sa notation exponentielle : ainsi écrit-il dans une lettre à Wallis que dans son calcul, on peut employer dy, d^2y, d^3y et que « même $d^{\frac{1}{2}}$ y apparaît utilement[88] ». Mais aucun de ces passages (rencensés par Dugowson 1994, p. 33–54) n'apporte d'élément nouveau.

2.6 Diffusion et postérité

La diffusion de ces innovations est contrastée. L'abréviation exponentielle pour les différentielles d'ordre supérieur (d^3 pour ddd, etc.) est rapidement et largement adoptée par les utilisateurs du calcul leibnizien. En revanche, l'analogie entre $d^n(xy)$ et $(x+y)^n$, et plus encore sa généralisation aux intégrales fondée sur la notation $d^{-1} = \int$, restent assez confidentielles. Lagrange, qui est le premier à les approfondir des décennies plus tard, les redécouvre en fait par lui-même – il est vrai sur la base de la notation exponentielle, qui parsème ses manuels comme un indice.

Penchons-nous tout d'abord sur l'abréviation exponentielle d^3, d^4, etc. À partir de 1695, Leibniz ne manque jamais de l'utiliser lorsqu'il présente son calcul[89] ; si elle est absente du

86. « un fondement dans la chose », c'est-à-dire un fondement en-dehors de notre esprit ; l'expression est d'origine scolastique. (Lettre de Leibniz à l'Hôpital du 15 janvier 1696, LAA, III.6B, p. 625 = LMS, II, p. 311.)

87. « Les différentielles qui ont pour exposant des nombres rompûs ne sont en effet que metaphysiques comme dit Mr. Leibnits, ou plutot imaginaires, qui ne sont pas determinables ou qui n'existent pas, et ainsi ce que est par exemple $\sqrt{-aa}$ parmi les quantités algebraiques, ou $a^{\sqrt{2}}$ parmi les puissances, la meme chose est aussy $d^{\frac{1}{2}}$ parmi les differences, sans pourtant que ces choses chimeriques fassent tort aux autres qui sont reelles ; c'est pourquoy il seroit inutile de demander une idée plus nette de ces sortes de differences, de meme qu'on ne peut pas avoir une autre idée de $\sqrt{-aa}$, ou de $a^{\sqrt{2}}$ que celle qu'on a d'une chose qu'on peut demontrer qu'elle n'existe pas in *rerum naturâ*, quoique on en puisse dire des verités reelles, par ex. que $\sqrt{-aa}$ est la racine de $-aa$, que $\sqrt{-aa}$ est plus petit en son espece d'*etre* que $3\sqrt{-aa}$; que $a^{\sqrt{2}}$ est la racine de $a^{2\sqrt{2}}$, que son logarithme est $\sqrt{2}a^{2\sqrt{2}}$; et ainsi que $d^{\frac{1}{2}}$ est l'integrale ou la somme de $d^{\frac{3}{2}}$ etc. » (lettre de Jean Bernoulli à l'Hôpital du 10 janvier 1696, BBW, I, p. 311).

88. « imo utiliter etiam occurrit $d^{\frac{1}{2}}y$ » (lettre à Wallis du 28 mai 1697, LMS, IV, p. 25).

89. Voici quelques exemples. Dès l'été 1695, il parle dans un article des « ddx ou d^2x, $dddx$ ou d^3x » (Leibniz 1695, p. 310 = LMS, V, p. 321) ; voir aussi sa lettre à Huygens du 21 juin 1695, LAA, III.6A, p. 421 = (LMS, II, p. 207). Dans une lettre à Wallis de 1697, il écrit : « judicavi, praeter affectiones quantitatis hactenus receptas y, y^2, y^3, $y^{\frac{1}{2}}$, $y^{\frac{1}{3}}$ etc. […] posse adhiberi novas differentiarum vel fluxionum affectiones dy, d^2y (seu ddy), d^3y (seu $dddy$) » (« j'ai jugé qu'outre les affections de quantités admises jusqu'ici y, y^2, y^3, $y^{\frac{1}{2}}$, $y^{\frac{1}{3}}$, etc. […] on pouvait employer de nouvelles affections des différences ou fluxions dy, d^2y (ou ddy), d^3y (ou $dddy$) » ; lettre de Leibniz à Wallis du 28 mai 1697, LMS, IV, p. 25) ; et dans l'*Historia et origo…*, composé vers 1715 : « ut

manuel très diffusé de l'Hôpital [90], rédigé trop tôt, elle se trouve en revanche dans celui de Reyneau quelques années plus tard [91].

Dans la première moitié du XVIIIe siècle, l'usage semble assez généralement être le suivant : les différentielles secondes sont écrites dd (ce qui n'est guère surprenant alors que les carrés sont plus volontiers écrits xx que x^2), mais dès qu'une différentielle d'ordre 3 ou plus apparaît, elle est notée exponentiellement. Cela reste plutôt rare, dans la mesure où la plupart des applications du calcul différentiel ne requièrent pas de dépasser l'ordre 2, mais cela arrive, que ce soit au détour d'un calcul, dans les présentations générales des différentielles d'ordre supérieur ou encore pour écrire des séries infinies de différentielles comme celle de Bernoulli. À titre d'échantillon, il en est ainsi des ouvrages qu'étudie le jeune Lagrange en 1753–1754, dans les mois qui précèdent sa redécouverte de l'analogie [92] : la notation figure dans le manuel de calcul différentiel de Maria Agnesi [93], dans la *Mécanique* d'Euler [94] et se trouve en bonne place dans le mémoire d'Euler *Methodus inveniendi lineas curvas…* qui joue pour Lagrange un rôle crucial [95].

L'analogie entre $d^n(xy)$ et $(x+y)^n$ a une diffusion plus restreinte. Comme nous l'avons vu, Leibniz finit par publier le cas des exposants positifs en 1710, sans évoquer la généralisation aux sommes [96]. L'analogie circule aussi *via* des extraits de lettres : Jean Bernoulli copie pour l'Hôpital les passages sur le sujet de sa correspondance avec Leibniz, y compris ceux qui traitent de son nouveau calcul [97], puis envoie Varignon consulter ces morceaux choisis chez le marquis [98]. Bernoulli en fait aussi publier quelques extraits en 1721 [99], dans le cadre de sa querelle de priorité avec Brook Taylor [100]. Finalement, une grande partie de

---

habuimus hactenus x, xx, x^3 etc. y, yy, y^3 etc. ita posse adhiberi dx, ddx, d^3x etc. dy, ddy, d^3y etc. » (« de la même manière que nous avions jusque là x, xx, x^3 etc., y, yy, y^3 etc., on peut employer dx, ddx, d^3x etc., dy, ddy, d^3y etc. » ; Leibniz 1846, p. 17).

90. Voir *supra*, p. 44 note 11.

91. Reyneau 1708, §544, p. 654.

92. Sur la formation de Lagrange, voir Borgato et Pepe 1990, p. 7 ou Pepe 2008, p. 38-40 ; leur source principale sur les lectures de Lagrange à cette période est Maurice 1814.

93. « La differenze seconde si sogliono marcare con doppia d, le terze con trè d ec. La differenza adunque di dx, cioè la differenza seconda di x si scriverà ddx, o pure d^2x […] ; la differenza terza farà $dddx$, o pure d^3x ec. » (Agnesi 1748, II, p. 438).

94. Euler 1736, p. ex. vol. 1, p. 377, 381.

95. Euler 1744, p. 9.

96. Voir *supra*, p. 55.

97. Lettre de Jean Bernoulli à l'Hôpital du 10 janvier 1696, BBW, I, p. 311.

98. Voir la lettre de Jean Bernoulli à Varignon du 24 décembre 1697 (BBW, II, p. 154), puis la lettre de Varignon du 18 février 1697 où il confirme que l'Hôpital les lui a communiquées (BBW, II, p. 159).

99. Burkhard [Mencke] 1721.

100. Sur cette querelle, voir Feigenbaum 1985 et Panza 1992, section III.2.d, et les références citées note 38 p. 30.

sa correspondance avec Leibniz, dont nos passages, paraît en 1745 [101].

Malgré ces différentes voies de diffusion, l'analogie n'est manifestement pas le genre de résultat que l'on trouve routinièrement dans les manuels [102]. Une étude plus poussée du corpus leibnizien de la première moitié du XVIIIe siècle, que je n'ai pas entreprise, serait nécessaire pour voir si et où elle est mentionnée ; les remarques qui suivent sont donc conjecturales. Mon impression, toutefois, est que tous ne partageraient pas le jugement enthousiaste de Bernoulli :

> Si on veut prendre la differentielle d'un certain degré, par ex. du quatrieme, d'une quantité composée de deux indeterminées comme de xy, selon la methode ordinaire vous prendriez d'abord la differentielle du premier degrés et vous auriez $xdy + ydx$, en differentiant cela vous trouveriez $xddy + 2dxdy + yddx$ pour la differentielle du second degrez, laquelle étant differentiée encore une fois, vous auriez celle du troisieme degré, qui enfin par sa differentiation vous donneroit la quatrieme que l'on cherche, mais quel long proces ? quelle grande peine ? au lieu que par une merveilleuse analogie l'affaire se fait dans un instant, fust ce meme la differentielle du centieme degrez que l'on cherche [...] [103].

À vrai dire, l'intérêt de calculer la différentielle du centième degré de xy ne saute pas aux yeux. Ce résultat pris isolément n'a, me semble-t-il, que peu d'utilité pour le praticien : tant que le calcul leibnizien est considéré avant tout comme un outil pour résoudre des problèmes géométriques ou physiques, il reste donc anecdotique. Il vaut surtout pour l'ouverture qu'il offre vers les formes générales que prennent les différentielles, et c'est par là qu'il devient central pour Lagrange.

C'est effectivement, écrit Lagrange à Euler, la recherche de « quelque ordre sûr [104] » dans les différentielles et dans les puissances qui le conduit, en juin 1694, à redécouvrir l'analogie de Leibniz y compris pour les intégrales. Il s'empresse d'en faire part non seulement à Euler, mais aussi à Fagnano, sur le conseil de qui [105] il la publie [106]. Euler, lui, connaissait la découverte de Leibniz : il ignore la lettre de Lagrange, et ce n'est que quand

101. Leibniz et Jean Bernoulli 1745.

102. Mon recensement n'a pas été très méthodique, mais le premier traité dans lequel j'aie trouvé l'analogie est la somme quasi-encyclopédique d'Euler sur le calcul différentiel, où elle n'est mentionnée qu'en passant et dans le cas d'exposants positifs.

103. Lettre de Jean Bernoulli à Varignon du 24 décembre 1697, BBW, II, p. 154.

104. « *certus aliquid ordo* » (lettre de Lagrange à Euler du 28 juin 1754, EOO IV.A.5, p. 361).

105. « Questo che sarà un bel monumento dell'acutezza del suo ingegno merita la pubblica luce [...] » (« Ce résultat, qui sera un beau monument à l'acuité de votre intelligence, mérite la lumière publique [...] » ; lettre de Fagnano à Lagrange du 3 juillet 1754, Fagnano 1911-1912, III, p. 179).

106. Sous la forme d'une lettre ouverte à Fagnano datée du 23 juillet : Lagrange 1754.

celui-ci lui envoie, un an plus tard, un résultat plus intéressant qu'Euler finit par répondre, lui signalant au passage que « quant à l'analogie des différentielles d'ordre quelconque de la formule xy avec les termes de la puissance du binôme $(a+b)^m$ que vous m'avez annoncée dans votre première lettre, je me rappelle qu'elle a déjà été remarquée par Leibniz, ce que, sauf erreur erreur de ma part, vous trouverez dans sa correspondance avec Bernoulli [107] ».

Lagrange, comme Leibniz avant lui [108], altère les notations usuelles et met les deux formules l'une sous l'autre, « afin que s'il y a entre elles quelque ressemblance, celle-ci puisse être perçue toute entière d'un seul coup d'œil [109] » :

$$\overline{a+b}^m = a^m b^0 + m a^{m-1} b^1 + \overline{m} \times \frac{\overline{m-1}}{2} a^{m-2} b^2 + \overline{m} \times \frac{\overline{m-1}}{2} \times \frac{\overline{m-2}}{3} a^{m-3} b^3$$
$$+ \overline{m} \times \frac{\overline{m-1}}{2} \times \frac{\overline{m-2}}{3} \times \frac{\overline{m-3}}{4} a^{m-4} b^4 \text{ etc.}$$
$$\overline{xy}^m = x^m y^0 + m x^{m-1} y^1 + \overline{m} \times \frac{\overline{m-1}}{2} x^{m-2} y^2 + \overline{m} \times \frac{\overline{m-1}}{2} \times \frac{\overline{m-2}}{3} x^{m-3} y^3$$
$$+ \overline{m} \times \frac{\overline{m-1}}{2} \times \frac{\overline{m-2}}{3} \times \frac{\overline{m-3}}{4} x^{m-4} y^4 \text{ etc.}$$

(Il écrit ici tout simplement x^m à la place $d^m x$.) Toujours comme Leibniz, il retrouve ensuite la « série universelle » de Bernoulli [110] en faisant $m = -1$.

À en croire Lagrange, c'est une authentique redécouverte [111], et c'est bien possible :

107. « Quod autem in prioribus litteris de analogia differentialium cuiusque ordinis formulae xy et terminorum bino[mii] potestatis $(a+b)^m$ attulisti, eam iam a Leibnizio obse[rvatam] esse memini, quod, nisi fallor, in ejus cum Bernoullio c[ommer]cio reperies » (lettre d'Euler à Lagrange du 6 septembre 1755, EOO IVA.5, p. 376 ; trad. fr. p. 377–378, modifiée).

108. Voir sa lettre à l'Hôpital puis sa publication de 1710, cf. p. 55 et 56.

109. « ut si quid in ipsis inest similitudinis, totum uno oculi ictu perspiciatur » (lettre de Lagrange à Euler du 28 juin [1754], *loc. cit.*) ; selon la formulation de la lettre ouverte à Fagnano : « Dunque primieramente propongo le due serie [. . .] sicchè in una sola occhiata se ne comprenda ogni possibile rapporto, e corrispondenza [. . .] » (OL, VII, p. 584).

110. Dans sa lettre à Euler, il attribue cette série à Leibniz ; dans sa lettre ouverte à Fagnano, qui est plus détaillée, il cite en revanche Jean Bernoulli 1694.

111. Quelques semaines après sa première lettre à Fagnano, son petit texte à peine publié, il affirme avoir trouvé par hasard sa découverte dans la correspondance Leibniz-Bernoulli : « [E]ssendo io l'altro dì a casa del Proffessor nostro di Teoria medica, Ignazio Somis e cercando non so che, in un certo libro che egli avea, contenente lettere latine private di Matematica, tra Leibnitz e Bernulli [. . .] [Leibniz et Jean Bernoulli 1745] il caso ha voluto che ivi io trovassi quella serie, che ho poco fa pubblicata per i differenziali ed integrali. [. . .] Questo veramente da principio m'ha crucciato non poco, considerando sopra tutto, che quando questo si spargesse [. . .] molti si troverebbero certo, che risguarderebbero me, come un plagiario, ed impostore. » (« [A]lors que j'étais, l'autre jour, chez notre professeur de théorie médicale Ignazio Somis et que je cherchais je ne sais quoi dans un certain livre qu'il avait, contenant des lettres latines privées de mathématiques entre Leibniz et Bernoulli [. . .] le sort a voulu que j'y trouve la série que j'ai fait publier il y a peu pour les différentielles et les intégrales [. . .]. Cela m'a tout de suite inquiété, et plus qu'un peu, considérant surtout que lorsque cela se saura [. . .] il s'en trouvera certainement beaucoup pour me considérer comme un plagiaire et un imposteur. » Lettre de Lagrange à Fagnano du 14 août 1754, Fagnano 1911-1912, III, p. 189.)

connaissant la formule du binôme et la série de Bernoulli, il avait les mêmes ingrédients que Leibniz cinquante ans plus tôt, et même davantage, puisque la notation exponentielle pour les différentielles d'ordre supérieur s'était entre temps imposée. Quoi qu'il en soit, c'est par l'intermédiaire des usages heuristiques très fructueux que Lagrange fait ultérieurement de cette analogie qu'elle devient, dans la seconde moitié du XVIII^e siècle, un important sujet de recherches, chez Laplace, Arbogast, Servois et d'autres [112].

112. Voir en particulier Lagrange 1774. Pour une discussion approfondie de ce mémoire et des tentatives de Laplace pour en justifier les méthodes, voir Panza 1992, section III.4. Pour un aperçu général des recherches suscitées par ces travaux de Lagrange, voir par exemple Koppelman 1971 ou Lubet 2010.

Chapitre 3

Analyse : le rôle des notations

Je reviendrai sur différents aspects de cet épisode tout au long de ce travail, en fonction des besoins de la discussion. Un point particulièrement délicat mérite cependant une analyse préalable : quel rôle les notations jouent-elles dans les découvertes de Leibniz et Jean Bernoulli ? Plus précisément, la question porte sur la notation exponentielle généralisée que Leibniz introduit dans sa lettre du 28 février 1695, et qui semble si intimement liée aux découvertes substantielles de nos auteurs : d^3x pour $dddx$, $d^{-2}x$ pour $\iint x$, d^nx pour une différentielle ou somme d'ordre quelconque, etc.

Il peut être tentant d'affirmer, comme Parmentier, que « l'analogie entre différences et puissances exploite une richesse propre à la notation leibnizienne, elle serait *impossible* sans la notation exponentielle de l'ordre de différentiation [1] ». L'idée sous-jacente est que certaines choses ne sont pas exprimables, voire pas pensables, en l'absence d'une écriture appropriée ; c'est par exemple dans cet esprit que Serfati soutient, à propos des avantages de notre notation usuelle des puissances, qu'« une substitution aussi simple pour nous à écrire et à opérer que celle de $A = x^2+x+2$ dans A^2+A demeura une opération inconcevable – elle ne pouvait *se penser* – dans le cadre de l'écriture médiévale des mathématiques [2] ». Dans notre cas, Parmentier affirme ainsi que la notation d^n est cruciale parce qu'elle « permet de représenter une différentielle d'ordre quelconque » ; d'ailleurs, ajoute-t-il, cette notation elle-même n'est possible que grâce à une spécificité antérieure de l'écriture leibnizienne, à savoir que « l'algorithme dx, à la différence de la notation newtonienne, sépare les variables et l'opération de différentiation, susceptible dès lors d'être considérée pour elle-même [3] ».

1. LNC, p. 411 (je souligne).
2. Serfati 2006, p. 113 (l'auteur souligne). Sa position est plus nuancée que cette citation peut le laisser croire ; j'y reviendrai ci-dessous, p. 84.
3. LNC, p. 411..

C'est, je crois, faire trop d'honneur aux notations de Leibniz ; plus généralement, ce n'est pas en termes de ce que les notations permettent ou ne permettent pas dans l'absolu qu'il faut poser le problème. C'est ce que je voudrais montrer d'abord, pour préparer une discussion plus positive.

3.1 Une comparaison : les notations de Newton et de Taylor

Pour commencer, il n'est pas vrai que les notations de Newton ne permettent rien d'analogue. Non seulement elles peuvent, en principe, être adaptées pour en faire autant que celles de Leibniz, mais nous verrons que Newton lui-même fait des essais dans cette direction dans un brouillon ; surtout, Brook Taylor (1715) introduit dans un contexte newtonien une notation qui est remarquablement parallèle à $d^{-1} = \int$.

Bien sûr, la comparaison entre les différentielles de Leibniz et les fluxions de Newton est délicate ; ce sont des objets différents, insérés dans des cadres conceptuels assez éloignés. Néanmoins, pour la question qui nous occupe, il y a un parallèle clair entre les deux calculs : Leibniz associe à toute quantité géométrique variable ses différences et ses sommes, selon un procédé qui peut être répété pour produire les différences des différences, les sommes des sommes, etc. ; de même Newton sait associer à une quantité variable sa fluxion et sa fluente, et, du moins [4] dans ses travaux plus approfondis des années 1691–1693, la fluxion de la fluxion, la fluente de la fluente, etc. De ce parallèle découle un problème notationnel similaire : comment noter toutes ces quantités que l'on peut former à partir d'une quantité variable x ?

Newton résout ce problème de diverses manières au fil des années, et certaines sont, il est vrai, très éloignées de la notation de Leibniz. Dans ses premiers manuscrits de 1664–1665 [5], il note généralement par des lettres, comme m, n, r etc. ce qu'il appellera plus tard les fluxions de x, y, z, etc. Leibniz critique ce genre de notation (quoique sans viser spécifiquement celle de Newton, qu'il ne connaît sans doute pas) dans une lettre à Huygens de 1690 :

> [C]e que j'apelle dx ou dy, vous le pouvés designer par quelque autre lettre, ainsi rien ne vous empeche d'exprimer les choses à vostre maniere. Cependant je m'imagine qu'il y a certaines vües qui ne viennent pas aussi aisément que par mon expression, et c'est à peu près comme si, au lieu des racines et

4. On trouve dès 1666 chez Newton des gestes vers une réitération du passage aux fluxions, mais seulement dans des cas particuliers ; voir Panza 2005, p. 490 et suivantes.

5. Publiés dans NMP, I ; voir l'analyse de Panza (2005).

> puissances, on vouloit toujours substituer des lettres, et au lieu de xx, ou x^3, prendre *m*, ou *n*, après avoir déclaré, que ce doivent estre les puissances de la grandeur x. Jugés Mons^r combien cela embarrasseroit. Il en est de meme de dx ou de ddx, et les differences ne sont pas moins des affections des grandeurs indeterminées dans leurs lieux [6], que les puissances sont des affections d'une grandeur prise à part. Il me semble donc qu'il est plus naturel de les designer ensorte qu'elles fassent connoistre immediatement la grandeur dont elles sont les affections [7].

Effectivement, la lettre n n'indique ni s'il s'agit d'une fluxion, ni, si oui, de quelle quantité elle provient. Dans d'autres textes, dont les *Philosophiæ naturalis principia mathematica* de 1687, Newton adopte une convention qui évite cette première critique : pour des quantités variables A, B, C, il note a, b, c (en minuscules) ce qu'il appelle dans ce texte leurs « moments » [8]. Il est alors clair que a vient de A. Leibniz, qui connaît les *Principia*, pointe cependant un autre problème : cette écriture-là ne se généralise pas aux différences d'ordre supérieur.

> Je voy que M. Neuton se sert des minuscules pour les differences, mais quand on vient aux differences des differences, et au de là, comme il peut arriver, il faudra encor changer, de sorte qu'il me semble qu'on fait mieux de se servir d'une expression qui s'etend à tout [9].

Mais précisément, les choix notationnels de Newton s'expliquent aussi par le fait qu'il ne s'intéresse guère, à l'époque, aux fluxions d'ordre supérieur. Comme le remarque Panza, c'est justement lorsqu'en 1691 il se met à réfléchir au problème de la résolution d'une équation fluxionnelle d'ordre quelconque qu'il introduit une notation plus générale [10]. À partir de là, le parallèle avec Leibniz devient plus étroit.

Dans ses manuscrits de 1691–1692 [11], donc, Newton introduit la notation suivante [12] :

6. Leibniz entend par là que les différences ne sont définies que pour ce que Bos appelle une « quantité géométrique variable » associée à une courbe (cf. *supra*, section 1.1).

7. Lettre de Leibniz à Huygens du 3/13 octobre 1690, LAA, III.4, p. 620 = LMS, II, p. 49.

8. Voir par exemple le lemme II du livre II : « Igitur sensus Lemmatis est, ut si quantitatum quarumcunq ; perpetuo motu crescentium vel decrescentium A,B,C, &c. [m]omenta vel mutationum velocitates dicantur a, b, c, &c. momentum vel mutatio rectanguli AB fuerit Ab+aB […] » (Newton 1687, p. 251 ; « Le sens de ce lemme est donc, que si A, B, C, etc. sont des quantités quelconques croissantes ou décroissantes par un mouvement continu, et que leurs moments ou vitesses de changement soient nommées a, b, c, etc. le moment ou le changement du rectangle AB sera $Ab + aB$ […] », trad. adaptée de celle d'Émilie du Châtelet : Newton 1759, vol. 1, p. 261).

9. *Id.*, LAA, III.4, p. 620 = LMS, II, p. 50.

10. Voir Panza 1992, vol. 1, note 97 p. 154.

11. Voir NMP, VII.

12. La notation semble apparaître vers novembre 1691 (cf. NMP, VII, p. 131, note 5), puis est utilisée massi-

« $\dot{v}$ est la fluxion de la quantité v, $\ddot{v}$ est la fluxion de la quantité $\dot{v}$, et $\dddot{v}$ est la fluxion de la quantité $\ddot{v}$ [13] ». (À la suite de Newton, cette notation reste largement dominante en Angleterre jusqu'au début du XIXe siècle, par opposition au reste de l'Europe qui adopte généralement la notation de Leibniz.) Quelques mois plus tard, il y ajoute $\acute{z}$ pour la fluente de z, de sorte que « $\acute{\acute{z}}$, $\acute{z}$, z, $\dot{z}$, $\ddot{z}$, $\dddot{z}$, $\ddddot{z}$, $\overset{\cdots}{\ddot{z}}$ etc. désignent une série de quantités dont chacune est la fluxion de la précédente et est la quantité fluente ayant pour fluxion celle qui la suit [14] ». On a ainsi une situation très analogue aux $\iint x$, $\int x$, x, dx, ddx, $dddx$ etc. de Leibniz avant la notation exponentielle d^n. D'ailleurs, comme l'écrit Niccolò Guicciardini, les praticiens des premières décennies du XVIIIe siècle n'ont aucune difficulté à passer d'une notation à l'autre et le font routinièrement [15].

Est-ce, alors, la possibilité d'indiquer l'ordre de différentiation par un chiffre, et donc aussi d'écrire d^n voire $d^{-1} = \int$, qui manque aux newtoniens ? À lire les légers sarcasmes d'Euler contre les « Anglais », on pourrait le croire : après avoir présenté la notation newtonienne, il construit un véritable gratte-ciel de points pour en faire voir l'inanité.

> Qui notandi modus, uti ab arbitrio pendens, etsi improbari nequit, si punctorum numerus fuerit parvus, ut numerando facile percipi queat ; tamen si plura puncta inscribi debeant, maximam confusionem plurimaque incommoda affert. Differentiale enim seu fluxio decima perquam incommode hoc modo $\overset{\vdots\vdots}{y}$ repraesentatur, cum nostro signandi modo $d^{10}y$ facillime comprehendatur. Oriuntur autem casus, quibus multo adhuc superiores differentialium ordines atque adeo indefiniti exprimi debent, ad quos Anglorum modus prorsus sit ineptus [16].

vement dans un manuscrit de décembre (cf. *id.*, p. 64, note 35).

13. « [...] $\dot{v}$ est fluxio quantitatis v, & $\ddot{v}$ fluxio ipsius $\dot{v}$, & $\dddot{v}$ fluxio ipsius $\ddot{v}$. » Ce passage vient d'un document que Newton prépare pour Wallis (cf. NMP, VII, p. 174, note 12), que ce dernier publie *verbatim* dans son *De Algebra* Wallis 1693, p. 392. *Via* Wallis, cette notation est donc diffusée dès 1693 au-delà du cercle restreint des proches de Newton.

14. « Designant igitur $\acute{\acute{z}}$, $\acute{z}$, z, $\dot{z}$, $\ddot{z}$, $\dddot{z}$, $\ddddot{z}$, $\overset{\cdots}{\ddot{z}}$ &c. seriem quantitatum quarum quælibet posterior est fluxio præcedentis et quælibet prior fluens quantitas fluxionem habens subsequentem » (Newton 1704, p. 171). Ce passage vient du *Tractatus de quadratura curvarum* que Newton publie en 1704 en appendice de son *Optique*, mais qu'il a pour l'essentiel rédigé dès 1693, dans le cadre d'une ébauche de grand traité de géométrie ; pour le passage parallèle de ce manuscrit, voir NMP, VII, p. 510.

15. Cf. Guicciardini 1999, p. 250-253.

16. « Cette notation, dans la mesure où elle dépend d'un choix arbitraire, ne peut être critiquée si le nombre de points est assez petit pour qu'on le saisisse facilement en comptant ; mais s'il faut inscrire davantage de points, elle occasionne une grande confusion et des embarras plus grands encore. En effet, il est tout à fait incommode de représenter la différentielle ou fluxion dixième de cette manière $\overset{\vdots\vdots}{y}$, alors qu'avec notre manière de noter $d^{10}y$ elle se saisit très facilement. D'autre part, il survient aussi des cas où il faut exprimer des ordres de différentiation encore beaucoup plus élevés et même indéfinis, ce dont la manière des Anglais est absolument incapable. », Euler 1755, p. 100-101. Voir aussi la traduction anglaise de John Blanton : Euler [1755] 2000,

En réalité, déjà dans ses brouillons de 1691, Newton se met à écrire en chiffres le nombre de points quand celui-ci devient trop élevé [17]. Certes, ce ne sont que des abréviations à usage privé, manifestement conçues pour faciliter ses longs calculs et non pour la publication. Mais l'idée est simple et on la voit réapparaître régulièrement, dans le contexte anglais, quand la nécessité s'en fait sentir. Ainsi Taylor, qui utilise la notation newtonienne, précise-t-il que « pour une plus grande généralité, j'écris parfois à la place des points des caractères qui désignent le nombre de points [18] ». Après avoir expliqué la notation newtonienne x́, x̋ des fluentes, il ajoute même :

> Et hæ lineolæ in indicibus fluentûm vim habent punctorum (ut ita dicam) negativorum in indicibus Fluxionum : Sic si sit $n = 2$, & $\overset{n}{x}$ designet $\ddot{x}$, mutato signo designabitur x̋ per $\overset{-n}{x}$ [19].

On a donc non seulement une écriture des ordres de fluxion par des chiffres et des lettres, mais même l'idée de traiter les fluentes comme des fluxions d'ordre négatif, exactement comme le $d^{-1} = \int$ de Leibniz. L'ouvrage de Taylor mériterait une analyse plus approfondie, mais remarquons seulement que ce n'est pas un hasard si cette notation apparaît précisément là : les résultats-clés de Taylor sont des séries infinies comportant des fluxions d'indice arbitraire, qui ont une certaine ressemblance avec la « série très générale » de Bernoulli [20]. Les mêmes causes semblent produire les mêmes effets notationnels, plutôt que l'inverse.

Qu'en est-il de l'« analogie des puissances et des différences » ? Serait-elle possible avec les notations de Taylor, par exemple ? L'ironie est que, si on s'en tient aux notations, elle serait peut-être même *plus facile* : $\overset{n}{x}$ ressemble davantage à x^n que $d^n x$. Cette dangereuse proximité avec les puissances est même le premier reproche qu'adresse Robert Woodhouse à la notation newtonienne lorsqu'un siècle plus tard, en 1803, il publie un argumentaire en faveur de l'adoption de la notation leibnizienne :

> Take a case from [Waring 1785, p. 299]
>
> $$P\left(\frac{\overset{n}{\dot{V}}}{\dot{x}^n}\right) + Q\left(\frac{\overset{n}{\dot{V}}}{\dot{x}^{n-1}\dot{y}}\right) + \&c.;$$

p. 64-65.

17. Cf. NMP, VII, p. 162–163.

18. « majoris generalitatis gratiâ, vice punctorum nonnunquam scribo characteres punctorum numeros designantes […] », B. Taylor 1715, p. 1.

19. « Et ces petites lignes dans les indices des fluentes ont le sens de points pour ainsi dire négatifs dans les indices des fluxions : ainsi, si $n = 2$ et si $\overset{n}{x}$ désigne $\ddot{x}$, alors en changeant le signe x̋ est désigné par $\overset{-n}{x}$ », B. Taylor 1715, p. 2.

20. Il y a d'ailleurs eu une controverse de priorité entre les deux auteurs. Voir *supra*, notes 37 et 38 p. 30.

> this by differential notation is,
>
> $$P.\frac{d^n V}{dx^n} + Q.\frac{d^n V}{dx^{n-1}dy} + \&c.$$
>
> in which it appears to me, that $\dot{x}^n$ is not so clearly distinguished from $\overset{n}{\dot{x}}$ as dx^n is from $d^n x$.[21]

(Waring 1785, que cite ici Woodhouse, utilise $\overset{n}{\dot{x}}$ au lieu du $\overset{n}{\dot{x}}$ de Taylor, mais la différence est minime ; dans tous les cas, il est clair que l'écriture de fluxions d'ordre quelconque ne pose pas de problème particulier aux mathématiciens anglais qui en ont besoin, quoi qu'en dise Euler[22].)

En fait, si les commentateurs de l'« analogie de Leibniz » tiennent souvent pour acquise l'importance de la notation d, c'est parce qu'ils lisent les travaux de Leibniz et Bernoulli à la lumière de développements mathématiques ultérieurs et considèrent donc que ce qui s'y cache, c'est un « calcul opérationnel » ou « calcul des opérations » qui manipulerait algébriquement les opérateurs d ou $\int$. Pourtant, ce n'est pas vraiment ce qu'on trouve dans nos textes. Le mathématicien Lionel Cooper ne s'y trompe pas : lorsqu'il tente de retracer l'histoire longue du calcul opérationnel, il a du mal à le reconnaître chez Leibniz.

> [Leibniz's] analogy seems to be not the one on which the modern symbolic calculus rests [...] but a less far-reaching and more awkward one between differentiation and raising a quantity to a power; to use suggestive notation, not that between $D^n u$ and $a^n u$, but that between $u^{(n)}$ and u^n [...].[23]

C'est si vrai que lorsqu'il présente son analogie, Leibniz semble plus encombré qu'autre chose par ses d, qui induisent un écart avec les puissances ; cela le conduit même à changer la notation usuelle x^n en $p^n x$[24]. Lagrange, comme nous l'avons vu, préférera plus simplement jeter les d par-dessus bord[25].

---

21. « Prenons un exemple tiré de [Waring 1785, p. 299] [...] ce qui en notation différentielle s'écrit [...] où il me semble que $\dot{x}^n$ ne se distingue pas aussi clairement de $\overset{n}{\dot{x}}$ que dx^n de $d^n x$ », Woodhouse 1803, p. xxvii.

22. Edward Waring, qui était l'un des mathématiciens anglais les plus éminents de sa génération (il a occupé pendant près de quarante ans la prestigieuse chaire « lucasienne » de mathématiques de Cambridge), discute amplement, dans ce traité, l'analyse continentale en la traduisant tout naturellement en notation fluxionnelle. Il écrit par exemple qu'Euler a résolu l'équation différentielle « $\overset{n}{\dot{y}} + a\overset{n-1}{\dot{y}}\dot{x} + b\overset{n-2}{\dot{y}}\dot{x}^2 + c\overset{n-3}{\dot{y}}\dot{x}^3 \ldots + Ay\dot{x}^n = 0$ » (Waring 1785, p. v).

23. « [L'analogie de Leibniz] ne semble pas être celle sur laquelle repose le calcul symbolique moderne [...] mais une autre, moins générale et plus maladroite, entre la différentiation et le fait d'élever une quantité à une puissance ; pour utiliser une notation suggestive, non pas l'analogie entre $D^n u$ et $a^n u$ mais celle entre $u^{(n)}$ et u^n [...]. », J. L. B. Cooper 1952, p. 6.

24. Dans une lettre à l'Hôpital, puis dans son mémoire de 1710 ; voir *supra*, p. 55 et 56.

25. Dans le cadre de l'analogie, il écrit x^m pour $d^m x$ (cf. p. 72).

Même le « genre singulier de calcul » de Bernoulli [26] est concevable, voire plus clair, sans les d. Il est bien sûr tentant, puisque Bernoulli manipule algébriquement les d, d'y voir une sorte de calcul des opérations dans lequel la séparation de l'opérateur est cruciale. La raison pour laquelle ce n'est pas si simple, c'est que Bernoulli ne fait pas abstraction des variables auxquels les d s'appliquent : dans son calcul, un d qui s'applique à x est différent d'un d qui s'applique à y [27] ; comme nous l'avons vu, Lenore Feigenbaum suggère de voir sa démarche comme un calcul algébrique avec des quantités d_x, d_y, en mettant x et y en indice. Certes, on pourrait s'appuyer sur cette notation pour donner une véritable interprétation opérationnelle du calcul de Bernoulli, selon laquelle d_x et d_y seraient des opérateurs de différentiation vis-à-vis de x et y respectivement (c'est-à-dire par exemple $d_x(x) = dx$ et $d_x(y) = 0$) et les formules de Bernoulli désigneraient des opérateurs complexes, qu'il suffirait d'appliquer systématiquement à xy pour retrouver les formules différentielles sur lesquelles Bernoulli croit travailler [28]. S'il s'agit de proposer une reconstruction de la méthode de Bernoulli pour en justifier les bons résultats, c'est parfaitement légitime. Toutefois, cela nous conduit assez loin du texte, et donne une importance qu'elle n'a pas à la notion d'opérateur : quitte à modifier les notations de Bernoulli, on resterait sans doute plus près de sa manière de penser en remplaçant $d^n x$ par $\overset{n}{x}$ et en décrétant qu'il faut calculer avec ces quasi-exposants placés à la verticale comme avec les exposants usuels [29]. Là encore, comme l'écrivait Cooper, l'analogie fondamentale est entre $d^n x$ et x^n et l'importance des d n'est qu'apparente.

Bien sûr, le d leibnizien a pu jouer un rôle essentiel plus tard au XVIIIe siècle, et on peut invoquer toutes sortes d'autres raisons de le défendre, par exemple la lisibilité. Mais il ne faut pas le fétichiser, surtout en ce qui concerne nos textes, dans lesquels son importance est en bonne part une illusion rétrospective. Plus largement, comme le montre notre comparaison avec Newton et Taylor, les avantages éventuels des notations de Leibniz ne doivent pas se comprendre en termes de ce qu'elles permettent d'exprimer et qui serait impossible ou impensable sans elles.

26. Cf. section 2.4.

27. « Les lettres x, y, qui dénotent habituellement une quantité, ne doivent pas être considérées comme telles, mais uniquement dans la mesure où elles déterminent d, d^2, d^3 etc. », écrit-il ainsi (voir *supra*).

28. Par exemple, les calculs que Bernoulli croit appliquer à la quantité $d^0xd^3y + d^1xd^2y$ porteraient en fait sur l'opérateur $d_x^0 d_y^3 + d_x^1 d_y^2 = d_y^3 + d_x d_y^2$, sachant que $(d_y^3 + d_x d_y^2)(xy) = xd^3y + dxd^2y$.

29. Les règles de Bernoulli ne se formulent pas plus mal ainsi qu'avec la notation leibnizienne : on aurait $(\overset{3}{y})^2 = \overset{6}{y}$, $\overset{2}{y} \times \overset{3}{y} = \overset{5}{y}$, etc.

3.2 La notation d^n permet-elle d'exprimer des choses auparavant impossibles ?

Admettons : écrire $d^n x$ plutôt que $\overset{n}{\dot{x}}$ n'est pas crucial pour l'analogie. Malgré tout, n'est-il pas indispensable de pouvoir noter n un ordre de différentiation ou de sommation quelconque, qu'on l'écrive en exposant de d, au-dessus des lettres ou ailleurs encore ? Au-delà des particularités de la notation de Leibniz, ce pourrait être cela, l'ingrédient notationnel sans lequel l'analogie serait impossible. Pourtant, là encore, les choses sont moins simples. En un certain sens, on peut réécrire tout ce que font Leibniz et Bernoulli sans la moindre notation générale de l'ordre de différentiation. La nouvelle notation apporte certes quelque chose, mais ce n'est pas, à proprement parler, un pouvoir expressif supérieur.

Commençons par l'analogie sous sa forme la plus simple, qui ne concerne que les différentielles et pas les sommes. Voici comment un résultat apparenté, dit « formule de Leibniz », est présentée dans un manuel de premier cycle contemporain [30] :

$$(fg)^{(n)} = \sum_{p=0}^{n} \binom{n}{p} f^{(p)} g^{(n-p)}$$

La notation $f^{(n)}$, qui indique ici la dérivée n-ième de la fonction f, y intervient partout et paraît indispensable. Il peut sembler évident qu'un tel résultat général, portant sur des dérivées ou différentielles d'ordre quelconque, ne peut être écrit sans une notation générale.

Il faut pourtant se garder de cette évidence : un résultat peut très bien être conçu comme général et pourtant présenté à travers ce qui semble au premier abord être des cas particuliers [31]. C'est exactement ainsi qu'Euler présente l'analogie dans son manuel de calcul différentiel, plus d'un demi-siècle après les investigations de Leibniz et Bernoulli. Le passage en question intervient au chapitre 8, « De la différentiation supérieure des formules différentielles [32] », au cours de la discussion des fonctions de deux variables. Le but d'Euler y est de montrer comment différentier plusieurs fois n'importe quelle formule ; il

30. Deschamps et Warusfel 2003, p. 387.

31. C'est même un phénomène fréquent, qui pose un problème spécifique à l'historien : la variabilité des conceptions et manières d'exprimer la généralité signifie qu'une lecture superficielle ne suffit pas forcément à évaluer la généralité visée, comme le souligne Chemla 2003, 2006 (pour davantage d'exemples, voir le collectif Chemla, Chorlay et Rabouin 2016). Le problème que posent les lettres de Leibniz est similaire, mais inverse : à première vue, sa manière d'exprimer la généralité peut nous paraître évidente et inévitable parce qu'elle est proche de la nôtre, alors qu'il s'agit d'un choix méthodologique très fort (voir *infra*). C'est ce sentiment d'évidence que je veux ébranler ici.

32. « De formularum differentialium ulteriori differentiatione », Euler 1755, p. 204 *sq.* ; cf. aussi la traduction anglaise de John Blanton (Euler [1755] 2000, p. 141 *sq.*) en prenant garde que celui-ci n'a aucun scrupule à moderniser les notations et remplace systématiquement dd par d^2.

commence d'ailleurs en introduisant V « une fonction finie quelconque de x et y » avec $dV = Pdx + Qdy$[33]. Mais après avoir calculé ddV, il délaisse ce niveau de généralité et préfère passer à des exemples :

> Simili modo differentialia tertia ac sequentia invenientur, quod in exemplo particulari ostendisse magis expediet quam formulas generales adhibendo.
>
> Sit igitur $V = xy$;
>
> $$\begin{aligned} \text{Erit } dV &= ydx + xdy \\ ddV &= yddx + 2dxdy + xddy \\ d^3V &= yd^3x + 3dyddx + 3ddydx + xd^3y \\ d^4V &= yd^4x + 4dyd^3x + 6ddxddy + 4dxd^3y + xd^4y \\ &\&c. \end{aligned}$$
>
> in quo exemplo coefficientes numerici legem potestatum binomii sequuntur indeque, quousque libuerit, continuari possunt[34].

Euler procède par l'exemple en un double sens : il traite des cas particuliers comme $V = xy$ plutôt qu'un V quelconque, et calcule dV, ddV, d^3V et d^4V plutôt que d^nV. Cet extrait, d'ailleurs, est représentatif du traitement qu'Euler réserve aux différentielles ; dans les près de 900 pages de ce traité, je n'ai trouvé *que deux passages* où figure un d^n avec n quelconque[35]. Il est vrai que Leibniz, lui, utilise cette écriture, à la fois lorsqu'il présente son analogie au marquis de l'Hôpital et lorsqu'il la publie[36], mais c'est plutôt lui l'exception : même son proche collaborateur Jean Bernoulli la présente à Varignon *via* un cas particulier[37]. En fait, Leibniz a des préoccupations méthodologiques très spécifiques, comme nous le verrons.

Or si, comme Euler, on présente l'analogie par l'exemple, écrire d^n n'est plus nécessaire. En fait, même la notation d^3, d^4 n'est pas clairement requise : contrairement à Leibniz qui, dans ce contexte, écrit soigneusement d^0x, d^1x, d^2x, Euler s'en tient aux notations usuelles x, dx, ddx, et on pourrait très bien imaginer qu'il écrive aussi $dddx$ plutôt que

33. « Denotet ergo V functionem quamcunque finitam ipsarum x et y sitque $dV = Pdx + Qdy$ », Euler 1755, p. 207.

34. « On trouvera de la même manière les différentielles troisièmes et les suivantes, ce qu'il sera plus à propos de montrer sur un exemple particulier qu'en exhibant des formules générales. Soit donc $V = xy$; on aura [...]. Dans cet exemple, les coefficients numériques suivent la loi des puissances d'un binôme et de ce fait peuvent être continués aussi loin qu'on voudra. », Euler 1755, §249, p. 209–210 ; Euler [1755] 2000, p. 144-145.

35. Cf. Euler 1755, p. 324-326, 587. En revanche, les puissances quelconques x^n y sont très nombreuses.

36. Voir *supra*, p. 55 et 56.

37. Cf. lettre de Jean Bernoulli à Varignon du 24 décembre 1697, BBW, II, p. 154–155.

d^3x. Certes, la notation exponentielle rend sans doute l'analogie plus visible ; mais pour bien comprendre cette idée, à laquelle je reviendrai, il faut d'abord voir qu'on peut faire sans.

Pourrait-on aussi éliminer la notation $d^\bullet$ (sous ses différentes formes : d^3, d^{-1}, d^n, etc.) des formules plus complexes de Leibniz, celles qui valent aussi pour les intégrales ? Prenons un exemple. Comme nous l'avons vu [38], Leibniz introduit la notation $d^{-1} = \int$ pour pouvoir écrire la formule [39]

$$\int \overline{\overset{e}{z} \overset{m}{d} n} = \overset{e}{z} \overset{m-1}{d} n - e \cdot \overset{e-1}{z} \overset{m-2}{d} n \cdot dz + e \cdot e - 1 \cdot \overset{e-2}{z} \overset{m-3}{d} n \cdot \overline{dz}^2 - e \cdot e - 1 \cdot e - 2 \cdot \overset{e-3}{z} \overset{m-4}{d} n \cdot \overline{dz}^3 \text{ etc.}$$

qu'il obtient par une méthode analogue à une application répétée de notre « intégration par parties ». Il y a une difficulté apparente pour éliminer la notation $d^{-1} = \int$, qui est que la règle de formation des termes devrait changer au bout de m étapes : quand on passe d'un terme au suivant, $d^{m-1}n$ devient $d^{m-2}n$ puis $d^{m-3}n$ et ainsi de suite jusqu'à atteindre dn puis n, et il faut ensuite passer à $\int n$, $\iint n$. Mais comme je viens de le faire, cette règle peut parfaitement s'expliquer en langue naturelle sans introduire de nouvelle convention notationnelle. Il y a plus : on peut éliminer non seulement d^n mais même d^3, d^4 etc. en procédant *via* des exemples, comme Euler plus haut. Leibniz lui-même, juste après avoir donné la formule générale, traite les cas $m = 1$ et $m = 2$ [40] :

$$\int \overline{\overset{e}{z} dn} = z^e n - e \cdot \overset{e-1}{z} \int \overline{n} + e \cdot e - 1 \cdot \overset{e-2}{z} \iint n - e \cdot e - 1 \cdot e - 2 \cdot \overset{e-3}{z} \int^3 n, \text{ etc.}$$

$$\int \overline{z^e ddn} = z^e dn - e \cdot \overset{e-1}{z} n + e \cdot e - 1 \cdot \overset{e-2}{z} \int n - e \cdot e - 1 \cdot e - 2 \cdot \overset{e-3}{z} \int^2 n, \text{ etc.}$$

Ici, la notation exponentielle est quasiment absente, et on pourrait en éliminer la dernière trace en remplaçant $\int^3$ par $\iiint$, $\int^2$ par $\iint$. Or ces deux exemples auraient certainement suffi à Bernoulli pour en saisir la généralité ; on peut d'ailleurs conjecturer que c'est ainsi que lui-même aurait procédé plutôt que d'introduire une notation exotique. Encore une fois, la notation exponentielle s'avère en principe dispensable.

À ce stade, nous pouvons mieux comprendre comment il faut interpréter une affirmation comme celle de Michel Serfati citée en tête de chapitre. Quand il écrit que « [la]

38. Cf. section 2.2.
39. Je donne ici la formule que donne Bernoulli dans sa réponse du 20–30 avril 1695, où il corrige les fautes de calcul de Leibniz ; cf. p. 48.
40. L'erreur de calcul initiale de Leibniz se répercute dans ces cas particuliers, cf. note 31 p. 49. J'ai ici corrigé les formules pour éviter une distraction inutile.

substitution […] de $A = x^2 + x + 2$ dans $A^2 + A$ demeura une opération inconcevable […] *dans le cadre de l'écriture médiévale des mathématiques* » (je souligne), la dernière clause introduit une nuance cruciale : l'opération est parfaitement concevable sans la notation moderne des puissances, mais il faut pour cela mobiliser des ressources extérieures aux signes utilisés [41]. C'est exactement ce que l'on observe dans le cas de l'intégrale $\int \overline{z^{e} \cdot dn}$. À se concentrer sur les notations prises isolément, on risque de conclure qu'il est impossible de l'exprimer sans écrire d^n ou quelque chose d'équivalent. Ce serait une erreur, parce que les notations ne fonctionnent jamais seules : elles peuvent être accompagnées d'explications en langue naturelle, et en tout cas sont toujours interprétées par un lecteur qui peut, à partir d'exemples, comprendre une règle générale.

Nous sommes maintenant en mesure de mieux poser le problème des notations. Il ne faut pas se demander si elles rendent possible quelque chose qui ne l'était pas : on peut toujours trouver moyen de faire sans, et il y a fort à parier qu'en cherchant bien, on trouvera même quelqu'un qui l'a fait. Je propose plutôt de comparer la manière de procéder *sans* et la manière de procéder *avec* telle nouvelle notation, pour mieux comprendre ce que la nouvelle notation apporte. En un sens, cette démarche est une instance de la méthodologie comparative mise en avant par Ken Manders [42]. Le but est alors de clarifier et préciser l'idée qu'une bonne notation peut rendre certaines choses plus « simples », certaines découvertes plus « faciles », etc.

3.3 Voir des régularités dans les formules

Dans notre épisode, le rôle le plus évident et le plus explicite des changements de notation est celui de rendre l'analogie plus claire (« *analogiæ clarioris causa* [43] », écrit Leibniz lorsqu'il introduit la notation $p^n x$ pour x^n) ou visible en un coup d'œil (« *ut […] totum uno oculi ictu perspiciatur* [44] », écrit Lagrange, qui utilise x^n au lieu de $d^n x$). Au contraire, comme l'écrit Bernoulli à Varignon, si après avoir trouvé $d^4(xy)$ *via* l'analogie on élimine les $d^\bullet$, cela dissimule le procédé :

> pour cacher maintenant l'artifice, pour $d^0 x$, $d^1 x$, $d^0 y$, $d^1 y$ j'écrirois simplement x, dx, y, dy, &c. ce qui me donneroit pour la differentielle cherchée une

41. C'est bien cela que constate Serfati dans son exemple, qui vient d'un passage de Michael Stifel : Serfati 2006, p. 107-111.
42. Lors de nombreux séminaires et communication personnelle ; voir aussi ses brouillons non publiés, par exemple Manders 1999.
43. Leibniz 1710b, p. 161.
44. Voir n. 109 p. 72.

expression en termes ordinaires $xddddy + 4dxdddy + 6ddxddy + 4dddxdy + yddddx$ [45].

Cela ne veut pas dire que la notation crée l'analogie : en un sens, il y a toujours une similarité de structure entre la formule précédente, malgré tous ses d, et

$$(m+y)^4 = n^0 m^4 + 4n^1 m^3 + 6n^2 m^2 + 4n^3 m^1 + n^4 m^0.$$

Il y a le même nombre de termes, les mêmes coefficients, et les nombres de d devant x ou y correspondent aux exposants de m ou n. La différence est que cette relation est bien moins apparente ; ce que les notations permettent, c'est de la rendre *plus saillante*.

Dans les découvertes de Leibniz aussi, les notations ont certainement joué un rôle semblable ; le phénomène n'est donc pas limité aux contextes expositoires des citations que je viens de donner. En effet, comme nous l'avons vu [46], une étape importante pour Leibniz a sans doute été de reconnaître une analogie entre

$$\int \overline{\overset{e}{z\cdot} \overset{m}{d\cdot} n} = \overset{e}{z\cdot} \overset{m-1}{d\cdot} n - e \cdot \overset{e-1}{z\cdot} \overset{m-2}{d\cdot} n \cdot dz + e \cdot e - 1 \cdot \overset{e-2}{z\cdot} \overset{m-3}{d\cdot} n \cdot \overline{dz}^2 \\ - e \cdot e - 1 \cdot e - 2 \cdot \overset{e-3}{z\cdot} \overset{m-4}{d\cdot} n \cdot \overline{dz}^3 + \text{ etc.}$$

et la formule du binôme de Newton. Les deux formules en question sont obtenues par des procédés très différents [47], ce qui rend d'autant plus probable que l'observation de Leibniz ait été authentiquement visuelle : l'analogie réside dans les formules elles-mêmes, pas dans leur méthode de construction. Or cette observation est certainement facilitée par la notation $d^\bullet$, exactement comme dans le cas de $d^4(xy)$ et $(n+m)^4$ que discute Bernoulli dans sa lettre à Varignon. Bien sûr, là encore, la découverte n'aurait pas été impossible sans la notation ; en principe, Leibniz aurait pu n'écrire que des cas particuliers en utilisant des suites de d, et pourtant se rendre compte qu'en passant d'un terme au suivant, le nombre de d devant n diminuait et le nombre de d devant z augmentait, et ainsi pressentir un rapport – que par ailleurs il appelait de ses vœux – avec la formule du binôme. Mais il faut bien admettre que la notation $d^\bullet$ facilite ce constat considérablement.

45. Lettre de Jean Bernoulli à Varignon du 24 décembre 1697, BBW, II, p. 155.

46. Cf. section 2.3.

47. Leibniz obtient sa formule pour $\int \overline{\overset{e}{z\cdot} \overset{m}{d\cdot} n}$ par la méthode de Bernoulli, apparentée à notre intégration par parties (cf. section 2.2). Le cas de la formule du binôme est plus complexe. Pour des exposants entiers positifs, elle peut être obtenue par des considérations combinatoires ; dans le cas d'exposants négatifs ou fractionnaires, on peut utiliser la méthode de Mercator (cf. *supra*, note 67 p. 61). Newton procède autrement (en s'inspirant des méthodes de Wallis, cf. Panza 2005), mais il est probable que Leibniz n'ait pas eu connaissance des détails de ses travaux. Quoi qu'il en soit, dans le cas général, Leibniz ne peut guère concevoir la formule que *via* la règle de formation de ses termes successifs.

3.4 Construire des théorèmes généraux

Les notations de Leibniz lui permettent aussi de rassembler en théorèmes généraux ce qui serait sans cela des formules éparses. Cet avantage est plus subtil qu'il n'y paraît. N'ai-je pas soutenu plus haut [48] que même sans $d^{-1} = \int$, il est possible, en passant par des cas particuliers, d'exprimer un résultat aussi général que sa formule pour $\int \overline{z^{\overset{e}{\cdot}}\, d^{\overset{m}{\cdot}}\, n}$? À ce stade de son investigation, c'est exact, mais si l'on continue l'exercice d'éliminer les $d^{\bullet}$, on recontre une difficulté instructive.

Nous l'avons vu, Leibniz introduit $d^{-1} = \int$ afin que les termes successifs de son développement en série pour $\int \overline{z^{\overset{e}{\cdot}}\, d^{\overset{m}{\cdot}}\, n}$ suivent une loi uniforme. Mais il a immédiatement l'idée d'utiliser cette nouvelle notation *dans l'expression même de l'intégrale à calculer*, c'est-à-dire de prendre $m = -r$ [49] :

$$\int \overline{z^e \int^{\overset{r}{\cdot}} \overline{n}} = z^{\overset{e}{\cdot}} \int^{\overset{r+1}{\cdot}} \overline{n - e \cdot z^{\frac{e-1}{\cdot}}} \int^{\frac{r+2}{\cdot}} \overline{n + e \cdot e - 1 \cdot z^{\frac{e-2}{}}} \int^{\frac{r+3}{\cdot}} \overline{n - e \cdot e - 1 \cdot e - 2 \cdot z^{\frac{e-3}{\cdot}}} \int^{\frac{r+4}{}} n, \text{ etc.}$$

On peut jouer au même petit jeu que ci-dessus et remplacer cette formule par des cas particuliers comme $r = 1$,

$$\int \overline{z^e \int \overline{n}} = z^{\overset{e}{\cdot}} \iint n - e \cdot z^{\frac{e-1}{\cdot}} \iiint n + e \cdot e - 1 \cdot z^{\frac{e-2}{}} \iiiint n, \text{ etc.}$$

puis aussi $r = 2$, et expliquer que le résultat est généralement valable pour autant d'intégrales qu'on veut ; ainsi, la notation exponentielle serait à nouveau éliminée. La question importante, cependant, est ailleurs : est-ce *le même résultat* que celui qui donne $\int \overline{z^{\overset{e}{\cdot}}\, dn}$ et $\int \overline{z^{\overset{e}{\cdot}}\, ddn}$? Au fond, la question de ce qui fait l'unité des formules contenant dn, ddn, $dddn$ se pose déjà, mais elle est beaucoup plus aiguë ici, où l'on y ajoute $\int n$, $\iint n$, etc. Dans les deux cas, on pourrait dire que ce qui les rassemble est la règle de formation des termes successifs : pour passer de l'intégrale à calculer au premier terme, il faut (a) supprimer le $\int$ initial et (b) supprimer un d ou ajouter un $\int$ au terme en n ; pour passer d'un terme du développement au suivant, il faut (a) supprimer un d ou ajouter un $\int$ au terme en n, et (b) différentier le terme en z. Si, comme je viens de le faire, on traite et écrit d et $\int$ différemment, on obtient alors une règle un peu complexe parce que disjonctive. Mais il y a plus. L'*énoncé même* du résultat devrait juxtaposer deux familles de cas différents : d'une part $\int \overline{z^{\overset{e}{\cdot}}\, dn}$ et ses analogues avec ddn, $dddn$, etc., d'autre part $\int \overline{z^e \int \overline{n}}$ et ses analogues avec $\iint$ etc. En bref, ce serait un drôle d'énoncé. Pourquoi en ferait-on un théorème unique ? On

48. Section 3.2.

49. Voir *supra*, p. 49. Encore une fois, j'ai corrigé les erreurs de calcul de Leibniz.

pourrait se contenter d'y voir diverses applications d'une même méthode, celle qui conduit Bernoulli à sa série très générale puis Leibniz aux généralisations de celle-ci.

Ce problème est encore plus clair pour l'analogie généralisée de Leibniz [50] :

$$d^e\overline{xy} = d^e x d^0 y + \frac{e}{1} d^{e-1} x \cdot d^1 y + \frac{e \cdot e - 1}{1 \cdot 2} d^{e-2} x \cdot d^2 y + \frac{e \cdot e - 1 \cdot e - 2}{1 \cdot 2 \cdot 3} d^{e-3} x \cdot d^3 y \text{ etc.}$$

Si on cherche à la traduire sans $d^\bullet$, on obtient d'un côté les formules finies donnant les différentielles supérieures d'un produit (cas $e > 0$), de l'autre des développements infinis qui rappellent la série universelle de Bernoulli (cas $e < 0$). Là encore, le problème n'est pas une impossibilité d'exprimer le résultat sans la nouvelle notation ; c'est plutôt qu'au lieu d'une formule unique, on en obtient plusieurs qui n'ont plus de rapport clair les unes avec les autres. Le contenu du théorème serait conservé… sauf qu'on n'aurait plus de théorème.

En fait, chercher autant que possible des formules générales est un choix méthodologique fort de Leibniz [51]. Il croit de longue date à l'importance de rassembler des cas apparemment différents en des formules uniques, comme le montre déjà un manuscrit de 1674, « De la methode de l'universalité », qui commence ainsi :

> La Methode de l'universalité nous enseigne de trouver par une seule operation des formules analytiques et des constructions geometriques generales pour des sujets ou cas differens dont chacun sans cela auroit besoin d'une analyse ou synthese particuliere [52].

Tout l'intérêt, écrit-il plus loin, de « la Reduction de plusieurs Cas differens à une seule formule, regle, equation ou construction » est qu'elle « diminue la peine [53] ». Cela peut faire écho à un thème très répandu, sous diverses formes, dans les mathématiques du XX^e^ siècle, où il est fréquent de louer le choix de « bons concepts » ou de « bonnes définitions » pour l'« économie de pensée » qu'ils procurent en synthétisant une multitude de phénomènes particuliers. La différence, c'est que c'est ici au choix des « caractères » ou « signes », donc à ce que nous appelons les notations, que Leibniz fait jouer ce rôle unificateur. Cette démarche est très nette dans notre exemple : c'est bien l'uniformité notationnelle permise par $d^{-1} = \int$ qui justifie de considérer comme un même résultat des formules qui, sans cela, seraient distinctes.

---

50. Cf. p. 55.

51. Pour plus de détails sur l'importance que Leibniz accorde à la généralité, et les valeurs qu'il lui associe (concision, beauté, mémorabilité, etc.) voir Knobloch 2016. Michel Serfati 2001, section 5 fait même de la recherche d'« harmonie », c'est-à-dire d'unité permettant de saisir par une formule unique une diversité de cas particuliers, le principe fondamental de la méthode leibnizienne.

52. Leibniz 1903, p. 97.

53. *Id.*, p. 98.

Il faut préciser que, chez Leibniz, la relation entre notations et formules générales est à double sens. D'un côté, les notations sont choisies pour permettre la généralité ; mais en retour, la possibilité même d'écrire des formules concises et générales suggère que les notations qui les permettent sont bien fondées. Dans notre cas, Leibniz pose $d^{-1} = \int$ simplement pour pouvoir écrire sa première formule pour $\int \overline{z^{\overset{e}{\cdot}} d^{\overset{m}{\cdot}} n}$; quand il se rend compte que la formule reste valable en remplaçant $d^{\overset{m}{\cdot}} n$ par $d^{\overset{-m}{\cdot}} n = \int^{\overset{m}{\cdot}} n$, c'est-à-dire que sa notation permet d'écrire une formule légèrement plus générale que ce qu'il avait initialement prévu, il écrit que cela « confirme » $d^{-1} = \int$ [54], et il déclare même à l'Hôpital quelques jours plus tard que le calcul des différences « est le même avec celuy des sommes, l'un estant seulement reciproque de l'autre [55] ». Pour Leibniz, la possibilité d'écrire une formule générale simple confère donc immédiatement à $d^{-1} = \int$ un statut privilégié. On explique souvent cette tendance leibnizienne à traiter la simplicité comme indice de la vérité en la liant à ses positions philosophiques et théologiques [56], et Serfati va même jusqu'à qualifier les pratiques symboliques leibniziennes d'« irrationalistes [57] ». Il me semble cependant qu'on pourrait gagner en perspective en établissant un parallèle avec les jugements de « naturalité » émis dans le cadre de mathématiques plus récentes [58] : la généralité peut être l'un des buts visés lorsqu'on élabore de nouvelles définitions, mais en retour, il n'est pas rare de considérer des définitions qui permettent des énoncés simples et généraux comme « naturelles ». Il y a là un brouillage tout à fait comparable entre considérations apparemment épistémiques et questions ontologiques, sauf qu'encore une fois, c'est dans notre cas pour les *notations* que des questions de généralité et de naturalité se posent.

3.5 Une réorganisation du paysage

En dernier lieu, une fois adoptées, les notations $d^{\bullet}$ changent ce qui peut s'exprimer de manière simple et par là réorientent les recherches de Leibniz et Bernoulli. C'est déjà vrai de la formule générale pour $d^e(xy)$, qui tire une grande partie de son intérêt du fait qu'elle peut s'exprimer simplement dans la nouvelle notation (tout en légitimant cette notation en retour) : comme nous l'avons vu, si on n'utilisait pas la notation $d^{\bullet}$, cette formule dégénérerait en un énoncé artificiel et disjonctif, sans intérêt au-delà de celui des cas particuliers qui la constituent. C'est plus vrai encore de la formule que donne Leibniz [59] pour $d^{\frac{1}{2}}$, qui tire

54. Voir *supra*, p. 49.
55. Cf. p. 50.
56. Voir par exemple Knobloch 2016.
57. Serfati 2008.
58. Jamie Tappenden, dans ses travaux, a clairement problématisé cette question ; cf. par exemple 2008a,b.
59. Cf. section 2.5.

toute sa plausibilité du fait qu'elle peut alors s'écrire de manière apparemment naturelle. Toutefois, le meilleur exemple est sans doute l'étrange calcul de Bernoulli [60].

Si on formulait les règles de calcul de Bernoulli sans utiliser la notation $d^\bullet$, on les trouverait certainement difficiles et contournées. On aurait par exemple

$$\frac{dddxdy}{dxddy} = ddx \int y, \quad dddxdy \times dy \int x = ddxddy, \quad \frac{dxdy}{dy \int x} = yddx,$$

et ainsi de suite. Certes, il reste possible de les formuler ainsi, quoique cela suppose sans doute plus ou moins de concevoir l'intégrale comme une sorte d'inverse algébrique de la différentielle. Mais on voit bien que la médiation de la notation $d^{-1} = \int$ joue un rôle important pour faire de ces règles une voie qu'il est raisonnable d'explorer.

60. Voir section 2.4.

Synthèse : un pari notationnel

L'importance des nouvelles notations de Leibniz n'est pas qu'elles permettent d'écrire des choses qui, sans elles, seraient impensables. Il n'est pas vrai non plus que la notation newtonienne n'aurait rien permis de similaire. Pour autant, écrire d^4, d^n ou d^{-3} plutôt que de s'en tenir à des $dddd$ et $\int\!\!\int\!\!\int$ (ou faire l'équivalent du côté newtonien), cela a des conséquences. Cela fait ressortir plus clairement des régularités dans des formules complexes qui, autrement, seraient beaucoup moins nettes. Cela change aussi le paysage mathématique ouvert à l'exploration : des résultats qui ne s'exprimaient que de manière complexe et disjonctive deviennent simples et paraissent plausibles ; de nouvelles voies deviennent prometteuses à explorer.

De la sorte, les notations introduites par Leibniz semblent effectivement aider nos auteurs dans leurs découvertes. Ce n'est pas un simple accident, dans la mesure où Leibniz avait des raisons de les choisir de préférence à d'autres. Mais ce n'est pas non plus un choix fait en pleine connaissance de cause, parce qu'en fin de compte, les motivations initiales de Leibniz n'ont qu'un lien indirect et imparfait avec les découvertes qui en sortent. Au fond, ses choix notationnels sont un pari, un pari informé sans doute, mais un pari tout de même, qui aurait aussi bien pu ne mener nulle part.

Deuxième partie

Équivalence informationnelle, différences computationnelles

Sommaire de la deuxième partie

Introduction : un slogan séduisant

Herbert Simon introduit en 1978 un slogan qui a, depuis, rencontré un certain succès tant en sciences cognitives qu'en philosophie : deux représentations peuvent être *informationnellement* équivalentes mais *computationnellement* différentes [61]. Pour lui, ce qui distingue avant tout images et phrases, par exemple, n'est donc pas qu'elles auraient un contenu différent, mais qu'elles rendraient différentes opérations plus ou moins faciles ou difficiles.

Pour une analyse des représentations utilisées en sciences ou en mathématiques, les idées de Simon ont deux attraits principaux. Tout d'abord, sa définition de l'équivalence informationnelle est séduisante parce qu'elle permet, du moins à première vue, de comparer des représentations en contournant la question délicate de savoir de quoi elles parlent. Voici comment Simon introduit son projet :

> It is impossible to find an entirely neutral language in which to describe representations of information, for a language is itself a form of representation. We can avoid this difficulty, at least in part, by not attempting to describe representations directly, but by talking instead in terms of the equivalence of representations. […] Specifically, I shall talk about *informational* equivalence and *computational* equivalence. […] Two representations are informationally equivalent if the transformation from one to the other entails no loss of information, i.e., if each can be constructed from the other. [62]

Cet usage du terme d'information s'inscrit dans le prolongement de la « théorie mathé-

61. Sa première publication sur le sujet est Simon 1978, mais son article le plus lu est Larkin et Simon 1987.

62. « Il est impossible de trouver un langage entièrement neutre pour décrire des représentations d'information, parce que tout langage est lui-même une forme de représentation. Nous pouvons éviter cette difficulté, au moins en partie, en ne tentant pas de décrire les représentations directement, mais en parlant à la place en termes d'équivalence de représentations. […] Plus précisément, je parlerai d'équivalence *informationnelle* et d'équivalence *computationnelle*. […] Deux représentations sont informationnellement équivalentes si la transformation de l'une à l'autre n'implique pas de perte d'information, c'est-à-dire si chacune peut être construite à partir de l'autre. » (Simon 1978, p. 4 ; l'auteur souligne.)

matique de la communication » (souvent appelée, justement, « théorie de l'information ») développée au milieu du XXe siècle, en particulier dans la formulation de Claude Shannon [63] :

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. The semantic aspects of communication are irrelevant to the engineering problem. [64]

Comme l'écrit George Markowsky dans son introduction au travail de Shannon,

> A key step in Shannon's work was his realization that, in order to have a theory, communication signals must be treated in isolation from the meaning of the messages that they transmit. This view is in sharp contrast with the common conception of information, in which meaning has an essential role. [65]

Quoiqu'il n'emploie pas l'appareillage technique de Shannon, Simon se repose sur cette intuition-là : on peut dire qu'une représentation contient « la même », ou « plus », ou « moins » d'information qu'une autre sans se demander ce qu'est cette information exactement ou sur quoi elle porte. Par exemple, si je peux, à partir d'une figure géométrique, en construire une description en phrases, puis à partir de cette description reconstruire la figure, alors je peux dire que ces deux représentations sont informationnellement équivalentes sans jamais avoir à me demander de quoi parle la géométrie.

En second lieu, à travers son idée de différences computationnelles, Simon offre un modèle clair pour comprendre comment deux représentations informationnellement équivalentes peuvent néanmoins différer. Ce modèle vient de l'informatique. D'une manière générale, quand on conçoit un logiciel (mettons un logiciel de courrier électronique), il y a de nombreuses manières d'organiser les données à traiter (par exemple des courriers électroniques) dans la mémoire de l'ordinateur. Quelle que soit la manière dont on les organise,

63. Pour une introduction claire à cette théorie, voir Markowsky 2017. Pour un manuel récent, voir par exemple Yeung 2008.

64. « Le problème fondamental de la communication est de reproduire à un endroit, de manière exacte ou approximative, un message sélectionné en un autre endroit. Fréquemment, les messages ont du *sens*, c'est-à-dire qu'ils renvoient ou sont corrélés, d'après un système ou un autre, à certaines entités physiques ou conceptuelles. Les aspects sémantiques de la communication n'ont aucune importance pour le problème d'ingénierie. » (Shannon 1948, p. 379 ; l'auteur souligne). Je remercie Gianni Gastaldi pour avoir attiré mon attention sur ce texte.

65. « Une étape-clé du travail de Shannon a été de se rendre compte que, pour avoir une théorie, les signaux de communication devaient être traités séparément du sens des messages qu'ils transmettent. Cette perspective est en nette opposition avec la conception usuelle de l'information, pour laquelle le sens joue un rôle essentiel. » (Markowsky 2017, p. 1).

ce seront les mêmes données ; on aura donc équivalence informationnelle. Cependant, le mode d'organisation choisi peut faire une différence cruciale : il peut rendre certaines opérations de traitement des données plus ou moins faciles ou difficiles, c'est-à-dire plus ou moins lentes ou rapides en temps de calcul ou plus ou moins gourmandes en mémoire. Pour rendre cette intuition précise, Simon se tourne vers des concepts venus de l'informatique théorique, ceux de *type de données* et de *structure de données*, puis leur assimile la notion de représentation. En d'autres termes, pour Simon, images et phrases (par exemple) sont fondamentalement deux manières différentes manières d'organiser les mêmes données, qui se prêtent bien à différentes opérations.

Les idées de Simon semblent donc promettre un cadre clair pour progresser sur notre problème. Il y a cependant des difficultés. Nous verrons au chapitre 4 que le travail de Simon s'ancre dans une méthodologie très spécifique d'analyse de la cognition à travers le modèle de l'ordinateur ; de surcroît, le contexte initial de ses articles est la psychologie cognitive et plus précisément le problème des « images mentales ». Or quel que soit l'intérêt heuristique de ses concepts, il n'est pas évident, du moins sans en sacrifier la rigueur, de les transposer à des représentations non plus mentales mais « externes », comme les figures ou les formules qui nous intéressent ici. Nous aurons l'occasion de nous en rendre compte au chapitre 5, qui passe en revue diverses tentatives récentes d'appliquer la terminologie de Simon aux représentations utilisées en sciences, et plus spécifiquement en mathématiques ; je tente aussi de m'en inspirer pour éclairer l'étude de cas de la partie 1. À l'issue de ce survol de la littérature, le chapitre 6, plus analytique, vise à faire le point : sous quelles conditions peut-on appliquer le modèle informatique de Simon, de manière rigoureuse, aux représentations utilisées en mathématiques ? Que nous apporte-t-il et quelles en sont les limites ?

Chapitre 4

Herbert Simon : représentations et types de données

L'idée d'une équivalence informationnelle associée à des différences computationnelles est régulièrement reprise dans la littérature récente. Toutefois, ce n'est souvent que de manière allusive, et par des auteurs qui sont loin de partager les postulats sur lesquels elle était initialement fondée. Avant toute discussion plus approfondie, il me semble donc important de comprendre ce que Simon lui-même cherchait à accomplir. C'est le but de ce chapitre.

4.1 Sciences cognitives et formes de représentation

Les travaux de Simon sur les représentations s'inscrivent dans le contexte de ce que nous appellerions aujourd'hui les sciences cognitives, et plus précisément d'un débat qui agite ce champ dans les années 1970 (et au-delà) : le débat sur l'« imagerie mentale ».

Il n'est pas question de retracer ici en détail l'histoire des sciences cognitives. Pour aller vite, disons seulement qu'à partir des années 1950 se développent de nouveaux programmes de recherche sur le comportement humain qui partagent certains postulats liés au modèle de l'ordinateur [1]. Tout d'abord, « ce qui se passe à l'intérieur de l'individu » peut

1. On parle parfois de « cognitivisme » (Haugeland 1978), ou d'un paradigme cognitiviste qui unirait des chercheurs venus de domaines divers. Le meilleur point de départ pour explorer le domaine ainsi délimité est certainement la collection d'articles Haugeland 1981a, à commencer par son excellente introduction Haugeland 1981b (pour une discussion plus approfondie, voir Haugeland 1985). Pour des introductions rapides en français, on pourra consulter Andler 1990, p. 99-102, Andler 2004 ou, pour une présentation qui souligne les liens avec la théorie mathématique de la calculabilité, P. Wagner 1998, chap. VII. D'un point de vue historique, il faut toutefois souligner la diversité des chercheurs qui seront ultérieurement rattachés aux sciences cogni-

être étudié et décrit au niveau physiologique (et en particulier neurophysiologique) *mais aussi* à un niveau plus abstrait en termes d'opérations de traitement de l'information, de la même manière qu'un ordinateur peut être décrit au niveau matériel mais aussi en termes de programmes. (C'est ce qu'en philosophie on appelle souvent *fonctionnalisme*.) Ensuite, cette description « informationnelle » est formulée en termes d'états mentaux représentationnels (ou « *représentations mentales* ») et de processus qui opèrent sur ces représentations. Beaucoup d'auteurs poussent plus loin encore l'analogie avec l'informatique, ou plus précisément avec les modèles logico-mathématiques abstraits des ordinateurs (comme les machines de Turing). Cela les conduit à un postulat supplémentaire : les représentations mentales en question sont des systèmes de symboles, et les processus qui les manipulent sont des processus *effectifs* au sens de la théorie mathématique de la calculabilité. Cette thèse forte, qui implique que la cognition humaine peut en théorie être simulée par un ordinateur, est particulièrement mise en avant dans une discipline qui joue un rôle structurant au sein des sciences cognitives naissantes : l'intelligence artificielle. Herbert Simon en est l'un des pionniers, et en écrit avec Allen Newell un manifeste resté canonique [2].

C'est dans ce cadre, où l'existence de représentations mentales est généralement admise, qu'intervient la controverse sur l'existence d'images mentales. L'intérêt pour mon propos est que ce débat soulève des questions de fond : qu'est-ce que cela veut dire, représenter le monde *sous forme d'images* ? Plus généralement, comment comprendre l'idée qu'on puisse représenter la même chose sous des formes différentes ? Ce sont exactement les questions que Simon se pose :

> The human brain encodes, modifies, and stores information that is received through its various sense organs, transforms that information by the processes that are called "thinking," and produces motor and verbal outputs of various kinds based on the stored information. So much is noncontroversial [...]. What *is* highly controversial is *how* information is stored in the brain—in the usual terminology, how it is "represented"—or even how we can describe representations, and what we mean when we say that information is represented in

tives, et noter que l'unité théorique de ces courants ne se construit que progressivement et n'apparaît au grand jour que dans les années 1970. Par ailleurs, le choix d'isoler comme essentiels les postulats que je reprends ici (qui sont fréquemment mis en avant) est marquée par une tendance à accorder une place théorique centrale aux travaux d'intelligence artificielle comme ceux de Newell et Simon, et répond moins à un objectif historique qu'à une volonté de systématisation philosophique, qui peut être déterminée par des débats extérieurs au champ (problème corps-esprit, matérialisme, etc.) ou par des polémiques pour ou contre diverses tentatives de constituer des paradigmes rivaux (connexionnisme, cognition incarnée, bayésianisme, etc.). Ma brève introduction ne vise donc qu'à fournir une reconstruction logique à grands traits du débat sur l'imagerie et à situer Simon, à la pensée duquel ma brève présentation me semble suffisamment fidèle.

2. Newell et Simon 1976, reproduit dans Haugeland 1981a.

FIGURE 4.1 – Exemples de paires de dessins utilisées par Shepard et Metzler (1971) : à gauche, deux figures qui peuvent s'obtenir l'une de l'autre par rotation (ici dans le plan de l'image), à droite deux figures qui ne le peuvent pas

one way rather than another. [3]

La controverse est déclenchée par certains résultats de psychologie expérimentale, dont sans doute le plus célèbre est dû à Roger Shepard et Jacqueline Metzler (1971). Ceux-ci ont présenté à leurs sujets expérimentaux des dessins d'objets tridimensionnels, par paires (figure 4.1), et leur ont demandé dans chaque cas si les deux objets dessinés pouvaient, ou non, s'obtenir l'un de l'autre par rotation. Shepard et Metzler ont alors mesuré le temps de réponse des sujets, et c'est là qu'arrive le résultat important : le délai nécessaire avant d'arriver à une réponse positive est en moyenne proportionnel à l'angle de la rotation mettant les deux objets en coïncidence, comme si les sujets pratiquaient une rotation mentale à une vitesse déterminée. L'interprétation de ce type de résultats a donné lieu à une controverse vive et durable, qui ne porte pas seulement sur l'existence d'images mentales mais aussi sur la cohérence même de cette notion, souvent accusée de reposer sur une analogie obscure avec les images externes [4].

Sans entrer en détail dans l'écheveau entremêlé d'ambiguïtés et de confusions qui traverse ces débats, il importe, pour situer Simon, de distinguer plusieurs oppositions possibles. Tout d'abord, la cognition est-elle entièrement symbolique au sens où l'est une machine de Turing, comme tendent à le penser les pionniers de l'intelligence artificielle ? On peut tenter d'utiliser les images mentales pour répondre par la négative, en soutenant qu'elles sont des représentations fondamentalement « analogiques » et non « digitales »,

3. « Le cerveau humain encode, modifie et enregistre des informations qu'il reçoit *via* ses divers organes sensoriels ; il transforme ces informations par des processus qu'on appelle “penser” ; et produit diverses sorties motrices ou verbales sur la base des informations enregistrées. Rien jusqu'ici ne prête à controverse [...]. Ce qui est controversé, en revanche, c'est la question de savoir *comment* les informations sont enregistrées dans le cerveau – dans la terminologie habituelle, comment elles sont “représentées” – ou même tout simplement comment nous pouvons décrire des représentations, et ce que nous voulons dire quand nous disons que des informations sont représentées d'une certaine manière plutôt que d'une autre. » (Simon 1978, p. 3)

4. Pour se faire une idée du débat, voir par exemple, par ordre chronologique, Pylyshyn 1973, Kosslyn et Pomerantz 1977, Anderson 1978, Kosslyn 1980 et Pylyshyn 1981. Plus récemment, de nouvelles méthodes d'étude de l'activité cérébrale par IRMf ont fait rebondir la controverse ; voir Kosslyn, Ganis et Thompson 2001 et Pylyshyn 2002.

distinction souvent glosée en invoquant l'opposition entre continu et discret [5]. Grossièrement, l'idée est alors que les images mentales seraient plutôt comme des photographies classiques que comme des photographies numériques, et qu'à ce titre, si l'esprit est peut-être susceptible d'une analyse comme dispositif physico-chimique complexe qui traite des signaux voire forme des représentations, il n'est en tout cas pas comparable à un ordinateur au sens de Turing [6].

Ce n'est pas à ce niveau-là que se présente le problème pour Simon : une photographie numérique et une description en français de la scène photographiée sont toutes deux symboliques au sens de la théorie de la calculabilité, et encodables dans un ordinateur ; la différence est à chercher ailleurs. Cette remarque simple est pourtant souvent occultée par une confusion entre « symbolique » et « logique ». L'idée, centrale en intelligence artificielle, que les représentations mentales sont formées de symboles – et donc réalisables par ordinateur – est régulièrement présentée en disant que ce sont « des formules d'un langage interne ou "mentalais" proche des langages formels de la logique [7] », pour citer par exemple Daniel Andler. Au sens large, c'est sans doute exact, dans la mesure où les langages formels de la logique ont, historiquement, servi de modèle pour le développement de la théorie de la calculabilité ; mais en un sens plus restreint, c'est trompeur. Il existe bien des thèses philosophiques spécifiques, comme celle de Fodor (1975), qui postulent l'existence d'un authentique « langage de la pensée » très semblable à la logique du premier ordre. Il existe aussi des chercheurs en intelligence artificielle qui expriment leurs représentations dans un langage logique. En revanche, l'opinion de Simon est sans ambiguïté :

> An influential coterie of contemporary artificial intelligence researchers, including Nils Nilsson, John McCarthy, and others, believe that formal logic provides the appropriate language for A.I. programs, and that problem solving is a process of proving theorems. They are horriby wrong on both counts [...]. [8]

En fait, à en croire l'autobiographie de Simon, si la logique formelle est importante pour l'intelligence artificielle, c'est seulement parce qu'elle fournit le paradigme d'une mani-

5. Il n'est pas si facile de formuler cette distinction de manière claire. Pour des tentatives particulièrement intéressantes, voir Goodman 1968, en part. chap. 4, Lewis 1971 et Haugeland [1981] 1998a.

6. Pour une introduction rapide à l'enjeu, qui est la distinction entre « ordinateur analogique » et ordinateur au sens habituel, voir Haugeland 1981b, section V. Pour une discussion plus approfondie, voir Haugeland [1981] 1998a.

7. Andler 2004, p. 18.

8. « Il y a aujourd'hui une coterie influente de chercheurs en intelligence artificielle, dont Nils Nilsson et John McCarthy, qui croient que la logique formelle fournit le langage adapté pour les programmes d'I.A., et que la résolution de problèmes est un processus de démonstration de théorèmes. Ce sont deux erreurs calamiteuses [...] » (Simon 1991, p. 192, note).

pulation réglée de symboles, que l'ordinateur généralise ensuite [9]. Il écrit même que c'est devant un programme produisant des *images bidimensionnelles* (en l'occurence, de fausses images radar destinées à la simulation d'un poste de défense anti-aérienne) qu'il a pour la première fois compris, quoique confusément encore, que les ordinateurs pouvaient manipuler des symboles quelconques et pas seulement faire des calculs numériques [10]. Il est vrai que le premier travail en intelligence artificielle de Simon – avec son collaborateur Allen Newell – est un programme qui prouve des théorèmes de logique [11], programme qu'il a immédiatement décrit, non sans quelque exagération, comme une « machine qui pense [12] ». Plus généralement, dans les travaux ultérieurs de Newell et Simon, la recherche de démonstration d'un théorème sert parfois de paradigme de problème à résoudre. Cela explique sans doute bien des malentendus sur leur position. Mais en réalité, les représentations internes des systèmes successifs de Newell et Simon sont très différentes des propositions d'un langage logique ; ce sont des « structures associatives », des sortes de graphes [13]. Ce qui est sûr, c'est que si les expériences sur l'imagerie suggèrent l'existence de représentations mentales différentes de formules d'un langage de la pensée, elles ne remettent pas en cause pour Simon l'idée d'une cognition symbolique en général.

On peut aussi placer le débat à un second niveau, qui est celui qui occupe principalement Zenon Pylyshyn, sans doute le critique le plus célèbre de l'idée d'images mentales [14] : dans quelle mesure les représentations mentales sont-elles *abstraites*, c'est-à-dire en gros séparées des détails des *stimuli* initialement reçus par de nombreuses étapes de catégorisation ? Sur ce plan-là et quoiqu'il garde une nette ouverture d'esprit, Simon se placerait plutôt du côté de Pylyshyn [15].

Mais ce qui rend la position de Simon intéressante, c'est que pour lui le problème de l'imagerie n'est pas là non plus. Même si toutes les représentations mentales conservées en mémoire de travail étaient abstraites (ou, dans les termes de Simon 1972, *sémantiques*), ce que Simon trouve plausible mais non certain, la distinction entre formes de représentation, par exemple entre images et phrases, aurait encore un sens. Son travail de clarifica-

9. Cf. Simon 1991, p. 193.

10. Simon 1991, p. 201. Voir aussi les entretiens rassemblés par Pamela McCorduck 2004, chap. 6.

11. Ce programme, appelé « *Logic Theory Machine* », démontre des théorèmes des *Principia Mathematica* de Russell et Whitehead. Newell et Simon l'ont mis au point fin 1955 et publié en 1956 ; voir Newell et Simon 1956, [1957] 1963. MacKenzie 2001, chap. 3 fait un récit agréable et bien contextualisé de ces travaux. Pour plus de références, voir ci-dessous, note 16.

12. Simon 1991, p. 206.

13. Voir par exemple Newell et Simon 1972.

14. Voir *supra*, note 4.

15. Pylyshyn 1973 (en part. p. 12–15) discute d'ailleurs les représentations en forme de graphes utilisées entre autres par Simon, et conclut que les différences avec des représentations par formules d'un langage logique ne sont pas essentielles pour le contraste qui l'intéresse.

tion conceptuelle, qui vise à comprendre de manière générale « ce que nous voulons dire quand nous disons que des informations sont représentées d'une certaine manière plutôt que d'une autre », est en fait largement transversal au débat tel qu'il existe en psychologie.

4.2 Représentations et heuristiques

Simon arrive face au problème des formes de représentation, que je viens de décrire, avec l'expérience issue d'un parcours très original. Je voudrais d'abord en dire quelques mots, qui éclaireront la suite [16].

C'est en fait en sciences politiques, et plus précisément en théorie des organisations, que Simon a commencé sa carrière. Aujourd'hui encore, il reste avant tout cité pour son concept de « rationalité limitée » (« *bounded rationality* »), qui s'oppose aux modèles idéalisés d'un agent parfaitement rationnel et parfaitement informé qui dominaient et dominent largement encore l'économie. (Ces travaux-là lui vaudront le « Nobel d'économie » en 1978, mais aussi, paradoxalement, une relative mise au ban de la profession.) Il écrit ainsi en 1947 :

> It is impossible for a single, isolated individual to reach any high degree of rationality. The number of alternatives he must explore is so great, the information he would need to evaluate them so vast that even an approximation to objective rationality is hard to conceive. [17]

Comment s'orienter dans cet immense espace de possibilités ? Dans le contexte de ses premiers travaux, Simon en conclut que les organisations doivent contraindre l'environnement du choix, en limitant les possibilités des agents et en leur fournissant les informations pertinentes ; plus tard, il en conclura d'une manière générale que la rationalité humaine implique de trouver des stratégies pour s'orienter dans un espace de possibilités trop vaste, stratégies qui, à défaut d'être optimales, peuvent tout de même être efficaces [18].

---

16. Les sources sur la carrière de Simon en intelligence artificielle sont nombreuses, à commencer par son autobiographie (Simon 1991). Le livre d'entretiens rassemblés par Pamela McCorduck 2004 offre un contrepoint utile en ce qu'il donne la parole à de nombreux autres acteurs (voir aussi Crevier 1993). La meilleure introduction rapide et accessible à ses travaux des années 1950 et 1960 sur la démonstration automatique est sans doute le chapitre 3 de MacKenzie 2001. Pour une vue d'ensemble sur sa carrière (qui englobe de nombreuses disciplines mais témoigne d'une forte cohérence méthodologique), voir Crowther-Heyck 2005, et pour une réévaluation originale et très stimulante, voir Mirowski 2002, en part. p. 452–479, 529-533.

17. « Il est impossible pour un individu seul et isolé d'atteindre un niveau quelque peu élevé que ce soit de rationalité. Le nombre d'alternatives qu'il doit explorer est si grand, l'information dont il aurait besoin pour évaluer celles-ci est si vaste, que même une approximation de la rationalité objective est difficile à concevoir. » (Simon 1976, p. 79 ; déjà dans la première édition : Simon 1947, p. 79, cité par MacKenzie 2001, p. 63–64.)

18. Sur la genèse du concept de « rationalité limitée », qui n'apparaît qu'en germe dans *Administrative Behavior* (1947) malgré les efforts de Simon pour l'y introduire (dans ses nouvelles introductions et même en l'insérant dans l'index, cf. Takahashi 2015), voir Crowther-Heyck 2005 ainsi que Mirowski 2002, p. 456 *sq.*

Lorsqu'en collaboration avec Allen Newell, il se tourne vers l'intelligence artificielle, c'est-à-dire pour commencer vers la conception de programmes capables de jouer aux échecs et de démontrer des théorèmes de logique, il s'exprime en des termes étonnamment proches :

> In a fundamental sense, proving theorems and playing chess involve the same problem: reasoning with heuristics that select fruitful paths of exploration in a space of possibilities that grows exponentially. [19]

Au-delà du terme d'heuristique (qui viendrait ici de George Pólya, dont Allen Newell avait suivi les cours à Stanford [20]), l'idée est analogue : l'important est de trouver des solutions qui nous soient accessibles malgré l'extrême limitation de nos capacités. Mais surtout, Simon en tire l'objectif de *simuler* les heuristiques et plus généralement les processus humains de résolutions de problème. Sa conception de l'intelligence artificielle implique donc certes de construire des machines capables de tâches complexes, mais de le faire en respectant l'esprit des limitations qui pèsent sur la résolution humaine de problèmes et puisse donc imiter la manière humaine de procéder. Le travail accompli sert alors autant à éclairer la cognition humaine qu'à la reproduire artificiellement.

Or comme le montre bien Donald MacKenzie [21], cette conception de l'intelligence artificielle était loin d'être partagée par beaucoup de ses collègues. La meilleure manière de résoudre un problème informatiquement peut être très différente de celle qui convient à un être humain. Dans les termes d'Allen Newell,

> In Herbert [Simon]'s and my stuff [...] the concern for artificial intelligence [is always] right there with the concern for cognitive psychology. Large parts of the rest of the AI field believe that is exactly the wrong way to look at it: you ought to know whether you're being a psychologist or an engineer. Herb and I have always taken the view that maximal confusion between those is the way to make progress. [22]

C'est cette originalité de Simon et de ses collaborateurs qui nous ramène au problème

19. « Fondamentalement, prouver des théorèmes et jouer aux échecs, cela implique le même problème : celui de raisonner avec des heuristiques permettant de choisir des voies fructueuses d'exploration dans un espace de possibilités qui croît exponentiellement. » (Newell, Shaw et Simon [1958] 1963, p. 50).

20. Voir Simon 1991, p. 199.

21. MacKenzie 2001, chap. 3.

22. « Dans nos trucs, à Herbert [Simon] et à moi [...] le souci de l'intelligence artificielle [est toujours] juste à côté du souci pour la psychologie cognitive. Une grande part du reste du domaine de l'IA pense que c'est précisément la mauvaise manière de voir les choses ; il faut savoir si on est en train de parler en psychologue ou en ingénieur. Herb et moi avons toujours considéré que la meilleure manière de progresser passait par la confusion maximale entre les deux. » (Crevier 1993, p. 258 ; l'interview date du début des années 1990).

des représentations. Newell et lui réfléchissent en effet dès les années 1950 à la démonstration automatique de théorèmes en géométrie, ce qui les conduit au rôle heuristique que peuvent jouer les figures. J'ai déjà mentionné que leur première réussite en intelligence artificielle avait été la « *Logic Theory Machine* » mise au point fin 1955 [23]. Il s'avère en fait que leur projet initial était de mettre au point un logiciel capable de découvrir des preuves en géométrie et non en logique ; ils ne changent de domaine que dans un second temps, parce que, écrit Simon, « confrontés à la question de savoir comment traiter les figures, nous nous sommes rendu compte que nous avions encore à affronter des problèmes élémentaires de perception [24] ». Ils n'abandonnent pourtant le sujet que temporairement. Quelques mois plus tard, le problème est discuté lors d'une rencontre d'été qui rassemble, outre Newell et Simon, plusieurs autres pionniers de l'intelligence artificielle, dont John McCarthy et Marvin Minsky [25]. Ce dernier y propose une méthode heuristique, fondée sur les figures, pour la recherche de preuves en géométrie [26] ; celle-ci donnera lieu à un programme relativement efficace mis au point par Herbert Gelernter, qui emploie des figures représentées sous forme de tableaux de pixels [27]. Simon est donc familier de longue date avec l'idée que des représentations bien adaptées peuvent fournir des heuristiques efficaces et donc un avantage computationnel dans la recherche de solutions.

Notre point de départ, cependant, était un problème de psychologie, celui des représentations mentales. Mais une seconde conséquence de la « confusion maximale » évoquée par Newell est qu'à un certain niveau d'analyse, Simon n'hésite pas à assimiler des processus souvent considérés comme appartenant à des ordres de phénomènes distincts. D'un côté, on a les processus inconscients, rapides et automatiques de la perception et de la mémoire. De l'autre, on a les processus délibérés et (partiellement) conscients de la résolution de problèmes mathématiques. La stratégie de Simon consiste à considérer que, conçus comme des problèmes généraux de traitement de l'information, ces deux situations sont comparables. En fait, sa vision d'ensemble est plus vaste encore : il défend l'idée de « sciences de l'artificiel » traitant non seulement de la psychologie et de la résolution de problèmes, mais aussi par exemple du fonctionnement des organisations, et dont la simulation informatique est la méthode centrale [28].

---

23. Voir *supra*, note 11.
24. « Facing the question of how to do the diagrams, we saw that we still had to deal with some basic problems of perception. » (Simon 1991, p. 205).
25. C'est la « conférence de Dartmouth », qui est parfois considérée comme un moment fondateur de l'intelligence artificielle. Pour quelques témoignages, voir McCorduck 2004, chap. 5.
26. Cf. McCorduck 2004, p. 126 et Simon 1991, p. 210.
27. Cf. Gelernter [1959] 1963.
28. Cf. Simon 1996. Pour se faire une idée de l'ampleur du projet, voir citation *infra*, p. 127.

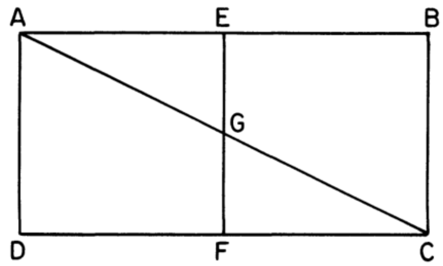


FIGURE 4.2 – Une figure géométrique qui, d'après Simon (1978, p. 6), permet de résoudre un problème élémentaire plus facilement qu'une description en phrases équivalente

C'est donc sans surprise que, pour clarifier les fondements conceptuels du débat sur l'imagerie, Simon (1978) se tourne vers la manière dont les figures peuvent faciliter la résolution de problèmes en géométrie. Voici l'un des exemples qu'il prend pour motiver son idée centrale. Soit ABCD un rectangle avec AB double de BC, soit E le milieu de AB et soit F le milieu de CD. Les droites EF et AC ont-elles un point d'intersection à l'intérieur de ABCD ? Pour Simon, la phrase énonçant les données du problème constitue une première *représentation* du problème, à partir de laquelle on pourrait, quoique difficilement, le résoudre (étant donnée une axiomatisation de la géométrie plane). Mais on pourrait aussi représenter les données du problème autrement : par des équations en coordonnées cartésiennes, ou plus simplement par la fig. 4.2. (Notons au passage qu'il ne se pose absolument pas la question de savoir s'il est *légitime* de conclure à l'existence d'une intersection sur la base de la figure.) Ce qui distingue ces différentes représentations n'est pas leur contenu (elles sont d'après lui « informationnellement équivalentes », comme je l'expliquerai), mais les opérations qu'elles facilitent.

4.3 Des représentations aux types de données

L'intuition fondamentale de Simon est donc la suivante : ce qui distingue avant tout entre elles les différentes manières de représenter les mêmes informations, c'est qu'elles rendent certaines *opérations* plus ou moins *faciles ou difficiles.*

Cette idée est naturelle si l'on réfléchit à l'expérience emblématique de Shepard et Metzler. Celle-ci nous renseigne en effet sur la difficulté de la tâche demandée, estimée *via* le temps de réponse des participants. Qu'est-ce qui nous permet d'en conclure quoi que ce soit sur l'existence d'images mentales, donc sur les formes de représentation qu'utiliserait le cerveau humain ? C'est précisément l'idée – certes vague – que seule une représentation « sous forme d'images » peut expliquer la distribution spécifique de niveaux de difficulté que révèle l'expérience (en l'occurence, le fait que la difficulté dépend de l'angle entre les deux objets présentés). Dans l'interprétation de cette expérience, la difficulté relative de certaines opérations sert donc, sinon de définition des images mentales, du moins de critère opérationnel pour les caractériser.

Pour clarifier et approfondir son idée, c'est encore une fois vers le modèle informatique que Simon se tourne : il assimile la notion de représentation à celle de stockage de données dans un ordinateur. Pour ce faire, il s'appuie sur la notion de « *type de données* » (« *data type* »), venue de l'informatique théorique. Pour simplifier, commençons par un exemple du même esprit que les siens mais qui ne requiert pas de familiarité avec la programmation.

Imaginons une université qui n'est pas informatisée, et qui conserve les dossiers complets de ses étudiants sur papier dans ses sous-sols. Le système de classement est très simple. Chaque étudiant reçoit un numéro de matricule à sa première inscription, et les dossiers sont simplement rangés sur des étagères par ordre de matricule. Lorsqu'un nouvel étudiant s'inscrit, on lui crée un nouveau dossier et un archiviste le range au bout de la dernière étagère utilisée. Ultérieurement, quand l'étudiant doit accomplir une formalité administrative, il donne son numéro, ce qui permet de remonter son dossier facilement. Tout fonctionne très bien dans notre université imaginaire, sauf dans un cas : quand un étudiant distrait perd son numéro, les archivistes sont obligés de parcourir les dossiers un par un jusqu'à trouver son dossier, une tâche très fastidieuse…

Un jour, une nouvelle équipe est élue à la tête de notre université. Pour régler le problème des étudiants qui oublient leur numéro (et aussi par opposition idéologique aux numéros d'étudiant, une invention technocratique déshumanisante), ils décident de changer de système. Désormais, les dossiers sont rangés par ordre alphabétique, et pour toute démarche, les étudiants doivent simplement indiquer leurs nom et prénom, et éventuellement leur date de naissance. Comme il y a peu d'étudiants assez négligeants pour oublier jusqu'à leur nom de famille, les archivistes n'ont plus guère besoin de parcourir tous les dossiers pour en retrouver un. L'inconvénient principal de ce nouveau système est que ranger un nouveau dossier est plus difficile : il ne suffit pas de l'ajouter au bout de la dernière étagère, mais il faut l'insérer au bon endroit selon l'ordre alphabétique, ce qui force régulièrement les archivistes à décaler beaucoup d'autres dossiers pour faire de la place.

Cette petite histoire permet d'illustrer de manière simple les deux idées centrales que Herbert Simon tire de l'informatique. Tout d'abord, il distingue deux manières pour des représentations (c'est-à-dire, si l'on suit son analogie, pour des manières de stocker des données) d'être « équivalentes ». En un premier sens, l'ancien et le nouveau système de rangement des dossiers étudiants sont équivalents parce que « la transformation de l'un à l'autre n'implique pas de perte d'information, c'est-à-dire que chacun peut être construit à partir de l'autre [29] ». C'est ce que Simon appelle l'équivalence *informationnelle*. Mais deux représentations (ou systèmes de stockage) peuvent aussi être équivalentes en un sens plus fort, si « les mêmes informations peuvent être extraites de chacune [...] avec à peu près la même quantité de calculs [30] » – Simon dit alors qu'elles sont *computationnellement* équivalentes. Ce n'est manifestement pas le cas dans notre exemple : l'ancien système (par numéro d'étudiant) et le nouveau (par ordre alphabétique) ne sont pas computationnellement équivalents parce que certaines tâches difficiles dans le premier système, par exemple trouver un dossier à partir du seul nom de l'étudiant, sont faciles dans le second et inversement. Certes, pour illustrer l'idée de Simon, je me repose ici sur une notion vague de quantité d'efforts requise. Mais si l'on suit Simon et que l'on se place strictement dans un cadre informatique, on peut rendre l'idée rigoureuse : la « quantité de calcul » peut alors se mesurer en nombre d'opérations élémentaires requises.

La seconde idée essentielle de Simon est qu'on peut décrire une représentation de manière *abstraite*, indépendamment de sa réalisation physique. Pour le comprendre, revenons encore une fois à notre université imaginaire. Les services administratifs, dans les étages, n'ont pas besoin de savoir en détail comment les dossiers sont rangés dans les sous-sols. Pour faire leur travail, une connaissance très approximative leur suffit : ils ont seulement besoin de savoir que les dossiers sont accessibles par numéro d'étudiant (dans le premier système), ou par nom (dans le second système). Ce qui compte pour eux, c'est de savoir quelles requêtes seront facilement satisfaites par les archivistes. En fait, on peut très bien imaginer que le service d'archivage, découragé par l'idée de réorganiser leurs dizaines de milliers de dossiers à cause d'une lubie de la présidence, choisisse une solution de contournement. Plutôt que de déplacer leurs dossiers, ils peuvent continuer à utiliser les numéros d'étudiants en interne, mais créer un nouveau petit fichier des étudiants rangé par ordre alphabétique (par exemple un genre de catalogue de bibliothèque à l'ancienne, fait de fiches cartonnées), indiquant pour chaque étudiant son numéro et permettant donc de retrouver son dossier. Les services administratifs n'ont pas besoin de savoir si, quand ils demandent le dossier de Sophie Michon, l'archiviste va directement le chercher sur les étagères ou s'il consulte d'abord la petite fiche de Sophie ; la seule chose importante pour eux est que le dossier soit sur leur bureau quelques minutes plus tard, ce qui n'était pas possible dans l'ancien système. Dans ce scénario, on peut dire que les archivistes réalisent ou émulent

29. « Two representations are informationally equivalent if the transformation from one to the other entails no loss of information, i.e., if each can be constructed from the other. » (Simon 1978, p. 4).

30. « Two representations are computationally equivalent if the same information can be extracted from each (the same inferences drawn) with about the same amount of computation. » (Simon 1978, p. 5).

un système alphabétique sur la base physique d'un système classé par numéro, de telle manière que les services administratifs et la présidence de l'université n'y voient que du feu.

En d'autres termes, ce qui définit un système de stockage de données (ou, si l'on suit Simon, une représentation), ce n'est pas les détails de sa mise en œuvre mais une caractérisation plus abstraite, fondée sur les requêtes auxquelles elle permet de répondre facilement. C'est sur cette base que Simon peut écrire que « les ordinateurs modernes ont libéré notre réflexion sur les représentations [31] ».

Nous pouvons, finalement, résumer le projet de Simon (1978). Son idée centrale consiste à caractériser les représentations par leurs propriétés computationnelles. Sur cette base, il cherche à opérationnaliser la notion de forme de représentation et donc à établir le cadre d'un programme de recherche sur les représentations mentales ; ce programme consiste à étudier systématiquement quelles opérations sont faciles ou difficiles pour les sujets expérimentaux et à comparer ces résultats à ceux que produiraient des modèles informatiques bien choisis. Simon tente aussi, dans cet article, un premier essai dans cette direction. Il imagine une structure de données qui pourrait expliquer nombre des résultats expérimentaux rattachés à l'imagerie mentale : computationnellement, cette structure serait caractérisée par le fait qu'à partir de chaque élément, elle donnerait un accès rapide aux éléments spatialement voisins. Mais ce n'est qu'une esquisse, qu'il reprend une dizaine d'années plus tard. Avant d'y venir, j'aimerais toutefois approfondir les notions informatiques qu'invoque Simon, qui sont beaucoup moins claires qu'il ne semble au premier abord.

4.4 Mais que sont les types de données de Simon ?

Simon précise son idée d'une caractérisation abstraite des représentations en invoquant la notion informatique de type de données (« *data type* »). Il faut y regarder de plus près, parce que s'il y a bien aujourd'hui une notion usuelle qui porte ce nom, Simon lui donne un sens assez différent, qui s'avère lourd de difficultés.

a) Une variante de la notion usuelle ?

Les histoires de l'informatique font souvent remonter l'idée de type de données (ou type de données abstrait, « *abstract data type* ») à des travaux du début des années 1970, en particulier de Tony Hoare et de David Parnas, un collègue de Simon à l'université Car-

31. « Modern computers have freed our thinking about representations. » (Simon 1978, p. 9).

negie Mellon [32]. Ces travaux ont une motivation commune : la nécessité de maintenir sous contrôle la complexité grandissante des logiciels ; pour ce faire, l'idée est de décomposer des logiciels complexes en composants plus simples dont on peut spécifier le comportement indépendamment les uns des autres, l'exemple paradigmatique d'une telle méthodologie étant la « programmation structurée » défendue à l'époque par Edsjer Dijkstra [33]. (Simon lui-même met fréquemment en avant le principe d'une organisation des systèmes complexes en sous-systèmes largement indépendants les uns des autres et insérés dans une structure hiérarchique [34].) Dans le cas des données, l'important est alors d'isoler du reste du programme les décisions concernant le détail du stockage, et donc codifier explicitement, par avance, les manières dont le reste du programme peut interagir avec les données. Pour reprendre mon exemple universitaire, l'important est de séparer les services administratifs du service d'archivage des dossiers, ce dernier pouvant se débrouiller comme il veut pour organiser ses rayonnages tant qu'il est capable de répondre aux requêtes qui lui sont faites, requêtes qui, pour cette raison, doivent être codifiées au préalable. La première définition explicite d'un type de données se trouve vraisemblablement chez Barbara Liskov et Stephen Zilles en 1974 :

> An *abstract data type* defines a class of abstract objects which is completely characterized by the operations available on those objects. This means that an abstract data type can be defined by defining the characterizing operations for that type. [35]

Les « opérations disponibles » sur un type de données forment ce qu'on appelle d'ordinaire son *interface*.

Cependant, pour Simon, codifier les opérations permises ne suffit pas : il est crucial de savoir quelles opérations sont *rapides*. Dans l'exemple de l'université, on pourrait définir le service d'archivage des dossiers par les opérations d'ajouter un dossier, d'obtenir un dossier par matricule, et d'obtenir un dossier par nom. Or les deux systèmes (numérique et alphabétique) répondent à cette spécification. Ce qui les distingue, c'est la rapidité relative des différentes requêtes, comme le comprend bien Simon :

32. En particulier Parnas 1972a,b et Hoare 1972. Voir par exemple Kutzler et Lichtenberger 1983, p. 1, et pour des travaux historiques, Priestley 2011, chap. 10 et Tucker 2018.

33. MacKenzie 2001, chap. 2 raconte très bien l'atmosphère qui prévalait et replace dans un contexte large les différents programmes de recherche qui allaient dans cette direction. Pour une histoire techniquement plus riche, voir Priestley 2011, chap. 9–11. Parnas 2002 explique rétrospectivement ses motivations de manière très claire.

34. Voir par exemple son article Simon 1962.

35. « Un *type de données abstrait* définit une classe d'objets abstraits qui est complètement caractérisée par les opérations disponibles sur ces objets. Cela veut dire qu'un type de données abstrait peut être défini en définissant les opérations caractéristiques de ce type. » (Liskov et Zilles 1974, p. 51).

> Defining a representation means (1) specifying one or more data types, and (2) specifying the primitive (i.e., "fast") operations that can be performed on information in those data types. [36]

À première vue, l'idée de Simon semble donc être de *compléter* la notion usuelle de type de données en spécifiant quelles opérations sont « rapides ». Pour rendre cette idée plus précise, on peut demander que chaque opération permise (par exemple obtenir un dossier par matricule) soit accompagnée d'une spécification explicite de sa vitesse, ou en termes techniques, de sa *complexité*[37]. En informatique même, l'idée que les spécifications des opérations permises devraient en fixer la complexité est maintenant répandue, mais à ma connaissance n'a pas cours dans les années 1970. Ainsi, Meng Lee et Alexander Stepanov – qui ont joué un rôle central dans le domaine en développant une « bibliothèque » standard de fonctions et structures de données pour le langage C++ – peuvent écrire en 1994 :

> It has been commonly assumed that the (time and space) complexity of an operation is part of its implementation and should not be specified at the interface level. This assumption is incorrect since it invalidates the main reason for the separation of interfaces and implementations, namely, ability to substitute one module for another with the conforming interface. [38]

b) « Structures » plutôt que « types » de données

L'idée de Simon est-elle donc de définir une représentation comme un type de données (au sens usuel) pour lequel on aurait spécifié quelles opérations sont rapides ? En fait, non.

---

36. « Définir une représentation, cela veut dire (1) spécifier un ou plusieurs types de données, et (2) spécifier les opérations primitives (c'est-à-dire "rapides") qui peuvent être exécutées sur les informations dans ces types de données. » Simon 1978, p. 7-8.

37. Pour comprendre la notion technique de complexité, il faut voir qu'en informatique, la question importante est d'habitude de savoir comment la vitesse d'une opération *dépend de la taille des données.* C'est à nouveau la même chose dans notre exemple universitaire. Pour une université qui n'aurait que trois étudiants, n'importe quel système de stockage ferait l'affaire ; c'est seulement quand le nombre d'étudiants augmente que les problèmes commencent, et qu'il devient important de distinguer entre les différents systèmes possibles. C'est à faire ce genre de distinctions que sert la notion de complexité. Avec un classement par matricule, par exemple, retrouver un dossier par numéro d'étudiant est aussi facile s'il y a trois étudiants ou trente mille : cette opération a une complexité en temps « constante » (à vrai dire, ce n'est pas tout à fait exact : plus il y a d'étudiants, plus il faudra marcher longtemps entre les rayonnages, mais laissons cela de côté pour simplifier !). Retrouver un dossier par nom, en revanche, requiert de parcourir tous les dossiers, une tâche dont la durée est en gros proportionnelle au nombre d'inscrits : l'opération a une complexité en temps « linéaire ».

38. « On est en général parti du principe que la complexité (en temps et en espace) d'une opération fait partie de son implémentation et n'a pas à être spécifiée au niveau de l'interface. Ce postulat est incorrect parce qu'il entre en contradiction avec la raison principale de séparer les interfaces et les implémentations, à savoir la possibilité de remplacer un module par un autre ayant une interface conforme. » (Lee et Stepanov 1994, p. 26). Je remercie Gabriel Scherer d'avoir attiré mon attention sur ce texte.

À y regarder de plus près, Simon semble entendre « type de données » en un sens moins abstrait.

> A *data type* is some particular way of organizing information in memory. For example, among the data types that are commonly used in computing are lists and arrays. [...] The declaration that information will be represented in lists or in arrays does not say anything about the physical location of the information in memory. [39]

Dans les termes de notre exemple universitaire, un type de données ne concerne donc pas uniquement les requêtes que l'on peut adresser au service d'archivage, mais aussi l'organisation interne de ce service. Cela dit, prise au pied de la lettre, la définition de Simon semble contradictoire : qu'est-ce que cela peut vouloir dire exactement, « une manière particulière d'organiser des informations en mémoire » qui néanmoins « ne dit rien sur la localisation physique des informations » ?

Pour comprendre l'intuition de Simon, il faut revenir à la pratique de la programmation, d'où viennent les « listes » et « tableaux » (« *arrays* ») de la citation précédente. À cette fin, examinons le premier exemple que traite Donald Knuth dans son traité classique de programmation, auquel Simon renvoie [40].

Imaginons que nous ayons à travailler avec une liste ordonnée d'éléments (par exemple des dossiers d'étudiants, des messages électroniques, ou n'importe quoi d'autre). On peut vouloir accomplir différentes opérations sur cette liste : consulter le premier élément, le dernier élément, ou en toute généralité l'élément numéro k ; insérer ou supprimer un élément au début, à la fin, ou plus généralement à une position k ; etc. Il y a de nombreuses manières d'organiser notre liste en mémoire, mais aucune qui rende toutes ces opérations également faciles. Toutefois, beaucoup d'applications informatiques peuvent se contenter de faire un usage intensif de certaines opérations seulement, que l'on peut alors optimiser aux dépens des autres.

Ainsi, on peut avoir besoin d'une liste pour mettre en œuvre une file d'attente, par exemple une file de messages électroniques à envoyer ; dans ce cas, deux opérations suffisent : ajouter un message à la fin (à l'arrière de la file) et récupérer le message qui se trouve au début (à l'avant de la file) pour le transmettre. De quoi a-t-on besoin pour les accomplir efficacement ? Tout d'abord, il faut pouvoir trouver facilement le premier élément de la file

39. « Un *type de données* est une manière particulière d'organiser des informations en mémoire. Il y a par exemple, parmi les types de données communément utilisées en informatique, les listes et les tableaux. [...] La déclaration que des informations seront représentées par des listes ou par des tableaux ne dit rien sur la localisation physique des informations en mémoire. » (Simon 1978, p. 8).

40. Voir Knuth [1968] 1997, chap. 2. L'exemple que je prends est expliqué plus en détail à la section 2.2.1.

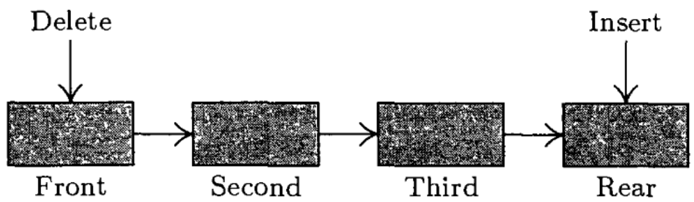


FIGURE 4.3 – Représentation d'une structure de file (« *queue* »), d'après Knuth ([1968] 1997, p. 241)

(le prochain message à transmettre) ainsi que le dernier élément (après lequel on devra insérer de nouveaux messages). Il faut ensuite pouvoir, à partir de chaque élément de la file, retrouver facilement le suivant, de telle manière qu'après chaque traitement d'un message on sache quel sera l'élément suivant à expédier et supprimer. En d'autres termes, les seules *relations* entre éléments qui doivent être faciles d'accès sont celles qui apparaissent sur la figure 4.3.

Ce genre de schéma, s'il spécifie ce que Knuth appelle les « relations structurelles [41] » pertinentes pour notre programme, ne spécifie en revanche pas la manière exacte dont les éléments doivent être disposés en mémoire : ils pourraient être disposés consécutivement ou loin les uns des autres, dans l'ordre ou dans le désordre. Pour rendre plus concrète cette multiplicité de possibilités, examinons rapidement une manière classique de faire qui permet de disperser les éléments de notre liste à travers toute la mémoire de l'ordinateur. L'idée est simple : on adjoint à chaque élément, c'est-à-dire par exemple à chaque mail à envoyer, une donnée supplémentaire qui indique l'*adresse* en mémoire de l'élément suivant ; on garde également trace des adresses du premier et du dernier élément de notre file de mails. Cela permet, au fur et à mesure de l'arrivée de nouveaux mails à envoyer, de les stocker à n'importe quel emplacement libre de la mémoire ; cela permet aussi, dès qu'on a transmis un mail de notre file, de libérer directement la mémoire qu'il occupait [42]. Or, que l'on fasse ainsi ou que l'on préfère s'en tenir à un stockage consécutif des mails, on peut considérer que l'on a mis en œuvre une file : tout ce qui compte est que l'on ait un moyen rapide de trouver les deux extrémités et, pour chaque élément, de trouver le suivant, selon le schéma de la fig. 4.3. C'est en ce sens que Simon peut parler d'une organisation en mémoire qui ne fixe pas « la localisation physique des informations » ; il est très familier de

41. Knuth parle de « relations structurelles » (« *structural relationships* ») au début de son chapitre (Knuth [1968] 1997, p. 232) ; il parle aussi régulièrement d'« information structurelle » (voir *infra*, note 48).

42. C'est ce que l'on appelle ordinairement une « liste chaînée » ; pour plus de détails, voir Knuth [1968] 1997, §2.2.1.

ce genre de considérations depuis les années 1950 [43].

Les « files » (« *queues* ») que je viens de décrire sont un exemple de ce que Knuth appelle *structure d'information* (« *information structure* »), et qu'il est aujourd'hui usuel d'appeler *structure de données* (« *data structure* »). Cette notion est moins abstraite que les types de données, qui spécifient seulement les opérations permises sans rien dire de leur mise en œuvre ; d'un autre côté, comme nous venons de le voir, cela reste une abstraction vis-à-vis des détails du programme, qui seraient plus précis sur l'organisation physique des données en mémoire. C'est bien cette notion, je crois, que Simon vise ; il faut tenir compte du fait que dans les années 1970, la terminologie n'est pas encore bien fixée et qu'il n'est pas lui-même informaticien. C'est en tout cas bien en termes de *structures* de données qu'il s'exprime dans son article suivant sur le sujet (Larkin et Simon 1987).

c) Un niveau d'abstraction problématique

Il me semble pourtant que l'hésitation de Simon entre « types de données » (comme il l'écrit en 1978) et « structures de données » (comme l'écrivent Larkin et Simon 1987) trahit une difficulté qui n'est pas seulement terminologique. Son objectif d'ensemble est clair : il lui faut une notion de représentation qui soit suffisamment abstraite pour s'appliquer indifféremment à sa réalisation physique, que ce soit « dans une tête humaine ou dans un ordinateur [44] ». Le problème, c'est que la discussion précédente nous fournit non pas une, mais *deux* approches possibles : *via* une interface d'opérations rapides ou *via* un modèle abstrait d'organisation des données. Quel est le lien entre ces deux approches ?

Dans son article de 1978, Simon semble hésiter. Dans un passage déjà cité, il commence par écrire :

> Defining a representation means (1) specifying one or more data types, and (2) specifying the primitive (i.e., "fast") operations that can be performed on information in those data types. [...] A *data type* is some particular way of organizing information in memory. [45]

43. Dès les années 1950, Allen Newell, Cliff Shaw et lui avaient en effet mis au point, pour leur *Logic Theory Machine*, le premier langage de programmation fournissant des opérations primitives de manipulation de données éloignées de l'organisation matérielle de la mémoire des ordinateurs sur lesquels ils travaillaient. Les langages de manipulation de listes qu'ils ont développé à cette époque ont d'ailleurs eu un impact important sur l'histoire ultérieure des langages de programmation, en particulier sur le développement du langage LISP et plus généralement sur la naissance de ce que l'on appelle aujourd'hui les langages de programmation « fonctionnels » ; sur ce point, voir Priestley 2017.

44. « Since we are concerned with representation at the symbolic or information-processing level, it will matter little whether the memory we are talking about resides in a human head or in a computer. » (Simon 1978, p. 4).

45. « Définir une représentation, cela veut dire (1) spécifier un ou plusieurs types de données, et (2) spécifier

Cela suggère clairement qu'organisation en mémoire et opérations rapides sont deux ingrédients distincts. Mais quelques lignes plus bas, il poursuit :

> [T]he declaration of a data type means that certain *primitive operations* are available for accessing, storing, and modifying data. A list representation requires a primitive operation of *find.next*, which, given a particular symbol as its argument, will access the succeeding symbol in the list to which the first belongs. [46]

Ici, Simon semble au contraire suggérer qu'une certaine organisation en mémoire (il parle de liste, mais on pourrait aussi bien penser aux files que nous avons discutées dans la partie précédente) est comme naturellement équipée de certaines opérations primitives. Reprenons le schéma que Knuth utilise pour illustrer la structure de file (fig. 4.3 p. 116). En suivant la terminologie de Knuth, j'ai dit que les flèches indiquaient les « relations structurelles » entre éléments qui sont pertinentes pour un programme utilisant une file. Les remarques de Simon que je viens de citer suggèrent de les interpréter autrement, à savoir comme indiquant les opérations primitives permises. Peut-être est-ce là l'idée de Simon pour unifier les deux descriptions abstraites de ce qu'est une représentation.

Il n'est pas si facile, cependant, de résoudre la tension entre opérations rapides et organisation en mémoire. La notion de type de données, telle qu'elle est comprise aujourd'hui en informatique, se laisse formaliser de manière précise [47], et peut être entièrement séparée de toute considération d'implémentation dans une architecture computationnelle particulière. En revanche, la notion de structure de données que semble invoquer Simon est beaucoup moins claire. Elle vient de la pratique de la programmation, et (à ma connaissance) n'a pas de définition pleinement satisfaisante. On peut être tenté de dire que choisir une structure pour certaines données revient à choisir quelles relations structurelles représenter « directement » (selon le terme qu'emploie Knuth [48]) ou encore « explicitement [49] »,

---

les opérations primitives (c'est-à-dire "rapides") qui peuvent être exécutées sur les informations dans ces types de données. […] Un *type de données* est une manière particulière d'organiser des informations en mémoire. » (Simon 1978, p. 7-8).

46. « La déclaration d'un type de données signifie que certaines *opérations primitives* sont disponibles pour accéder aux données, pour les enregistrer et pour les modifier. Une représentation en liste requiert une opération primitive *trouver.suivant* qui, à partir d'un symbole particulier donné en argument, accède au symbole suivant dans la liste auquel appartient le premier. » (Simon 1978, p. 8 ; l'auteur souligne.)

47. Voir par exemple Dale et Walker 1996.

48. Cf. par exemple Knuth [1968] 1997, p. 238 (je souligne) : « Data usually has much more structural information than we actually want to represent *directly* in a computer […]. We must decide in each case how much structure to represent […]. » (« En général, les données comportent beaucoup plus d'informations structurelles que nous voulons n'en représenter *directement* dans un ordinateur […]. Dans chaque cas, nous devons décider combien de structure représenter […]. »)

49. Larkin et Simon 1987 s'expriment parfois ainsi (par ex. p. 67).

mais ce sont des expressions peu claires [50] ; à moins de se rabattre sur la notion de type de données, en considérant (comme je l'ai suggéré plus haut) que parler de relations structurelles représentées directement n'est qu'une périphrase pour stipuler les primitives de l'interface, on a une vraie difficulté. En fait, la manière dont on comprend et présente habituellement les structures de données présuppose certaines caractéristiques spécifiques aux ordinateurs usuels, par exemple l'existence d'adresses en mémoire pour accéder aux données ; Knuth, par exemple, se place exclusivement dans ce cadre. Or, si l'on travaille avec des ordinateurs conçus différemment sur le plan matériel, par exemple équipés de mémoire accessible par le contenu (« *content-addressable memory* ») plutôt que par adresse, les liens entre organisation des données et opérations rapides seront complètement différents. Mais précisément, Simon a besoin de pouvoir traiter simultanément l'IBM 7090 et le cerveau humain ! Pour cela, le niveau d'abstraction intermédiaire auquel correspond la notion usuelle de structure de données (plus précise qu'une simple stipulation d'interface, moins précise que les détails de l'implémentation) ne semble pas satisfaisante.

Cette discussion, qui peut paraître oiseuse, a un enjeu crucial. Le point de départ de Simon était celui des représentations mentales : quelles formes ont-elles ? Y a-t-il des images mentales ? Face à ce problème, son but était clair : opérationnaliser le concept de représentation en termes de l'efficacité computationnelle de certaines opérations, ce qui permet d'étudier la question des formes de représentations mentales au moyen de modèles informatiques bien choisis. Or s'il faut séparer nettement organisation des données et opérations rapides sur ces données, parce que le lien entre les deux dépend profondément des spécificités matérielles de l'ordinateur ou du cerveau dont on traite, ce projet-là s'effondre. Cela ne signifie pas que les idées de Simon sont sans valeur, mais cela veut dire qu'elles ne suffisent pas à résoudre complètement un problème comme celui des images mentales : la question des formes de représentations mentales ne peut pas se réduire sans reste à une affaire d'efficacité computationnelle. À mon avis, les hésitations de l'article de Simon pointent donc un problème essentiel, qu'il n'a pas résolu.

4.5 Retour aux représentations externes

Simon revient à ces questions plus en détail dans les années 1980, en collaboration avec la psychologue Jill Larkin (Larkin et Simon 1987). (Dans la même veine, il poursuivra au cours des années 1990 une activité de recherche intense sur la question du raisonnement

50. Dans un article fouillé, Kirsh 1990 montre bien que dans un cadre informatique, la notion de représentation « explicite » n'est pas du tout claire.

avec de multiples formes de représentation [51].) Son travail avec Larkin traite cette fois uniquement de ce qu'il appelle des représentations « externes » et est apparemment prudent sur la généralisation aux représentations mentales, mais il faut bien voir que cela reste pour Simon fondamentalement le même problème ; d'ailleurs, il écrira deux ans plus tard :

> In order to deal with the difficulties one by one, [Jill Larkin and I] fudged a bit, alleging that we were talking about diagrams on paper rather than mental pictures; but most of our argument carries over in a straightforward way. [52]

Pourtant, il y a une inflexion de problématique. Larkin et Simon (1987) évitent toute discussion théorique d'une notion abstraite de représentation : ils ne cherchent plus explicitement à *caractériser* ou à *définir* les représentations d'après les opérations qu'elles rendent faciles, ce qui était l'aspect le plus ambitieux de l'article de 1978 mais aussi, comme l'a montré la discussion précédente, le plus problématique. Ils se contentent d'étudier les avantages ou les inconvénients d'utiliser telle représentation plutôt que telle autre, comme l'indique leur titre : « Why a Diagram Is (Sometimes) Worth Ten Thousand Words [53] ». Leur démarche est de proposer, pour des problèmes simples, des modèles informatiques de deux procédures différentes de résolution, l'une étant censée correspondre à l'usage d'une figure et l'autre à une résolution utilisant uniquement des phrases. Les différentes représentations utilisées sont cette fois modélisées au moyen de « structures de données » (expression qui remplace ici celle de « types de données » de 1978) et d'opérations qui agissent sur elles, ainsi que d'un mécanisme attentionnel qui explicite, à chaque étape, quel élément de la structure de données est examiné :

> A representation consists of both data structures and programs operating on them to make new inferences. [. . .] Since data structures for a problem are complex, we must also provide for an attention management system that determines what portion of the data structure is currently attended to and can trigger the productions [54] of the program. The nature of attention management depends crucially on the linkages provided in the data structure since

51. Les articles Qin et Simon 1995 et Tabachneck-Schijf, Leonardo et Simon 1997 sont représentatifs de ses deux projets de recherche principaux sur le sujet.

52. « Dans l'idée de traiter les difficultés une par une, [Jill Larkin et moi] avons été un peu évasifs et avons prétendu que nous parlions de figures sur papier plutôt que d'images mentales ; mais l'essentiel de notre argument se transpose sans problèmes. » (Simon 1989, p. 383).

53. « Pourquoi une figure vaut (parfois) dix mille mots. »

54. Ce que Simon appelle les « productions » du programme sont des sortes de règles d'inférence qui peuvent être appliquées si certaines prémisses sont disponibles. À chaque étape de l'exécution du programme, seules les données sur lesquelles porte le mécanisme d'attention peuvent servir de prémisses à l'application éventuelle d'une telle règle.

> this is the only information available for guiding shifts in attention. […] In general the computational efficiency of a representation depends on all three of these factors (data structure, program, attention management) and on how well they work together. [55]

L'hésitation que j'ai notée dans l'article de 1978, sur la question de savoir s'il y a un lien intrinsèque entre une structure de données et les opérations rapides qu'elle permet, est ici éliminée : Larkin et Simon séparent clairement données et opérations qui agissent sur elles, même s'ils admettent bien sûr que l'organisation des données impose des contraintes sur les opérations efficaces.

Dans ce cadre, ils expliquent alors l'efficacité supérieure (dans certains contextes) de représentations « diagrammatiques » de deux manières principales, qu'ils illustrent sur deux problèmes élémentaires : un problème de statique et un problème de géométrie. Examinons-les tour à tour.

a) Un problème de statique

Leur première idée reprend dans les grandes lignes une suggestion que Simon avait déjà ébauchée en 1978 [56] ; ils l'illustrent sur un problème de statique à base de poulies et de poids. Ce problème est posé en phrases, du genre de « Soient trois poulies *A*, *B* et *C*, deux poids et des cordes telles que le premier poids est suspendu à l'extrémité gauche d'une corde qui passe sur la poulie *A* et dont l'autre extrémité est attachée à l'autre poids, etc. ». Larkin et Simon commencent par formaliser ces phrases pour obtenir une première structure de données (fig. 4.4). Pour résoudre le problème, ils fournissent et formalisent ensuite un certain nombre de règles, par exemple que les deux côtés d'une corde attachée à une poulie ont la même tension.

Une méthode naïve de résolution consiste alors à tenter d'appliquer ces règles pour obtenir des propositions supplémentaires, en parcourant phrase par phrase, dans l'ordre, les éléments de la structure de données de la fig. 4.4, puis de recommencer. Cette méthode

55. « Une représentation consiste à la fois en des structures des données et en des programmes qui opèrent sur celles-ci pour faire de nouvelles inférences. […] Puisque les structures de données pour un problème sont complexes, nous devons aussi fournir un système de gestion de l'attention qui détermine quelle portion de la structure de données est examinée à un moment donné et peut déclencher les productions du programme. La nature de la gestion de l'attention dépend crucialement des liens fournis dans la structure de données, parce que ce sont les seules informations disponibles pour guider des déplacements de l'attention. […] En général, l'efficacité computationnelle d'une représentation dépend à la fois de ces trois facteurs (structure de données, programme, gestion de l'attention) et de la manière dont ils travaillent ensemble » (Larkin et Simon 1987, p. 67-68).

56. Cf. *supra*, p. 112.

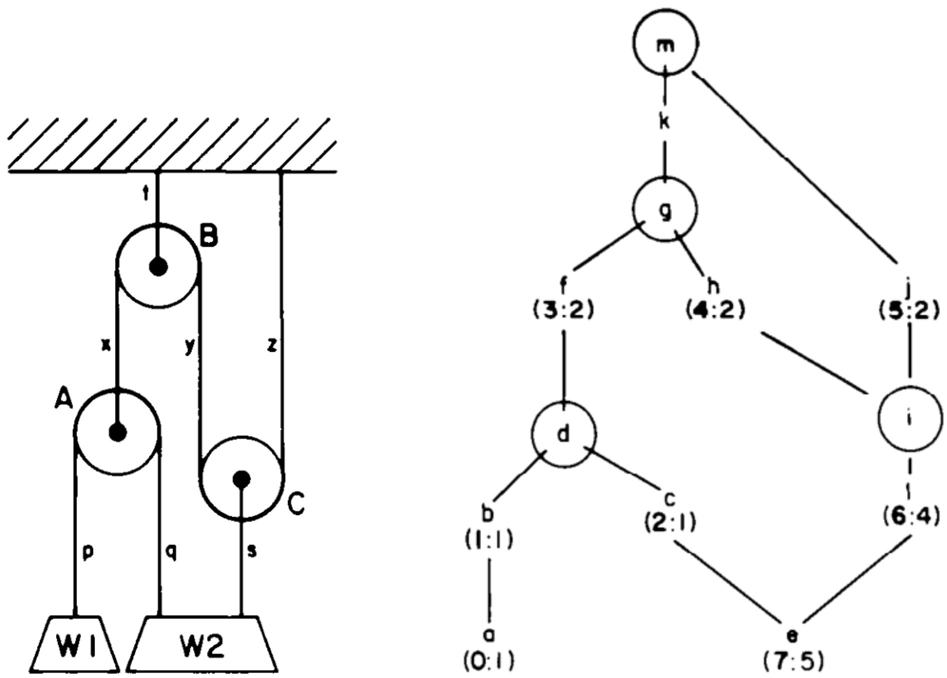
```
          (Weight W1) (Rope Rp) (Rope Rq) (Pulley Pa)
(1a.1)    (hangs W1 from Rp)
(1a.2)    (pulley-system Rp Pa Rq)

          (Weight W2)
(1b.1)    (hangs W2 from Rq)

          (Rope Rx) (Pulley Pb) (Rope Ry) (Pulley Pc) (Rope Rz)
          (Rope Rt) (Rope Rs) (Ceiling c)
(2a.1)    (hangs Pa from Rx)
(2a.2)    (pulley-system Rx Pb Ry)
(2a.3)    (pulley-system Ry Pc Rz)
(2b.1)    (hangs Pb from Rt)
(2b.2)    (hangs Rt from c)

(3a.1)    (hangs Rx from c)
(3a.2)    (hangs Rs from Pc)
(3b.3)    (hangs W2 from Rs)

(4.1)     (value W1 1)
```

FIGURE 4.4 – Structure de données correspondant au problème de statique qu'étudient Larkin et Simon (1987, p. 74)

est coûteuse, parce qu'à chaque étape, on risque de devoir parcourir une bonne partie des données dont on dispose avant d'en trouver une qui permette d'appliquer une de nos règles et donc de progresser.

D'après Larkin et Simon, si une figure (cf. fig. 4.5(a)) nous facilite le travail, c'est parce qu'elle permet de limiter cette recherche. Plus précisément, on peut résoudre le problème de proche en proche en parcourant la figure élément par élément. Si on connaît la valeur du poids $W1$, on peut en déduire la tension de la corde p, puis celle de la corde q, etc. Ainsi, on n'a pas besoin de parcourir tout ce que l'on sait à chaque étape.

Pour modéliser cette méthode humaine de résolution, Larkin et Simon suggèrent une structure de données dans laquelle les données, au lieu d'être organisées séquentiellement dans un ordre arbitraire, sont organisées par localisation. Cela permet un mécanisme attentionnel qui procède de proche en proche. Intuitivement, leur idée est la suivante. Il faut imaginer que tout ce que l'on sait sur la poulie A est rassemblé dans un même emplacement en mémoire, et que cet emplacement est muni de liens vers trois autres emplacements en mémoire : ceux qui contiennent tout ce que l'on sait sur, respectivement, les cordes p, q et x. Le mécanisme attentionnel peut alors suivre les liens d'un emplacement mémoire à un autre qui y est relié, et ainsi procéder de proche en proche. Pour clarifier l'organisation de cette structure de données, ils la représentent à la fig. 4.5(b). (Il faut prendre garde au fait que cette figure n'est qu'une manière de clarifier les liens de la structure de données sous-jacente, qui n'est pas elle-même une figure au sens usuel du terme !) L'avantage est



(a) (b)

FIGURE 4.5 – Figures (a) d'un problème de statique, (b) de la structure de données que Larkin et Simon utilisent pour modéliser une résolution de ce problème s'appuyant sur la figure précédente (Larkin et Simon 1987, p. 73, 79)

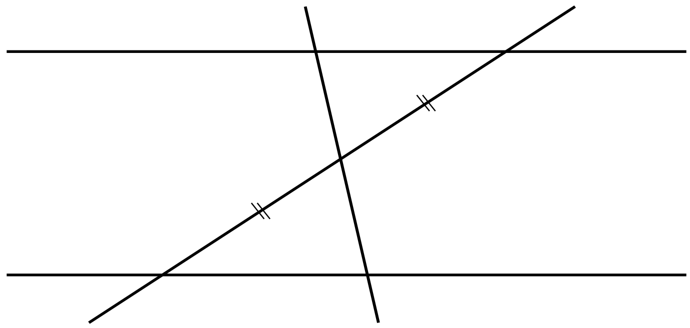


FIGURE 4.6 – Une figure pour le problème de géométrie qu'étudient Larkin et Simon (1987, section 3)

que cette structure de données permet de localiser efficacement les composants du système dont on a besoin à chaque étape de résolution, et donc de raccourcir la recherche, à chaque étape, des données qui peuvent être utiles.

b) Un problème de géométrie

Leur seconde idée, très différente, se fonde sur l'exemple d'un problème de géométrie élémentaire, que l'on peut paraphraser ainsi (voir aussi fig. 4.6) :

> Soient deux droites sécantes qui coupent deux droites parallèles et se rencontrent en un point x situé entre ces deux droites parallèles ; montrer que si l'une de ces droites sécantes bisecte le segment de l'autre qui se trouve entre les deux parallèles, alors les deux triangles formés par les sécantes sont semblables.

Il faut noter d'emblée que ce problème peut se résoudre sans faire appel ni à une figure ni à des axiomes sophistiqués : quelques propositions élémentaires de géométrie suffisent [57]. Ce n'était pas le cas de l'exemple que Simon avait utilisé pour motiver son article de 1978, dans lequel la figure permettait d'observer l'existence d'un point d'intersection [58]. Ici, au contraire, tous les points d'intersection entre les droites mentionnées sont explicitement introduits par l'énoncé : Larkin et Simon choisissent donc un cas particulièrement simple. Cela leur permet de proposer d'abord une méthode de résolution sans figure, puis de spéculer sur les étapes de celles-ci que l'usage d'une figure rend « faciles et automatiques [59] » grâce à nos capacités perceptuelles.

57. On a besoin de l'égalité des angles alternes internes et des angles opposés par le sommet, et de la superposabilité de deux triangles si deux de leurs angles, et le côté entre ces deux angles, sont égaux.

58. Voir *supra*, p. 109.

59. Larkin et Simon 1987, p. 91.

Plus précisément, ils partent d'une structure de données initiale ne mentionnant que des points et des droites, qui est inutilisable telle quelle puisque les théorèmes à appliquer parlent d'angles et de triangles. Ils expliquent alors comment déterminer algorithmiquement triangles et angles de manière à produire une structure de données « enrichie », sur la base de laquelle le problème peut être résolu. Ils se contentent ensuite de suggérer que c'est le passage de la structure de données initiale à la structure enrichie (c'est-à-dire l'identification des angles et triangles) qui est immédiate par inspection de la figure :

> We have seen that formally producing perceptual elements [60] does most of the work of solving the geometry problem. But we have a mechanism—the eye and the diagram—that produces exactly these "perceptual" results with little effort. We believe the right assumption is that diagrams and the human visual system provide, at essentially zero cost, all of the inferences we have called "perceptual." [61]

En d'autres termes, ils ne proposent pas vraiment *deux* modèles informatiques, mais un seul ; seulement, ils postulent qu'on peut simuler le travail d'un être humain utilisant une figure directement sur la base de la structure de données « enrichie ».

On est loin, dans ce second exemple, de la rigueur programmatique de l'article de 1978. Il y a deux difficultés principales. La première est qu'au lieu de proposer un modèle computationnel pleinement explicite, ils invoquent ici des processus perceptuels non élucidés, qu'ils ne modélisent pas directement. Une conséquence en est que transposer leurs idées au cas des représentations mentales supposerait tout un travail supplémentaire, à moins de postuler un « œil de l'esprit » inspectant des images mentales, ce qui est précisément la stratégie que moquent traditionnellement les critiques de l'imagerie en psychologie [62], et que l'article de 1978 évitait. On peut se demander, d'ailleurs, si ce n'est pas pour cela que malgré les convictions de Simon, nos auteurs sont ici contraints à battre en retraite vers les images externes [63].

La seconde difficulté est qu'ici la figure géométrique n'est pas identifiable à une « représentation » (c'est-à-dire, pour eux, une structure de données munie d'opérations et d'un mécanisme attentionnel) bien déterminée. Larkin et Simon en définissent en fait deux :

60. Il s'agit des angles et des triangles du problème.

61. « Nous avons vu que l'essentiel du travail, pour résoudre le problème de géométrie, est de produire formellement des éléments perceptuels. Or nous avons un mécanisme – l'œil et la figure – qui produit précisément ces résultats "perceptuels" pour peu d'efforts. La bonne hypothèse, d'après nous, est que les figures et le système visuel humain fournissent, de manière gratuitement gratuite, toutes les inférences que nous avons appelées "perceptuelles". » (Larkin et Simon 1987, p. 92).

62. Voir par ex. Pylyshyn 1973.

63. Voir toutefois leur brève note sur les images mentales : Larkin et Simon 1987, p. 97-98.

une structure de données initiale, qui parle de points et de droites, et une structure « perceptuellement enrichie » qui parle aussi de segments, d'angles et de triangles, dont les éléments « sont perceptuellement évidents pour toute personne éduquée qui regarde [la] figure [64] ». C'est cette seconde structure de données qui est censée correspondre à une représentation géométrique du problème. Or il est clair qu'elle ne capture pas la figure elle-même, mais plutôt ce qu'un utilisateur compétent, quoique pas non plus expert, peut en tirer facilement ; en effet, comme le reconnaissent les auteurs, « nous pourrions ajouter des configurations [*patterns*] comme les angles alternes-internes, mais il n'est pas clair que ces configurations soient évidentes pour tout le monde [...] [65] ». Il faudrait en fait concevoir des structures enrichies différentes pour chaque utilisateur et chaque contexte. Loin de caractériser ou de définir les figures géométriques elles-mêmes par les opérations rapides qu'elles permettent, nos auteurs se contentent donc ici de capturer le rôle fonctionnel de la figure dans un contexte particulier et pour un agent particulier.

c) Simon : pionnier paradoxal de la « cognition étendue » ?

La discussion qui précède soulève une question très intriguante. Ce que Larkin et Simon modélisent informatiquement, ce sont en fait des processus entiers de résolution de problèmes. Or il s'avère (au moins dans le second exemple) que les composants de leurs programmes qui sont *a priori* censés modéliser les figures capturent en fait les figures *telles que perçues et utilisées par un agent particulier, dans un contexte particulier*. En d'autres termes, les composants de leur programme ne respectent pas les frontières étroites du crâne. Faut-il en conclure que Simon est un praticien convaincu de la cognition « étendue », selon le slogan « *Cognitive processes ain't (all) in the head!* [66] » de David Chalmers et Andy Clark ?

Ce serait, bien sûr, paradoxal. C'est en opposition explicite à l'intelligence artificielle traditionnelle, dont Simon sert souvent d'emblème, qu'ont été conçus les mots d'ordre d'une cognition « étendue », « incarnée » ou « imbriquée [67] », ainsi que les travaux de

64. « We argue that all elements currently in the perceptually enhanced representation (including points, segments, angles, and triangles) are perceptually obvious to any educated person looking at a diagram like that in Figure 3. » (Larkin et Simon 1987, p. 88).

65. « Although we could add patterns like alternate interior angles, it is not clear that such patterns are obvious to everyone [...] » (Larkin et Simon 1987, p. 88).

66. « Les processus cognitifs ne sont pas (tous) dans la tête ! » (Clark et Chalmers 1998, p. 8). La tournure est un clin d'œil à la célèbre formule de Putnam : « 'meanings' just ain't in the *head* ! » (Putnam 1975, p. 227).

67. En anglais respectivement « *extended* », « *embodied* » et « *embedded* ». Ces termes ne sont pas du tout synonymes (le premier met plutôt l'accent sur l'importance de supports de raisonnement externes, le second sur le du corps, le dernier sur l'environnement) mais les distinctions n'ont pas grande importance pour moi ici. On rencontre aussi une multitude d'expressions apparentées, comme celle de « computationnalisme large » (« *wide computationalism* », cf. R. A. Wilson 1994). Pour une introduction, voir par exemple le bel essai de John

sciences cognitives sur lesquels ils s'appuient [68]. Il est vrai que, par rapport à ce que proposent ces auteurs, on peut trouver le travail de Larkin et Simon plutôt conventionnel. Edwin Hutchins, l'un des défenseurs de ce genre d'approche, écrit par exemple :

> I think it is fair to say that in the twenty years since the publication of *Human Problem Solving* [Newell and Simon 1972] the use of material structure in the problem-solving environment has not been a central topic in the physical-symbol-system research agenda. Some recent work within this tradition takes the "external world" into account [Larkin 1989 [69] ; Vera et Simon 1993] but treats the world only as an extra memory on which the same sorts of operations are applied as are applied to internal memories. Structure in the world can be much more than an augment to memory. [70]

Peut-être Simon lui-même aurait-il accepté cette description de son travail. Pour mon propos, toutefois, l'essentiel est ailleurs : du point de vue de Simon, la distinction même entre mémoire externe et mémoire interne n'a que peu de sens. Les représentations ne sont qu'un rôle fonctionnel au sein d'un processus de résolution, dont la réalisation matérielle peut être diffuse. Chez Larkin et Simon, comme nous l'avons vu, il est bien difficile de localiser d'un côté ou de l'autre la « structure de données enrichie » de leur problème de géométrie.

En réalité, c'est la méthode même de simulation informatique de Simon qui est intrinsèquement radicale et tend à brouiller les frontières usuelles. Y a-t-il d'ailleurs meilleur exemple de ce brouillage systématique que son projet de « sciences de l'artificiel », qu'il introduit, dans un passage célèbre, en comparant la recherche d'une démonstration au chemin d'une fourmi sur une plage ?

> We watch an ant make his laborious way across a wind- and wave-molded beach. He moves ahead, angles to the right to ease his climb up a steep dunelet, detours around a pebble, stops for a moment to exchange information with

Haugeland, « *Mind embodied and embedded* » : Haugeland [1995] 1998b.

68. Parmi les travaux les plus souvents évoqués dans ce contexte, mentionnons Gibson 1979 et Brooks 1991, qui insistent sur l'importance de l'environnement, ainsi que Kirsh et Maglio 1994 et Hutchins 1995 qui se concentrent sur l'emploi de supports externes au cours du raisonnement.

69. Cet article de Larkin est d'un esprit très proche de celui que nous avons étudié, que Hutchins aurait tout aussi bien pu citer.

70. « Il est juste de dire, je pense, que dans les vingt ans écoulés depuis la publication de *Human Problem Solving* [Newell et Simon 1972], cela n'a pas été un thème central du programme de recherche sur les systèmes physiques de symboles que d'étudier comment la structure matérielle de l'environnement de résolution de problème était utilisée. Il y a certes des travaux récents de cette tradition qui prennent en compte le "monde extérieur" [Larkin 1989 ; Vera et Simon 1993], mais ils ne traitent le monde que comme une mémoire supplémentaire qui permet les mêmes opérations que les mémoires internes. La structure disponible dans le monde peut être bien plus qu'un auxiliaire à la mémoire. » (Hutchins 1995, p. 369).

> a compatriot. [...] I sketch the path on a piece of paper. It is a sequence of irregular, angular segments—not quite a random walk, for it has an underlying sense of direction, of aiming toward a goal.
>
> I show the unlabelled sketch to a friend. Whose path is it? An expert skier, perhaps, slaloming down a steep and somewhat rocky slope. Or a sloop, beating upwind in a channel dotted with islands or shoals. Perhaps it is a path in a more abstract space: the course of search of a student seeking the proof of a theorem in geometry. [71]

À titre d'exemple, Simon suggère dans cet ouvrage que pour simuler la résolution humaine de problème, il peut être utile de considérer la mémoire (au sens habituel du terme) comme une partie de l'environnement [72] ; c'est dire si les frontières du crâne le préoccupent peu.

4.6 Conclusion : la démarche de Simon

Résumons la démarche de Simon. Il part d'un modèle informatique, celui du stockage de données dans un ordinateur. C'est dans ce contexte qu'il introduit les notions d'équivalence informationnelle et d'équivalence computationnelle, qui ont alors un sens clair. Il tente ensuite d'aborder, à l'aide de ce modèle informatique, les représentations mentales mais aussi les représentations « externes » dans les sciences ou les mathématiques.

Le passage du modèle informatique aux représentations en général se justifie de deux manières. Il est tout d'abord fondé sur un postulat fondamental qui parcourt toute l'œuvre de Simon : au bon niveau d'analyse, le niveau « informationnel » ou « fonctionnel », tout processus de résolution de problème est identifiable à un programme informatique. D'autre part et plus spécifiquement, Simon tente d'élaborer une notion abstraite de représentation qui fasse sens à ce niveau d'analyse-là ; celle-ci est censée légitimer sa démarche en montrant que d'un point de vue de traitement de l'information, représentations externes,

---

71. « Observons une fourmi qui progresse laborieusement sur une plage façonnée par le vent et par les vagues. Elle avance, tourne à droite pour escalader plus facilement une dunelette raide, contourne un galet, s'arrête un moment pour échanger des informations avec une congénère. [...] J'esquisse son chemin sur un bout de papier. C'est une ligne brisée et irrégulière – pas vraiment une marche aléatoire, parce qu'elle témoigne d'un sens de la direction, de la visée d'un but. Je montre l'esquisse à un ami, sans légende. De qui est-ce le chemin ? D'un skieur expérimenté, peut-être, qui descend en slalomant une pente raide et un peu caillouteuse, ou d'un petit voilier, luttant contre le vent dans un chenal étroit parsemé d'îles ou de récifs. Peut-être est-ce un chemin dans un espace plus abstrait : l'itinéraire exploratoire d'un étudiant qui cherche la démonstration d'un théorème de géométrie. » (Simon 1996, chap. 3, p. 51. Il faut prendre garde au fait que Simon a profité de chaque nouvelle édition des *Sciences de l'artificiel* pour réécrire de larges parties de l'ouvrage. L'image de la fourmi est toutefois présente dès la première édition de 1969.)

72. Simon 1996, chap. 4 : « Remembering and Learning: Memory As Environment for Thought ». Ce chapitre, absent de la première édition (Simon 1969) a été introduit en 1981.

représentations mentales et organisation des données dans un ordinateur peuvent être assimilées. Nous avons vu, cependant, que ses tentatives de clarifier sa notion abstraite de représentation conduisaient à des difficultés de fond, parce que le lien entre organisation des données et efficacité computationnelle des opérations portant sur elles n'est pas clair et semble dépendre de l'architecture computationnelle utilisée.

En collaboration avec la psychologue Jill Larkin (1987), Simon traite de manière plus précise deux exemples d'usage de figures, l'un en statique, l'autre en géométrie. Dans le prolongement de la méthodologie générale de Simon, ils proposent des programmes informatiques censés modéliser des procédures de résolution de problème et identifient les figures à des composants de ces programmes. Cela permet de conserver un sens précis aux notions d'équivalence informationnelle et d'équivalence (ou de différences) computationnelle, mais induit un décalage avec les figures concrètes qui étaient le point de départ de l'analyse. C'est cette tension qu'il importera avant tout de garder à l'esprit lorsque, au prochain chapitre, nous lirons les travaux d'auteurs qui reprennent les concepts de Simon mais pas sa méthodologie très particulière.

Chapitre 5

Différences computationnelles partout ?

Que peut-on faire des concepts de Simon ? La littérature existante offre de nombreuses pistes. Ils ont été utilisés pour décrire et unifier les phénomènes les plus divers, de la visualisation de données aux différentes formes des équations de la mécanique en passant par les notations en logique propositionnelle. Je voudrais ici en dresser un panorama, puis revenir, sur cette base, à l'étude de cas de la partie I.

Avant de commencer, toutefois, un avertissement s'impose : il faut garder, vis-à-vis de ces travaux, une certaine distance critique. Souvent, les concepts de Simon y sont réinterprétés ou utilisés de manière purement heuristique ; il arrive aussi que les tensions déjà présentes chez Simon y atteignent le point de rupture. Je tente ici de suivre les intuitions des auteurs que je traite, et me contente de noter au passage les difficultés les plus claires. Une évaluation sérieuse devra attendre le prochain chapitre.

5.1 Diagrammes et figures en mathématiques

Un travail récent de Jessica Carter (2018) fournit un premier exemple d'application des idées de Simon aux mathématiques. Son but est de comparer l'efficacité de représentations « diagrammatiques » et « formelles » de graphes (en l'occurence dans le contexte d'une méthode de calcul de K-groupes, mais les détails n'ont pas d'importance ici). Elle prend l'exemple d'un graphe simple, que l'on peut représenter « diagrammatiquement » (fig. 5.1) mais aussi de manière « formelle », c'est-à-dire en spécifiant l'ensemble $E^0 = \{v_1, v_2, v_3\}$ de ses sommets, l'ensemble $E^1 = \{e_1, e_2, e_3, e_4, e_5\}$ de ses arêtes, et les fonctions $r, s : E^1 \to E^0$

$$\begin{tikzcd} v_1 \arrow[loop left, "e_1"] \arrow[loop above, "e_2"] \arrow[r, "e_3"] & v_2 \arrow[r, "e_4"] & v_3 \arrow[loop above, "e_5"] \end{tikzcd}$$

FIGURE 5.1 – Un diagramme représentant un graphe orienté (Carter 2018, p. 4)

donnant respectivement la destination et l'origine de chaque arête :

$$\begin{aligned} &r(e_1) = r(e_2) = v_1, && r(e_3) = v_2, && r(e_4) = r(e_5) = v_3, \\ &s(e_1) = s(e_2) = s(e_3) = v_1, && s(e_4) = v_2, && s(e_5) = v_3. \end{aligned}$$

« Le contenu informationnel », écrit-elle, « est le même [1] », et le calcul qui l'intéresse peut être fait sur la base de ces deux représentations. Mais, comme elle l'écrit en faisant référence à Larkin et Simon (1987), partir de la représentation formelle « requiert davantages de ressources cognitives, c'est-à-dire davantage d'observations de faits et de déplacements d'attention qu'en regardant le diagramme [2] ». Selon son analyse, l'avantage de la représentation diagrammatique est essentiellement dû au fait que toutes les arêtes liées à un même sommet sont accessibles au même endroit, ce qui diminue les efforts requis pour localiser les données nécessaires à chaque étape. On est donc très proche du premier avantage des diagrammes repéré par Larkin et Simon [3].

Bon nombre de ce que l'on appelle « diagrammes » dans les mathématiques contemporaines seraient susceptibles d'une analyse similaire, parce que leur élimination systématique, c'est-à-dire leur traduction en énoncés, ne pose aucune difficulté de principe. Les « diagrammes commutatifs » largement utilisés en algèbre, géométrie et topologie ou encore en théorie des catégories en sont un exemple clair [4]. Le diagramme commutatif suivant, par exemple,

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "\phi_A"'] & B \arrow[d, "\phi_B"] \\ A' \arrow[r, "f'"'] & B' \end{tikzcd}$$

1. « […] the informational content is the same […] », Carter 2018, p. 8.
2. « […] it requires more cognitive resources, that is, recording a greater number of facts and more shifts of attention than by looking at the diagram. » (Carter 2018, p. 9).
3. Voir *supra*, section 4.5.
4. Pour une introduction aux diagrammes commutatifs, voir par exemple Lang 2002, p. ix-x, ou à peu près n'importe quel manuel raisonnablement récent de deuxième cycle en géométrie, topologie ou algèbre. Silvia de Toffoli 2017 en donne une présentation destinée à un public de philosophes dans le contexte plus spécifique de l'algèbre homologique élémentaire.

se laisse traduire sans reste par le fait que f est une application [5] de A dans B, f' de A' dans B', ϕ_A de A dans A', ϕ_B de B dans B', et par la relation de commutation $f'\phi_A = \phi_B f$. On pourrait en dire autant des nombreuses variantes de ce genre de diagrammes qui permettent de les enrichir au moyen de conventions supplémentaires [6]. La grande facilité de ce genre de traduction (qui explique, au passage, pourquoi ce genre de diagrammes ne posent pas les problèmes de rigueur souvent associés à l'usage de figures au cours de démonstrations et sont admis sans inquiétude [7]) appelle tout naturellement une analyse dans le style de Simon de ce que les diagrammes commutatifs apportent. À vrai dire, leur avantage essentiel est peut-être ailleurs. Parler d'équivalence informationnelle entre le diagramme et une liste de relations de commutation risque d'occulter que l'on sait normalement beaucoup plus que ce que le diagramme représente : on en sait plus sur les applications que leurs seules relations de commutation, et on a en général affaire à bien plus d'applications que celles qui figurent sur le diagramme. Le diagramme peut en fait servir à *isoler* les informations pertinentes autant qu'à les présenter sous une forme efficace ; j'y reviendrai. Néanmoins, le fait que les diagrammes commutatifs permettent de localiser efficacement, à chaque étape de la preuve, les flèches dont on peut avoir besoin fait indéniablement partie de leurs qualités. Lorsque de Toffoli (2017) écrit que les diagrammes commutatifs fonctionnent « comme des cartes [8] », elle ne dit pas autre chose, et on pourrait en rendre compte à la manière de Carter ou de Larkin et Simon.

À première vue, le cas des figures en géométrie élémentaire est plus délicat. Quand on parle aujourd'hui de figures géométriques en mathématiques, on pense d'abord à une question très différente de celle de Simon : leur usage est-il légitime ? N'introduisent-elles pas des éléments qui, précisément, ne sont pas dans le texte ? Dans son premier article, Simon évacue la question en supposant qu'on dispose d'une « axiomatisation appropriée

5. Ou un morphisme ou plus généralement une flèche, en fonction du contexte, mais cela n'a pas d'importance ici.

6. On peut par exemple décorer les flèches pour indiquer si ce sont des isomorphismes ($\xrightarrow{\sim}$), des surjections ($\twoheadrightarrow$) ou des injections ($\hookrightarrow$) ; d'autres conventions permettent de suggérer diagrammatiquement que certaines suites de flèches forment des « suites exactes » ; et ainsi de suite.

7. Il faudrait nuancer ce jugement. Si en principe les diagrammes commutatifs ne posent aucun problème intrinsèque de rigueur et qu'ils sont largement acceptés, en pratique ils peuvent s'avérer dangereux et sont une source récurrente d'erreurs et d'omissions dans les preuves. Ils permettent en effet de représenter simultanément plusieurs relations de commutation et donc potentiellement de ne pas trop attirer l'attention sur le fait que toutes n'ont pas été justifiées explicitement… Jean-Pierre Serre se plaint par exemple dans un entretien de ces « preuves fondées sur des diagrames prétendument commutatifs, des flèches qui sont prétendument les mêmes et des arguments laissés au lecteur. » (« proofs based on diagrams which are claimed to commute, arrows which are supposed to be the same, and arguments which are left to the reader. », Raussen et Skau 2004, p. 212-213.)

8. De Toffoli 2017, p. 177-178, section 4.2 : « CDs function as *maps* ».

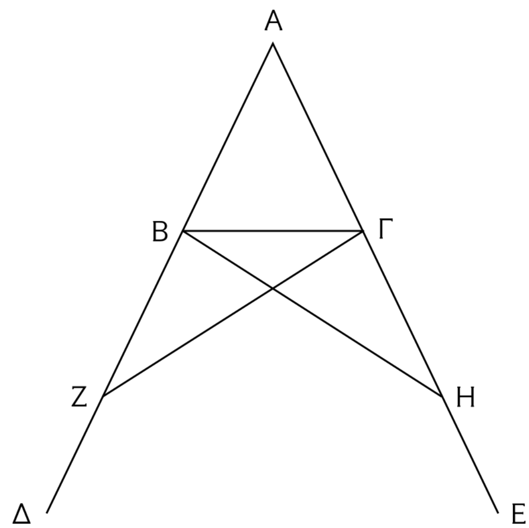


FIGURE 5.2 – Figure de la prop. I.5 des *Éléments* d'Euclide, d'après Netz (1999, p. 175)

de la géométrie plane [9] » ; il tient ensuite pour acquis qu'une figure géométrique peut nous montrer l'existence d'un point d'intersection entre deux droites [10]. Dans son second article écrit avec Jill Larkin, il prend soin de choisir un exemple où aucun détour par des axiomes complexes n'est nécessaire pour écrire le raisonnement sans figure [11].

Mais en réalité, même dans ce cas l'approche de Simon est pertinente, comme le montre par exemple l'étude de Reviel Netz sur la géométrie grecque (1999). Netz y examine les diverses manières dont les figures interviennent dans les preuves [12]. Plus précisément, il s'intéresse aux assertions qu'il appelle les « points de départ » (« *starting points* ») : celles qui, au cours d'une preuve, ne sont pas justifiées par celles qui précèdent. Certains de ces points de départ sont manifestement fondés sur une inspection de la figure. Certes, écrit Netz, il y a des cas où on peut identifier des postulats manquants, sans lesquels la preuve ne peut pas être reformulée sans figure [13]. Dans un grand nombre de cas, ajoute-t-il cependant, si l'appel à la figure n'est pas indispensable en ce sens, il offre pourtant un genre de « raccourci » très utile.

Penchons-nous sur deux des exemples de Netz. Le premier concerne la proposition I.5 des *Éléments* d'Euclide, qui montre (entre autres) que les angles à la base d'un triangle

9. « an appropriate axiomatization of plane geometry », Simon 1978, p. 6.
10. Voir *supra*, p. 109.
11. Cf. p. 124 ci-dessus.
12. Cf. Netz 1999, chap. 5, section 1, en part. p. 175–182.
13. Cf. p. 180

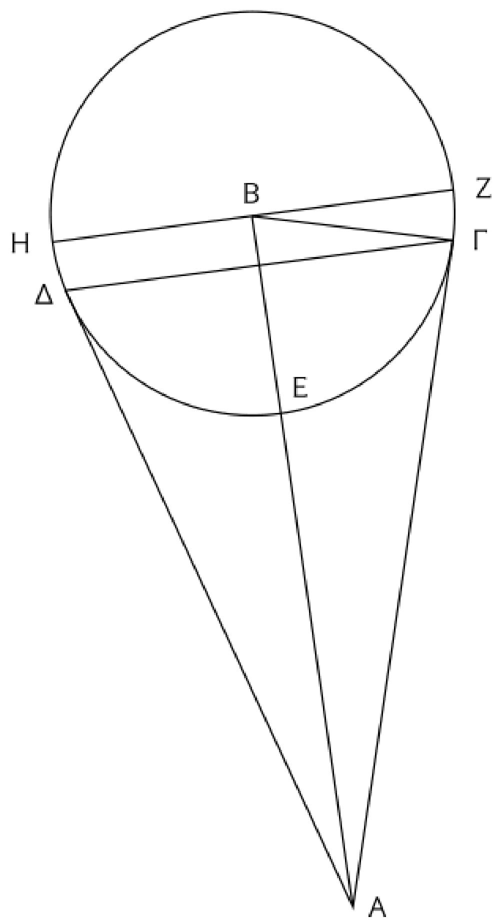


Figure 5.3 – Figure de la prop. 12 du *Traité sur les grandeurs et les distances du Soleil et de la Lune* d'Aristarque de Samos, d'après Netz (1999, p. 179)

isocèle sont égaux [14] (voir fig. 5.2). Une étape de la preuve est de reconnaître que les angles ZAΓ et HAB sont en fait le même angle, à savoir ZAH. Cela peut se faire sans la figure ; tout au plus faudrait-il ajouter une justification, à savoir que Z, B, A d'une part, H, Γ, A d'autre part sont alignés, ce qui est vrai par définition de Z et H. La figure offre cependant un raccourci manifeste pour vérifier cette égalité, qui permet d'en laisser la justification implicite.

Le second exemple de Netz est tiré de la proposition 12 du *Traité sur les grandeurs et les distances du Soleil et de la Lune* d'Aristarque de Samos [15] (fig. 5.3). L'étape qui intéresse Netz est que (l'arc) ΔEΓ est double de (l'arc) ΓE. Là encore, l'affirmation n'est pas justifiée,

14. Cf. EEV, I, p. 204–205 = EEHea, I, p. 251–252 = EEHei, I, p. 20–23
15. Cf. Heath 1913, p. 388-391.

cette fois parce que la figure présente une symétrie évidente ; pour autant, elle pourrait l'être au moyen d'un argument implicite facile à reconstituer. Netz conclut en des termes que Simon n'aurait pas reniés :

> This does not mean that the diagram adds information, but it does mean that the diagram saves 'logical space', as it were, and thus makes deduction at least easier. [16]

5.2 Visualisation de données

En philosophie des sciences, John Kulvicki (2010) a proposé de penser aussi en termes de différences computationnelles certains phénomènes assez différents, caractéristiques de la *visualisation de données*. L'enjeu de ce domaine est bien résumé par la figure 5.4, tirée d'un ouvrage fondateur de Jacques Bertin [17] : on a des colonnes de chiffres à n'en plus finir ; comment présenter ces données d'une manière éclairante, c'est-à-dire qui permette facilement de les comparer, de remarquer des corrélations, etc. ? Il peut sembler naturel de dire que les nombreuses manières de représenter les mêmes données diffèrent par les informations auxquelles elles nous donnent un accès *facile*, donc par leurs propriétés computationnelles plutôt qu'informationnelles.

a) Le cadre de Kulvicki

Kulvicki prend un exemple simple. Partons d'une liste de températures relevées à différents endroits d'un pays (pour simplifier, on supposera que ce pays a la forme d'une grille rectangulaire de 4 cases sur 4 et que les relevés sont indexés par deux coordonnées entre

16. « Cela ne signifie pas que la figure ajoute de l'information ; ce que cela signifie en revanche, c'est que la figure épargne pour ainsi dire de l''espace logique', et ainsi rend la déduction pour le moins plus facile. » (Netz 1999, p. 179-180).

17. Bertin 1967, p. 100. Jacques Bertin, cartographe à l'EPHE, a fortement marqué la réflexion de générations de chercheurs français sur les modes de représentation, en particulier en géographie (pour une introduction illustrée à son travail, voir par exemple Palsky 2017 ; pour une contextualisation, voir Palsky et Robic 2000 et Palsky 2012). À l'échelle mondiale, il a été un acteur parmi d'autres d'un renouveau de la visualisation après une période de reflux pendant laquelle, entre autres sous l'influence des nouvelles méthodes statistiques pour estimer rigoureusement les corrélations, les méthodes visuelles avaient quelque peu perdu de leur légitimité scientifique (une certaine méfiance vis-à-vis du manque de rigueur et du caractère potentiellement trompeur des graphiques persiste d'ailleurs encore). Dans le monde anglo-saxon et en-dehors de la cartographie, un auteur essentiel de ce renouveau a été le statisticien John Tukey (voir en particulier Tukey 1972, pour une présentation rapide, et son *magnum opus* Tukey 1977), qui a mis l'accent sur la productivité des visualisations pour l'analyse exploratoire de données. Les beaux livres de son collègue et collaborateur Edward Tufte ont ensuite beaucoup fait pour attirer l'attention d'un public large sur le potentiel de représentations graphiques rigoureuses et bien réfléchies (voir avant tout Tufte 1983, et aussi Tufte 1990, 1997). Pour une introduction à l'histoire de la visualisation de données, voir par exemple Friendly 2008.

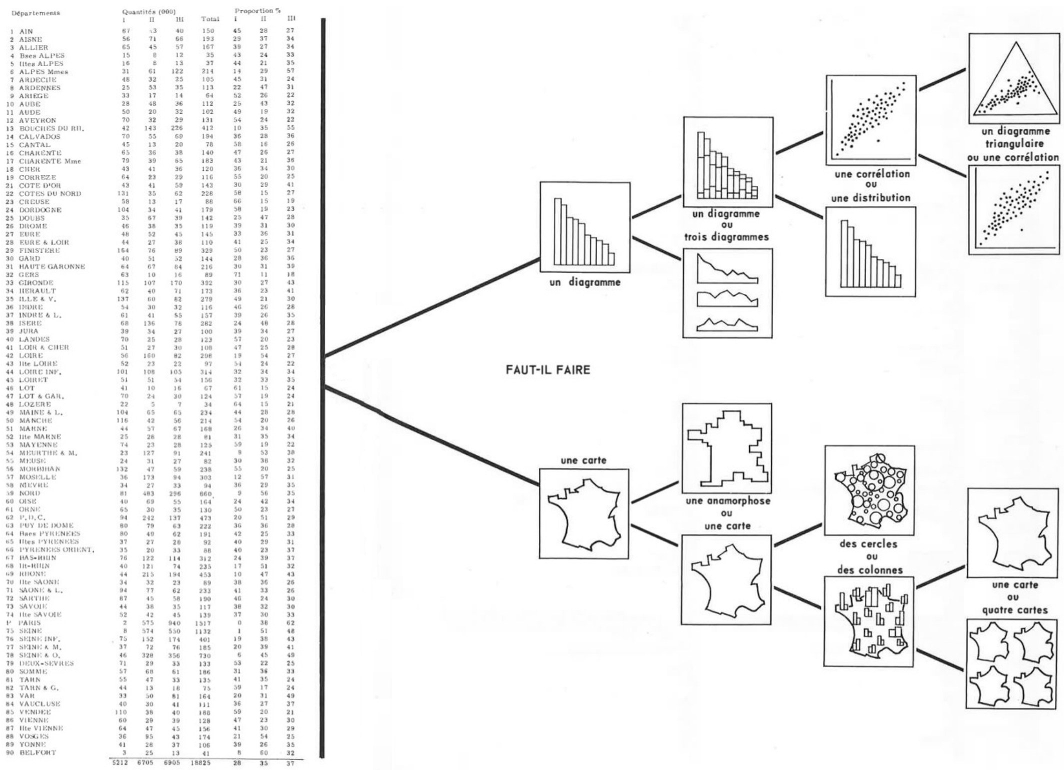
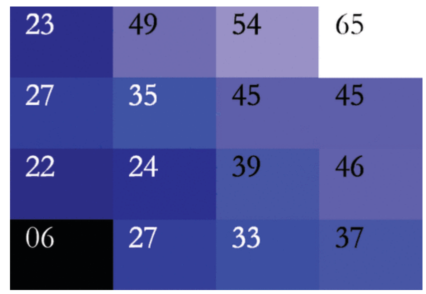


FIGURE 5.4 – « Le problème graphique » d'après Bertin (1967, p. 100) : « La décision de transcrire graphiquement une information devrait reposer sur une appréciation de l'efficacité de chaque langage, de chaque système d'expression. » Les données à illustrer donnent la population active en France en 1954, par département et grand secteur d'activité (primaire, secondaire, tertiaire).

1, 1, 06	3, 1, 33
1, 2, 22	3, 2, 39
1, 3, 27	3, 3, 45
1, 4, 23	3, 4, 54
2, 1, 27	4, 1, 37
2, 2, 24	4, 2, 46
2, 3, 35	4, 3, 45
2, 4, 49	4, 4, 65

(a)



(b)

FIGURE 5.5 – Deux représentations de relevés de température (adaptées de Kulvicki 2010, p. 303, 305)

1 et 4). On peut représenter ces températures sous la forme d'une liste (fig. 5.5(a)) ou les disposer sur une carte ; une variante possible est de colorer la carte en fonction de la température, disons en choisissant des nuances de bleu plus ou moins foncées [18] (fig. 5.5(b)). Il est clair ici que l'on peut reconstituer la liste à partir de la carte et inversement ; selon le critère de Simon, elles sont donc informationnellement équivalentes. Pourtant, elles ne nous facilitent pas les mêmes observations. La seule chose que la liste (fig. 5.5(a)) nous offre immédiatement, d'après Kulvicki, c'est « la température en tel endroit est de tant », alors que la carte colorée (fig. 5.5(b)) nous donne immédiatement non seulement les valeurs exactes, mais aussi leur ordre de grandeur, et bien davantage : le fait que les températures maximales et minimales sont dans les coins, que les températures sont plus basses dans le sud-ouest et plus hautes dans le nord-est, et ainsi de suite. Plus généralement, Kulvicki écrit que l'avantage central des graphiques, diagrammes, etc. est qu'ils nous donnent immédiatement accès à des informations *à différents niveaux d'abstraction* : des valeurs plus ou moins précises, mais aussi leurs ordres de grandeur, des propriétés générales de leur répartition, etc.

Kulvicki parle comme je l'ai fait d'informations « immédiatement disponibles », ce qui lui sert à fonder les différences computationnelles entre représentations sur quelque chose d'apparenté à un contenu explicite. Il définit les informations immédiatement disponibles par trois conditions. Tout d'abord, il faut qu'il y ait une caractéristique de la représentation qui porte cette information et aucune autre plus précise (« extractibilité »). Ensuite, il faut que cette caractéristique soit facile à reconnaître pour l'observateur (« saillance syntaxique »). Enfin, il faut qu'il soit facile à l'observateur d'interpréter cette caractéristique

18. Les températures représentées sont données en degrés Fahrenheit et vont donc de -14 °C à 18 °C environ. Kulvicki enseigne en Nouvelle-Angleterre, où les hivers sont rudes. En tout cas, on comprend le choix du bleu.

(« saillance sémantique »).

b) Une application des concepts de Simon ?

Cette adaptation de l'idée de Simon requiert des modifications profondes. Une première différence importante avec Simon est que celui-ci cherchait initialement à caractériser les représentations au moyen d'opérations élémentaires clairement définies. Le passage à la visualisation de données suppose au contraire d'invoquer des processus perceptuels vagues et mal délimités ; ce qui « saute aux yeux » dépend à la fois des détails du système perceptuel humain en général et de l'expertise particulière de l'observateur. Nous avons déjà rencontré le problème de l'expertise dans la partie la plus spéculative de l'article de Larkin et Simon [19] : sur une figure de géométrie, faut-il considérer que les angles alternes-internes sont immédiatement évidents ? Cela dépend de qui la regarde [20]. Mais même sans entrer dans ce genre de différences individuelles, il est bien difficile de décrire précisément ce que le système visuel humain rend facile à identifier.

Les méthodes de visualisation les plus folkloriques rendent ce problème particulièrement clair. Une illustration frappante, que j'emprunte à Mark Wilson [21], en est les « visages » introduits par Hans Chernoff (1973) pour visualiser des données en de nombreuses variables. Les statisticiens suisses Bernhard Flury et Hans Riedwyl (1981) les emploient par exemple pour distinguer billets authentiques et faux billets de 1000 francs suisses [22] : ils font six mesures sur chaque billet (fig. 5.6) puis encodent ces six variables dans des déformations de visages schématiques (fig. 5.7). Ils écrivent :

> Though the univariate frequency distributions of all six variables overlap considerably, real and forged notes can easily be distinguished by a first glance at the faces. This is because of the human ability to perceive not only details of an object but its gestalt. The real faces seem to look at us in a way that is hard to describe but that is quite striking. Beside this, a lot of detailed information can be taken from the faces [...]. [23]

19. Voir *supra*, section 4.5.
20. Cf. ci-dessus, section 4.5.b), en particulier note 65 p. 126.
21. M. Wilson 2006, p. 492.
22. Voir aussi leur manuel Flury et Riedwyl [1983] 1988, qui utilise l'exemple des billets de banque comme fil directeur pour présenter diverses méthodes d'analyse de données (dont les visages de Chernoff).
23. « Bien qu'il y ait beaucoup de chevauchements entre les distributions statistiques de chacune des six variables prise individuellement, on distingue les billets authentiques des faux au premier coup d'œil sur les visages. Cela s'explique par la capacité humaine à percevoir non seulement les détails d'un objet mais aussi sa *gestalt*. Les visages authentiques semblent nous regarder d'une manière difficile à décrire, mais qui est très frappante. Par ailleurs, on peut tirer des visages beaucoup d'informations détaillées [...] » (Flury et Riedwyl 1981, p. 761-762).

FIGURE 5.6 – Ce billet de 1000 francs suisses est-il un faux ? Pour le déterminer, Flury et Riedwyl mesurent les longueurs $x_1, \ldots, x_6$ et les encodent dans des visages (Flury et Riedwyl [1983] 1988, p. 4).

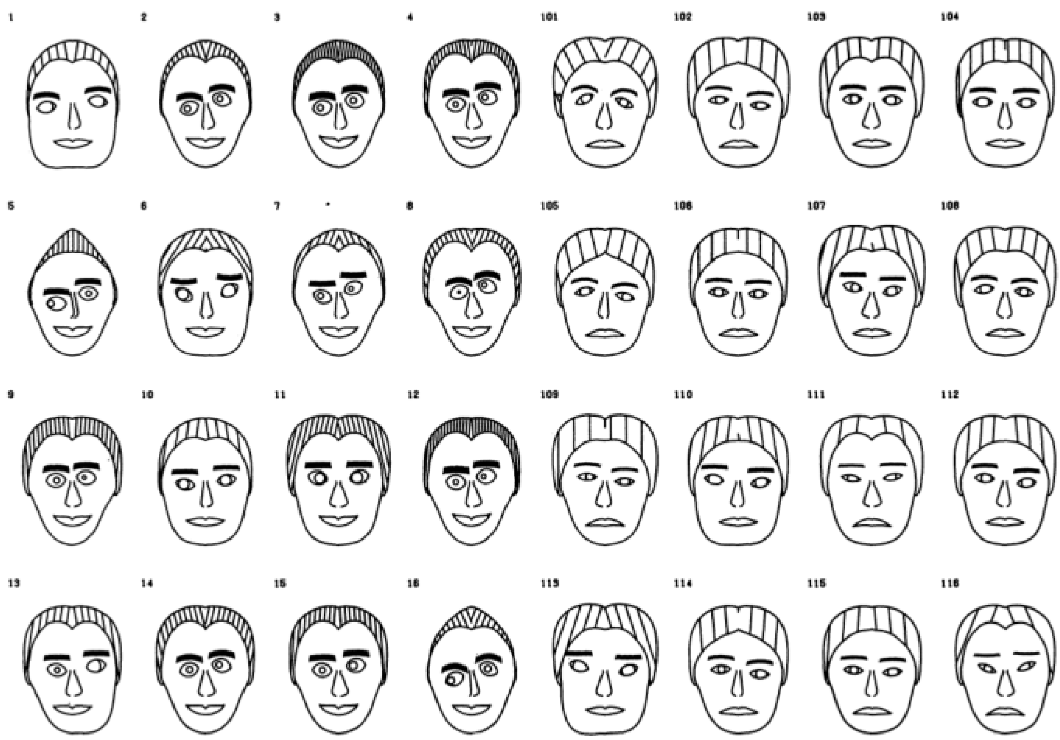


FIGURE 5.7 – Visages encodant les mesures réalisées sur 16 billets authentiques, à gauche, et 16 billets produits par un même faussaire, à droite (Flury et Riedwyl 1981, p. 762)

En d'autres termes, toute la difficulté est qu'aucune des mesures prise individuellement n'est suffisante pour distinguer vrais et faux billets, parce que pour chacune d'entre elles, la variation entre billets authentiques est déjà très grande. C'est quand on les considère ensemble qu'on commence à voir apparaître des régularités ; les visages de la fig. 5.7 attirent par exemple notre attention sur l'épaisseur des cheveux, qui correspond ici en gros à la somme des marges du haut et du bas de nos billets. C'est pour nous aider dans cette exploration que les visages sont particulièrement efficaces. Or s'ils le sont, c'est uniquement à cause d'une spécificité de notre système visuel, qui est extrêmement sensible aux expressions faciales. Le projet de délimiter précisément ce qui peut sauter aux yeux d'un être humain s'annonce drôlement complexe.

Parler de « différences computationnelles » dans le cas de la visualisation de données soulève un second problème, plus sérieux : dans ce contexte, il est délicat de délimiter non seulement ce qu'il est *aisé*, mais même ce qu'il est *légitime* de lire sur un graphique. D'une manière générale, le projet de Simon implique de présupposer que les opérations permises par les différentes représentations sont fixées pour comparer ensuite leur efficacité ; cela le conduit à laisser complètement de côté la question classique de savoir si un diagramme peut nous induire en erreur. Kulvicki ne fait pas autrement. Il considère que ce qui est lisible sur un graphique est fixé dès le départ par nos pratiques (c'est-à-dire par des conventions implicites) et qu'on peut le tenir pour acquis :

> [T]here are doubtless many ways to interpret any given artifact as a representation. Nelson Goodman [1968, p. 229], for example, used the example of treating a Hokusai woodblock print as an electrocardiogram. Doing so means that one takes on different commitments concerning how nonsemantic features of the graph relate to what they represent. This article offers no account of how these practices conspire to generate such norms of use. Given that they are established, this article asks questions about those practices. [24]

Pourtant, si ce problème est minoré, chez Simon, par les exemples qu'il prend soin de choisir, il est omniprésent en visualisation de données.

Pour être plus précis, il faut distinguer deux usages des graphiques. Il arrive qu'on les utilise pour faire voir de manière particulièrement persuasive une conclusion que l'on tient

24. « Il est clair qu'il y a de nombreuses manières d'interpréter n'importe quel artefact donné comme une représentation. Nelson Goodman [1968, p. 229], par exemple, donne l'exemple d'une gravure de Hokusai que l'on traite comme un électrocardiogramme. Faire cela revient à adopter des engagements différents concernant le lien entre caractéristiques non sémantiques du graphe et ce qu'elles représentent. Cet article ne cherche pas à rendre compte de la manière dont de telles pratiques engendrent de telles normes d'usage ; il suppose que ces pratiques sont établies et s'interroge sur elles. » (Kulvicki 2010, p. 299-300).

pour acquise, par exemple « le déficit budgétaire de l'État explose » ; ce que l'on est censé lire sur le graphique est alors très limité et le contexte le rend limpide. John Tukey, lorsqu'il entreprend de réhabiliter les méthodes de visualisation [25], parle dans ce cas de « graphiques de propagande [26] », qu'il oppose aux graphiques qui l'intéressent, dont l'usage est essentiellement *exploratoire*. C'est bien de ceux-là que traite aussi Kulvicki : on représente des données pour favoriser leur inspection, sans savoir exactement par avance ce qu'on pourra y voir. Or dans ce contexte, le lien entre les observations que l'on fait sur le graphique et leur interprétation est intrinsèquement problématique. Nous examinons cinquante visages et croyons reconnaître chez certains une physionomie récurrente, aux longs cheveux et au regard torve. D'accord, mais qu'est-ce que cela veut dire ? Y a-t-il vraiment un groupe de billets qui se détache, qui sont plus proches entre eux que de la majorité des autres ? Y a-t-il vraiment un faussaire à l'œuvre ? Ce sont là des questions substantielles, qui, *pace* Kulvicki, ne se réduisent pas à une affaire de « saillance sémantique » donnée d'avance.

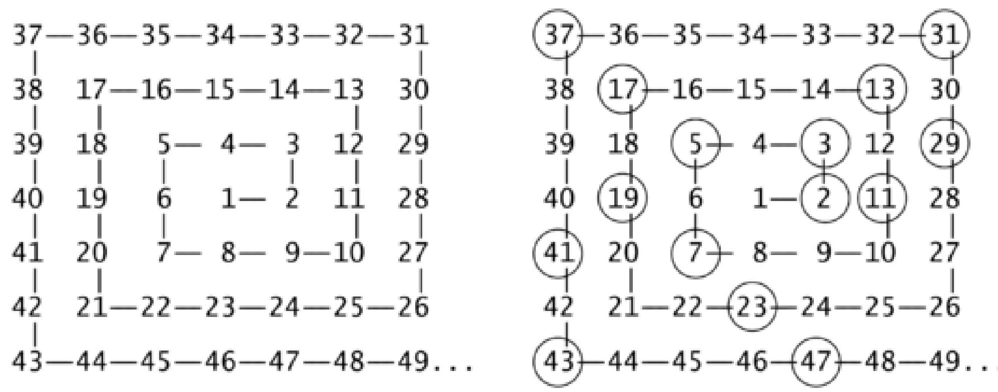
c) Un exemple mathématique

Afin d'y voir plus clair, examinons un exemple un peu moins exotique, qui a l'avantage supplémentaire de montrer que la visualisation exploratoire de données a parfois sa place en mathématiques, et ce jusque dans leurs branches les plus pures. En 1963, Stanislaw Ulam découvre par hasard une propriété étrange de la distribution des nombres premiers. À en croire Martin Gardner, qui rend cet épisode célèbre [27], c'est initialement pour tromper l'ennui lors d'une conférence qu'Ulam se met à représenter les entiers sous forme de spirale, comme à la figure 5.8(a). En marquant machinalement les nombres premiers (figure 5.8(b)), il s'aperçoit qu'ils semblent se regrouper le long de diagonales. Comme il l'écrit lorsqu'il publie sa découverte :

> A glance at [such] a picture [. . .] reveals many such lines. It is a property of the visual brain which allows one to discover such lines at once and also notice many other peculiarities of distribution of points in two dimensions. [28]

Au passage, cela montre une nouvelle fois la difficulté de codifier ce qui peut, sur une figure, sauter aux yeux d'un être humain. Mais ce qui m'importe ici, c'est autre chose. Que faut-il

25. Voir *supra*, note 17, pour la position de Tukey dans l'histoire de la visualisation de données.
26. Tukey 1972, p. 293.
27. Voir Gardner 1964, p. 122. (En ce qui me concerne, j'ai découvert cette anecdote dans Delahaye 2000, p. 29-31 au hasard de la préparation d'un cours d'arithmétique.)
28. « Le premier coup d'œil à une [telle] image [. . .] révèle de nombreuses droites de ce genre. C'est une propriété du cerveau visuel qui permet de découvrir immédiatement de telles droites, et aussi de remarquer bien d'autres bizarreries dans la distribution de points en deux dimensions. » (Stein, Ulam et Wells 1964, p. 517).

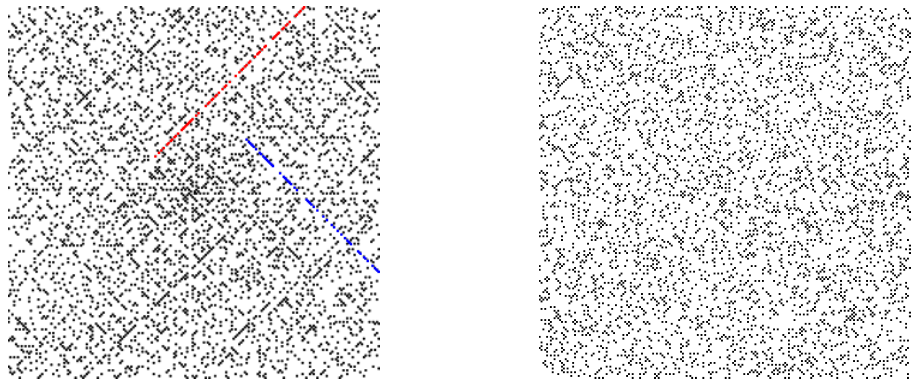


(a) (b)

FIGURE 5.8 – Début d'une spirale d'Ulam ; à droite, les nombres premiers sont entourés.

conclure de ces droites qui nous apparaissent ? Faut-il en déduire que les nombres premiers aiment les diagonales ? que les nombres premiers *inférieurs à 64* aiment les diagonales ? ou peut-être que c'est notre cerveau qui aime les lignes bien régulières et croit volontiers en apercevoir dans une répartition tout ce qu'il y a de plus aléatoire ? D'ailleurs, à quels ensembles de nombres ces diagonales correspondent-elles donc ? En pratique, reconnaître des alignements dans notre spirale n'est qu'un début.

Remarquons tout d'abord, à la suite de Gardner, que les nombres *pairs* et *impairs* se répartissent selon des diagonales : si l'on traçait notre spirale sur un plateau d'échecs, les uns occuperaient les cases noires, les autres les cases blanches. Comme les nombres premiers sont impairs, ils apparaîtront donc naturellement le long de diagonales à condition d'être suffisamment nombreux parmi les nombres impairs pour ne pas laisser trop de trous. Or il y a beaucoup de nombres premiers parmi les entiers de 1 à 64 (environ un impair sur deux) ; cela pourrait donc suffire à expliquer les alignements de la figure 5.8(b). Mais s'il en est ainsi, l'effet devrait s'atténuer à mesure que l'on prolonge la spirale, puisque les nombres premiers se font de plus en plus rares. Ulam, qui travaillait aux laboratoires atomiques de Los Alamos, avait accès à l'un des tout premiers ordinateurs électroniques, le « MANIAC II » ; avec deux collègues, il fait alors produire à cette machine des spirales allant jusqu'à environ 65000, et constate que l'effet reste très net (Stein, Ulam et Wells 1964). C'est sans doute la première visualisation assistée par ordinateur de toute l'histoire des mathématiques. La fig. 5.9(a) présente une spirale de 4000 entiers où les nombres premiers sont marqués par des points noirs ; deux diagonales particulièrement frappantes y sont mises en couleur. (À titre de comparaison, la fig. 5.9(b) présente la même spirale de 4000



(a) (b)

FIGURE 5.9 – Spirale d'Ulam pour les 4000 premiers entiers, avec en noir (a) les nombres premiers, (b) des entiers impairs choisis au hasard, en même nombre que les premiers entre 1 et 4000

entiers, où ont été marqués en noir non pas les nombres premiers, mais une quantité égale de nombres impairs choisis au hasard.)

L'effet est donc robuste, et il faut poursuivre la réflexion. Un examen plus poussé montre que toutes les diagonales correspondent à des ensembles d'entiers de la forme $4n^2+an+b$. (C'est d'ailleurs aussi le cas des droites horizontales et verticales, dont certaines semblent également concernées par le phénomène ; seulement, elles sont moins visibles sur les figures parce qu'elles sont entrecoupées de nombres pairs et apparaissent donc comme en pointillés.) Parmi ces expressions quadratiques, il y en a qui peuvent se factoriser et produisent donc des diagonales vides de nombres premiers, ce qui peut déjà aider à faire ressortir les diagonales dont le polynôme ne se factorise pas ; mais il semble bien que même parmi ces dernières, certaines soient étonnamment riches en nombres premiers, ce qu'Ulam et ses collègues tâchent de quantifier par quelques statistiques à la fin de leur article [29]. Pour autant que je sache, la question reste aujourd'hui non élucidée, quoiqu'on l'ait reliée à des conjectures plus générales.

Ce qui m'importe ici, c'est que rien dans ce processus interprétatif ne va de soi : je peux comprendre comment tracer la spirale sans savoir à quoi correspondent les alignements que je crois remarquer sur mon diagramme, ni même s'ils correspondent à quoi que ce soit.

29. Voir Stein, Ulam et Wells 1964, p. 520.

5.3 Un élargissement : Humphreys et Vorms

Marion Vorms[30] reprend les idées de Simon et de Kulvicki mais leur donne une extension plus large encore, en les combinant à une problématique apparemment différente, venue du travail du philosophe des sciences Paul Humphreys.

Humphreys (2004) reproche à une bonne part de la philosophie des sciences héritière de l'empirisme logique de se concentrer sur ce que les théories scientifiques affirment et permettent « en principe », et de négliger les limitations de nos méthodes mathématiques, qui « en pratique » contraignent drastiquement les calculs et prédictions possibles[31] ; pourtant, des équations différentielles que nous ne savons pas résoudre n'ont guère d'utilité. Il est par exemple habituel, écrit-il, de considérer que les différentes formulations de la mécanique classique (la formulation de Newton lui-même et ce qu'on appelle aujourd'hui les reformulations lagrangiennes et hamiltoniennes) sont de simples variantes d'une même théorie, parce qu'on peut mathématiquement en démontrer l'équivalence et qu'on sait donc qu'en un sens, elles disent la même chose du monde. Or il peut être très facile de prédire l'évolution d'un système en partant d'équations hamiltoniennes, alors qu'on ne sait pas résoudre les équations lagrangiennes correspondantes et qu'en termes newtoniens, il est pratiquement impossible ne serait-ce que de formuler le problème ; Vorms, qui approfondit cet exemple, le montre en détail[32]. Plus généralement, Humphreys souligne combien l'organisation toute entière des sciences est déterminée par nos limitations computationnelles. Celles-ci expliquent entre autres l'apparente bizarrerie que l'on ne cesse de rencontrer les quelques mêmes modèles mathématiques (comme celui de l'oscillateur harmonique) dans les domaines les plus divers de la physique : c'est en fait parce que ce sont les rares équations différentielles que l'on sache résoudre[33]. Ces analyses lui servent à préparer le terrain pour sa thèse principale, qui est que la généralisation des *simulations* par ordinateur, en permettant de contourner les limitations de nos méthodes mathématiques de résolution, est en passe de transformer profondément la méthode scientifique.

L'idée de Vorms est que l'on peut parler des exemples de Humphreys (équations newtoniennes contre lagrangiennes) dans les mêmes termes que de ceux de Kulvicki (liste de relevés de température contre carte colorée) :

30. Le travail que je vais discuter est issu de sa thèse de doctorat (Vorms 2009b) ; je citerai de préférence les textes publiés Vorms 2009a, 2011, 2012, mais renverrai aussi à l'occasion à sa thèse, qui est souvent plus approfondie.

31. Voir en particulier Humphreys 2004, section 3.12, p. 95–100.

32. Vorms 2011, chap. 1 ou Vorms 2009b, partie I.

33. Pour une analyse plus poussée de cet argument, voir Barberousse et Cyrille Imbert 2013 ou (pour une version en anglais, un peu plus brève) Barberousse et Cyrille Imbert 2014.

> Although they contain, at least partially, the same information, these representations do not convey this information the same way. The different pieces of information they contain are not equally accessible to agents. […] *These different representations do not play the same role in the agent's inferential processes.*[34]

Ces différences computationnelles sur fond d'équivalence informationnelle sont ce que Vorms appelle des différences de « format[35] », choisissant délibérément un terme qui rappelle le modèle informatique initial de Simon (ou plus précisément de Larkin et Simon 1987, qui est l'article qu'elle cite).

Cet élargissement supplémentaire, cependant, creuse encore l'écart avec les idées initiales de Simon. Son extension à la visualisation de données rendait déjà flou le sens des « différences computationnelles », dans la mesure où les opérations permises n'étaient plus bien délimitées. Cette fois, c'est l'idée d'« équivalence informationnelle » qui prend un sens moins précis. Chez Simon, cette équivalence a un sens clair et opératoire : deux représentations sont informationnellement équivalentes si chacune peut être reconstruite à partir de l'autre. Pour les visualisations de la section précédente, ce critère reste plus ou moins respecté, du moins jusqu'à un certain niveau de précision des données. Vorms, en revanche, n'affirme pas que l'on puisse systématiquement traduire les équations newtoniennes d'un système en équations lagrangiennes ou inversement.

C'est qu'à y bien regarder, Vorms combine l'idée d'équivalence informationnelle, empruntée à Simon, avec des définitions d'« information » et de « contenu informationnel » qui sont complètement étrangères à celui-ci :

> [A] piece of information is a proposition that can be object of belief. One can start by defining the informational content of a representation as the set of all the pieces of information that an agent mastering its symbol system could in principle extract from it.[36]

Avant toute chose, ces définitions méritent clarification. À première vue, elles induisent une dissymétrie gênante entre phrases et figures, que Simon a pris soin d'éviter : les phrases

34. « Bien qu'elles contiennent, au moins en partie, les mêmes informations, ces représentations ne les transmettent pas de la même manière. Les différentes informations qu'elles contiennent ne sont pas également accessibles aux agents. […] *Ces différentes représentations ne jouent pas le même rôle dans les processus inférentiels de l'agent.* » (Vorms 2012, p. 257 ; l'auteure souligne.)

35. Sur cette notion qui est introduite par Vorms, voir Vorms 2012 (dont est issu la citation précédente), Vorms 2009b, section 2 ou Vorms 2009b, section 7.2.

36. « [U]ne information est une proposition qui peut faire l'objet d'une croyance. À titre de point de départ, on peut définir le contenu informationnel d'une représentation comme l'ensemble des informations qu'un agent qui maîtrise son système symbolique pourrait en principe en extraire. » (Vorms 2012, p. 260 ; cf. aussi Vorms 2009b, p. 360-361).

auraient un contenu réduit à une seule proposition ou information, de la manière la plus classique ; les figures auraient pour contenu l'ensemble des propositions que l'on peut y lire, en un sens ou un autre. Mais ce n'est sans doute pas cela que Vorms veut dire. Tout d'abord, elle semble employer le terme de proposition en un sens inhabituellement large, qui l'identifie à tout ce qui peut faire l'objet, par exemple, d'une croyance, ce qui n'exclut pas que l'on puisse croire directement en une figure [37]. Surtout, elle défend en fait un inférentialisme radical [38]. Quand elle parle d'« extraire » des informations de représentations, elle est prête à inclure sous cette rubrique une large gamme d'inférences (ce terme aussi étant compris en un sens très large, qui est, pour le coup, similaire à celui que lui donne Simon) : lire une valeur dans l'un des tableaux de températures de Kulvicki peut compter au même titre que faire une déduction simple, ou même résoudre une équation différentielle [39].

À vrai dire, une définition aussi radicale du contenu (qu'on l'appelle informationnel ou non) semble difficilement tenable : cela en fait une notion extrêmement ouverte et vague. Elle mériterait à tout le moins d'être précisée. De toute manière, chez Vorms, elle sert surtout de point de référence pour définir ensuite la notion de contenu « cognitivement accessible », qui dépend du « format » de la représentation et qui est seule pertinente en pratique [40]. Quoi qu'il en soit, la seule chose qui importe pour mon propos est que, quand Vorms dit que deux représentations contiennent les mêmes informations, elle ne pense pas à l'équivalence informationnelle au sens que lui donne Simon : elle veut plutôt dire que les deux représentations ont le même « contenu informationnel » au sens de sa propre définition. Les deux définitions ont, certes, un point commun : toutes deux sont (au moins en

37. « [J]e me contenterai », écrit-elle dans sa thèse, « de définir une information comme ce qui peut faire l'objet d'une attitude propositionnelle (que la proposition en question soit exprimée linguistiquement ou non) » (Vorms 2009b, p. 360).

38. Elle s'appuie en fait, de manière plus ou moins assumée (cf. Vorms 2009b, chap. 6, §1.3.2, p. 309–314), sur les sémantiques dites « du rôle conceptuel » (« *conceptual role semantics* ») défendues entre autres par Ned Block et Gilbert Harman (pour une introduction aux versions de cette position que Vorms a en tête, voir M. Greenberg et Harman 2008 ou Block 1998). D'une manière plus générale, c'est une idée que l'on peut faire remonter au slogan wittgensteinien que « le sens, c'est l'usage », *via* en particulier Wilfrid Sellars (cf. surtout Sellars 1963). Il en existe aujourd'hui diverses variantes ; outre les auteurs déjà cités, il faudrait mentionner l'œuvre difficile de Robert Brandom 1994, 2000.

39. Elle écrit en effet : « [I]l serait déraisonnable d'affirmer que la maîtrise du système symbolique dans lequel les équations différentielles fonctionnent n'inclut pas la maîtrise de tout un ensemble de règles de calcul ; si je suis simplement capable, face à des équations newtoniennes, d'en tirer les informations qui y sont explicitement présentées (la valeur des variables représentant la force et la masse, et la relation de proportionnalité entre force et masse), on ne peut pas dire que j'en maîtrise vraiment le système. Si l'on s'en tient à ce qui est explicitement représenté par ces équations, leur contenu est bien pauvre. » (Vorms 2009b, p. 365). Elle semble donc même prête à intégrer les solutions d'une équation différentielle au « contenu informationnel » de celle-ci.

40. Voir Vorms 2012 ou Vorms 2009b, chap. 7, section 2.2, p. 396–406.

partie) motivées par le souhait de parler d'équivalence entre représentations en mettant entre parenthèses la question de savoir ce qui est représenté exactement ; Simon le fait par un critère d'intertraductibilité, Vorms le fait en ne parlant que des inférences que la représentation permet. Au-delà de ce point commun, toutefois, leurs définitions de l'équivalence informationnelle sont différentes et même incompatibles.

5.4 Dirk Schlimm et les notations

Un dernier genre d'exemples que le cadre de Simon est *a priori* bien adapté pour traiter est celui des notations : l'écriture d'une même formule dans deux notations différentes semble un cas paradigmatique d'équivalence informationnelle. C'est effectivement dans ce contexte que l'on trouve certaines des applications les plus claires de la méthodologie de Simon aux mathématiques, à travers les travaux de Dirk Schlimm. Avec Hansjörg Neth [41], il a par exemple comparé l'efficacité computationnelle des chiffres romains et des chiffres arabes, sur la base d'un modèle assez sophistiqué prenant en compte les déplacements attentionnels nécessaires, les ressources requises en mémoire à long terme et en mémoire de travail, etc.

Plus récemment, Schlimm a appliqué une démarche analogue à différentes notations pour la logique propositionnelle : les arbres syntaxiques, les graphes existentiels de Peirce, la notation dite « polonaise » (dans laquelle les constantes logiques comme la conjonction, la disjonction ou l'implication sont placées avant les expressions auxquelles elles s'appliquent), ainsi que les notations de Frege, Peano et Russell [42]. Dans ce second exemple, l'intérêt de la stratégie de Simon est particulièrement net. Schlimm peut en effet montrer l'équivalence « informationnelle » des formules écrites dans ces différentes notations (ou dans des fragments bien choisis de celles-ci) en indiquant comment les traduire les unes dans les autres, sans jamais se poser la question de savoir de quoi elles parlent ; après avoir délimité plusieurs notations qui, en ce sens clair, sont équivalentes, Schlimm peut alors tenter d'en comparer les avantages et les inconvénients, en particulier computationnels.

Dans ces travaux, le critère d'intertraductibilité joue un rôle essentiel. Loin d'être trivial, il permet de distinguer des différences strictement notationnelles et des différences plus substantielles. Par exemple, le logicien anglais du XIX^{e} siècle Augustus De Morgan

41. Schlimm et Neth 2008, Voir.
42. Dans l'ensemble, ce travail n'est pas encore publié, mais Schlimm le présente régulièrement dans des conférences. J'ai pu en entendre des versions au congrès de l'APMP à Paris le 4 novembre 2015 et lors d'un atelier à Nancy le 5 avril 2018. Pour certains aspects de ce travail concernant la notation de Frege, voir Schlimm 2018.

propose d'écrire la négation de X par x, en passant simplement de la majuscule à la minuscule ou inversement [43]. Or cette écriture-là ne permet pas d'itérer la négation, et n'est donc pas informationnellement équivalente à nos notations usuelles pour la logique propositionnelle.

5.5 La notation exponentielle de Leibniz

Tournons-nous, pour finir, vers l'étude de cas de la première partie. Face aux changements notationnels introduit par Leibniz (d^3 pour ddd, par exemple, ou d^{-1} pour $\int$, et plus généralement d^n), les travaux que nous venons de voir invitent à se poser les questions suivantes. Y a-t-il équivalence informationnelle, au sens où l'on peut traduire ce qui est écrit dans une notation en l'autre ? Et si oui, y a-t-il néanmoins des différences que l'on peut caractériser computationnellement ?

Commençons par un cas simple. La différentielle quatrième de xy peut s'écrire

$$d^4(xy) = d^0xd^4y + 4d^1xd^3y + 6d^2xd^2y + 4d^3xd^1y + d^0yd^4x$$

en utilisant la notation exponentielle des différences, ou

$$dddd(xy) = xddddy + 4dxdddy + 6ddxddy + 4dddxdy + yddddx$$

sans elle. Cela semble un cas clair d'équivalence informationnelle, parce qu'on peut manifestement passer d'une notation à l'autre de manière réversible. Pourtant, ce changement de notation fait une différence : comme l'écrit Bernoulli en parlant de l'exemple précédent [44], éliminer la notation exponentielle des différences « cache » l'analogie de $d^4(xy)$ avec

$$(x+y)^4 = x^0y^4 + 4x^1y^3 + 6d^2xd^2y + 4d^3xd^1y + d^0yd^4x.$$

En d'autres termes, la notation de Leibniz fait ressortir une régularité qui, sans elle, est moins manifeste. On pourrait en dire autant des cas plus complexes que nous avons étudiés à la section 3.3.

Il est tentant de rapprocher ces exemples de la visualisation de données [45]. D'ailleurs, il y a les mêmes difficultés : on peut certes parler, de manière vague, de différences « computationnelles », mais il serait bien difficile de délimiter exactement ce que l'on peut déceler

43. « Let x, y, z be the contrary names of X, Y, Z [...] » (De Morgan 1847, p. 60). Schlimm a évoqué cet exemple lors de l'une de ses conférences.

44. Cf. *supra*, p. 86.

45. Voir ci-dessus, section 5.2.

facilement dans nos formules. Une complication supplémentaire qui apparaît ici est qu'il s'agit d'un exemple d'analogie, c'est-à-dire que la régularité qui frappe Leibniz est en fait une similarité avec des formules qu'il connaît déjà. Ce qu'une formule peut faire apparaître est donc relatif, non seulement, peut-être, aux spécificités du système visuel humain, mais aussi aux formules préalablement familières.

Plus généralement, j'ai défendu dans la partie précédente [46] que la notation exponentielle de Leibniz n'avait pas véritablement un pouvoir expressif supérieur aux notations qui lui préexistaient.

5.6 Un premier bilan

Dans tous ces travaux, les concepts de Simon jouent avant tout un rôle heuristique. D'abord, l'idée d'équivalence informationnelle fournit un critère clair permettant de comparer des représentations sans avoir à se demander ce qu'elles représentent au juste. Une fois que deux représentations ont été identifiées comme informationnellement équivalentes, l'idée de différences computationnelles sert ensuite de vocabulaire pour regrouper les diverses manières dont elles peuvent pourtant rendre la résolution de certains problèmes plus ou moins facile ou difficile.

Les concepts de Simon, toutefois, auraient-ils perdu au passage beaucoup de leur précision voire de leur substance ? La différence essentielle est que les auteurs discutés ici reprennent certes la terminologie de Simon, mais pas du tout sa méthodologie. Cela introduit un décalage crucial.

Pour le voir, revenons d'abord à l'article de Larkin et Simon (1987) discuté plus haut [47]. Fondamentalement, ce que ceux-ci proposent, ce sont des *simulations informatiques* de processus de résolution de problème. Ils appellent alors « représentations » non pas des figures concrètes dessinées sur du papier, mais les composants de leur programme qui sont censés leur correspondre (à savoir les structures de données utilisées, les opérations qui agissent sur celles-ci et un mécanisme attentionnel associé [48]). Se placer directement sur ce terrain informatique leur permet de donner aux notions d'équivalence informationnelle et de différences computationnelles un sens rigoureux.

En contrepartie de cette rigueur, l'usage que Larkin et Simon font du terme de représentation n'est, en un sens, pas très intuitif. Par exemple, lorsqu'ils modélisent des procédures

46. Voir en particulier section 3.2.
47. Cf. *supra*, section 4.5.
48. Voir p. 120 ci-dessus.

de résolution d'un problème de géométrie élémentaire [49], ce qu'ils appellent représentation n'est pas un ensemble de marques sur du papier ; c'est plutôt le rôle fonctionnel que les éléments du programme correspondant à ces marques jouent dans la résolution du problème. Or si, comme il semble, la figure joue un rôle différent pour un collégien et pour un mathématicien expérimenté, il faudra modéliser leurs démarches respectives de résolution par des programmes différents et donc considérer qu'on a affaire à des représentations différentes. Cela peut sembler contredire la notion intuitive de représentation, d'après laquelle la figure serait la même pour le collégien et pour l'expert : les marques sur le papier ne sont-elles pas identiques ? Tant pis, répondraient sans doute Larkin et Simon, pour la notion intuitive.

Au contraire de Larkin et Simon, les auteurs que nous venons d'étudier s'épargnent le détour par des simulations informatiques : ils appliquent les concepts de Simon *directement* aux représentations externes que l'on rencontre dans la pratique scientifique ou mathématique. Dans le modèle computationnel qu'ils ont en tête, la représentation est la figure concrète tracée sur du papier, et le scientifique qui l'utilise joue le rôle du programme qui interagit avec la représentation. C'est un point de vue plus naïf que celui de Simon, qui, comme nous l'avons vu, fait passer la simulation informatique avant tout [50]. De ce fait, ils sont beaucoup plus proches de la notion intuitive de représentation, mais fatalement, les concepts de Simon prennent alors un sens plus vague. Trop vague ? Toute la question est de savoir s'il est malgré tout possible de les adapter rigoureusement à ce nouvel usage. C'est l'objet du prochain chapitre.

---

49. Voir section 4.5.b).
50. Voir ci-dessus, section 4.5.c).

Chapitre 6

De Simon à la pratique des mathématiques

Les concepts de Simon peuvent-ils nous aider à comprendre les représentations utilisées en mathématiques ? Le chapitre précédent a ouvert de nombreuses pistes, mais soulève un problème. Prenons par exemple l'idée d'équivalence informationnelle. Les auteurs que nous avons discutés l'appliquent directement à des graphes, figures ou formules, sans passer par les modèles informatiques qui servaient chez Simon d'intermédiaire et lui donnaient un sens précis. Peut-on malgré tout l'appliquer rigoureusement, et sous quelles conditions ? Les mêmes questions se posent pour l'idée de différences computationnelles. Mon premier objectif ici est d'apporter les clarifications requises. Celles-ci me permettront ensuite de montrer que, malgré leur intérêt dans certains cas, les idées de Simon sont insuffisantes pour beaucoup des différences représentationnelles que l'on rencontre en mathématiques.

6.1 Un prérequis sémiologique de la notion d'information

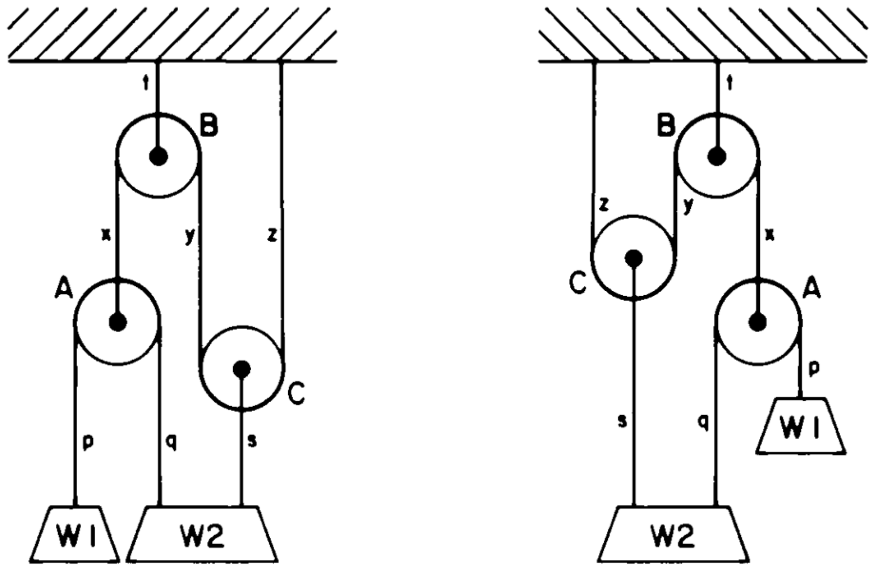
Commençons par le concept d'équivalence informationnelle. Dans le cadre informatique initial de Simon, il est clair : deux structures de données A et B sont informationnellement équivalentes si l'on a un algorithme de conversion qui permet de passer d'une instance a de A à une instance b de B, et un autre en sens inverse. De surcroît, ce critère semble parfaitement syntaxique, c'est-à-dire qu'il ne fait aucune référence à ce sur quoi les données portent.

Les choses se compliquent, toutefois, lorsque l'on veut passer aux représentations externes utilisées en mathématiques ou en sciences. L'idée même d'un critère syntaxique d'équivalence entre représentations suppose que celles-ci soient, en un certain sens, munies d'une syntaxe. Plus précisément, il faut que l'on sache distinguer quels aspects des représentations concrètes que l'on examine sont pertinents informationnellement, c'est-à-dire que l'on sache quelles représentations on peut considérer comme « (essentiellement) les mêmes ». C'est ce que je voudrais montrer ici.

a) Une distinction préalable *type-token*

Revenons pour commencer au cas informatique. Supposons que j'aie une structure de données A qui décrit comment encoder en mémoire une liste alphabétique de dossiers d'étudiants, et une structure de données B qui fait de même pour une liste triée par numéro d'étudiant. J'ai des algorithmes qui permettent de convertir une liste alphabétique donnée a en la liste par numéro correspondante b et vice versa. Dans la mesure où je peux toujours reconstruire a à partir de b et inversement, on peut alors parler d'équivalence informationnelle. Ce que je veux souligner ici, c'est que a et b, en tant que valeurs d'entrée et de sortie sur lesquelles agissent ces algorithmes, *sont déjà des entités abstraites*, pas des réalités physiques particulières. En pratique, si on transforme une liste alphabétique stockée dans un serveur de Paris 1 en une liste par numéro avant de refaire l'opération inverse, la liste alphabétique d'arrivée ne sera sans doute pas la même, *au niveau matériel*, que la liste de départ (elle sera par exemple très probablement à un autre endroit de la mémoire de l'ordinateur). Tout l'intérêt d'une description « fonctionnelle » est précisément qu'on peut faire abstraction de ce genre de différences-là, et donc affirmer que la liste alphabétique peut être reconstruite à partir de la liste par numéro et inversement.

Examinons, au contraire, le problème des poulies étudié par Larkin et Simon (1987). En quel sens la fig. 6.1(a) est-elle informationnellement équivalente à une description (« Le poids W1 est suspendu à une corde pendant de la poulie A dont l'autre extrémité soutient partiellement le poids W2, etc. ») ? Sans doute puis-je, à partir de la figure, produire une description semblable à celle que donnent Larkin et Simon. Mais si ensuite, sur la seule base de cette description, vous traciez une figure, elle serait bien différente de celle de départ. Non seulement ces deux figures seraient des objets concrets différents, tracés avec des stylos différents, etc. mais il se pourrait même que vous ayiez mis le poids W1 à droite de W2 plutôt qu'à gauche, par exemple, ou les poulies à des hauteurs différentes (fig. 6.1(b)). Bien sûr, ces différences sont inessentielles pour résoudre le problème. Encore faut-il le savoir. On ne peut qualifier figure et description d'« informationnellement équivalentes » que relativement au choix préalable des figures que l'on considère comme les mêmes. Or ce choix est un problème substantiel (en l'occurence, il dépend entre autres de notre connais-



(a) (b)

FIGURE 6.1 – Deux figures pour le même problème de statique (la figure (a) vient de Larkin et Simon 1987, p. 73)

sance d'arrière-plan qu'une symétrie droite-gauche ne fait pas de différence pour notre problème).

D'emblée, Simon fait disparaître cette difficulté en remplaçant figures et phrases par des structures de données définies de manière rigoureuse ; l'équivalence informationnelle prend alors un sens parfaitement précis. C'est une démarche tout à fait légitime. Il faut bien voir, cependant, qu'elle s'appuie nécessairement sur un choix préalable, celui de la structure de données qu'il utilise pour modéliser, par exemple, les figures. Ce choix, encore une fois, revient à décider quels aspects de la figure sont pertinents et donc quelles figures on peut traiter comme essentiellement les mêmes. Il ne faut pas s'imaginer que cette question-là peut être résolue de manière purement syntaxique, en évacuant toute réflexion sur ce que la figure est censée représenter.

En d'autres termes, pour pouvoir parler d'équivalence informationnelle entre deux formes de représentation, on a besoin d'une distinction *type-token* pour chacune. Dans une terminologie différente, le philosophe américain Nelson Goodman conduit une analyse très semblable dans son ouvrage *Langages de l'art*, dans le contexte d'un problème apparenté : quelles conditions doit remplir un système de notation de performances de musique ou de danse pour pouvoir adéquatement identifier une même œuvre à travers ses réalisations successives ? Il faut au moins, écrit Goodman, que l'on sache quand deux partitions peuvent être considérées comme les mêmes. Le reste de son travail montre bien que ce genre de considérations ne sont pas limitées aux suites de symboles des langages formels [1].

b) Un sens minimal du terme d'information

Sous quelles conditions peut-on parler de l'« information » que porte une représentation ? La leçon de ce qui précède, c'est qu'on a au moins besoin de savoir quels aspects de la représentation sont pertinents, c'est-à-dire au fond de savoir quelles représentations traiter comme les mêmes. On peut aller plus loin : cette seule condition est non seulement nécessaire, mais même suffisante pour parler, comme Simon, comme Simon, de « la même », de « plus » ou de « moins » d'information, en un sens du terme que l'on pourrai dire *minimal*. On peut ainsi parler d'information sans donner à notre représentation une véritable sémantique.

Cette remarque vaut, d'une manière générale, pour tout emploi du terme qui s'inspire de près ou de loin de la stratégie de Claude Shannon [2]. On peut effectivement parler d'information sans introduire de sémantique, mais il est pour cela indispensable de bien saisir le *type* dont la représentation qui nous intéresse est un *token*, et plus largement l'ensemble des autres *types* qui seraient possibles. Comme l'écrit Shannon,

> [The] semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. [3]

Cela peut créer beaucoup de confusion que de négliger ce point, et donc de présupposer que l'information que porte une représentation en est une propriété intrinsèque.

Ken Manders, dans un article peu connu, diagnostique bien ce problème en préalable d'une discussion sur l'emploi de figures en mathématiques :

1. Voir Goodman 1968, en part. chap. IV, section 2. Il parle d'*inscriptions* (qui peuvent être visuelles, auditives, etc.) et de *caractères*, qui sont des « classes » d'inscriptions ; il parle de *schème notationnel* lorsqu'inscriptions et caractères sont tels que chaque inscription peut être assignée de manière unique à un caractère, c'est-à-dire lorsqu'inscriptions et caractères jouent respectivement le rôle de nos *tokens* et *types*. Il s'emploie alors à analyser plus exactement ce qui est requis pour qu'on ait un schème notationnel : il faut que les caractères forment des classes disjointes, mais aussi qu'on puisse déterminer par une inspection matérielle finie à quel caractère une inscription donnée correspond. Pour une présentation plus approfondie, voir aussi Kulvicki 2006, chap. 1 ou encore Vorms 2009b, section 7.1.

2. Voir l'introduction de cette partie, en part. p. 97–98.

3. « Les aspects sémantiques de la communication ne sont pas pertinents pour le problème d'ingénierie. L'aspect important est que le message effectivement transmis n'est qu'une possibilité *sélectionnée dans un ensemble* de messages possibles. » (Shannon 1948, p. 379).

> The notion of representational content is a difficult one, for many reasons. Many of its problems are inherited by contemporary notions of information. […]
>
> For the term 'information', [a] systematic ambiguity is abetted by the communication engineer's quantitative conception of information […], which succeeded marvellously in separating engineering concerns from those of the user. […]
>
> Genuine application of the Shannon theory requires individuating the events of interest, to which the statistical measures are applied, at each stage of the communication channel. On the user side, for example, telephone carriers put major research into picking out components of a voice signal most important to understanding transmitted speech […].
>
> Such specialized tasks figure little in an outsider's conception of the Shannon theory, leaving the mis-impression that individuation of events in statistical ensembles at various stages of the channel is intrinsically given, and that information in the Shannon sense can be unproblematically indentified between various stages of a channel. [4]

En d'autres termes, toute mesure de l'information au sens de Shannon suppose que l'on sache identifier les signaux concrets que l'on traite comme des instances de l'un des signaux possibles [5].

---

4. « La notion de contenu représentationnel est délicate, et ce pour de nombreuses raisons. Les notions contemporaines d'information héritent de beaucoup de ses difficultés. […] Pour le terme d''information', une ambiguïté systématique est encouragée par la conception quantitative de l'information des ingénieurs en télécommunications […], qui a merveilleusement bien réussi à séparer les questions d'ingénierie de celles qui intéressent l'utilisateur. […] Une authentique application de la théorie de Shannon exige d'individuer les événements d'intérêt auxquels on appliquera les mesures statistiques à chaque étape du canal de communication. Du côté de l'utilisateur, par exemple, les opérateurs de téléphonie consacrent de grands efforts de recherche à identifier quels composants d'un signal vocal sont les plus importants pour comprendre la parole retransmise […]. De telles tâches spécialisées font rarement partie de la conception que se font les profanes de la théorie de Shannon ; cela laisse l'impression fausse que l'individuation des événements au sein d'ensembles statistiques [de possibilités] à différentes étapes du canal [de communication] est donnée intrinsèquement, et que l'on peut identifier sans problèmes l'information au sens de Shannon présente à différentes étapes d'un canal. » (Manders 1996, p. 389-390).

5. C'est en fait l'hypothèse essentielle sur laquelle repose la théorie mathématique de la communication de Shannon : « [The] semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. » (« Les aspects sémantiques de la communication ne sont pas pertinents pour le problème d'ingénierie. L'aspect important est que le message effectivement transmis n'est qu'une possibilité *sélectionnée dans un ensemble* de messages possibles. » Shannon 1948, p. 379.)

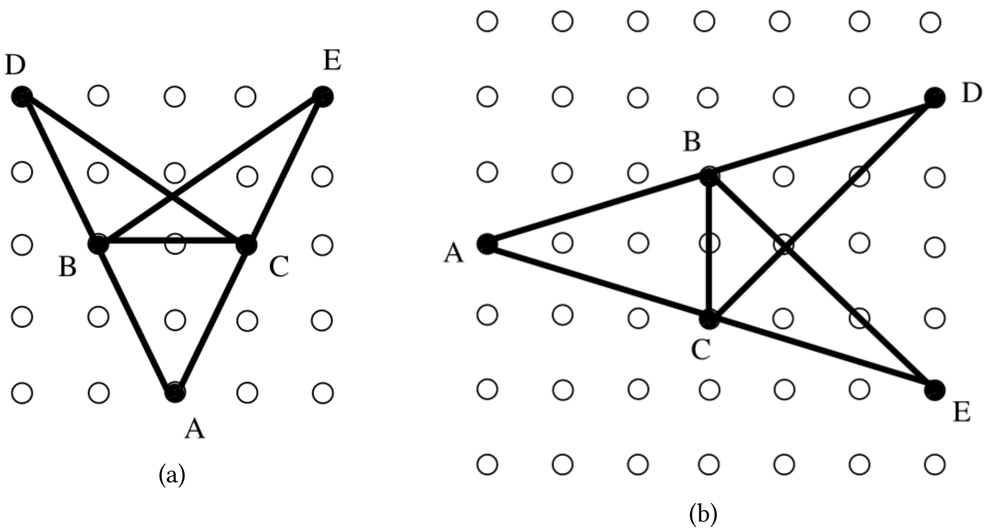


FIGURE 6.2 – Deux « figures étiquetées » équivalentes du système **Eu** de Mumma (Mumma 2006, p. 21)

c) Un exemple : Miller et Mumma sur les figures d'Euclide

De manière révélatrice, c'est précisément en un tel sens minimal que John Mumma et Nathaniel Miller emploient le terme d'information dans leurs études sur la géométrie d'Euclide [6].

Tous deux proposent des systèmes formels (respectivement nommés **Eu** et **FG**) destinés à rendre compte des démonstrations d'Euclide, figures comprises ; tous deux commencent donc par expliquer comment ils encodent les figures. Mumma, par exemple, part de grilles de n sur n cases ($n \geq 1$) dont certaines positions marquées désignent des points, des segments, des demi-droites, des droites et des cercles, comme à la fig. 6.2(a) ; il les appelle *figures étiquetées* (« *labeled diagrams* »). Or juste après avoir les définies, il écrit :

> [L]abeled diagrams carry too much information. For any one labeled diagram there will be an infinite number of others which [...] support the same inferences in a Euclidean proof. For example the differences between the two labeled diagrams [of fig. 6.2] would have no bearing on what one could do with them in a Euclidean proof. [...] To group such diagrams together, **Eu** possesses an equivalence relation ~ which abstracts away all irrelevant information. [7]

6. Cf. Miller 2001, 2007 et Mumma 2006. J'aurai l'occasion de revenir sur ces travaux à plusieurs reprises dans la suite de ce travail ; pour une présentation générale, voir *infra*, p. 220–221. Plus largement, cf. chap. 8 et 12.

7. « Les figures étiquetées portent trop d'informations. Pour chaque figure étiquetée, il y en aura une infi-

Mumma ne donne pas de véritable sémantique à ses figures [8] ; isoler les « informations » pertinentes revient ici uniquement à indiquer quelles figures on peut considérer comme les mêmes.

Au départ, Miller définit ses figures autrement ; mais à la même étape de son développement, il s'exprime en des termes très proches :

> We have now defined a primitive diagram to be a particular kind of geometric object. These diagrams contain somewhat too much information, though. [...] As it stands, if we have a diagram containing only a single dot, that dot could be at any one of an infinite number of points [...]. We don't actually care exactly where the point is, however. [...] Instead, we're only going to use our diagrams to show the topology of how lines and circles might lie in the plane. For example, all of the diagrams that only contain a single dot have the same topology, even though each one is slightly different, so we'd like to consider them to all be the same diagram in some sense. In other words, we'd really like to look at equivalence classes of diagrams that contain the same topological information. [9]

Contrairement à Mumma, Miller donne une sémantique à ses figures, mais elle joue un rôle limité dans son système et, à vrai dire, elle dépend elle-même de l'équivalence entre figures, ce qui lui donne un goût de circularité : en substance, un modèle d'une figure F est une configuration du plan euclidien qui, si elle était utilisée comme diagramme, serait équivalente à F [10]. On le voit bien, la notion cruciale est celle d'équivalence entre figures.

En l'espèce, tant Miller que Mumma (quoique ce dernier soit plus clair sur ce point) choisissent de considérer comme équivalentes informationnellement les figures que l'on

nité d'autres qui [...] permettent les mêmes inférences dans une démonstration Euclidienne. Par exemple, les différences entre les deux figures étiquetées [de la fig. 6.2] n'auraient pas d'influence sur ce que l'on pourrait faire avec eux dans une démonstration Euclidienne. [...] Pour grouper de tels diagrammes ensemble, **Eu** possède une relation d'équivalence ~ qui permet de faire abstraction de toutes les informations non pertinentes. » (Mumma 2006, p. 20-21).

8. Du moins dans la version originale de son système ; pour une version ultérieure incluant une sémantique, développée à la suite des commentaires de Miller 2012, voir Mumma 2014.

9. « Selon la définition que nous venons de donner, une figure primitive est un objet géométrique d'une certaine sorte. Cependant, ces figures contiennent un peu trop d'information. [...] En l'état, si nous avons une figure qui ne contient qu'un unique point, ce point pourrait se trouver à une infinité de positions différentes [...]. La position exacte du point, cependant, n'a pas d'importance pour nous. [...] Au contraire, nous n'utiliserons nos figures que pour montrer la topologie de la manière dont des droites et des cercles pourraient être disposées dans le plan. Par exemple, toutes les figures qui ne contiennent qu'un unique point ont la même topologie, même si elles sont toutes un peu différentes ; en un certain sens, nous aimerions donc toutes les considérer comme la même figure. Autrement dit, ce que nous voudrions en réalité, c'est considérer les classes d'équivalence de figures qui contiennent les mêmes informations topologiques. » (Miller 2007, p. 27).

10. Cf. Miller 2007, section 2.3, p. 31–33.

pourrait substituer sans faire de différences pour les démonstrations. Ce choix fait écho à la célèbre définition du « contenu conceptuel » que donne Frege dans sa *Begriffsschrift* :

> [D]ie Inhalte von zwei Urtheilen [können] in doppelter Weise verschieden [sein]: erstens so, dass die Folgerungen, die aus dem einen in Verbindung mit bestimmten andern gezogen werden können, immer auch aus dem zweiten in Verbindung mit denselben andern Urtheilen folgen; zweitens so, dass dies nicht der Fall ist. Die beiden Sätze: „bei Plataeae siegten die Griechen über die Perser" and „bei Plataeae wurden die Perser von den Griechen besiegt" unterscheiden sich in der erstern Weise. [...] Ich nenne nun denjenigen Theil des Inhaltes, der in beiden *derselbe* ist, den *begrifflichen Inhalt*. [11]

Le point de départ de Miller et Mumma est donc la pratique inférentielle qu'est la géométrie euclidienne élémentaire, ou plutôt certaines hypothèses descriptives sur la manière dont les figures y interviennent dans les preuves. Quand ils parlent d'élucider l'« information » présente dans les figures, ce qu'ils font est ni plus ni moins qu'identifier les figures que l'on peut considérer comme identiques sans dommage pour les démonstrations.

6.2 Équivalence informationnelle et intertraductibilité algorithmique

Un autre aspect de l'équivalence informationnelle, qui va de soi dans le cas informatique mais requiert prudence au-delà, est qu'elle doit être algorithmique [12]. Ce point est essentiel parce que, pour mon propos, l'intérêt principal de l'idée d'équivalence informationnelle est d'offrir un critère précis et opératoire.

Le chapitre précédent fournit un exemple dans lequel ce critère ne semble pas rempli : c'est celui de l'équivalence entre équations newtoniennes, lagrangiennes ou encore hamil-

11. « [L]es contenus de deux jugements peuvent différer de deux manières : selon la première, les conclusions qui peuvent être tirées à partir de l'un d'eux en relation avec d'autres déterminés découlent toujours du second en relation avec ces mêmes autres jugements ; selon la deuxième, ce n'est pas le cas. Les deux propositions : "les Grecs ont vaincu les Perses à Platée" et "les Perses ont été vaincus par les Grecs à Platée" diffèrent suivant la première manière. [...] J'appelle alors *contenu conceptuel*, cette partie du contenu qui est *la même* dans les deux. » (Frege 1879, p. 2-3 ; l'auteur souligne. Trad. fr. adaptée de Frege 1999, p. 16-17).

12. Un argument intéressant et apparenté se trouve chez Haugeland [1991] 1998c. D'après lui, toute l'idée qu'images et phrases seraient différentes manières de représenter un même contenu (nous dirions ici, à la suite de Simon, les mêmes informations) est une erreur. Or, il considère que la source de cette erreur est précisément une confusion entre des processus purement mécaniques comme ceux qui permettent de transcrire une image en liste de pixels (donc avec identité de contenu) et des processus beaucoup plus exigeants comme ceux qui permettent de *décrire* une scène en phrases (sans identité de contenu).

toniennes pour un même système, dont parle Marion Vorms[13]. Pour rappel, Vorms part du constat de Humphreys que, si ces équations sont équivalentes « en principe », elles ne le sont pas du tout « en pratique » : résoudre les équations différentielles qui gouvernent l'évolution d'un système peut par exemple être très difficile dans un cadre newtonien mais très facile dans un cadre lagrangien. Sa reformulation de ce contraste est que ces équations ont le même « contenu informationnel », mais un « contenu cognitivement accessible » très différent, c'est-à-dire qu'il y a entre elles une différence radicale de « format ». Mon propos n'est pas de critiquer ce contraste général, qui me semble convaincant, ni de contester qu'il y a un sens en lequel les formulations de la mécanique classique sont équivalentes : on peut effectivement démontrer que des équations sous ces différentes formes pour un même système prédiront la même évolution. En outre, il est clair que Vorms a tout intérêt à définir sa notion initiale d'équivalence de la manière la plus large possible, puisque son but est de montrer que cette équivalence n'a d'intérêt que théorique. Mes objectifs, cependant, sont différents, et pour exigent de distinguer la notion d'équivalence qu'utilise Vorms de celle de Simon.

Or, dans l'exemple de Vorms, on n'a pas équivalence informationnelle au sens de Simon. Sans doute n'existe-t-il pas de méthode pleinement générale pour passer d'équations lagrangiennes à des équations newtoniennes du même problème. En tout cas, ce passage est souvent délicat et n'a rien d'automatique ; c'est même l'un des arguments principaux qu'utilise Vorms pour affirmer que ces deux formulations sont différentes en pratique[14]. L'identité de « contenu informationnel » au sens de Vorms est donc une notion nettement plus large que celle d'équivalence informationnelle. (En réalité, et malgré la terminologie informationnelle et la référence à Simon, il s'agit plutôt d'une identité sémantique comme nous en verrons dans la partie suivante.)

Pour clarifier le problème, prenons un exemple mathématique apparenté à ceux de Vorms, mais plus simple. Une même courbe plane, disons un même cercle, peut être représenté de diverses manières : par exemple en coordonnées cartésiennes, comme l'ensemble des points $(x, y) \in \mathbf{R}^2$ tels que $x^2 + y^2 = 1$; en coordonnées polaires, par l'équation $\rho(\theta) = 1$; ou encore de manière paramétrique, comme l'image de la fonction

$$\begin{array}{rcl} \mathbf{R} & \to & \mathbf{R}^2 \\ t & \mapsto & (\cos t, \sin t). \end{array}$$

13. Cf. *supra*, section 5.3.

14. Voir Vorms 2011, p. 24 *sq.* Elle écrit par exemple que pour certains systèmes, « l'étape de la mise en équation [est] pratiquement impossible » dans un cadre newtonien, alors qu'écrire des équations lagrangiennes est très facile.

On peut démontrer que ces représentations définissent la même courbe plane, mais elles peuvent en rendre différentes propriétés plus ou moins faciles d'accès. Vorms dirait là encore que l'on a le même contenu informationnel, mais une différence de format.

Confrontons maintenant cet exemple à celui de Kulvicki, qui compare deux manières de présenter des relevés de température : sous la forme d'une liste, ou disposés sur une grille bidimensionnelle [15]. Intuitivement, on a bien là les mêmes données sous des formes différentes, ou organisées différemment. En revanche, il est tentant d'aborder tout autrement l'exemple des équations du cercle, en s'inspirant plutôt des concepts usuels de la philosophie du langage : il est naturel de voir ces différentes équations comme des descriptions ayant, dans la terminologie de Frege, la même *référence* mais des *sens* différents. Que l'on accepte ou non cette intuition (que la prochaine partie nous donnera des outils pour préciser), ce qui importe ici est de prendre la mesure du problème : on a un fouillis de situations dans lesquelles il est tentant de dire de manière vague que deux représentations « représentent la même chose », « contiennent les mêmes informations », etc. et il n'est pas du tout clair qu'il faille les mettre sur le même plan.

C'est précisément là que le critère de Simon, s'il est utilisé avec rigueur, peut permettre de progresser. Il invite à se poser des questions précises : peut-on délimiter deux classes de représentations entre lesquelles on peut définir des algorithmes de traduction dans chaque sens ? De ce point de vue, beaucoup d'équivalences apparentes s'avèrent problématiques. Ainsi, même la conversion pourtant simple entre coordonnées polaires aux coordonnées cartésiennes n'est pas entièrement triviale [16] ; dans beaucoup de cas plus complexes, par exemple le passage d'une équation en x et y à une équation paramétrique ou inversement, il n'y a pas du tout de méthode générale. À l'inverse, ce critère peut légitimer l'idée qu'il y a effectivement des différences purement notationnelles. Tout cela, cependant, requiert de donner un sens précis à l'intertraductibilité requise par la définition de Simon.

6.3 Différences computationnelles et opérations

L'idée de différences computationnelle est plus délicate, et nous place face à un dilemme. Soit on conserve la méthode de Simon et on propose des modèles informatiques ; alors, comme nous l'avons dans le cas de Larkin et Simon [17], on ne capturera pas la représentation elle-même mais plutôt son rôle fonctionnel dans un contexte particulier. Soit,

15. Voir *supra*, section 5.2.

16. Les (petites, mais réelles) difficultés sont liées au fait qu'il y plusieurs choix possibles pour exprimer l'angle θ en fonction de x et y.

17. Voir *supra*, section 4.5.

comme Kulvicki, on se retrouve contraint d'utiliser le terme en un sens très large.

Commençons par la première possibilité. Pour rendre plus clair la difficulté décelé chez Larkin et Simon, prenons un autre exemple, celui du système formel *E* d'Avigad, Dean et Mumma (2009) pour les *Éléments* d'Euclide. À première vue, leur but est différent de celui de Simon : ils ne cherchent pas à simuler informatiquement des procédures de résolution de problème, mais seulement à formaliser des démonstrations. Il y a pourtant une parenté. Les démonstrations d'Euclide s'appuient sur des figures, et *E* vise à capturer leur rôle dans les démonstrations. Certains éléments d'une preuve de *E*, à savoir les « assertions diagrammatiques [18] », sont censés correspondre à la figure. Mais, tout comme chez Larkin et Simon, ces assertions, loin de capturer la figure elle-même, en capturent seulement un certain rôle. Vis-à-vis des figures utilisées par Euclide, il y a en fait un double décalage.

Tout d'abord, les assertions diagrammatiques d'une preuve de *E* ne peuvent pas être reconstruites simplement en examinant la figure. Comme l'écrivent les auteurs,

> in *E*, the diagram *is* nothing more than the collection of generally valid diagrammatic features [19] that are *guaranteed by the construction.* [20]

Les assertions diagrammatiques qui apparaissent dans une preuve dépendent donc aussi du texte, qui indique comment la figure a été construite. J'y reviendrai plus en détail au chapitre 11 (section 11.2).

Le second décalage est que ces assertions diagrammatiques ne peuvent pas rendre compte de tous les usages possibles de la figure. J'ai repris au chapitre précédent un exemple de Netz, tiré d'Aristarque, dans lequel la figure permet d'observer une symétrie [21]. Certes, on peut considérer qu'il y a, à cet endroit de la démonstration, un argument implicite dont *E* pourrait parfaitement rendre compte (bien que ce système n'ait jamais prétendu traiter Aristarque !). Mais précisément, la manière dont la figure permet à Aristarque son raccourci (en rendant une symétrie manifeste) n'apparaît pas dans *E*. Encore une fois, ce que *E* capture, de même que les systèmes de Larkin et Simon, c'est un *rôle fonctionnel particulier* des figures.

Bien sûr, cela n'empêche pas *E* d'accomplir sa fonction, qui est de formaliser les preuves d'Euclide sans trop s'éloigner du texte d'origine. Cela ne condamne pas non plus une stratégie comme celle de Larkin et Simon. Les pistes ébauchées par Jessica Carter ou encore

18. Voir Avigad, Dean et Mumma 2009, p. 710

19. Ce que nos auteurs appellent ici « *diagrammatic features* », ce sont, dans la terminologie de Manders 2008b, les « attributions co-exactes » qui seraient permises par la figure. Pour plus de détails, voir *infra*, chap. 8.

20. « [D]ans *E*, la figure n'*est* rien de plus que la collection des caractéristiques diagrammatiques généralement valables qui sont garanties par sa construction. » (Avigad, Dean et Mumma 2009, p. 706).

21. Cf. p. 135–136.

Dirk Schlimm que j'ai évoquées au chapitre précédent [22] montrent que l'on peut poursuivre dans cette voie avec profit. Il faut toutefois garder conscience des limites intrinsèques d'une telle approche.

Examinons maintenant l'autre possibilité. Si l'on abandonne la méthodologie de Simon et qu'on veut appliquer ses concepts directement à des représentations externes, sans simulation ou modèle informatique, l'expression de « différences computationnelles » est condamnée à recouvrir n'importe quelle différence que l'on pourra déceler entre représentations informationnellement équivalentes. Les exemples de visualisations de données de la section 5.2 le montrent bien : la tâche de délimiter tout ce qui peut apparaître sur une représentation semble désespérée. En réalité, on ne fait que nommer la difficulté, qui est que deux représentations équivalentes au sens clair que nous fournit Simon peuvent néanmoins faciliter des observations différentes à notre système visuel.

6.4 Le vrai problème : une équivalence avec quoi ?

Ces clarifications en main, venons-en au vrai problème : dans les cas les plus intéressants de différences représentationnelles, on pourrait sans doute parler d'équivalence informationnelle, mais ce ne serait pas très éclairant.

Pour commencer, tournons-nous vers une difficulté qui frappe déjà le tout premier exemple de Simon. Au début de son article de 1978, lorsqu'il compare l'énoncé en phrases d'un problème de géométrie et la fig. 4.2 (p. 109), il écrit que la figure nous permet d'obtenir plus facilement la solution, qui est que les droites EF et AC se coupent à l'intérieur du rectangle ABCD. Simon tient pour acquise l'équivalence informationnelle entre les deux, mais est-elle si évidente ? Admettons avec lui que l'on puisse construire la figure à partir de l'énoncé du problème. L'autre direction en revanche pose un problème inattendu : si l'on traduit naïvement la figure en phrases, la traduction comprendra non seulement l'énoncé, mais aussi… la solution du problème, c'est-à-dire l'existence et la position du point d'intersection de EF et AC. Avec quel ensemble de phrases la figure est-elle donc « informationnellement équivalente » ?

Peut-être cet exemple, que Simon n'utilise que pour motiver initialement ses concepts, est-il trop vague. En voici un autre plus précis, sur lequel je reviendrai dans la troisième partie. Soient trois ensembles A, B et C tels que $A \setminus C = \varnothing$ et $C \cap (A \cup B) = \varnothing$. Je peux les représenter par un diagramme de Venn (fig. 6.3), c'est-à-dire représenter chaque ensemble par une courbe fermée avec la convention que les régions grisées correspondent à des

22. Voir respectivement sections 5.1 et 5.4.

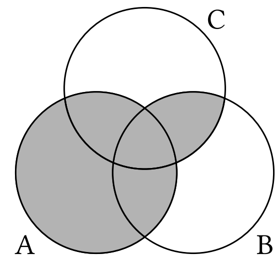


FIGURE 6.3 – Diagramme de Venn pour trois ensembles A, B et C tels que $A \setminus C = \varnothing$ et $C \cap (A \cup B) = \varnothing$

ensembles vides[23]. Ce diagramme peut ensuite m'aider à voir certaines *conséquences* de mes hypothèses, par exemple que $A = \varnothing$.

L'avantage qu'offre ce diagramme peut-il être décrit dans les termes de Simon ? Il faut d'abord se demander en quel sens le diagramme est équivalent aux énoncés. Or si les deux énoncés dont je suis parti me permettent bien de construire le diagramme, le sens inverse pose problème : une traduction systématique des diagrammes de Venn en énoncés en fournira davantage, par exemple $A = \varnothing$ ou $A \setminus B = \varnothing$. Bien sûr, avec quelques efforts, nous pourrions concocter un algorithme capable de retrouver exactement nos énoncés de départ *dans ce cas particulier*, mais il serait nécessairement très artificiel. Pour le voir, remarquons par exemple que B et C jouent des rôles symétriques : le même diagramme aurait donc pu être construit à partir des prémisses $A \setminus B = \varnothing$ et $B \cap (A \cup C) = \varnothing$. Pour retrouver spécifiquement les énoncés de départ $A \setminus C = \varnothing$ et $C \cap (A \cup B) = \varnothing$ plutôt que $A \setminus B = \varnothing$ et $B \cap (A \cup C) = \varnothing$, il faudrait donc un algorithme qui ne traite pas les différents ensembles représentés de manière symétrique. C'est dire à quel point il faudrait tricher pour retrouver exactement certaines hypothèses particulières.

Ces deux exemples soulignent le même problème. Pour parler d'équivalence informationnelle, il ne suffit pas de dire de manière vague que l'on peut traduire telle figure en phrases ou telles phrases en une figure. Il faut pouvoir préciser *quel* ensemble de phrases, ou *quelle* figure, et pouvoir donner des algorithmes permettant d'accomplir cette traduction dans les deux sens de manière systématique. Or si l'on se livre soigneusement à cet exercice, on s'aperçoit que l'avantage qu'apportent la figure dans l'exemple de Simon, ou le diagramme de Venn dans le mien, ne se laissent pas formuler en termes de différences computationnelles sur fond d'équivalence informationnelle.

Une difficulté apparentée se manifeste dans un autre genre d'exemples. Revenons au changement notationnel de Leibniz étudié dans la première partie. Au chapitre précédent,

23. Pour plus de détails sur les diagrammes de Venn, voir *infra*, section 7.1.b).

j'ai tenté de le reformuler en termes d'équivalence informationnelle [24]. J'ai été délibérément vague ; soyons maintenant plus précis. Il est clair que l'on pourrait passer algorithmiquement de formules comme

$$dddd(xy) = xddddy + 4dxdddy + 6ddxddy + 4dddxdy + yddddx$$

à des formules comme

$$d^4(xy) = d^0xd^4y + 4d^1xd^3y + 6d^2xd^2y + 4d^3xd^1y + d^0yd^4x$$

et inversement. Plus précisément, si l'on restreint la notation $d^\bullet$ à des exposants entiers finis (positifs ou négatifs), comme d^2, d^3, d^{-1}, d^{-2}, etc., on peut définir un ensemble de formules possibles écrites en employant la nouvelle notation, un ensemble de formules possibles ne l'employant pas, et un algorithme de traduction simple entre les deux.

Même si l'on autorise des exposants indéterminés, comme d^n, la notation de Leibniz reste éliminable ; c'est du moins ce que j'ai soutenu au chapitre 3 [25]. Certes, c'est plus délicat, parce que cela implique de remplacer une formule générale par plusieurs cas particuliers. Par exemple, au lieu d'écrire, à la manière de Leibniz,

$$d^e(xy) = d^exd^0y + \frac{e}{1}d^{e-1}x \cdot d^1y + \frac{e \cdot e-1}{1 \cdot 2}d^{e-2}x \cdot d^2y + \frac{e \cdot e-1 \cdot e-2}{1 \cdot 2 \cdot 3}d^{e-3}x \cdot d^3y \text{ etc.}$$

on peut écrire à la manière d'Euler

$$\begin{aligned} d(xy) &= ydx + xdy \\ dd(xy) &= yddx + 2dxdy + xddy \\ ddd(xy) &= ydddx + 3dyddx + 3ddydx + xdddy \end{aligned}$$

et aussi, pour y ajouter le cas d'exposants négatifs (qu'Euler ne traite pas),

$$\int(xy) = \left(\int x\right)y - \left(\iint x\right) \cdot dy + \left(\iiint x\right) \cdot ddy \text{ etc.}$$

et ainsi de suite. Pour rendre précise l'idée d'une intertraductibilité systématique, il faudrait introduire une syntaxe pour exprimer « le résultat général dont voici des cas particuliers ». Pour pouvoir passer des cas particuliers à la formule générale, il faudrait aussi concevoir un algorithme capable d'inférer une formule générale à partir de quelques exemples. En l'occurence, ce serait certainement délicat à cause des coefficients binomiaux, mais on pourrait

24. Cf. section 5.5.
25. Voir section 3.2.

tricher et aider un peu l'algorithme, puisque, après tout, Euler lui-même accompagne ses exemples d'explications en langue naturelle pour souligner que les coefficients numériques qui apparaissent sont ceux du binôme.

Admettons (ne serait-ce que pour l'argument) que l'on soit parvenu à régler ces difficultés et que l'on puisse donc parler d'équivalence informationnelle. Restent les problèmes que j'ai soulignés à la fin du chapitre 3. Quand on élimine la notation d^n, on se retrouve avec une série de cas particuliers au lieu d'une formule générale. Mais qu'est-ce qui justifie de traiter ces cas particuliers ensemble (section 3.4) ? C'est encore plus clair dans le cas du calcul de Bernoulli (section 3.5). La notation de Leibniz permet à Bernoulli de concevoir une règle que l'on pourrait écrire $(d^k x d^m y)(d^l x d^n y) = d^{k+l} x d^{m+n} y$; sans cette notation, cette règle se décompose en des instances aussi hétéroclites que

$$\frac{dddxdy}{dxddy} = ddx \int y, \quad dddxdy \times dy \int x = ddxddy, \quad \text{ou encore} \quad \frac{dxdy}{dy \int x} = yddx.$$

Sans doute pourrait-on, là encore, les rassembler plus ou moins artificiellement pour maintenir une traductibilité théorique. Mais on voit bien que c'est précisément l'intérêt central de la notation leibnizienne que de les rapprocher. Ce serait donc passer à côté de l'essentiel que de voir, dans ce changement notationnel, une manière de représenter les *mêmes* informations sous une autre forme.

Comme je l'ai suggéré en passant au chapitre précédent [26], on peut penser que les diagrammes commutatifs présentent un problème similaire. Certes, on peut traduire tout diagramme donné en une liste de relations de commutation. Pour autant, la situation n'est pas comparable à l'exemple de visualisation de données de Kulvicki, où l'on a des données de températures délimitées d'avance que l'on peut représenter sous différentes formes. En effet, l'une des fonctions des diagrammes commutatifs peut être, précisément, de nous aider à délimiter quels morphismes et relations de commutation sont intéressants. Il s'agit sans doute d'un phénomène répandu ; on pourrait multiplier les exemples. Pour le traiter adéquatement, il nous faudra toutefois d'autres outils.

26. Cf. section 5.1, p. 132–133.

Conclusion : avancées et limites

Les concepts de Simon nous ont permis de progresser. Tout d'abord, nous avons vu qu'il était possible de parler de manière précise d'information, et d'équivalence informationnelle, sans définir de véritable sémantique. Ensuite, nous avons pu mettre le doigt avec précision sur un phénomène général : des représentations informationnellement équivalentes peuvent néanmoins présenter des différences cruciales, comme le montre l'exemple de la visualisation de données, ou encore celui de l'analogie de Leibniz de la première partie.

Toutefois, nous avons aussi vu que nombre des différences représentationnelles que l'on rencontre en mathématiques ne se laissent pas penser de cette manière ; c'est en particulier le cas de certains des aspects les plus intéressants du changement notationnel de Leibniz auquel est consacrée la première partie. Il nous faut donc poursuivre notre enquête.

Troisième partie

Syntaxe et sémantique des diagrammes

Sommaire de la troisième partie

Introduction : logique et information

Il y a un quart de siècle, Jon Barwise et John Etchemendy ont ébauché un programme radical : refonder la logique sur les notions de *représentation* et d'*information* pour l'élargir aux raisonnements qui emploient des diagrammes. C'est à leur travail, et à celui de leurs élèves, que cette troisième partie est consacrée. Voici comment ils présentent leur projet :

> Valid deductive inference is often described as the extraction or making explicit of information that is only implicit in information already obtained. Modern logic builds on this intuition by modeling inference as a relation between sentences, usually sentences of a formal language like the first-order predicate calculus. [...] But of course language is just one of the many forms in which information can be couched. Visual images, whether in the form of geometrical diagrams, maps, graphs, or visual scenes of real-world situations, are other forms. [27]

Diagrammes, figures, plans, graphes, etc. entrent alors dans la même catégorie que les énoncés du calcul des prédicats, celle de représentation portant de l'information, et la logique doit les traiter de la même manière.

Au premier abord, Barwise, Etchemendy et leurs élèves semblent se contenter d'intégrer les diagrammes à la logique sans transformer celle-ci, en étendant aux diagrammes les concepts usuels de syntaxe et de sémantique développés pour les langages formels. C'est ainsi, par exemple, que procède Sun-Joo Shin (1994) dans un ouvrage devenu emblématique

27. « On dit souvent que l'inférence déductive valide consiste à extraire ou à rendre explicite de l'information qui n'est qu'implicite dans l'information déjà obtenue. La logique moderne développe cette intuition en modélisant l'inférence comme une relation entre des énoncés, d'habitude des énoncés d'un langage formel comme le calcul des prédicats du premier ordre. [...] Mais bien sûr, le langage n'est que l'une des nombreuses formes sous lesquelles on peut exprimer de l'information. Les images visuelles, que ce soit sous la forme de diagrammes géométriques, de plans, de graphes, ou de portraits visuels de situations du monde réel, en sont d'autres. » (Barwise et Etchemendy [1991] 1996b, p. 4).

tique de tout ce programme de recherche, qui me servira d'entrée en matière (chapitre 7) : elle définit, pour certains diagrammes logiques (apparentés aux « diagrammes de Venn »), des règles de formation et de transformation et leur donne une sémantique en termes de modèles.

Cette approche, qui peut paraître simple et naturelle, implique en réalité un important changement de perspective. Elle s'oppose en effet à une position par défaut, qui était au départ celle de nos auteurs, d'après laquelle les diagrammes seraient fondamentalement non discursifs : on aurait d'un côté des suites de raisons données en langue naturelle (ou dans des langages formels conçus comme des enrégimentations de la langue naturelle), qui seraient l'objet de la logique, et de l'autre des diagrammes qui peuvent certes être utiles parce qu'ils nous donnent un accès plus direct aux objets du discours, mais qui n'expriment pas de proposition et ne peuvent faire partie intégrante de raisonnements. Nos auteurs, au contraire, accordent aux diagrammes le même statut qu'aux énoncés, ce qui veut dire qu'on peut avoir d'authentiques *inférences* entre diagrammes, ou entre énoncés et diagrammes. C'est à ce changement de perspective et à ses conséquences que je consacre le chapitre 8. Nous verrons que ce cadre permet de comparer, eu égard à leur contenu, des représentations très différentes, par exemple des figures géométriques et des équations, et de clarifier certaines idées sur la « granularité » des représentations développées par Ken Manders (1996).

Le projet initial de Barwise et Etchemendy, cependant, va plus loin qu'une simple généralisation aux diagrammes des sémantiques en termes de modèles. Comme je le montrerai au chapitre 9, ils ébauchent une nouvelle théorie de l'inférence autour du concept de *contenu d'information* qu'ils empruntent à la sémantique des situations, une approche novatrice de la sémantique des langues naturelles développée quelques années plus tôt par Barwise en collaboration avec John Perry. Sur cette base, Barwise et son élève Atsushi Shimojima développent une analyse originale des vertus inférentielles des diagrammes, qui fait l'objet du chapitre 10.

En un sens très général, leur projet affronte donc la même question que Simon : comment comprendre les notions de représentation et d'information ? Comment concevoir la possibilité de représenter les mêmes informations de différentes manières ? Mais Barwise et Etchemendy s'inscrivent dans une tradition intellectuelle différente, celle de la logique, et abordent ces questions avec d'autres outils ; là où Simon se contente de donner une caractérisation de l'*équivalence* informationnelle, eux cherchent à employer les outils de la logique mathématique pour proposer une analyse substantielle de ce qu'est l'information.

Pour situer les problèmes auxquels nos auteurs font face, j'aimerais donc, à titre de préliminaire, revenir sur la notion d'information en logique. Comme le remarquent van Benthem et Martinez (2008), le terme est souvent employé informellement dans les cours d'introduction à la logique, mais de manière vague et pas toujours cohérente. Ainsi, on entend que l'inférence consiste à « extraire de l'information » des prémisses, comme l'écrivent Barwise et Etchemendy plus haut ; pourtant, on dit aussi que ce qui caractérise l'inférence déductive, c'est qu'elle n'apporte « aucune nouvelle information ». Comment comprendre cette notion ?

Commençons par l'idée qu'une inférence déductive n'apporte « aucune nouvelle information ». John Corcoran (1998) explicite soigneusement comment il faut employer le terme si l'on suit cette intuition-là :

> First, a given proposition follows from, is a consequence of, a given postulate set if all of the information contained in the proposition is contained within the set. Second, a given proposition is independent of, not a consequence of, a given postulate set if the proposition contains any information outside of the information content of the set. Third, a proposition is tautological if it is devoid of information [...]. Fourth, a proposition is contradictory if it contains all information (pertaining to the domain of investigation) [...]. Fifth, no proposition has any information in common with its own negation [...]. Sixth, the disjunction of one given proposition with a second contains exactly the information that the first has in common with the second [...].[28]

Cet usage est en harmonie avec la métaphore classique d'après laquelle la conclusion d'une inférence déductive est « contenue » dans ses prémisses. Dans ce cadre, les propositions ont à la fois une forme (leur forme logique) et un contenu, et l'information est du côté du contenu. Ainsi, deux propositions logiquement équivalentes contiennent la même information mais ont des formes différentes. L'information est alors conçue comme une sorte de substance amorphe que l'on peut mettre dans différents récipients[29]. Pour filer une compa-

28. « Premièrement, une proposition donnée découle de (est conséquence de) un ensemble donné de postulats si toute l'information contenue dans la proposition est contenue dans cet ensemble. Deuxièmement, une proposition donnée est indépendante de (n'est pas une conséquence de) un ensemble donné de postulats si la proposition contient quelque information que ce soit en-dehors de l'information contenue dans l'ensemble. Troisièmement, une proposition est tautologique si elle est dépourvue d'information [...]. Quatrièmement, une proposition est contradictoire si elle contient toutes les informations (concernant le domaine d'investigation) [...]. Cinquièmement, aucune proposition n'a d'information en commun avec sa propre négation [...]. Sixièmement, la disjonction d'une proposition donnée avec une seconde contient exactement les informations que la première a en commun avec la deuxième [...]. » (Corcoran 1998, p. 115).

29. Cet usage est facilité par une particularité grammaticale de l'anglais, qui pose d'ailleurs un problème de traduction : le terme d'« *information* » est un nom indénombrable (« *mass noun* »), qui n'admet pas de singulier et de pluriel ; on ne peut pas dire « une information » (sinon *via* une périphrase comme « *a piece of*

raison suggérée par Corcoran[30], faire une inférence serait comme tirer de l'eau d'un puits qui contiendrait l'information des prémisses ; passer de la conclusion ainsi obtenue à une autre, logiquement équivalente, serait comme transvaser dans une bouteille le seau tiré du puits.

À elles seules, ces élucidations ne disent rien de substantiel sur ce qu'est l'information (Corcoran, d'ailleurs, propose d'en faire une notion primitive). Néanmoins, elles s'accordent bien avec les sémantiques en termes de modèles ou de mondes possibles. L'idée centrale, qui remonte à Carnap et Bar-Hillel (1952), est d'assimiler le contenu d'information d'une proposition à l'ensemble des états du monde avec lesquels elle est compatible : plus une proposition est informative, moins elle laisse de possibilités ouvertes. Il est alors facile de traduire le vocabulaire informationnel en termes d'ensemble de possibilités, à condition de veiller à inverser la relation d'inclusion au passage : par exemple, le contenu d'information d'une proposition Q *est inclus* dans celui d'une proposition P (i.e. $P \to Q$) si l'ensemble des possibilités permises par Q *contient* celui des possibilités permises par P. Pour van Benthem et Martinez (2008), il s'agit de la première de trois grandes conceptions de l'information en logique[31]. Ils l'appellent « l'information comme éventail de possibilités » (« *information as range* »), et identifient toute une famille d'approches qui en sont dérivées.

Mais alors, comment comprendre l'autre expression informelle dont nous sommes partis, d'après laquelle l'inférence déductive est une « extraction d'information » ? Elle suggère qu'une inférence déductive peut nous apprendre quelque chose ; or, dans le cadre précédent, on rencontre ce que Jaakko Hintikka a appelé le « scandale » de la déduction : celle-ci n'est jamais informative. Une pirouette pour s'en sortir est de dire que la déduction ne nous donne pas de nouvelles informations *simpliciter*, mais seulement de nouvelles informations explicites. (C'est d'ailleurs ce que font Barwise et Etchemendy dans la citation ci-dessus.) Toutefois, c'est seulement déplacer le problème. Le fait même de parler de nouvelles *informations* explicites – plutôt que de changement de forme de l'information qu'on a déjà – montre bien qu'on vise ici une autre conception de l'information, liée à la forme des propositions.

information ») ni « des informations » (en toute rigeur, le pluriel « *informations* » est incorrect). Pour rester au plus près de l'anglais, il faudrait donc traduire par « de l'information », ce qui n'est guère idiomatique ni toujours possible. Je l'ai fait là où cela me semblait tolérable ; dans les autres cas, j'ai choisi le pluriel (traduire « *information* » par « une information » me semble en effet plus trompeur, parce que cela introduit une idée d'atomicité complètement absente de l'original).

30. Voir Corcoran 1998, p. 122.

31. La typologie de van Benthem et Martinez 2008 en trois grandes approches est reprise dans Martinez et Sequoiah-Grayson 2018, qui est plus complet et plus à jour techniquement, mais peut-être moins clair concernant les motivations et les difficultés des différentes approches.

Ce constat conduit naturellement à une autre des trois grandes familles de conceptions de l'information identifiées par van Benthem et Martinez, « l'information comme code », appuyée sur la syntaxe plutôt que sur des modèles. Bien sûr, il ne serait pas raisonnable de prendre en compte tous les détails syntaxiques de tel ou tel système formel : l'enjeu est alors de capturer des invariances pertinentes pour l'information portée sans quitter le domaine de la syntaxe [32]. La difficulté principale, toutefois, est d'expliquer comment des structures purement formelles de règles de formation et de transformation peuvent capturer de l'information sur le monde. Une première stratégie est de soutenir que la donnée de telles règles suffit à déterminer la signification des termes de nos langues ; c'est la voie suivie par les sémantiques preuve-théoriques [33]. Une autre stratégie est d'essayer de combiner une notion syntaxique d'information avec une approche en termes de modèles [34].

Ce qui est clair, c'est que les intuitions liées au concept d'information s'accordent mal avec la division habituelle des tâches entre syntaxe et modèles. On peut voir le travail de Barwise et Etchemendy comme une tentative d'échapper à ce dilemme en important en logique les idées développées pour la langue naturelle par la sémantique des situations. Dans la typologie de van Benthem et Martinez, celle-ci est justement à la base de la troisième grande approche, « l'information comme corrélation ». Je la présenterai au chap. 9. D'ici-là, les systèmes logiques des chapitres 7 et 8 nous permettront de mieux comprendre pourquoi, au-delà des tensions générales que je viens d'exposer, les concepts habituels de syntaxe et de modèles posent des problèmes particulièrement aigus pour les diagrammes et invitent à une reconceptualisation.

32. Pour un survol de différentes approches récentes, voir Martinez et Sequoiah-Grayson 2018, section 3.

33. Pour une introduction, voir Sundholm [1986] 2002 ou encore Schroeder-Heister 2018. De ce point de vue, on peut concevoir ces sémantiques comme une sorte de version logique des positions inférentialistes brièvement évoquées dans la partie précédente (cf. *supra*, note 38 p. 147).

34. C'est une solution fréquemment ébauchée en philosophie du langage, où un problème proche est connu sous un autre nom : celui de la nature et de l'individuation des propositions. Pour une introduction au débat contemporain, voir Soames 2012 ; cf. aussi King, Soames et Speaks 2014.

Chapitre 7

Sun-Joo Shin et les diagrammes logiques

Pour comprendre comment Barwise, Etchemendy et leurs élèves étendent aux diagrammes les notions de syntaxe et de sémantique, commençons par un exemple, celui des diagrammes de Venn. C'est l'un des tout premiers cas traités en détails par leur groupe de recherche [1] : Sun-Joo Shin l'étudie dès la fin des années 1980 pour sa thèse de doctorat [2], et son travail est prolongé par d'autres jeunes chercheurs, en particulier Eric Hammer [3]. Il a pour eux valeur paradigmatique.

7.1 Diagrammes d'Euler, de Venn et de Peirce

L'expression « diagramme de Venn » est souvent employée abusivement pour toutes les représentation d'ensembles ou de notions par des courbes fermées, alors qu'il existe plusieurs conventions bien distinctes [4]. Pour lever toute ambiguïté, je commence ici par présenter informellement les deux systèmes les plus connus, celui d'Euler et celui de Venn (discuté et amélioré par Peirce, comme nous le verrons). Shin concentre ses efforts de formalisation sur le système de Venn, mais discute aussi en détail les diagrammes d'Euler,

1. L'autre grand exemple de Barwise et Etchemendy à cette époque est leur système *Hyperproof*, qu'ils développent pour l'enseignement de la logique. Nous le discutons plus en détail dans la section suivante.

2. Son travail circule d'abord sous la forme d'un rapport interne daté de 1990, correspondant à une conférence, puis est diffusé par un article Shin [1991] 1996. Sa thèse Shin 1991 sera ultérieurement publiée sous forme de livre (Shin 1994).

3. Hammer 1994 ; Hammer et Danner 1996.

4. Pour une introduction générale à l'histoire des diagrammes en logique, voir Moktefi et Shin 2012. Pour une présentation efficace des systèmes les plus courants, voir Shin, Lemon et Mumma 2013, section 2.

qu'elle finit par rejeter : ils seraient ambigus, non rigoureux et même trompeurs. Cette différence de traitement rend la comparaison entre les deux systèmes particulièrement instructive. J'aurai l'occasion, dans la dernière partie de ce travail, de réfléchir aux raisons pour lesquelles la pratique informelle d'Euler semble échapper au genre de formalisation que Shin propose [5].

a) Les diagrammes d'Euler

La convention la plus répandue, et sans doute la plus naturelle, a été popularisée par le mathématicien Leonhard Euler [6] (qui n'est toutefois pas le premier à utiliser de tels diagrammes en logique [7]). Euler représente une « notion générale », par exemple la notion d'*homme*, par l'intérieur d'un cercle « dans lequel on conçoit que tous les hommes sont compris [8] ». Les relations entre cercles représentent alors les relations entre notions : si A correspond à la notion d'*homme* et B à la notion de *mortel*, la figure 7.1(a) fait voir que « tous les hommes sont mortels ». Plus généralement, Euler utilise ce principe pour illustrer les quatre types de propositions [9] reconnues par la logique aristotélicienne, à savoir « Tout A est B », « Nul A n'est B », « Quelque A est B » et « Quelque A n'est pas B » (fig. 7.1).

5. Cf. section 11.1.

6. Leonhard Euler emploie ces diagrammes en 1761 dans une série de lettres didactiques adressées à la jeune princesse Frédérique Charlotte, fille du prince Frédéric Henri de Brandenburg-Schwedt (sur cette correspondance, voir Speiser 1960, Calinger 2015, p. 417). Les lettres d'Euler à la princesse sont publiées dans leur intégralité à partir de 1768 (ELP) ; rapidement traduites dans la plupart des langues européennes (voir la longue liste des éditions dans EOO, III/11, p. LXI–LXX), elles connaissent d'après Andreas Speiser un succès « sans égal dans toute l'histoire des sciences » (Speiser 1960, p. VIII). Les diagrammes en question se trouvent dans les lettres CII-CV (ELP, vol. 2, p. 95–126), souvent numérotées lettres XXXIV à XXXVII de la deuxième partie dans les éditions ultérieures (p. ex. Euler 1843, p. 258–273). Ces lettres assureront une telle diffusion à nos diagrammes qu'en 1881, John Venn les retrouvera dans 34 traités de logique sur 60 qu'il a consultés (Venn 1881, p. 100, n. 1).

7. L'emploi de cercles s'intersectant pour représenter des collections partageant certains éléments a pu être redécouverte plusieurs fois ; quoi qu'il en soit d'éventuelles transmissions, on en trouve déjà trace dans un manuscrit musical au XIe siècle (Edwards 2006), et peut-être chez Raymond Lulle (Baron 1969, p. 115). En logique, l'emploi de diagrammes proches de ceux d'Euler pour illustrer la syllogistique remonte au moins à un traité de 1661 de Johann Christian Sturm, *Universalia Euclidea* (Sturm 1661, p. 84–96 ; un extrait est reproduit dans Risse 1970, p. 168, n. 678). D'après Maarten Bullynck, Sturm aurait pu les développer à partir d'autres représentations venues de son maître Erhard Weigel (Bullynck 2013, p. 31). Leibniz les emploie à son tour vers 1686 dans un manuscrit longtemps resté inédit (Leibniz 1903, p. 292–321), sans toutefois prétendre à la nouveauté ; il les tenait sans doute de Sturm, dont il connaissait bien le traité (il le cite entre autres dans les *Essais de Théodicée*, §212 : LPS, vol. VI, p. 245 ; voir aussi De Risi 2007, p. 223, n. 90). On retrouve ensuite un usage systématique de ces diagrammes dans un traité de logique du début du XVIII^e siècle (Lange 1712, en part. p. 249–268 ; voir aussi Risse 1970, p. 561–564). Il est vraisemblable qu'Euler ait eu connaissance, directement ou indirectement, de l'une de ces sources.

8. ELP, vol. 2, p. 98.

9. Notons qu'Euler, selon l'usage de son époque, entend par proposition l'expression linguistique d'un jugement, et non une entité abstraite qui serait le sens d'une phrase ou le contenu d'un jugement. La différence n'a pas d'importance pour nous ici.

(a) Tout A est B. (b) Nul A n'est B. (c) Quelque A est B. (d) Quelque A n'est pas B.

FIGURE 7.1 – Représentation diagrammatique des propositions par Euler, d'après ELP, vol. 2, p. 99–100

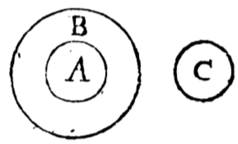


FIGURE 7.2 – Illustration du syllogisme « Tout A est B ; nul B n'est C ; donc nul A n'est C » d'après ELP, vol. 2, p. 106

Il s'en sert ensuite pour illustrer des syllogismes, c'est-à-dire des raisonnements à deux prémisses et une conclusion. La figure 7.2 illustre par exemple les deux prémisses « Tout A est B » et « Nul B n'est C » ; elle est obtenue en combinant leurs diagrammes et permet d'observer la conclusion « Nul A n'est C ».

Cette idée, à première vue très claire, souffre cependant d'une ambiguïté fondamentale. Euler le remarque lui-même : la proposition « Quelque B est A », par exemple, est compatible non seulement avec la figure 7.1(c), mais aussi avec la figure 7.1(a). Le problème est que les diagrammes d'Euler en disent toujours davantage que les propositions de la logique aristotélicienne. Ainsi, le diagramme 7.1(c) montre non seulement que « Quelque B est A », mais aussi « Quelque A est B », « Quelque A n'est pas B » et « Quelque B n'est pas A » ; de même, le diagramme 7.1(a) montre non seulement « Quelque B est A », mais aussi « Quelque B n'est pas A » et « Tout A est B [10] ». En d'autres termes, on ne peut pas représenter « Quelque B est A », par exemple, sans trancher d'autres questions, comme celle de savoir s'il y a des A qui ne sont pas B. Les diagrammes sont *plus spécifiques* que les propositions. Ce problème a un analogue bien connu en géométrie : on ne peut pas, écrit-on souvent, dessiner un triangle sans décider s'il a ou non un angle obtus (c'est-à-dire plus grand qu'un angle droit).

Le même problème, à savoir que les diagrammes sont plus spécifiques que les propositions, se repose dès que l'on veut combiner deux prémisses. Pour illustrer « Tout A est B » et « Quelque C est A », Euler est par exemple contraint de considérer deux cas (fig. 7.3) avant de conclure que « Quelque C est B ». En se limitant au premier diagramme, on risquerait

10. ELP, vol. 2, p. 102–103.

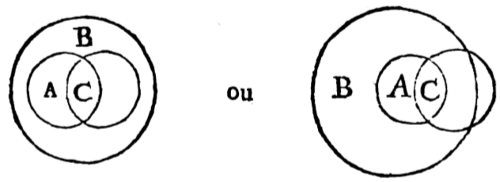


FIGURE 7.3 – Illustrations du syllogisme « Tout A est B ; quelque C est A ; donc quelque C est B » d'après ELP, vol. 2, p. 105

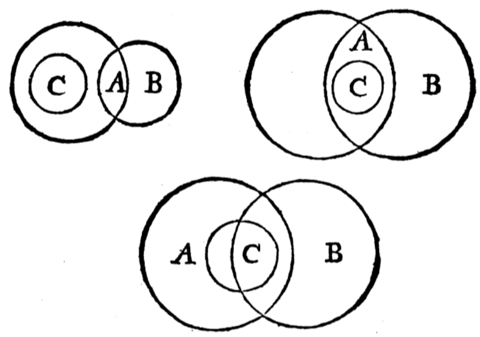


FIGURE 7.4 – Illustrations des prémisses « Quelque A est B » et « Tout C est A » d'après ELP, vol. 2, p. 112

de conclure à tort que « Tout C est B ».

Une solution simple pour contourner ces difficultés est de représenter tous les cas possibles. Il faut tout d'abord envisager tous les diagrammes correspondant à chaque prémisse : par exemple, pour « Quelque A est B », il faut envisager les diagrammes 7.1(c) et 7.1(a) mais aussi le cas où les deux cercles sont confondus. Puis, pour chaque choix d'un diagramme par prémisse, il faut envisager toutes les manières dont ces deux diagrammes peuvent se combiner.

Cependant, Euler ne fait pas ainsi : il ne varie jamais la représentation des prémisses. Il représente systématiquement « Quelque A est B », par exemple, par le diagramme 7.1(c). Certes, il envisage ensuite les différents cas qui résultent de la combinaison des diagrammes des prémisses. Mais cela ne suffit pas. Si l'on ne varie pas aussi la représentation des prémisses, on peut tirer des conclusions fausses. Considérons par exemple les trois diagrammes qu'Euler utilise pour illustrer « Quelque A est B » et « Tout C est A » (fig. 7.4). Il affirme qu'« on n'en saurait rien conclure, puisqu'il serait possible que la notion C fut dans la notion B toute entière, ou en partie, ou point du tout [11] ». Ses diagrammes représentent pourtant tous « Quelque B n'est pas C » ! Pour éviter cette erreur en procédant par cas, il

11. ELP, vol. 2, p. 113.

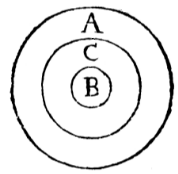


FIGURE 7.5 – Diagramme où « Quelque A est B » et « Tout C est A » mais où on n'a pas « Quelque B n'est pas C »

faudrait envisager un diagramme où tout B est A (fig. 7.5). Puisque Euler n'envisage pas ce diagramme, qu'est-ce qui l'empêche de tirer cette conclusion fausse ?

Hammer et Shin (1998) envisagent la possibilité qu'Euler utilise la position des lettres pour lever l'ambiguïté de ses diagrammes [12]. De fait, et quoiqu'il ne le dise pas explicitement, il ne place pas les lettres aux mêmes endroits pour « Quelque A est B » et pour « Quelque A n'est pas B » (cf. fig. 7.1 ci-dessus). Toutefois, Hammer et Shin montrent en détail que systématiser cet artifice est délicat (il faut par exemple aussi tenir compte de l'ordre des lettres, pour distinguer « Quelque A n'est pas B » de « Quelque B n'est pas A ») et qu'en définitive il ne suffit pas à lever toute ambiguïté. D'ailleurs, il arrive à Euler de ne pas l'utiliser, comme le prouve le troisième diagramme de la fig. 7.4.

Hammer et Shin en concluent que la démarche d'Euler n'est pas rigoureuse. Son système, d'après eux, ne fonctionne correctement que pour les propositions universelles. Pour pouvoir traiter aussi les propositions existentielles, Shin se fonde sur un autre système, mis au point au siècle suivant par John Venn.

b) Les diagrammes de Venn-Peirce

C'est entre autres pour pallier les défauts du système d'Euler [13] que John Venn développe les diagrammes qui portent aujourd'hui son nom [14]. Comme nous l'avons vu, la source des difficultés d'Euler est que ses diagrammes sont plus spécifiques que les propositions qu'ils sont censés représenter : on ne peut pas figurer « Quelque B est A » sans trancher d'autres questions, comme celle de savoir si « Quelque A n'est pas B ». Venn est très explicite :

12. C'est une idée assez répandue dans la littérature ; voir par exemple Stenning 2002, chap. 4 ou Moktefi et Shin 2012, p. 617.

13. En réalité, Venn cherche avant tout à mettre au point un système de représentation plus adapté à la nouvelle logique de Boole, mais cet aspect de l'histoire n'est pas essentiel pour nous.

14. Venn publie son idée d'abord dans un article (Venn 1880), ensuite dans le traité de logique issu de son enseignement à l'université de Cambridge (Venn 1881, 1894). Dans ces textes, Venn n'explique pas comment représenter les propositions existentielles ; il le fera brièvement au détour d'une digression dans un compte-rendu (Venn 1883).

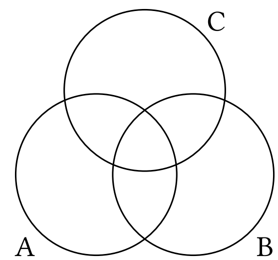


Figure 7.6 – Diagramme de Venn générique pour trois termes

> The weak point about [Euler's circles] consists in the fact that they only illustrate in strictness the actual relations of classes to one another, rather than the imperfect knowledge of these relations which we may possess, or wish to convey, by means of the proposition. [...] If we want to represent our uncertainty about the correct employment of a diagram, the only consistent way is to draw *all* the figures which are covered by the assigned propositions and say frankly that we do not know which is the appropriate one. [15]

Lui-même évite cette difficulté en adoptant des conventions radicalement différentes.

Tout d'abord, à la différence d'Euler, les relations d'intersection et d'inclusion entre courbes ne sont pas signifiantes : dans un diagramme de Venn, toutes les régions délimitées par des courbes doivent s'intersecter (cf. fig. 7.6), et les informations à représenter sont portées en grisant certaines régions ou en les marquant par des croix. En d'autres termes, chez Venn, on ne peut jamais rien conclure de la simple disposition des courbes. Clarifions ce que cela veut dire que toutes les intersections possibles doivent être représentées. Considérons n courbes fermées. Notons $A_1, \ldots, A_n$ leurs intérieurs et $\overline{A_1}, \ldots, \overline{A_n}$ leurs extérieurs (les complémentaires de leurs intérieurs). Alors tous les

$$X_1 \cap X_2 \cap \ldots \cap X_n \quad \text{où pour } 1 \leq i \leq n,\ X_i = A_i \text{ ou } X_i = \overline{A_i}$$

doivent être *non vides* et *connexes*. Un diagramme de Venn à n courbes découpe donc le plan en 2^n régions. Ce n'est pas immédiatement évident, mais il est possible de construire de tels diagrammes pour tout entier n : Venn (1880) en donne une construction générale,

15. « Le point faible [des cercles d'Euler] est qu'ils ne font qu'illustrer de manière stricte les relations existant entre certaines classes, plutôt que la connaissance imparfaite de ces relations que nous pouvons posséder ou vouloir transmettre au moyen de la proposition. [...] Si nous voulons représenter notre incertitude sur l'emploi correct d'un diagramme, la seule manière cohérente de procéder est de dessiner *toutes* les figures couvertes par les propositions assignées et de dire franchement que ne ignorons laquelle convient » (Venn 1881, p. 424–425, 1894, p. 510–511).

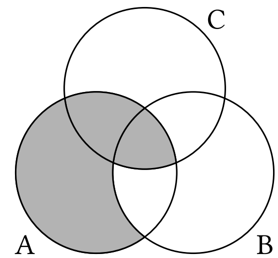


FIGURE 7.7 – Diagramme de Venn représentant « Tout A est B » et « Nul C n'est A »

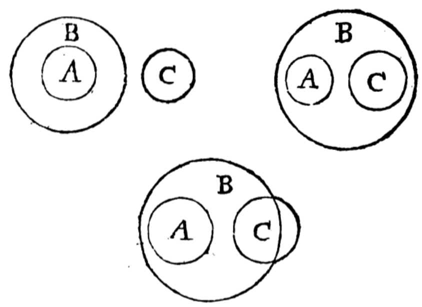


FIGURE 7.8 – Illustrations des prémisses « Tout A est B » et « Nul C n'est A » d'après ELP, vol. 2, p. 105–106

et il y en a de nombreuses autres [16].

On représente ensuite les informations dont on dispose en enrichissant ces diagrammes génériques. En premier lieu, on marque les régions dont on sait qu'elles sont vides en les grisant. Ainsi, pour représenter « Tout A est B » on grise la partie de A en-dehors de B ; pour « Nul C n'est A », la partie commune à C et A (fig. 7.7). On peut donc représenter ces deux prémisses simultanément en un unique diagramme, là où Euler avait besoin de trois diagrammes distincts (fig. 7.8). En effet, un diagramme d'Euler doit inévitablement trancher la question de savoir si B et C ont des éléments en commun, alors que le système de Venn peut laisser cette question ouverte.

On peut représenter les propositions existentielles, comme « Quelque B est A », de manière similaire. Venn est assez vague sur ce point [17] ; Sun-Joo Shin reprend la solution proposée par Peirce dans un manuscrit de 1903 [18], qui est conforme à l'esprit des remarques de Venn. Cette solution consiste à représenter « Quelque A est B » en marquant l'intersection des deux régions A et B par une croix (fig. 7.9 – nous utilisons le signe ⊗, comme

16. Voir par exemple Ruskey et Weston 2005.

17. Il ne l'explique que brièvement dans un compte-rendu ; voir note 14 *supra.*

18. C'est le manuscrit Ms. 479, qui n'a été publié qu'en 1933 (PCP, vol. 4, §347–371). Pour une discussion détaillée des différents manuscrits de Peirce sur les diagrammes d'Euler et de Venn, voir Pietarinen 2016.

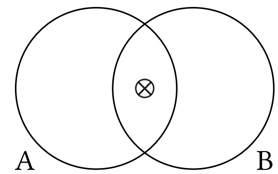


FIGURE 7.9 – Diagramme de Venn-Peirce pour « Quelque A est B »

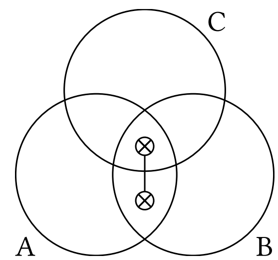


FIGURE 7.10 – Diagramme de Venn-Peirce pour « Quelque A est B », s'il y a un troisième ensemble C

Shin). S'il y a trois ensembles, Peirce fait deux croix reliées par une ligne (fig. 7.10) pour indiquer qu'il existe un élément *dans l'une au moins des deux régions marquées.* On peut joindre autant de croix qu'on veut, sachant que les lignes se lisent disjonctivement : il doit exister un élément correspondant à l'une des croix ainsi reliées. En d'autres termes, pour indiquer qu'une certaine région est non vide, il faut placer une croix dans chacune de ses sous-régions minimales et les relier. Par ailleurs, rien n'empêche plusieurs groupes de croix reliées de coexister, par exemple lorsqu'on représente « Quelque A n'est pas B » et « Quelque C est B » (fig. 7.11 ; on ne peut rien en conclure). Dans ces « diagrammes de Venn-Peirce [19] », les ambiguïtés du système d'Euler sont éliminées : aucune distinction de cas n'est plus nécessaire. En effet, chaque proposition de la logique aristotélicienne reçoit une représentation univoque, et la combinaison de plusieurs propositions se fait sans difficulté en les superposant sur le même diagramme de base.

En contrepartie, raisonner n'est plus aussi simple qu'avec les diagrammes d'Euler. Certes, dans les cas les plus simples, il suffit encore de combiner les diagrammes des prémisses pour lire la conclusion : ainsi, le diagramme pour « Tout A est B » et « Nul B n'est C » (fig. 7.12) montre immédiatement que « nul A n'est C », parce que l'intersection de A et C est grisée. Mais souvent, il faut en outre transformer le diagramme obtenu. Représen-

19. C'est Shin qui nomme ainsi ce système, qui est hybride parce qu'il associe l'idée initiale de Venn aux croix reliées introduites par Peirce, sans pour autant reprendre toutes les modifications proposées par ce dernier : Peirce marque les zones vides par des o (que l'on peut ensuite relier avec d'autres o ou avec des croix) au lieu de les griser. Pour plus de détails, voir Shin 1994, chap. 2 ou Shin, Lemon et Mumma 2013, sec. 2.

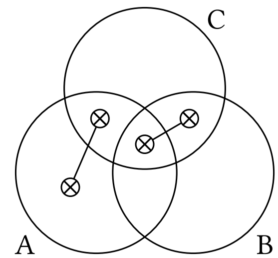


FIGURE 7.11 – Diagramme de Venn-Peirce pour « Quelque A n'est pas B » et « Quelque C est B »

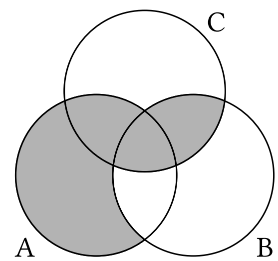


FIGURE 7.12 – Diagramme de Venn pour « Tout A est B » et « Nul B n'est C »

tons par exemple « Tout A est B » et « Quelque C est A » (fig. 7.13). Pour conclure, il faut remarquer que l'une des deux croix reliées est dans une région grisée, c'est-à-dire vide : on peut donc la supprimer (fig. 7.14), puis déduire de la croix restante qu'il existe un élément dans l'intersection des trois ensembles et donc que « Quelque C est B ». Pour cette raison, Peirce se rend compte qu'il faut introduire des règles de transformation des diagrammes [20]. Comme nous le verrons, cette idée est essentielle pour Shin.

Ces diagrammes de Venn-Peirce sont ceux dont Sun-Joo Shin formalise l'usage dans son système « Venn-I », à une différence près. S'inspirant là encore d'une remarque de Peirce [21], Shin encadre ses diagrammes dans un rectangle qui figure l'univers du discours. Cette modification permet de représenter des propositions comme « Tout est A » (fig. 7.15).

Pour compléter ce passage en revue des diagrammes logiques fondés sur des courbes fermées, je voudrais mentionner une dernière variante. Le système de Venn est souvent présenté comme moins « visuellement clair », moins « intuitif » ou moins « naturel » que celui d'Euler [22]. Pour cette raison, Hammer et Shin 1996, 1998 proposent un système hy-

20. PCP, Vol. 4, §361–362 (cf. note 18 *supra*).
21. PCP, vol. 4, §365.
22. Cf. par exemple Shin, Lemon et Mumma 2013, sec. 2.3, Giardino 2013, p. 143-144, ou pour une discussion plus approfondie, Giardino et G. Greenberg 2014, p. 9-14. Dans tous les cas, l'idée est qu'en passant du système

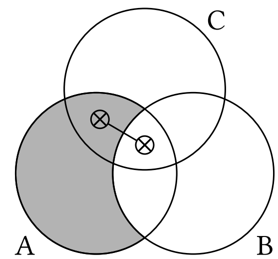


Figure 7.13 – Diagramme de Venn-Peirce pour « Tout A est B » et « Quelque C est A »

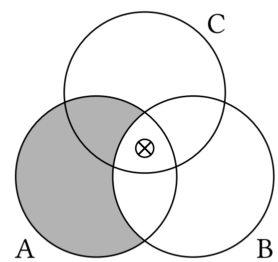


Figure 7.14 – Diagramme de Venn-Peirce pour « Tout A est B » et « Quelque C est A » après transformation

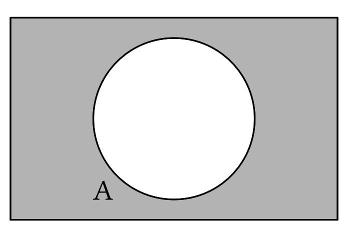


Figure 7.15 – Diagramme de Venn-Peirce-Shin pour « Tout est A »

bride. Ils conservent une partie du système d'Euler, à savoir que deux cercles délimitant des régions disjointes correspondent à des ensembles disjoints et que l'inclusion de cercles correspond à l'inclusion des ensembles correspondants. En revanche, lorsque deux cercles s'intersectent, on ne peut rien conclure sur la relation des ensembles correspondants, qui peuvent donc être disjoints, inclus l'un dans l'autre ou n'avoir en commun qu'une partie de leurs éléments respectifs. Ce système, d'après eux, conserve certains des avantages intuitifs du système d'Euler sous une forme rigoureuse et cohérente (voir aussi Hammer 1995, chap. 6, qui appelle « cercles d'Euler » sans autre précision ce système modifié). Ce système ne permet cependant pas d'exprimer de propositions existentielles, contrairement à celui que formalise Shin dans son travail principal.

7.2 Syntaxe et sémantique des diagrammes de Venn-Peirce

C'est sur la base des diagrammes de Venn-Peirce que Shin élabore ses systèmes formels diagrammatiques. Ce qui rend ces diagrammes adaptés, c'est que contrairement aux diagrammes d'Euler on peut clarifier leur contenu de manière univoque, indépendante du contexte. Cela permet à Shin de les munir d'une syntaxe et d'une sémantique, comme on fait d'ordinaire pour les langages formels. Voici comment Shin, Lemon et Mumma (2013) résument cette démarche :

> Since the development of modern logic, important concepts, e.g., syntax, semantics, inference, logical consequence, validity, and completeness, have been applied to sentential representation systems only. However, none of these turned out to be intrinsic to these traditional symbolic logics only. For any representation system, whether it is sentential or diagrammatic, we can discuss two levels, a syntactic and a semantic level. What inference rules tell us is how to manipulate a given unit, whether symbolic or diagrammatic, to another. [23]

Énoncés comme diagrammes peuvent donc être munis d'une syntaxe et d'une sémantique. La syntaxe consiste en règles de formation qui délimitent les représentations bien for-

---

d'Euler à celui de Venn, l'on perd en naturalité ce que l'on gagne en pouvoir expressif.

23. « Depuis que la logique moderne a été développée, ses concepts importants, par exemple la syntaxe, la sémantique, l'inférence, la conséquence logique, la validité et la complétude, n'ont été appliqués qu'aux systèmes de représentation phrastiques [*sentential representation systems*]. Cependant, aucun d'eux ne s'est avéré intrinsèque à ces seules logiques symboliques traditionnelles. Pour tout système de représentation, qu'il soit phrastique [*sentential*] ou diagrammatique, nous pouvons discuter deux niveaux, un niveau syntaxique et un niveau sémantique. Ce que les règles d'inférence nous disent, c'est comment manipuler une unité donnée, qu'elle soit symbolique ou diagrammatique, pour la transformer en une autre. » (Shin, Lemon et Mumma 2013).

mées, et en règles d'inférence qui indiquent comment transformer une représentation en une autre. La sémantique définit une classe d'interprétations et une relation de satisfaction entre une représentation et une interprétation. La satisfaction permet de définir une relation de conséquence logique entre représentations. Examinons brièvement comment Shin applique chacune de ces notions aux diagrammes de Venn-Peirce.

a) Préliminaires : comment raisonne-t-on dans le système de Shin ?

Le système de Shin est purement diagrammatique : un raisonnement doit partir de diagrammes et aboutir à un diagramme. Or les syllogismes sont des raisonnements qui conduisent de prémisses exprimées par des phrases à une conclusion exprimée par une phrase. Pour les traiter dans le système de Shin, il faut donc d'abord remplacer chaque prémisse, ainsi que la conclusion, par un diagramme. Cette première étape est extérieure au système (même si, comme nous le verrons plus loin, rien n'empêche de mettre au point un système *hétérogène*, dont la syntaxe contienne à la fois des phrases et des diagrammes ainsi que des règles permettant de passer des unes aux autres). On *combine* ensuite les diagrammes des prémisses et le cas échéant on *transforme* le diagramme résultant. Enfin, la conclusion découle des prémisses si le diagramme correspondant à la conclusion est *inclus* dans le diagramme final.

b) Diagrammes bien formés

Les diagrammes bien formés (DBFs) sont définis à partir des « objets primitifs » suivants : rectangles, courbes fermées [24], grisages, croix, segments. Une *région basique* d'un diagramme est l'intérieur d'un rectangle ou d'une courbe fermée ; les *régions* d'un diagramme sont définies inductivement à partir des régions basiques : l'intersection, l'union et la différence de deux régions sont encore des régions. Une *région minimale* est une région dans laquelle aucune autre région du diagramme n'est comprise. Les diagrammes bien formés sont alors définis inductivement comme suit :

1. un rectangle est un diagramme bien formé ;
2. à partir d'un diagramme bien formé, chacune des opérations suivantes conduit à un nouveau diagramme bien formé :
   (a) ajouter une courbe qui est intérieure au rectangle, de telle manière qu'elle ne rencontre aucune croix et que l'intersection de son intérieur avec chaque région

24. Pour s'assurer qu'on ne puisse pas confondre une courbe fermée et un rectangle, il faut imposer aux courbes fermées une forme arrondie (par exemple en demandant qu'elles soient dérivables).

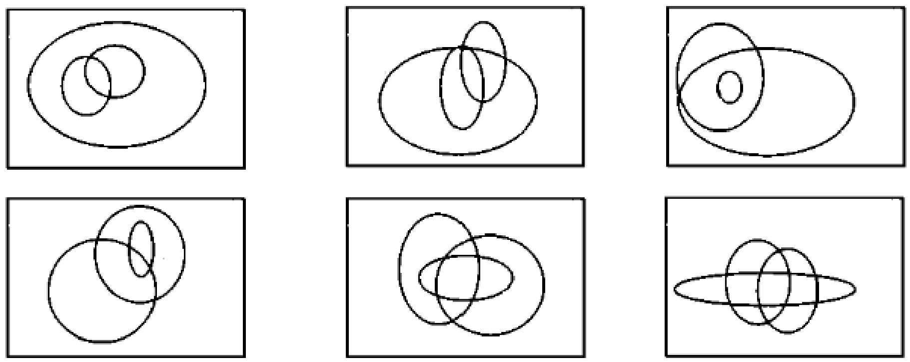


FIGURE 7.16 – Diagrammes mal formés (adapté de Shin 1994, p. 59)

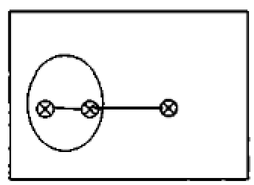


FIGURE 7.17 – Diagramme mal formé parce que deux croix reliées sont dans la même région minimale, d'après Shin 1994, p. 62

minimale soit un sous-ensemble strict, connexe et non vide de celle-ci ;

(b) griser une région entière (minimale ou non) ;

(c) ajouter une croix à l'intérieur d'une région minimale ;

(d) relier par un segment deux croix situées dans des régions minimales différentes.

Un premier effet de ces règles est que toutes les intersections possibles entre courbes doivent correspondre à une région non vide et connexe, ce qui est précisément la condition qui caractérise les diagrammes de Venn [25]. La raison en est qu'on ne peut introduire de nouvelle courbe que par la condition 2a. Les diagrammes suivants (fig. 7.16) sont donc mal formés.

Une seconde conséquence à noter est qu'aucune croix ne peut chevaucher une courbe : toute croix est à l'intérieur d'une région minimale bien déterminée. De plus, si une même région minimale peut contenir plusieurs croix différentes, deux croix d'une même région minimale ne peuvent pas être reliées (la fig. 7.17 est donc mal formée). Par conséquent, une famille de croix reliées – nous dirons comme Shin une « suite de x » – est déterminée par la plus petite région qui contient toutes les croix de la famille. La suite de x de la figure

25. Pour plus de précisions, voir *supra*, section 7.1.b).

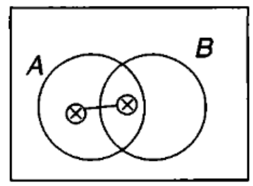


FIGURE 7.18 – « Suite de x » dans la région A, d'après Shin 1994, p. 69

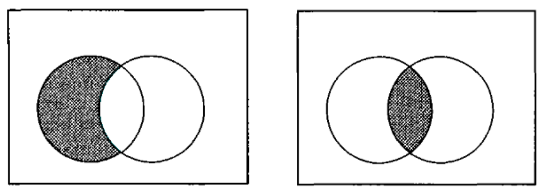


FIGURE 7.19 – Diagrammes pour « Tous les A sont B » (à gauche) et « Aucun B n'est C » (à droite)

7.18 correspond par exemple à la région A : il y a exactement une croix dans chaque région minimale incluse dans A et elles sont reliées.

Le diagramme précédent, comme beaucoup de ceux qu'utilise Shin, contient des lettres qui servent à étiqueter les courbes fermées. Ces lettres facilitent la référence à telle ou telle partie du diagramme, mais *ne font pas partie de la syntaxe officielle du système de Shin.* Elles font seulement partie du métalangage.

La syntaxe se complique d'un élément supplémentaire dès lors qu'on considère plusieurs diagrammes simultanément. Par exemple, si l'on veut traiter le syllogisme « Tous les A sont B ; aucun B n'est C ; donc aucun A n'est C » on commence par représenter séparément chaque prémisse (fig. 7.19). Ces deux diagrammes ne sont pas indépendants : la courbe de droite du premier est censée représenter le même ensemble que la courbe de gauche du second. Il faut le savoir, sans quoi on serait bien en peine de combiner ces diagrammes comme voulu (fig. 7.20). Or, jusqu'ici, rien dans la syntaxe de notre système ne permet de l'indiquer. La méthode la plus naturelle, déjà utilisée par Euler, serait d'utiliser des lettres pour étiqueter les courbes et d'employer la même lettre pour les courbes censées représenter le même ensemble [26]. Ce n'est pas la solution de Shin, qui a choisi d'exclure les lettres-étiquettes de la syntaxe de son système (probablement pour éviter le problème délicat d'éviter toute amiguïté dans la position des lettres par rapport aux courbes qu'elles étiquettent). Elle est donc forcée d'ajouter à son système un dernier élément : une relation

26. Cette solution plus naturelle est en général utilisée par les auteurs ultérieurs qui reprennent le système de Shin, par exemple Hammer 1995, chap. 4 et Hammer et Danner 1996.

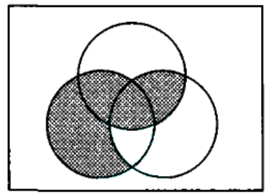


FIGURE 7.20 – Diagramme pour représenter conjointement « Tous les A sont B » et « Aucun B n'est C »

d'équivalence qui relie les courbes fermées censées représenter les mêmes ensembles [27]. Elle l'appelle *relation de contrepartie* [28]. Dans notre exemple précédent, la courbe de droite du diagramme de gauche de la figure 7.19 serait en relation avec la courbe de gauche du diagramme de droite [29]. On en vient cependant vite à des formulations compliquées pour exprimer quelles courbes sont en relation, comme le montre la phrase précédente. L'emploi de lettres-étiquettes, *extérieures au système*, permet à Shin d'indiquer plus facilement de quelles courbes elle parle. Les deux diagrammes correspondant à nos prémisses pourraient par exemple être représentés, avec des étiquettes, comme à la figure 7.21. Au lieu de parler de la courbe de gauche du diagramme de droite, et ainsi de suite, on peut tout simplement dire que les courbes A_2 et A_3 sont en relation [30]. Remarquons bien que les étiquettes servent ici de noms propres pour des courbes, et n'indiquent pas directement – comme c'était le cas chez Euler – quelles courbes sont censées représenter le même ensemble. Cette manière contournée de procéder a été largement abandonnée par la suite au

27. Plus précisément, c'est une relation d'équivalence entre *régions basiques*, c'est-à-dire entre intérieurs de rectangles ou de courbes fermées. Cette relation indique donc aussi si deux diagrammes ont ou non le même univers du discours. Cette relation doit satisfaire les conditions suivantes : un rectangle ne peut être en relation qu'avec un rectangle et une courbe fermée avec une courbe fermée ; de plus, deux éléments distincts ne peuvent être en relation que s'ils appartiennent à des diagrammes distincts (en d'autres termes, on ne peut pas décider de représenter un même ensemble par deux courbes d'un même diagramme). Cette relation peut naturellement s'étendre en une relation d'équivalence entre régions (et plus seulement régions basiques). Par exemple, si l'on a deux diagrammes, le premier avec deux courbes A et B, le second avec deux courbes A' et B', et que A et A' d'une part, B et B' d'autre part sont en relation, l'intersection de A et B et celle de A' et B' devront aussi représenter le même ensemble. Notons que nous utilisons ici les lettres-étiquettes comme Shin : elles servent de noms pour certaines courbes, ce qui nous permet de dire lesquelles sont censées représenter le même ensemble, mais ces lettres-étiquettes font partie du métalangage et pas du système lui-même.

28. Le mot anglais est « *counterpart* », que l'on emploie par exemple dans une phrase comme « *the French prime minister met his British counterpart in London yesterday* », c'est-à-dire « le premier ministre français a rencontré son homologue britannique à Londres hier ». On pourrait donc traduire « relation d'*homologie* » si ce terme français n'était pas déjà surchargé de sens techniques.

29. De plus, les deux rectangles seraient en relation pour indiquer qu'on fait référence au même univers de discours.

30. De même pour les rectangles D_1 et D_2.

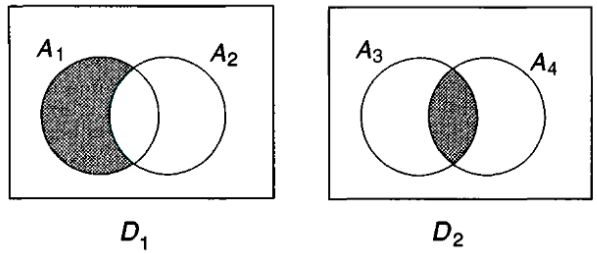


FIGURE 7.21 – Diagrammes avec lettres-étiquettes pour « Tous les A sont B » (à gauche) et « Aucun B n'est C » (à droite)

profit de lettres-étiquettes, y compris par Shin elle-même [31].

Les diagrammes bien formés sont ainsi définis inductivement, comme les formules bien formées des langages formels habituels. Cependant, on peut construire un même diagramme bien formé de différentes manières [32] : à première vue, il n'y a donc pas, contrairement au cas des langages formels, de théorème de lecture unique. En conséquence, il est difficile de définir inductivement des propriétés des diagrammes en suivant l'ordre d'application des règles de formation : de telles définitions supposeraient en effet des preuves fastidieuses d'invariance pour toutes les constructions possibles, ce qui explique certainement que Shin n'en utilise pas. Il serait cependant possible de modifier le système de Shin pour remédier assez largement à ce défaut [33].

c) Règles de transformation

Sur le modèle des règles d'inférence usuelles, Shin introduit ensuite des *règles de transformation* permettant de passer de certains diagrammes à d'autres [34]. Comme nous l'avons vu, l'idée essentielle est déjà présente chez Peirce ; Shin clarifie et complète les règles de celui-ci, et les décrit sous une forme qui distingue soigneusement syntaxe et sémantique [35].

Il y a par exemple une « règle d'unification » qui permet de combiner deux diagrammes

31. Voir par exemple Hammer 1995 ; Hammer et Danner 1996 ; Howse, Molina, Shin et J. Taylor 2002.

32. Il est en effet souvent possible de changer l'ordre des règles appliquées et même leur nombre (du fait de la règle 2b qui permet de griser des régions non minimales).

33. Pour ce faire, il faudrait exiger que l'on applique les règles dans l'ordre : d'abord la règle 1, puis la règle 2a autant de fois que nécessaire, puis de même la 2b, la 2c (modifiée de manière à ce qu'on ne puisse griser qu'une région minimale à la fois) et enfin la 2d. On aurait alors lecture unique à l'ordre près des applications de chaque règle : on pourrait donc utiliser des définitions inductives, à condition de démontrer leur invariance par permutation des applications de chaque règle. Voir aussi Scotto di Luzio 2002.

34. Signalons que l'une des règles de transformation de Shin, qui permet d'effacer une courbe d'un diagramme, comporte une difficulté : son application indiscriminée lorsqu'il y a quatre courbes ou plus peut parfois conduire à des diagrammes qui ne sont pas bien formés (Scotto di Luzio 2002). Voir *infra*, section 7.4.a).

35. Voir la discussion des règles de Peirce dans Shin 1994, p. 27-40.

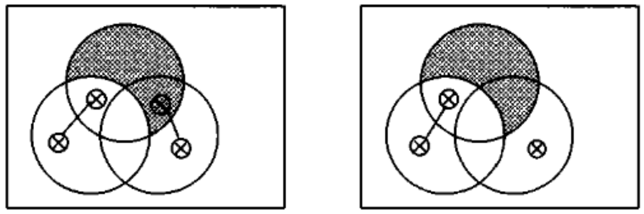


FIGURE 7.22 – Application correcte de la règle d'élimination des x, d'après Shin 1994, p. 85

en un seul. Cette règle est cruciale pour raisonner, puisque le principe du système est de représenter chaque prémisse séparément puis de les combiner. Une autre règle permet d'effacer les croix qui se trouvent dans des régions grisées. Nous avons déjà discuté un cas où cette règle est nécessaire : celui du syllogisme « Tout A est B ; quelque C est A ; donc quelque C est B » (voir figures 7.13 et 7.14).

Je me contenterai ici d'illustrer la démarche de Shin en présentant plus en détails la règle d'élimination des croix. Shin écrit :

> We may copy a *wfd* omitting any subpart of an x-sequence [36] if that part is in a shaded region. The number of x-sequences in a diagram does not increase. That is, if ⊗ in a shaded region is at the end of an x-sequence, we may erase −⊗ or ⊗− so that the remaining part is one x-sequence. If ⊗ in a shaded region is in the middle of an x-sequence, after erasing this ⊗ we should connect both of the remaining parts so that we have one x-sequence (not two x-sequences by disconnection). [37]

On peut donc supprimer toute croix qui se trouve dans une région grisée dès lors qu'elle est reliée à d'autres croix (fig. 7.22). Il faut cependant s'assurer que les croix restantes qui faisaient partie d'une même suite de x restent reliées après la transformation (fig. 7.23). Shin illustre sa règle par des exemples d'application correcte, que nous venons de donner, ainsi que par des contre-exemples.

Les règles de transformation permettent à Shin de définir une relation $\vdash$ analogue à la relation de prouvabilité en logique ordinaire : si Δ est un ensemble de diagrammes et D un autre diagramme (tous bien formés), alors $\Delta \vdash D$ si l'on peut obtenir D à partir des

36. C'est-à-dire une famille de croix reliées.

37. « Nous pouvons recopier un diagramme bien formé en omettant toute sous-partie d'une suite de x qui est dans une région grisée. Le nombre de suites de x dans le diagramme n'augmente pas. En d'autres termes, si le ⊗ dans une région grisée est à la fin d'une suite de x, nous pouvons supprimer −⊗ ou ⊗− de sorte que la partie restante ne constitue qu'une unique suite de x. Si le ⊗ dans une région grisée est au milieu d'une suite de x, après avoir effacé ce ⊗ nous devons relier les deux parties restantes de manière à avoir une unique suite de x et non deux suites de x déconnectées. » (Shin 1994, p. 85).

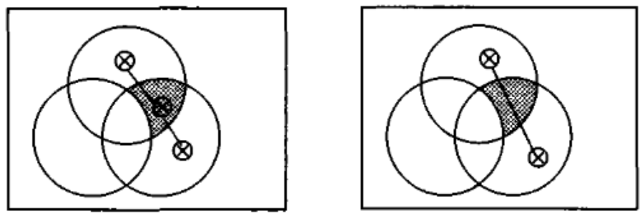


FIGURE 7.23 – Élimination correcte d'une croix située au milieu d'une suite de x, d'après Shin 1994, p. 85

diagrammes de Δ *via* une suite finie d'applications de règles de transformation.

d) Interprétations

Pour Shin, interpréter un diagramme, c'est associer à chaque région du diagramme un certain ensemble. Plus précisément, une interprétation est la donnée

— d'un ensemble U qui correspond à l'univers du discours,

— d'une fonction qui assigne à chaque région du diagramme une partie de U.

Déterminer une interprétation revient en fait à associer chaque courbe fermée (c'est-à-dire, dans la terminologie de Shin, chaque *région basique* du diagramme) à une partie de U : par union et intersection, les interprétations des autres régions du diagramme en découlent [38].

Quand on interprète simultanément plusieurs diagrammes, il faut en outre que les régions basiques associées par la relation de contrepartie correspondent aux mêmes ensembles.

On peut alors définir ce que c'est pour une interprétation de *satisfaire* un diagramme : il suffit que les régions grisées correspondent à des ensembles vides et que les régions contenant entièrement une suite de croix reliées correspondent à des ensembles non vides. [39].

e) Relation de conséquence ; correction et complétude

Une fois qu'on a défini ce qu'est une interprétation d'un diagramme, et ce que c'est pour une interprétation de satisfaire un diagramme, on peut définir une relation de conséquence

38. Shin 1994, section 3.3.1.

39. Cette sémantique est définie directement à partir de propriétés du diagramme (les régions grisées ou contenant des suites de croix reliées) et non inductivement en suivant les règles de formation, comme on le fait d'ordinaire en logique ; j'ai déjà signalé plus haut (section 7.2.b)) que les règles de formation de Shin se prêtent mal à des définitions de ce type. En faisant les modifications suggérées à la note 33, on pourrait cependant donner une définition inductive de la sémantique.

comme on le fait habituellement en logique : Shin dit qu'un diagramme D est conséquence d'un ensemble P de diagrammes (ce qu'elle note $P \vDash D$) si toute interprétation qui satisfait les diagrammes P satisfait aussi le diagramme D.

On peut alors démontrer la correction des règles de transformation du système. En d'autres termes, si les règles permettent de dériver D de certains autres diagrammes P (c'est-à-dire si $P \vdash D$), alors D est bien conséquence de P (c'est-à-dire $P \vDash D$[40]). À l'inverse, Shin montre aussi la complétude de son système : si $P \vDash D$, alors $P \vdash D$[41]. Shin montre même qu'une variante de son système (permettant de considérer des disjonctions de diagrammes bien formés) est équivalente à un calcul des prédicats monadique simple[42]. Le système de Shin a donc les propriétés habituellement attendues d'un système logique.

7.3 Un système hétérogène pour les diagrammes de Venn

La manière dont Shin étend aux diagrammes les notions de représentation bien formée, de règle d'inférence, de structure d'interprétation, de satisfaction et de conséquence logique se généralise. Barwise et Hammer ([1994] 1996) en déduisent un concept élargi de « système logique », qui ne se distingue des habitudes de la logique que sur un point : ils abandonnent l'idée que « les représentations bien formées sont de forme finie et linéaire et peuvent donc être modélisées par des suites finies de symboles[43] ». On peut donc avoir des systèmes logiques fondés sur des diagrammes.

Dans ce cadre, rien n'empêche de concevoir également des systèmes formels qui combineraient diagrammes et formules logiques. Pour reprendre la terminologie de Barwise et Etchemendy ([1995] 1996a), de tels systèmes sont non plus diagrammatiques, mais « hétérogènes ». Leur syntaxe juxtapose deux (ou davantage) types de représentations différentes dont chacun a ses propres règles de formation, mais qui ont les mêmes structures d'interprétation, de manière à ce qu'on puisse définir une relation de conséquence sémantique entre représentations de types différents. On peut alors codifier des règles de transformation permettant de passer de formules à des diagrammes ou réciproquement, et ainsi de se rapprocher de la pratique mathématique.

Hammer (1994, 1995) développe par exemple un système hétérogène pour les dia-

40. Shin 1994, §3.6.

41. Plus précisément, Shin ne démontre ce résultat que pour un ensemble *fini* P de diagrammes de départ (Shin 1994, §3.7). Hammer et Danner 1996 étendent le théorème au cas général. (Voir aussi Miller 2006 pour une preuve plus simple.)

42. Shin 1994 présente cette extension de son système, qu'elle appelle Venn-II, au chap. 4. Elle en montre l'équivalence avec un calcul des prédicats monadique sans symboles de fonction ni de constante au chap. 5.

43. Barwise et Hammer [1994] 1996, p. 54.

grammes de Venn. Côté formules, Hammer reprend une logique du premier ordre usuelle ; côté diagrammes, il reprend essentiellement le système Venn-I de Shin [44]. Mais à ces deux composants, il ajoute des règles d'inférence hétérogènes. Ainsi, si une région d'un diagramme est grisée, on peut déduire que l'ensemble correspondant est vide par une règle formelle qui a pour prémisse un diagramme et pour conclusion une formule logique [45]. À l'inverse, à partir d'une formule affirmant qu'un certain ensemble est vide et d'un diagramme qui représente cet ensemble par une courbe, on peut déduire un nouveau diagramme où l'intérieur de cette courbe est grisé (la règle a cette fois pour prémisses un diagramme et une formule, et pour conclusion un diagramme [46]). La sémantique permet dans les deux cas de vérifier que ces règles d'inférence sont correctes, au sens où leur conclusion est toujours conséquence de leurs prémisses. Nous avions remarqué que le système de Shin s'écartait de l'usage habituel des diagrammes de Venn parce qu'il était purement diagrammatique et que, par conséquent, la traduction des prémisses en diagrammes lui échappait. Le système de Hammer montre qu'il est facile de remédier à ce défaut.

7.4 Difficultés sémiologiques et formalité

Peut-on vraiment qualifier de *formels* le système de Shin et ses variantes ? Certains ingrédients essentiels semblent manquer. Tout d'abord, pour parler de système formel au sens habituel du terme, il faut pouvoir distinguer *types* et *tokens* de diagrammes ; or je n'ai pas expliqué dans quels cas considérer deux diagrammes comme les mêmes, et comme nous le verrons, cette question cache une difficulté assez sérieuse. En second lieu, on s'attend d'ordinaire à ce que les règles de transformation de nos systèmes formels soient mécanisables ; mais comment comprendre cette idée dans le cas d'entités syntaxiques qui ne sont pas de simples suites de symboles ?

---

44. Il faut signaler une différence. Diagrammes et formules doivent avoir les mêmes structures d'interprétation, ce qui n'est pas immédiatement le cas si l'on combine le système de Shin avec un langage du premier ordre : les interprétations d'un langage du premier ordre sont des fonctions qui assignent des ensembles à des *symboles* (de constante et de prédicat), alors que les interprétations d'un diagramme de Shin assignent des ensembles à des *courbes* (ou à des classes d'équivalence de courbes pour la relation de contrepartie, cf. 7.2.b)). Hammer modifie donc la syntaxe du système de Shin pour que les courbes soient étiquetées par des termes de son langage du premier ordre (en fait des termes à une variable libre, c'est-à-dire qui définissent un ensemble). La relation de contrepartie n'est alors plus nécessaire, et une interprétation des symboles de constante et de prédicat suffit à interpréter aussi les courbes. (Une solution encore plus simple, mais moins flexible, aurait été d'étiqueter les courbes par des symboles de prédicat unaire du langage du premier ordre utilisé.)

45. C'est la règle « $\forall$–Observe » (Hammer 1994, p. 82).

46. C'est la règle « $\forall$–Apply » (Hammer 1994, p. 82).

a) *Types* et *tokens* de diagrammes

Les langages formels dont nous avons l'habitude sont fondés sur une distinction très robuste entre *types* et *tokens* de formules. Cette distinction est cruciale si l'on veut justifier l'idée que différentes personnes, travaillant sur des inscriptions matérielles différentes, puissent procéder de manière uniforme : si l'on veut affirmer que deux personnes ont appliqué la même règle pour passer d'une inscription à une autre, il faut (au minimum) pouvoir reconnaître leurs points de départ comme identiques, et de même pour leurs points d'arrivée. La distinction *type-token* est donc essentielle pour remplir l'objectif d'explicitation complète des procédures qui est consubstantiel au projet même de mathématiques formalisées.

Il est vrai que d'ordinaire, on n'applique la distinction *type-token* qu'aux symboles et suites de symboles caractéristiques des langages formels. Toutefois, rien n'empêche de la généraliser, ce que fait déjà Nelson Goodman (1968) quoique dans une terminologie différente, comme nous l'avons déjà vu [47].

Or dans son livre, Shin n'introduit pas vraiment de distinction analogue. Pour formuler ses règles de transformation, elle se contente d'une notion floue de « copie » d'un diagramme [48] : par exemple, la règle d'élimination des croix se trouvant dans une région grisée permet de « *recopier* un diagramme bien formé en omettant toute suite de x qui est dans une région grisée [49] ». Quand on recopie un diagramme en ce sens, a-t-on le droit de réorganiser spatialement les courbes ou faut-il respecter aussi exactement que possible leurs positions ? La définition de Shin est trop vague pour répondre avec certitude, mais ses exemples suggèrent qu'elle pense plutôt à une copie exacte. Cette interprétation poserait problème, parce qu'en pratique, tous les diagrammes concrets seront un peu différents, et on se retrouverait sans critère permettant de dire s'il faut les traiter ou non comme identiques.

Shin définit en outre une notion d'« équivalence syntaxique » entre diagrammes. Même si Shin ne l'utilise pas ainsi, cette notion est plus adaptée pour remplir le rôle d'une distinction *type-token* ; d'ailleurs, Howse, Molina, Shin et J. Taylor (2002) la reprennent ultérieurement en ces termes-là [50]. L'idée est de caractériser un diagramme par la liste de ses courbes, de ses grisures et de ses suites de croix reliées. Formellement, Shin introduit pour

47. Voir *supra*, section 6.1.a), en part. n. 1 p. 156.
48. Voir Shin 1994, p. 81.
49. Shin 1994, p. 85 (je souligne). Pour plus de détails sur cette règle, voir *supra*, section 7.2subsec :shin-transformations.
50. Le cadre formel de Howse, Molina, Shin et J. Taylor 2002 est assez différent de celui de Shin et ils s'en tiennent à des diagrammes de Venn sans croix, mais l'idée est bien la même.

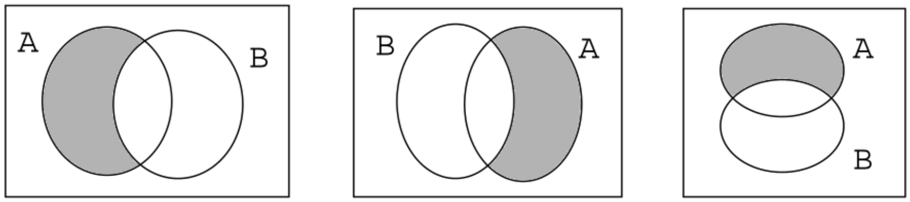


FIGURE 7.24 – Diagrammes syntaxiquement équivalents, adaptés de Howse, Molina, Shin et J. Taylor (2002, p. 149)

chaque diagramme un nouvel objet, qu'elle appelle sa « suite » (« *sequence* »), formé de la liste des éléments suivants :

1. ses régions basiques (c'est-à-dire son rectangle et ses courbes) ;
2. ses régions grisées ;
3. pour chaque suite de croix reliées, la plus petite région qui la contient entièrement.

Elle considère ensuite deux diagrammes comme syntaxiquement équivalents (je dirai : comme deux *tokens* du même *type*) si leurs listes sont les mêmes [51]. Le résultat est de considérer les diagrammes de la fig. 7.24 comme des *tokens* du même *type*, comme le sont par exemple les deux formules

$$\mathrm{A} \wedge \mathrm{B} \qquad \text{et} \qquad \mathrm{A} \wedge \mathrm{B}.$$

Au contraire, les diagrammes de la fig. 7.25 sont certes *sémantiquement* équivalents, mais ne sont pas des *tokens* d'un même *type* : leur cas serait donc plutôt analogue à celui des formules $\mathrm{A} \wedge \mathrm{B}$ et $\mathrm{B} \wedge \mathrm{A}$. Il peut sembler étonnant de traiter ces deux formules comme de *types* distincts et les deux premiers diagrammes de la fig. 7.24, pourtant pas plus différents, comme deux *tokens* du même *type*. Mais à y bien regarder, l'identité de *type* entre formules des langages formels repose elle aussi sur des habitudes et des conventions, qui nous font négliger des différences typographiques parfois considérables et, au contraire, nous font insister sur une distinction essentielle entre des formules comme $x + y$ et $y + x$ dont un

51. Voir Shin 1994, section 3.3.2, en part. p. 70. En fait, il faut faire une correction mineure à la définition de Shin ; celle-ci comporte en effet une bizarrerie qui s'explique par le fait que Shin ne pensait manifestement pas faire jouer le rôle de distinction *type-token* à sa notion d'équivalence syntaxique. Le problème est que dans son système, une même région minimale peut comporter plusieurs croix, mais celles-ci sont redondantes sémantiquement si elles ne font pas partie de suites de x différentes (c'est-à-dire si elles ne sont pas reliées à des croix différentes dans d'autres régions). De même, on peut avoir plusieurs suites de croix qui relient exactement les mêmes régions minimales. Or quand Shin définit la « suite » d'un diagramme, c'est-à-dire ce qu'elle voit comme la liste de ses éléments syntaxiques pertinents, elle ne compte qu'une croix ou qu'une suite de croix à chaque fois. Deux diagrammes qui diffèrent par le nombre de croix dans une certaine région minimale, par exemple, seront donc considérés comme syntaxiquement équivalents (voir l'exemple qu'elle donne p. 77). Il faut corriger cela si l'on veut faire jouer à l'équivalence syntaxique le rôle d'une distinction *type*-token.

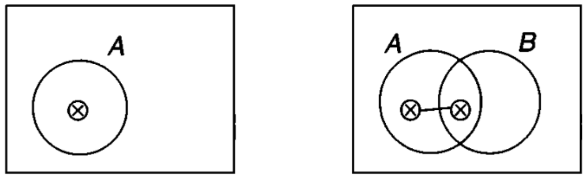


FIGURE 7.25 – Diagrammes sémantiquement, mais non syntaxiquement équivalents, adaptés de Shin (1994, p. 74)

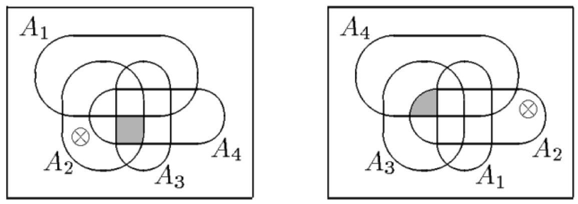


FIGURE 7.26 – Diagrammes à quatre courbes syntaxiquement équivalents, adaptés de Scotto di Luzio (2002)

mathématicien comme Descartes aurait sans doute eu du mal à comprendre l'intérêt. La notion de *type* de diagramme définie ici ne semble donc pas insensée.

Malheureusement, cette distinction *type-token*, qui fonctionne bien tant qu'on s'en tient à des diagrammes à deux ou trois courbes, pose problème dès qu'on veut aller plus loin. Tout d'abord, reconnaître que deux diagrammes comme ceux de la fig. 7.26 sont du même *type*, quoique possible, devient plus difficile. Surtout, deux diagrammes du même *type* peuvent être topologiquement différents : les régions colorées en rouge et en vert sur les deux diagrammes de la fig. 7.27, par exemple, se correspondent [52] mais sont contiguës dans un cas et non dans l'autre. Ces différences topologiques ont des conséquences sur l'utilisation du système. Comme l'a remarqué Scotto di Luzio (2002), l'une des règles de transformation de Shin peut induire en erreur dès qu'on a quatre courbes ou plus. Cette règle permet d'effacer une courbe quelconque d'un diagramme. Essayons de l'appliquer pour supprimer la courbe A_2 des deux diagrammes précédents, en procédant naïvement, c'est-à-dire en les recopiant (sans les réorganiser) à l'exception de A_2 (fig. 7.28). À droite, il n'y a pas de problème, mais à gauche, on se retrouve avec un diagramme mal formé parce que certaines des intersections possibles se retrouvent en deux morceaux : en notant (par abus de notation) A_i les intérieurs des courbes du même nom, $A_1 \cap \overline{A_3} \cap A_4$ correspond par exemple aux deux régions marquées en rouge, et $\overline{A_1} \cap \overline{A_2} \cap A_4$ aux deux régions marquées en vert.

52. Dans chacun des deux diagrammes, la région rouge correspond à $A_1 \cap A_2 \cap \overline{A_3} \cap A_4$ et la région verte à $A_1 \cap \overline{A_2} \cap \overline{A_3} \cap A_4$ (en employant, par abus de langage, la même notation pour les courbes et pour leur intérieur).

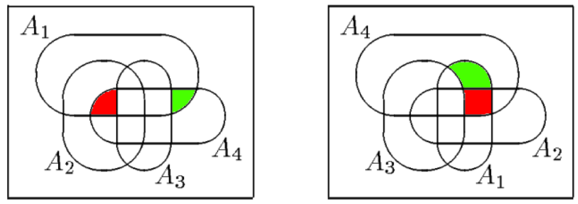


FIGURE 7.27 – Diagrammes syntaxiquement équivalents mais topologiquement différents, adaptés de Scotto di Luzio (2002)

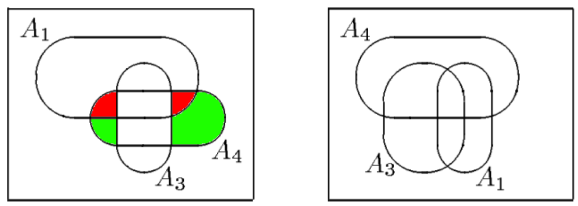


FIGURE 7.28 – Diagrammes précédents après suppression de la courbe A_2, adaptés de Scotto di Luzio (2002). Le diagramme de gauche n'est pas bien formé parce que les régions rouges et vertes ne sont pas connexes.

(Le diagramme de droite, lui, comporte bien seulement $2^3 = 8$ régions.) En fait, la distance entre *type* et *token* devient intolérable dès lors que l'on va au-delà de trois courbes.

Pour régler ce problème, Howse, Molina, Shin et J. Taylor (2002) introduisent une distinction entre règles de transformation au niveau des *tokens* et règles au niveau des *types*. Leur idée est que, contrairement à la plupart des règles de Shin, la règle d'élimination des courbes ne peut être correctement définie qu'au niveau des *types* : il faudrait d'abord passer du *token* particulier que l'on considère au diagramme abstrait correspondant, éliminer une courbe, puis tracer un *token* bien formé correspondant au diagramme abstrait résultant, qui peut ne pas ressembler du tout au *token* de départ.

Leur emploi ici de la terminologie *type-token* me semble inadapté. Le rôle fondamental que doit jouer une distinction *type-token* dans le cadre d'un système formel est de garantir l'uniformité de procédure en indiquant quelles inscriptions doivent être traitées comme identiques. En toute rigueur, une règle « au niveau des *tokens* » est donc une contradiction dans les termes, puisque par définition une règle doit pouvoir être appliquée de manière uniforme à des *tokens* différents. Dès lors que l'on a plus de quatre courbes, la distinction définie plus haut introduit trop de distance entre prétendus *types* et prétendus *tokens* pour qu'on puisse continuer à employer cette terminologie de manière cohérente. Je suggère donc d'employer une terminologie différente dont ils font également usage, et de parler de *diagrammes abstraits* contre *diagrammes concrets*.

Reformulons leur idée dans cette nouvelle terminologie. Certaines règles peuvent être formulées au niveau des diagrammes concrets, écrivent-ils, mais d'autres doivent passer par les diagrammes abstraits. Examinons ces deux cas tour à tour. Une règle sur les diagrammes concrets, ce serait, comme dans le système de Shin, une règle qui implique de « recopier » le diagramme en ajoutant ou supprimant certains éléments, mais sans changer la disposition des autres courbes. Le problème est ici le même que plus haut : on n'a pas de critères clairs sur ce qu'est une bonne copie, et le piège des diagrammes topologiquement différents montre que l'on ne peut pas utiliser notre première tentative de distinction *type-token*. On peut s'en sortir en introduisant une nouvelle distinction *type-token* plus fine ; une idée pourrait être de ne considérer comme du même *type* que les diagrammes de Venn topologiquement équivalents, c'est-à-dire que l'on peut continûment déformer l'un en l'autre. On aurait alors une nouvelle distinction *type-token*, coïncidant avec celle introduite plus haut jusqu'à trois courbes, mais plus fine au-delà. À partir de quatre courbes, il serait en effet possible pour des diagrammes « syntaxiquement équivalents » au sens initial de Shin – c'est-à-dire pour des instances d'un même diagramme abstrait – d'être pourtant de *types* distincts. Certaines règles de transformation seraient définies de manière à ne pas toucher à la disposition des courbes, et on pourrait donc les considérer comme des règles sur les diagrammes concrets bien qu'elles ne soient pas définies au niveau des *tokens*, ce qui n'aurait guère de sens. En fait, une fois cette nouvelle distinction *type-token* en main, ce sont tout simplement des règles au sens le plus habituel.

Que faire alors des autres règles, celles qui, d'après Howse, Molina, Shin et J. Taylor (2002), ne peuvent être formulées qu'au niveau des diagrammes abstraits ? L'exemple qu'ils donnent est bien sûr la règle d'élimination des courbes, mais il y en a d'autres, comme la règle d'unification (permettant de combiner deux diagrammes en un seul). Pour eux, ces règles-là impliquent de faire un détour par une représentation abstraite, qui est invariante par rapport aux particularités topologiques du diagramme concret considéré. Mais s'il en est ainsi, peut-on encore les considérer comme formelles ? À tout le moins, il faudrait les décomposer, et expliquer en détail comment on passe (sur la seule base de l'apparence du diagramme) du diagramme concret à sa description abstraite, comment on transforme cette description abstraite, puis comment on construit un nouveau diagramme. Si l'on procédait ainsi, on aurait une authentique règle formelle, mais au prix d'un nouveau problème : elle ne serait plus vraiment diagrammatique !

Une solution différente permet d'éviter cette conclusion désagréable : on peut introduire une règle de *réécriture* permettant de passer d'un diagramme à n'importe quelle autre instance du même diagramme abstrait. Si passer d'un diagramme concret à sa représentation abstraite implique de quitter le domaine du diagrammatique, en revanche évaluer si deux diagrammes correspondent au même diagramme abstrait peut se faire de manière formelle sur la seule base des diagrammes, en mettant en correspondance les courbes deux à deux et en examinant si cette correspodance préserve les grisures, etc. Les règles problématiques, comme celle d'élimination des courbes, pourraient alors être restreintes aux cas où elle ne posent pas de problème topologique. En fin de compte, on aurait un système diagrammatique et formel et la distinction entre deux sortes de règles ne serait plus nécessaire.

Ces difficultés peuvent donc être résolues. Au fond, elles viennent du fait que le système de Shin n'est jamais utilisé en pratique pour plus de quatre courbes ; le problème de la formalisation se superpose donc à la simple question de savoir comment procéder. Une autre leçon de ces difficultés est qu'il est hasardeux de définir *types* et *tokens* par stipulation : une telle distinction n'est opérante que s'il est facile de traiter les *tokens* d'un même *type* de manière invariante, ce sur quoi on ne peut légiférer sans expérience du système considéré.

b) Ces systèmes sont-ils formels ?

À première vue, il semble clair que les systèmes de Shin sont *formels* au sens minimal suivant : les règles de formation et de transformation données par Shin ne font référence qu'à l'*apparence* des diagrammes et pas à leur signification [53] ; elles sont rédigées avec précision, manifestement dans le but qu'elles puissent être suivies par quelqu'un qui n'aurait aucune notion de logique et ne saurait pas comment ces diagrammes sont censés être interprétés. John Macfarlane, dans son étude sur la formalité de la logique, nomme « formalité syntaxique [54] » ce sens du terme.

Il faut toutefois faire une distinction supplémentaire. La formalité de Shin est « hu-

53. Un problème particulier est posé par la relation de « contrepartie » (cf. section 7.2.b)) qui, dans le système original de Shin, associe certaines courbes de différents diagrammes (celles qui sont destinées à représenter le même ensemble, quoiqu'il n'y ait bien sûr pas besoin de faire référence à ce concept sémantique pour comprendre l'idée d'une relation entre courbes). La difficulté est que Shin considère cette relation comme faisant partie de la syntaxe de son système alors qu'elle n'est pas visible sur les diagrammes. On a donc besoin de quelque chose de plus que la simple apparence des diagrammes pour appliquer les règles correctement. Toutefois, ce problème n'est pas très profond. D'une part, cette relation relie des éléments visibles des diagrammes et ne fait donc pas entrer de concepts sémantiques par la petite porte. D'autre part, on pourrait la remplacer par l'usage de lettres-étiquettes directement visibles sur les diagrammes, comme le font en général les variantes ultérieures du système de Shin (Hammer 1995 ; Hammer et Danner 1996, Voir). Ce n'est donc pas une menace sérieuse pour la formalité du système.

54. MacFarlane 2000, section 2.1. Il range cette notion de formalité dans la catégorie des « leurres » (« *decoys* »), non parce qu'elle serait inutile ou incohérente, mais seulement parce qu'elle ne donne pas de critère de logicité : on peut avoir des systèmes « formels » (au sens syntaxique) destinés à capturer les mouvements autorisés aux échecs, par exemple, donc la formalité en ce sens est inutile pour délimiter la logique.

maine » : l'important est qu'un être humain puisse comprendre et appliquer les règles sur la seule base de la forme des diagrammes. Or nous avons l'habitude d'exiger davantage de nos langages formels ; nous souhaitons qu'ils puissent être manipulés non seulement par un être humain ignorant tout de leur signification, mais aussi par une machine. Cette exigence est fréquemment interprétée à travers la théorie de la calculabilité, qui cherche à réduire toutes les opérations mécanisables à des transformations élémentaires de suites finies de symboles [55]. Les diagrammes posent alors un problème spécifique, puisque ce ne sont pas des suites de symboles et que leurs règles de manipulation reposent sur des opérations plus complexes que simplement reconnaître les instances d'un symbole ; dans le cas des diagrammes de Shin, il faut par exemple reconnaître les inclusions de régions. A priori, le critère de calculabilité mécanique ne leur est donc pas applicable. Le problème est qu'on y perd en précision : la formalité « humaine », contrairement à la calculabilité, ne se laisse pas délimiter rigoureusement.

Faut-il s'en tenir là, ou peut-on dire que les systèmes de Shin sont formels en un sens plus serré ? Dans son travail sur la géométrie d'Euclide, Nathaniel Miller propose une stratégie pour réconcilier diagrammes et critère de calculabilité, en prenant d'ailleurs les diagrammes de Venn-Peirce comme exemple :

> The simplest possible definition that we can adopt here is that a system is completely formal if it can be completely implemented on a computer. [...] [This] means that the objects manipulated by the formal system can be anything at all as long as they can be translated into finite objects that a computer can manipulate. In the case of Euclidean geometry, this means that our formal system can manipulate diagrams, as long as we have some way representing the diagrams as finite objects in a computer. [...] For example, a Venn diagram consists of a finite number of regions each of which can contain an x, an o [56], both, or neither, and in which some of the xs can be connected to each other. If we know what is in each region and how they are connected, then we have all the information contained in the diagram. It is clear that we can capture all of this information in a finite array. [57]

55. Comme le souligne Dutilh Novaes 2012, chap. 1, il faut donc distinguer deux notions au sein de ce que Macfarlane appelle « formalité syntaxique » : la formalité comme « dé-sémantification » (c'est-à-dire que les manipulations se font sans référence au sens) et la formalité comme calculabilité.

56. Miller fait ici référence aux symboles que Peirce utilise à la place des grisages ; cf. p. 188 n. 19.

57. « La définition la plus simple possible que nous puissions adopter ici est qu'un système est complètement formel s'il peut être complètement implémenté dans un ordinateur. [...] [Cela] veut dire que les objets manipulés par le système formel peuvent être absolument n'importe quoi, dès lors que l'on peut les traduire en objets finis qui peuvent être manipulés par un ordinateur. Dans le cas de la géométrie euclidienne, cela veut

À vrai dire, les choses sont moins claires que ne l'affirme Miller. Certes, étant donné un diagramme de Venn-Peirce, ce que Shin appelle sa « suite » (la liste de ses régions, grisures et suites de croix, cf. sous-section précédente) en contient « toute l'information » au sens où elle suffit à en définir la sémantique. Il est moins clair, en revanche, qu'elle lui soit informationnellement équivalente en un sens plus strict comme celui de Simon : s'il y a trois courbes ou moins, elle suffit certes à reconstruire, sinon le diagramme lui-même, du moins un diagramme du même *type* ; mais au-delà de quatre courbes, comme nous l'avons vu, il y a des différences topologiques qui peuvent jouer un rôle pour l'application des règles et que pourtant la « suite » fait disparaître. Miller, sans doute, n'était pas conscient de cette subtilité. Dans tous les cas, cela ne suffit pas à remettre en cause la légitimité générale de son critère. Si l'on s'en tient aux diagrammes de moins de trois courbes qu'on utilise le plus souvent, sa remarque est légitime : on peut remplacer les diagrammes par leurs suites et formuler les règles de transformation au niveau des suites plutôt que des diagrammes eux-mêmes.

La difficulté de fond, cependant, est différente. Si l'on s'intéresse au système diagrammatique de départ, ce critère de formalité par traduction ne fait que déplacer le problème. Il revient à dire qu'on peut *juxtaposer* au système diagrammatique de départ un système formel au sens le plus classique du terme, constitué de suites de symboles et de règles définies sur celles-ci. Mais le lien entre les diagrammes et leur traduction symbolique ne peut être défini qu'informellement, ou plutôt selon le genre de formalité humaine qui caractérise le système de Shin : pour traduire un diagramme de Venn-Peirce en sa suite, par exemple, il faut l'inspecter, reconnaître des inclusions de régions, etc. En d'autres termes, pour reconnaître la légitimité de la traduction symbolique proposée, il faut avoir exactement les capacités humaines mal délimitées qu'on cherchait à éliminer de l'analyse. La calculabilité permet de juger la formalité d'une reconstruction non diagrammatique du système de départ, et non celle du système diagrammatique lui-même. Pour cela, il faut en rester à un critère de formalité plus vague, qui se contente d'exiger que les manipulations ne fassent référence qu'à l'apparence des diagrammes.

dire que notre système formel peut manipuler des diagrammes, tant que nous avons une manière de représenter les diagrammes comme des objets finis dans un ordinateur. [...] Un diagramme de Venn, par exemple, consiste en un nombre fini de régions dont chacune peut contenir un *x*, un *o*, les deux, ou ni l'un ni l'autre, et où certains des xs peuvent être reliés. Si nous savons ce qu'il y a dans chaque région et comment les différents éléments sont reliés, alors nous avons toute l'information contenue dans le diagramme. Il est clair que nous pouvons capturer toute cette information dans un tableau fini. » (Miller 2007, p. 13).

7.5 Conclusion : diagrammes, formalisation et rigueur

Le travail de Shin montre que l'on peut étendre la sémantique en termes de modèles au-delà des langages formels.

Le but principal de Shin, cependant, est plus large : elle veut montrer qu'on peut raisonner rigoureusement avec des diagrammes. À la suite de Barwise et Etchemendy ([1991] 1996b), elle cherche ainsi à contrer une « attitude négative envers les diagrammes[58] » de la part des logiciens et mathématiciens, qui cantonneraient les diagrammes à un rôle heuristique et rejetteraient leur emploi dans les preuves.

Mais réussit-elle ? Elle montre de manière convaincante qu'on peut raisonner *sans risque d'erreur* avec les diagrammes de Venn-Peirce. Cet argument convaincrait certainement un mathématicien qui rejetterait en bloc de ses preuves tout ce qui ne ressemble pas une suite de mots ou de symboles. Toutefois, il me semble que cette attitude est rare ; c'est d'ailleurs le reproche que Larvor (1996) adresse à Shin. On rencontre plus fréquemment une attitude sélective, qui emploie sans hésitation certaines sortes de diagrammes (diagrammes commutatifs en algèbre et théorie des catégories, par exemple, ou diagrammes de nœuds en topologie) et en traite d'autres avec davantage de méfiance (comme en géométrie et en analyse). Dans un tel contexte, le travail de Shin montre seulement que les diagrammes de Venn-Peirce sont à ranger parmi ceux qui sont sans danger. Ce qui pose ici problème, c'est la notion même de diagramme, qui n'est pas bien définie : peut-être est-il plus pertinent, précisément parce qu'on peut les utiliser de manière rigoureuse, de rapprocher les diagrammes de Venn-Peirce des formules que des diagrammes euclidiens. En l'absence d'une distinction claire entre représentations diagrammatiques et symboliques, le projet général de Shin perd son sens. C'est l'une des raisons pour lesquelles nos auteurs, à commencer par Barwise et Shin, chercheront par la suite à clarifier cette distinction. J'y reviendrai au chapitre 10.

58. Shin 1994, p. 1.

Chapitre 8

Les diagrammes comme texte : un cadre général

Une idée particulièrement importante est sous-jacente aux systèmes diagrammatiques du chapitre précédent : c'est celle de traiter les diagrammes, à égalité avec le texte, comme des représentations susceptibles d'avoir une sémantique et de permettre des inférences.

Sans doute l'idée n'est-elle pas neuve : Ken Manders et John Mumma, qui comme nous le verrons développent des idées apparentées [1], invoquent respectivement le précédent de Leibniz [2] et de Wittgenstein [3] ; et l'on pourrait mentionner bien d'autres auteurs [4].

1. Voir *infra*, section 8.2 ; concernant Manders, cf. aussi chapitre 12.

2. Manders (1996, p. 391) renvoie à un dialogue de Leibniz de 1677 : « A. [...] si characteres abessent nunquam quicquam distincte cogitaremus, neque ratiocinaremur. B. At quando figuras Geometriae inspicimus saepe ex accurata earum meditatione veritates eruimus. A. Ita est, sed sciendum etiam has figuras habendas pro characteribus, neque enim circulus in charta descriptus verus est circulus, neque id opus est, sed sufficit eum a nobis pro circulo haberi. » (« A. [...] s'il n'y avait pas de caractères, nous ne penserions jamais rien distinctement, ni ne raisonnerions. B. Pourtant, quand nous inspectons les figures de la Géométrie, il arrive souvent que nous découvrions des vérités par une réflexion soigneuse sur elles. A. Certes, mais il faut savoir que ces figures aussi doivent être considérées comme des caractères ; en effet, le cercle tracé sur du papier n'est pas un vrai cercle et il n'est pas nécessaire qu'il le soit, mais il suffit que nous le tenions pour un cercle. », LPS, VII, p. 191 = LAA, VI.4, p. 23).

3. Mumma (2006, p. 14) signale un passage d'un cours de Wittgenstein : « The figure of the Euclidean proof [of the constructibility of a regular pentagon] as used in mathematics is just as rigorous as writing—because it has nothing to do with whether it is drawn well or badly. The main difference between a proof by drawing lines and a proof in writing is that it doesn't matter how you draw lines, or whether the r's and l's and m's and e's are written well. » (« La figure de la preuve Euclidienne (de la constructibilité d'un pentagone régulier) telle qu'on l'utilise en mathématiques est tout aussi rigoureuse que l'écriture – parce qu'elle n'a rien à voir avec la question de savoir si elle est bien ou mal dessinée. La différence principale entre une preuve faite en dessinant des droites et une preuve faite en écrivant est que [dans le premier cas], la manière dont on dessine les droites n'a pas d'importance, [alors que dans le second cas] si on écrit bien les r et l et m et e n'a pas d'importance. », Wittgenstein 1976, p. 134, traduction libre pour clarifier le sens).

4. Dans le cas des diagrammes logiques, nous avons recontré l'idée assez clairement chez Venn, lorsqu'il dit

Cependant, généralisée et combinée avec une sémantique en termes de modèles, cette idée permet d'avancer sur les problèmes qui guident ce travail. Elle fournit une première manière de décrire précisément comment deux représentations qui, en un sens, « parlent de la même chose » peuvent avoir un contenu différent, et offre un modèle pour comprendre comment ce genre de différences représentationnelles peut permettre de progresser.

8.1 De *Tarski's World* à *Hyperproof*

Pour mieux comprendre l'idée de traiter les diagrammes comme du texte, revenons tout d'abord en arrière. La première étape du projet de Barwise et Etchemendy sur le raisonnement diagrammatique, avant les travaux de Shin, est le système *Hyperproof*, qu'ils développent à partir de 1988 pour l'enseignement de la logique. Retracer leur cheminement jusqu'à ce système permet de mieux comprendre l'important changement de perspective qui les conduit à leur vision des diagrammes.

Au départ, Barwise et Etchemendy conçoivent en effet leurs diagrammes comme *ce dont* parlent les énoncés de leur système logique, c'est-à-dire comme une sorte de sémantique. À l'issue de leur travail sur *Hyperproof*, ils y voient plutôt des représentations qui, à l'instar des énoncés de la logique du premier ordre, servent à parler d'autre chose : le diagramme n'est plus objet du discours, mais fait lui-même partie du discours ; il n'est plus la sémantique, mais admet lui-même une sémantique.

a) *Tarski's World* ou les diagrammes comme sémantique

À l'origine d'*Hyperproof*[5], il y a le logiciel *Tarski's World*, mis au point par Barwise et Etchemendy à Stanford à partir du milieu des années 1980. L'université de Stanford, située

rechercher des diagrammes qui ne soient pas le reflet exact de certaines relations entre classes, mais expriment la « connaissance imparfaite » que nous pouvons en avoir (cf. *supra*, p. 186) ; Peirce en comprenait parfaitement les enjeux, jusqu'à mettre au point divers systèmes logiques tant symboliques que diagrammatiques pour lesquels il formule des règles de transformation explicites. Dans un autre registre, le théoricien W. J. T. Mitchell a pu écrire dans les années 1980 que « le lieu commun des travaux modernes sur les images est qu'elles doivent être comprises comme une sorte de langage » (Mitchell 1986, p. 8), dans des domaines aussi différents que la théorie de l'art, la critique littéraire et jusqu'à la philosophie analytique à travers l'œuvre de Nelson Goodman (dans cette littérature, comme le suggère Mitchell, l'assimilation des images au texte répond souvent au souhait de dévoiler, sous la transparence trompeuse des images, des dispositifs de représentation conventionnels voire idéologiquement chargés). Du reste, ranger phrases et images dans la même catégorie n'est pas inhabituel non plus dans le contexte intellectuel analytique et imprégné de cognivisme de Barwise et Etchemendy : comme nous l'avons vu dans la première partie, c'est pour les sciences cognitives des années 70 et 80 une question centrale que de savoir sous quelle forme le cerveau représente le monde, et un emploi très large du terme de représentation est omniprésent dans ces travaux-là.

5. Mon aperçu historique s'appuie essentiellement sur le récit rétrospectif de Barwise et Etchemendy 1998.

en pleine « Silicon Valley » et étroitement associée à l'informatique naissante, employait des logiciels d'enseignement de la logique depuis les années 1960. Développés sous l'impulsion du philosophe Patrick Suppes, ceux-ci portaient sur les aspects syntaxiques de la logique : application des règles d'inférences, construction de preuves. Le but de *Tarski's World* était d'y ajouter un enseignement de concepts sémantiques comme la vérité dans un modèle. Publié pour la première fois en 1987 et intégré depuis lors à plusieurs manuels de logique [6], ce logiciel reste populaire aux États-Unis. Aujourd'hui encore, il est régulièrement mis à jour par l'ancienne équipe de recherche de Barwise et Etchemendy à Stanford [7].

Tarski's World permet d'interagir avec un langage du premier ordre simple. Les structures d'interprétation de ce langage sont des « mondes » de 8 cases sur 8, faciles à représenter diagrammatiquement (figure 8.1), dont chaque case peut être occupée par un objet caractérisé par sa forme – tétraédrique, cubique ou dodécaédrique – et sa taille – petit, moyen ou grand. Les objets représentés sur les diagramme peuvent en outre porter des étiquettes, qui permettent de spécifier l'interprétation de symboles de constante a, b, c, etc. Le langage du premier ordre associé comprend divers prédicats et relations pour parler de ces mondes, par exemple le prédicat unaire « x est un tétraèdre » ou les relation binaires « x est à gauche de y », « x est plus grand que y ».

L'intérêt pédagogique du logiciel est qu'on peut y poser des exercices. Les plus simples demandent aux étudiants de dire si un certain énoncé est vrai dans un monde affiché à l'écran [8], ou de construire un monde vérifiant ou falsifiant certains énoncés ; le logiciel permet de vérifier immédiatement les réponses. Des exercices plus sophistiqués permettent d'apprendre à traduire des phrases de la langue naturelle en énoncés du premier ordre : étant donné par exemple « tous les objets sont des cubes ou sont à côté d'un tétraèdre », le but du jeu est de trouver une traduction formelle ayant les mêmes valeurs de vérité dans différents mondes, le logiciel permettant d'expérimenter. Le manuel de *Tarski's World*

6. D'abord testé à l'université de Stanford, le logiciel *Tarski's World* est diffusé en 1987 pour la première fois (Barwise et Etchemendy 1987a). Il est ensuite intégré aux éditions successives du manuel Barwise et Etchemendy 1990b, puis au manuel Barwise et Etchemendy 1999, qui reste régulièrement réédité et est utilisé dans de nombreuses universités américaines. À côté de ces manuels de logique, il existe aussi une version séparée du logiciel, qui n'est accompagnée que d'un petit livre explicatif (Barwise et Etchemendy 1991 ; Barker-Plummer, Barwise et Etchemendy 2008). On peut également acheter le logiciel seul.

7. Il s'agit du « *Center for the Study of Language and Information* » (CSLI). John Etchemendy en est encore membre et supervise l'effort de développement des logiciels pédagogiques, quoique ses responsabilités administratives ne lui laissent guère le temps d'y travailler lui-même (il est « *Provost* », c'est-à-dire numéro 2, de l'université de Stanford depuis 2000).

8. Si l'étudiant se trompe en déclarant vrai un énoncé qui est faux dans le monde présenté (par exemple), le logiciel sait décomposer cet énoncé en composants de plus en plus simples, demandant à chaque fois à l'étudiant s'ils sont vrais ou faux. C'est un genre de jeu, qui s'arrête lorsque l'erreur de l'étudiant lui devient manifeste.

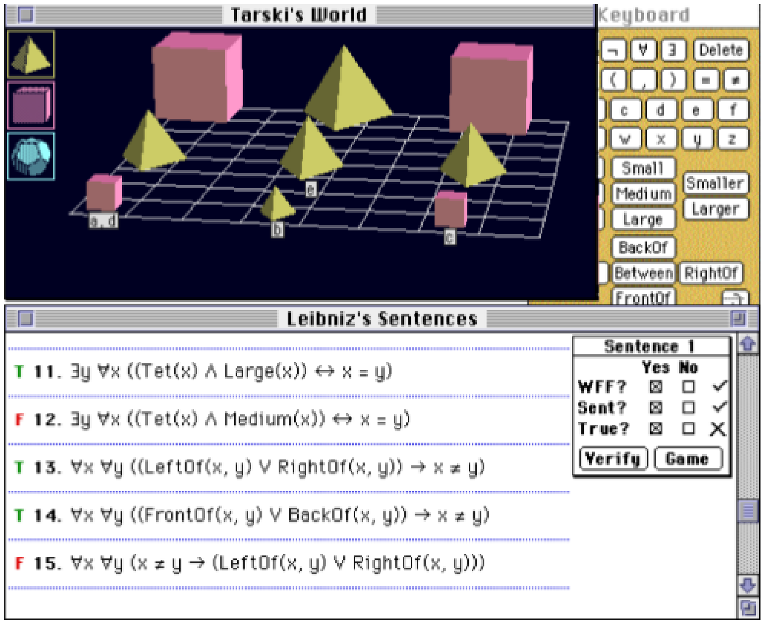


FIGURE 8.1 – Capture d'écran de *Tarski's World* dans une version pour Macintosh des années 1990, d'après Barwise et Etchemendy (1998, p. 98)

propose de nombreux exercices de ce genre.

À première vue, il n'y a là rien de révolutionnaire. On a d'un côté un langage du premier ordre, de l'autre ses modèles, ici représentés par des diagrammes affichés à l'écran. Le langage du premier ordre vise à modéliser le raisonnement sous forme d'inférences entre énoncés. Les diagrammes, de leur côté, tiennent lieu des modèles de ce langage. En plus de permettre l'enseignement de la sémantique, ils rendent donc possible de démontrer des résultats inaccessibles dans le langage associé, comme la cohérence de certains énoncés ou le fait qu'un énoncé ne découle pas de certains autres. Énoncés et diagrammes (c'est-à-dire modèles) jouent donc des rôles fondamentalement différents.

Mais certains exercices que posent Barwise et Etchemendy dans *Tarski's World* les conduisent à rejeter cette distinction nette entre énoncés et diagrammes. Outre les exercices déjà mentionnés, comme « cet énoncé est-il vrai de ce diagramme ? », le logiciel permet en effet aussi d'en poser d'autres, qui exigent un raisonnement plus complexe. Par exemple, étant donnée la figure 8.2, et sachant que « les cases voisines de *b* sont vides », « *c* n'est pas un cube et n'est pas de taille moyenne », « *b* est à l'arrière des deux tétraèdres » et « *b* est plus grand que *c* », pouvez-vous déterminer la forme de *c* ? Un raisonnement simple, qui s'appuie sur le diagramme, montre que *b* est le dodécaèdre au fond à droite, et que *c* ne peut être que l'un des deux petits dodécaèdres (fig. 8.3). On peut donc conclure que *c*

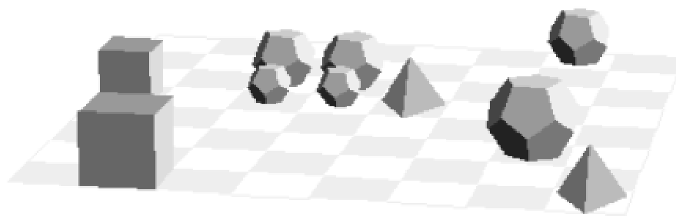


FIGURE 8.2 – Un monde de *Tarski's World* dans lequel il faut identifier deux objets b et c (d'après Barwise et Etchemendy 1998, p. 99)

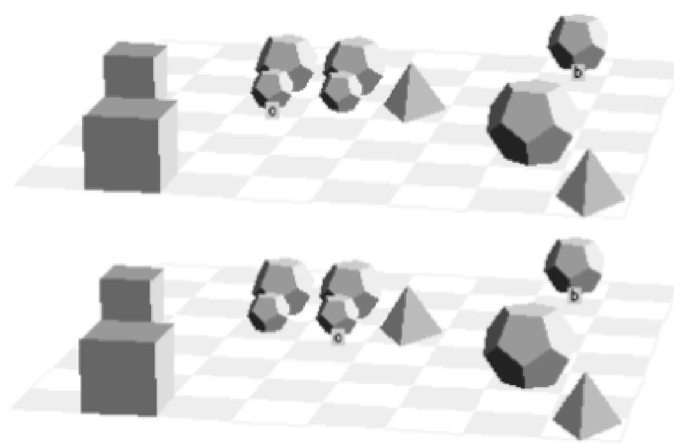


FIGURE 8.3 – Les deux identifications possibles de c (d'après Barwise et Etchemendy 1998, p. 101)

est un dodécaèdre. Dans cet exercice, les diagrammes semblent jouer un rôle essentiel, non pour prouver une propriété métathéorique comme la cohérence ou la non-conséquence, mais au cours d'un raisonnement banal qui part de données et en déduit une conséquence.

Face à ce constat, deux réponses sont possibles. La première est de remplacer le diagramme de départ par des énoncés (par exemple « il y a deux tétraèdres de taille moyenne, 5 dodécaèdres dont deux petits, trois moyens et un grand et deux cubes dont un moyen et un grand, et rien d'autre ; l'un des tétraèdres moyens est derrière tous les autres objets ; etc. »), et de maintenir que tout le raisonnement peut être capturé par des inférences entre énoncés. Barwise et Etchemendy rejettent cette solution, principalement parce qu'elle conduit à des formalisations vraiment trop éloignées du raisonnement original. Ils avancent deux arguments. Tout d'abord, « il faut faire beaucoup de choix arbitraires pour passer d'une représentation visuelle à un ensemble d'énoncés plus ou moins équivalent [9] », et pour chacun de ces choix, « une mauvaise décision rendrait le problème insoluble, donc en pratique la conversion vers la logique traditionnelle ne pourrait se faire qu'*après* que le problème a été résolu [10] ». En second lieu, les méthodes suivies en raisonnant sur le diagramme ne se traduisent pas fidèlement :

> If you examine the steps in the above reasoning [11], it becomes evident that the methods are quite different from those we usually teach in logic. They have a distinctly "semantic" character [...]. For example, you are as likely to break into cases based on atomic, negated or quantified sentences as you are on the basis of a disjunctive claim. As a consequence, attempts to represent the reasoning using sentences do not faithfully model the real character of the reasoning. Simple proofs explode into proofs involving hundreds of steps, proofs in which the key steps in the original reasoning are obscured or obliterated. [12]

Barwise et Etchemendy choisissent donc une autre solution : ils mettent diagrammes et énoncés sur le même plan, les considérant comme deux manières différentes de représenter

9. « [T]here are many arbitrary choices to be made in going from a a visual representation to a roughly equivalent set of sentences » (Barwise et Etchemendy 1998, p. 102).

10. « In each case, the wrong decision would make the problem unsolvable, and so practically speaking, the conversion to traditional logic could only be accomplished *after* the problem had been solved. » (Barwise et Etchemendy 1998, p. 102 ; les auteurs soulignent).

11. Ils parlent de la résolution d'exercices comme celui que j'ai présenté.

12. « Si vous examinez les étapes du raisonnement précédent, il vous apparaîtra que les méthodes en sont bien différentes de celles que nous enseignons habituellement en logique. Elles ont un caractère nettement "sémantique" [...]. Par exemple, vous êtes tout aussi susceptible d'introduire des distinctions de cas sur la base d'énoncés atomiques, niés ou quantifiés que sur la base d'une assertion disjonctive. Il en découle que les tentatives de représenter le raisonnement en utilisant des phrases ne modélise pas fidèlement le véritable caractère du raisonnement. Des preuves simples explosent en preuves ayant des centaines d'étapes, dans lesquelles les étapes-clé du raisonnement original sont cachées ou éliminées. » (Barwise et Etchemendy 1998, p. 102).

de l'information sur ce dont on veut parler (en l'occurence des mondes de 8 cases sur 8). Certes, on peut traduire les diagrammes en énoncés, mais, écrivent nos auteurs, « nous nous sommes aperçus qu'en fait, ces énoncés n'étaient que des *conséquences* du diagramme, et donc que l'inférence permettant de passer du diagramme aux énoncés *était elle-même affaire de logique*, qu'elle pouvait elle-même être valide ou invalide [13] ». En d'autres termes, on n'a plus de distinction entre le raisonnement au sens strict, qui ferait l'objet de la logique et consisterait en inférences entre énoncés, et une éventuelle lecture du diagramme qui pourrait fournir des prémisses aux inférences. L'usage du diagramme devient une étape du raisonnement à part entière et peut être étudié avec les outils de la logique.

Dans cette perspective, diagrammes comme énoncés peuvent être munis d'une sémantique sous la forme d'un *ensemble* de modèles. À y bien regarder, le diagramme de départ de l'exercice ci-dessus (fig. 8.2), par exemple, ne suffit pas à déterminer un modèle unique. Pour fixer un modèle bien défini, il faudrait en effet spécifier l'interprétation de tous les symboles de constante, or ce diagramme ne dit pas quels blocs sont *b* et *c*. Il ne correspond donc pas à *un* modèle particulier, mais plutôt à un *ensemble* de modèles, comme c'est en général le cas des énoncés d'un langage du premier ordre.

Entre diagrammes et énoncés subsiste pourtant une différence importante. Même si l'on incorpore les diagrammes à la syntaxe, ils fonctionnent encore comme des modèles au sens où on peut les utiliser pour démontrer la *cohérence* ou la *non-conséquence* de certains ensembles d'énoncés. (On peut par exemple démontrer qu'un certain ensemble d'énoncés est cohérent en construisant un diagramme qui les vérifie tous, ou qu'un énoncé ne découle pas de certains autres en construisant un diagramme qui vérifie ces derniers mais pas leur conséquence présumée.) L'explication qu'en donnent Barwise et Etchemendy est qu'au contraire des énoncés, les diagrammes ont une relation privilégiée à leurs modèles, plus précisément un isomorphisme [14]. En fait, c'est cette relation privilégiée qui, entre autres, distinguerait le graphique du linguistique. Je reviendrai brièvement sur cette idée au chapitre 10.

b) *Hyperproof* ou la sémantique des diagrammes

Certains exercices posés aux étudiants dans le cadre de *Tarski's World* conduisent donc Barwise et Etchemendy à voir les diagrammes autrement, non plus comme de simples

13. « we realized that those sentences were in fact merely consequences of the diagram, and consequently that the inference from the diagram to the sentences was itself a matter of logic, could itself be valid or invalid. » (Barwise et Etchemendy 1998, p. 103 ; je souligne).

14. Du moins en l'absence de symboles de constante, ou lorsque ceux-ci figurent tous sur le diagramme ; sinon, il y a plusieurs modèles possibles.

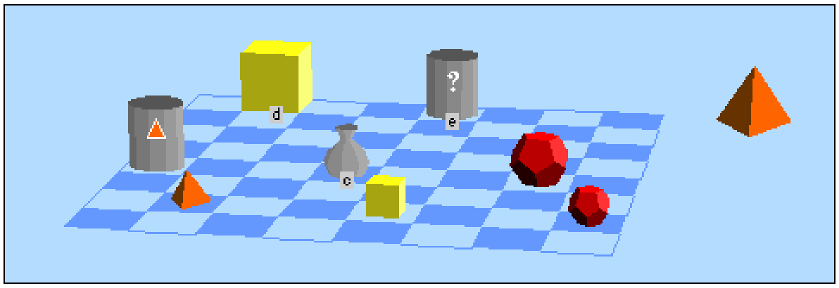


FIGURE 8.4 – Exemple de diagramme du système *Hyperproof*

images des modèles mais comme des représentations qui *elles-mêmes* délimitent des ensembles de modèles :

> When thinking about reasoning and problem solving, a useful metaphor is that of exploration: to solve a reasoning problem, we explore a space of possible situations or worlds consistent with the initial information we are given. [...] Sentences, whether of English or first-order logic, partition this space of possibilities, dividing it up into fiefdoms with a multitude of overlapping claims. [...] Diagrams, like sentences, carry information: they carve up the same space of possibilities, though perhaps in very different ways. [15]

Dans le cas de *Tarski's World*, le diagramme détermine un modèle quasiment unique : il spécifie la forme, la taille et la position de tous les objets. La seule chose qui peut rester indéterminée, comme nous l'avons vu, est l'interprétation d'éventuels symboles de constante, et cela uniquement dans le contexte très particulier de certains exercices de raisonnement. Pourtant, une fois qu'on traite le diagramme non comme un modèle, mais comme une représentation munie de modèles, rien n'empêche d'utiliser des diagrammes beaucoup plus incomplets.

C'est en tirant les conséquences de cette idée que Barwise et Etchemendy sont conduits à *Hyperproof*. Dans ce nouveau système, les diagrammes ne donnent plus qu'une information très partielle sur leurs modèles. Comme l'écrivent Barwise et Etchemendy,

> Tarski's World only allows you to construct 'complete' or 'total' worlds [16],

15. « Lorsque l'on réfléchit au raisonnement et à la résolution de problèmes, la métaphore de l'exploration est utile : résoudre un problème de raisonnement, c'est explorer un espace de situations ou mondes possibles cohérents avec les informations initialement données. [...] Les phrases, qu'elles soient en anglais ou en logique du premier ordre, partitionnent cet espace de possibilités, le divisant en fiefs dont les revendications territoriales se chevauchent. [...] Les diagrammes, comme les phrases, portent de l'information : ils découpent le même espace de possibilités, quoique peut-être de manière différente. » (Barwise et Etchemendy 1998, p. 109).

16. Il y a une exception : les exercices de raisonnement comme celui que j'ai présenté dans la section précédente, qui sont en fait posés aux étudiants en-dehors du logiciel.

> worlds in which every object has a particular size, shape, and location, and in which all the names in use are assigned to specific objects.
>
> In Hyperproof you reason about complete worlds, but the pictorial information you are given about the worlds is, in general, only partial. For example, you may know that there is an object located at a certain point on the grid, but not know the size or shape of that object. Or you may know the size and shape of an object, but not know where on the grid it is located. [17]

La fig. 8.4 fait voir les différentes manières dont les diagrammes d'*Hyperproof* peuvent représenter de l'incertitude. Le cylindre marqué d'un point d'interrogation est un objet dont on ignore la taille et la forme ; le cylindre marqué d'un triangle est un tétraèdre dont on ignore la taille ; le sac de petite taille est un objet de petite taille dont on ignore la forme ; le tétraèdre qui se trouve en-dehors de la grille est un objet dont on connaît forme et taille, mais dont on ignore la position.

Dans *Hyperproof*, les diagrammes font ainsi partie de la syntaxe du système et ont pour sémantique des ensembles de modèles, comme les énoncés du langage du premier ordre associé. Comme dans *Tarski's World*, les diagrammes gardent néanmoins une relation suffisamment étroite avec les modèles pour qu'on puisse les utiliser pour démontrer qu'un certain ensemble d'énoncés est cohérent.

Cette position ambiguë des diagrammes, intégrés à la syntaxe mais néanmoins plus proches des modèles que les énoncés, a une conséquence importante. Dans les présentations habituelles de la logique du premier ordre, il y a une asymétrie entre prouver qu'une conclusion découle de certaines hypothèses (ce qui peut se faire syntaxiquement, en construisant une preuve) et prouver que certaines hypothèses sont cohérentes (ce qui se fait sémantiquement, en construisant un modèle [18]). Dans *Hyperproof*, au contraire, les deux sortes de raisonnement sont sur le même plan. Les exercices posés au sein du système peuvent ainsi avoir les formes les plus diverses [19] : démontrer un énoncé donné à partir de

17. « *Tarski's World* ne permet de construire que des mondes "complets" ou "totaux", c'est-à-dire des mondes dans lesquels la taille, la forme et la position de chaque objet sont déterminés et dans lesquels tous les noms utilisés sont assignés à des objets particuliers. Dans *Hyperproof*, le raisonnement porte encore sur des mondes complets, mais l'information graphique donnée sur ces mondes n'est en général que partielle. Ainsi, on peut savoir qu'il y a un objet à une certaine position sur la grille, mais ne pas connaître sa taille ou sa forme. Ou alors, on peut connaître la forme et la taille d'un objet mais ne pas savoir où il se trouve sur la grille. » (Barwise et Etchemendy 1994, p. 4).

18. Cette remarque ne concerne que les présentations habituelles de la logique du premier ordre dans les manuels. En se donnant une machinerie adaptée, rien n'empêche de démontrer (syntaxiquement) dans un système du premier ordre que certains ensembles d'énoncés de ce système formel lui-même sont cohérents ; cette possibilité n'a rien à voir avec les diagrammes.

19. Une liste récapitulative des énoncés possibles est donnée en appendice du livre qui accompagne le logiciel : Barwise et Etchemendy 1994, app. 2, p. 239–243.

certaines prémisses, montrer qu'un énoncé donné ne découle pas de certaines prémisses, montrer qu'un ensemble d'énoncés est cohérent ou contradictoire, déterminer la forme ou la taille d'un objet, déterminer la référence d'un symbole de constante comme « a », etc.

Par ailleurs, *Hyperproof* est pleinement hétérogène : il permet d'utiliser les diagrammes au cours du raisonnement. Comme le système de Hammer pour les diagrammes de Venn-Peirce discuté plus haut [20], il admet en effet des règles d'inférence mélangeant diagrammes et énoncés. Une règle *Observer* permet ainsi de déduire un énoncé d'un diagramme, et une règle *Appliquer* d'enrichir un diagramme d'après un énoncé (si le diagramme comporte un sac c, c'est-à-dire un objet de forme indéterminée, l'énoncé « c est un tétraèdre » permet par exemple de remplacer ce sac par un tétraèdre de même taille). Ces règles sont spécifiées par des algorithmes : contrairement aux diagrammes de Venn-Peirce, les diagrammes dans *Hyperproof* se présentent en effet d'emblée comme des objets discrets, faciles à représenter par des suites de symboles et donc sur lesquels on peut sans problème définir des algorithmes au sens habituel.

8.2 Généralisation : les figures géométriques

L'idée de Barwise et Etchemendy est donc d'assimiler leurs diagrammes au texte. Il en découle deux conséquences importantes. D'une part, cela introduit une distance entre le diagramme et les objets du discours : le diagramme peut ne pas refléter exactement une configuration particulière d'objets, et ne donner que des informations partielles. Les diagrammes d'*Hyperproof*, par exemple, ne suffisent pas à déterminer un monde de blocs particulier mais ne font que restreindre les mondes possibles. D'autre part, cela conduit le chercheur à prêter une attention systématique aux transformations des diagrammes et aux transitions entre diagramme et texte, en les considérant toutes comme des inférences susceptibles d'être codifiées comme le sont les inférences entre énoncés.

Tout l'intérêt de ce changement de perspective apparaît quand on passe à des cas moins simples que celui d'*Hyperproof* ou des diagrammes de Venn. L'exemple des figures géométriques est particulièrement éclairant et a suscité une littérature assez abondante, qui commence elle aussi chez nos auteurs et dont il faut d'abord dire quelques mots. Les figures utilisées en géométrie élémentaire étant paradigmatiques de l'usage de diagrammes ou figures en mathématiques (et, depuis la fin du XIXe siècle, de leur usage répréhensible) Barwise et Etchemendy cherchent rapidement à s'y attaquer. Une étudiante de Barwise,

20. Cf. section 7.3.

Isabel Luengo, rédige sa thèse sur la question quelques années après Shin [21]. Il s'avère cependant que le problème est plus difficile que celui des diagrammes de Venn et que, pour des raisons fondamentales, le système de Luengo est incohérent [22]. À la même époque, Ken Manders (qui a certainement connaissance du regain d'intérêt pour le raisonnement diagrammatique en général, et des travaux de Barwise et Etchemendy en particulier [23]) écrit un texte non formel sur le rôle des figures dans la géométrie plane d'Euclide [24]. Plusieurs autres systèmes formels portant spécifiquement sur les premiers livres des *Éléments* d'Euclide voient le jour dans les années qui suivent [25] : celui de Miller (2001, 2007), dont le point de départ est le travail de Barwise, Etchemendy et Luengo ; celui de Mumma (2006, 2010), qui s'inspire plutôt des idées de Manders ; et celui d'Avigad, Dean et Mumma (2009), dérivé du précédent.

Examinons donc l'effet du changement de perspective de nos auteurs dans ce cas. Traditionnellement, on considère que les figures de la géométrie élémentaire exhibent les objets dont parlent les énoncés [26]. Certes, ils n'en sont qu'un reflet imparfait : nos traits sont épais, nos cercles ne sont pas exactement circulaires et nos droites, pas vraiment droites. Il n'en reste pas moins qu'à condition de ne pas se laisser tromper par ces imperfections, on peut considérer que notre dessin de triangle nous donne un accès fiable à un certain triangle idéal sur lequel on peut raisonner. Deux problèmes philosophiques principaux se posent alors : celui de la généralité et celui de la nature des objets géométriques. Comment puis-je démontrer un théorème général valable pour tous les triangles à partir d'une figure particulière, qui exhibe un triangle particulier possédant des propriétés que tous les triangles ne partagent pas ? Et quelle est la nature de ces triangles idéaux que mes figures sont censées refléter imparfaitement ?

Mettre figures et énoncés sur le même plan conduit à des questions toutes différentes. Le double effet du changement de persective opéré par nos auteurs est ici très clair. Tout

21. Voir Luengo 1995, 1996. (Comme Shin, elle se place pour simplifier dans un cadre purement diagrammatique, le but ultime restant un système hétérogène.)

22. Voir Miller 2007, appendice C.

23. Manders connaissait les travaux de Barwise, Etchemendy et leurs élèves. En fait, il développe une critique générale de la sémantique des situations (qui est l'un des points de départ du projet de nos auteurs sur le raisonnement diagrammatique, voir prochain chapitre) à l'époque où il écrit son article sur les diagrammes chez Euclide (voir Manders 1996).

24. Voir en particulier Manders 2008b. (En substance, cet article date de 1995 ; il a longtemps circulé informellement avant d'être publié.) L'article Manders 1996 contient une discussion plus brève des mêmes idées.

25. Pour une comparaison des approches assez différentes de Miller et de Mumma, voir Mumma 2008, Miller 2012 ainsi que Manders 2008a, en part. p. 70–71. Il faut noter que comme le montre Miller 2012, le système de Mumma est incohérent sous sa forme originale. Mumma en prépare une version corrigée, non encore publiée ; voir déjà Mumma 2014.

26. Pour une analyse plus poussée, voir *infra*, section 12.1.a)

d'abord, la distance nouvellement introduite entre la représentation et son contenu a pour conséquence qu'on peut facilement interpréter un figure de triangle, non plus comme le reflet imparfait d'un triangle idéal particulier, mais comme la représentation simultanée de toute une famille de configurations triangulaires. John Mumma, l'un des auteurs qui ont récemment tenté de formaliser l'usage des figures dans les premiers livres des *Éléments* d'Euclide, écrit ainsi :

> Once it is understood how the diagrams function as proof symbols, any particular diagram can comprise a general mathematical claim. [27]

D'autre part, l'attention se porte alors sur les transitions entre figures et énoncés, désormais conçues comme des inférences. Il devient central de savoir ce qu'on a le droit d'observer sur une figure, c'est-à-dire quels énoncés on peut en inférer.

La perspective de Barwise et Etchemendy conduit alors à formaliser l'usage de figures géométriques au moyen de règles syntaxiques hétérogènes. Les constructions d'une figure *ab initio* correspondraient par exemple à des règles ayant pour prémisses des énoncés et pour conclusion une figure. Les constructions sur une figure déjà existante correspondraient à des règles ayant pour prémisses à la fois une figure et des énoncés, et pour conclusion une figure. Les inférences faites sur la base d'une figure, quant à elles, correspondraient à des règles ayant pour prémisse une figure et pour conclusion un énoncé. Le programme est clair, quoique difficile à mettre en œuvre.

L'idée générale dépasse cependant le cadre étroit de la perspective syntaxico-sémantique de Barwise et Etchemendy. Ken Manders, dans son travail sur les figures dans la géométrie d'Euclide, accomplit un geste similaire (c'est en fait de Manders que s'inspire John Mumma, cité plus haut) :

> I treat diagrams as *textual* components of a traditional geometrical text or argument, rather than semantic counterparts. [28]

Pour lui aussi, la figure n'est donc pas ce dont parle le texte (une « contrepartie sémantique ») mais est sur le même plan que le texte ; l'important est alors d'étudier systématiquement les inférences qu'elle permet. La convergence, il est vrai, s'arrête là. Dans une perspective à la Barwise et Etchemendy, figures comme énoncés portent de l'information sur un domaine commun, et c'est cette sémantique commune qui fonde toutes nos inférences, c'est-à-dire à la fois le passage de certains énoncés à d'autres et le passage d'un

27. « Dès lors que l'on comprend comment les figures fonctionnent en tant que symboles de preuves, n'importe quelle figure particulière peut englober une affirmation mathématique générale » (Mumma 2010, p. 267).

28. « Je traite les diagrammes comme des composants *textuels* des textes ou arguments géométriques traditionnels, plutôt que comme des contreparties sémantiques. » (Manders 1996, p. 391). Sa position est la même dans l'article plus connu Manders 2008b.

figure à un énoncé ou inversement. Manders, en revanche, cherche à se dispenser de toute référence à une quelconque sémantique et tente de décrire la pratique euclidienne comme un jeu à deux supports – textes et diagrammes – sans donner à ces supports de signification extérieure. J'y reviendrai plus en détail dans la dernière partie [29].

8.3 Conclusion : comparer des représentations différentes

Les exemples des diagrammes de Venn et d'*Hyperproof* montrent ainsi comment on peut appliquer en pratique le slogan de Barwise et Etchemendy selon lequel diagrammes comme énoncés sont des « représentations » qui portent de l'« information ». On peut alors comparer des représentations de différentes sortes (par exemple des énoncés et des diagrammes) au moyen de leur sémantique commune ; on peut par exemple donner un sens précis à l'idée qu'un certain diagramme et une certaine liste d'énoncés ont « le même contenu » ou portent « les mêmes informations ». On peut imaginer analyser ainsi les avantages *systématiques* de certaines formes de représentation par rapport à d'autres : on peut imaginer qu'elles redécoupent de manière plus efficace le même espace de modèles possibles.

Toutefois, concernant le travail de Barwise et Etchemendy, ces exemples peuvent donner une impression trompeuse : identifier l'« information » que porte une représentation à un ensemble de modèles leur pose de nombreux problèmes, et ils cherchent à faire autrement. C'est ce que nous verrons au prochain chapitre.

29. Voir section 12.1.

Chapitre 9

Le contenu d'information, une nouvelle sémantique ?

Pour Barwise et Etchemendy, diagrammes comme énoncés sont fondamentalement des porteurs d'*information*. Mais que faut-il entendre ici par information ? En pratique, comme nous l'avons vu, nos auteurs associent à leurs diagrammes des ensembles de modèles. Ils ont pourtant des objections de fond vis-à-vis de cette approche et conçoivent l'information tout autrement, reprenant en cela des idées de la « sémantique des situations », une théorie de la sémantique des langues naturelles mise au point par Barwise en collaboration avec John Perry dans la décennie qui précède. Sur la base de ces idées, ils tentent de développer une nouvelle sémantique pour l'étude des diagrammes et plus généralement de l'inférence hétérogène (Barwise et Etchemendy 1990a).

Shin, par exemple, adopte cette nouvelle approche dans son premier article sur les diagrammes de Venn [1]. Elle y définit le « contenu [2] » du diagramme 9.1 non pas comme l'ensemble de ses interprétations possibles, mais comme un ensemble de « faits représentés », à savoir que les régions A et B représentent des ensembles, que l'ensemble représenté par A est vide et que l'ensemble représenté par B est non vide. Empruntant notations et terminologie à la sémantique des situations et à Barwise et Etchemendy (1990a), Shin modélise

1. Shin [1991] 1996, en part. p. 94–97.

2. « [T]he content of a *wfd* [well-formed diagram] D, $Cont(D)$, is defined as the set of the represented facts. [...] [O]ur definition of the content of a *wfd* tells us the information that the diagram conveys. » (« Le contenu d'un *dbf* [diagramme bien formé] D, $Cont(D)$, est défini comme l'ensemble des faits représentés [...] notre définition du contenu d'un DBF nous dit quelle est l'information que le diagramme transmet. » Shin [1991] 1996, p. 97.)

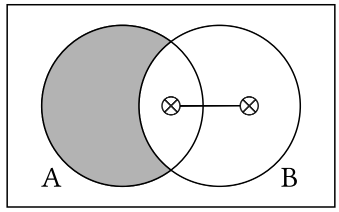


FIGURE 9.1 – Exemple de diagramme de Shin-Venn

ces faits par des « infons », ou unités d'information, qu'elle note comme ceci :

$$\langle\langle \text{Set}, f(A); 1 \rangle\rangle, \langle\langle \text{Set}, f(B); 1 \rangle\rangle, \langle\langle \text{Empty}, f(A); 1 \rangle\rangle, \langle\langle \text{Empty}, f(B); 0 \rangle\rangle$$

(où $f(A), f(B)$ désignent les objets représentés par les régions A et B respectivement). À première vue et malgré ces notations déroutantes, tout semble donc se passer comme si nos auteurs identifiaient l'information portée par un diagramme à un ensemble de propositions, ce qui mérite d'être clarifié.

Le but de ce chapitre est d'exposer cette sémantique exotique et surtout d'en comprendre les motivations, ce qui exigera, au passage, un examen sommaire de la sémantique des situations de Barwise et Perry. Il est vrai que l'approche précise de Barwise et Etchemendy (1990a) a ensuite été abandonnée, même par nos auteurs. Cependant, certaines idées en ont été reprises dans des théories ultérieures de l'information développées par Barwise, et, par ce biais, ont servi de base à d'intéressants travaux sur le raisonnement diagrammatique, ceux d'Atsushi Shimojima, auxquels est consacré le prochain chapitre. Plus profondément, l'intérêt d'étudier le cheminement de Barwise et Etchemendy est qu'il met bien en évidence les difficultés inhérentes à toute modélisation du raisonnement diagrammatique qui ne laisse pas de place au raisonnement explicite sur le sens ou sur les conventions de représentation des diagrammes.

9.1 Pourquoi une nouvelle sémantique ?

Le travail de Barwise et Etchemendy (1990a) sur les diagrammes s'inscrit dans le cadre de la *sémantique des situations* de Barwise et Perry[3]. Cette théorie a pour but de clarifier

3. Jon Barwise et John Perry commencent à travailler à cette nouvelle approche à la fin des années 70 (pour quelques éléments sur la genèse de leur projet, voir Barwise et Perry 1983, p. xiv-xxii), et publient leurs idées sous forme d'articles (Barwise et Perry 1980, 1981b), puis du livre Barwise et Perry 1983 (cf. aussi Devlin 2004 pour une bibliographie complète des travaux de Barwise en sémantique). La sémantique des situations se développe ensuite dans plusieurs directions. Certains travaux portent sur la logique, les fondements de la sémantique et la nature de l'information (Barwise 1989 ; Barwise et Etchemendy 1987b ; Devlin 1991 ; Israel et

les notions de *signification* et d'*information*, en particulier telles qu'elles s'appliquent à la langue naturelle. Sur de nombreux plans, elle rompt radicalement avec les sémantiques des mondes possibles en vogue dans les années 1970, et aura un impact important sur la recherche ultérieure en sémantique des langues naturelles, en particulier par sa prise en compte massive du contexte d'énonciation.

Plutôt que de résumer en toute généralité les multiples motivations de la sémantique des situations, je voudrais ici me concentrer sur un parti pris de Barwise et Perry qui, à mon sens, joue pour l'étude des diagrammes un rôle crucial : pour eux, la signification est une chose accessible aux utilisateurs de la langue, sur laquelle ceux-ci peuvent raisonner. Ce parti pris rend inadéquates, comme théories de la signification, non seulement la sémantique telle qu'utilisée en logique, mais aussi ses dérivés pour les langues naturelles que sont les sémantiques des mondes possibles (section a). Cette critique générale qu'adressent Barwise et Perry aux sémantiques en termes de modèles trouve une résonance toute particulière dans le cas des diagrammes, parce que l'usage des diagrammes suggère assez naturellement l'idée d'un raisonnement au niveau de l'information plutôt que de la syntaxe (section b). Barwise et Etchemendy (1990a) motivent leur travail en écrivant que la sémantique usuelle a été conçue uniquement pour les langages formels et n'est donc pas assez générale pour les systèmes hétérogènes ; cette idée prend un sens plus clair dans le cadre de ce projet général d'un raisonnement au niveau de l'information (section c).

a) La sémantique des situations : information contre mondes possibles

Pour Barwise et Perry, la signification d'un énoncé et l'information qu'il porte sont des choses auxquelles les utilisateurs de la langue ont accès et sur lesquelles ceux-ci peuvent

---

Perry 1990), tandis que d'autres utilisent ces nouveaux outils pour l'étude de divers aspects de la sémantique des langues naturelles (R. Cooper et Poesio 1994 ; Gawron et Peters 1990 ; Ginzburg et Sag 2000). Pour se faire une idée de la diversité des travaux liés à la sémantique des situations, on pourra consulter les recueils Aczel, Israel, Katagiri et Peters 1993 ; Barwise, Gawron, Plotkin et Tutiya 1991 ; R. Cooper, Kuniaki et Perry 1990. Stojanovic 2012 dresse une présentation générale utile de l'approche de Barwise et Perry en insistant sur ses motivations philosophiques. Ginzburg 2011b est également une introduction très claire, qui s'adresse toutefois plutôt à des linguistes et suppose une certaine familiarité avec la sémantique contemporaine. Enfin, Devlin 2006 donne, d'un point de vue interne à cette tradition, une présentation synthétique des fondements techniques de la théorie sous sa forme aboutie des années 1990.

S'il n'y a plus guère de chercheurs qui travaillent encore dans le cadre original de Barwise et Perry, nombreuses sont leurs idées qui ont été intégrées à la sémantique et à la pragmatique contemporaines. Pour s'en faire une idée, on pourra consulter Ginzburg 2011b, axé sur des applications de leurs idées à l'étude du dialogue, et Kratzer 2014, qui passe en revue les avantages, en sémantique, d'employer des situations (ou portions de mondes) plutôt que des mondes possibles complets.

raisonner [4]. Ce *desideratum* sur les théories de la signification ne va pas de soi : de nombreux auteurs comprennent autrement le but principal de la sémantique.

Pour bien saisir ces divergences, il est utile de revenir à Frege. Celui-ci [5] distingue la *référence* (« *Bedeutung* [6] ») d'une expression, qui est un objet ou une propriété du monde, de son *sens* (« *Sinn* »), qui est la manière dont cette référence est donnée au locuteur. Le sens, écrit Frege, est saisi [7] par les locuteurs de la langue, mais ceux-ci n'ont qu'un accès partiel à la référence : on peut comprendre les expressions « le point d'intersection des droites a et b » et « le point d'intersection des droites b et c », c'est-à-dire en saisir le sens, sans savoir si elles ont la même référence. Entre ces deux pôles du sens et de la référence, où situer la notion intuitive de « signification », qui doit faire l'objet de la sémantique ? La question est certes terminologique, mais cache aussi des enjeux techniques et conceptuels, comme nous allons le voir.

Une première possibilité est de comprendre la notion intuitive de signification comme le fait Michael Dummett, qui dans son commentaire de Frege écrit que le but d'une théorie de la signification doit être de rendre compte « de ce qu'une personne sait quand elle sait ce que signifie un mot ou une expression, c'est-à-dire quand elle le comprend [8] ». Selon cette interprétation du terme, le sens (« *Sinn* ») de Frege fait partie de la signification, *mais pas la référence*, puisque comme nous l'avons dit on peut comprendre parfaitement une expression sans en connaître la référence. Cet usage du terme de signification est assez fréquent aujourd'hui en philosophie analytique, où il est usuel de considérer que la tâche d'une théorie de la signification est d'assigner aux expressions linguistiques un « contenu sémantique » qui joue un rôle analogue au « *Sinn* » de Frege en ce qu'il est distinct de la référence mais la détermine [9].

Toutefois, de nombreux auteurs emploient les termes de signification et de sémantique

4. Dans son article d'introduction, Ginzburg 2011b, p. 854-855 voit dans ce parti pris, qu'il appelle « réification du sens », l'une des innovations principales de leur théorie : « Meaning reification : the reification of meanings as entities on which humans reason (rather than as metatheoretical entities, as standard in logic). »

5. Frege 1892. Pour un commentaire approfondi, voir Dummett 1973, chap. 5–6.

6. Contrairement à Claude Imbert (voir Claude Imbert 1971, p. 15), je ne traduirai pas « *Bedeutung* » par « dénotation » mais par « référence », de manière à pouvoir, sans créer de confusion, utiliser le même terme aussi pour sa traduction anglaise usuelle « *reference* ».

7. Le verbe allemand qui apparaît chez Frege est « *erfassen* » : « Der Sinn eines Eigennamens wird von jedem *erfaßt*, der die Sprache oder das Ganze von Bezeichnungen hinreichend kennt, der er angehört; damit ist die Bedeutung aber, falls sie vorhanden ist, doch immer nur einseitig beleuchtet. » (Frege 1892, p. 27 ; je souligne.)

8. « What we have to give an account of is what a person knows when he knows what a word or expression means, that is, when he understands it » (Dummett 1973, p. 92).

9. Voir par ex. Speaks 2014, ou encore l'introduction à la philosophie du langage, largement diffusée, de Soames 2010, sections 1.11–1.12, qui assimile sans autre forme de procès signification (« *meaning* ») et « *Sinn* » fregéen.

en un sens plus large, qui inclut aussi la référence. (Pour lever l'ambiguïté, je distinguerai ici entre signification ou sémantique *au sens strict*, qui exclut la référence, et *au sens large*.) En fait, lorsque le terme de sémantique commence à se diffuser dans les années 1930, il recouvre tout ce qui concerne le lien entre expressions et objets, donc aussi et même surtout les notions de référence et de vérité : la question centrale est alors celle de savoir à quels objets du monde renvoient nos mots. C'est en particulier l'usage qu'en fait Alfred Tarski [10], dont la terminologie est devenue usuelle en logique.

Il importe donc de souligner que ce qu'on appelle sémantique en logique (à la suite de Tarski) n'est pas une théorie de la signification au sens strict. Un système logique vise à modéliser une certaine forme de raisonnement en codifiant des règles d'inférence. La sémantique d'un système logique consiste alors en deux choses. La première est ce qu'on pourrait appeler une théorie de la référence pour les énoncés du langage étudié : on explique comment les références des expressions complexes (y compris les valeurs de vérité des énoncés) dépendent des références de leurs composants. La seconde est ce qu'on pourrait appeler une théorie des références possibles : on fait un modèle mathématique (d'habitude ensembliste) de toutes les références imaginables des expressions de notre langage [11]. Cette théorie des références possibles permet alors d'évaluer la validité des règles d'inférences, en examinant si elles préservent bien la vérité *quelles que soient les références choisies* pour les composants des énoncés considérés. Vis-à-vis de notre problème, deux

10. Tarski écrit ainsi : « Nous entendons par sémantique l'ensemble des études qui traitent des concepts qui, en gros, expriment certaines relations entre les expressions d'un langage et les objets et états de choses auxquels ces expressions se réfèrent. Comme exemples typiques de concepts sémantiques, nous pouvons mentionner les concepts de *désignation*, de *satisfaction* et de *définition* […]. Le concept de *vérité* également […] doit être considéré comme un concept sémantique » (Tarski 1936a, p. 1, trad. fr. Tarski 1972-1974, vol. 2, p. 133 ; voir aussi Tarski [1933] 1935, 1936b). Tarski emprunte sans doute le terme à Charles Morris, qui a introduit la terminologie aujourd'hui courante d'une tripartition syntaxe, sémantique, pragmatique. Celui-ci écrit : « L'analyse révèle que les signes linguistiques entretienennt trois types de relations (à d'autres signes de la langue, à des objets signifiés, à des personnes qui les utilisent et comprennent) qui définissent trois dimensions de la signification [*meaning*]. Ces dimensions, à leur tour, sont les objets d'investigation pour la syntaxe, la sémantique et la pragmatique […] » (Morris 1937, p. 4). À son tour, Carnap reprend cette terminologie, qu'il attribue à Tarski et à Morris (Carnap 1936-1937, p. 2, n. 2 : « Notre problème de la signification appartient au domaine que Tarski appelle *Sémantique* ; c'est la théorie des relations entre les expressions d'une langue et les choses, propriétés, faits etc. décrits dans cette langue » ; pour la référence à Morris, voir par exemple Carnap 1942, §4.)

11. Etchemendy 1990 insiste sur la différence entre deux manières de comprendre cette seconde étape. On peut tout d'abord considérer que les modèles correspondent à des états possibles du monde. En faisant varier la référence du mot *vert* d'un modèle à l'autre, on fait alors varier quels objets sont verts : dans un monde possible voisin du nôtre, il se pourrait par exemple que les pêches soient vertes. On peut au contraire considérer que les modèles correspondent à des interprétations des termes non logiques du langage, en laissant le monde fixé. Le mot *vert* pourra alors correspondre, dans certains modèles, à l'ensemble des objets triangulaires ou à n'importe quoi d'autre. Le but d'Etchemendy, qui s'intéresse à l'analyse tarskienne de la conséquence logique, est de montrer que celle-ci est fondée sur une confusion entre ces deux interprétations (sur ce point, voir Sher 1996). Pour la question qui nous occupe, cependant, cette distinction peut être ignorée.

conclusions s'imposent alors. Non seulement la sémantique ainsi comprise n'a pas pour ambition de capturer la signification au sens strict, mais elle joue même un rôle explicitement méta-théorique : elle ne sert pas à modéliser le raisonnement – tâche qui revient aux règles d'inférence –, elle sert plutôt à en évaluer la validité de l'extérieur. On est donc très loin d'une conception de la signification comme quelque chose sur quoi les agents peuvent raisonner.

Derrière ce problème terminologique sur ce qu'on entend par « sémantique » se cache néanmoins un problème plus profond : de nombreux auteurs, à partir de Carnap (1947), ont tenté d'utiliser des outils issus de la sémantique tarskienne pour modéliser quelque chose d'apparenté au « *Sinn* » fregéen, c'est-à-dire à la signification au sens strict. Les approches de ce genre, enrichies par les travaux de Hintikka et Kripke sur la sémantique des modalités et généralisées de manière décisive par le logicien Richard Montague [12], ont permis de grandes avancées : ils ont fourni une théorie systématique et souvent convaincante liant la signification de phrases complexes de la langue naturelle à celle de leurs constituants, et expliquant quelles inférences entre phrases sont valides. La tradition sémantique qui en est issue peut être appelée, de manière plus ou moins vague et selon l'extension qu'on lui donne, sémantique de Montague, sémantique des mondes possibles ou sémantique modèle-théorique [13]. Pourtant, comme nous allons le voir, si le but est une théorie de la signification au sens strict, les outils mathématiques que ces approches utilisent sont inadaptés ou du moins insuffisants.

L'idée de départ de toute cette tradition, idée qui fonde le lien entre signification au sens strict et sémantique tarskienne, est que le sens d'un énoncé n'est rien d'autre que ses conditions de vérité [14]. Selon ce point de vue, connaître le sens d'un énoncé, c'est connaître sa valeur de vérité dans toutes les circonstances possibles. C'est ainsi que Carnap (1947) remplace le *Sinn* de Frege par ce qu'il appelle l'*intension* : schématiquement, l'intension d'une expression est une fonction associant à chaque monde possible la référence, dans ce monde, de cette expression [15]. En particulier, comme Carnap conserve la thèse fregéenne

12. Richard Montague, un élève d'Alfred Tarski, est connu pour une poignée d'articles sur la sémantique des langues naturelles rédigés autour de 1970, tous reproduits à titre posthume dans Montague 1974b (voir en particulier Montague [1970] 1974a et Montague [1973] 1974c).

13. Partee 1989 décrit de manière très claire et synthétique le développement de ces approches. Pour une discussion des différentes variantes de cette approche et des postulats philosophiques qui les sous-tendent, voir Zimmermann 2011.

14. Cette thèse, apparentée au vérificationnisme défendu par Carnap plus tôt dans sa carrière, est souvent attribuée à Wittgenstein, qui écrit dans le *Tractatus* que « la proposition est l'expression de ses conditions de vérité » (Wittgenstein [1921] 1993, §4.431).

15. Pour alléger mon propos, je me permets ici quelques approximations. D'une part, Carnap ne remplace pas seulement la notion fregéenne de sens par celle d'intension, mais aussi la notion de référence par celle

d'après laquelle la référence d'un énoncé est sa valeur de vérité, l'intension d'un énoncé correspond à la donnée de sa valeur de vérité dans chaque monde possible.

Or plusieurs problèmes rendent difficile de faire le lien entre l'intension carnapienne d'une expression et la compréhension que peuvent en avoir les agents. Ces problèmes sont par exemple rassemblés dans un article de Barbara Partee (1979), dont le titre révélateur (« La sémantique – mathématiques ou psychologie ? ») est symptômatique du malaise, dans les années 1970, de certains praticiens de cette nouvelle sémantique, qui ne comprenaient plus bien le statut de leurs propres théories.

Le problème principal n'est *pas* que les ensembles de mondes possibles seraient des entités trop riches pour qu'on puisse en attribuer la connaissance aux locuteurs. Certes, il serait absurde de supposer que les utilisateurs du langage aient la totalité des mondes possibles dans la tête. Cependant, cette hypothèse n'est pas indispensable : on peut considérer que les mondes possibles fournissent un bon modèle mathématique des conditions de vérité, mais que les agents n'en maîtrisent qu'une approximation dont la psychologie devra élucider les détails. L'objectif serait alors de faire une théorie de la compétence sémantique idéale, celle qu'aurait un agent omniscient, ou, comme l'écrit Partee, une théorie de « l'anglais tel que le comprendrait Dieu ».

Mais il y a un problème plus grave, celui des « attitudes propositionnelles ». Considérons par exemple deux phrases logiquement équivalentes ϕ et ψ. Elles sont vraies exactement dans les mêmes mondes possibles, et ont donc la même intension au sens de Carnap. Est-il plausible de leur attribuer la même signification ? On pourrait peut-être le défendre en invoquant encore une fois l'idée d'une compétence sémantique idéale, celle d'un agent omniscient. Mais considérons à présent « John croit que ϕ » et « John croit que ψ ». En sémantique des mondes possibles, on n'a pas le choix : le principe de compositionnalité – d'après lequel la signification d'un énoncé est déterminée par celle de ses parties – force à attribuer la même signification aussi à ces deux phrases. Montague l'admet sans ambages :

> [I]f ϕ and ψ are logically equivalent sentences [...], then $\ulcorner$John believes that $\phi\urcorner$ and $\ulcorner$John believes that $\psi\urcorner$ will turn out also to be logically equivalent [...]. This may at first appear strange, but is a conclusion that I believe we should accept [...]. [16]

d'extension, ce qui transforme le traitement des prédicats en éliminant la distinction qu'établit Frege entre la *propriété* qui est référence d'un prédicat et l'*extension* de cette propriété (pour une introduction, voir Recanati 2008, chap. 1). D'autre part, Carnap représente ses mondes possibles non par des modèles comme nous en avons depuis l'habitude, mais par des « descriptions d'états » (« *state descriptions* ») qui correspondent à l'attribution de valeurs de vérité à toutes les propositions atomiques.

16. « si ϕ et ψ sont des phrases logiquement équivalentes [...], alors $\ulcorner$John croit que $\phi\urcorner$ et $\ulcorner$John croit que $\psi\urcorner$ vont aussi s'avérer logiquement équivalentes [...]. Au premier abord, cela peut sembler étrange, mais je pense

Pourtant, si John n'est pas omniscient, il peut croire ϕ sans croire ψ. Les phrases « John croit que ϕ » et « John croit que ψ » ne sont donc apparemment pas vraies dans les mêmes mondes possibles. Contrairement à Montague, la plupart des auteurs l'admettent, même au sein de cette tradition. Dowty, Wall et Peters (1981), dans leur manuel d'introduction à la sémantique de Montague, écrivent qu'il s'agit d'« un problème fondamental pour la sémantique des mondes possibles » qui pourrait bien conduire à l'abandonner ou à la transformer radicalement [17].

À vrai dire, Carnap (1947) est conscient de ce problème (des exemples du même genre se trouvent en fait déjà chez Frege [18]) et tente de le résoudre. Pour ce faire, il propose d'ajouter à la notion d'intension celle de *structure intensionnelle* : deux phrases sont substituables dans des contextes d'attitudes propositionnelles (comme « John croit que… ») si, en plus d'avoir la même intension, elles sont *construites de la même manière* à partir d'expressions qui ont la même intension [19]. David Lewis reprend cette idée dans sa célèbre défense de la sémantique des mondes possibles (Lewis 1970). Les intensions, écrit-il, font partie de la signification (« *meaning* ») mais ne l'épuisent pas ; il faut leur ajouter d'autres ingrédients, analogues à la structure intensionnelle carnapienne [20]. Au cours des années 1970 et 1980, des auteurs comme Cresswell (1985) tentent d'intégrer pleinement des solutions de ce genre à la sémantique des mondes possibles. Il semble toutefois que les problèmes techniques à résoudre soient nombreux et que les théories hybrides qui en résultent perdent la clarté, la simplicité et la généralité qui font l'attrait de la sémantique de Montague [21].

Barwise et Perry, quant à eux, tentent de refonder la sémantique sur de nouvelles bases, cognitivement plus réalistes, plutôt que de réparer ou compléter la sémantique des mondes possibles. Pour autant, ils reconnaissent l'avantage qu'ont les sémantiques en termes de mondes possibles d'établir un lien direct entre expressions du langage et monde extérieur. Ils positionnent donc leur propre théorie non seulement contre les variantes usuelles de la

---

qu'il faut accepter cette conclusion. » (Montague [1970] 1974a, p. 218).

17. Dowty, Wall et Peters 1981, p. 175.

18. Frege, bien sûr, ne développe pas une sémantique des mondes possibles. Pour lui, le problème se pose donc différemment : la difficulté est que, dans une phrase comme « John croit que ϕ », la référence (« *Bedeutung* ») de ϕ, c'est-à-dire sa valeur de vérité, ne suffit pas à déterminer la référence de la phrase complète, puisque John peut croire certaines vérités et pas d'autres. L'idée de Frege est que dans ces contextes, la référence d'une phrase ϕ doit être son *sens* (« *Sinn* ») et non sa référence habituelle qui est sa valeur de vérité. Reste alors à clarifier la notion vague de sens. Le problème que j'ai exposé ci-dessus est que si, comme Carnap, on remplace la notion de sens par celle d'intension, la solution de Frege ne peut plus fonctionner : l'intension de ϕ ne suffit à déterminer ni la valeur de vérité, ni l'intension de « John croit que ϕ ».

19. Carnap 1947, §13–15.

20. Voir Lewis 1970, p. 25, 31-35.

21. Pour un survol des solutions proposées, voir Bäuerle et Cresswell [1989] 2003. Pour un traitement plus récent, voir Fox et Lappin 2005.

sémantique des mondes possibles, mais aussi contre les sémantiques qui se contentent d'associer aux mots et aux phrases des représentations mentales [22]. Pour eux, ces approches-là ne font que déplacer le problème : elles traduisent la langue naturelle en un langage de la pensée, mais laissent ouvert le problème du lien entre ce langage de la pensée et le monde extérieur [23]. Ils cherchent donc une voie moyenne. Le concept de signification qu'ils visent doit permettre de parler *à la fois* de l'état du monde et de l'état mental des locuteurs.

b) Diagrammes et raisonnements sémantiques

Jusqu'ici, j'ai parlé de la sémantique en général. Mais le cas particulier des diagrammes fournit aussi des raisons d'abandonner les sémantiques en termes de modèles. Dans cette section, je voudrais examiner un argument qui est à mon avis déterminant pour Barwise et Etchemendy, et éclaire leur choix de faire appel à la sémantique des situations pour étudier le raisonnement diagrammatique. (Je reviendrai au prochain chapitre sur la question de savoir si cet argument est pleinement convaincant [24].)

Le problème est qu'il y a un fossé particulièrement profond entre les systèmes formels diagrammatiques que j'ai présenté plus haut (par exemple pour les diagrammes d'Euler ou de Venn) et nos pratiques informelles avec les diagrammes. À en croire ces systèmes formels, raisonner avec des diagrammes consiste à suivre des règles syntaxiques de manipulation données d'avance. Pourtant, ce n'est pas du tout ce que nous faisons informellement, comme le montrent bien certains exemples traités par Barwise et Etchemendy eux-mêmes. Examinons un cas simple [25].

Imaginez que vous deviez asseoir quatre personnes, notées A, B, C et D, sur une rangée de cinq chaises, en respectant les contraintes suivantes : A et C doivent être assis à côté de la chaise vide (de part et d'autre) ; C doit être plus près du centre que D ; D doit être à côté de B. Pouvez-vous déduire de ces contraintes que C doit être assis au centre ? Voici une solution simple, qui s'appuie sur des diagrammes. Comme A et C doivent être assis de

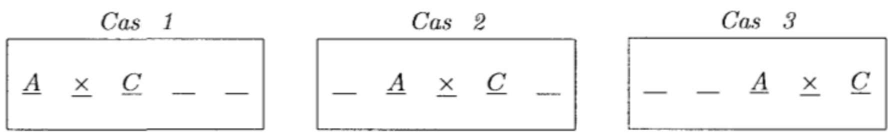
22. Aujourd'hui, on parle volontiers de « sémantiques cognitives » pour décrire ce genre de travaux. À l'époque, Barwise et Perry font surtout référence à un programme de recherche concurrent de celui de Montague, associé entre autres au nom de Jerrold Katz ; pour plus de détails, voir Kempson 2011, qui le décrit en ces termes (p. 218 ; l'auteure souligne) : « Much of the focus was on exploring appropriate *semantic representations* in some internalised *language of thought* to assign to words to secure a basis for predicting such entailment relations as *John killed Bill*, *Bill died*, or *John is a bachelor*, *John is an unmarried man* [...]. There was no detailed mapping defined from such constructs onto the objects/events which the natural language expression might be presumed to depict. »

23. Cette critique n'a rien d'original à l'époque et est par exemple dévelopée vigoureusement par Lewis 1970.

24. Cf. *infra*, section 10.4.

25. Il s'agit d'un une version abrégée d'un exemple de Barwise et Etchemendy 1990a.

part et d'autre de la chaise vide, que nous noterons par une croix, seuls sont possibles les trois cas suivants (et leurs symétriques, où A et C sont inversés) :



Dans le second cas, B ne peut être voisin de D faute de deux places libres côte à côte ; dans le troisième cas, C ne peut être plus près du centre que D puisqu'il est aussi loin du centre que possible. Le seul cas possible est donc le premier, où C est assis au centre, comme voulu.

Vous n'aviez sans doute jamais vu ce type particulier de diagrammes, et ne sauriez sans doute pas dire sur quelles règles syntaxiques le raisonnement repose. Pourtant, vous avez certainement pu le suivre sans problème et vous convaincre qu'il était correct. Comment est-ce possible ? Barwise et Etchemendy semblent parfois suggérer que c'est parce que le raisonnement a lieu au niveau de l'*information* contenue dans les diagrammes, et non au niveau de leur forme. Une fois que l'on a compris selon quelles conventions ces diagrammes représentent de l'information (les segments horizontaux représentent des chaises ; un segment surmonté d'une croix indique une chaise vide ; etc.), on sait évaluer les inférences : par exemple, on reconnaît qu'il est légitime de déduire du premier diagramme que C est assis au centre, parce que cela fait partie des informations que ce diagramme porte. En d'autres termes, tout se passe comme si, pour décider si le passage d'une représentation à une autre est légitime, on se laissait guider par le *contenu* de ces représentations plutôt que par leur forme : par leur *sémantique* plutôt que par leur *syntaxe*.

Soulignons ce que ce problème a de spécifique aux diagrammes. Dans le cas des systèmes logiques habituels, les règles d'inférence spécifiées syntaxiquement peuvent être considérées comme correspondant à des lois de la pensée, ou comme constituant le sens des constantes logiques : on peut défendre l'idée qu'elles ont une justification intrinsèque. Dans le cas des diagrammes de Shin ou de nos diagrammes de chaises, ce serait inconcevable. Non seulement la seule justification possible des règles du système de Shin étudié plus haut passe par le sens attribué aux diagrammes, mais encore ces règles seront différentes pour chaque nouvelle sorte de diagramme que l'on considérera. C'est ainsi que Barwise et Etchemendy justifient leur décision de modéliser le raisonnement au niveau de l'information plutôt que de la syntaxe :

> When we look at a particular formal language, such as the first-order predicate calculus, it is often possible to find a dozen or so basic syntactic rules that suffice for all valid reasoning that can be carried out within that language.

> […] [W]hen we go beyond traditional formal languages and start admitting all possible forms of visual representation, the idea of finding a few fundamental rules, stated in terms of "syntactic" operations on representations, becomes implausible.
>
> Our strategy is to work at the level of information, rather than its representation, and try to find a few basic principles of information flow that suffice to account for all valid inference. [26]

Cette idée semble rompre avec la division des tâches entre syntaxe et sémantique qui est habituelle en logique. En effet, la syntaxe sert d'ordinaire à modéliser nos inférences, c'est-à-dire les étapes élémentaires de nos raisonnements. La sémantique, au contraire, sert à définir une relation de conséquence entre propositions qui est indépendante des moyens inférentiels dont on cherche à rendre compte. Certes, dans de nombreux systèmes logiques, un théorème de complétude garantit qu'une proposition est démontrable syntaxiquement à partir de certaines hypothèses si et seulement si elle en est conséquence sémantique, de sorte qu'en définitive dérivabilité syntaxique et conséquence sémantique définissent la même relation de conséquence logique. Mais malgré cette coïncidence occasionnelle, syntaxe et sémantique jouent des rôles fondamentalement distincts. Il ne serait par exemple pas raisonnable d'introduire, dans un système logique d'arithmétique, une règle d'inférence permettant de déduire directement le grand théorème de Fermat des axiomes de Peano : une telle règle, bien que valide sémantiquement, ne correspondrait pas à une étape élémentaire de raisonnement. Dans ce cadre familier, une inférence « d'après le sens », c'est-à-dire sémantique plutôt que syntaxique, n'a de place qu'à l'extérieur du système formel considéré : c'est une inférence qui va au-delà de la pratique inférentielle que l'on a capturée.

Une réponse naturelle serait de dire que ce problème apparent repose sur une confusion : les systèmes diagrammatiques formels des chapitres précédents ne serviraient qu'à justifier *a posteriori* les raisonnements diagrammatiques, et n'auraient rien à dire sur les processus mentaux des utilisateurs des diagrammes. Barwise et Etchemendy eux-mêmes

26. « Lorsque nous examinons un langage formel particulier, comme le calcul des prédicats du premier ordre, il est souvent possible de trouver mettons une douzaine de règles syntaxiques de base, qui suffisent pour tous les raisonnements valides que l'on peut conduire dans ce langage. […] [L]orsque nous allons au-delà des langages formels traditionnels et que nous commençons à admettre toutes les formes possibles de représentations visuelles, l'idée de trouver quelques règles fondamentales, formulées en termes de manipulations "syntaxiques" sur les représentations, devient peu plausible. Notre stratégie est de travailler au niveau de l'information plutôt que de sa représentation, et d'essayer de trouver quelques principes de base d'écoulement de l'information qui suffisent à rendre compte de toutes les inférences valides. » (Barwise et Etchemendy 1990a, p. 59–60).

décrivent parfois ainsi leur travail, comme nous le verrons. Il me semble cependant que leur projet est plus ambitieux. Ils cherchent manifestement à rapprocher la logique du raisonnement informel, et l'un de leurs buts centraux est de mieux rendre compte de l'usage de diagrammes *ad hoc* pour résoudre de petits problèmes de raisonnement comme le précédent, typique de ceux qui sont posés à l'examen « GRE » aux États-Unis [27]. Ils parlent souvent de la manière dont les diagrammes facilitent la découverte d'une solution [28]. Surtout, beaucoup de leurs formulations montrent qu'ils conçoivent l'information portée par un diagramme (ou, d'ailleurs, par un énoncé) comme quelque chose à quoi la personne qui raisonne doit avoir accès.

À mon sens, c'est bien là qu'il faut chercher la raison la plus profonde pour laquelle, dans leur unique article plus technique sur le raisonnement hétérogène (Barwise et Etchemendy 1990a), nos auteurs empruntent à la théorie des situations une notion de « contenu d'information » plus fine que la sémantique en termes de modèles : cette nouvelle sémantique ouvre une piste pour clarifier l'idée que l'on raisonne sur les diagrammes d'après leur sens.

c) Une notion d'information indépendante du type de représentation

Dans leur introduction, Barwise et Etchemendy (1990a) justifient autrement leur projet d'une nouvelle sémantique :

> The most widely accepted semantic analysis of consequence represents the information content of a sentence with the collection of structures in which the sentence is true, and declares one sentence to be a consequence of another if the information content of the first is contained in the content of the second, that is, if all the models of the second are models of the first. [...] These familiar techniques have been tailor-made for the homogeneous, linguistic case. When we turn our attention to heterogeneous inference, we are naturally drawn to look for a model of information that is independent of the particular form in which the information is represented. Thus, if we are studying inferences that involve both sentences and diagrams, we need a way to combine the information provided by both forms of representation. If our model of information is too closely tied to the syntax of any particular system of representation, as the

27. Le GRE, ou « *Graduate Record Examination* », est un test standardisé utilisé aux États-Unis pour évaluer les candidats à des formations universitaires de deuxième ou troisième cycle ; Barwise et Etchemendy l'utilisent régulièrement comme source d'exemples. Pour un autre problème dans le même esprit que celui des chaises, voir Barwise et Etchemendy 1998, p. 103–104.

28. Voir en particulier Barwise et Etchemendy [1991] 1996b, 1998.

> traditional model is, then the requisite operations on information will not be available. [29]

Effectivement, la sémantique habituelle telle que présentée en logique est étroitement liée à la syntaxe des langages logiques associés : la définition même des structures d'interprétation d'un langage du premier ordre fait intervenir les symboles du langage considéré. Barwise s'en plaint très explicitement dans un autre article :

> [T]he starting point of the standard theory is that structures are defined only relative to a specific language. It is as though God first created language and then the world, not the other way around. [30]

De fait, lorsque Shin munit les diagrammes de Venn d'une sémantique, elle se trouve obligée de définir les interprétations différemment, en partant des courbes ; lorsque Hammer enrichit ce système pour le rendre hétérogène, il doit ajouter des mécanismes de coréférence *ad hoc* pour que diagrammes et énoncés puissent parler de la même chose [31].

Cette objection pointe un problème technique plutôt qu'un argument de fond contre l'usage de modèles pour décrire la signification ou l'information. En réalité, les défenses philosophiques de l'usage de mondes possibles en sémantique (par exemple celle de Stalnaker 1984) affirment souvent que l'un des avantages de capturer une proposition par un ensemble de mondes possibles est précisément que cette description est indépendante de la forme linguistique particulière sous laquelle la proposition se présente.

Au mieux, cet argument de Barwise et Etchemendy peut motiver l'idée d'introduire l'« information » (telle que Barwise et Etchemendy la définissent) comme strate intermédiaire entre représentations et modèles, et ainsi d'éviter ces difficultés de manière systématique en séparant deux étapes dans l'interprétation : tout d'abord, on assignerait à

29. « L'analyse sémantique la plus largement acceptée de la conséquence représente le contenu d'information d'une phrase par la collection des structures dans lesquelles la phrase est vraie, et décrète qu'une phrase est conséquence d'une autre si le contenu d'information de la première est inclus dans le contenu de la seconde, c'est-à-dire si tous les modèles de la seconde sont modèles de la première. [...] Ces techniques familières ont été faites sur mesure pour le cas homogène et linguistique. Lorsque nous nous tournons vers l'inférence hétérogène, nous sommes naturellement enclins à chercher un modèle de l'information qui soit indépendant de la forme particulière dans laquelle l'information est représentée. Ainsi, lorsque nous étudions des inférences qui comportent à la fois des phrases et des diagrammes, il nous faut une manière de combiner l'information fournie par les deux formes de représentation. Si notre modèle de l'information est trop étroitement lié à la syntaxe d'un quelconque système représentationnel particulier, comme l'est le modèle traditionnel, alors les opérations sur l'information dont nous avons besoin ne seront pas disponibles. » (Barwise et Etchemendy 1990a, p. 33–34).

30. « [L]e point de départ de la théorie standard est que les structures ne sont définies que relativement à un langage particulier. C'est comme si Dieu avait créé le langage d'abord et le monde seulement ensuite, plutôt que l'inverse. » (Barwise 1983, p. 65).

31. Voir note 44 ci-dessus.

chaque représentation son contenu d'information, ce qui serait fait différemment pour un diagramme et pour une phrase ; ensuite, on ferait le lien entre informations et modèles, on définirait la relation de conséquence, etc., cette seconde étape étant alors indépendante de la forme particulière de la représentation utilisée. En ce sens, le contenu d'information jouerait le rôle d'un intermédiaire entre les représentations elles-mêmes et les modèles dont elles parlent.

Il n'est cependant pas clair qu'un tel intermédiaire soit nécessaire. Le système de Hammer, par exemple, montre que des ensembles de modèles peuvent remplir le rôle de sémantique commune entre des représentations de nature différente et permettent de définir une relation de conséquence logique entre elles. Certes, la manière dont on détermine les modèles associés à une représentation donnée dépendra du type de celle-ci, mais il faut bien que cette différence-là réapparaisse quelque part ; introduire un nouveau concept pour cette seule raison ne semble pas se justifier.

En définitive, il me semble que si Barwise et Etchemendy ont besoin d'un nouveau concept d'information, différent d'ensembles de modèles, ce n'est pas vraiment à cause de ce problème de neutralité ; c'est fondamentalement pour respecter les intuitions de la sémantique des situations et surtout, dans le contexte de leur article de 1990, parce que c'est *au niveau de l'information* qu'ils veulent modéliser l'inférence. Comme nous le verrons, les « opérations sur l'information » qu'ils visent vont au-delà de la simple conséquence logique définissable en termes de modèles et exigent donc un nouveau cadre.

9.2 La sémantique des situations

Le travail de Barwise et Etchemendy sur les diagrammes partage non seulement certaines motivations, mais aussi nombre de présupposés et de concepts avec la sémantique des situations. Avant de poursuivre, il est donc utile de survoler rapidement les fondements de celle-ci [32].

Un organisme vivant, écrivent Barwise et Perry, n'est jamais confronté qu'à ce qu'ils appellent des *situations*, c'est-à-dire des portions de monde limitées dans l'espace et dans le temps [33]. Pour s'orienter dans cet environnement changeant, chaque organisme catégorise les situations qu'il rencontre en identifiant parmi elles certaines régularités (« *uniformi-*

32. Ce survol est très rudimentaire ; je me concentre sur les aspects les plus utiles pour comprendre les travaux de Barwise et Etchemendy sur les diagrammes. Pour une introduction plus détaillée, voir références citées *supra*, n. 3, p. 226.

33. Le choix de partir de situations plutôt que de mondes possibles, qui donne son nom à la théorie, prend sa source dans le travail d'Austin (en particulier Austin 1950) et dans l'analyse de certaines constructions particulières de l'anglais ; voir Barwise 1981 et, pour une introduction, Ginzburg 2011a, sect. 2.

ties ») ; les individus, propriétés, relations et lieux qui figurent dans les descriptions humaines du monde sont ultimement de telles régularités. Nos auteurs s'inspirent ici d'une perspective « écologique », mise à l'époque au centre des discussions par le travail de Gibson sur la perception visuelle (Gibson 1979).

S'il y a de l'*information*, c'est parce que les organismes peuvent identifier des régularités dans les *relations entre situations*, par exemple savoir que s'il y a de la fumée dans une situation, il y a du feu à proximité. Ces régularités d'ordre supérieur, que Barwise et Perry appellent des contraintes (« *constraints* »), permettent aux organismes qui y sont sensibles d'utiliser une situation (par exemple catégorisée comme une situation où il y a de la fumée) pour apprendre quelque chose sur une autre situation (par exemple, qu'une situation voisine doit être catégorisée comme contenant du feu). L'information que peut porter un fait est ainsi toujours relative à la contrainte considérée et donc à l'organisme qui y est sensible [34].

Les concepts de base sont ainsi ceux de situation, de type de situation et de contrainte. Les types de situation sont parfois appelés *états de faits* (« *states of affairs* ») ou « *infons* ». En pratique, Barwise et Perry modélisent souvent les situations par des constructions ensemblistes formées d'objets et de relations, appelées *situations abstraites*, et les infons par des sortes de propositions structurées dont les composants sont ces mêmes objets et relations. Il importe toutefois de comprendre que ces modèles ensemblistes ne sont pas consubstantiels à la théorie ; ils reposent sur des hypothèses supplémentaires et substantielles sur l'ontologie du langage naturel, c'est-à-dire sur les uniformités effectivement repérées entre situations par les humains [35].

Dans ce cadre, la signification (« *meaning* ») linguistique n'est qu'une forme de contrainte, qui, pour caricaturer, ne relie pas fumée et feu mais plutôt situations comportant le cri « au feu ! » et situations où il y a du feu. La signification est donc fondamentalement une *relation* entre situations d'élocution et situations décrites. Une conséquence en est que cette théorie accorde d'emblée une place centrale au contexte. Les porteurs d'information ne sont pas les phrases abstraites, mais seulement les phrases prononcées dans un contexte particulier. Hors contexte, une phrase a certes un sens, déterminé par les conventions de la langue, mais ce sens est relationnel : ce n'est qu'un potentiel de porter certaines informations, qui dépendront de la situation d'élocution.

Intéressons-nous de plus près au cas d'une phrase déclarative, prononcée dans un certain contexte d'élocution [36]. En vertu de son sens, qui est relationnel, cette phrase peut

34. Pour une analyse plus approfondie, voir Israel et Perry 1990.
35. Voir Ginzburg 2011a.
36. Les facteurs qui peuvent intervenir dans la détermination de l'interprétation d'une phrase vont en fait

donner à l'auditeur toutes sortes d'informations : certaines portent sur le monde (la situation décrite), d'autres sur la situation d'élocution et sur le locuteur. Concentrons-nous, comme le fait la sémantique traditionnelle, sur les liens de la phrase avec la situation décrite. Barwise et Perry suivent ici Austin (1950) et en distinguent deux aspects. Le premier aspect, qu'Austin appelait *démonstratif*, est ce dont la phrase parle, c'est-à-dire une situation particulière ou éventuellement plusieurs si la phrase est ambiguë. Nos auteurs parlent d'*interprétation* de la phrase. L'interprétation joue un rôle analogue à celui de la référence dans d'autres approches – les composants de la phrase ont aussi une interprétation, par exemple celle d'un nom est un objet, etc. – à la différence capitale que l'interprétation d'une phrase est une situation et non une valeur de vérité [37]. En second lieu, il y a l'aspect qu'Austin appelait *descriptif* : de la situation décrite, la phrase affirme un certain nombre de choses. (*Un certain nombre* de choses, et non une chose précise : Barwise et Perry soulignent la multiplicité de ce qui est affirmé par une phrase, même simple. Ainsi, si je vous dis que « Shachar est assise par terre sur la terrasse », vous pouvez apprendre où ma nièce Shachar se trouve, mais aussi – grâce à l'accord au féminin – que c'est une fille.) Cet aspect descriptif est ce que nos auteurs appellent parfois *contenu d'information* : c'est un ensemble d'infons, c'est-à-dire un ensemble de *types de situations* sous lesquels je peux apprendre que la situation décrite tombe.

Toujours à la suite d'Austin (1950), nos auteurs distinguent alors soigneusement information et proposition. Le contenu d'information, argumentent-ils, ne suffit pas à déterminer la situation décrite, qui dépend elle aussi du contexte et des intentions du locuteur ; or sans la donnée d'une situation dont on peut juger si elle vérifie nos infons, on ne saurait attribuer à ceux-ci de valeur de vérité. Pour évaluer la valeur de vérité d'une phrase déclarative, il faut donc tenir compte des deux aspects, la situation décrite et le contenu descriptif ou contenu d'information. Pour eux, une proposition – ils parlent parfois, pour clarifier, de *proposition austinienne* – est donc toujours de la forme « s est de type ϕ », pour s une situation et ϕ un type de situation ou infon [38].

Beaucoup de ces idées ne s'appliquent pas immédiatement au cas du raisonnement diagrammatique ou hétérogène. En particulier, dans les exemples que Barwise et Etche-

---

au-delà du contexte au sens étroit du terme : ils incluent les intentions du locuteur, prises en charge par ce que Barwise et Perry appellent les « connexions » (« *connections* »).

37. Cette position, très hétérodoxe, force Barwise et Perry à réfuter plusieurs arguments classiques d'après lesquels la référence d'une phrase ne peut être que sa valeur de vérité. L'argument principal est celui qu'ils appellent, pour des raisons obscures, « la fronde » (« *the slingshot* ») ; voir Barwise et Perry 1983, p. 24-26 et, pour une discussion plus approfondie, Barwise et Perry 1981a ; Perry 1996.

38. Cette distinction, entre type de situation (c'est-à-dire information possible) et proposition qui affirme qu'une situation donnée est d'un certain type, est développée en détail par Barwise et Etchemendy 1987b et sert de base à leur solution au paradoxe du menteur.

mendy (1990a) discutent, toute dépendance au contexte est éliminée. Ce qu'ils retiennent principalement de la sémantique des situations, c'est l'idée d'associer à une représentation un *contenu d'information* formé d'infons. Le but principal de leur article, comme nous le verrons, est en fait de clarifier cette notion de contenu d'information et de l'enrichir suffisamment pour qu'elle permette l'étude de l'inférence ; pour eux, il s'agit donc de contribuer non seulement à l'étude du raisonnement hétérogène, mais aussi aux fondements de la sémantique des situations en général.

9.3 Information et inférence d'après Barwise et Etchemendy (1990a)

Venons-en à Barwise et Etchemendy (1990a), qui tentent d'appliquer certaines idées de la sémantique des situations à l'inférence hétérogène. Ils procèdent en trois étapes. Tout d'abord, ils développent une théorie mathématique très générale de l'information en suivant une méthode axiomatique. La structure d'*algèbre d'infons* qu'ils dégagent est applicable tant à la sémantique habituelle en termes de modèles qu'à leurs nouvelles notions non standard de contenu, qu'ils présentent dans un second temps. Pour finir, ils se tournent vers l'inférence.

a) Infons et algèbres d'infons

Barwise et Etchemendy (1990a) commencent par proposer une théorie mathématique très générale de l'information. Le point de départ est un ensemble Sit de *situations* dont on veut parler ; ces situations jouent un rôle entièrement analogue aux modèles de la sémantique habituelle. Une information possible, c'est-à-dire un type de situation ou infon, détermine une partie de Sit, à savoir l'ensemble des situations dont cet infon est vrai [39]. Pour eux, la sémantique en termes de modèles ou de mondes possibles revient à prendre comme infons tout simplement les parties de Sit, mais Barwise et Etchemendy cherchent un cadre plus général et n'interdisent pas à deux infons distincts d'être vrais du même ensemble de situations [40]. En général, les infons ne sont donc pas entièrement déterminés par l'ensemble des situations dont ils sont vrais.

39. Nos auteurs, qui évitent la terminologie traditionnelle pour lever toute ambiguïté, parlent des situations qui *supportent* un infon.

40. Ils appellent *fortement équilibrée* (« *strongly balanced* ») une algèbre d'infons qui respecte cette condition supplémentaire, comme l'algèbre d'infons formée des parties de Sit.

Pour trouver un tel cadre plus général, nos auteurs ne cherchent pas à décrire la structure interne des infons, mais adoptent une méthode axiomatique. Ils se demandent quelle structure l'*ensemble* I des infons doit avoir. Tout d'abord, on doit avoir une relation d'implication $\Rightarrow$ entre infons, qui capture le fait que certains infons découlent logiquement d'autres infons. Il faut pour cela que si $\sigma \Rightarrow \tau$, toute situation vérifiant σ vérifie aussi τ. Barwise et Etchemendy imposent en outre à la relation $\Rightarrow$ d'être réflexive, transitive, et anti-symétrique, c'est-à-dire d'être une relation d'ordre (partiel) sur I. Ensuite, il faut qu'on puisse former des conjonctions et disjonctions d'informations. La conjonction $\sigma \wedge \tau$ doit être l'information la plus faible qui est impliquée (au sens de $\Rightarrow$) par σ et τ, la disjonction $\sigma \vee \tau$ doit être l'information la plus forte qui implique tant σ que τ. En d'autres termes, il faut que deux infons admettent toujours une borne supérieure et une borne inférieure (relativement à la relation d'ordre $\Rightarrow$), c'est-à-dire que $(I, \Rightarrow, \wedge, \vee)$ forme ce qu'on appelle un *treillis* (« *lattice* » en anglais), dont on demande de plus qu'il soit *distributif*[41]. Il faut de surcroît que ces conjonctions et disjonctions se comportent comme attendu vis-à-vis des situations, c'est-à-dire que toute situation vérifiant $\sigma \wedge \tau$ vérifie aussi σ et τ, et que toute situation vérifiant σ ou τ vérifie aussi $\sigma \vee \tau$. Pour finir, Barwise et Etchemendy imposent l'existence d'un infon trivial **1**, élément maximal du treillis et vrai de toutes les situations, et d'un infon contradictoire **0**, élément minimal du treillis et vrai d'aucune situation.

Ces conditions définissent la structure d'*algèbre d'infons*[42]. Comme nous l'avons dit, un exemple simple en est $I = \mathcal{P}(Sit)$, où les infons sont les parties de Sit. Une situation s vérifie alors l'infon σ si $s \in \sigma$, $\sigma \Rightarrow \tau$ se réduit à $\sigma \subseteq \tau$, et on a tout simplement $\sigma \wedge \tau = \sigma \cap \tau$, $\sigma \vee \tau = \sigma \cup \tau$, $\mathbf{1} = Sit$ et $\mathbf{0} = \varnothing$. Mais tout l'intérêt de ce cadre général est de permettre des notions moins classiques d'information, dont nos auteurs développent ensuite un exemple.

b) Infons élémentaires et contenu d'information

La notion de *contenu d'information* utilisée par Barwise et Etchemendy pour modéliser l'inférence (et en particulier l'inférence hétérogène) repose non sur la notion d'algèbre d'infons dans toute sa généralité, mais sur des algèbres d'infons particulières. Leur idée est de construire les infons à partir d'un ensemble B dont ils appellent les éléments *infons*

41. La distributivité correspond aux axiomes $\sigma \wedge (\tau_1 \vee \tau_2) = (\sigma \wedge \tau_1) \vee (\sigma \wedge \tau_2)$ et $\sigma \vee (\tau_1 \wedge \tau_2) = (\sigma \vee \tau_1) \wedge (\sigma \vee \tau_2)$. Cela revient à demander que nos conjonctions et disjonctions se comportent comme attendu l'une vis-à-vis de l'autre.

42. Pour l'étude de l'inférence, une condition supplémentaire peut s'avérer utile : c'est l'existence, pour tous $\sigma, \tau \in I$, d'un infon $[\sigma \to \tau]$ qui, conjointement avec σ, implique τ, et qui soit l'infon le plus faible ayant cette propriété (c'est-à-dire que pour tout π vérifiant $\sigma \wedge \pi \Rightarrow \tau$, on a $\pi \Rightarrow [\sigma \to \tau]$). Dans ce cas, $(I, \Rightarrow, \wedge, \vee, \to)$ est une *algèbre de Heyting* et l'algèbre d'infons associée est dite *de Heyting*. Les algèbres d'infons construites par Barwise et Etchemendy pour capturer leur notion de contenu d'information sont de ce type.

élémentaires (« *basic infons* »). Techniquement, nous verrons que ces infons élémentaires ne font pas eux-mêmes partie de l'algèbre d'infons, mais sont les briques élémentaires qui permettent de la composer.

Qu'est-ce qu'un infon élémentaire ? L'important pour la construction qui suit est seulement qu'il détermine un ensemble de situations, celles dont il est vrai : il faut donc seulement que soit donné un ensemble B et une application $B \to \mathcal{P}(Sit)$. En pratique, cependant, Barwise et Etchemendy utilisent comme infons élémentaires des sortes de formules atomiques, c'est-à-dire des relations appliquées à des objets. Reprenons par exemple le problème des chaises discuté ci-dessus [43]. Les infons élémentaires de nos auteurs y sont formés à partir des relations suivantes : les relations unaires d'être assis (pour une personne) et d'être occupée (pour une chaise), les relations binaires (entre personnes) d'être assis immédiatement à gauche de, immédiatement à droite de, à gauche de, à droite de, la relation entre une personne et une chaise d'être assis sur, et la relation ternaire (entre personnes) d'être plus proche de quelqu'un que quelqu'un d'autre. Un infon élémentaire, ce sera donc par exemple « A est assis immédiatement à gauche de B ».

L'idée essentielle est ensuite de construire les infons à partir d'*ensembles d'infons élémentaires.* Par exemple, l'une des données du problème des chaises dit que quatre personnes doivent être assises : l'infon que Barwise et Etchemendy y associent doit comprendre les infons élémentaires « A est assis », « B est assis », « C est assis » et « D est assis ».

Barwise et Etchemendy ne veulent cependant pas accepter comme infon n'importe quel ensemble d'infons élémentaires. Ils proposent de restreindre les ensembles acceptables selon deux critères intuitifs. Tout d'abord, il est naturel de chercher à ce que les composants d'un infon soient *cohérents* entre eux : dans le problème des chaises, on ne veut par exemple pas d'infon contenant à la fois « A est assis à gauche de B » et « A est assis à droite de B ». Ensuite, écrivent-ils, on peut vouloir que nos ensembles d'infons élémentaires soient *clos* eu égard à certaines contraintes. Dans notre exemple, on peut exiger que tout ensemble qui contient « A est assis à gauche de B » contienne aussi « B est assis à droite de A ». C'est une manière de capturer l'idée qu'on ne peut pas avoir certaines informations sans en avoir automatiquement d'autres, qu'on n'a pas besoin d'inférence pour posséder explicitement. Pour capturer ces deux notions, ils proposent de compléter la donnée de B (et de l'application $B \to \mathcal{P}(Sit)$) par le choix d'un ensemble de parties de B correspondant aux ensembles qu'on appellera *cohérents* et *clos* d'infons élémentaires [44] (comme mes auteurs, j'abrégerai en *ensembles c.c.* pour la suite). En fait, leur cadre est très souple : en fonction

43. Voir section 9.1.b).

44. On appelle alors cohérente (mais pas nécessairement close) n'importe quelle sous-partie d'un tel ensemble cohérent et clos.

de l'application en vue, on peut se fixer à peu près n'importe quelles contraintes et même considérer que n'importe quel ensemble d'infons élémentaires est c.c. ; la seule exigence de Barwise et Etchemendy est que pour toute situation s, l'ensemble des infons élémentaires vrais dans s soit c.c.

Mais les ensembles d'infons élémentaires, même cohérents et clos, ne suffisent pas encore à définir les éléments de l'algèbre d'infons qu'ils veulent construire. Dans l'exemple des chaises, comment exprimer l'information « A est assis à côté de B » ? Aucun infon élémentaire ne le permet. Ce n'est possible que disjonctivement : ou bien « A est assis à gauche de B », ou bien « A est assis à droite de B ». D'une manière générale, écrivent nos auteurs, une information peut être réalisée par différents ensembles cohérents et clos d'infons élémentaires. Pour cette raison, ils définissent finalement un infon comme un *ensemble d'ensembles c.c.*, qu'il faut comprendre comme une disjonction d'ensembles c.c. dont chacun correspond à une possibilité.

Une dernière complication se présente. Pour des raisons techniques [45], Barwise et Etchemendy ne prennent pas comme infons tous les ensembles d'ensembles c.c., mais seulement ceux qui sont maximaux en un certain sens [46] : un ensemble σ d'ensembles c.c. est un infon si n'importe quel sur-ensemble c.c. Y d'un ensemble c.c. X de σ est aussi dans σ. Il est alors facile de munir ces éléments d'une structure d'algèbre d'infons. Pour σ, τ deux infons,

$$\sigma \Rightarrow \tau \quad \text{si} \quad \forall X \in \sigma,\ \exists Y \in \tau,\ Y \subseteq X$$

et comme dans le cas des sémantiques en termes de modèles, on a tout simplement $\sigma \wedge \tau = \sigma \cap \tau$ et $\sigma \vee \tau = \sigma \cup \tau$.

c) Modéliser l'inférence

Barwise et Etchemendy utilisent ensuite leur notion générale d'algèbre d'infons pour développer une théorie de l'inférence. Cette théorie est donc compatible tant avec une sémantique en termes de modèles ou mondes possibles qu'avec leur sémantique non standard en termes de contenu d'information, que je viens de présenter. Elle se distingue des conceptions plus usuelles de l'inférence en deux points principaux. D'une part, elle est fondée sur une distinction inhabituelle entre informations explicites et informations im-

45. Le problème est que si l'on accepte n'importe quel ensemble d'ensembles c.c., plusieurs infons distincts peuvent s'impliquer mutuellement au sens de la relation $\Rightarrow$ définie plus loin. La solution de Barwise et Etchemendy revient à choisir un unique ensemble d'ensembles c.c. dans chaque classe d'équivalence pour la relation d'implication mutuelle.

46. Barwise et Etchemendy appellent *ouverts* ces ensembles d'ensembles c.c. maximaux, parce qu'ils forment les ouverts d'une topologie.

plicites. D'autre part, elle emploie cinq principes sémantiques différents d'inférence, alors qu'on considère d'habitude que sur le plan sémantique, toutes les inférences sont justifiées de la même manière (par le principe que l'inférence est valide dès lors que tout modèle des prémisses est modèle de la conclusion). Cette seconde différence vise à mieux rendre compte des preuves hétérogènes, comme nous le verrons.

Tout d'abord, pour Barwise et Etchemendy, faire une inférence, c'est partir de certaines informations dont on dispose *explicitement* et en extraire de nouvelle informations qui n'étaient qu'*implicitement* présentes dans les prémisses :

> We imagine that we have been given some pieces of information (or purported information) about something, call it a situation s. Our task is to extract additional information, information that is implicit in the information given.
>
> We will model the information about s present at some stage of an inference by a set Σ of infons. We use a set, rather than a single infon, because we are interested in the pieces of information explicitly present. We want to study the move from Σ to some larger set $\Sigma' \supseteq \Sigma$ that has additional infons present, all of which were implicit in Σ taken as a whole. [47]

Plus loin, ils réitèrent qu'à chaque étape d'une inférence, les informations explicitement présentes correspondent à un ensemble d'infons et non à un infon unique qui serait, par exemple, leur conjonction (au sens de l'opération $\wedge$ de l'algèbre d'infons). Dans le cas contraire, on « brouillerait la distinction entre information explicite et implicite [48] ».

Ici, une clarification s'impose. Comme je l'ai expliqué en introduction de cette partie [49], la conception de l'information qu'il est naturel d'associer à l'idée que « l'inférence déductive valide consiste à extraire ou à rendre explicite de l'information qui n'est qu'implicite dans l'information déjà obtenue » est *syntaxique* : ce qui est explicite à chaque étape d'un raisonnement, ce sont les propositions supposées ou déjà déduites, individuées par leur forme syntaxique ; ce qui est implicite, ce sont leurs conséquences valides (quelle que soit la manière, syntaxique ou sémantique, dont on les caractérise). Dans ce genre de perspective, le terme d'« information » ne peut pas faire référence à une sémantique en termes

47. « Imaginons qu'on nous a donné des informations (ou de prétendues informations) sur quelque chose, disons une situation s. Notre tâche est d'en extraire des informations additionnelles, qui sont implicites dans l'information donnée. Nous modéliserons l'information sur s présente à une certaine étape d'une inférence par un ensemble Σ d'infons. Nous utilisons un ensemble, plutôt qu'un infon unique, parce que nous nous intéressons aux informations présentes explicitement. Ce que nous voulons étudier, c'est le passage de Σ à un ensemble plus étendu $\Sigma' \supseteq \Sigma$ contenant des infons supplémentaires, qui étaient tous implicites dans l'ensemble Σ pris comme un tout. » (Barwise et Etchemendy 1990a, p. 36-37).

48. Barwise et Etchemendy 1990a, p. 62.

49. Cf. p. 176–179.

de modèles. En effet, ce qu'on a « explicitement » ne saurait être l'ensemble des modèles des propositions déjà dérivées : cela conduirait à présupposer une forme d'omniscience logique, toutes les propositions logiquement équivalentes ayant les mêmes modèles ; or c'est tout le contraire de ce que la distinction entre explicite et implicite est censée accomplir.

Pourtant, quand nos auteurs parlent d'« information » explicite ou implicite, ils comprennent le terme sémantiquement : l'information portée par une représentation (énoncé ou diagramme), c'est l'ensemble d'infons associé, et non la forme syntaxique de la représentation. Cela se comprend bien si l'on adopte leur nouvelle notion de contenu d'information. Mais si l'on choisit comme infons des ensembles de situations, on retombe sur l'idée que ce qu'on a explicitement est un ensemble de modèles. Les affirmations de Barwise et Etchemendy sont alors difficiles à comprendre. Supposons par exemple que nos prémisses soient les axiomes de Peano (disons au premier ordre). Quelle différence y a-t-il à les considérer comme un groupe d'axiomes séparés ou à considérer leur conjonction ? Dans le premier cas, il faut connaître les ensembles de modèles de chaque axiome séparément ; dans le second cas, il faut connaître l'intersection de ces ensembles de modèles, c'est-à-dire les modèles de l'arithmétique, ce qui en un sens est plus exigeant. Cependant, ce que veut dire connaître un ensemble de modèles n'est pas clair. En réalité, il me semble que leur idée n'a guère de sens que si l'on adopte une sémantique non standard, conçue pour modéliser le raisonnement et pas seulement la validité.

À chaque étape d'un raisonnement, on a donc un ensemble d'infons correspondant aux informations explicites. Barwise et Etchemendy modélisent ensuite le raisonnement dans son ensemble par un graphe orienté [50] dont chaque nœud correspond à une étape, et donc à un ensemble d'infons. Plus précisément, une étape peut être un cas, raison pour laquelle la structure est en graphe plutôt que linéaire (voir le graphe correspondant à la solution du problème des chaises, fig. 9.2). Chaque nœud peut être « ouvert » (représenté avec un intérieur blanc, comme le dernier nœud en bas sur la fig. 9.2) ou « fermé » (représenté avec un intérieur noir). Les nœuds ouverts correspondent aux conclusions du raisonnement et les nœuds fermés aux étapes intermédiaires. Par ailleurs, un nœud peut aussi être marqué « réalisé », ce qui signifie qu'on est parvenu à reconnaître que les informations qu'il présente sont cohérentes (c'est-à-dire qu'il existe une situation où elles sont toutes vraies simultanément). Les nœuds réalisés sont représentés par des losanges, les autres par des cercles.

Barwise et Etchemendy expliquent ensuite comment construire inductivement les graphes correspondant à des raisonnement valides. Ces graphes valides, qu'ils appellent Gra-

50. Plus précisément, un graphe orienté sans cycles ni branches infinies.

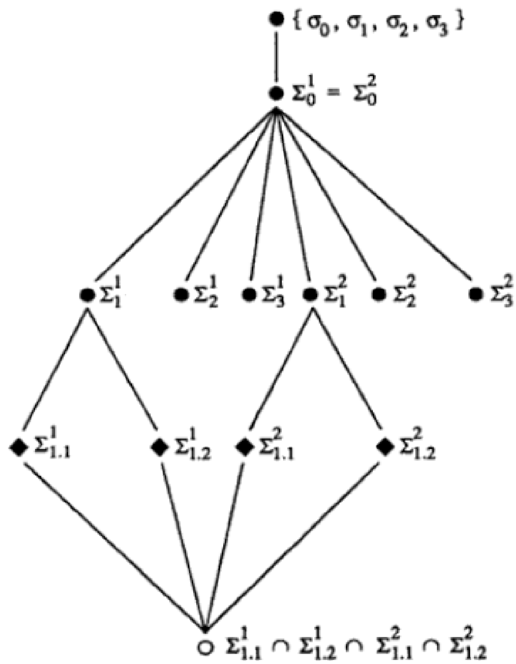


FIGURE 9.2 – Graphe inférentiel correspondant à la solution du problème des chaises, d'après Barwise et Etchemendy (1990a, p. 74)

phes d'Écoulement de l'Information (« *Information Flow Graphs* », IFGs), se lisent de la manière suivante : un graphe dont un seul nœud est ouvert (i.e. blanc), comme l'exemple ci-dessus, signifie que l'information portée par le nœud initial du graphe – celui qui est parent de tous les autres – permet de conclure l'information portée par le nœud ouvert. S'il y a plusieurs nœuds ouverts, alors l'information portée par le nœud initial permet seulement de conclure la *disjonction* des informations portées par les différents nœuds ouverts. Les Graphes d'Écoulement de l'Information sont construits selon cinq règles : *Admettre comme donné*, *Supposer*, *Subsumer*, *Fusionner*, *Reconnaître comme possible*. La première règle ne sert qu'à commencer un raisonnement. La règle *Fusionner* ne sert que pour les distinctions de cas, et la règle *Reconnaître comme possible* sert pour constater que certaines hypothèses sont cohérentes (ce que l'on peut faire sur la base du diagramme dans *Hyperproof*). Les deux règles les plus importantes sont *Supposer* et *Subsumer*. Pour les clarifier, commençons par un exemple.

La figure 9.3 indique comment construire en trois étapes le graphe d'un raisonnement très simple, qui part des infons σ_0 et σ_1 et en déduit l'infon σ_2. Tout d'abord, la règle *Admettre comme donné* permet d'introduire un premier nœud ouvert portant l'ensemble d'infons $\{\sigma_0, \sigma_1\}$. Ensuite, la règle *Supposer* permet de lui ajouter un nœud enfant portant un

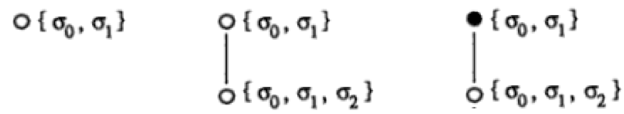


FIGURE 9.3 – Construction en trois étapes d'un Graphe d'Écoulement de l'Information correspondant à une inférence simple

infon supplémentaire, σ_2. À ce stade, rien ne dit que σ_2 est conséquence de nos hypothèses σ_0 et σ_1 : comme les deux nœuds sont ouverts, notre graphe affirme seulement la chose triviale que $\{\sigma_0, \sigma_1\}$ a pour conséquence la disjonction des informations portées par les nœuds ouverts, c'est-à-dire $(\sigma_0 \wedge \sigma_1) \vee (\sigma_0 \wedge \sigma_1 \wedge \sigma_2)$. La règle *Subsumer* permet alors de fermer le nœud initial sur la base du fait que $(\sigma_0 \wedge \sigma_1) \Rightarrow (\sigma_0 \wedge \sigma_1 \wedge \sigma_2)$.

Il peut paraître étrange de séparer la règle *Supposer*, qui dans notre exemple introduit la conclusion comme si c'était une nouvelle hypothèse, de la règle *Subsumer*. La raison en est que chacune de ces règles a d'autres usages. La règle *Supposer* sert aussi à introduire les différentes branches d'un raisonnement par cas. Dans ce contexte, la règle *Subsumer* sert à fermer le nœud parent des différents cas que l'on considère : l'appliquer revient alors à constater que ces cas sont exhaustifs. La règle *Subsumer* permet aussi de fermer un nœud dont les informations sont contradictoires, ce qui est utile pour raisonner par l'absurde.

Voici une description complète des différentes règles :

1. *Admettre comme donné* : on peut commencer un graphe par un nœud unique, ouvert, étiqueté par n'importe quel ensemble d'infons.

2. *Supposer* : on peut ajouter un fils à n'importe quel nœud, à condition que l'ensemble d'infons portés par ce fils contienne l'ensemble d'infons portés par son père (typiquement, il comprendra en outre des infons supplémentaires). Cette règle a deux usages principaux : elle permet d'introduire une conclusion, comme dans l'exemple précédent, mais aussi d'ouvrir les différentes branches d'une distinction de cas.

3. *Subsumer* : on peut marquer un nœud ouvert comme fermé s'il y a d'autres nœuds ouverts qui en sont conséquence, au sens suivant. Notons a le nœud que nous voulons fermer, Σ_a l'ensemble des infons qu'il porte, et $\widehat{\Sigma_a}$ la conjonction de ces infons (dans l'exemple traité ci-dessus, $\Sigma_a = \{\sigma_0, \sigma_1\}$ et $\widehat{\Sigma_a} = \sigma_0 \wedge \sigma_1$). Notons E un ensemble de nœuds ouverts qui ne contient pas a ; pour chaque nœud $e \in E$, notons Σ_e l'ensemble des infons qu'il porte et $\widehat{\Sigma_e}$ leur conjonction. (L'ensemble E peut se réduire à un seul élément, comme dans l'exemple précédent, ou contenir plusieurs éléments, par exemple plusieurs fils de a correspondant à des cas dont on veut montrer qu'ils épuisent toutes les possibilités.) La règle *Subsumer* nous permet de marquer a comme

fermé dès lors que $\widehat{\Sigma_a} \Rightarrow \bigvee_{e \in E} \widehat{\Sigma_e}$.

4. *Fusionner* : on peut ajouter un fils commun à plusieurs nœuds ouverts. L'ensemble d'infons portés par ce nouveau nœud sera l'intersection des ensembles d'infons portées par ses parents.

5. *Reconnaître comme possible* : on peut marquer comme « réalisé » un nœud s'il existe une situation dans laquelle tous les infons qu'il porte sont conjointement vrais.

Pour Barwise et Etchemendy, ces règles de construction des graphes correspondent à des principes généraux censés permettre de justifier, au niveau informationnel, n'importe quelle règle d'inférence syntaxique. Ils écrivent ainsi que « nous avançons, du moins à titre provisoire, la thèse que *tout raisonnement déductif valide peut être justifié par une combinaison de ces cinq principes*[51] ».

Pourquoi ce cadre exotique, avec des graphes et cinq principes sémantiques ? Essentiellement parce que nos auteurs veulent se rapprocher du genre de raisonnement informel avec diagrammes que leur a révélé l'examen de *Tarski's World*, et qui a motivé leur développement du système *Hyperproof*[52]. Les embranchements que permettent *Supposer* et *Subsumer* servent ainsi à rendre compte du raisonnement par cas ; le principe *Reconnaître comme Possible* autorise à constater que certaines hypothèses sont cohérentes, comme le permettent tant *Tarski's World* que *Hyperproof*[53].

Au-delà de ces innovations, il faut toutefois remarquer que la règle cruciale, *Subsumer*, est très proche du traitement habituel de la conséquence sémantique. Imaginons en effet que l'on prenne pour infons des ensembles de modèles, et que l'on veuille justifier une inférence simple, comme le passage, dans notre exemple ci-dessus[54], de σ_0 et σ_1 à σ_2. La règle *Subsumer* nous autorise à fermer le nœud initial si tout modèle appartenant à σ_0 et σ_1 appartient aussi à σ_2, c'est-à-dire si σ_2 est conséquence sémantique de $\sigma_0 \wedge \sigma_1$ au sens le plus banal.

S'il y a une différence de fond entre cette théorie de l'inférence et le traitement sémantique habituel, elle ne peut donc apparaître que si, à la suite de Barwise et Etchemendy, on choisit de prendre comme infons quelque chose d'autre que des ensembles de modèles, par exemple leur nouvelle notion de contenu d'information.

51. « We tentatively put forward the thesis that *all valid deductive reasoning can be justified by a combination of these five principles.* » (Barwise et Etchemendy 1990a, p. 61).

52. Voir ci-dessus, section 8.1.

53. Voir ci-dessus, p. 217 et p. 219–220.

54. Cf. p. 247–248 et fig. 9.3.

9.4 La position ambiguë du contenu d'information

Le cadre de Barwise et Etchemendy admet des situations, qui correspondent aux modèles de la sémantique usuelle, et des représentations – phrases ou diagrammes – qui correspondent à la syntaxe. Entre les deux, le contenu d'information occupe une position ambiguë et problématique, sur laquelle je voudrais revenir.

À première vue, le but de nos auteurs est de clarifier l'idée d'un raisonnement conduit au niveau de l'information plutôt que selon des règles syntaxiques rigides. Ils écrivent ainsi :

> Our strategy is to work at the level of information, rather than its representation, and to try to find a few basic principles of information flow that suffice to account for all valid inference. [55]

Les principes de base en question sont ceux que j'ai présentés plus haut : *Admettre comme donné*, *Supposer*, *Subsumer*, *Fusionner*, *Reconnaître comme possible*.

Il y a cependant une vraie ambiguïté sur le statut de ces principes. Barwise et Etchemendy affirment que ce sont des principes sémantiques selon lesquels on peut *justifier* un raisonnement. Ils écrivent par exemple :

> [W]hen one works with any given form or forms of representation, be they visual, linguistic, or whatever, one can search for rules that can be *justified* by appeal to these basic principles. [56]

Plus loin, quand ils appliquent leurs outils à l'analyse du problème des chaises, ils insistent :

> We emphasize that we are presenting a mathematical model that shows the reasoning above [57] to be valid. This is a distinct enterprise from modeling the reasoning itself. It is analogous to a model-theoretic proof of the soundness of principles used in a piece of syntactic reasoning. Needless to say, this is not something that the reasoner does in the course of the reasoning. [58]

55. « Notre stratégie est de travailler au niveau de l'information plutôt que de sa représentation, et d'essayer de trouver un petit nombre de principes de base d'écoulement de l'information qui suffisent à rendre compte de toutes les inférences valides » (Barwise et Etchemendy 1990a, p. 60).

56. « [Q]uand on travaille avec une ou plusieurs formes de représentations, quelles qu'elles soient (visuelles, linguistiques, ou autre), on peut alors chercher des règles qui peuvent être *justifiées* d'après ces principes de base. » (Barwise et Etchemendy 1990a, p. 60 ; je souligne).

57. Il s'agit de la solution du problème des chaises.

58. « Soulignons que ce que nous présentons est un modèle mathématique qui montre que le raisonnement ci-dessus est valide. Ce n'est pas la même chose que de modéliser le raisonnement lui-même. Ce que nous faisons est analogue à une preuve par la théorie des modèles que les principes utilisés au cours d'un raisonnement syntaxique sont corrects. Inutile de préciser que ce n'est pas quelque chose que celui qui raisonne fait au cours de son raisonnement. » (Barwise et Etchemendy 1990a, p. 68).

Comme je l'ai déjà dit, on risque alors d'en arriver à la conclusion décevante qu'il n'y a que peu de différence entre le cadre de nos auteurs et la sémantique habituelle en termes de modèles. Au fond, les seules différences substantielles seraient les règles *Fusionner* (qui permet un traitement allégé des distinctions de cas – un avantage mineur) et surtout *Reconnaître comme possible* (qui admet que certaines représentations puissent avoir un rapport d'isomorphisme assez étroit avec leurs modèles pour qu'on puisse reconnaître syntaxiquement que certaines propositions sont compatibles, c'est-à-dire admettent un modèle). La règle *Subsumer*, qui sert le plus fréquemment aux inférences, se rapproche en revanche d'une notion banale de conséquence logique ; elle s'y réduit même entièrement, comme nous l'avons vu, si on prend comme infons des ensembles de situations.

Pourtant, le travail de nos auteurs devient largement incompréhensible si le contenu d'information ne sert qu'à la justifications *a posteriori* d'inférences conduites syntaxiquement, et ne vise pas au moins en partie à modéliser les inférences elles-mêmes. D'abord, nous avons vu que Barwise et Etchemendy insistent sur la distinction entre informations explicites et implicites à une étape d'un raisonnement. Or que sont les informations explicites sinon des informations directement accessibles à l'agent qui raisonne, c'est-à-dire accessibles sans inférence ? Cette limitation aux informations explicites n'a pas d'intérêt du strict point de vue de la justification ; elle n'a aucune place dans les formulations habituelles de la conséquence logique en termes de modèles. En second lieu, s'il s'agissait simplement de justification, Barwise et Etchemendy pourraient s'en tenir à prendre comme infons des ensembles de situations. Leur développement d'une notion sophistiquée de contenu d'information, dans laquelle les infons sont définis à partir d'ensembles de formules atomiques ou « infons élémentaires », n'a d'intérêt que si l'on cherche une sémantique plus proche de ce qui est accessible à l'agent qui raisonne. Si le contenu d'information ne sert qu'à justifier les inférences, il faudrait conclure que ces longs développements, empruntés à la sémantique des situations, n'ont en fait rien à faire dans le contexte logique de cet article.

Sun-Joo Shin, d'ailleurs, interprète spontanément les principes informationnels de Barwise et Etchemendy comme des règles d'inférence :

> A major role Barwise and Etchemendy assigned to the information theory in the heterogeneous reasoning project is that their *inference rules* are written at the level of information, not at the level of representation. They allow us to manipulate pieces of information, not to manipulate units of representation. [...] I find the main framework presented in "Information, Infons, and Inference" [59] intuitive and convincing as well as clever and revolutionary for the

59. Barwise and Etchemendy 1990a.

> following reasons. At a naïve level, the reason why we judge certain steps in our reasoning valid and others not valid has nothing to do with any form of representation. In many cases, we are not aware of what kind of representation has been used in the reasoning process. […] Instead, we intuitively think of valid reasoning in terms of information […]. [60]

Bien sûr, cette interprétation n'est pas sans difficultés. La règle « Subsumer », qui est la plus importante, fait appel à la conjonction des infons explicitement disponibles, donc repose sur un genre d'omniscience quant aux conséquences de nos prémisses – surtout, mais pas seulement, si l'on interprète les infons comme des ensembles de situations.

Cette ambiguïté permanente vient de leur double objectif. D'un côté, leur tâche la plus urgente et la plus claire était de montrer que l'on peut raisonner rigoureusement en combinant différentes formes de représentation ; pour cela, il leur fallait pouvoir parler de conséquence logique. D'un autre côté, le but plus général de leur projet sur les diagrammes était bien de rapprocher nos modèles logiques du raisonnement réel, ce qui impliquait de clarifier l'idée de raisonnement d'après le sens.

9.5 Conclusion : une impasse seulement en apparence

Le concept de contenu d'information de Barwise et Etchemendy (1990a) répond ainsi à deux motivations difficilement conciliables : d'une part, à la suite de la sémantique des situations, décrire ce que savent les agents qui raisonnent ; d'autre part, à l'image de la sémantique usuelle, rendre compte de la validité des inférences. Ces deux tâches sont en tension, et le résultat n'est véritablement adapté ni à l'une ni à l'autre. Si les infons de la sémantique des situations semblent raisonnablement adaptés à la première, la seconde accule essentiellement Barwise et Etchemendy à reconstruire les ensembles de modèles sous forme d'ensembles d'ensembles d'infons. À ce compte-là, pourquoi ne pas se contenter de modèles, qui sont bien plus simples et (comme nous l'avons vu dans le premier chapitre) parfaitement capables de traiter l'inférence hétérogène ? Peut-être est-ce la conclusion à la-

60. « Un rôle majeur que Barwise et Etchemendy attribuent à la théorie de l'information dans leur projet sur le raisonnement hétérogène est que leurs *règles d'inférence* sont écrites au niveau de l'information, non au niveau de la représentation. Elles nous permettent de manipuler des informations et non des représentations. […] Je trouve le cadre principal présenté dans l'article "Information, Infons, and Inference" non seulement intuitif et convaincant, mais aussi astucieux et révolutionnaire, pour les raisons suivantes. À un niveau naïf, la raison pour laquelle nous jugeons certaines étapes de nos raisonnements valides et d'autres non n'a rien à voir avec une forme de représentation en particulier. Dans de nombreux cas, nous n'avons même pas conscience de la forme de représentation utilisée au cours du raisonnement. […] Au contraire, c'est en termes d'information que nous concevons intuitivement le raisonnement valide […]. » (Shin 2004, p. 98–100 ; je souligne).

quelle sont arrivés nos auteurs eux-mêmes à la suite des travaux de Shin sur les diagrammes de Venn. En tout cas, dans les travaux qui visent à montrer que l'on peut raisonner rigoureusement avec des diagrammes (comme ceux de Shin, Hammer ou encore Luengo) on ne trouve plus guère d'infons.

Néanmoins, ces deux motivations peuvent être dissociées : on peut conserver les infons sans la machinerie complexe étudiée ci-dessus. C'est ce que Barwise tente de faire tout au long des années 1990. Cela le conduit à une nouvelle théorie mathématique de l'information, développée en collaboration avec Jerry Seligman [61].

Dans le cas particulier des diagrammes, utiliser des ensembles d'infons plutôt que de modèles s'est avéré très fructueux, non pour justifier la validité de raisonnements hétérogènes – tâche pour laquelle les modèles conviennent mieux – mais pour expliquer pourquoi un diagramme peut être inférentiellement plus efficace que des phrases. C'est ce que fait Atsushi Shimojima, l'un des étudiants de Barwise, sur la base de la nouvelle théorie de l'information de Barwise et Seligman. Son travail fait l'objet du prochain chapitre.

61. Barwise 1993 ; Barwise et Seligman 1994, 1997.

Chapitre 10

Dynamique du raisonnement diagrammatique

10.1 Introduction : la distinction linguistique-graphique

À la fin de leur article programmatique ([1991] 1996b), Barwise et Etchemendy se demandent ce qui distingue les « représentations visuelles » et les rend « si utiles, en mathématiques, pour la découverte, la démonstration et l'enseignement [1] ». En quoi sont-elles différentes des « représentations linguistiques » dont traite d'ordinaire la logique ?

Une idée classique, explicitée par Peirce, est de distinguer deux sortes de signes : les *symboles*, arbitraires et conventionnels, et les *icônes*, qui ont une ressemblance (peut-être seulement de structure) avec ce qu'elles représentent [2]. Suivant une intuition apparentée, Barwise et Etchemendy écrivent qu'« un bon diagramme est isomorphe, ou du moins homomorphe, à la situation qu'il représente [3] » (les structures d'interprétation tenant lieu, dans leurs systèmes formels, de situations représentées). Pour les diagrammes de Venn, on peut par exemple établir un morphisme entre régions du plan munies de la relation

1. « In the concluding section, we present some thoughts about what it is that makes diagrams and other forms of visual representation so useful for mathematical discovery, proof, and pedagogy. » (Barwise et Etchemendy [1991] 1996b, p. 4).

2. Sur cette distinction, voir pour commencer le texte assez précoce Peirce 1885 (où il parle de « *tokens* » plutôt que de symboles) ; sur la sémiotique de la maturité, voir PE, vol. 2, p. 289–299 et plus largement les textes rassemblés dans PCP, vol. 2, §227–308. Pour des commentaires, on pourra consulter Short 2007, chap. 8, Hookway 1985, chap. IV ou encore Chauviré 2008. D'une manière générale, l'œuvre de Peirce est assez difficile d'accès, entre autres pour des raisons éditoriales : l'édition classique PCP est un collage thématique assez inextricable d'extraits de manuscrits, et la future édition de référence Peirce 1982- n'a encore atteint que l'année 1892. L'anthologie PE peut servir de point de départ.

3. « a good diagram is isomorphic, or at least homomorphic, to the situation it represents, at least along certain crucial dimensions » (Barwise et Etchemendy [1991] 1996b, p. 24).

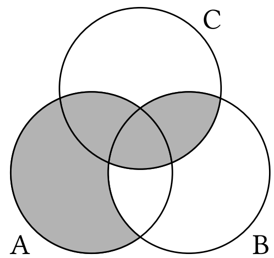


FIGURE 10.1 – Diagramme de Venn pour « Tout A est B » et « Nul B n'est C »

d'inclusion (spatiale) et ensembles munis de l'inclusion (ensembliste). Les diagrammes de *Tarski's World* ou d'*Hyperproof* ont aussi, comme nous l'avons vu, une relation suffisamment étroite avec leurs structures d'interprétation pour que l'on puisse établir sur leur base des résultats de cohérence [4]. Cependant, clarifier ce critère d'homomorphisme ne va pas de soi. Il faut entre autres le rendre suffisamment précis et restrictif pour ne pas intégrer du même coup tous les langages formels, puisque la fonction d'interprétation habituellement utilisée en sémantique, qui est définie récursivement sur la structure des formules, peut aussi être considérée comme un morphisme [5] ; c'est ce que tentent ultérieurement Barwise et Hammer ([1994] 1996, p. 71-74).

Le critère qui m'intéresse ici est différent, mais peut également être relié à une remarque de Peirce. Pour clarifier l'intérêt pour le raisonnement des « icônes », c'est-à-dire des signes qui ressemblent à leur signifié, il écrit :

> [A] great distinguising property of the icon is that by the direct observation of it other truths concerning its object can be discovered than those which suffice to determine its construction. [6]

Les diagrammes de Venn (ou d'Euler) en sont un bon exemple : pour construire le diagramme 10.1, les prémisses « Tout A est B » et « Nul B n'est C » suffisent ; pourtant, on peut ensuite obtenir la conclusion (« Nul A n'est C ») par simple observation. En un sens, comme l'écrit le peircéen Frederik Stjernfelt, ce critère n'est qu'une « élaboration du concept de ressemblance [7] », l'idée étant que c'est parce qu'elle a une structure commune avec son objet que l'icône peut nous en apprendre davantage sur lui. D'un autre point de vue toutefois,

4. Cf. *supra*, section 8.1, en part. p. 217.
5. Cette objection est par exemple soulevée par Stenning 2000, p. 137.
6. « Une grande propriété caractéristique de l'icône est qu'en l'observant directement, on peut découvrir d'autres propriétés de son objet que celles qui suffisent à déterminer sa construction. » (PCP, vol. 2, §279. Manuscrit intitulé « That Categorical and Hypothetical Propositions are One in Essence, with Some Connected Matters », dont divers extraits sont éparpillés dans PCP ; l'éditeur, Arthur Burks, date notre passage de 1895, cf. PCP, vol. 8, p. 286.)
7. Cf. Stjernfelt 2006, p. 71 ; Stjernfelt 2007, p. 78 et plus généralement chap. 4.

cela donne de la substance à l'idée vague de ressemblance en fournissant un « critère opérationnel [8] » pour reconnaître l'icônicité. On pourrait même aller plus loin et considérer que cela permet de ne plus parler de ressemblance du tout.

C'est une idée apparentée que Barwise et Etchemendy formulent dans le langage de la sémantique des situations :

> Diagrams are physical situations. [...] As such, they obey their own set of constraints. In our example from *Hyperproof*, when we represent tetrahedron *a* as large, a host of other facts are thereby supported by the diagram. By choosing a representational scheme appropriately, so that the constraints on the diagrams have a good match with the constraints on the described situation, the diagram can generate a lot of information that the user never need infer. Rather, the user can simply read off facts from the diagram as needed. [9]

Rappelons qu'une « contrainte », en sémantique des situations, est une régularité de la forme suivante : si telle situation est de tel type (vérifie tel infon), alors telle situation est de tel autre type (vérifie tel autre infon). En l'occurence, les contraintes dont il s'agit sont « locales [10] », c'est-à-dire ne portent que sur une seule situation, le diagramme : elles disent que dès qu'un diagramme vérifie tels infons, il en vérifie aussi d'autres.

Illustrons leur idée sur un exemple plus simple qu'*Hyperproof*. Reprenons le diagramme de Venn ci-dessus (fig. 10.1), mais cette fois, pour simplifier, dans le cadre ensembliste de Shin plutôt que dans un cadre syllogistique. En utilisant les notations de Shin [11] (qui utilise brièvement le formalisme des infons [12]), on peut dire que notre diagramme vérifie par

8. Stjernfelt 2006, p. 71.

9. « Les diagrammes sont des situations physiques. [...] En tant que telles, ils obéissent à un ensemble de contraintes qui leur est propre. Dans notre exemple tiré d'*Hyperproof*, lorsque nous représentons le tétraèdre *a* comme étant grand, une pléthore d'autres faits sont par là même supportés par le diagramme. Si l'on choisit un schème représentationnel approprié, de telle manière que les contraintes qui pèsent sur les diagrammes s'alignent bien avec les contraintes qui pèsent sur la situation décrite, le diagramme peut engendrer beaucoup d'informations que leur utilisateur n'a jamais besoin d'inférer. Au contraire, il peut se contenter de les lire sur le diagramme en fonction de ses besoins. » (Barwise et Etchemendy [1991] 1996b, p. 23).

10. C'est le terme qu'utilisent, par exemple, Barwise et Shimojima 1995, p. 15.

11. Si R_1 et R_2 désignent deux régions du diagramme, Shin note $R_1 + R_2$ leur union, R_1-*and*-R_2 leur intersection, et $R_1 - R_2$ la partie de R_1 qui n'est pas dans R_2 (voir Shin 1994, p. 52-53). Les infons que je donne sont alors faciles à comprendre. Notons seulement que le dernier argument d'un infon, qui est toujours un 0 ou un 1, indique ce que les théoriciens des situations appellent sa « polarité », c'est-à-dire s'il est affirmatif ou négatif. Ainsi, $\langle\langle \text{Shading}, A; 0 \rangle\rangle$ signifierait que la région A n'est *pas* grisée.

12. Elle emploie des infons dans son premier article, Shin [1991] 1996 ainsi que dans son livre : Shin 1994, section 3.3.2, p. 68–71. Dans le livre, toutefois, leur rôle est largement cosmétique et je n'ai pas eu de mal à présenter son travail sans les mentionner.

construction les infons ou « faits représentants [13] » suivants :

$$\langle\!\langle \text{Shading}, A - B; 1 \rangle\!\rangle \quad \text{et} \quad \langle\!\langle \text{Shading}, B\text{-}and\text{-}C; 1 \rangle\!\rangle,$$

qui, selon les conventions de lecture des diagrammes de Venn, sont équivalents aux assertions $A \setminus B = \varnothing$ et $B \cap C = \varnothing$. Or – c'est là qu'il y a une « contrainte » – un diagramme qui vérifie ces infons vérifie aussi

$$\langle\!\langle \text{Shading}, A\text{-}and\text{-}C; 1 \rangle\!\rangle$$

et nous donne donc la conclusion $A \cap C = \varnothing$. Celle-ci, dans les termes de Barwise et Etchemendy, peut être *lue directement* (« *read off* ») sur le diagramme, alors que sans lui, elle ne s'obtient à partir des assertions $A \setminus B = \varnothing$ et $B \cap C = \varnothing$ que *via* une inférence.

Ce qui est spécifique aux diagrammes ou figures, dans cette perspective, c'est donc qu'en exploitant des contraintes comme la précédente, ils nous permettent d'en apprendre davantage par simple observation. Atsushi Shimojima approfondit cette idée dans la deuxième moitié des années 1990 [14] ; c'est à ses travaux qu'est consacré ce chapitre.

Toutefois, il est hasardeux de fonder une distinction diagrammatique-linguistique sur quelque chose comme le critère « opérationnel » d'icônicité, dont l'idée précédente n'est au fond qu'une variante. Le cas de Peirce le montre bien. J'ai sorti de son contexte la citation ci-dessus sur l'avantage des icônes, mais voici les phrases qui la précèdent :

> [A]n algebraic formula is an icon, rendered such by the rules of commutation, association, and distribution of the symbols. It may seem at first glance that it is an arbitrary classification to call an algebraic expression an icon; that it might as well, or better, be regarded as a compound conventional sign. But it is not so. For a great distinguishing property of the icon [etc., cf. *supra* p. 256]. [15]

Si ce qui distingue les icônes est seulement qu'en les observant, on peut en apprendre davantage sur ce qu'elles représentent, alors la catégorie d'icône est extrêmement large et ne

13. Cf. Shin 1994, p. 68.

14. Pour un aperçu du projet général tel que le concevait Barwise, un bon point de départ est Barwise et Shimojima 1995. La meilleure introduction à la version actuelle de la théorie de Shimojima est l'exposé Shimojima 2014 ou son ouvrage de 2015. Shimojima 1999c discute la distinction linguistique-graphique. Il a publié toute une série d'articles qui permettent de retracer l'évolution de ses idées à partir de sa thèse (Shimojima 1996a) ; voir en particulier Shimojima 1996b, 1999a,b, 2003, 2008.

15. « Une formule algébrique est une icône, rendue telle par les règles de commutation, d'association, et de distribution des symboles. À première vue, il peut sembler que c'est une classification arbitraire que d'appeler une expression algébrique une icône ; que l'on pourrait aussi bien, ou mieux, la considérer comme un sign conventionnel composé. Mais il n'en est rien. En effet, une grande propriété caractéristique de l'icône [...]. » (PCP, vol.2, §279. Pour plus de détails sur ce texte, voir *supra*, note 6 p. 256.)

respecte guère l'idée intuitive de ce qu'est un diagramme. Cette tension, comme le montre Stjernfelt (2006), est présente chez Peirce lui-même : lorsque celui-ci met au point des systèmes logiques diagrammatiques [16], il se trouve obligé, pour les distinguer des logiques symboliques, d'élaborer une définition plus restrictive de l'icônicité. Shimojima se trouve lui aussi confronté à ce problème. Il tente de le résoudre en distinguant différentes sortes de contraintes : les contraintes « nomiques » (par exemple, écrit-il, « géométriques » ou « topologiques »), comme celle de notre diagramme de Venn, et les contraintes « conventionnelles », qui seraient seules présentes dans le cas des formules [17]. Il essaie ultérieurement d'établir une gradation plus progressive entre degrés de « graphicalité [18] ». En fin de compte, il n'est pas clair qu'il parvienne à une distinction convaincante.

Je voudrais ici laisser cette question de côté, comme le fait d'ailleurs Shimojima lui-même dans ses publications récentes [19]. L'intérêt pour moi de son travail est plutôt qu'il propose, au passage, une analyse fine de l'utilité de certaines représentations et une typologie des différentes manières dont l'inspection d'une représentation peut nous apprendre du nouveau.

10.2 L'insuffisance des approches précédentes

Le diagramme de Venn de la section précédente est caractéristique des exemples qui intéressent Shimojima : son travail porte sur des cas où tracer un diagramme permet de progresser par simple inspection. À titre de préalable, je voudrais souligner que les approches que nous avons abordées jusqu'ici ne rendent pas bien compte de ces cas-là.

Tout d'abord, il n'est pas très éclairant de parler de différences « computationnelles » sur fond d'équivalence « informationnelle ». Certes, on peut traduire un diagramme de Venn en une suite de phrases ; sans doute pourrait-on aussi, dans l'esprit de Larkin et Simon [20], défendre l'idée que le diagramme permet de regrouper efficacement en un seul endroit tout ce que l'on sait, mettons, sur l'ensemble A. Mais dans l'exemple précédent, ce qui importe est que l'on part de *moins* que ce que le diagramme permet d'observer directement. Sachant seulement que $A \setminus B = \varnothing$ et $B \cap C = \varnothing$, quelles sont les relations entre

16. Peirce a en particulier développé un système logique diagrammatique, les « graphes existentiels », capable de traiter ce que nous appelons logique propositionnelle et logique du premier ordre ; voir Shin 2002.

17. Cf. par exemple Shimojima 1999c.

18. Voir Shimojima 2003.

19. De manière significative, la question est largement absente de son livre Shimojima 2015. Ses exemples paradigmatiques vont d'ailleurs au-delà des diagrammes ; il discute beaucoup les tableaux, par exemple. Il est néanmoins clair que, de son point de vue, les formules algébriques ou logiques ne présentent que peu, ou pas du tout, les phénomènes qui l'intéressent (sur ce point, voir Shimojima 2003).

20. Voir sections 4.5 et 5.1.

A et C ? Ce que le diagramme facilite, d'après Barwise, Etchemendy et Shimojima, c'est de répondre à cette question, c'est-à-dire de passer de nos hypothèses à $A \cap C = \varnothing$. Or, comme je l'ai déjà signalé [21], on ne peut pas raisonnablement dire que le diagramme soit informationnellement équivalent *à nos hypothèses* ; il l'est plutôt à un ensemble de phrases qui comprend *déjà* $A \cap C = \varnothing$.

Il y a bien un sens en lequel le diagramme a le même contenu que les hypothèses $A \smallsetminus B = \varnothing$ et $B \cap C = \varnothing$: tous deux admettent les mêmes modèles, si l'on associe des modèles tant aux diagrammes qu'aux énoncés à la manière des chapitres 7 et 8. Mais précisément, cela montre bien qu'à ce niveau sémantique-là, il n'y a pas de différence entre eux et donc qu'on ne peut pas expliquer l'avantage du diagramme. Face à cette difficulté, Shimojima suit l'idée de Barwise et Etchemendy, présentée au chapitre précédent, d'introduire une sémantique plus fine pour capturer ce que diagramme ou énoncés disent « explicitement ».

10.3 La théorie de Shimojima

a) Le cadre formel

À l'arrière-plan du travail de Shimojima, il y a la théorie de l'information de Barwise et Seligman [22], aussi appelée « théorie des canaux » (« *channel theory* »). Heureusement pour nous, Shimojima n'en utilise qu'une version simplifiée, et nous pouvons donc nous limiter ici à quelques rudiments [23].

Un système représentationnel, pour Shimojima, comporte deux niveaux [24]. Au premier niveau, il y a les situations « sources », qui sont les représentations elles-mêmes (par exemple les diagrammes bien formés du système de Shin), et les situations « cibles », qui sont les situations représentées (par exemple, les modèles possibles des diagrammes de Shin) ; lorsque l'on utilise un diagramme pour représenter certains ensembles particuliers, on établit une relation de représentation entre une situation source et une situation cible. Au second niveau, il y a les « types » de situations sources et les « types » de situations cibles (que Shimojima n'exprime pas dans le formalisme des infons, mais l'idée sous-jacente est la même). C'est à ce second niveau qu'opèrent les conventions de représentation, que

21. Cf. section 6.4.

22. Cf. Barwise et Seligman (1997). Cet ouvrage n'est pas encore paru lorsque Shimojima publie ses premiers articles, mais Barwise et Seligman avaient déjà développé l'essentiel de leur théorie (voir par exemple Barwise 1993 ; Barwise et Seligman 1994).

23. Barwise et Seligman 1997, chap. 20 reformulent le travail de Shimojima dans le cadre de leur théorie générale ; Shimojima et Barker-Plummer 2014 tentent de motiver les concepts généraux de Barwise et Seligman et d'expliquer pourquoi ils sont utiles pour analyser les représentations.

24. Cf. par exemple Shimojima 2015, section 2.2.

$$\begin{tikzcd}
\{\ A \setminus B = \varnothing \arrow[from=d, "\text{indique}"] & B \cap C = \varnothing\ \} \arrow[from=d, "\text{indique}"] \arrow[r, "\text{contrainte}"] & A \cap C = \varnothing \arrow[from=d, "\text{indique}"] \\
\left\{\begin{array}{c}\text{La région } A - B\\ \text{est grisée}\end{array}\right. & \left.\begin{array}{c}\text{La région } B\textit{-et-}C\\ \text{est grisée}\end{array}\right\} \arrow[r, "\text{contrainte}"'] & \begin{array}{c}\text{La région } A\textit{-et-}C\\ \text{est grisée}\end{array}
\end{tikzcd}$$

FIGURE 10.2 – Structure d'un exemple de « *free ride* » (schéma inspiré de Shimojima 2014, p. 8, 2015, §2.3)

Shimojima modélise par une relation d'« indication » entre types sources et types cibles. Ainsi, dans le cas des diagrammes de Venn, le type source « la région A est grisée » *indique* le type cible « l'ensemble A est vide ». Il y a donc deux relations représentationnelles distinctes : une relation entre *situations* et une relation entre *types*. Cette distinction s'inscrit dans le prolongement de la sémantique des situations et ultimement de la distinction d'Austin entre les aspects démonstratif et descriptif d'une assertion [25].

b) « Passagers clandestins » (« *Free rides* »)

L'avantage de certains diagrammes pour le raisonnement vient alors de ce qu'ils sont soumis à des contraintes : un diagramme de Shin bien formé, par exemple, ne peut pas être de certains types sans être du même coup d'autres types, comme nous l'avons vu plus haut. Si ces contraintes sur les diagrammes correspondent à des contraintes sur les situations cibles, le diagramme permet des inférences correctes, que Shimojima appelle des « passagers clandestins » (« *free rides* »). Le diagramme de Venn ci-dessus en est un bon exemple ; on peut schématiser les contraintes en jeu comme à la fig. 10.2. (Pour souligner que les types du bas portent sur les diagrammes alors que les types du haut portent sur les ensembles qui interviennent dans leurs modèles, j'ai conservé la notation de Shin et noté par exemple A-et-B l'intersection des *régions* A et B d'un diagramme [26].) Remarquons que les relations d'indication entre types sources et types cibles (flèches verticales) sont toutes, ici, des instances d'un même principe : si une région est grisée, l'ensemble correspondant est vide.

Ce phénomène est répandu. L'exemple introductif favori de Shimojima exploite non pas

25. Voir ci-dessus, section 9.2, en part. p. 239–240.
26. Voir ci-dessus, note 11 p. 257.

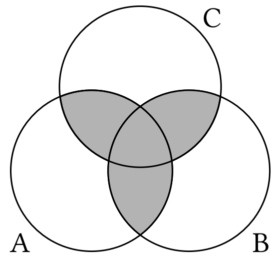


FIGURE 10.3 – Un diagramme de Venn symétrique par permutation de A, B, C

des contraintes sur l'inclusion entre régions du plan, mais sur l'ordre d'éléments disposés le long d'une ligne. Imaginons, écrit-il, que l'on représente l'ordre d'arrivée des participants à une course en écrivant le nom des participants sur une ligne, selon la convention que le nom A est placé à gauche du nom B si A est arrivé avant B. Si je sais que A est arrivé avant B et B avant C et que j'applique la convention précédente, j'obtiendrai le « diagramme positionnel »

$$A \qquad B \qquad C$$

qui montre nécessairement (du fait d'une contrainte à laquelle est soumis le système des diagrammes positionnels) que A est arrivé avant C. Ce passager clandestin, d'après Shimojima, nous évite de faire intervenir la transitivité de l'ordre d'arrivée *via* une inférence explicite [27]. Mumma (2010, p. 278-279) considère ce genre d'exemples comme caractéristiques des avantages que fournissent les figures en géométrie élémentaire.

c) « Signification dérivée » (« *Derivative meaning* »)

Les passagers clandestins de Shimojima n'épuisent pas les bénéfices que l'on peut tirer d'un diagramme : on fait souvent des observations qui vont au-delà des conventions sémantiques élémentaires qui le définissent. Par exemple, avec un peu d'habitude des diagrammes de Venn, on peut remarquer directement sur celui de la fig. 10.3 que les ensembles A, B et C sont deux à deux disjoints. Plus simplement, on peut remarquer que ce diagramme est symétrique, et donc que les ensembles A, B et C jouent des rôles invariants par permutation. On peut aussi penser aux visualisations de données de la section 5.2. Dans tous ces cas, on observe sur le graphique ou sur le diagramme des choses qui vont au-delà des

27. Remarquons au passage que la même contrainte qui permet le passager clandestin a ici une autre conséquence : il est impossible de représenter que A est arrivé avant B et avant C sans trancher la question supplémentaire de savoir si B est arrivé avant C ou inversement. C'est un phénomène que nous avons déjà rencontré pour les diagrammes d'Euler et que Shimojima appelle « sur-spécificité » (« *over-specificity* ») ; voir par exemple Shimojima 2015, chap. 3.

	USA	CHN	KOR	BRA	TUR	SRB
USA	–	○	○	○	○	○
CHN	●	–	○	●	○	○
KOR	●	●	–	○	●	○
BRA	●	○	●	–	○	○
TUR	●	●	○	●	–	○
SRB	●	●	●	●	●	–

FIGURE 10.4 – Un tableau présentant les résultats d'une phase initiale de compétition sportive, d'après Shimojima (2014, p. 21) : un disque blanc à la ligne KOR et à la colonne BRA indique que l'équipe coréenne a battu l'équipe brésilienne ; le nombre de disques blancs à la ligne BRA indique le nombre de matchs gagnés par l'équipe brésilienne ; le fait qu'il y ait davantage de disques blancs à la ligne BRA qu'à la ligne KOR indique que l'équipe brésilienne est arrivée avant l'équipe coréenne dans cette phase initiale

conventions élémentaires qui définissent la représentation (par exemple que les régions grisées correspondent à des ensembles vides), mais semblent malgré tout en découler.

Notons d'ailleurs que ce phénomène se présente aussi dans des cas moins évidemment diagrammatiques. L'exemple préféré de Shimojima[28] est les tableaux que l'on utilise parfois dans les compétitions sportives pour présenter les résultats des phases initiales ou « phases de poule » (voir par exemple fig. 10.4). On peut y lire le résultat de chaque match individuellement, mais aussi l'utiliser de manière plus globale : on peut par exemple compter directement le nombre de victoires ou de défaites d'une équipe, ou comparer les équipes entre elles. Pour prendre un exemple mathématique, voici une table décrivant une loi de composition interne sur un ensemble à quatre éléments notés 1, a, b, et c (il s'agit en fait du *Vierergruppe* de Klein).

	1	a	b	c
1	1	a	b	c
a	a	1	c	b
b	b	c	1	a
c	c	b	a	1

Si je l'ignorais, je pourrais remarquer que l'opération représentée est inversible (parce que

28. Cf. p. ex. Shimojima 2015, p. 112-114.

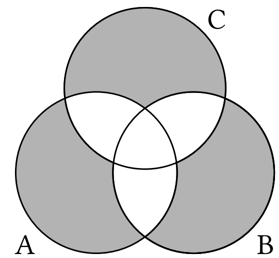


FIGURE 10.5 – Un autre diagramme de Venn symétrique par permutation de A, B, C

chaque élément est représenté une et une seule fois dans chaque ligne et chaque colonne). Je peux aussi remarquer, par symétrie, qu'elle est commutative. Dans le même esprit, l'histoire du calcul matriciel semble présenter des exemples plus substantiels [29] d'exploitation de propriétés visuelles de matrices, qui au fond sont des tableaux.

En un sens, ces phénomènes ne sont pas mystérieux. Observer sur le diagramme de Venn 10.3 que les ensembles A, B et C sont deux à deux disjoints se réduit immédiatement à trois usages successifs de la convention de lecture de base, conduisant à $A \cap B = \varnothing$, $B \cap C = \varnothing$ et $A \cap C = \varnothing$. Mais si l'on veut expliquer la possibilité de lire, de manière directe et systématique, des aspects du diagramme sans revenir à chaque fois aux conventions de base liant régions grisées et ensembles vides, il faut un peu plus. En effet, une même propriété, par exemple la symétrie par permutation observée sur la fig. 10.3, peut être réalisée par de nombreux diagrammes de Venn différents (la fig. 10.5 en présente un autre).

C'est cela que Shimojima veut expliquer : il cherche à rendre compte de la manière dont des conventions de lecture élémentaires peuvent en légitimer d'autres, plus sophistiquées. Il parle, à partir de 1999 [30], de « signification dérivée » (« *derivative meaning* »). Le schéma général de son analyse est donné à la fig. 10.6. Son idée est très simple. On se donne des conventions sémantiques de départ, qui définissent une relation d'indication entre certains types sources et certains types cibles. On a alors une relation d'indication *dérivée* entre un type source σ et un type cible θ sous deux conditions. D'une part, il faut que les types σ et θ soient équivalents à des disjonctions de conjonctions de types qui apparaissent dans les conventions sémantiques de départ ; chacune des conjonctions peut être conçue comme une manière particulière, pour une situation source (resp. cible), d'être du type σ (resp. θ). D'autre part, il faut que les conventions sémantiques de départ établissent une bijection entre types sources et types cibles de chacune de ces conjonctions (voir fig. 10.6). (Rien

29. Voir Brechenmacher 2006. Je remercie David Rabouin pour cette remarque.
30. Shimojima 1999a présente des phénomènes apparentés, mais sans les distinguer clairement des cas de « *free rides* » ; la distinction n'est thématisée explicitement qu'à partir de Shimojima 1999b.

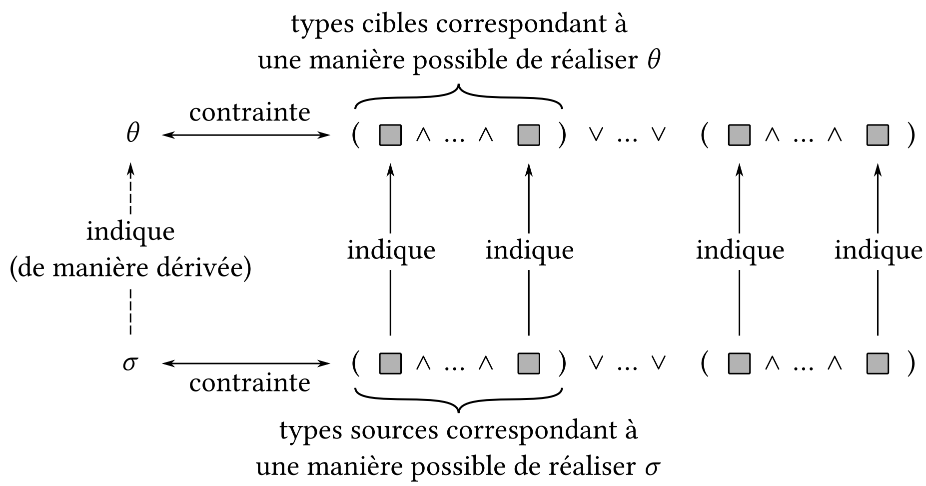


FIGURE 10.6 – Principe du « *derived meaning* » (cf. Shimojima 2014, p. 24, 2015, p. 114-124)

n'empêche, bien sûr, de démontrer que ces conditions sont vérifiées sans traiter tous les cas un par un ; Shimojima se contente ici d'expliciter une condition théorique qui doit être vérifiée pour que le type source θ indique le type cible σ, étant données nos conventions sémantiques.)

Comme le remarque Shimojima, la notion de signification dérivée est relative [31] : elle ne dépend pas seulement du système représentationnel, mais aussi des conventions sémantiques qui ont été choisies au départ ; or pour un même système, plusieurs choix différents de conventions sont possibles. Reprenons par exemple les « diagrammes positionnels » introduits dans la section précédente pour représenter l'ordre d'arrivée de coureurs, en nous limitant aux diagrammes où figurent tous les participants à la course. Si nous partons de la convention sémantique d'après laquelle A est écrit à gauche de B si A est arrivé avant B, alors nous obtenons, comme « signification dérivée », la convention supplémentaire que si C est écrit en troisième position, alors il était troisième à l'arrivée. Mais on pourrait très bien adopter comme convention sémantique de départ que si le nom A est écrit en position n, alors A était n^{e} à l'arrivée de la course ; c'est dans ce cas notre première convention qui devient « dérivée ».

10.4 Sémantique ou syntaxe ?

Pour mieux comprendre le but de Shimojima, il est utile de reformuler son travail en éliminant complètement la notion d'information. Reprenons une nouvelle fois l'exemple des

31. Cf. Shimojima 2015, p. 123.

diagrammes de Shin. Comme nous l'avons vu à la section 7.3, on peut les insérer dans des systèmes logiques hétérogènes. Ce que j'ai appelé, à la suite de Shimojima, les « conventions » des diagrammes de Venn correspondent alors à des règles d'inférence : dans le système de Hammer, par exemple, on a une règle ∀−Observe permettant d'inférer qu'un ensemble est vide sur la base d'une région grisée, et une règle ∀−Apply permettant de griser une région si l'on sait que l'ensemble correspondant est vide [32].

Dans ce cadre, les passagers clandestins de Shimojima sont des énoncés que je peux inférer en n'utilisant que des règles de construction et d'observation (comme ∀−Apply et ∀−Observe). Étudier les avantages des diagrammes, c'est alors faire deux choses : d'une part, distinguer certaines règles d'inférence comme particulièrement simples, peut-être parce qu'elles semblent relever d'une simple observation ; d'autre part, étudier systématiquement quelles inférences l'emploi d'un diagramme permet de faire en n'utilisant que ces règles simples. C'est ainsi, en substance, que Shimojima reformule l'idée de passagers clandestins dans certains travaux récents [33].

Dans le même esprit, on peut modéliser les cas de signification dérivée : il suffit d'introduire, pour chacune, une nouvelle règle (syntaxique) de lecture des diagrammes. Si on souhaite en montrer la validité, on peut le faire simplement en termes de modèles plutôt qu'en suivant le schéma 10.6 ci-dessus.

Pourquoi, alors, utiliser un cadre hétérodoxe comme celui de la section précédente ? Une première réponse nous ramène aux préjugés de Barwise, et ultérieurement de Shimojima, sur ce qu'il faut entendre par *information*. J'ai écrit plus haut [34] qu'une analyse des diagrammes de Venn en termes de modèles ne pouvait être satisfaisante, parce que, par exemple, le diagramme 10.1 et les énoncés $A \setminus B = \varnothing$ et $B \cap C = \varnothing$ avaient les mêmes modèles, et qu'on ne pouvait donc pas capturer ainsi la différence entre eux. L'idée de Barwise est qu'il faut pour cette raison une autre sémantique, plus fine, qui capture l'information « explicite » supplémentaire que porte le diagramme. Mais c'est omettre une autre possibilité : on peut comprendre la différence en termes *syntaxiques*, comme je viens de le suggérer ; on rejoint ainsi une conception de l'« information comme code », selon l'expression de van Benthem et Martinez [35], à la manière des sémantiques preuve-théoriques.

Il y a cependant une objection, déjà présentée au chapitre précédent [36] : notre usage informel des diagrammes ne semble pas guidé par des règles syntaxiques rigides, mais plutôt

32. Voir *supra*, section 7.3, en part. notes 45 et 46.
33. Voir Stapleton, Jamnik et Shimojima 2017 ; Stapleton, Shimojima et Jamnik 2018.
34. Section 10.2.
35. Voir l'introduction de cette partie, en part. p. 176–179.
36. Cf. section 9.1.b).

par l'information que les diagrammes portent d'après leurs conventions de représentation. Si l'on s'en tient à des règles simples comme ∀−Apply et ∀−Observe, l'argument n'est pas très convaincant. Certes, au lieu d'avoir une unique convention de représentation, on a *deux* règles, mais ces règles sont inverses l'une de l'autre ; un défenseur des sémantiques preuve-théoriques pourrait sans mal justifier le lien intrinsèque de ces deux règles en invoquant leur « harmonie [37] », et considérer que ces règles sont constitutives de l'information portée par le diagramme.

En fait, ce sont les cas de signification dérivée qui donnent toute sa force à l'objection précédente. Comme le montre le travail de Shimojima, on lit souvent davantage sur les diagrammes que ce que permettent nos conventions initiales de représentation, mais ces lectures supplémentaires peuvent être *justifiées* par nos conventions initiales. C'est là que la perspective générale de Barwise et Shimojima est cruciale : pour arriver à ces lectures supplémentaires, on raisonne non seulement sur le diagramme, mais *sur ses conventions de représentation*. Or si l'on examine l'usage informel des diagrammes, on s'aperçoit qu'un tel raisonnement sur leurs conventions de représentation est souvent nécessaire. C'était, d'ailleurs, l'une des leçons de notre étude de la visualisation de données (section 5.2). Un cadre plus classique ne donne pas vraiment d'espace à cette possibilité ; celui de Shimojima le permet, parce qu'il hérite du choix de la sémantique des situations de traiter la signification comme quelque chose sur quoi les locuteurs peuvent raisonner [38].

10.5 Le cas de l'analogie de Leibniz

Il est également éclairant de voir quels cas les concepts de Shimojima ne permettent pas de traiter. Au chapitre 5, nous avons rencontré divers exemples où traduire un tableau de nombres en graphique, changer de notation, etc. peut rendre visible des choses qui ne l'étaient guère [39]. Beaucoup de ces exemples, en particulier en visualisation de données, peuvent être traités au titre de la « signification dérivée » de Shimojima. Son analyse rend précise l'idée que l'on peut *justifier*, par exemple, l'usage d'un graphique pour lire davantage que ce que permettent les règles élémentaires d'après lesquelles il a été construit. (Dans le contexte d'une visualisation de données à des fins exploratoires, cette justification ne viendra peut-être qu'après coup, mais elle est en tout cas possible.)

Toutefois, on ne peut pas traiter ainsi tous nos exemples. Un cas particulièrement in-

37. Pour une introduction à l'usage de ce terme en sémantique preuve-théorique, voir par exemple Schroeder-Heister 2018, sect. 2.2.1.
38. Voir *supra*, section 9.1.a).
39. Voir en particulier sections 5.2 et 5.5.

téressant qui y échappe est celui de l'analogie de Leibniz étudiée dans la première partie. La structure que Leibniz remarque dans ses formules [40] ne se laisse pas décrire en termes d'une conjonction de propriétés de la formule pouvant être interprétées individuellement, comme le veut Shimojima. Cette remarque est vraie, plus généralement, des structures que l'on peut observer sur les formules, parce que celles-ci ont une syntaxe complexe qui rend impossible d'en interpréter isolément certains morceaux [41] ; on ne peut donc pas en rendre compte en revenant à des conventions sémantiques simples.

10.6 Conclusion

La théorie de Shimojima est silencieuse sur un point essentiel. Comment savons-nous que les conventions sémantiques que nous adoptons sont cohérentes et qu'elles produiront des résultats fiables ?

Cependant, dès lors que l'on admet certaines conventions de représentation, Shimojima montre bien que les diagrammes peuvent offrir une authentique normativité : étant données nos conventions, si l'on représente certains faits, on en représente d'autres par là-même (ce sont les « passagers clandestins »). Par ailleurs, son travail montre qu'il est tout à fait possible de lire davantage, sur un diagramme, que ce que les conventions sémantiques de départ permettent, et que cela n'a rien d'intrinsèquement mystérieux : on peut en rendre compte sans faire appel à une similarité de structure non élucidée entre représentation et représenté.

Au-delà des détails techniques, qui mériteraient peut-être d'être reformulés, la force du cadre de Shimojima (héritée des réflexions de Barwise et Etchemendy présentées au chap. 9) est surtout qu'il permet de rendre compte de la manière dont on peut utiliser un diagramme en raisonnant explicitement sur ses conventions de représentation.

---

40. Voir en particulier sections 3.3 et 5.5.

41. C'est en fait là que Stenning (2000) localise la distinction essentielle entre graphique et linguistique ; mais comme le montre Shimojima (2003), il y a là aussi des gradations.

Conclusion : une sémantique en quel sens ?

À travers le travail de Barwise, Etchemendy et leurs élèves, nous avons rencontré non pas une, mais deux conceptions substantielles de la sémantique applicables dans un cadre général qui inclut énoncés, formules, diagrammes et figures. En premier lieu, nous avons vu comment généraliser aux diagrammes une sémantique en termes de modèles. En second lieu, nous avons vu comment définir une sémantique plus fine, que l'on peut appeler, à la suite de nos auteurs, « informationnelle ».

Ces deux sémantiques nous apportent des bénéfices différents. La première permet de comparer des représentations différentes eu égard à leur contenu, et peut-être, dans certains cas, d'expliquer les avantages systématiques dont jouissent certaines formes de représentations par rapport à d'autres. Elle est aussi adaptée à parler de vérité et de conséquence logique.

La seconde permet de modéliser de plus près le raisonnement des agents qui utilisent des diagrammes : elle montre en effet comment rendre compte de la manière dont les agents qui les utilisent *comprennent* ce que leurs représentations représentent, et comment ils peuvent les utiliser pour acquérir des « informations explicites » supplémentaires, c'est-à-dire pour dériver de nouvelles propositions. Elle permet aussi de rendre compte de l'impression de normativité que donnent certains diagrammes. Celle-ci est due à ce que Shimojima appelle les « passagers clandestins » : lorsqu'on utilise des diagrammes, il arrive que représenter certaines prémisses selon certaines conventions en fasse immédiatement apparaître des conséquences *selon les mêmes conventions.* Cette normativité, toutefois, dépend du fait que les conventions adoptées sont cohérentes, ce que les diagrammes eux-mêmes sont bien incapables de justifier.

Quatrième partie

Faut-il parler de représentation ?

Sommaire de la quatrième partie

Introduction : plusieurs objections

Les parties précédentes sont fondées sur un double postulat. Tout d'abord, les nombreux éléments non discursifs rencontrés en mathématiques (formules, figures, etc.) doivent être vus comme des *représentations* qui, en un sens à clarifier, portent de l'information. D'autre part, c'est à partir de leur contenu informationnel qu'il faut comprendre l'usage qu'on en fait. Cette dernière partie, un peu hétéroclite, examine différents arguments contre cette idée.

Tout d'abord, on peut soupçonner qu'en pratique, les figures utilisées en mathématiques sont trop floues pour se prêter à ce genre de traitement : peut-on vraiment leur identifier un contenu clair ? Le chapitre 11 examine cette objection et cherche à y répondre. Je reviens, en particulier, sur les deux exemples que sont les diagrammes d'Euler, dont nous avons vu dans la partie précédente qu'ils semblent peu rigoureux [42], et les figures de la géométrie élémentaire. Je montre que l'on peut rendre compte de leur usage, mais que cela exige de leur donner une sémantique *contextuelle*, c'est-à-dire de reconnaître qu'ils ne sont pas indépendants du texte qui les accompagne, contrairement aux diagrammes de Shin de la partie précédente.

Une seconde objection est due à Manders : il est vain, écrit-il, de chercher à rendre compte des figures géométriques qu'utilise Euclide en termes sémantiques. Le chapitre 12 vise à comprendre cette objection, et fait le lien avec certains des phénomènes observés dans l'étude de cas de la première partie.

Bien sûr, il existe en philosophie du langage des positions philosophiques plus générales, que l'on peut appeler *inférentialistes*, d'après lesquelles il importe de partir des inférences et non d'une supposée signification donnée d'avance. C'est, par exemple, la position de Robert Brandom (1994, 2000). Une véritable discussion de ces positions dépasse le cadre de ce travail. Ici encore, ma stratégie est de partir du cas particulier des représentations utilisées en mathématiques, quitte à rejoindre, en fin de compte, des discussions plus gé-

42. Il s'agit ici des diagrammes d'Euler, par opposition aux diagrammes de Venn avec lesquels ils sont souvent confondus ; voir ci-dessus, section 7.1.

nérales de philosophie du langage.

Chapitre 11

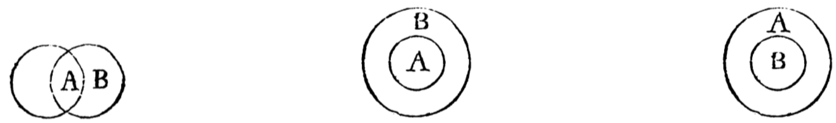
Usages informels des diagrammes et contextualité

Les systèmes formels de Barwise, Etchemendy et Shin fournissent un modèle clair de raisonnement rigoureux fondé sur des diagrammes. Mais en pratique, les usages de diagrammes en mathématiques ne sont-ils pas trop vagues pour entrer dans ce cadre ? C'est un soupçon légitime, auquel je voudrais ici répondre.

Les diagrammes d'Euler sont à première vue un bon exemple : Shin les étudie, mais conclut qu'ils souffrent d'un « problème fatal d'ambiguïté » et les laisse finalement de côté au profit des diagrammes de Venn. N'est-ce pas un cas typique d'emploi « peu rigoureux » de diagrammes ? Effectivement, comme nous l'avons vu [1], à y regarder de près, ils risquent d'induire en erreur ; on peut se demander s'il est vraiment possible de les utiliser de manière fiable sans les contrôler par une bonne connaissance – non diagrammatique ! – des syllogismes. Euler, pourtant, ne commet pas d'erreur. Je voudrais ici montrer que l'on peut rendre raison de sa démarche [2]. Des travaux récents de John Mumma arrivent à une conclusion similaire pour l'emploi des figures dans la géométrie élémentaire d'Euclide.

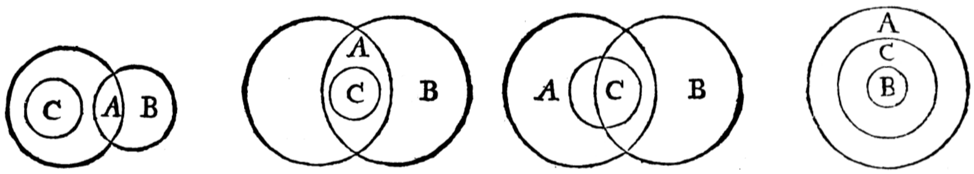
Ces solutions soulèvent toutefois une difficulté : il s'avère que l'on ne peut interpréter correctement les diagrammes d'Euler, ni les figures d'Euclide, indépendamment des prémisses qu'ils sont censés représenter. S'il n'est pas nécessaire d'abandonner l'idée de donner aux diagrammes ou figures une sémantique précise, il faut en revanche abandonner le présupposé que celle-ci peut forcément être définie, à la manière de Shin, sur la seule base de leur apparence. Souvent, il faut aussi tenir compte de leur contexte.

1. Cf. section 7.1.a).
2. Une version préliminaire de cette première section a été publiée en anglais : Waszek 2018.



(a) (b) (c)

Figure 11.1 – Trois diagrammes compatibles avec « Quelque A est B »



(a) (b) (c) (d)

Figure 11.2 – Pour représenter « Quelque A est B » et « Tout C est A », Euler trace les diagrammes (a)-(c) (ELP, vol. 2, p. 112) mais omet le (d).

11.1 Retour sur les diagrammes d'Euler

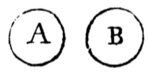
a) Le problème

Pour rappel, le problème fondamental du système d'Euler est que les diagrammes sont plus spécifiques que les propositions aristotéliciennes qu'ils sont censés représenter [3]. La proposition « Quelque A est B », par exemple, est compatible avec les trois diagrammes de la fig. 11.1 ; pourtant, Euler utilise toujours le 11.1(a). En toute logique, cela devrait l'induire en erreur. Par exemple, pour les prémisses « Quelque A est B » et « Tout C est A », il trace les diagrammes (a)-(c) de la fig. 11.2 et omet le (d), qui exigerait de représenter « Quelque A est B » par 11.1(c) plutôt que par 11.1(a). Ses trois diagrammes semblent donc représenter la conclusion fausse « Quelque B n'est pas C ». Pourtant, Euler ne commet pas l'erreur et déduit le bon résultat, qui est que ses deux prémisses n'admettent pas de conclusion. Comment fait-il ? Faut-il tout simplement en conclure qu'il connaît déjà la bonne réponse, et que les diagrammes ne lui servent que d'illustration ? C'est au fond le jugement auquel sont acculés les commentateurs précédents, dont Hammer et Shin (1998).

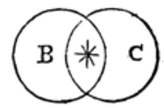
b) La démarche d'Euler

Pour mieux comprendre comment Euler procède, tournons-nous vers sa lettre cv (ELP, vol. 2, p. 119–126). Il y réexplique sa méthode sur un exemple de la forme « Nul A n'est

3. Pour plus de détails, voir *supra*, section 7.1.a).

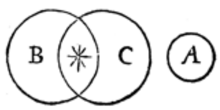


(a) Nul A n'est B.

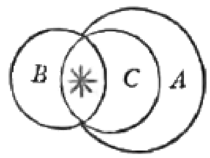


(b) Quelque B est C.

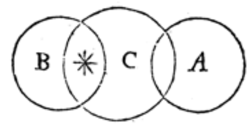
FIGURE 11.3 – Diagrammes tracés par Euler dans la lettre CV (ELP, vol. 2, p. 120-121)



(a)



(b)



(c)

FIGURE 11.4 – Diagrammes d'Euler avec étoiles pour « Nul A n'est B ; quelque B est C ; donc quelque C n'est pas A » (ELP, vol. 2, p. 121–122). Dans le diagramme (b), il faut imaginer que la frontière de l'espace A longe celle de l'espace B sans la couper ; ce diagramme n'est pas toujours bien reproduit dans les éditions ultérieures.

B ; quelque B est C ; donc quelque C n'est pas A ». Pour « Nul A n'est B », il utilise le diagramme habituel, mais pour « Quelque B est C », il ajoute une étoile au diagramme attendu, et explique :

> [...] dans la premiere proposition [...] on soutient que [...] l'espace A se trouve tout entier hors de l'espace B, en cette sorte [fig. 11.3(a)]. Mais dans la seconde proposition [...] on dit qu'une partie de l'espace B se trouve dans l'espace C, ensorte [fig. 11.3(b)] où la partie de l'espace B comprise dans C est marquée d'une étoile * qui sera donc aussi une partie de l'espace C. Donc puisque une partie de l'espace C est en B & que tout l'espace B se trouve hors de l'espace A, il est évident que la même partie de l'espace C doit aussi être hors de l'espace A [...].
>
> Il faut bien remarquer que cette conclusion ne regarde que la partie * de la notion C qui est plongée dans la notion B. Pour le reste [de la notion C] il est incertain, s'il est aussi exclû de la notion A, comme dans cette figure [fig. 11.4(a)], ou s'il y est renfermé tout entier comme [fig. 11.4(b)] ou seulement en partie comme dans cette figure : [fig. 11.4(c)]. Or puisque cela est incertain, le reste de l'espace C n'entre dans aucune consideration : la conclusion se borne uniquement à ce qui est certain, c'est à dire, que la même partie de l'espace C, qui

> est contenuë dans l'espace B se trouve certainement hors de l'espace A [...] [4].

L'objectif d'Euler, ici, ne semble pas être d'introduire une version modifiée de son système ; si c'était le cas, il tenterait d'expliquer plus systématiquement comment utiliser sa nouvelle étoile pour toutes les formes de prémisses, ce qu'il ne fait pas du tout. À mon sens, il cherche plutôt à expliciter la manière dont il comprend la démarche qu'il utilisait jusque là : l'étoile lui sert à attirer l'attention de sa lectrice sur la partie pertinente du diagramme, ou plus précisément à expliquer comment il *regarde* ses diagrammes.

C'est ici que les explications d'Euler nous fournissent un indice crucial : il utilise ses diagrammes *conjointement avec les prémisses qu'ils sont censés illustrer.* Lorsqu'il regarde ses cercles, il garde à l'esprit quelles sont les régions dont les prémisses affirment quelque chose ; dans le passage cité ci-dessus, Euler explique clairement que les autres régions du diagrammes ne peuvent pas servir à tirer de conclusion. C'est cette manière de regarder le diagramme qu'il cherche à rendre visible à la princesse en ajoutant une croix.

Cela explique pourquoi Euler ne fait pas les erreurs auxquelles ses diagrammes invitent. Reprenons l'exemple ci-dessus (fig. 11.2). Dans chacun des diagrammes (a)-(c), les régions A et B illustrent la prémisse « Quelque A est B », donc n'affirment quelque chose que sur l'intersection de A et B : sur la région de B qui est en-dehors de A, en revanche, on ne sait rien. Or c'est de cette région qu'on aurait besoin pour affirmer « Quelque B n'est pas C ». Si l'on garde cela à l'esprit, on peut éviter de tirer cette conclusion fausse sans tracer de diagramme supplémentaire.

Au passage, cela éclaire le rôle de la position des lettres. Euler semble l'utiliser pour distinguer les diagrammes de « Quelque A est B » et « Quelque A n'est pas B », quoiqu'il ne l'écrive nulle part explicitement, qu'il ne s'y tienne pas toujours, et que cela ne fonctionne pas comme convention systématique de représentation [5]. Ce qui précède permet de lever le mystère. Pour utiliser ses diagrammes, Euler a besoin de garder à l'esprit de quelle région sa prémisse existentielle affirme quelque chose. La position des lettres peut servir d'aide-mémoire commode, même si ce n'est ni indispensable, ni toujours suffisant.

Bien sûr, tout cela n'est possible que parce qu'Euler travaille dans le cadre (de notre point de vue restreint et rigide) de la syllogistique, c'est-à-dire qu'il n'étudie que des raisonnements à deux prémisses aristotéliciennes ayant un seul terme en commun. Dès que l'on veut aller un peu plus loin, on rencontre des difficultés ; on voit mal, par exemple, comment représenter simultanément « Tout A est B » et « Tout B est A », ou encore « Tout A

4. ELP, vol. 2, pp. 121–122.
5. C'est ce que montrent Hammer et Shin, comme nous l'avons vu plus haut (voir p. 185).

est B » et « Aucun A n'est B » [6]. J'ai choisi ici de juger le système d'Euler dans ses propres termes, c'est-à-dire vis-à-vis des problèmes qu'il est censé résoudre. Or contrairement aux apparences, dans ce cadre restreint, il est adéquat et on peut s'y fier.

Que la méthode d'Euler soit fiable ne signifie pas qu'elle soit simple. Ce que fait Euler est d'autant plus subtil qu'il n'explicite pas pleinement les règles selon lesquelles il inspecte ses diagrammes. Peut-être ce genre de pratique, dont les règles sont à la fois implicites et contextuelles, est-il intrinsèquement difficile à transmettre. Il est d'ailleurs révélateur, comme l'ont noté Hammer et Shin [7], que lorsqu'Euler tente de réexpliquer sa démarche dans la lettre CV, il se retrouve à enrichir ses diagrammes (ne serait-ce que temporairement) par un nouveau symbole : une étoile, qui rappelle les croix que Peirce utilisera plus tard. En fait, ce que l'on gagne lorsque l'on passe aux diagrammes de Venn ou de Peirce, c'est précisément que ceux-ci sont pleinement autonomes et indépendants du contexte de lecture.

c) Formaliser la méthode

À ce stade, puisque les diagrammes ne peuvent être utilisés seuls, il est tentant de conclure qu'ils sont seulement illustratifs, c'est-à-dire sans rôle réel dans le raisonnement, ou seulement heuristiques, parce que les conclusions qu'ils suggèrent devraient être continuellement vérifiées par d'autres voies.

Il n'en est rien. En premier lieu, Euler s'appuie bien sur ses diagrammes, pour y observer l'inclusion de certaines régions dans d'autres. En second lieu, il n'a pas besoin de vérifier systématiquement ses conclusions en utilisant ce qu'il sait par ailleurs des syllogismes. Pour les conclusions universelles, il peut se reposer entièrement sur les diagrammes. Pour les conclusions existentielles, il doit certes faire intervenir les prémisses (qu'il a besoin de garder à l'esprit quand il examine les diagrammes, pour savoir quelles régions peuvent permettre une conclusion), mais seulement de manière bien précise.

Pour clarifier, explicitons les règles qu'Euler semble suivre. Partons de deux prémisses syllogistiques et d'un diagramme censé les illustrer. Alors :

1. si le diagramme comporte deux régions étiquetées A et B et que la région A est incluse dans la région B, on peut conclure « Tout A est B » ;
2. si le diagramme comporte deux régions étiquetées A et B qui sont disjointes, on peut conclure « Nul A n'est B » ;

6. Pour des critiques de ce genre, voir par exemple Shin 1994, p. 14-16 ou Shin, Lemon et Mumma 2013, section 2.1.
7. Hammer et Shin 1998, p. 11.

3. si le diagramme comporte trois régions étiquetées A, B et C, que l'intersection des régions A et C est contenue dans la région B et que le diagramme illustre la prémisse « Quelque A est C », on peut conclure « Quelque A est B » ;
4. si le diagramme comporte trois régions étiquetées A, B et C, que l'intersection des régions A et C est hors de la région B et que le diagramme illustre la prémisse « Quelque A est C », on peut conclure « Quelque A n'est pas B » ;
5. si le diagramme comporte trois régions étiquetées A, B et C, que la partie de la région A hors de la région C est comprise dans la région B et que le diagramme illustre la prémisse « Quelque A n'est pas C », on peut conclure « Quelque A est B » ;
6. si le diagramme comporte trois régions étiquetées A, B et C, que la partie de la région A hors de la région C est hors de la région B et que le diagramme illustre la prémisse « Quelque A n'est pas C », on peut conclure « Quelque A n'est pas B ».

Les deux premières règles (celles qui mènent à des conclusions universelles) s'appuient uniquement sur le diagramme. Les règles 3 à 6 (celles qui mènent à des conclusions existentielles) s'appuient conjointement sur le diagramme et sur l'une des prémisses. Dans les deux cas, le diagramme est utile : il sert à observer l'inclusion de certaines régions dans d'autres.

Ces six règles pourraient être formalisées avec les outils de Barwise, Etchemendy et Shin. Dans leur terminologie, ce seraient tout simplement des règles d'inférence *hétérogènes*[8]. À vrai dire, cela requiert de pousser leur idée un peu plus loin qu'ils ne l'ont fait : en pratique, dans leurs travaux, ils se limitent à des règles passant de diagrammes à des énoncés ou d'énoncés à des diagrammes, alors que les règles 3 à 6 ci-dessus ont pour prémisses *à la fois* un diagramme et un énoncé. Mais leur cadre s'en accomoderait sans difficultés.

On pourrait, sur cette base, formaliser complètement la pratique d'Euler. Pour déduire une conclusion de deux prémisses aristotéliciennes ayant un terme en commun, il trace d'abord un diagramme par prémisse selon les conventions de la fig. 7.1 p. 183. Ensuite, il les superpose, traçant autant de diagrammes qu'il y a de relations possibles entre les termes « extrêmes », c'est-à-dire les termes qui n'apparaissent pas dans les deux prémisses. (Il faudrait donc décrire un algorithme d'énumération des cas. Une stratégie plus simple, correspondant à la pratique de Leibniz et Lange, qui utilisent des diagrammes similaires avant Euler[9], serait de ne tracer qu'un seul diagramme, en choisissant – lorsqu'il y en a plusieurs possibles – celui où les courbes des termes extrêmes s'intersectent, pour éviter

8. Voir *supra*, section 7.3.
9. Voir note 7 p. 182.

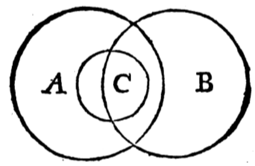


FIGURE 11.5 – Un diagramme d'Euler, d'après ELP, vol. 2, p. 106. Sans information supplémentaire, ce diagramme est ambigu.

les conclusions universelles incorrectes.) Pour finir, Euler ne tire une conclusion que si tous les diagrammes qu'il a tracés le permettent d'après les règles hétérogènes ci-dessus.

d) Une sémantique contextuelle ?

On peut donc rendre compte de ce que fait Euler au moyen d'un système formel hétérogène, dont j'ai ébauché les règles. Mais quelle sémantique donner aux diagrammes ?

À première vue, selon ma reconstruction, on ne peut pas lire un diagramme d'Euler sans savoir quelles prémisses il est censé représenter. Si vous me montrez la figure 11.5 sans rien me dire de plus, je serais bien en peine de dire s'il permet de conclure « Quelque A est B », « Quelque B n'est pas A », « Quelque C n'est pas B » ou que sais-je encore. Sans éléments supplémentaires, on ne pourrait donc pas lui donner de sémantique. En ce sens, il est tentant de dire que les diagrammes d'Euler n'acquièrent de sémantique qu'en contexte.

Pourtant, on peut aussi donner une sémantique non contextuelle aux diagrammes d'Euler pour laquelle les règles du système formel hétérogène précédent sont correctes [10]. L'idée serait que les diagrammes d'Euler ne représentent jamais que des propositions universelles ; la fig. 11.5 ci-dessus, par exemple, ne représenterait que « Tout C est A ». Plus précisément, un diagramme d'Euler serait équivalent à la conjonction des propositions « Aucun X n'est Y » pour toutes courbes X et Y qui ne s'intersectent pas et des propositions « Tout X est Y » pour toutes courbes X et Y telles que X est entièrement incluse dans Y. On a donc une division du travail entre diagrammes et texte : le diagramme prend en charge les prémisses universelles, et le texte les prémisses existentielles. C'est bien ce que l'on observe dans nos règles hétérogènes. Pour tirer une conclusion universelle de deux prémisses universelles (règles 1 et 2), le diagramme seul suffit. Pour tirer une conclusion existentielle (règles 3 à 6), il faut en revanche une prémisse universelle, portée par le diagramme, et une prémisse existentielle, qui exige le recours au texte.

10. « Correct » est ici à prendre au sens technique : vis-à-vis d'une sémantique, une règle d'inférence est correcte (« *sound* ») si tout modèle des prémisses est modèle de la conclusion.

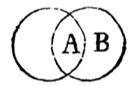


FIGURE 11.6 – Le diagramme d'Euler pour « Quelque A est B »

Cette sémantique (et les règles hétérogènes qu'elle fonde) permet de rendre raison de la pratique d'Euler, mais il faut bien voir qu'elle est en décalage avec la manière dont lui-même comprend et décrit son système. Elle conduit par exemple à conclure qu'essentiellement, la fig. 11.6 *ne représente rien*. Nos efforts de formalisation nous acculent ici à un dilemme. Ou bien on considère que la position des lettres est une authentique convention de représentation ; on peut alors accepter que la fig. 11.6 représente « Quelque A est B », mais on doit conclure que le système d'Euler est incorrect voire incohérent. Ou bien on accepte la sémantique précédente et les règles hétérogènes associées, mais alors les diagrammes ne peuvent plus représenter que les propositions universelles, au contraire de ce qu'écrit Euler. Informellement, pourtant, la pratique d'Euler est à la fois fiable et claire, comme j'ai tenté de le montrer : ses figures doivent être lues conjointement avec le texte et la position des lettres lui sert d'aide-mémoire.

Cela nous conduit au cœur de ce qui distingue une pratique informelle, comme celle d'Euler, de formalisations à la Barwise-Etchemendy : dans la pratique informelle, en rester à la forme des représentations utilisées, sans tenir compte du contexte et des intentions de l'utilisateur, ne suffit pas. Pour atteindre une formalisation qui respecte la pratique informelle, il faudrait admettre une sémantique authentiquement contextuelle.

11.2 Un parallèle chez Euclide

L'exemple des diagrammes d'Euler peut, du fait de sa grande simplicité, servir de modèle pour éclairer des cas plus complexes. Je voudrais ici montrer qu'il permet de mieux comprendre les multiples tentatives récentes de formaliser l'usage des figures dans les premiers livres des *Éléments* d'Euclide [11].

a) Manders et Miller : une décontextualisation des figures

Un poncif de la littérature récente sur les figures géométriques est qu'on ne peut pas tracer deux segments ou deux angles exactement égaux, deux droites exactement parallèles, etc. et qu'en conséquence, il faut distinguer entre deux classes de propriétés géométriques :

11. Pour une présentation rapide de cette littérature, voir *supra*, p. 220–221.

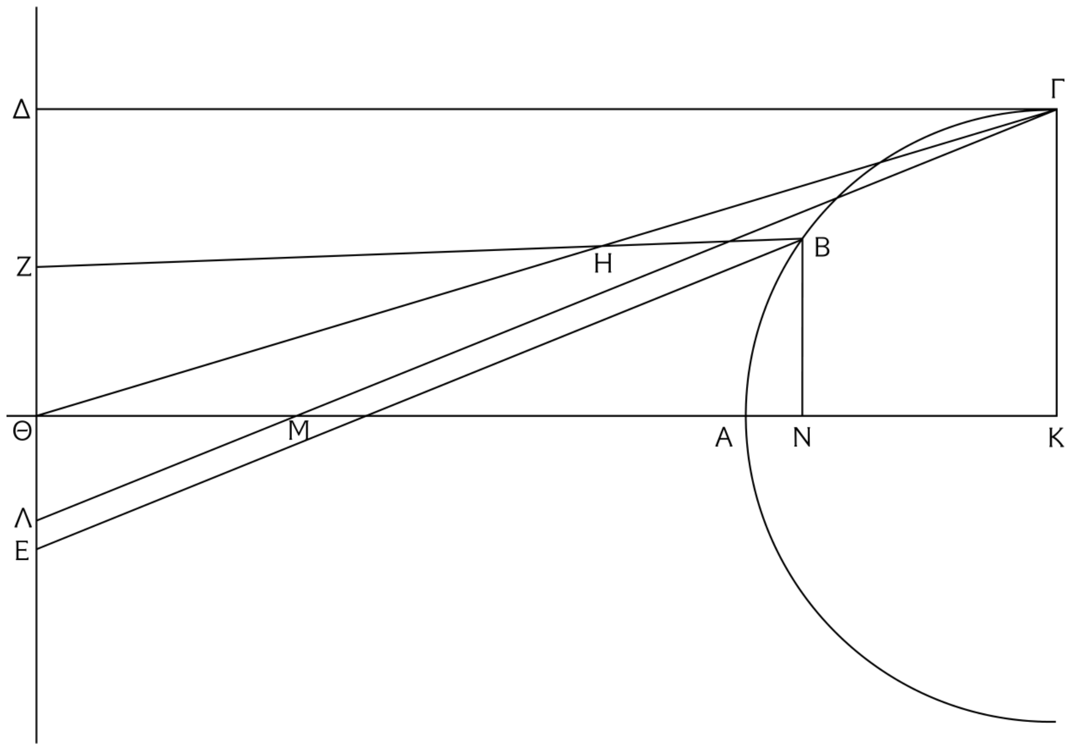


FIGURE 11.7 – Figure de la proposition I.45 des *Coniques* d'Apollonius (cas de l'hyperbole) d'après Netz (1999, p. 28)

celles que l'on peut et celles que l'on ne peut pas observer sur une figure raisonnablement bien construite. Miller (2007) écrit, dans un passage déjà cité, que ses figures ne sont censées montrer que la « topologie » de la disposition de droites et de cercles dans le plan [12]. Manders (2008b) distingue entre attributs « exacts » et « co-exacts » (ou, dans des textes antérieurs, « non-exacts [13] »); Panza distingue les attributs « diagrammatiques » et non diagrammatiques [14].

Un examen naïf montre pourtant que lorsqu'on utilise des figures, on se repose sans cesse, par exemple, sur le parallélisme de certaines droites. Un expert identifie immédiatement les angles alternes-internes sur une figure, comme nous l'avons déjà remarqué [15] en étudiant le problème géométrique discuté par Larkin et Simon (1987). Reviel Netz en donne un exemple clair tiré de la proposition I.45 d'Apollonius [16] (fig. 11.7). Au cours de

12. Voir *supra*, p. 159.
13. Cf. Manders 1996, p. 392.
14. Panza 2012, p. 75 *sq.*
15. Cf. ci-dessus, p. 126.
16. Voir Netz 1999, p. 27-29. Il s'agit plus précisément du cas de l'hyperbole de la proposition I.45, cette

la preuve, Apollonius s'appuie implicitement sur le fait que les triangles MKΓ et ΓΔΛ sont semblables, comme si ce fait était plus ou moins évident. D'où vient cette évidence ? Pour le genre de raisons évoqué plus haut, « les figures ne peuvent pas, à elles seules, montrer de manière satisfaisante que des triangles sont semblables [17] ». Il semble bien, cependant, que la figure nous rende disponible les différentes relations de parallélisme nécessaires ; comme le remarque Netz :

> [I]t is not the case that the diagram asserts information such as 'ΓK is parallel to ΔΘ'. Such assertions cannot be shown to be true in a diagram. But once the *text* secures that the lines are parallel, this piece of knowledge may be encoded into the reader's representation of the diagram. When necessary, such pieces of knowledge may be mobilised to yield, as an ensemble, further results. [18]

Ce que l'on utilise en géométrie, ce n'est pas la figure telle qu'elle est dessinée mais la figure « perçue d'une certaine manière [19] », soumise aux hypothèses que l'on a faite sur elle.

En d'autres termes, les figures géométriques – exactement comme les diagrammes d'Euler – doivent être lues conjointement avec le texte. Cela nous permet de mieux comprendre l'exercice auquel se livre, par exemple, Manders. Ce qu'il fait revient à isoler, de manière un peu artificielle, les contributions justificationnelles de la figure et du texte, exactement comme quand nous avons dit qu'un diagramme d'Euler ne représente que des propositions universelles et que les propositions existentielles sont seulement portées par le texte.

Cette démarche est très productive, mais a des conséquences contre-intuitives. De même qu'il est étrange de dire que, contrairement à ce qu'affirme Euler, son diagramme pour « Quelque A est B » ne représente au sens strict rien du tout, de même il est étrange d'affirmer qu'une figure ne représente jamais de droites ni de cercles, mais seulement des courbes ouvertes et fermées, parce que la propriété pour une ligne d'être droite ou pour

---

dernière couvrant toutes les côniques à centre (Heiberg 1891-1893, I, p. 136–140 = Rashed 2008-2010, 1.1, p. 416–420, 1.2, p. 156–161 ; le diagramme que donne Netz est semblable à celui de Heiberg). Ici n'est pas le lieu de discuter précisément ce que démontre cette proposition ni à quoi elle sert. Disons seulement qu'on peut la considérer comme un lemme pour les propositions qui suivent et qui montrent comment exprimer la propriété caractéristique d'une conique par référence à un diamètre quelconque. Pour mieux comprendre le texte, on peut se référer au commentaire de Roshdi Rashed 2008-2010, 1.1, p. 162–172 ou encore à Heath 1896, p. 31-41.

17. « Now diagrams cannot, in themselves, show satisfactorily the similarity of triangles. » (Netz 1999, p. 27).

18. « [I]l ne faut pas croire que la figure affirme une information comme 'ΓK est parallèle à ΔΘ'. De telles assertions ne peuvent pas être démontrées sur la base du diagramme. Mais une fois que le *texte* assure que les droites sont parallèles, cette connaissance peut être encodée dans la représentation que le lecteur se fait de la figure. De telles connaissances peuvent, au besoin, être mobilisées et combinées pour fournir de nouveaux résultats. » (Netz 1999, p. 29 ; l'auteur souligne).

19. Netz 1999, p. 33.

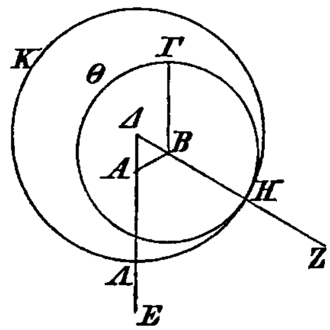


FIGURE 11.8 – Figure de la proposition I.2 des *Éléments* d'après Heiberg (EEHei, I, p. 15)

une courbe fermée d'être un cercle sont « exactes » au sens de Manders. C'est toutefois le prix à payer pour donner une interprétation de la figure qui soit entièrement indépendante de son contexte textuel.

b) Mumma : une recontextualisation à un autre niveau

L'étude du rôle des figures dans les *Éléments* d'Euclide réserve toutefois un rebondissement. Comme le montre bien John Mumma, la distinction établie par Manders entre attributs exacts et co-exacts ne suffit pas à rendre l'interprétation de la figure indépendante du contexte.

Le parallèle avec les diagrammes d'Euler est ici très étroit. La difficulté avec ces diagrammes, comme nous l'avons vu, est qu'ils sont plus spécifiques que les prémisses à traiter. On ne peut pas représenter « Quelque A est B » sans décider s'il est vrai aussi que « Quelque A n'est pas B », par exemple. Or Euler, au lieu de dessiner tous les cas possibles compatibles avec ses prémisses, n'en trace que certains, mais les contrôle de telle manière que les conclusions restent vraies en général.

Un problème similaire se pose chez Euclide : il ne fait pas toujours les distinctions de cas qui semblent nécessaires. Certes, si l'on admet la distinction de Manders entre attributs exacts et co-exacts (ou une autre apparentée), beaucoup de distinctions qui peuvent à première vue paraître requises s'avèrent inutiles. Par exemple, une proposition qui traite de triangles n'a pas besoin de traiter le cas des triangles isocèles, équilatéraux, rectangles ou scalènes *via* des figures séparées : les longueurs relatives des différents côtés ou la valeur exacte des angles ne sont pas le genre de choses que l'on peut observer sur une figure ; tous ces cas ayant les mêmes attributs co-exacts, on peut se contenter d'en traiter un seul. De plus, Euclide évite de formuler des propositions dont les énoncés recouvriraient d'emblée des cas aux attributs co-exacts différents.

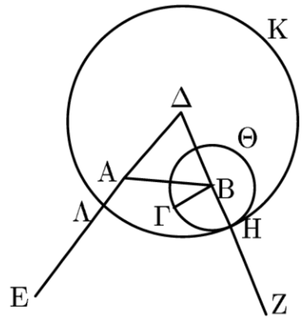


FIGURE 11.9 – Figure de la proposition I.2 des *Éléments* dans le codex b, d'après la reproduction de Saito (2006, p. 98)

Pourtant, il reste de nombreuses situations, chez Euclide, où des distinctions de cas que le texte ne fait pas seraient apparemment nécessaires. La raison en est que la plupart des propositions d'Euclide requièrent de faire une construction ; or deux figures ayant les mêmes attributs co-exacts au sens de Manders peuvent, après une même construction, présenter des attributs co-exacts différents. Mumma (2010) prend l'exemple de la proposition I.2, qui indique comment « placer, en un point donné, une droite égale à une droite donnée [20] ». En termes modernes, étant donné un segment ΒΓ et un point A, le but est de construire un segment AΛ de même longueur que ΒΓ. La construction que propose Euclide est assez complexe, et requiert de tracer successivement deux cercles. Le problème est que, pour différentes positions relatives initiales de ΒΓ et de A, cette construction produit des résultats notablement différents. La plupart des éditions récentes présentent une variante de la fig. 11.8, qui est celle de Heiberg, mais on pourrait aussi bien tracer la fig. 11.9, qui vient d'un manuscrit byzantin du XIe siècle [21].

Si l'on considère que les preuves euclidiennes s'appuient sur les attributs co-exacts de leurs figures, et qu'il y a des cas présentant des attributs co-exacts distincts qu'Euclide n'a pas envisagé, ne faut-il pas conclure que les démonstrations d'Euclide sont incorrectes ? On peut, il est vrai, simplement considérer qu'il y a là un (léger) manque de rigueur de la part d'Euclide. Proclus, dans son commentaire ancien du premier livre des *Éléments*, complète par exemple la proposition I.2 en traitant une pléthore de cas supplémentaires [22]. Heath, de son côté, affirme que « la méthode d'Euclide est de ne donner qu'un seul cas – tant qu'à choisir, le plus difficile – et de laisser le reste au lecteur [23] ». La rigueur complète

20. Cf. EEV, I, p. 197–199 = EEHea, I, p. 244–246 = EEHei, I, p. 12–15.
21. Sur les figures dans la traduction manuscrite des *Éléments*, voir *infra*, note 7 p. 294.
22. Cf. Friedlein 1873, p. 222-228 = Proclus 1970, p. 175-179 ; voir aussi EEHea, I, p. 245–246.
23. « Euclid's method is to give one case only, for choice the most difficult, leaving the reader to supply the

exigerait alors de les traiter tous ; dans cet esprit, le système formel de Miller (2007) requiert systématiquement une énumération exhaustive de tous les cas possibles.

La réponse de Mumma est différente [24]. D'après lui, Euclide peut en toute rigueur se contenter d'une seule figure parce qu'il prend soin de ne s'appuyer que sur des attributs co-exacts qui resteraient vrais dans tout autre cas possible. Mumma montre que l'on peut formaliser cette démarche, en adjoignant à la figure un objet supplémentaire – qu'il appelle précisément son « contexte » – qui en retrace la construction et permet de contrôler l'usage qu'il est légitime d'en faire. (On peut considérer que les règles d'inférences de son système sont en fait des règles hétérogènes, qui s'appuient à la fois sur la figure et sur ce contexte textuel, comme celles que j'ai proposées pour les diagrammes d'Euler.) Avigad, Dean et Mumma (2009) reprennent l'idée de Mumma, mais se dispensent entièrement de la figure et dérivent directement les attributs co-exacts qui seraient vrais pour n'importe quelle figure construite selon les mêmes règles à partir de données vérifiant les hypothèses.

L'important, ici, est que la solution de Mumma force à *contextualiser* les figures, c'est-à-dire à admettre à nouveau qu'on ne peut les lire indépendamment du texte (plus précisément, indépendamment de la manière dont elles ont été construites). Comme j'ai tenté de le faire pour Euler, Mumma montre que si l'on procède ainsi, on peut rendre compte adéquatement des preuves d'Euclide sans les trouver déficientes.

11.3 Conclusion

Nous avons examiné deux exemples apparemment paradigmatiques de l'usage peu rigoureux de diagrammes ou figures dans la pratique des mathématiques : les diagrammes d'Euler et les figures en géométrie élémentaire, en particulier chez Euclide. Dans les deux cas, il s'avère qu'on peut en rendre compte de manière satisfaisante, mais qu'il faut abandonner l'un des présupposés des travaux de Barwise, Etchemendy et Shin, qui est que l'on doit pouvoir donner aux diagrammes une sémantique indépendante de leur contexte.

En fait, ce phénomène n'est sans doute pas limité aux diagrammes. Le développement de la pragmatique en linguistique ces dernières décennies a bien montré que le langage courant présente un grand nombre de dépendances contextuelles systématiques. Il serait surprenant que les mathématiques informelles en soient exemptes. M. Wilson (1994) suggère une piste intéressante. Il propose de comprendre en termes de dépendances contex-

rest for himself. » (EEHea, I, p. 246).

24. Pour une introduction, voir Mumma 2010 ou encore Shin, Lemon et Mumma 2013, sect. 4.2.2 et 4.3 ; le système complet de Mumma se trouve dans sa thèse (2006). (Voir aussi Mumma 2014 pour des corrections rendues nécessaires par les critiques de Miller 2012.)

tuelles certaines pratiques mathématiques fréquemment jugées peu rigoureuses, comme l'emploi de « points infiniment proches » dans la géométrie algébrique du XIX^{e} siècle ou les infinitésimaux de l'analyse leibnizienne. Là encore, les praticiens semblent globalement capables de les utiliser sans erreur, mais prises au pied de la lettre et utilisées sans discernement, leurs méthodes conduisent facilement aux pires incohérences. L'idée de Wilson est de capturer l'usage qu'en font les praticiens par des « grammaires contraintes » qui, essentiellement, explicitent des contraintes contextuelles tacites. Est-il pertinent d'y voir une variante de ce que nous avons observé ici pour les diagrammes ? La piste, me semble-t-il, mériterait d'être creusée.

Chapitre 12

Manipulations sans représentation ?

Faut-il penser les figures de la géométrie euclidienne en termes sémantiques, c'est-à-dire, en un sens ou en un autre, comme des *représentations* ? Non, d'après Ken Manders, qui soutient que ce n'est pas une approche prometteuse, ni dans le cas de la géométrie d'Euclide ni en général [1]. Mon premier but ici est d'évaluer ses arguments. Cela me forcera, au passage, à tenter de clarifier certaines ambiguïtés dans l'usage des termes de représentation et de sémantique [2]. J'aborderai ensuite un autre genre d'exemple : celui des manipulations symboliques non justifiées ou mal justifiées, comme on en rencontre dans mon étude de cas.

12.1 Manders contre la sémantique

Manders commence ainsi :

> Artifacts in a practice that gives us a grip on life are sometimes thought of in semantic terms—say, as representing something in life. There is, of course, an age-old debate on how geometrical diagrams are to be treated in this regard. [3]

1. Concernant la géométrie d'Euclide, voir Manders 2008b, section 4.1.2, p. 84–87 ; quant à ses thèses plus générales, voir Manders 1999, 2017.

2. L'ensemble de cette section doit beaucoup à mes discussions avec David Rabouin, que je tiens d'autant plus à remercier qu'il risque d'être en désaccord avec l'essentiel.

3. « Parfois, on conçoit les artefacts d'une pratique qui nous donne prise sur la vie en termes sémantiques – par exemple, comme représentant quelque chose dans la vie. Il y a, bien sûr, un débat ancien sur la manière dont il faut, à cet égard, traiter les figures géométriques. » (Manders 2008b, p. 84).

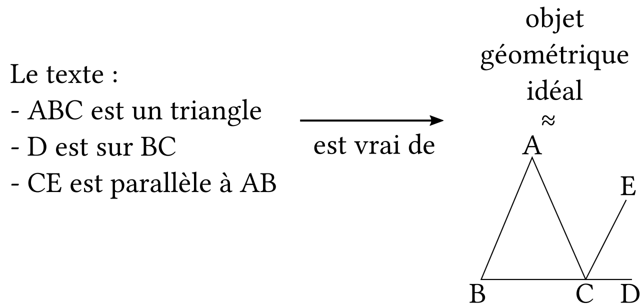


FIGURE 12.1 – La « vision classique » des figures

Les « artefacts » qui intéressent ici Manders sont les figures de la géométrie euclidienne. Le problème est de savoir s'il faut leur donner une sémantique, et si oui laquelle ; l'expression vague de représenter quelque chose « dans la vie » suggère que cette sémantique serait à chercher du côté des situations réelles auxquelles la géométrie peut être appliquée (champs à mesurer, tunnels à creuser, bâtiments à construire ou autres). La stratégie de Manders n'est toutefois pas de réfuter frontalement cette idée générale. Il tente plutôt d'en saper la plausibilité en détruisant une thèse plus spécifique, comme nous allons le voir.

a) Une vision classique des figures

La thèse principale que Manders réfute est en fait la suivante :

> Long-standing philosophical difficulties, on the nature of geometric objects and our knowledge of them, arise from *the assumption that the geometrical text is in an ordinary sense true of the diagram or a 'perfect counterpart'*.[4]

Cette thèse, que j'appellerai la « vision classique » des figures, peut être schématisée comme à la fig. 12.1.

Il faut bien voir que c'est une vision intuitivement plausible de la géométrie élémentaire. Celle-ci porterait sur des objets géométriques parfaits auxquelles nos figures dessinées nous donneraient un accès certes approximatif (nos traits sont épais, parfois un peu tordus, etc.), mais fiable, du moins en ce qui concerne certaines de leurs propriétés. Panza (2012) défend par exemple une position de ce genre.

4. « Des difficultés philosophiques anciennes, sur la nature des objets géométriques et de la connaissance que nous en avons, proviennent de ce que *l'on suppose que le texte géométrique est, en un sens ordinaire, vrai de la figure ou d'une 'contrepartie parfaite' de celle-ci.* » (Manders 2008b, p. 84 ; je souligne).

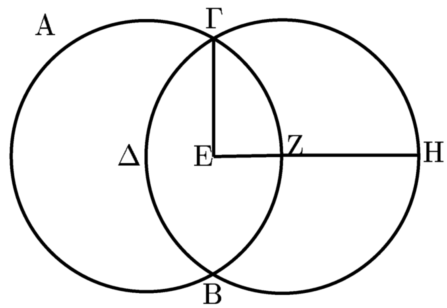


FIGURE 12.2 – Figure de la proposition III.5 des *Éléments* dans le codex B, d'après la reproduction de Saito (2011, p. 54)

b) L'objection des démonstrations par l'absurde

Contre cette vision classique, Manders invoque les démonstrations par l'absurde :

> [A] genuinely semantic relationship between the geometrical diagram and text is incompatible with the successful use of diagrams in proof by contradiction: reductio contexts serve precisely to assemble a body of assertions which patently could not together be true; hence no genuine geometrical situation could in a serious sense be pictured in which they were. This simple-minded objection has nothing to do specifically with geometry: proofs by contradiction never admit of semantics in which each entry in the proof sequence is true (in any sense which entails their joint compatibility). [5]

Pour rendre manifeste le problème que pose Manders, un bon exemple est la proposition III.5 (proposition 5 du livre III) : « si deux cercles se coupent l'un l'autre, ils n'auront pas le même centre [6] ».

Après avoir introduit deux cercles ABΓ et ΓΔH se coupant en deux points B et Γ, Euclide suppose qu'ils ont le même centre E. Or ces hypothèses, comme il le démontre ensuite, sont contradictoires : il est donc impossible de dessiner correctement deux cercles concentriques qui se couperaient en deux points. Le problème, c'est que malgré tout, et comme toujours chez Euclide, il y a une figure (fig. 12.2). Seulement, elle n'obéit manifestement pas aux

5. « [U]ne relation authentiquement sémantique entre la figure géométrique et le texte est incompatible avec le fait que l'on puisse utiliser des figures, avec succès, dans les preuves par l'absurde. Les contextes de *reductio* servent précisément à assembler un ensemble d'assertions qui, manifestement, ne peuvent être vraies simultanément ; on ne saurait donc représenter (en aucun sens raisonnable du terme) une authentique situation géométrique dans lesquels elles le seraient. Cette objection naïve n'a rien de spécifique à la géométrie : les preuves par l'absurde n'admettent jamais de sémantique dans laquelle tous les énoncés successifs de la preuve sont vrais en même temps (en aucun sens du terme qui implique leur compatibilité conjointe). » (Manders 2008b, p. 84).

6. Cf. EEV, I, p. 399–400 = EEHea, II, p. 12 = EEHei, I, p. 176–177.

hypothèses : il y a bien deux cercles qui se coupent, mais le point E ne semble guère être le centre ni de l'un ni de l'autre. (À vrai dire, nous ignorons quelles figures pouvait bien utiliser Euclide lui-même ; j'ai choisi, dans ce chapitre, de reproduire les figures du « codex B », l'un des plus anciens manuscrits des *Éléments* dont nous disposions [7].) Il est donc clair, ici comme dans toutes les démonstrations par l'absurde des *Éléments*, que le texte n'est pas vrai de la figure.

D'après Manders, si cela pose un vrai problème, c'est parce que la vision classique des figures remplit traditionnellement un rôle essentiel : elle explique pourquoi nous pouvons nous appuyer sur la figure au cours de nos raisonnements.

> If diagram imperfections only were in play, one might well hold that the function of diagrams could fruitfully be approached by first elaborating a notion of perfect geometricals of which the text is literally true, then treating diagrams actually drawn in geometrical demonstrations as approximations to perfect ones; finally deriving from all this an understanding of the bearing of the imperfect diagram on inferences in the text. But no detour through ontology and semantics which treats of truth in a diagram in a sense which entails joint compatibility of all claims in force in the reductio context can speak to the difficulty with the role of diagrams in reductio arguments, which are pervasive in Euclid. [8]

Toutes les figures géométriques que nous dessinons sont (au moins un peu) fausses. Il serait sans espoir de chercher à mesurer les angles d'un triangle qu'on a construit pour savoir si ceux d'une version idéale présumée sont exactement égaux. Il est plausible, en revanche, de

7. Comme l'a récemment montré Ken Saito, les manuscrits byzantins des textes mathématiques grecs présentent souvent des figures assez différentes de celles des éditions modernes. (Pour une introduction, voir Saito et Sidoli 2012, Saito 2009, p. 817-825 ou encore Saito 2012, qui s'intéresse entre autres à la propostion que je discute ici. Pour un aperçu plus complet des figures dans les manuscrits des premiers livres des *Éléments*, voir Saito 2006, 2011.) Dans le cas des démonstrations par l'absurde, la différence est particulièrement intéressante : les figures des manuscrits sont souvent plus caricaturalement impossibles que celles des éditions modernes. C'est la raison pour laquelle j'ai choisi, dans toute cette section, de reproduire les figures d'un manuscrit daté de 888, qui (bien que postérieur à Euclide de près de 1200 ans !) est l'un des plus anciens que nous possédions : le « codex B ». (Les sigles utilisés pour désigner les différents manuscrits des *Éléments* ont été introduits par Heiberg ; pour des références complètes, voir Heiberg 1883 ou Saito 2006, p. 95-96.)

8. « S'il ne s'agissait que d'imperfections des figures, on pourrait bien soutenir que la fonction des figures pourrait être abordée avec profit en commençant par élaborer une notion d'[objets] géométriques parfaits dont le texte est littéralement vrai, puis en traitant les figures dessinées en pratique au cours des démonstrations géométriques comme des approximations des figures parfaites ; et enfin, en dérivant de tout cela une compréhension du rôle de la figure imparfaite pour les inférences conduites dans le texte. Mais aucun détour par l'ontologie et la sémantique qui traite la vérité dans une figure en un sens qui implique la compatibilité simultanée de toutes les assertions en vigueur dans le contexte d'une démonstration par l'absurde ne peut répondre à la difficulté que pose le rôle des figures dans les arguments par l'absurde, qui sont omniprésents chez Euclide. » (Manders 2008b, p. 85-86).

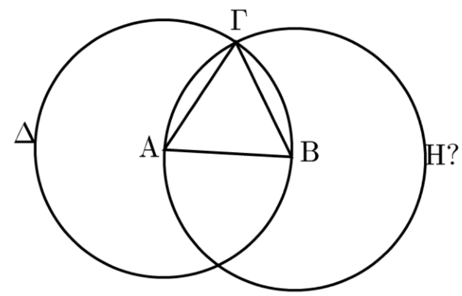


FIGURE 12.3 – Figure de la proposition I.1 des *Éléments* dans le codex B (Saito 2006, p. 97)

penser que certaines propriétés du diagramme ne sont pas affectées par ces approximations, et qu'elles reflètent les propriétés de l'objet géométrique idéal visé ; c'est en particulier la position que défend Panza. Prenons un exemple célèbre, celui de la proposition I.1 des *Éléments* (fig. 12.3), qui montre comment construire un triangle équilatéral sur un segment donné [9]. D'après Panza, il est légitime, dans le cadre de la géométrie plane d'Euclide, de conclure sur la base de la figure que les deux cercles s'intersectent [10].

Tout le problème est que les démonstrations par l'absurde aussi s'appuient sur leurs figures. Reprenons par exemple la proposition III.5 (cf. fig. 12.2 ci-dessus). Après avoir introduit les deux cercles ΑΒΓ, ΓΔΗ, leurs intersections Β, Γ et leur centre Ε, Euclide écrit :

> [Q]ue soit jointe ΕΓ et que soit conduite ΕΖΗ, au hasard. Et puisque le point Ε est le centre du cercle ΑΒΓ, ΕΓ est égale à ΕΖ. Ensuite puisque le point Ε est le centre du cercle ΓΔΗ, ΕΓ est égale à ΕΗ. Or il a été démontré aussi que ΕΓ est égale à ΕΖ. Et donc ΕΖ est égale à ΕΗ, la plus petite à la plus grande. Ce qui est impossible. [11]

La démonstration repose sur la ligne ΕΖΗ, tirée « au hasard », qui est supposée rencontrer ΑΒΓ en Ζ et ΓΔΗ en Η (dans cet ordre). Seule la figure nous fournit une ligne remplissant ces conditions. Si l'on voulait, d'un point de vue moderne, réécrire la preuve sans figure, il serait nécessaire de faire appel à un axiome supplémentaire du même genre que pour le point d'intersection de la proposition I.1 [12].

9. Cf. EEV, I, p. 194–195 = EEHea, I, p. 241–242 = EEHei, I, p. 10–13

10. D'une manière générale, Panza appelle « rôle local » des figures dans la géométrie plane d'Euclide le fait qu'elles permettent d'attribuer certaines propriétés aux objets géométriques correspondants ; Panza appelle « *diagrammatic attributes* » les propriétés que les objets géométriques héritent de leurs figures. (Voir Panza 2012, en part. p. 72–82.)

11. Trad. de EEV, I, p. 400, où (par cohérence avec ma figure) j'ai réinséré les lettres grecques que l'on trouve chez Heiberg. Cf. aussi EEHea, II, p. 12 = EEHei, I, p. 177.

12. Voici une possibilité qui peut se formaliser dans le système d'Avigad, Dean et Mumma 2009 (j'indique entre crochets, à chaque étape, la règle de ce système que j'invoque ; le point crucial est qu'on a besoin des

L'argument de Manders est alors clair. Les démonstrations par l'absurde des *Éléments*, tout comme les démonstrations directes, font un usage essentiel de la figure. Or si dans le cas des preuves directes, cela peut se justifier en supposant que la figure reflète certaines au moins des propriétés d'objets géométriques idéaux, cette réponse n'est plus disponible pour les preuves par l'absurde, puisque précisément il ne pourrait pas exister d'objets géométriques répondant aux hypothèses. À moins de postuler que les figures fonctionnent de manière fondamentalement différente dans les preuves directes et dans les preuves par l'absurde, il faut donc chercher une autre interprétation de leur rôle.

c) Est-ce réellement un problème ?

À première vue, l'argument de Manders n'est pas convaincant. Comme il l'écrit lui-même,

> It does not follow that there could not be a picturing-like relationship between the diagram and *some* claims in force within a reductio context. [13]

En d'autres termes, on pourrait conserver une version *partielle* de la vision classique (schématisée à la fig. 12.4) d'après laquelle seulement certaines des assertions du texte seraient vraies de la figure. Dans notre exemple de III.5, la figure montre bien deux cercles qui se coupent en deux points ; seulement, le point E n'est pas leur centre.

De ce point de vue, l'usage de la figure en III.5 s'explique sans mal. La figure représente adéquatement deux cercles et un point *qui leur est intérieur* ; sur cette base inattaquable, on peut construire Z et H. On peut ensuite obtenir la contradiction en faisant intervenir l'*hypothèse supplémentaire* que E est le centre des deux cercles, mais là, plus aucune référence à la figure n'est nécessaire. Pour citer Panza (qui parle d'un autre exemple de preuve

règles de construction d'intersection 4 et 5). Puisque les deux cercles ΑΒΓ et ΓΔΗ sont différents [hypothèse], on peut prendre un point H de ΓΔΗ qui n'est pas sur ΑΒΓ [règle de construction de point n°7]. De deux choses l'une. Ou bien H est à l'extérieur de ΑΒΓ, et alors, puisque E est à l'intérieur de ΑΒΓ, la droite EH intersecte ΑΒΓ en un point Z situé entre E et H [règle de construction d'intersection n°4] ; on peut alors terminer comme Euclide [plusieurs possibilités, par exemple axiomes de transfert figure-segment n°1 et 3]. Ou bien H est à l'intérieur de ΑΒΓ, et alors la droite EH intersecte ΑΒΓ en un point Z tel que H est entre E et Z [règle de construction d'intersection n°5] ; on peut alors terminer comme Euclide en inversant les rôles de Z et H [comme précédemment]. En réalité, dans le système d'Avigad, Dean et Mumma 2009, cette preuve est inutile : ce système comporte un principe plus fort que ceux d'Euclide, l'axiome de transfert figure-segment n°2, qui rend triviale la proposition III.5 (voir *infra*, p. 302–303). Cela n'a toutefois pas d'importance pour mon propos ici.

13. « Il ne s'ensuit pas qu'il ne peut pas y avoir un genre de relation de dépiction entre la figure et *certaines* des assertions en vigueur dans un contexte de reductio. » (Manders 2008b, p. 85). Les « hypothèses en vigueur dans un contexte de reductio », ce sont les hypothèses sous lesquelles on arrive à une contradiction ; le terme de « contexte » vient ici de la déduction naturelle.

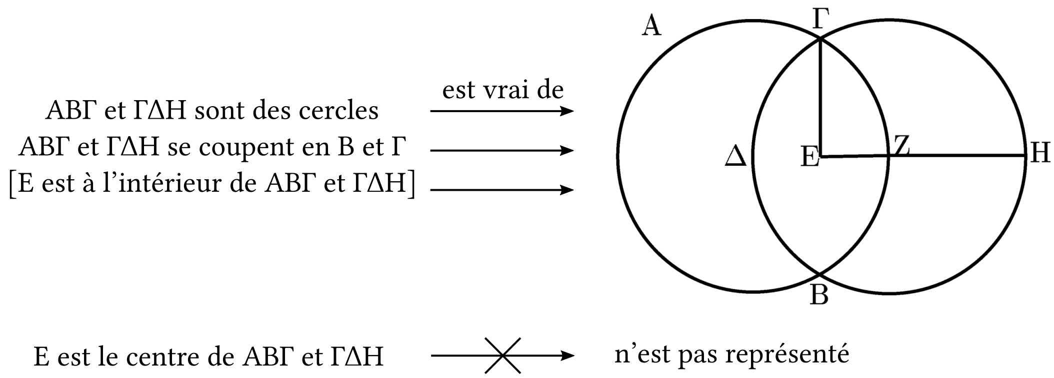


FIGURE 12.4 – Une version partielle de la vision classique des figures

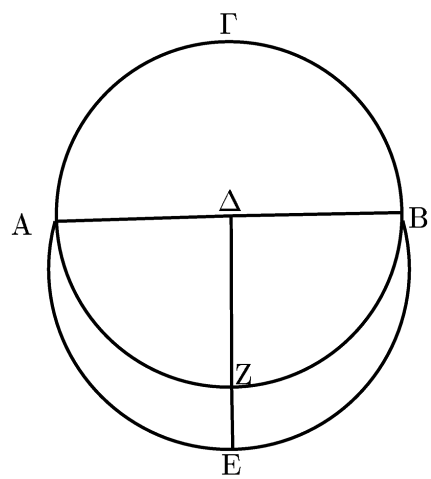


FIGURE 12.5 – Figure de la proposition III.2 dans le codex B (Saito 2011, p. 52)

par l'absurde), « cet argument ne démontre aucune impossibilité relative à la figure qui y intervient [14] ».

Malheureusement, le problème de Manders ne se laisse pas écarter si vite. Tout d'abord, il y a des preuves par l'absurde plus problématiques, qui exigent une figure impossible en un sens plus fort. Examinons par exemple III.2, « si deux points sont pris au hasard sur la circonférence d'un cercle, la droite joignant ces points tombera à l'intérieur du cercle [15] » (fig. 12.5) ou encore III.10, « un cercle ne coupe pas un cercle en plus de deux points [16] » (fig. 12.6). Dans le premier cas, la ligne droite AB ne peut être droite ; dans le second, au

14. « Clearly, this argument does not show any impossibility relative to the diagram involved in it. » (Panza 2012, 83, n. 53). Cette remarque porte sur la démonstration par l'absurde de I.6.
15. Cf. EEV, I, p. 394–395 = EEHea, II, p. 8–9 = EEHei, I, p. 168–171.
16. Cf. EEV, I, p. 412–413 = EEHea, II, p. 23–24 = EEHei, I, p. 192–195.

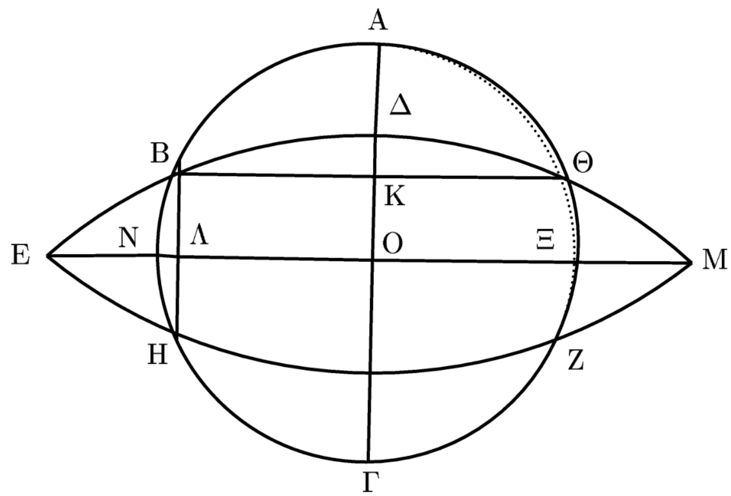


FIGURE 12.6 – Figure de la proposition III.10 dans le codex B (Saito 2011, p. 59)

moins l'un des deux cercles ne peut être un cercle. Cette fois, il semble bien y avoir une authentique impossibilité relative aux figures. Plus exactement, si l'on voulait adopter la même stratégie que pour III.5 ci-dessus, il faudrait considérer que certaines des assertions essentielles du texte (que AB est une ligne droite, que ABΓ et ΔEZ sont des cercles) ne sont pas représentées. Or contrairement à III.5, ce qui reste n'est alors plus un assemblage familier de cercles et de lignes droites, mais un dessin exotique fait de lignes courbes mal définies. Est-ce un problème ? Notre but était de justifier l'usage démonstratif des figures en soutenant qu'elles reflétaient adéquatement certaines des propriétés d'objets géométriques idéaux. Les propositions III.2 et III.10 (par exemple) ne rendent pas ce projet impossible, mais il y a un prix à payer : il faut admettre que les figures peuvent nous permettre, au détour d'une preuve, de découvrir les propriétés d'objets géométriques plus généraux que les cercles et les droites d'Euclide, par exemple des lignes courbes quelconques.

On peut ajouter une seconde objection, plus générale et plus profonde, contre les tentatives précédentes de sauver la vision classique des figures *via* une variante partielle. On peut bien suggérer, comme je l'ai fait (cf. fig. 12.4), que la figure de III.5 ne représente pas E comme le centre des deux cercles, mais seulement comme un point à l'intérieur de ceux-ci ; ou encore que la figure de III.2 représente ne représente pas AB comme une ligne droite. Il faut cependant admettre que le texte d'Euclide ne dit rien de précis sur ce que la figure représente ou ne représente pas. Pour expliciter le rôle de la figure dans III.5 (fig. 12.4 ci-dessus), j'ai même été conduit à introduire une assertion fantôme, absente du texte et que j'ai notée entre crochets ([E est à l'intérieur de ABΓ et ΓΔH]). Autant la vision classique (fig. 12.1) était simple, autant sa variante partielle est soumise à des normes peu claires : qu'a-t-on le droit de ne pas représenter ? Quelles déformations peut-on faire subir à nos cercles et à nos droites ?

On comprend donc pourquoi Manders peut considérer qu'une fois qu'on a admis que la figure n'est pas purement et simplement un modèle (peut-être approximatif) du texte, on a concédé l'essentiel : le fonctionnement des figures dans les preuves par l'absurde n'est pas transparent et mérite une analyse plus approfondie. Il écrit ainsi :

> Thus one is forced back to a direct attack on the way diagrams are used in reductio argument; the problem of the relationship between diagram and geometric inference here turns out to be one of standards of inference not reducible in a straightforward way to an interplay of ontology, truth, and approximate representation. But once this is admitted, there seems to be no reason why direct inferential analysis of diagram-based geometrical reasoning should not be the approach of choice to characterizing geometrical reasoning overall, with or without reductio. [17]

L'« analyse inférentielle directe » de Manders consiste avant tout à mettre au premier plan, non plus ce que les figures représentent, mais l'exigence que différents praticiens puissent *se mettre d'accord* sur ce que les figures montrent. Par exemple, si l'on ne peut pas observer sur une figure que deux lignes sont égales, c'est fondamentalement parce que deux praticiens traçant des figures différentes pourraient arriver à des conclusions différentes. J'ai déjà évoqué les résultats de son analyse plus en détail [18].

d) Une solution sémantique

L'argument des démonstrations par l'absurde ne suffit cependant pas à réfuter l'idée d'une sémantique *en général* : s'il est convaincant contre la vision classique des figures (fig. 12.1), il ne dit rien contre l'idée que les figures elles-mêmes pourraient avoir une sémantique (fig. 12.7). La conclusion naturelle de l'argument est, dans les termes de Manders lui-même, qu'il faut traiter les figures comme des « composants *textuels* » de la preuve

17. « On est donc contraint de revenir à une attaque directe de la manière dont les figures sont utilisées dans les arguments par l'absurde ; le problème du lien entre figure et inférence géométrique s'avère ici être celui de normes d'inférences qui ne sont pas réductibles de manière directe à des questions d'ontologie, de vérité et de représentation approximative. Mais une fois cela admis, on ne voit pas pourquoi une analyse inférentielle directe ne serait pas l'approche privilégiée pour caractériser le raisonnement géométrique en général, avec ou sans preuve par l'absurde. » (Manders 2008b, p. 86).

18. Voir *supra*, en part. section 11.2.

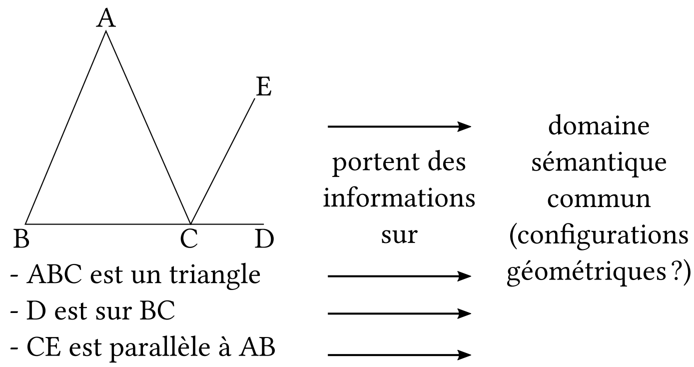


FIGURE 12.7 – Une conception des figures géométriques dans l'esprit de Barwise et Etchemendy

« plutôt que comme des contreparties sémantiques [19] ». Mais précisément, cela n'exclut pas de donner une sémantique à la fois aux figures *et* au texte, à la manière de Barwise et Etchemendy [20].

À quoi une telle sémantique des figures pourrait-elle ressembler, et comment rendrait-elle intelligibles les démonstrations par l'absurde ? Pour y voir plus clair, examinons d'abord un exemple en arithmétique qui ne requiert pas de figure [21], la proposition 22 du livre VII : « les nombres les plus peits parmi ceux qui ont le même rapport qu'eux sont premiers entre eux [22] ». Euclide raisonne par l'absurde et se donne deux nombres A et B sur lesquels il fait deux hypothèses : A et B sont les plus petits parmi ceux qui ont le même rapport qu'eux (nous noterons cette hypothèse p), et ne sont pas premiers entre eux (ce que nous noterons q). Rien n'empêche de donner une sémantique à ces deux hypothèses, ni même à leur conjonction $p \wedge q$. Certes, dans ce dernier cas, l'ensemble de modèles correspondant sera vide, puisqu'il n'existe pas de nombres A et B qui vérifient à la fois p et q. On peut toutefois utiliser $p \wedge q$ comme point de départ pour une véritable démonstration par l'absurde, parce qu'il n'y a pas de *règle d'inférence* (syntaxique) permettant d'en déduire une contradiction de manière immédiate : les hypothèses, pour ainsi dire, ne s'auto-détruisent pas.

Revenons maintenant à une démonstration par l'absurde comme celle de la proposition III.10. Elle peut très bien être pensée sur le même modèle. À vrai dire, on peut concevoir

19. L'auteur souligne. Cette citation vient d'une version précédente de son article sur la géométrie d'Euclide ; voir *supra*, p. 222.

20. Voir *supra*, en particulier chap. 8.

21. À vrai dire, les propositions arithmétiques des *Éléments* comportent aussi des sortes de figures, qui représentent les nombres dont il est question dans le texte par des segments ; mais contrairement aux propositions des livres de géométrie, ces figures ne jouent guère de rôle dans les démonstrations (cf. par ex. Mueller 1981, p. 67).

22. Cf. EEV, II, p. 328 = EEHea, II, p. 323–324 = EEHei, II, p. 234–237.

au moins deux approches sémantiques différentes pour la fig. 12.6 qui l'illustre. Dans l'esprit des discussions qui précèdent, on peut considérer qu'elle représente seulement deux courbes fermées s'intersectant en quatre points. Elle serait alors comparable à l'une des hypothèses de la *reductio* de VII.22, par exemple p : elle n'est pas sémantiquement contradictoire à elle seule, mais le devient si on lui adjoint des hypothèses supplémentaires, à savoir que ΑΒΓ et ΔΕΖ sont des cercles. On peut aussi en faire une lecture plus naïve, qui, dans la terminologie du chapitre précédent, ne cherche pas à décontextualiser les figures [23], et considérer qu'elle représente bien deux *cercles* qui s'intersectent en quatre points. Cette fois, c'est à la conjonction $p \wedge q$ plutôt qu'à l'une des hypothèses de VII.22 qu'elle serait comparable, c'est-à-dire qu'elle aurait pour sémantique un ensemble vide de modèles. Mais même dans cette hypothèse, rien n'empêche de l'utiliser pour une démonstration par l'absurde : tout ce qui compte est qu'il ne soit pas permis d'en déduire une contradiction directement (par exemple en observant que les cercles n'en sont pas). En bref, comme pour les hypothèses de VII.22, il suffit que la figure ne s'auto-détruise pas. Le fait qu'elle soit contradictoire au sens sémantique (c'est-à-dire qu'elle n'ait pas de modèles) n'est pas davantage un problème pour la figure qu'elle ne l'est pour $p \wedge q$.

Dans une note, Manders semble rejeter une nouvelle fois la possibilité d'une sémantique des figures qui permette les démonstrations par l'absurde, mais son argument est, là encore, fondé sur une équivoque : il critique une forme particulière de « sémantique » (en fait la vision classique des figures, selon laquelle elles sont la sémantique du texte), mais parle de sémantique des figures en général, comme s'il avait les moyens de réfuter aussi une sémantique des figures à la Barwise-Etchemendy, ce qui n'est pas le cas. Voici ce qu'il écrit :

> Jerry Seligman has suggested that one might avoid the argument [of *reductio* proofs] via some kind of 'compositional semantics' of diagrams. To avoid the argument, however, geometrically incompatible sets of sentences would have to be 'compositionally true' in the same diagram. That is to say (and the point has nothing to do with diagrams or even geometry), a 'semantics' that fills the bill here thereby lacks *minimal soundness*, the minimum requirement for notions of (weak) truth, that a set of sentences all 'true' in the same situation, cannot trivially entail a contradiction. [24]

23. Voir ci-dessus, section 11.2.

24. « Jerry Seligman a suggéré que l'on pourrait échapper à l'argument [des démonstrations par l'absurde] *via* une 'sémantique compositionnelle' des figures, sous une forme ou une autre. Pour échapper à l'argument, cependant, il faudrait que des ensembles d'énoncés géométriquement incompatibles soit 'compositionnellement vrais' dans la même figure. En d'autres termes (et cela n'a rien à voir avec les figures ni même avec la

La remarque de Jerry Seligman, un collaborateur de Jon Barwise évoqué plus haut [25], peut être reconstituée : la fig. 12.6 de la proposition III.10 pourrait représenter, *via* différents de ses éléments (c'est là l'aspect compositionnel), à la fois que ABΓ et ΔEZ sont des cercles et qu'ils se coupent en quatre points B, H, Z, Θ ; ces énoncés sont certes incompatibles (la figure n'admet donc pas de modèles) mais ont un sens individuellement, de sorte que la figure n'est pas plus absurde que l'énoncé $p \wedge q$ de notre exemple arithmétique. Le malentendu est que Manders persiste à traiter la figure comme une sorte de modèle, alors que Seligman la traite manifestement comme une représentation susceptible d'admettre une sémantique.

En bref, la polémique de Manders contre un traitement sémantique des figures est fondée sur un malentendu. Il n'envisage pas de traiter les figures comme représentations munies de sémantique ou porteuses d'information, une possibilité qui est parfaitement claire, non seulement pour Barwise, Etchemendy et Seligman, mais aussi pour Netz dans son étude sur la géométrie grecque :

> Diagrams, just like words, may be a way of encoding information. If, then, diagrams are seen in this way, to ask 'how can diagrams be true?' is like asking 'how can language be true?' – not a meaningless question, but clearly a different question from that we started from. [26]

L'argument de Manders est donc insuffisant.

Pour finir, et quoique la question n'ait qu'une importance tangentielle pour mon propos, je voudrais souligner que les considérations qui précèdent ne suffisent pas à régler le problème des démonstrations par l'absurde ; elles montrent seulement que Manders en tire une conclusion exagérée. Il reste en effet une difficulté : quelles sont les figures qui sont acceptables dans une démonstration ? Comme le remarque Rabouin (2015), si l'on pouvait, en toutes circonstances, tracer des cercles écrasés ou des droites tordues, on pourrait facilement écrire des preuves fausses [27]. Le système formel d'Avigad, Dean et Mumma (2009)

---

géométrie), une 'sémantique' qui conviendrait ici ne satisferait pas à la *correction minimale*, l'exigence minimale pour les notions de vérité (faible), qui est qu'un ensemble d'énoncés tous 'vrais' dans la même situation ne peuvent pas trivialement impliquer une contradiction. » (Manders 2008b, p. 86, n. 3).

25. Ils ont développé ensemble la théorie de l'information de Barwise et Seligman 1997, mentionnée aux chapitres 9 et 10.

26. « Les diagrammes, comme les mots, peuvent être une manière d'encoder de l'information. Or si l'on voit les diagrammes de cette manière, demander "comment les diagrammes peuvent-ils être vrais ?" est comme demander "comment la langue peut-elle être vraie ?" – une question qui n'est pas dépourvue de sens, mais qui est clairement différente de celle dont nous sommes partis. » (Netz 1999, p. 33).

27. Voir Rabouin 2015, en part. p. 115–118 et 126–131. Manders discute assez longuement la question de savoir comment les normes qui exigent qu'un figure doive être raisonnablement précis doivent être relâchées sélectivement pour les démonstrations par l'absurde ; cf. Manders 2008b, p. 109-118.

permet d'explorer la question plus précisément et confirme la difficulté. Si l'on peut y reconstruire toutes les démonstrations par l'absurde des trois premiers livres des *Éléments*[28], même celles qui exigent apparemment une figure impossible[29], un certain nombre de celles du livre III sont rendues inutiles par les axiomes ou règles du système[30], et ce pour une bonne raison : ces règles, trop fortes pour le livre III, sont requises pour certaines propositions du livre I[31]. On peut reformuler ce problème en disant que le livre I suppose des normes d'acceptabilité des figures qui excluent certaines des démonstrations par l'absurde du livre III.

e) Le vrai problème : une sémantique, pour quoi faire ?

En un sens cependant, la discussion précédente passe à côté de la vraie leçon du texte de Manders. Peut-être une sémantique des diagrammes n'est-elle pas impossible, mais elle serait sans intérêt. Comme il l'écrit :

> If this order of analysis [a direct inferential analysis of diagram-based geometrical reasoning] proves fruitful, ontological and semantic considerations will seem decidedly less central to the philosophical project of appreciating geometry as a means of understanding. For in their then remaining role of making the standards of geometrical reasoning seem appropriate, ontological-and-semantic pictures will have to compete with other types of considerations which we will find have potential to shape a reasoning practice.[32]

En d'autres termes, on peut peut-être donner une sémantique aux figures comme je l'ai proposé, mais celle-ci ne serait pas très éclairante parce qu'elle ne vient qu'après coup. Ce qui explique la manière dont les figures sont utilisées dans la pratique euclidienne, ce n'est pas fondamentalement ce qu'elles représentent, d'après Manders ; ce sont plutôt d'autres

28. Il s'agit là d'un avantage du système d'Avigad, Dean et Mumma (2009) sur ceux de Miller et de Mumma, qui obligent à sacrifier certaines de ces démonstrations, comme le souligne Rabouin 2015, note 33 p. 121.

29. Les auteurs semblent affirmer le contraire dans le cas de la proposition III.5 (Avigad, Dean et Mumma 2009, p. 740) mais c'est sans doute une erreur (cf. ci-dessus, note 12 p. 295).

30. La proposition III.2 est trivialisée par l'axiome de cercle n°2 (ainsi que par l'axiome de transfert figure-segment n°4) et les propositions III.5 et III.6 par l'axiome de transfert figure-segment n°2.

31. Voir par exemple la proposition I.22.

32. « Si cet ordre d'analyse [c'est-à-dire une analyse inférentielle directe du raisonnement géométrique fondé sur des figures] s'avère fructueux, les considérations ontologiques et sémantiques apparaîtront résolument moins centrales pour le projet philosophique d'évaluer la géométrie comme mode de compréhension. En effet, pour le rôle qui leur reste, celui de donner aux normes du raisonnement géométrique l'apparence de la plausibilité, les perspectives ontologiques-et-sémantiques seront en compétition avec d'autres genres de considérations dont nous verrons qu'elles peuvent influer sur la forme d'une pratique de raisonnement. » (Manders 2008b, p. 86).

genres de facteurs, par exemple l'exigence que les praticiens de la géométrie puissent toujours se mettre d'accord sans difficultés sur leurs inférences. On peut certes donner une sémantique qui rende compte des normes de la pratique, par exemple, comme le fait Miller, en disant que les diagrammes euclidiens ne représentent que les propriétés topologiques (en un sens à préciser) des configurations géométriques, mais ce n'est qu'une reconstruction *a posteriori*, qui vient légitimer des normes dont la source est ailleurs.

Face à une pratique mathématique, conclut Manders, ce qu'il importe d'étudier avant tout, ce sont ses normes inférentielles ; on peut peut-être y ajouter une analyse sémantique, mais chercher à l'utiliser comme point de départ risque surtout d'induire en erreur. C'est d'ailleurs la raison principale pour laquelle il préfère, à partir de la fin des années 1990, parler d'« utilisation d'artefacts [33] » ou plus récemment de « moyens expressifs [34] ». Dans des notes de cours récentes, il clarifie sa position comme suit :

> The accont we will be developing would probably be classified by Brandom [35] as "expressivist functionalism"; except that we remain agnostic as to whether or not our focus on expression could be usefully supplemented by what he calls representationalist accounts of the role of expressions. [36]

Il est donc clair que ce qui lui importe n'est pas tant d'affirmer l'impossibilité d'une approche « représentationnelle » (qui, par exemple, donnerait une sémantique aux figures) que de *marginaliser* une telle approche au profit d'une analyse inférentielle.

12.2 Les règles de Bernoulli, un calcul sans justification ?

Mon étude de cas présente un autre genre d'exemple dans lequel parler de représentation pose problème. Les règles du « genre singulier de calcul » de Bernoulli [37] ne peuvent pas être justifiées d'après la manière dont Leibniz et Bernoulli comprennent jusqu'alors les symboles d, d^2, etc. En dernière instance, ce qui les légitime est seulement leur succès à produire des résultats corrects, vérifiables par ailleurs.

La réaction de Pierre Varignon face au calcul de Bernoulli rend cela très clair [38]. Vari-

33. « Artifact use » ; voir par exemple Manders 1999, p. 9, n. 12 pour une discussion explicite.
34. Cf. par exemple Manders 2012, 2017.
35. Cf. Brandom 2000.
36. « La manière de rendre compte des mathématiques que nous allons développer serait probablement classifiée par Brandom comme un "fonctionnalisme expressiviste", à ceci près que nous restons agnostique sur la question de savoir s'il serait utile d'ajouter à notre étude centrée sur les expressions ce que Brandom appelle une analyse représentationnaliste du rôle des expressions. » (Manders 2017, p. 4).
37. Voir première partie, en part. section 2.4.
38. Varignon découvre cette méthode à travers des extraits de la correspondance entre Leibniz et Jean Ber-

gnon commence par interpréter les règles de Bernoulli comme des égalités valables selon l'interprétation habituelle des symboles d, d^2, etc. et elles lui paraissent donc tout simplement fausses :

> [E]n faisant ainsy $d^1x \cdot d^1x = d^2x = ddx$, ce seroit confondre [...] une difference seconde avec le quarré d'une premiere [...]. Il est vray que 1, ddx ou ddx & $dx \times dx$ sont grandeurs homogenes ; mais elles ne sont pas égales, à moins [...] que 1, dx, ddx, ne soient en proportion continue [39].

Après plusieurs élucidations de Bernoulli, Varignon finit par comprendre que ces règles ne sont véritablement justifiées que par leurs résultats :

> Je voy presentement tout l'artifice de votre nouvelle maniére d'integrer. Ce qui me fesoit de la peine, c'est que je ne pouvois me résoudre au double sens que vous donnez à la lettre d, n'étant pas permis dans un raisonnement juste de passer ainsy d'un sens à l'autre. Mais je voy bien presentement que vous ne concluez pas l'un de l'autre : vous les employez seulement tous deux comme arbitrairement d'abord & sans consequence ; & à la fin vous vous trouvez comme sans y penser, à une grandeur que vous sçavez d'ailleurs, & non pas par ce calcul, être l'intégrale cherchée. D'où vous concluez, non en vertu d'aucun raisonnement qu'il y ait dans ce calcul, mais par le seul succés de l'intégrale generale qui se trouve au bout [...] [40].

Certes, Bernoulli pourrait justifier ses règles non seulement par leur succès dans certains cas, mais aussi, heuristiquement, par l'analogie proposée par Leibniz entre puissances et différences. Toutefois, il ne dispose pas d'une interprétation de ses symboles qui permettrait de les justifier véritablement. Bien sûr, rien n'empêche de donner une telle interprétation *a posteriori*, mais à quoi cela sert-il ? Cela permet seulement d'expliquer après coup le succès de ses calculs, mais pas de comprendre, pour reprendre les termes de Manders, les pressions qui les ont façonnés.

Ce n'est pas un cas isolé. À vrai dire, la tension entre interprétation et manipulations guidées par la syntaxe est fréquemment mise en avant dans la littérature récente, et la manipulation algébrique d'opérateurs (dont on considère parfois Bernoulli comme le pionnier)

---

noulli qu'il consulte chez le marquis de l'Hôpital (cf. *supra*, note 98 p. 70). Il écrit ensuite plusieurs fois à Bernoulli pour lui demander des éclaircissements, ce qui donne lieu à un échange de plusieurs mois : voir les lettres de Varignon à Bernoulli des 4 septembre (BBW, II, p. 201–202) et 16 décembre 1698 (p. 207–208) et des 19 février (p. 217) et 19 mars 1699 (p. 223), ainsi que les réponses de Bernoulli du 4 octobre 1698 (p. 202–204) et du 24 janvier 1699 (p. 213–214).

39. Lettre de Varignon à Bernoulli du 4 octobre 1698, BBW, II, p. 201–202.

40. Lettre de Varignon à Bernoulli du 19 mars 1699, BBW, II, p. 223.

en sert régulièrement d'exemple paradigmatique [41].

12.3 Conclusion : le rôle de la sémantique

Ces deux exemples sont très différents. À travers Euclide, Manders (2008b) cherche à analyser une pratique démonstrative relativement homogène et stable (quoique, comme je l'ai mentionné en passant, certaines des difficultés qu'il rencontre viennent peut-être de décalages entre les différents livres des *Éléments* [42]) ; les facteurs non sémantiques qu'il met en avant pour expliquer cette pratique sont sociaux, à savoir l'importance de pouvoir résoudre tous les désaccords sans ambiguïté. Mon étude de cas traite au contraire d'un contexte de découverte, où les manipulations sont plutôt légitimées par les résultats auxquelles elles conduisent. L'intérêt de les mettre malgré tout côte à côte est de montrer les limites d'une approche sémantique. Dans ces exemples, attribuer une sémantique aux figures ou aux formules permet certes d'en justifier *a posteriori* l'usage, mais ne l'éclaire pas.

Il est clair, toutefois, qu'il ne faut pas pousser cet argument trop loin. Les phénomènes que Shimojima appelle « signification dérivée », discutés au chapitre 10, seraient incompréhensibles si l'on n'attribuait pas aux utilisateurs des diagrammes concernés une compréhension des conventions sémantiques qui les gouvernent.

41. Du côté des mathématiciens, voir par exemple Cartier 2000 ; du côté des philosophes, on peut citer M. Wilson 2006, en part. section 8.viii et plus largement R. Wagner 2017, en part. chap. 4. En un sens, on pourrait aussi ajouter Ferreirós 2016, chap. 4.

42. Cf. ci-dessus, p. 302–303.

Conclusion générale

Comment comprendre l'idée de différences *représentationnelles*, et plus précisément le fait que l'on puisse progresser en changeant de représentation ? Pour répondre à cette question, j'ai commencé par étudier un changement notationnel apparemment anodin introduit par Leibniz, et essayé de comprendre en détail ce qu'il apportait (première partie) ; en particulier, nous avons vu qu'il rendait plus manifeste certaines structures dans les formules et qu'en transformant ce qui pouvait s'exprimer de manière simple, il rendait certaines voies plus accessibles à l'exploration (chap. 3). J'ai ensuite tenté de clarifier différentes manières dont les notions d'information ou de sémantique sont comprises et appliquées à des représentations de différentes sortes dans la littérature contemporaine de philosophie des mathématiques.

Le premier résultat de mon enquête est qu'il faut distinguer soigneusement différents sens en lesquels on peut parler d'information. On peut le faire en un sens minimal qui ne présuppose aucune véritable sémantique (deuxième partie). On peut identifier l'information portée par une représentation à un ensemble de modèles (troisième partie, chap. 7 et 8). On peut enfin viser une notion d'information explicite, modélisant ce que les utilisateurs des représentations comprennent des conventions qui gouvernent celles-ci (troisième partie, chap. 9 et 10).

Employer cette notion en ces différentes acceptions sert des fonctions différentes. Parler d'information au sens minimal de la deuxième partie revient à identifier quelles caractéristiques de nos représentations nous considérons comme pertinentes, c'est-à-dire quelles paires de représentations nous traiterions comme essentiellement identiques. Cela permet aussi de comparer des représentations quant à leur contenu d'information, mais seulement de manière grossière : on peut dire qu'une représentation contient les mêmes, ou plus, ou moins d'informations qu'une autre, mais rien de plus. Définir une sémantique en termes de modèles permet de parler de vérité et de se doter d'une relation de conséquence entre représentations, c'est-à-dire de justifier un certain usage de représentations « de l'extérieur », pour ainsi dire, comme le fait usuellement la sémantique dans le cas des langages logiques. (Les justifications de ce genre, qui peuvent être *a posteriori* et éloignées de la pratique, ne sont toutefois pas toujours très éclairantes, comme nous l'avons vu au chap. 12.) Cela permet aussi de comparer le contenu d'information de représentations de manière plus fine, en comparant leurs ensembles de modèles. Enfin, la dernière approche permet de rester au plus près de la manière dont les utilisateurs de représentations les interprètent et les emploient pour progresser.

Ces clarifications préalables en main, nous pouvons revenir au problème de départ, celui des différences représentationnelles. Trois réponses se dégagent des pages qui précèdent.

Tout d'abord, la deuxième partie nous a montré que deux représentations informationnellement équivalentes au sens minimal – le plus exigeant – peuvent néanmoins différer « computationnellement », comme l'écrit Herbert Simon. Cependant, nous avons vu que son modèle informatique ne fonctionne que si l'on fixe d'avance toutes les opérations permises sur nos représentations (par exemple, tout ce qu'on a le droit d'observer sur une figure) : il ne permet d'éclairer que des *reconstructions* de l'usage de certaines représentations. En pratique, on ne peut pas délimiter *a priori* tout ce que l'on peut observer sur une représentation donnée, comme le montrent les exemples de la visualisation de données (sect. 5.2), mais aussi les structures que Leibniz observe dans ses formules, étudiées dans la première partie. Sur ce point, nous n'avons guère fait que mettre le doigt sur la difficulté ; mais peut-être n'y a-t-il rien de plus à en dire.

Le chapitre 10 nous a montré que certaines des différences importantes entre formes de représentations, et en particulier certains des avantages essentiels des diagrammes, ne peuvent se comprendre que si l'on conduit une analyse informationnelle (ou, si l'on veut, « sémantique ») au troisième sens du terme, c'est-à-dire si l'on s'en tient aux conventions que comprennent les utilisateurs des représentations. À travers ce qu'il appelle les les « passagers clandestins » et la « signification dérivée », Shimojima montre comment certains diagrammes peuvent être précieux au cours du raisonnement.

La troisième et dernière réponse est la plus délicate. L'étude de cas de la première partie montre que l'une des contributions essentielles du changement notationnel de Leibniz est de transformer ce qui peut s'écrire sous une forme simple. Beaucoup des cas les plus intéressants de différences représentationnelles importantes en mathématiques pourraient sans doute être pensés sur ce modèle ; or il ne semble pas éclairant d'utiliser les outils précédents pour les comprendre. Ils méritent une réflexion plus poussée, pour un travail ultérieur.

Table des matières détaillée

Table des figures

Références bibliographiques

Abréviations

BBW Bernoulli, Jean (1955–1992). *Der Briefwechsel von Johann I Bernoulli.* Hrsg. von der Naturforschenden Gesellschaft in Basel. Basel: Birkhäuser.
Bd. I: Hrsg. von O. Spiess. 1955.
Bd. II: *Der Briefwechsel mit Pierre Varignon, Erster Teil: 1692–1702.* Hrsg. von Pierre Costabel und Jeanne Peiffer. 1988.
Bd. III: *Der Briefwechsel mit Pierre Varignon, Zweiter Teil: 1702–1714*. Hrsg. von Pierre Costabel und Jeanne Peiffer. Basel: Springer. 1992.

BOO Bernoulli, Jean (1742). *Opera omnia. Tam antea sparsim edita, quam hactenus inedita.* Éd. par Gabriel Cramer. 4 vol. Lausannæ et Genevæ : M.-M. Bousquet & soc.
Vol. I: *Quo continentur ea quæ ab Anno 1690 ad Annum 1713 prodierunt*.
Vol. II: *Quo continentur ea quæ ab Anno 1714 ad Annum 1726 prodierunt*.
Vol. III: *Quo continentur ea quæ ab Anno 1727 ad hanc usque diem prodierunt. Accedunt Lectiones Mathematicæ de Calculo Integralium.*
Vol. IV: *Quo continentur Ανεκδοτα*.

EEHea *The Thirteen Books of Euclid's Elements* (1908). *Translated from the text of Heiberg*. Trans. and comm., with an introd., by Thomas L. Heath. 3 vols. Cambridge: at the University Press.
Vol. I: *Introduction and Books I, II*.
Vol. II: *Books III–IX*.
Vol. III: *Books X–XIII and Appendix*.

EEHei Heiberg, J. L., éd. et trad. (1883-1888). *Euclidis Elementa.* 5 vol. Lipsiæ [Leipzig] : B. G. Teubner.
Vol. I : *Libros I–IV continens.* Lipsiæ [Leipzig] : B. G. Teubner. 1883.
Vol. II : *Libros V–IX continens.* 1884.
Vol. III : *Librum X continens.* 1886.
Vol. IV : *Libros XI–XIII continens.* 1885.
Vol. V : *Continens Elementorum qui peruntur libros XIV–XV et scholia in Elementa cum prolegomenis criticis et appedicibus.* 1888.

EEV Euclide d'Alexandrie (1990-2001). *Les Éléments. Traduits du texte de Heiberg.* Trad. et notes par Bernard Vitrac. Avec une introd. de Maurice Caveing. Paris : Presses Universitaires de France. (Bibliothèque d'histoire des sciences).
Vol. I : *Introduction générale. Livres I-IV : Géométrie plane.* 1990.
Vol. II : *Livres V-VI : Proportions et similitudes. Livres VII-IX : Arithmétique.* 1994.
Vol. III : *Livre X : Grandeurs commensurables et incommensurables, classification des lignes irrationnelles.* 1998.
Vol. IV : *Livres XI-XIII : Géométrie des solides.* 2001.

ELP Euler, Leonhard (1768-1772). *Lettres à une princesse d'Allemagne sur divers sujets de physique & de philosophie.* 3 vol. Saint Pétersbourg : Imprimerie de l'Académie Impériale des Sciences. Réimpr. in EOO, vol. III.11–12.
Vol. I : *[Lettres I–LXXIX].* 1768.
Vol. II : *[Lettres LXXX–CLIV].* 1768.
Vol. III : *[Lettres CLV–CCXXXIV].* 1772.

EOO Euler, Leonhard (1911-). *Opera Omnia.* Societas Scientiarum Naturalium Helvetica.
Vol. I.10 : *Institutiones Calculi Differentialis.* Éd. par Gerhard Kowalewski. Lipsiae [Leipzig] et Berolini [Berlin] : B.G. Teubner. 1913.
Vol. I.24 : *Methodus inveniendi […].* Éd. par Constantin Carathéodory. Turici [Zürich] : Orell Füssli. 1952.
Vol. III.11–12 : *Lettres à une princesse d'Allemagne. Rettung der göttlichen Offenbarung. Éloge d'Euler par le Marquis de Condorcet.* Éd. par Andreas Speiser. Turici [Zürich] : Orell Füssli. 1960.
Vol. IVA.5 : *Correspondance avec A. C. Clairaut, J. D'Alembert et J. L. Lagrange.* Éd. par Adolf P. Juškevič et René Taton. Avec la coll. de Charles Blanc, Ašot T. Grigorijan, Walter Habicht et Guy Picolet. Bâle : Birkhäuser. 1980.

LAA Leibniz, Gottfried Wilhelm (1923–). *Sämtliche Schriften und Briefe (Die Akademie-Ausgabe).* Hrsg. von den Berlin-Brandenburgischen und Göttinger Akademien der Wissenschaften. Berlin: Akademie Verlag.

Bd. III.4: *Mathematischer, naturwissenschaftlicher und technischer Briefwechsel. 4. Band: Juli 1683–1690.* Hrsg. von Heinz-Jürgen Hess, James G. O'Hara und Herbert Breger. 1995.

Bd. III.5: *Mathematischer, naturwissenschaftlicher und technischer Briefwechsel. 5. Band: 1691–1693.* Hrsg. von Heinz-Jürgen Hess und James G. O'Hara. 2003.

Bd. III.6: *Mathematischer, naturwissenschaftlicher und technischer Briefwechsel. 6. Band: 1694–Juni 1696.* Hrsg. von Heinz-Jürgen Hess und James G. O'Hara. 2004.

Bd. III.7: *Mathematischer, naturwissenschaftlicher und technischer Briefwechsel. 7. Band: Juli 1696–Dezember 1698.* Hrsg. von James G. O'Hara, Charlotte Wahl, Ralf Krömer und Heike Sefrin-Weis. 2011.

Bd. VI.4: *Philosophische Schriften. 4. Band: 1677–Juni 1690.* Hrsg. von Heinrich Schepers u. a. 1999.

Bd. VII.3: *Mathematische Schriften. 3. Band: 1672–1676. Differenzen, Folgen, Reihen.* Hrsg. von Siegmund Probst, Eberhard Knobloch und Nora Gädeke. 2003.

Bd. VII.5: *Mathematische Schriften. 5. Band: 1674–1676. Infinitesimalmathematik.* Hrsg. von Siegmund Probst, Uwe Mayer und Heike Sefrin-Weis. 2008.

LMS Leibniz, Gottfried Wilhelm (1849–1863). *Mathematische Schriften.* Hrsg. und mit einer Einl. vers. von Carl Immanuel Gerhardt. 7 Bde. Berlin: A. Ascher & Comp. (Bde. I-II); Halle: H. W. Schmidt (Bde. III-VII). (Réimp. en fac-similé : Hildesheim: Georg Olms, 1962, avec renum. des vol. : I-II rassemblés au vol. 1, III/1 au vol. 2, III/2 au vol. 3.)

Bd. I: *Briefwechsel zwischen Leibniz und Oldenburg, Collins, Newton, Galloys, Vitale Giordano.* 1849.

Bd. II: *Briefwechsel zwischen Leibniz, Hugens van Zulichem und dem Marquis de l'Hospital.* 1850.

Bd. III/1: *Briefwechsel zwischen Leibniz, Jacob Bernoulli und Nicolaus Bernoulli. [Jacob Bernoulli, Johann Bernoulli bis zum 7. Juni 1697].* 1855.

Bd. III/2: *Briefwechsel zwischen Leibniz, Jacob Bernoulli und Nicolaus Bernoulli. [Johann Bernoulli ab dem 15. Juni 1697, Nicolaus Bernoulli].* 1856.

Bd. IV: *Briefwechsel zwischen Leibniz, Wallis, Varignon, Guido Grandi, Zendrini, Hermann und Freiherrn von Tschirnhaus.* 1859.

Bd. V (o. 2. Abt., Bd. I): *[Dissertatio de Arte Combinatoria. De Quadratura Arithmeti-*

ca Circuli, Ellipseos et Hyperbolae. Characteristica Geometrica. Analysis Geometrica propria. Calculus situs. Analysis Infinitorum]. 1858.
Bd. VI (o. 2. Abt., Bd. II): *[Dynamica]*. 1860.
Bd. VII (o. 2. Abt., Bd. III): *[Initia mathematica. Mathesis universalis. Arithmetica. Algebraica. Geometrica]*. 1863.

LNC Leibniz, Gottfried Wilhelm (1989). *La naissance du calcul différentiel : 26 articles des « Acta Eruditorum »*. Trad., introd. et notes par Marc Parmentier. Avec une préf. de Michel Serres. Paris : Vrin.

LPS Leibniz, Gottfried Wilhelm (1875–1890). *Philosophische Schriften*. Hrsg. von Carl Immanuel Gerhardt. 7 Bde. Berlin: Weidmannsche Buchhandlung.
Bd. VI: *[Essais de théodicée. Philosophische Abhandlungen, 1702–1716]*. 1885.
Bd. VII: *[Scientia Generalis. Characteristica. Philosophische Abhandlungen. Streitschriften zwischen Leibniz und Clarke. Ergänzungen [...]]*. 1890.

NMP Newton, Isaac (1967–1981). *The Mathematical Papers of Isaac Newton*. Ed. by D. T. Whiteside. 8 vols. Cambridge: Cambridge University Press.
Vol. I: *1664–1666*. In collab. with Michael A. Hoskin. 1967.
Vol. VII: *1691–1695*. In collab. with Adolf Prag. 1976.

OL Lagrange, Joseph-Louis (1867-1892). *Œuvres*. Éd. par Joseph-Alfred Serret. 14 vol. Paris : Gauthier-Villars.
Vol. III : 1869.
Vol. VII : 1877.

PCP Peirce, Charles Sanders (1931–1958). *Collected Papers*. 8 vols. Cambridge, MA: The Belknap Press of Harvard University Press.
Vol. II: *Elements of Logic*. Ed. by Charles Hartshorne and Paul Weiss. 1931.
Vol. III: *Exact Logic*. Ed. by Charles Hartshorne and Paul Weiss. 1933.
Vol. IV: *The Simplest Mathematics*. Ed. by Charles Hartshorne and Paul Weiss. 1933.
Vol. VIII: *Reviews, Correspondence, and Bibliography*. Ed. by Arthur W. Burks. 1958.

PE Peirce, Charles Sanders (1992–1998). *The Essential Peirce. Selected Philosophical Writings.* 2 vols. Bloomington and Indianapolis: Indiana University Press.
Vol. 1: *1867–1893.* Ed. by Nathan Houser and Christian Kloesel. 1992.
Vol. 2: *1893–1913.* Ed. by Nathan Houser, André De Tienne, Jonathan R. Eller, et al. 1998.

Autres références

Aczel, Peter, Israel, David J., Katagiri, Yasuhiro, and Peters, Stanley, eds. (1993). *Situation Theory and Its Applications.* Vol. 3. Stanford: CSLI Publications. (CSLI Lecture Notes, 37).

Agnesi, Maria Gaetana (1748). *Instituzioni analitiche. Ad uso della gioventù italiana.* 2 voll. Milano: nella Regia-Ducal Corte.

Allwein, Gerard and Barwise, Jon, eds. (1996). *Logical Reasoning with Diagrams.* New York and Oxford: Oxford University Press. (Studies in Logic and Computation, 6).

Anderson, John R. (1978). Arguments Concerning Representations for Mental Imagery. *Psychological Review* 85.4, pp. 249–277.

Andler, Daniel (1990). Connexionnisme et cognition. À la recherche des bonnes questions. *Revue de synthèse.* Sér. IV 111.1–2 : *Sciences cognitives. Quelques aspects problématiques*, p. 95-127.

— (2004). Introduction. Calcul et représentation : les sources. In Andler, Daniel, éd. *Introduction aux sciences cognitives.* 2e éd. [Paris] : Gallimard, p. 13-50. (Folio Essais, 179).

Auchter, Heinrich (1937). *Brook Taylor, der Mathematiker und Philosoph. Beiträge zur Wissenschaftsgeschichte der Zeit des Newton-Leibniz-Streites.* Würzburg: K. Triltsch.

Austin, John L. (1950). Truth. *Aristotelian Society Supplementary Volume* 24.1, pp. 111–128.

Avigad, Jeremy, Dean, Edward, and Mumma, John (2009). A Formal System for Euclid's Elements. *The Review of Symbolic Logic* 2.4, pp. 700–768.

Barberousse, Anouk et Imbert, Cyrille (2013). Le tournant computationnel et l'innovation théorique. In Le Bihan, Soazig, éd. *Précis de philosophie de la physique.* Paris : Vuibert, p. 244-273. (Philosophie des sciences).

Barberousse, Anouk and Imbert, Cyrille (2014). Recurring Models and Sensitivity to Comptutational Constraints. *The Monist* 97.3: *Models and Simulations*, pp. 259–279.

Barker-Plummer, Dave, Barwise, Jon, and Etchemendy, John (2008). *Tarski's World: Revised and Expanded Edition.* In collab. with Albert Liu. Stanford: CSLI Publications. (CSLI Lecture Notes, 169).

Baron, Margaret E. (1969). A Note on the Historical Development of Logic Diagrams: Leibniz, Euler and Venn. *The Mathematical Gazette* 53.384, pp. 113–125.

Barwise, Jon (1981). Scenes and Other Situations. *The Journal of Philosophy* 78.7, pp. 369–397.

— (1983). Information and semantics. *Behavioral and brain sciences* 6 (1 1983), p. 65.

— (1989). *The Situation in Logic.* Stanford: Center for the Study of Language and Information. (CSLI lecture notes, 17).

— (1993). Constraints, Channels, and the Flow of Information. In Aczel, Peter, Israel, David J., Katagiri, Yasuhiro, and Peters, Stanley, eds. *Situation Theory and Its Applications.* Vol. 3. Stanford: CSLI Publications, pp. 3–27. (CSLI Lecture Notes, 37).

Barwise, Jon and Etchemendy, John (1987a). *Tarski's World.* Santa Barbara: Kinko's Academic Courseware Exchange.

— (1987b). *The Liar. An Essay on Truth and Circularity.* New York and Oxford: Oxford University Press.

— (1990a). Information, Infons, and Inference. In Cooper, Robin, Kuniaki, Mukai, and Perry, John, eds. *Situation Theory and Its Applications.* Vol. 1. Stanford: Center for the Study of Language and Information, pp. 33–78. (CSLI Lecture Notes, 22).

— (1990b). *The Language of First-Order Logic, Including the Program Tarski's World.* Stanford: Center for the Study of Language and Information. (CSLI Lecture Notes, 23).

— (1991). *Tarski's World.* Stanford: Center for the Study of Language and Information. (CSLI Lecture Notes, 25).

— (1994). *Hyperproof.* Stanford: CSLI Publications. (CSLI Lecture Notes, 42).

— [1995] (1996a). Heterogeneous Logic. In Allwein, Gerard and Barwise, Jon, eds. *Logical Reasoning with Diagrams.* New York and Oxford: Oxford University Press, pp. 179–200. (Studies in Logic and Computation, 6). Réimpr. de *Diagrammatic Reasoning.* Ed. by T. I. Glasgow, N. H. Narayanan, and B. Chandrasekaran. With a forew. by Herbert A. Simon. Menlo Park and Cambridge: AAAI Press and MIT Press, 1995, pp. 209–232 .

— [1991] (1996b). Visual Information and Valid Reasoning. In Allwein, Gerard and Barwise, Jon, eds. *Logical Reasoning with Diagrams.* New York and Oxford: Oxford University Press, pp. 3–25. (Studies in Logic and Computation, 6). Réimpr. de *Visualization in Teaching and Learning Mathematics.* Ed. by Walter Zimmermann and Steve Cunningham. Washington, D.C.: Mathematical Association of America, 1991, pp. 9–24. (MAA Notes, 19).

— (1998). Computers, Visualization, and the Nature of Reasoning. In Bynum, Terrell Ward and Moor, James H., eds. *The Digital Phoenix: How Computers Are Changing Philosophy.* Oxford and Malden, MA: Blackwell, pp. 93–116.

— (1999). *Language, Proof, and Logic.* In collab. with Gerard Allwein, Dave Barker-Plummer, and Albert Liu. Stanford: CSLI Publications.

Barwise, Jon, Gawron, Jean Mark, Plotkin, Gordon, and Tutiya, Syun, eds. (1991). *Situation Theory and Its Applications.* Vol. 2. Stanford: Center for the Study of Language and Information. (CSLI Lecture Notes, 26).

Barwise, Jon and Hammer, Eric M. [1994] (1996). Diagrams and the Concept of Logical System. In Allwein, Gerard and Barwise, Jon, eds. *Logical Reasoning with Diagrams.* New York and Oxford: Oxford University Press, pp. 49–78. (Studies in Logic and Computation, 6). Réimpr. de *What Is a Logical System?* Ed. by Dov M. Gabbay. Oxford: Clarendon Press, 1994, pp. 73–106. (Studies in Logic and Computation, 4).

Barwise, Jon and Perry, John (1980). The Situation Underground. In Barwise, Jon and Sag, Ivan A., eds. *Stanford Working Papers in Semantics.* Vol. 1. Stanford: Stanford Cognitive Science Group, pp. 1–55.

— (1981a). Semantic Innocence and Uncompromising Situations. *Midwest Studies in Philosophy* 6, pp. 387–404.

— (1981b). Situations and Attitudes. *The Journal of Philosophy* 78.11, pp. 668–691.

— (1983). *Situations and Attitudes.* Cambridge, Mass. and London: Bradford Books–MIT Press.

Barwise, Jon and Seligman, Jerry (1994). The Rights and Wrongs of Natural Regularity. *Philosophical Perspectives* 8: *Logic and Language*, pp. 331–364.

— (1997). *Information Flow. The Logic of Distributed Systems.* Cambridge and New York: Cambridge University Press. (Cambridge Tracts in Theoretical Computer Science, 44).

Barwise, Jon and Shimojima, Atsushi (1995). Surrogate Reasoning. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society* 2.4, pp. 7–26.

Bäuerle, Rainer and Cresswell, Max J. [1989] (2003). Propositional Attitudes. In Gabbay, Dov M. and Guenthner, Franz, eds. *Handbook of Philosophical Logic.* 2nd ed. Vol. 10. Dordrecht: Kluwer Academic Publishers, pp. 121–141. 1st ed. 1989.

Benis-Sinaceur, Hourya (1988a). *Ars inveniendi* et théorie des modèles. *Dialogue* XXVII.4, p. 591-613.

— (1988b). La logique comme *ars inveniendi*. In Robinet, André, éd. *Doctrines et concepts, 1937–1957. Cinquante ans de philosophie de langue française.* Actes du colloque pour le cinquantenaire de l'Association des sociétés de philosophie de langue française, Paris, 6–7 juillet 1987. Paris : Vrin, p. 319-334. (Problèmes et controverses).

— (1989). « *Ars inveniendi* » aujourd'hui. *Les Études Philosophiques* 20.2 : *Leibniz*, p. 201-214. [Sous le nom Sinaceur, H. B.]

— (2002). Modernité mathématique : Quelques invariants épistémologiques. *Revue d'Histoire des Sciences* 55.1, pp. 83–100. [Sous le nom Sinaceur, Hourya].

Bernoulli, Jacques (1690). Analysis problematis antehac propositis. De Inventione Lineæ descensus a corpore gravi percurrendæ uniformiter, sic ut temporibus æqualibus æquales altitudines emetiatur : et alterius cujusdam Problematis Propositio. *Acta Eruditorum* (mai 1690), p. 217-219. Réimpr. in Jacques Bernoulli 1744, vol. 1, p. 421–426.

— (1744). *Opera.* 2 vol. Genevæ : Hæredes Cramer & Fratri Philibert.

Bernoulli, Jean (1694). Additamentum effectionis omnium quadraturam et rectificationum curvarum per seriem quandam generalissimam. *Acta Eruditorum* (nov. 1694), p. 437-441. Réimpr. in BOO, I, p. 125–128.

Bertin, Jacques (1967). *Sémiologie graphique. Les diagrammes, les réseaux, les cartes.* La Haye et Paris : Mouton et Gauthier-Villars.

Block, Ned (1998). Semantics, Conceptual Role. In *Routledge Encyclopedia of Philosophy.* Taylor and Francis. URL: https://www.rep.routledge.com/articles/thematic/semantics-conceptual-role/v-1.

Borgato, Maria Teresa e Pepe, Luigi (1990). *Lagrange. Appunti per una biografia scientifica.* Torino: La Rosa Editrice. (Piemontesi illustri, 1).

Bos, Hendrik Jan Maarten (1974). Differentials, Higher-Order Differentials and the Derivative in the Leibnizian Calculus. *Archive for History of Exact Sciences* 14.1, pp. 1–90.

Brandom, Robert B. (1994). *Making It Explicit. Reasoning, Representing, and Discursive Commitment.* Cambridge, Mass. and London: Harvard University Press.

— (2000). *Articulating Reasons. An Introduction to Inferentialism.* Cambridge, Mass. and London: Harvard University Press.

Brechenmacher, Frédéric (2006). *Les matrices. Formes de représentations et pratiques opératoires (1850–1930).* CultureMATH – Site expert, ENS Ulm / DESCO. URL : http://culturemath.ens.fr/content/les-matrices-formes-de-repr%C3%A9sentation-et-pratiques-op%C3%A9ratoires-1850-1930.

Brooks, Rodney (1991). Intelligence without Representation. *Artificial Intelligence* 47.1–3: *Foundations of A.I.* Ed. by David Kirsh, pp. 139–159.

Brown, James Robert (2008). *Philosophy of Mathematics. A Contemporary Introduction to the World of Proofs and Pictures.* 2nd ed. New York and London: Routledge.

Bullynck, Maarten (2013). Erhard Weigel's Contributions to the Formation of Symbolic Logic. *History and Philosophy of Logic* 34.1, pp. 25–34.

Burkhard [Mencke], Johann (1721). Epistola ad Virum Clarissimum Broock Taylor. *Acta Eruditorum* (mai 1721), p. 195-228. Réimpr. in BOO, II, p. 483–512.

Cajori, Florian (1928–1929). *A History of Mathematical Notations.* 2 vols. Chicago and London: The Open Court.
Vol. I: *Notations in Elementary Mathematics.* 1928.
Vol. II: *Notations Mainly in Higher Mathematics.* 1929.

Calinger, Ronald S. (2015). *Leonhard Euler: Mathematical Genius in the Enlightenment.* Princeton and Oxford: Princeton University Press.

Carnap, Rudolf (1936–1937). Testability and Meaning. *Philosophy of Science* 3.4, p. 419–471; 4.1, p. 1–40.

— (1942). *Introduction to Semantics.* Cambridge, Mass.: Harvard University Press.

— (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic.* Chicago: University Chicago Press.

Carnap, Rudolf and Bar-Hillel, Yehoshua (1952). *An Outline of a Theory of Semantic Information.* Tech. rep. 247. Cambridge, Mass.: Research Laboratory of Electronics, MIT, Oct. 27, 1952. Réimpr. in Bar-Hillel, Yehoshua. *Language and information. Selected Essays on their Theory and Application.* Reading, Mass. and Jerusalem: Addison-Wesley and Jerusalem Academic Press, 1964, pp. 221–274. (Addison-Wesley Series in Logic).

Carter, Jessica (2018). Graph-Algebras—Faithful Representations and Mediating Objects in Mathematics. *Endeavour*. Pre-published.

Cartier, Pierre (2000). Mathemagics. In Planat, Michel, ed. *Noise, Oscillators and Algebraic Randomness. From Noise in Communication Systems to Number Theory.* Lectures of a School Held in Chapelle des Bois, France, April 5–10, 1999. Berlin and Heidelberg: Springer, pp. 6–67. (Lecture Notes in Physics, 550).

Chauviré, Christiane (2008). *L'Œil mathématique. Essai sur la philosophie mathématique de Peirce.* Paris: Kimé.

Chemla, Karine (2003). Generality above Abstraction. The General Expressed in Terms of the Paradigmatic in Mathematics in Ancient China. *Science in Context* 16.3, pp. 413–458.

— (2006). La généralité, valeur épistémologique fondamentale des mathématiques de la Chine ancienne. *Comptes-rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 150.4, p. 1927-1960.

— ed. (2012). *The History of Mathematical Proof in Ancient Traditions.* Cambridge and New York: Cambridge University Press.

Chemla, Karine, Chorlay, Renaud, and Rabouin, David, eds. (2016). *The Oxford Handbook of Generality in Mathematics and the Sciences.* Oxford: Oxford University Press.

Chernoff, Herman (1973). The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of the American Statistical Association* 68.342, pp. 361–368.

Clark, Andy and Chalmers, David (1998). The Extended Mind. *Analysis* 58.1, pp. 7–19.

Collins, John (1712). *Commercium Epistolicum D. Johannis Collins et aliorum de Analysi promota.* Londini: Pearson.

Colyvan, Mark (2012). *An Introduction to the Philosophy of Mathematics.* Cambridge University Press.

Cooper, Jacob Lionel Bakst (1952). Heaviside and the Operational Calculus. *The Mathematical Gazette* 36.315 (Feb. 1952), pp. 5–19.

Cooper, Robin, Kuniaki, Mukai, and Perry, John, eds. (1990). *Situation Theory and Its Applications.* Vol. 1. Stanford: Center for the Study of Language and Information. (CSLI Lecture Notes, 22).

Cooper, Robin and Poesio, Massimo (1994). Situation Semantics. In The FraCaS Consortium, ed. *Describing the Approaches.* FraCaS Deliverable D8. Edinburgh: Centre for Cognitive Science, University of Edinburgh, pp. 114–166.

Corcoran, John (1998). Information-Theoretic Logic. In Martínez, Concha, Rivas, Uxia, and Villegas-Forero, Luis, eds. *Truth in Perspective. Recent Issues in Logic, Representation and Ontology.* Aldershot: Ashgate Publishing, pp. 113–135. (Avebury Series in Philosophy).

Corry, Leo (1997). The Origins of Eternal Truth in Modern Mathematics. Hilbert to Bourbaki and Beyond. *Science in Context* 10.2, pp. 253–296.

Costabel, Pierre (1949). Deux inédits de la correspondance indirecte Leibniz-Reyneau. *Revue d'histoire des sciences et de leurs applications* 2.4, p. 311-332.

Couturat, Louis (1901). *La logique de Leibniz d'après des documents inédits.* Paris : Félix Alcan.

Cresswell, Max J. (1985). *Structured Meanings: The Logic of Propositional Attitudes.* Cambridge, Mass. and London: Bradford Books—MIT Press.

Crevier, Daniel (1993). *AI. The Tumultuous History of the Search for Artificial Intelligence.* New York: Basic Books.

Crowther-Heyck, Hunter (2005). *Herbert A. Simon. The Bounds of Reason in Modern America.* Baltimore and London: The Johns Hopkins University Press.

Dahan-Dalmedico, Amy (2001). An Image Conflict in Mathematics after 1945. In Bottazzini, Umberto and Dahan-Dalmedico, Amy, eds. *Changing Images in Mathematics. From the French Revolution to the New Millenium.* Routledge, pp. 223–254. (Studies in the History of Science, Technology and Medicine).

Dale, Nell and Walker, Henry Mackay (1996). *Abstract Data Types. Specifications, Implementations, and Applications.* Lexington, Mass.: D.C. Heath and Company.

Davis, Harold T. (1936). *The Theory of Linear Operators, from the Standpoint of Differential Equations of Infinite Order.* Bloomington, Indiana: The Principia Press.

De Risi, Vincenzo (2007). *Geometry and Monadology. Leibniz's* Analysis Situs *and Philosophy of Space.* Basel, Boston, and Berlin: Birkhäuser.

De Morgan, Augustus (1847). *Formal Logic. Or, The Calculus of Inference, Necessary and Probable.* London: Taylor and Walton.

De Toffoli, Silvia (2017). ‘Chasing’ the Diagram—The Use of Visualization in Algebraic Reasoning. *The Review of Symbolic Logic* 10.1, pp. 158–186.

De Toffoli, Silvia and Giardino, Valeria (2015). An Inquiry into the Practice of Proving in Low-Dimensional Topology. In Lolli, Gabriele, Panza, Marco, and Venturi, Giorgio, eds. *From Logic to Practice. Italian Studies in the Philosophy of Mathematics.* Cham: Springer, pp. 315–336. (Boston Studies in the Philosophy and History of Science, 308).

Delahaye, Jean-Paul (2000). *Merveilleux nombres premiers. Voyage au cœur de l'arithmétique.* Paris : Belin – Pour la science. (Bibliothèque scientifique).

Descartes, René (1637). *Discours de la Methode pour bien conduire sa raison, & chercher la verité dans les sciences. Plus la Dioptrique, les Météores et la Géométrie, qui sont des essais de cete Methode.* Leyde : Ian Maire.

— (1657-1667). *Lettres de M. Descartes.* Éd. par Claude Clerselier. 3 vol. Paris : Charles Angot.
Vol. I: *Où sont traittées les plus belles questions de la Morale, de la Physique, de la Médecine et des Mathématiques.* 1657.
Vol. II: *Où sont expliquées plusieurs belles difficultez touchant ses autres Ouvrages.* 1659.
Vol. III: *Où il répond à plusieurs difficultez qui luy ont esté proposées sur la Dioptrique, la Géométrie, et sur plusieurs autres sujets.* 1667.

Deschamps, Claude et Warusfel, André (2003). *Mathématiques tout-en-un, 1re année.* 2e éd. Paris : Dunod.

Devlin, Keith (1991). *Logic and Information.* Cambridge and New York: Cambridge University Press.

— (2004). Jon Barwise's Papers on Natural Language Semantics. *The Bulletin of Symbolic Logic* 10.1, pp. 54–85.

— (2006). Situation Theory and Situation Semantics. In Gabbay, Dov M. and Woods, John, eds. *Handbook of the History of Logic.* Vol. 7: *Logic and the Modalities in the Twentieth Century.* Amsterdam and Oxford: Elsevier—North-Holland, pp. 601–664.

Dowty, David R., Wall, Robert E., and Peters, Stanley (1981). *Introduction to Montague Semantics.* Dordrecht, Boston, and London: D. Reidel. (Synthese Language Library, 11).

Dugowson, Stéphane (1994). *Les différentielles métaphysiques. Histoire et philosophie de la généralisation de l'ordre de différentiation.* Thèse de doct. Villetaneuse : Université Paris XIII.

Dummett, Michael (1973). *Frege: Philosophy of Language.* New York: Harper & Row.

Dutilh Novaes, Catarina (2012). *Formal Languages in Logic. A Philosophical and Cognitive Analysis.* Cambridge and New York: Cambridge University Press.

Eckes, Christophe and Giardino, Valeria (2018). The Classificatory Function of Diagrams. Two Examples from Mathematics. In Chapman, Peter et al., eds. *Diagrammatic Representation and Inference.* 10th International Conference, Diagrams 2018. Cham: Springer, pp. 120–136. (Lecture Notes in Artificial Intelligence, 10871).

Edwards, A. W. F. (2006). An Eleventh-Century Venn Diagram. *Bulletin of the British Society for the History of Mathematics* 21.2, pp. 119–121.

Etchemendy, John (1990). *The Concept of Logical Consequence.* Cambridge, Mass. and London: Harvard University Press.

Euler, Leonhard (1736). *Mechanica. Sive motus scientia analytice exposita.* 2 vol. Petropoli [Saint-Pétersbourg] : Typographia Academiae Scientiarum.

— (1744). *Methodus inveniendi Lineas Curvas Maximi Minimive proprietate gaudentes. Sive solutio problematis isoperimetrici latissimo sensu accepti.* Lausannæ & Genevæ : Marcus-Michaelis Bousquet & Socii. Réimpr. in EOO, I.24.

— (1755). *Institutiones calculi differentialis. Cum eius usu in analysi finitorum ac doctrina serierum.* [Saint-Pétersbourg] : Academia imperialis scientiarum petropolitana. Réimpr. in EOO, I.10.

— (1843). *Lettres à une princesse d'Allemagne sur divers sujets de physique et de philosophie, précédées de l'éloge d'Euler par Condorcet.* Éd., introd. et notes par Émile Saisset. Paris : Charpentier. Rééd. de ELP.

— [1755] (2000). *Foundations of Differential Calculus.* Trans. by John D. Blanton. New York: Springer. Trad. de *Institutiones calculi differentialis. Cum eius usu in analysi finitorum ac doctrina serierum.* [Saint-Pétersbourg] : Academia imperialis scientiarum petropolitana, 1755.

Fagnano, Giulio Carlo de' Toschi di (1911–1912). *Opere Matematiche.* A cura di V. Volterra, G. Loria e D. Gambioli. 3 voll. Milano, Roma e Napoli: Società Editrice Dante Alighieri. Vol. III: *Altri scritti scientifici – Scritti polemici – Carteggio – Biografia.* 1912.

Feigenbaum, Lenore (1985). Brook Taylor and the Method of Increments. *Archive for History of Exact Sciences* 34.1–2, pp. 1–140.

— (1986). Leibniz and the Taylor Series. In Heinekamp, Albert, Hrsg. *300 Jahre „Nova Methodus" von G. W. Leibniz*. Symposion der Leibniz-Gesellschaft im Congresscentrum „Leewenhorst" in Noordwijkerhout (Niederlande), 28. bis 30. August 1984. Stuttgart: Franz Steiner Wiesbaden, S. 258–267. (Studia Leibnitiana Sonderheft, 14).

Ferraro, Giovanni (2008). *The Rise and Development of the Theory of Series Up to the Early 1820s*. New York: Springer. (Sources and Studies in the History of Mathematics and Physical Sciences).

Ferreirós, José (2016). *Mathematical Knowledge and the Interplay of Practices*. Princeton and Oxford: Princeton University Press.

Fleckenstein, J.O. (1946). Die Taylorsche Formel bei Johann I Bernoulli. *Elemente der Mathematik* 1.1, S. 13–17.

Flury, Bernhard and Riedwyl, Hans (1981). Graphical Representation of Multivariate Data by Means of Asymmetrical Faces. *Journal of the American Statistical Association* 76.376, pp. 757–765.

— [1983] (1988). *Multivariate Statistics. A Practical Approach*. London and New York: Chapman and Hall. Trad. de *Angewandte multivariate Statistik. Computergestützte Analyse Mehrdimensionaler Daten*. Stuttgart und New York: Gustav Fischer, 1983.

Fodor, Jerry A. (1975). *The Language of Thought*. New York: Thomas Y. Crowell. (The Language and Thought series).

Fox, Chris and Lappin, Shalom (2005). *Foundations of Intensional Semantics*. Malden, MA and Oxford: Blackwell.

Frege, Gottlob (1879). *Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle an der Saale: Louis Nebert.

— (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*. N. F. 100.1, S. 25–50. Trad. Sens et dénotation. In *Écrits logiques et philosophiques*. Trad. et introd. par Claude Imbert. Paris : Seuil, 1971, p. 102-126. (L'Ordre philosophique).

— (1999). *Idéographie*. Trad., préf. et notes par Corine Besson. Avec une postf. de Jonathan Barnes. Paris : Vrin.

Friedlein, Gottfried, éd. (1873). *Procli Diadochi in primum Euclidis Elementorum librum commentarii*. Lipsiae [Leipzig] : B. G. Teubner.

Friendly, Michael (2008). A Brief History of Data Visualization. In Chen, Chun-houh, Härdle, Wolfgang K., and Unwin, Antony, eds. *Handbook of Data Visualization*. Berlin and Heidelberg: Springer, pp. 15–56. (Handbooks of Computational Statistics, 3).

Gardner, Martin (1964). Mathematical Games. The Remarkable Lore of the Prime Numbers. *Scientific American* 210.3 (Mar. 1964), pp. 120–130. Réimpr. sous le titre Patterns and Primes. In *Martin Gardner's Sixth Book of Mathematical Diversions from* Scientific American. Chicago and London: University of Chicago Press, 1983, pp. 79–90.

Gawron, Jean Mark and Peters, Stanley (1990). *Anaphora and Quantification in Situation Semantics*. Stanford: CSLI Publications. (CSLI Lecture Notes, 19).

Gelernter, Herbert [1959] (1963). Realization of a Geometry-Theorem Proving Machine. In Feigenbaum, Edward A. and Feldman, Julian, eds. *Computers and Thought*. New York: McGraw-Hill, pp. 134–152. Réimpr. d'*Information Processing*. Paris, Munich, and London: UNESCO, Oldenbourg, and Butterworths, 1959, pp. 273–282 .

Giaquinto, Marcus (2015). The Epistemology of Visual Thinking in Mathematics. In Zalta, Edward N., ed. *The Stanford Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab, Stanford University, Win. 2015. URL: http://plato.stanford.edu/archives/win2015/entries/epistemology-visual-thinking/.

Giardino, Valeria (2013). A Practice-Based Approach to Diagrams. In Moktefi, Amirouche and Shin, Sun-Joo, eds. *Visual Reasoning with Diagrams*. Basel: Birkhäuser, pp. 135–151. (Studies in Universal Logic).

Giardino, Valeria and Greenberg, Gabriel (2014). Introduction: Varieties of Iconicity. *Review of Philosophy and Psychology* 6.1: *Pictorial and Diagrammatic Representation*. Ed. by Valeria Giardino and Gabriel Greenberg, pp. 1–25.

Giardino, Valeria et Hamami, Yacin (2019). Au-delà des preuves formelles. Philosophie de la pratique mathématique. In Arana, Andy, Panza, Marco, Poggiolesi, Francesca et Wagner, Pierre, éds. *Précis de philosophie des mathématiques et de la logique*. Paris : Éditions de la Sorbonne. À paraître.

Gibson, James J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Ginzburg, Jonathan (2011a). Situation Semantics and the Ontology of Natural Language. In Maienborn, Claudia, von Heusinger, Klaus, and Portner, Paul, eds. *Semantics: An*

International Handbook of Natural Language Meaning. Vol. 1. Berlin and Boston: De Gruyter Mouton, pp. 830–851. (Handbücher zur Sprach- und Kommunikationswissenschaft, 33.1).

Ginzburg, Jonathan (2011b). Situation Semantics: From Indexicality to Metacommunicative Interaction. In Maienborn, Claudia, von Heusinger, Klaus, and Portner, Paul, eds. *Semantics: An International Handbook of Natural Language Meaning.* Vol. 1. Berlin and Boston: De Gruyter Mouton, pp. 852–872. (Handbücher zur Sprach- und Kommunikationswissenschaft, 33.1).

Ginzburg, Jonathan and Sag, Ivan A. (2000). *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives.* Stanford: CSLI Publications. (CSLI Lecture Notes, 123).

Goodman, Nelson (1968). *Languages of Art. An Approach to a Theory of Symbols.* Indianapolis: Bobbs-Merrill.

Granger, Gilles-Gaston (1981). Philosophie et mathématique leibniziennes. *Revue de Métaphysique et de Morale* 86.1, pp. 1–37.

Grattan-Guinness, Ivor [1980] (2000). The Emergence of Mathematical Analysis and its Foundational Progress, 1780–1880. In Grattan-Guinness, Ivor, ed. *From the Calculus to Set Theory, 1630-1910. An Introductory History.* Princeton and Oxford: Princeton University Press, pp. 94–148. 1st pub. London: Duckworth, 1980.

Gray, Jeremy (2001). Symbols and Suggestions: Communication of Mathematics in Print. *The Mathematical Intelligencer* 23.2, pp. 59–64.

— (2007). *Worlds Out of Nothing. A Course in the History of Geometry in the 19th Century.* London: Springer. (Springer Undergraduate Mathematics Series).

— (2008). *Plato's Ghost. The Modernist Transformation of Mathematics.* Princeton and Oxford: Princeton University Press.

— (2009). Modernism in Mathematics. In Robson, Eleanor and Stedall, Jacqueline, eds. *The Oxford Handbook of the History of Mathematics.* Oxford and New York: Oxford University Press, pp. 663–683.

Greenberg, Mark and Harman, Gilbert (2008). Conceptual Role Semantics. In Lepore, Ernest and Smith, Barry C., eds. *The Oxford Handbook of Philosophy of Language.* Oxford and New York: Oxford University Press, pp. 295–322.

Guedj, Denis (1981). Claude Chevalley. Un des fondateurs de Bourbaki. *Dédales* 1 (nov. 1981) : *Les jeux et enjeux demain tenant*. Éd. par Denis Guedj, p. 23-24. Trad. Nicholas Bourbaki, Collective Mathematician. An Interview with Claude Chevalley. Trans. by Jeremy Gray. *The Mathematical Intelligencer* 7.2 (1985), pp. 18–22.

Guicciardini, Niccolò (1999). *Reading the Principia. The Debate on Newton's Mathematical Methods for Natural Philosophy from 1687 to 1736*. Cambridge and New York: Cambridge University Press.

Guisnée, Nicolas (1705). *Application de l'Algebre a la Geometrie. Ou Methode de Démontrer par l'Algebre, les Theorêmes de Geometrie, & d'en résoudre & construire tous les Problêmes*. Paris : Jean Boudot et Jacque Quillau.

Haffner, Emmylou (2014). *The "Science of Numbers" in Action in Richard Dedekind's Works. Between Mathematical Explorations and Foundational Investigations*. PhD thesis. Paris: Université Paris VII Diderot. HAL: tel-01144626.

Hammer, Eric M. (1994). Reasoning with Sentences and Diagrams. *Notre Dame Journal of Formal Logic* 35.1, pp. 73–87.

— (1995). *Logic and Visual Information*. Stanford: CSLI Publications and FoLLI. (Studies in Logic, Language and Information).

Hammer, Eric M. and Danner, Norman (1996). Towards a Model Theory of Diagrams. *Journal of Philosophical Logic* 25.5, pp. 463–482. Repr. in Allwein and Barwise 1996, pp. 109–128.

Hammer, Eric M. and Shin, Sun-Joo (1996). Euler and the Role of Visualization in Logic. In Seligman, Jerry and Westerståhl, Dag, eds. *Logic, Language and Computation*. Vol. 1. Stanford: CSLI Publications, pp. 271–286. (CSLI Lecture Notes, 58).

— (1998). Euler's Visual Logic. *History and Philosophy of Logic* 19.1, pp. 1–29.

Hatcher, Allen (2002). *Algebraic Topology*. Cambridge: Cambridge University Press.

Haugeland, John (1978). The Nature and Plausibility of Cognitivism. *Behavioral and Brain Sciences* 1.2, pp. 215–260.

— ed. (1981a). *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Cambridge, Mass. and London: Bradford Books—MIT Press.

— (1981b). Semantic Engines. An Introduction to Mind Design. In Haugeland, John, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Cambridge, Mass. and London: Bradford Books—MIT Press, pp. 1–34.

Haugeland, John (1985). *Artificial Intelligence: The Very Idea.* Cambridge, Mass. and London: Bradford Books—MIT Press.

— [1981] (1998a). Analog and Analog. In *Having Thought. Essays in the metaphysics of mind.* Cambridge, Mass. and London: Harvard University Press, pp. 75–88. Réimpr. *Philosophical Topics* 12. 1981, pp. 213–225.

— [1995] (1998b). Mind Embodied and Embedded. In *Having Thought. Essays in the metaphysics of mind.* Cambridge, Mass. and London: Harvard University Press, pp. 207–237. Réimpr. *Acta Philosophica Fennica* 58: *Mind and Cognition. Philosophical Perspectives on Cognitive Science and Artificial Intelligence*, 1995. Ed. by Leila Haaparanta and Sara Heinämaa, pp. 233–267.

— [1991] (1998c). Representational Genera. In *Having Thought. Essays in the metaphysics of mind.* Cambridge, Mass. and London: Harvard University Press, pp. 171–206. Réimpr. de *Philosophy and Connectionist Theory.* Ed. by William Ramsey, Stephen Stich, and David Rumelhart. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1991, pp. 61–89 .

Heath, Thomas L., ed. and introd. (1896). *Apollonius of Perga,* Treatise on Conic Sections. *Edited in modern notation with introductions including an essay on the earlier history of the subject.* Cambridge: University Press.

— (1913). *Aristarchus of Samos, the Ancient Copernicus. A History of Greek Astronomy to Aristarchus Together with Aristarchus's Treatise on the Sizes and Distances of the Sun and Moon.* Ed. and trans. by Thomas L. Heath. Oxford: Clarendon Press.

Heiberg, J. L. (1883). Praefatio. In *Euclidis Elementa.* Vol. I : *Libros I–IV continens.* Éd. et trad. par J. L. Heiberg. 5 vol. Lipsiæ [Leipzig] : B. G. Teubner, p. V-X.

— éd. et trad. (1891-1893). *Apollonii Pergaei quae Graece exstant. Cum commentariis antiquis.* 2 vol. Lipsiae [Leipzig] : B. G. Teubner.

Hilbert, David (1899). Grundlagen der Geometrie. In *Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen.* Leipzig: B. G. Teubner.

— (2004). *Lectures on the Foundations of Geometry, 1891–1902.* Ed. by Michael Hallett and Ulrich Majer. Berlin and Heidelberg: Springer. (David Hilbert's Lectures on the Foundations of Mathematics and Physics, 1891–1933, 1).

Hilbert, David und Cohn-Vossen, Stephan (1932). *Auschauliche Geometrie.* Berlin: Julius Springer. (Die Grundlehren der Mathematischen Wissenschaften, 37).

Hoare, C. A. R. (1972). Proof of Correctness of Data Representations. *Acta Informatica* 1.4, pp. 271–281.

Hookway, Christopher (1985). *Peirce*. London and Boston: Routledge & Kegan Paul. (The Arguments of the Philosophers).

Howse, John, Molina, Fernando, Shin, Sun-Joo, and Taylor, John (2002). On Diagram Tokens and Types. In Hegarty, Mary, Meyer, Bernd, and Narayanan, N. Hari, eds. *Diagrammatic Representation and Inference*. Second International Conference, Diagrams 2002. Berlin and Heidelberg: Springer, pp. 146–160. (Lecture Notes in Artificial Intelligence, 2317).

Høyrup, Jens (1990). Algebra and Naive Geometry. An Investigation of Some Basic Aspects of Old Babylonian Mathematical Thought. *Altorientalische Forschungen* 17.1–2, pp. 27–69, 262–354.

— (2002). *Lengths, Widths, Surfaces. A Portrait of Old Babylonian Algebra and Its Kin*. New York: Springer.

Humphreys, Paul (2004). *Extending Ourselves. Computational Science, Empiricism, and Scientific Method*. Oxford and New York: Oxford University Press.

Hutchins, Edwin (1995). *Cognition in the Wild*. Cambridge, Mass. and London: Bradford Books—MIT Press.

— (2005). Material Anchors for Conceptual Blends. *Journal of Pragmatics* 37, pp. 1555–1577.

Imbert, Claude (1971). Introduction. In Frege, Gottlob. *Écrits logiques et philosophiques*. Trad. et introd. par Claude Imbert. Paris : Seuil, p. 11-59. (L'Ordre philosophique).

Israel, David J. and Perry, John (1990). What Is Information? In Hanson, Philip P., ed. *Information, Language and Cognition*. Vancouver: University of British Columbia Press, pp. 1–19. (Vancouver Studies in Cognitive Science, 1).

Johansen, Mikkel Willum (2010). *Naturalism in the Philosophy of Mathematics*. PhD thesis. Copenhagen, Denmark: University of Copenhagen. URL: http://www.nbi.ku.dk/english/research/phd_theses/phd_theses_2011/mikkel_willum_johansen/.

— (2014). What's in a Diagram? On the Classification of Symbols, Figures and Diagrams. In Magnani, Lorenzo, ed. *Model-Based Reasoning in Science and Technology*. Berlin and Heidelberg: Springer, pp. 89–108. (Studies in Applied Philosophy, Epistemology and Rational Ethics, 8).

Johansen, Mikkel Willum, Misfeldt, Morten, and Pallavicini, Josephine (2018). A Typology of Mathematical Diagrams. In Chapman, Peter et al., eds. *Diagrammatic Representation and Inference.* 10th International Conference, Diagrams 2018. Cham: Springer, pp. 105–119. (Lecture Notes in Artificial Intelligence, 10871).

Kempson, Ruth (2011). Formal Semantics and Representationalism. In Maienborn, Claudia, von Heusinger, Klaus, and Portner, Paul, eds. *Semantics: An International Handbook of Natural Language Meaning.* Vol. 1. Berlin and Boston: De Gruyter Mouton, pp. 216–241. (Handbücher zur Sprach- und Kommunikationswissenschaft, 33.1).

King, Jeffrey C., Soames, Scott, and Speaks, Jeff (2014). *New Thinking about Propositions.* Oxford and New York: Oxford University Press.

Kirsh, David (1990). When Is Information Explicitly Represented? In Hanson, Philip P., ed. *Information, Language and Cognition.* Vancouver: University of British Columbia Press, pp. 340–365. (Vancouver Studies in Cognitive Science, 1).

Kirsh, David and Maglio, Paul (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science* 18.4, pp. 513–549.

Knobloch, Eberhard (1973). *Die Mathematischen Studien von G. W. Leibniz zur Kombinatorik. Auf Grund fast ausschliesslich handschriftlicher Aufzeichnungen dargelegt und kommentiert.* Wiesbaden: F. Steiner. (Studia Leibnitiana, 11).

— (2002). Le calcul leibnizien dans la correspondance entre Leibniz et Jean Bernoulli. In Abel, Günter, Engfer, Hans-Jürgen und Hubig, Christoph, Hrsg. *Neuzeitliches Denken: Festschrift für Hans Poser zum 65. Geburtstag.* Berlin und New York: Walter de Gruyter, S. 173–193.

— (2016). Generality in Leibniz's Mathematics. In Chemla, Karine, Chorlay, Renaud, and Rabouin, David, eds. *The Oxford Handbook of Generality in Mathematics and the Sciences.* Oxford: Oxford University Press, pp. 90–109.

Knuth, Donald E. [1968] (1997). *The Art of Computer Programming.* Vol. 1: *Fundamental Algorithms.* 3rd ed. Reading, Mass.: Addison Wesley. 1st ed. 1968.

Koppelman, Elaine (1971). The Calculus of Operations and the Rise of Abstract Algebra. *Archive for History of Exact Sciences* 8.3, pp. 155–242.

Kosslyn, Stephen M. (1980). *Image and Mind.* Harvard University Press: Cambridge, Mass. and London.

Kosslyn, Stephen M., Ganis, Giorgio, and Thompson, William L. (2001). Neural Foundations of Imagery. *Nature Reviews in Neuroscience* 2, pp. 635–642.

Kosslyn, Stephen M. and Pomerantz, James R. (1977). Imagery, Propositions, and the Form of Internal Represenations. *Cognitive Psychology* 9.1, pp. 52–76.

Kratzer, Angelika (2014). Situations in Natural Language Semantics. In Zalta, Edward N., ed. *The Stanford Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab, Stanford University, Spr. 2014. URL: https://plato.stanford.edu/archives/spr2014/entries/situations-semantics/.

Kulvicki, John (2006). *On Images. Their Structure and Content*. Oxford and New York: Clarendon Press.

— (2010). Knowing with Images: Medium and Message. *Philosophy of Science* 77.2, pp. 295–313.

Kutzler, B. and Lichtenberger, F. (1983). *Bibliography on Abstract Data Types*. Berlin and Heidelberg: Springer-Verlag. (Informatik-Fachberichte, 68).

L'Hôpital, Guillaume de (1696). *Analyse des infiniment petits pour l'intelligence des lignes courbes*. Paris : Imprimerie Royale.

Lagrange, Joseph-Louis (1754). *Lettera di Luigi de la Grange Tournier, torinese, all'illustrissimo signor conte Giulio Carlo da Fagnano [...] contenente una nuova serie per i differenziali ed integrali di qualsivoglia grado, corrispondente alla Newtoniana per le potestà e le radici*. Torino: Stamperia reale. Rist. in OL, VII, pp. 583–588; Fagnano 1911–1912, III, pp. 181–185.

— (1774). Sur une nouvelle espèce de calcul relatif à la différentiation et à l'intégration des quantités variables. *Nouveaux mémoires de l'Académie des Sciences et Belles-Lettres de Berlin* : *Année 1772*, p. 185-221. Réimpr. in OL, III, p. 441–476.

Lang, Serge (2002). *Algebra*. 3rd ed. New York, Berlin, and Heidelberg: Springer. (Graduate Texts in Mathematics, 211).

Lange, Johann Christian (1712). *Nucleus Logicae Weisianae*. Gissae-Hassorum [Giessen]: Henning Müller.

Larkin, Jill H. (1989). Display-Based Problem Solving. In Klahr, David and Kotovsky, Kenneth, eds. *Complex Information Processing. The Impact of Herbert A. Simon*. 21st

Carnegie-Mellon Symposium on Cognition. Hillsdale, New Jersey: Lawrence Erlbaum Associates, pp. 319–341.

Larkin, Jill H. and Simon, Herbert A. (1987). Why a Diagram Is (Sometimes) Worth Ten Thousand Words. *Cognitive science* 11.1, pp. 65–100.

Larvor, Brendan (1996). Review of Sun-Joo Shin, *The Logical Status of Diagrams*. *International Studies in the Philosophy of Science* 10.2, pp. 177–179.

— (2010). Syntactic Analogies and Impossible Extensions. In Löwe, Benedikt and Müller, Thomas, eds. *Philosophy of Mathematics : Sociological Aspects and Mathematical Practice*. London: College Publications, pp. 197–208. (Texts in Philosophy, 11).

Lee, Meng and Stepanov, Alexander (1994). Science of C++ Programming. (Slides of a presentation, Hewlett-Packard Laboratories). Jan. 1994. URL: http://stepanovpapers.com/Stepanov-Science_of_C++_Programming-1994.pdf.

Leibniz, Gottfried Wilhelm (1684a). De dimensionibus figurarum inveniendis. *Acta Eruditorum* (mai 1684), p. 233-236. Réimpr. in LMS, V, p. 123–126. Trad. in LNC, p. 82-92.

— (1684b). Nova methodus pro maximis et minimis, itemque tangentibus, quæ nec fractas, nec irrationales quantitates moratur, et singulare pro illis calculi genus. *Acta Eruditorum* (oct. 1684), p. 467-473. Réimpr. in LMS, V, p. 220–226. Trad. in LNC, p. 96-117.

— (1686). De geometria recondita et analysi indivisibilium atque infinitorum. *Acta Eruditorum* (juin 1686), p. 292-300. Réimpr. in LMS, V, p. 226–233. Trad. in LNC, p. 126-143.

— (1693). Supplementum geometriæ practicæ sese ad problemata transcendentia extendens, ope novæ methodi generalissimæ per series infinitas. *Acta Eruditorum* (avr. 1693), p. 178-180. Réimpr. in LMS, V, p. 285–288. Trad. in LNC, p. 236-246.

— (1694a). Considerations sur la différence qu'il y a entre l'analyse ordinaire et le nouveau calcul des transcendantes. *Journal des Sçavans* (août 1694), p. 404-406. Réimpr. in LMS, V, p. 306–308.

— (1694b). Constructio propria problematis de curva isochrona paracentrica. *Acta Eruditorum* (août 1694), p. 364-375. Réimpr. in LMS, V, p. 309–318. Trad. in LNC, p. 282-305.

— (1695). Responsio ad nonnullas difficultates a Dn. Bernardo Nieuwentijt circa methodum differentialem seu infinitesimalem motas. *Acta Eruditorum* (juil. 1695), p. 310-316. Réimpr. in LMS, V, p. 320–326. Trad. in LNC, p. 316-334.

— (1702). Specimen novum analyseos pro scientia infiniti circa summas et quadraturas. *Acta Eruditorum* (mai 1702), p. 210-219. Réimpr. in LMS, V, p. 350–361. Trad. in LNC, p. 381-401.

— (1710a). Monitum de Characteribus Algebraicis. *Miscellanea Berolinensia* 1, p. 155-160. Réimpr. in LMS, VII, p. 218–223.

— (1710b). Symbolismus memorabilis calculi algebraici & infinitesimalis, in comparatione potentiarum & differentiarum; & de lege homogeneorum transcendentali. *Miscellanea Berolinensia* 1, p. 160-165. Réimpr. in LMS, V, p. 377–382. Trad. in LNC, p. 414-421.

— (1846). *Historia et origo calculi differentialis*. Hrsg. und mit einer Einl. vers. von Carl Immanuel Gerhardt. Hannover: Hahn.

— (1899). *Der Briefwechsel von Gottfried Wilhelm Leibniz mit Mathematikern*. Hrsg. und mit einer Einl. vers. von Carl Immanuel Gerhardt. Berlin: Mayer & Müller.

— (1903). *Opuscules et fragments inédits de Leibniz : Extraits des manuscrits de la Bibliothèque royale de Hanovre*. Éd., préf. et notes par Louis Couturat. Paris : F. Alcan.

— (1976). *Die Mathematischen Studien von G. W. Leibniz zur Kombinatorik*. Textband im Anschluß an den gleichnamigen Abhandlungsband zum ersten Mal nach den Originalhandschriften herausgegeben. Hrsg. und mit einer Einl. vers. von Eberhard Knobloch. Wiesbaden: F. Steiner. (Studia Leibnitiana, 16).

— (2011). *Die mathematischen Zeitschriftenartikel*. Übers. und komm. von Heinz-Jürgen Hess und Malte-Ludolf Babin. Hildesheim, Zürich und New York: Georg Olms.

Leibniz, Gottfried Wilhelm et Bernoulli, Jean (1745). *Commercium philosophicum et mathematicum*. Vol. 1. Lausanæ et Genevæ : M.-M. Bousquet & soc.

Lewis, David (1970). General Semantics. *Synthese* 22.1-2, pp. 18–67.

— (1971). Analog and Digital. *Noûs* 5.3, pp. 321–327.

Liskov, Barbara and Zilles, Stephen (1974). Programming with Abstract Data Types. *ACM SIGPLAN Notices* 9.4, pp. 50–59.

Lubet, Jean-Pierre (2010). Calcul symbolique et calcul intégral de Lagrange à Cauchy. *Revue d'Histoire des Mathématiques* 16.1, p. 63-131.

Luengo, Isabel (1995). *Diagrams in Geometry*. PhD thesis. Bloomington, In.: Indiana University.

— (1996). A Diagrammatic Subsystem of Hilbert's Geometry. In Allwein, Gerard and Barwise, Jon, eds. *Logical Reasoning with Diagrams*. New York and Oxford: Oxford University Press, pp. 149–176. (Studies in Logic and Computation, 6).

MacFarlane, John Gordon (2000). *What Does It Mean to Say That Logic Is Formal?* PhD thesis. Pittsburgh: University of Pittsburgh. URL: https://www.johnmacfarlane.net/dissertation.pdf.

MacKenzie, Donald (2001). *Mechanizing Proof. Computing, Risk, and Trust.* Cambridge, Mass. and London: MIT Press. (Inside Technology).

Mancosu, Paolo (2005). Visualization in Logic and Mathematics. In Mancosu, Paolo, Jørgensen, Klaus Frovin, and Pedersen, Stig Andur, eds. *Visualization, Explanation and Reasoning Styles in Mathematics.* Springer, pp. 13–30. (Synthese Library, 327).

— (2008a). Introduction. In Mancosu, Paolo, ed. *The Philosophy of Mathematical Practice.* Oxford and New York: Oxford University Press, pp. 1–21.

— ed. (2008b). *The Philosophy of Mathematical Practice.* Oxford and New York: Oxford University Press.

Manders, Kenneth (1996). Diagram Contents and Representational Granularity. In Seligman, Jerry and Westerståhl, Dag, eds. *Logic, Language and Computation.* Vol. 1. Stanford: CSLI Publications, pp. 389–404. (CSLI Lecture Notes, 58).

— (1999). Euclid or Descartes? Representation and Responsiveness. (Unpublished but widely circulated draft). Aug. 1999.

— (2008a). Diagram-Based Geometric Practice. In Mancosu, Paolo, ed. *The Philosophy of Mathematical Practice.* Oxford and New York: Oxford University Press, pp. 65–79.

— (2008b). The Euclidean Diagram (1995). In Mancosu, Paolo, ed. *The Philosophy of Mathematical Practice.* Oxford and New York: Oxford University Press, pp. 80–133. Draft first circulated in 1995.

— (2012). Expressive Means and Mathematical Understanding. (Unpublished draft). May 2012.

— (2017). Expressive Usage and Intelligibility, Preliminary Outline. (Syllabus of a graduate course at the University of Pittsburgh). Spr. 2017.

Markowsky, George (2017). Information Theory. In *Encyclopædia Britannica.* June 16, 2017. URL: https://www.britannica.com/science/information-theory.

Maronne, Sébastien (2007). *La théorie des courbes et des équations dans la Géométrie cartésienne : 1637–1661.* Thèse de doct. Paris : Université Paris 7 Diderot. HAL : tel-00203094.

Martinez, Maricarmen and Sequoiah-Grayson, Sebastian (2018). Logic and Information. In Zalta, Edward N., ed. *The Stanford Encyclopedia of Philosophy.* Stanford: Metaphysics Research Lab, Stanford University, Sum. 2018. URL: https://plato.stanford.edu/archives/sum2018/entries/logic-information.

Maurice, Jean Frédéric Théodore (1814). Lettre à M. le Redacteur du *Moniteur Universel,* sur l'Eloge de Lagrange, par M. Delambre […] suivie de quelques remarques, et d'un supplément à cet Eloge. *Moniteur Universel* 54 (26 fév. 1814), p. 226-228. URL : http://diginole.lib.fsu.edu/islandora/object/fsu:449329.

McCorduck, Pamela (2004). *Machines Who Think. A Personal Inquiry into the History and Prospects of Artificial Intelligence.* 2nd ed. Natick, MA: A K Peters. 1st ed. New York: W. H. Freeman, 1979.

Mercator, Nicolaus (1668). *Logarithmotechnia: sive methodus construendi logarithmos nova, accurata et facilis.* Londini : G. Godbid & M. Pitt.

Miller, Nathaniel (2001). *A Diagrammatic Formal System for Euclidean Geometry.* PhD thesis. Ithaca: Cornell University. URL: http://hopper.unco.edu/faculty/personal/miller/diagrams/thesis.pdf.

— (2006). A Brief Proof of the Full Completeness of Shin's Venn Diagram Proof System. *Journal of Philosophical Logic* 35.3, pp. 289–291.

— (2007). *Euclid and His Twentieth Century Rivals. Diagrams in the Logic of Euclidean Geometry.* Stanford: CSLI Publications. (Studies in the Theory and Applications of Diagrams).

— (2012). On the Inconsistency of Mumma's Eu. *Notre Dame Journal of Formal Logic* 53.1, pp. 27–52.

Mirowski, Philip (2002). *Machine Dreams. Economics Becomes a Cyborg Science.* Cambridge and New York: Cambridge University Press.

Mitchell, W. J. Thomas (1986). *Iconology: Image, Text, Ideology.* Chicago and London: The University of Chicago Press.

Moktefi, Amirouche and Shin, Sun-Joo (2012). A History of Logic Diagrams. In Gabbay, Dov M., Pelletier, Francis Jeffry, and Woods, John, eds. *Handbook of the History of Logic.* Vol. 11: *Logic: A History of its Central Concepts.* Amsterdam and Boston: Elsevier—North-Holland, pp. 611–682.

Montague, Richard (1970). English as a Formal Language. In Visentini, Bruno, ed. *Linguaggi nella Società e nella Tecnica.* Milano: Edizioni di Comunità, pp. 189–224.

— [1970] (1974a). English as a Formal Language. In *Formal Philosophy: Selected Papers of Richard Montague.* Ed., with an introd., by Richmond H. Thomason. New Haven and London: Yale University Press, pp. 188–221. Réimpr. de *Linguaggi nella Società e nella Tecnica.* Ed. by Bruno Visentini. Milano: Edizioni di Comunità, 1970, pp. 189–224 .

Montague, Richard (1974b). *Formal Philosophy: Selected Papers of Richard Montague.* Ed., with an introd., by Richmond H. Thomason. New Haven and London: Yale University Press.

— [1973] (1974c). The Proper Treatment of Quantification in Ordinary English. In *Formal Philosophy: Selected Papers of Richard Montague.* Ed., with an introd., by Richmond H. Thomason. New Haven and London: Yale University Press, pp. 247–270. Réimpr. d'*Approaches to Natural Language.* Ed. by Jaakko Hintikka, Julius M. E. Moravcsik, and Patrick Suppes. Dordrecht: D. Reidel, 1973, pp. 221–242. (Synthese Library).

Morris, Charles W. (1937). Logical Positivism, Pragmatism and Scientific Empiricism. *Actualités scientifiques et industrielles* 449: *Exposés de philosophie scientifique.*

Mueller, Ian (1981). *Philosophy of Mathematics and Deductive Structure in Euclid's* Elements. Cambridge, Mass. and London: MIT Press.

Mumford, David, Series, Caroline, and Wright, David (2002). *Indra's Pearls. The Vision of Felix Klein.* Cambridge: Cambridge University Press.

Mumma, John (2006). *Intuition Formalized. Ancient and Modern Methods of Proof in Elementary Geometry.* PhD thesis. Pittsburgh: Carnegie Mellon University. URL: http://johnmumma.org/Writings_files/Thesis.pdf.

— (2008). Review of Nathaniel Miller, *Euclid and His Twentieth Century Rivals: Diagrams in the Logic of Euclidean Geometry. Philosophia Mathematica.* 3rd ser. 16.2, pp. 256–281.

— (2010). Proofs, Pictures, and Euclid. *Synthese* 175.2, pp. 255–287.

— (2014). Technical Notions of Eu. (Unpublished draft, available online). Aug. 3, 2014. URL: http://johnmumma.org/Writings_files/Technical%20notions%20of%20Eu.pdf.

Netz, Reviel (1999). *The Shaping of Deduction in Greek Mathematics. A Study in Cognitive History.* Cambridge: Cambridge University Press. (Ideas in Context, 51).

Newell, Allen, Shaw, J. Clifford, and Simon, Herbert A. [1958] (1963). Chess Playing Programs and the Problem of Complexity. In Feigenbaum, Edward A. and Feldman, Julian, eds. *Computers and Thought.* New York: McGraw-Hill, pp. 39–70. Réimpr. *IBM Journal of Research and Development* 2.4. 1958, pp. 320–335.

Newell, Allen and Simon, Herbert A. (1956). The Logic Theory Machine. A Complex Information Processing System. *IRE Transactions on Information Theory* 2.3, pp. 61–79.

— [1957] (1963). Empirical Explorations with the Logic Theory Machine. In Feigenbaum, Edward A. and Feldman, Julian, eds. *Computers and Thought.* New York: McGraw-Hill, pp. 109–133. Réimpr. de *Proceedings of the Western Joint Computer Conference.* New York: Institute of Radio Engineers, 1957, pp. 218–230 .

— (1972). *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.

— (1976). Computer Science as Empirical Inquiry. Symbols and Search. *Communications of the ACM* 19.3, pp. 113–126.

Newton, Isaac (1687). *Philosophiæ Naturalis Principia Mathematica.* Londini : J. Streater.

— (1704). *Opticks. Or, a Treatise of the Reflexions, Refractions, Inflexions and Colours of Light. Also, Two Treatises of the Species and Magnitude of Curvilinear Figures.* London: Sam. Smith and Benj. Walford.

— (1759). *Principes mathématiques de la philosophie naturelle.* Trad. et comm. par Émilie du Châtelet. Paris : Desaint & Saillant et Lambert.

Palsky, Gilles (2012). Map Design *vs.* sémiologie graphique. Réflexions sur deux courants de la cartographie théorique. *Cartes & géomatique (revue du Comité Français de Cartographie)* 212, p. 7-12.

— (2017). *La Sémiologie graphique de Jacques Bertin a cinquante ans !* Visionscarto. 7 juin 2017. URL : https://visionscarto.net/la-semiologie-graphique-a-50-ans.

Palsky, Gilles et Robic, Marie-Claire (2000). Aux sources de la sémiologie graphique. *Cybergeo (European Journal of Geography)* : *Dossier « 30 ans de sémiologie graphique ».* URL : http://journals.openedition.org/cybergeo/554.

Panza, Marco (1992). *La forma della quantità.* 2 vol. Paris : Société Française d'Histoire des Sciences et des Techniques. (Cahiers d'histoire et de philosophie des sciences, 38–39).

— (2005). *Newton et les origines de l'analyse, 1664–1666.* Paris : A. Blanchard.

— (2012). The Twofold Role of Diagrams in Euclid's Plane Geometry. *Synthese* 186.1: *Diagrams in Mathematics: History and Philosophy.* Ed. by John Mumma, Marco Panza, and Gabriel Sandu, pp. 55–102.

Parnas, David L. (1972a). A Technique for Software Module Specification with Examples. *Communications of the ACM* 15.5, pp. 330–336.

— (1972b). On the Criteria to Be Used in Decomposing Systems into Modules. *Communications of the ACM* 5.12, pp. 1053–1058.

— (2002). The Secret History of Information Hiding. In Broy, Manfred and Denert, Ernst, eds. *Software Pioneers. Contributions to Software Engineering.* Berlin and Heidelberg: Springer, pp. 399–409.

Partee, Barbara H. (1979). Semantics — Mathematics or Psychology? In Bäuerle, Rainer, Egli, Urs, and von Stechow, Arnim, eds. *Semantics from Different Points of View.* Berlin, Heidelberg, and New York: Springer, pp. 1–14. (Springer Series in Language and Communication, 6).

— (1989). Possible Worlds in Model-Theoretic Semantics: A Linguistic Perspective. In Allén, Sture, ed. *Possible Worlds in Humanities, Arts and Sciences.* Proceedings of Nobel Symposium 65, August 1986. Berlin and New York: Walter de Gruyter, pp. 93–123. (Research in Text Theory, 14).

Peirce, Charles Sanders (1885). On the Algebra of Logic: A Contribution to the Philosophy of Notation. *American Journal of Mathematics* 7.2, pp. 180–196. Repr. in PCP, vol. 3, §359–403; Peirce 1982–, vol. 5, pp. 162–190. Exc. in PE, vol. 1, pp. 225–228.

— (1982–). *Writings of Charles S. Peirce. A Chronological Edition.* Bloomington and Indianapolis: Indiana University Press.
Vol. 5: *1884–1886.* Ed. by Christian J. W. Kloesel, Max H. Fisch, Nathan Houser, et al. 1993.

Pepe, Luigi (2008). Il giovane Lagrange e i fondamenti dell'analisi. In Sacchi Landriani, Giannantonio e Giorgilli, Antonio, cur. *Sfogliando la Méchanique Analitique.* Giornata di studio su Louis Lagrange, Milano, 19 ottobre 2006. Milano: Edizioni Universitarie di Lettere Economia Diritto, pp. 37–49. (Incontri di studio – Istituto lombardo, Accademia di scienze e lettere, 46).

Perry, John (1996). Evading the Slingshot. In Clark, Andy, Ezquerro, Jesús, and Larrazabal, Jesús M., eds. *Philosophy and Cognitive Science: Categories, Consciousness and Reasoning.* Proceedings of the Second International Colloquium on Cognitive Science. Kluwer Academic Publishers, pp. 95–114. (Philosophical Studies Series, 69).

Petkovšek, Marko, Wilf, Herbert S., and Zeilberger, Doron (1996). *A=B.* Wellesley, Mass.: A K Peters.

Pietarinen, Ahti-Veikko (2016). Extensions of Euler Diagrams in Peirce's Four Manuscripts on Logical Graphs. In Jamnik, Mateja, Uesaka, Yuri, and Elzer Schwartz, Stephanie, eds. *Diagrammatic Representation and Inference.* 9th International Conference, Diagrams 2016. Cham: Springer, pp. 139–154. (Lecture Notes in Artificial Intelligence, 9781).

Pólya, George (1954). *Mathematics and Plausible Reasoning.* Vol. I: *Induction and Analogy in Mathematics.* Princeton: Princeton University Press.

Priestley, Mark (2011). *A Science of Operations. Machines, Logic and the Invention of Programming*. London and Dordrecht: Springer. (History of Computing).

— (2017). AI and the Origins of the Functional Programming Language Style. *Minds & Machines* 27.3, pp. 449–472.

Pringsheim, A. (1900). Zur Geschichte des Taylor'schen Lehrsatzes. *Bibliotheca Mathematica* 3.1, S. 433–479.

Proclus (1970). *A Commentary on the First Book of Euclid's Elements.* Trans. and annot., with an introd., by Glenn R. Morrow. Princeton: Princeton University Press.

Putnam, Hilary (1975). The Meaning of 'Meaning'. In *Philosophical Papers.* Vol. 2: *Mind, Language and Reality.* Cambridge: Cambridge University Press, pp. 215–271. Réimpr. de *Language, Mind and Knowledge.* Ed. by Keith Gunderson. Minneapolis: University of Minnesota Press, 1975, pp. 131–193. (Minnesota Studies in the Philosophy of Science, 7).

Pylyshyn, Zenon W. (1973). What the Mind's Eye Tells the Mind's Brain. A Critique of Mental Imagery. *Psychological Bulletin* 80.1, pp. 1–24.

— (1981). The Imagery Debate: Analogue Media versus Tacit Knowledge. *Psychological Review* 88.1, pp. 16–45.

— (2002). Mental Imagery: In Search of a Theory. *Behavioral and Brain Sciences* 25.2, pp. 157–182.

Qin, Yulin and Simon, Herbert A. (1995). Imagery and Mental Models. In Glasgow, T. I., Narayanan, N. H., and Chandrasekaran, B., eds. *Diagrammatic Reasoning. Cognitive and Computational Perspectives.* With a forew. by Herbert A. Simon. Menlo Park and Cambridge: AAAI Press and MIT Press, pp. 403–434.

Rabouin, David (2015). Proclus' Conception of Geometric Space and Its Actuality. In De Risi, Vincenzo, ed. *Mathematizing Space. The Objects of Geometry from Antiquity to the Early Modern Age.* Basel: Birkhäuser, pp. 105–142. (Trends in the History of Science).

— (2017). Styles in Mathematical Practice. In Chemla, Karine and Fox Keller, Evelyn, eds. *Cultures Without Culturalism. The Making of Scientific Knowledge.* Durham and London: Duke University Press, pp. 196–225.

Rashed, Roshdi, dir. (2008-2010). *Apollonius de Perge,* Coniques. *Texte grec et arabe établi,* *traduit et commenté sous la direction de Roshdi Rashed.* 7 vol. Berlin et New York : Walter de Gruyter. (Scientia Graeco-Arabica, 1).

Vol. 1.1 : *Livre I. Commentaire historique et mathématique, édition et traduction du texte arabe par Roshdi Rashed.* 2008.

Vol. 1.2 : *Livre I. Édition et traduction du texte grec par Micheline Decorps-Foulquier et Michel Federspiel.* 2008.

Raussen, Martin and Skau, Christian (2004). Interview with Jean-Pierre Serre. *Notices of the AMS* 51.2, pp. 210–214.

Recanati, François (2008). *Philosophie du langage (et de l'esprit).* [Paris] : Gallimard. (Folio essais, 509).

Reyneau, Charles-René (1708). *Analyse Démontrée.* Vol. 2 : *Usage de l'Analyse. Ou la manière de l'appliquer à découvrir les propriétés des figures [...].* Paris : Jacque Quillau.

Risse, Wilhelm (1970). *Die Logik der Neuzeit.* Bd. 2: *1640–1780.* Stuttgart–Bad Cannstatt: Friedrich Frommann.

Ross, Bertram (1977). The Development of Fractional Calculus, 1695–1900. *Historia Mathematica* 4.1, pp. 75–89.

Ruskey, Frank and Weston, Mark (2005). A Survey of Venn Diagrams. *Electronic Journal of Combinatorics* (June 2005). URL: http://www.combinatorics.org/files/Surveys/ds5/VennWhatEJC.html.

Saito, Ken (2006). A Preliminary Study in the Critical Assessment of Diagrams in Greek Mathematical Works. *SCIAMVS* 7, pp. 81–144.

— (2009). Reading Ancient Greek Mathematics. In Robson, Eleanor and Stedall, Jacqueline, eds. *The Oxford Handbook of the History of Mathematics.* Oxford and New York: Oxford University Press, pp. 801–826.

— (2011). The Diagrams of Book II and III of the *Elements* in Greek Manuscripts. In Saito, Ken, ed. *Diagrams in Greek Mathematical Texts.* Version 2.03. Greekmath.org, Apr. 3, 2011, pp. 39–80. URL: http://greekmath.org/diagrams/Diagrams_in_Greek_Mathematical_Texts_Report_Ver_2_03_20110403.pdf.

— (2012). Traditions of the Diagram, Tradition of the Text. A Case Study. *Synthese* 186.1: *Diagrams in Mathematics: History and Philosophy.* Ed. by John Mumma, Marco Panza, and Gabriel Sandu, pp. 7–20.

Saito, Ken and Sidoli, Nathan (2012). Diagrams and Arguments in Ancient Greek Mathematics. Lessons Drawn from Comparisons of the Manuscript Diagrams with Those in Modern Critical Editions. In Chemla, Karine, ed. *The History of Mathematical Proof in Ancient Traditions.* Cambridge and New York: Cambridge University Press, pp. 135–162.

Schafheitlin, Paul (1922). Johannis (I) Bernoullii *Lectiones de calculo differentialium. Verhandlungen der Naturforschenden Gesellschaft in Basel* 34, S. 1–32.

Schlimm, Dirk (2018). On Frege's Begriffsschrift Notation for Propositional Logic: Design Principles and Trade-Offs. *History and Philosophy of Logic* 39.1, pp. 53–79.

Schlimm, Dirk and Neth, Hansjörg (2008). Modeling Ancient and Modern Arithmetic Practices. Addition and Multiplication with Arabic and Roman Numerals. In Love, B. C., McRae, K., and Sloutsky, V. M., eds. *Proceedings of the 30th Annual Conference of the Cognitive Science Society.* CogSci 2008 (Washington, D.C. July 23–26, 2008). CD-ROM, archived online. Austin: Cognitive Science Society, pp. 2097–2102. URL: http://csjarchive.cogsci.rpi.edu/Proceedings/2008/pdfs/p2097.pdf.

Schroeder-Heister, Peter (2018). Proof-Theoretic Semantics. In Zalta, Edward N., ed. *The Stanford Encyclopedia of Philosophy.* Stanford: Metaphysics Research Lab, Stanford University, Spr. 2018. URL: https://plato.stanford.edu/archives/spr2018/entries/proof-theoretic-semantics/.

Scotto di Luzio, Patrick (2002). Patching Up a Logic of Venn Diagrams. In Vermeulen, Kees and Copestake, Ann, eds. *Algebras, Diagrams and Decisions in Language, Logic and Computation.* Stanford: CSLI Publications, pp. 119–134. (CSLI Lecture Notes, 144).

Sellars, Wilfrid (1963). Some Reflections on Language Games. In *Science, Perception and Reality.* New York: The Humanities Press, pp. 321–358. Rpt. (with rev.) from *Philosophy of Science* 21. 1954, pp. 204–228.

Serfati, Michel (2001). Mathématiques et pensée symbolique chez Leibniz. *Revue d'histoire des sciences* 54.2, p. 165-222.

— (2005). *La révolution symbolique. La constitution de l'écriture symbolique mathématique.* Avec une préf. de Jacques Bouveresse. Paris : Pétra. (Transphilosophiques).

— (2006). La constitution de l'écriture symbolique mathématique. Symbolique et invention. *Gazette de la SMF* 118 (avr. 2006), p. 101-118.

Serfati, Michel (2008). Symbolic Inventiveness and “Irrationalist” Practices in Leibniz’s Mathematics. In Dascal, Marcelo, ed. *Leibniz: What Kind of Rationalist?* Dordrecht: Springer, pp. 125–139. (Logic, Epistemology, and the Unity of Science, 13).

Serres, Michel (1968). *Le système de Leibniz et ses modèles mathématiques*. Paris : PUF.

Shannon, Claude E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal* 27.3, pp. 379–423.

Shepard, Roger N. and Metzler, Jacqueline (1971). Mental Rotation of Three-Dimensional Objects. *Science*. New ser. 171.3972, pp. 701–703.

Sher, Gila Y. (1996). Did Tarski Commit “Tarski’s Fallacy”? *The Journal of Symbolic Logic* 61.2, pp. 653–686.

Shimojima, Atsushi (1996a). *On the Efficacy of Representation*. PhD thesis. Bloomington: Indiana University. URL: http://shimo-pro.com/a-pro/Publications_in_English_files/Thesis.pdf.

— (1996b). Reasoning with Diagrams and Geometrical Constraints. In Seligman, Jerry and Westerståhl, Dag, eds. *Logic, Language and Computation*. Vol. 1. Stanford: CSLI Publications, pp. 527–540. (CSLI Lecture Notes, 58).

— (1999a). Constraint-Preserving Representations. In Moss, Lawrence S., Ginzburg, Jonathan, and de Rijke, Maarten, eds. *Logic, Language and Computation*. Vol. 2. Stanford: CSLI Publications, pp. 296–317. (CSLI Lecture Notes, 96).

— (1999b). Derivative Meaning in Graphical Representations. In *Proceedings of the 1999 IEEE Symposium on Visual Languages* (Tokyo, Japan, Sept. 13–16, 1999). Los Alamitos, Calif.: IEEE Computer Society, pp. 212–219.

— (1999c). The Linguistic-Graphic Distinction: Exploring Alternatives. *Artificial Intelligence Review* 13.4, pp. 313–335.

— (2003). What Makes a System Less Graphical? *Machine Graphics and Vision* 12.1, pp. 99–116.

— (2008). Sketching Formal Semantics of Graphical Meaning Derivation. *Tetsugaku Ronso* 35, pp. 1–21.

— (2014). Tutorial: Semantic Properties of Diagrams and Their Cognitive Potentials. (Slides from a tutorial given at the conference “Diagrammatic Representation and Inference: Eighth International Conference, Diagrams 2014”). URL: http://shimo-pro.com/a-pro/Publications_in_English_files/Tutorial_Diagrams2014.pdf.

— (2015). *Semantic Properties of Diagrams and Their Cognitive Potentials.* Stanford: CSLI Publications. (Studies in the Theory and Applications of Diagrams).

Shimojima, Atsushi and Barker-Plummer, Dave (2014). The Barwise-Seligman Model of Representation Systems. A Philosophical Explication. In Dwyer, Tim, Purchase, Helen, and Delaney, Aidan, eds. *Diagrammatic Representation and Inference.* 8th International Conference, Diagrams 2014. Berlin and Heidelberg: Springer, pp. 231–245. (Lecture Notes in Artificial Intelligence, 8578).

Shin, Sun-Joo (1991). *Valid Reasoning and Visual Representation.* PhD thesis. Stanford: Stanford University.

— (1994). *The Logical Status of Diagrams.* Cambridge and New York: Cambridge University Press.

— [1991] (1996). A Situation-Theoretic Account of Valid Reasoning with Venn Diagrams. In Allwein, Gerard and Barwise, Jon, eds. *Logical Reasoning with Diagrams.* New York and Oxford: Oxford University Press, pp. 81–108. (Studies in Logic and Computation, 6). Réimpr. de *Situation Theory and Its Applications.* Ed. by Jon Barwise, Jean Mark Gawron, Gordon Plotkin, and Syun Tutiya. Vol. 2. Stanford: Center for the Study of Language and Information, 1991, pp. 581–606. (CSLI Lecture Notes, 26).

— (2002). *The Iconic Logic of Peirce's Graphs.* MIT Press.

— (2004). Heterogeneous Reasoning and Its Logic. *Bulletin of Symbolic Logic* 10.1, pp. 86–106.

Shin, Sun-Joo, Lemon, Oliver, and Mumma, John (2013). Diagrams. In Zalta, Edward N., ed. *The Stanford Encyclopedia of Philosophy.* Stanford: Metaphysics Research Lab, Stanford University, Aut. 2013. URL: https://plato.stanford.edu/archives/fall2013/entries/diagrams/.

Short, T. L. (2007). *Peirce's Theory of Signs.* Cambridge and New York: Cambridge University Press.

Simon, Herbert A. (1947). *Administrative Behavior. A Study of Decision-Making Processes in Administrative Organization.* With a forew. by Chester I. Barnard. London and New York: The Free Press—Macmillan. 2nd ed. 1957; 3rd ed. = Simon 1976; 4th ed. 1997.

— (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society* 106.6, pp. 467–482. Repr. in Simon 1969, chap. 4 (2nd ed., chap. 7; 3rd ed. = Simon 1996, chap. 8).

Simon, Herbert A. (1969). *The Sciences of the Artificial.* Cambridge, Mass. and London: MIT Press. 2nd ed. 1981; 3rd ed. = Simon 1996.

— (1972). What Is Visual Imagery? An Information Processing Interpretation. In Gregg, Lee W., ed. *Cognition in Learning and Memory.* New York: Wiley, pp. 183–204.

— (1976). *Administrative Behavior. A Study of Decision-Making Processes in Administrative Organization.* With a forew. by Chester I. Barnard. 3rd ed. London and New York: The Free Press—Macmillan.

— (1978). On the Forms of Mental Representation. In Savage, C. Wade, ed. *Perception and Cognition. Issues in the Foundations of Psychology.* Minneapolis: University of Minnesota Press, pp. 3–18. (Minnesota Studies in the Philosophy of Science, IX).

— (1989). The Scientist as Problem Solver. In Klahr, David and Kotovsky, Kenneth, eds. *Complex Information Processing. The Impact of Herbert A. Simon.* 21st Carnegie-Mellon Symposium on Cognition. Hillsdale, New Jersey: Lawrence Erlbaum Associates, pp. 375–398.

— (1991). *Models of My Life.* New York: Basic Books. (The Alfred P. Sloan Foundation series).

— (1996). *The Sciences of the Artificial.* 3rd ed. Cambridge, Mass. and London: MIT Press.

Soames, Scott (2010). *Philosophy of Language.* Princeton and Oxford: Princeton University Press. (Princeton Foundations of Contemporary Philosophy).

— (2012). Propositions. In Russell, Gillian and Graff Fara, Delia, eds. *The Routledge Companion to the Philosophy of Language.* New York and London: Routledge, pp. 209–220. (Routledge Philosophy Companions).

Speaks, Jeff (2014). Theories of Meaning. In Zalta, Edward N., ed. *The Stanford Encyclopedia of Philosophy.* Stanford: Metaphysics Research Lab, Stanford University, Sum. 2014. URL: https://plato.stanford.edu/archives/sum2014/entries/meaning/.

Speiser, Andreas (1960). Einleitung zu den *Lettres à une princesse d'Allemagne.* In Euler, Leonhard. *Opera Omnia.* Bd. III.11–12: *Lettres à une princesse d'Allemagne. Rettung der göttlichen Offenbarung. Éloge d'Euler par le Marquis de Condorcet.* Hrsg. von Andreas Speiser. 2 Bde. Turici [Zürich]: Orell Füssli, S. VII–XLIII.

Stalnaker, Robert C. (1984). *Inquiry.* Cambridge, Mass. and London: Bradford Books—MIT Press.

Stapleton, Gem, Jamnik, Mateja, and Shimojima, Atsushi (2017). What Makes an Effective Representation of Information: A Formal Account of Observational Advantages. *Journal of Logic, Language and Information* 26.2, pp. 143–177.

Stapleton, Gem, Shimojima, Atsushi, and Jamnik, Mateja (2018). The Observational Advantages of Euler Diagrams with Existential Import. In Chapman, Peter et al., eds. *Diagrammatic Representation and Inference.* 10th International Conference, Diagrams 2018. Cham: Springer, pp. 313–329. (Lecture Notes in Artificial Intelligence, 10871).

Stein, Myron L., Ulam, Stanislaw M., and Wells, Mark B. (1964). A Visual Display of Some Properties of the Distribution of Primes. *American Mathematical Monthly* 71.5, pp. 516–520.

Stenning, Keith (2000). Distinctions with Differences. Comparing Criteria for Distinguishing Diagrammatic from Sentential Systems. In Anderson, Michael, Cheng, Peter, and Haarslev, Volker, eds. *Theory and Application of Diagrams.* First International Conference, Diagrams 2000. Berlin, Heidelberg, and New York: Springer, pp. 132–148. (Lecture Notes in Artificial Intelligence, 1889).

— (2002). *Seeing Reason. Image and Language in Learning to Think.* Oxford: Oxford University Press. (Oxford Cognitive Science Series).

Stjernfelt, Frederik (2006). Two Iconicity Notions in Peirce's Diagrammatology. In Schärfe, Henrik, Hitzler, Pascal, and Øhrstrøm, Peter, eds. *Conceptual Structures. Inspiration and Application.* Proceedings of the 14th International Conference on Conceptual Structures, ICCS 2006. Berlin and Heidelberg: Springer, pp. 70–86. (Lecture Notes in Artificial Intelligence, 4068).

— (2007). *Diagrammatology. An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics.* Dordrecht: Springer. (Synthese Library, 336).

Stojanovic, Isidora (2012). Situation Semantics. In Newen, Albert and van Riel, Raphael, eds. *Identity, Language, and Mind: An Introduction to the Philosophy of John Perry.* Stanford and Paderborn: CSLI Publications; Mentis, pp. 67–86. (CSLI Lecture Notes, 203).

Sturm, Johann Christoph (1661). *Universalia Euclidea.* Hagæ-Comitis [La Haye] : Adrian Vlacq.

Sundholm, Göran [1986] (2002). Proof Theory and Meaning. In Gabbay, Dov M. and Guenthner, Franz, eds. *Handbook of Philosophical Logic.* 2nd ed. Vol. 9. Dordrecht: Kluwer Academic Publishers, pp. 165–198.

Tabachneck-Schijf, Hermina J. M., Leonardo, Anthony M., and Simon, Herbert A. (1997). CaMeRa: A Computational Model of Multiple Representations. *Cognitive Science* 21.3, pp. 305–350.

Takahashi, Nobuo (2015). Where is Bounded Rationality From? *Annals of Business Administrative Science* 14, pp. 67–82.

Tappenden, Jamie (2008a). Mathematical Concepts and Definitions. In Mancosu, Paolo, ed. *The Philosophy of Mathematical Practice.* Oxford and New York: Oxford University Press, pp. 256–275.

— (2008b). Mathematical Concepts: Fruitfulness and Naturalness. In Mancosu, Paolo, ed. *The Philosophy of Mathematical Practice.* Oxford and New York: Oxford University Press, pp. 276–301.

Tarski, Alfred [1933] (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. Übers. von L. Blaustein. *Studia Philosophica* 1, S. 261–405. Réimpr. in Tarski 1986, S. 53–178. Trad. ang. rév. in Tarski 1956, S. 152–278. Trad. fr. in Tarski 1972–1974, vol. 1, p. 157–269. Trad. de Pojęcie prawdy w językach nauk dedukcyjnych. *Prace Towarzystwa Naukowego Warszawskiego, Wydział III, Matematyczno-fizycznych* 34 (1933).

— (1936a). Grundlegung der wissenschaftlichen Semantik. In *Actes du Congrès International de Philosophie Scientifique.* Vol. III. Paris : Hermann et C^ie^, p. 1-8. (Actualités scientifiques et industrielles, 390). Réimpr. in Tarski 1986, p. 261-268. Trad. ang. rév. in Tarski 1956, p. 401-408. Trad. fr. in Tarski 1972-1974, vol. 2, p. 131–139. Trad. de O ugruntowaniu naukowej semantyki. *Przegląd Filozoficzny* 39.1 (1936), s. 50–57.

— (1936b). Über den Begriff der logischen Folgerung. In *Actes du Congrès International de Philosophie Scientifique.* Vol. VII. Paris: Hermann et C^ie^, pp. 1–11. (Actualités scientifiques et industrielles, 394). Réimpr. in Tarski 1986, pp. 271–281. Trad. ang. rév. in Tarski 1956, pp. 409–420. Trad. fr. in Tarski 1972–1974, vol. 2, p. 141–152. Trad. de O pojęciu wynikania logicznego. *Przegląd Filozoficzny* 39.1 (1936), s. 58–68.

— (1956). *Logic, Semantics, Metamathematics. Papers from 1923 to 1938.* Trans. by J. H. Woodger. Oxford: Clarendon Press.

— (1972-1974). *Logique, sémantique, métamathématique. 1923–1944.* Sous la dir. de Gilles-Gaston Granger. 2 vol. Paris : Armand Colin. (Philosophies pour l'âge de la science).

— (1986). *Collected Papers.* Vol. 2: *1935–1944.* Ed. by Steven R. Givant and Ralph N. McKenzie. Basel, Boston, and Stuttgart: Birkhäuser. (Contemporary Mathematicians).

Taylor, Brook (1715). *Methodus incrementorum directa et inversa.* Londini : Pearson.

Tefera, Akalu (2010). What Is… a Wilf-Zeilberger Pair? *Notices of the AMS* 57.4, pp. 508–509.

Tucker, John V. (2018). Theorising Data. A History of Abstract Data Types and their Specification. (Slides of a talk given at the Fourth Symposium on the History and Philosophy of Programming (HaPoP), March 2018, Oxford).

Tufte, Edward R. (1983). *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

— (1990). *Envisioning Information.* Cheshire, CT: Graphics Press.

— (1997). *Visual Explanations. Images and Quantities, Evidence and Narrative.* Cheshire, CT: Graphics Press.

Tukey, John W. (1972). Some Graphic and Semigraphic Displays. In Bancroft, T.A., ed. *Statistical Papers in Honor of George W. Snedecor.* Ames, Iowa: Iowa State University Press, pp. 293–316.

— (1977). *Exploratory Data Analysis.* Reading, Mass. and Menlo Park, Calif.: Addison-Wesley. (Addison-Wesley Series in Behavioral Science: Quantitative Methods).

Van Benthem, Johan and Martinez, Maricarmen (2008). The Stories of Logic and Information. In Adriaans, Pieter and van Benthem, Johan, eds. *Handbook of the Philosophy of Science.* Vol. 8: *Philosophy of Information.* Amsterdam and Boston: North-Holland, pp. 225–288.

Venn, John (1880). On the Diagrammatic and Mechanical Representation of Propositions and Reasonings. *Philosophical Magazine.* 5th ser. 10.59 (July 1880), pp. 1–18.

— (1881). *Symbolic Logic.* London: Macmillan.

— (1883). Review of Studies in Logic. *Mind* 8.32, pp. 594–603.

— (1894). *Symbolic Logic.* 2nd ed. London: Macmillan.

Vera, Alonso H. and Simon, Herbert A. (1993). Situated Action: A Symbolic Interpretation. *Cognitive Science* 17.1, pp. 7–48.

Volkert, Klaus Thomas (1986). *Die Krise der Anschauung. Eine Studie zu formalen und heuristischen Verfahren in der Mathematik seit 1850.* Göttingen: Vandenhoeck & Ruprecht. (Studien zur Wissenschafts-, Sozial- und Bildungsgeschichte der Mathematik, 3).

Vorms, Marion (2009a). Le pendule, la licorne et la figure. Comment les modèles scientifiques représentent les phénomènes. *Philonsorbonne* 3, pp. 75–95. URL: http://journals.openedition.org/philonsorbonne/237.

— (2009b). *Théories, mode d'emploi. Une perspective cognitive sur l'activité théorique dans les sciences empiriques.* Thèse de doct. Université Paris 1 Panthéon-Sorbonne. HAL : tel-00462403.

— (2011). *Qu'est-ce qu'une théorie scientifique ?* Avec une préf. de Paul Humphreys. Paris : Vuibert. (Philosophie des sciences).

— (2012). Formats of Representation in Scientific Theorizing. In Humphreys, Paul and Imbert, Cyrille, eds. *Models, Simulations, and Representations.* New York and Abingdon: Routledge, pp. 250–269. (Routledge Studies in the Philosophy of Science, 9).

Wagner, Pierre (1998). *La machine en logique.* Paris : Presses Universitaires de France. (Science, histoire et société).

Wagner, Roi (2017). *Making and Breaking Mathematical Sense: Histories and Philosophies of Mathematical Practice.* Princeton: Princeton University Press.

Wallis, John (1693). *Opera Mathematica.* Vol. 2 : *De Algebra Tractatus; Historicus & Practicus. Anno 1685 Anglice editus; Nunc Auctus Latine.* Cum variis appendicibus; partim prius editis Anglice, partim nunc primum editis. Oxoniæ [Oxford] : E Theatro Sheldoniano.

Waring, Edward (1785). *Meditationes Analyticæ.* 3e éd. Cantabrigiæ [Cambridge] : Typis Academicis. 1re éd. 1773.

Waszek, David (2018). Rigor and the Context-Dependence of Diagrams. The Case of Euler Diagrams. In Chapman, Peter et al., eds. *Diagrammatic Representation and Inference.* 10th International Conference, Diagrams 2018. Cham: Springer, pp. 382–389. (Lecture Notes in Artificial Intelligence, 10871).

Weil, André (1978). Who Betrayed Euclid? (Extract from a Letter to the Editor). *Archive for History of Exact Sciences* 19.2, pp. 91–93.

Wilson, Mark (1994). Can We Trust Logical Form? *The Journal of philosophy* 91.10, pp. 519–544.

— (2006). *Wandering Significance.* Oxford and New York: Clarendon Press.

Wilson, Robert A. (1994). Wide Computationalism. *Mind.* New ser. 103.411, pp. 351–372.

Wittgenstein, Ludwig (1976). *Lectures on the Foundations of Mathematics. Cambridge, 1939.* From the notes of R. G. Bosanquet, Norman Malcolm, Rush Rhees, and Yorick Smythies. Ed. by Cora Diamond. Hassocks, Sussex: The Harvester Press.

— [1921] (1993). *Tractatus Logico-Philosophicus.* Trad., préf. et notes par Gilles Gaston Granger. Avec une introd. de Bertrand Russell. [Paris] : Gallimard. Trad. de Logisch-Philosophische Abhandlung. *Annalen der Naturphilosophie* 14 (1921). Hrsg. von Wilhelm Ostwald, S. 185–262.

Woodhouse, Robert (1803). *The Principles of Analytical Calculation.* Cambridge: University Press.

Yeung, Raymond W. (2008). *Information Theory and Network Coding*. New York: Springer. (Information Technology: Transmission, Processing, and Storage).

Zimmermann, Thomas Ede (2011). Model-Theoretic Semantics. In Maienborn, Claudia, von Heusinger, Klaus, and Portner, Paul, eds. *Semantics: An International Handbook of Natural Language Meaning*. Vol. 1. Berlin and Boston: De Gruyter Mouton, pp. 762–802. (Handbücher zur Sprach- und Kommunikationswissenschaft, 33.1).

Index des noms

David WASZEK — Thèse de doctorat en philosophie — Université Paris 1 Panthéon-Sorbonne

LES REPRÉSENTATIONS EN MATHÉMATIQUES

Résumé : Pour résoudre un problème de mathématiques ou comprendre une démonstration, une figure bien choisie est parfois d'un grand secours. Ce fait souvent remarqué peut être vu comme un cas particulier d'un phénomène plus général. Utiliser une figure plutôt que des phrases, reformuler un problème sous la forme d'une équation, employer telles notations plutôt que telles autres : dans tous ces cas, en un sens, on ne fait que représenter sous une nouvelle forme ce qu'on sait déjà, et pourtant, cela peut permettre d'avancer. Comment est-ce possible ? Pour répondre à cette question, la première partie de cette thèse étudie ce qu'apporte un changement notationnel précis introduit par Leibniz à la fin du XVII^e^ siècle. La suite de ce travail analyse, et confronte à l'exemple précédent, plusieurs manières de penser les différences représentationnelles proposées dans la littérature philosophique récente. Herbert Simon, étudié dans la deuxième partie, s'appuie sur le modèle informatique des structures de données : deux représentations peuvent être « informationnellement » équivalentes, mais « computationnellement » différentes. Les logiciens Barwise et Etchemendy, étudiés dans la troisième partie, cherchent à élargir les concepts de la logique mathématique (en particulier ceux de syntaxe et de sémantique) aux diagrammes et figures. Enfin, certains philosophes des mathématiques contemporains, comme Kenneth Manders, remettent en cause la notion même de représentation, en soutenant qu'elle n'est pas éclairante pour comprendre l'usage de figures, formules ou autres supports externes en mathématiques. C'est à ces critiques qu'est consacrée la quatrième et dernière partie.

REPRESENTATIONS IN MATHEMATICS

Abstract: When solving a mathematical problem or reading a proof, drawing a well-chosen diagram may be very helpful. This well-known fact can be seen as an instance of a more general phenomenon. Using a diagram rather than sentences, reformulating a problem as an equation, choosing a particular notation rather than others: in all these cases, in a sense, we are only representing in a new form what we already knew; and yet, it can help us make progress. How is this possible? To address this question, the first part of this thesis explores the benefits afforded by a specific notational change introduced by Leibniz in the late seventeenth-century. The rest of this work analyses, and puts to the test of the preceding case study, several ways of understanding representational differences which have been put forward in the recent philosophical literature. Herbert Simon, studied in the second part, relies on a comparison with the notion of data structures in computer science: two representations, he writes, can be "informationally" equivalent yet "computationnally" different. The logicians Barwise and Etchemendy, studied in the third part, try to broaden the concepts of mathematical logic (in particular those of syntax and semantics) to cover diagrams and figures. Finally, some contemporary philosophers of mathematics, for instance Ken Manders, argue that the notion of representation itself is not helpful to understand the use of diagrams, formulas or other external reasoning tools in mathematics. Such arguments are the focus of the fourth (and last) part.

Mots-clés : Philosophie des mathématiques – Histoire des mathématiques – Logique – Notation – Représentation – Figure – Diagramme – Signe

École Doctorale de Philosophie (ED 280)
Université Paris 1 Panthéon-Sorbonne
1 rue d'Ulm
75005 Paris

Institut d'histoire et de philosophie des sciences et des techniques (UMR 8590)
13 rue du Four
75006 Paris