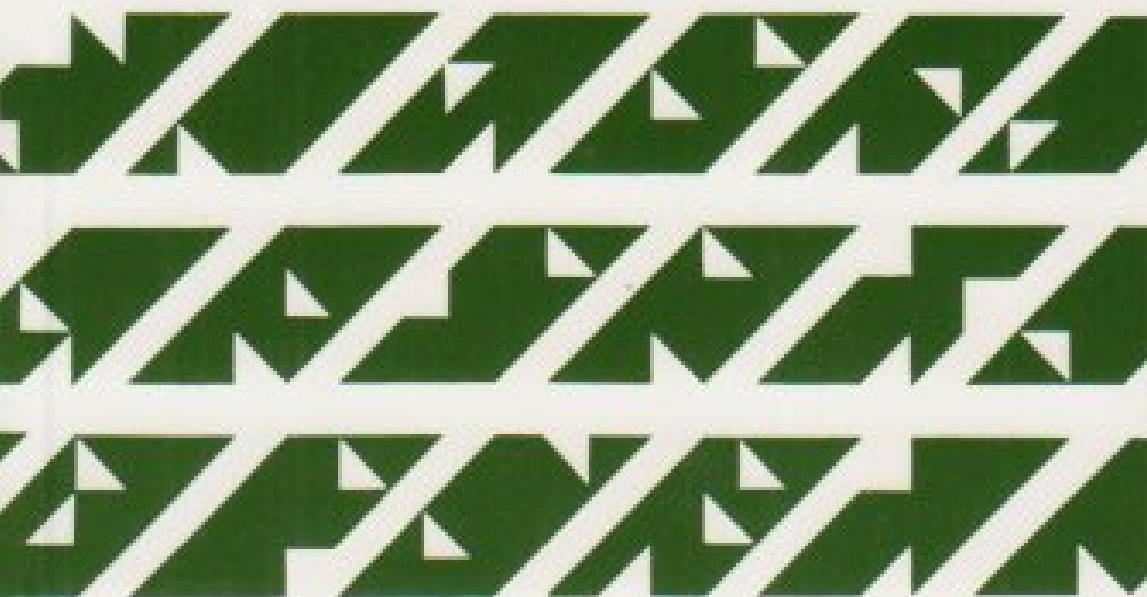


Méthodes logiques pour les sciences cognitives

coordonnateurs

Jacques Dubucs & François Lepage



HERMES

Facebook : La culture ne s'hérite pas elle se conquiert

Méthodes logiques pour les sciences cognitives

Langue, Raisonnement, Calcul

Collection dirigée par Mario Borillo et Frédéric Nef

La collection Langue, Raisonnement, Calcul rassemble des ouvrages consacrés à l'étude des langages et des raisonnements, qu'ils soient naturels ou formels, ou de programmation. Au carrefour des sciences cognitives, de la linguistique, de l'informatique, de la théorie formelle des activités symboliques (dialogue, argumentation, apprentissage) cette collection a pour ambition de contribuer à faire émerger une nouvelle problématique scientifique.

Titres parus

Jean-Marie PIERREL, *Dialogue oral homme-machine*, 1987.

Jacques JAYEZ, *L'inférence en langue naturelle*, 1988.

Gérard SABAH, *L'intelligence artificielle et le langage*, vol. 1 : *Représentation des connaissances*, 1988 ; vol. 2 : *Processus de compréhension*, 1989.

Frédéric NEF, *Logique et Langage, Essais de logique intensionnelle*, 1988.

Stanislaw LESNIEWSKI, *Sur les fondements de la mathématique*, trad. G. KALINOWSKI, 1989.

Jacques MOESCHLER, *Modélisation du dialogue. Représentation de l'inférence argumentative*, 1989.

Philip MILLER et Thérèse TORRIS, éd., *Formalismes syntaxiques pour le traitement automatique du langage naturel*, 1990.

M. LOTHAIRE, éd., *Mots, Mélanges offerts à M.P. SCHUTZENBERGER*, 1990.

Jacques PITRAT, *Métaconnaissance, futur de l'intelligence artificielle*, 1990.

Paul GOCHET, Pascal GRIBOMONT, *Logique, méthodes pour l'informatique fondamentale*, vol. 1, 1990.

Paul GOCHET, Pascal GRIBOMONT, *Logique, méthodes formelles pour l'étude des programmes*, vol. 2, 1994.

Jean-Pierre DESCLÉS, *Langages applicatifs, langues naturelles et cognition*, 1990.

Eric GRÉGOIRE, *Logiques non monotones et intelligence artificielle*, 1990.

Jean-François NOGIER, *Génération automatique de langage et graphes conceptuels*, 1991.

Archibald MICHIELS, *Traitement du langage naturel et Prolog*, 1991.

Stéphane BOUCHERON, *Théorie de l'apprentissage — de l'approche formelle aux enjeux cognitifs —*, 1992.

Eric BILANGE, *Dialogue personne-machine, modélisation et réalisation informatique*, 1992.

Emmanuel SAINT-JAMES, *La programmation applicative (de LISP à la machine en passant par le lambda-calcul)*, 1993.

Jean-Paul DELAHAYE, *Information, complexité et hasard*, 1994.

Philippe BALBIANI, Vincent DUGAT, Luis FARINAS DEL CERRO, Anne LOPEZ, *Éléments de géométrie mécanique*, 1994.

Bruno BACHIMONT, *Le contrôle dans les systèmes à base de connaissances — contribution à l'épistémologie de l'intelligence artificielle*, 2^e éd., 1994.

coordonnateurs
Jacques Dubucs et François Lepage

Méthodes logiques pour les sciences cognitives

HERMES

© Hermès, Paris, 1995
Éditions Hermès
14, rue Lantiez
75017 Paris

ISBN 2-86601-461-8
ISSN 0988-0569

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, d'une part, que les "copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective" et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, "toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite" (article L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Table des matières

Liste des auteurs	11
Introduction	13
PARTIE I. LOGIQUE ET RAISONNEMENT	23
Chapitre 1. Logique, raisonnement et psychologie	25
1. Etude expérimentale et théories du raisonnement déductif élémentaire	28
2. Théories du deuxième ordre.....	35
3. Questions théoriques	40
4. Une catégorie de raisonnements inductifs : le jugement en situation d'incertitude	54
5. Bibliographie.....	71
Chapitre 2. Confirmation non probabiliste	77
1. Introduction	77
2. Principes d'adéquation.....	80
2.1. Principes d'adéquation de la confirmation absolue	81
2.2. Principes d'adéquation de la confirmation relative	83
3. Critères probabilistes.....	84
4. Confirmation absolue.....	85
4.1. Syntaxe : confirmation absolue et logique non monotone.....	85
4.2. Sémantique : confirmation absolue et théorie des possibilités	87
5. Théorie des possibilités et confirmation relative.....	89
6. Conclusion.....	94
Annexe	95
Bibliographie.....	95

Chapitre 3. Logique possibiliste, modèles préférentiels et objets conditionnels	99
1. Introduction	100
2. Logique possibiliste et bases de connaissances stratifiées	101
3. L'inférence préférentielle et ses propriétés	105
4. Représentation de connaissances génériques	108
5. Des objets conditionnels à la logique possibiliste	112
6. Conclusion	115
Bibliographie	117

Chapitre 4. Une application de la logique linéaire : la planification déductive	121
1. Généralités	121
1.1. La planification en général	122
1.2. La planification déductive	122
1.3. Le contenu de cet article	124
2. Planification déterministe	125
2.1. Planification déterministe "concrète"	126
2.2. Définition des actions concrètes d'un système	130
3. Planification déductive	134
3.1. Tentative dans un cadre traditionnel	135
3.2. Planification en logique linéaire	142
3.3. Une géométrie de l'action	149
4. Calcul des séquents linéaire	154
4.1. Le groupe de l'identité	155
4.2. Le groupe multiplicatif	156
4.3. Le groupe additif	158
4.4. Le groupe de l'exponentielle	160
Bibliographie	161
Annexe des familles	162

Chapitre 5. Approches logiques du raisonnement taxinomique	165
Introduction	165
1. Le raisonnement taxinomique	167
1.1. La taxinomie en tant qu'outil conceptuel	167
1.2. Trois formes de prédication	169
1.3. Les taxinomies et la logique	172
2. La formalisation logique des hiérarchies d'héritage	174
2.1. Les hiérarchies d'héritage	175
2.2. Les solutions non monotones	178
2.3. Les solutions dans le cadre de la logique linéaire	184
3. Conclusion	190
Bibliographie	192

Chapitre 6. Changements minimaux constructifs.....	195
1. Introduction	195
2. Langage et sémantique.....	197
3. Axiomatique.....	198
4. Mise en forme normale	199
5. Correction et complétude.....	200
6. Dédution automatique dans ASSUME.....	200
7. La représentation du changement dans ASSUME.....	201
8. Discussion	202
Bibliographie.....	203
 PARTIE II. LOGIQUE ET REPRÉSENTATION	 205
 Chapitre 7. Logique de la connaissance : un panorama	 207
1. Connaissance, sens et référence.....	208
1.1. L'intuitionnisme, une logique du sens	209
1.2. La logique épistémique, une logique de la référence indexée	210
1.3. Sémantiques de la référence indexée : convergences et controverses	214
1.4. Critiques des sémantiques de la référence indexée.....	218
1.5. Les paradoxes du représentationnalisme	220
2. Perspectives actuelles en logique de la connaissance.....	221
2.1. La connaissance en société.....	222
2.2. Rationalité limitée : vers la fin des logiques immaculées	223
 Chapitre 8. “Mondes de croyance” et probabilité bivalentes	
Une interprétation logique du théorème de Mertens et Zamir	235
1. Introduction	235
2. Probabilités à deux valeurs et structures de voisinage	240
3. Les hiérarchies de probabilité bivalentes - un analogue simplifié du théorème des Mertens et Zamir	246
4. Une interprétation logique des hiérarchie de probabilités bivalentes.....	252
5. Conclusion.....	256
Bibliographie.....	258
 Chapitre 9. Applications épistémiques de la théorie des ensembles anti-fondés	 261
1. Introduction	261
2. Une classe universelle anti-fondée pour les croyances	266
3. Logique modale anti-fondée	271
4. Logique modale anti-fondée et connaissance commune	276
5. Conclusion et perspectives.....	278
Bibliographie.....	279

Chapitre 10. Problèmes de logique de la connaissance 281

1. Introduction 281

2. Les formes du problème de l'omniscience logique..... 282

2.1. La fermeture sur l'implication matérielle 282

2.2. La fermeture sur l'implication valide..... 283

2.3. La fermeture sur la règle de nécessité épistémique 283

2.4. La fermeture sur la conjonction..... 283

2.5. L'atténuation de la croyance : $Bp \Rightarrow B(p \vee q)$ 283

2.6. La cohérence de la croyance : $Bp \Rightarrow \neg B\neg p$ 283

2.7. Bp implique p . Ceci ne fait pas difficulté dans le système KD45
qui ne contient pas l'axiome T dont 2.7 est une instance..... 283

2.8. $B(Bp \Rightarrow p)$ 283

2.9. B true. Le sujet croit toutes tautologies 284

2.10. Un plus grand nombre d'alternatives épistémiques fournit
un plus petit nombre de croyances (principe de l'antimonotonie
des alternatives épistémiques) 284

2.11. Si $p \Leftrightarrow q$, alors $Bp \Leftrightarrow Bq$ 284

3. La logique générale de la conscience et la logique du raisonnement local
de Fagin et Halpern..... 284

3.1. La logique générale de la conscience..... 284

3.2. La logique du raisonnement local 286

4. La contribution d'E. Gillet 286

4.1. Les degrés de conscience 286

4.2. L'enchâssement de l'analyse dans la conscience implicite..... 288

4.3. La mesure de la profondeur..... 290

4.4. La fonction d'analyse..... 290

4.5. Exemple d'application de la fonction d'analyse à une formule..... 290

4.6. L'élimination de l'omniscience logique dans le système de E. Gillet
292

4.7. Recherches futures..... 293

Bibliographie 293

**Chapitre 11. L'analyse de la forme logique des propositions
en grammaire universelle** 295

1. Principes généraux de la logique de la quantification 298

2. La langue-objet idéale..... 308

3. La sémantique formelle 315

4. Un système axiomatique 317

5. Lois valides 321

Bibliographie 325

PARTIE III. LOGIQUE ET LANGAGE	327
Chapitre 12. Logique partielle et inférence valide	329
1. La théorie des types propositionnels.....	331
2. La logique partielle	336
3. Deux notions de validité	343
4. Un système déductif pour la logique partielle	344
Bibliographie	348
Chapitre 13. Conditionnels, quantificateurs et inférences	351
Introduction	351
1. Quantificateurs et conditionnels.....	352
1.1. Qu'est-ce qu'un quantificateur généralisé ?	352
1.2. Analogie entre les conditionnels et les quantificateurs	355
2. Quantificateurs et inférences.....	360
3. Les logiques C , E et QD	362
4. Les logiques CB , EB et QDB	374
5. Conclusion.....	374
Bibliographie	374
Chapitre 14. Logique constructive et donkey-anaphore	377
1. Introduction	377
1.1. L'exigence d'une théorie de la compréhension.....	377
1.2. Le problème des donkey sentences	378
1.3. Une stratégie de solution	380
2. La logique constructive LC	381
2.1. Syntaxe	381
2.2. Sémantique	383
2.3. Quelques remarques	386
3. Un fragment d'anglais	388
4. Traduction	389
5. Quelques exemples	390
6. Conclusion.....	392
Bibliographie.....	394
Chapitre 15. Approximations et négation	395
1. Introduction	395
2. La problématique fondamentale des approximations et de la négation	395
3. Logique classique et négation	402
4. Conclusion.....	408
Bibliographie.....	408

Chapitre 16. Paradoxes sémantiques et programmation logique..... 411

1. Introduction 411

1.1. Quel est le problème posé par les paradoxes sémantiques ?..... 412

1.2. Quel est le problème posé par les langages logiques à base
de règles ? 412

1.3. Quel rapport y a-t-il entre ces deux problèmes ? 413

2. Le parallèle..... 413

2.1. Dire le plus possible sans tomber sur l'inconsistance 413

2.2. Sur-détermination et sous-détermination..... 414

2.3. Rechercher des principes généraux 414

2.4. Les principes généraux s'expriment naturellement avec des
équivalences..... 415

2.5. Les logiques partielles (ou trivaluées) s'imposent..... 416

2.6. Les principes opératoires sont jugés importants 416

3. Conclusion..... 418

3.1. Non-monotonie, remplissage maximum cohérent..... 418

3.2. Les schémas de sens wittgensteinien et gödélien..... 419

Annexe 421

Bibliographie..... 425

Conclusion 429

Liste des auteurs

Partie I. Logique et raisonnement

ANDLER Daniel - Université de Paris X
Chap. 1. *Logique, raisonnement et psychologie* (p. 23-76)

ZWIRN Denis et ZWIRN Hervé - Université de Paris VI. Chap. 2. *Confirmation non probabiliste* (p. 77-98)

DUBOIS Didier et PRADE Henri - IRIT - Université Paul Sabatier, Toulouse
Chap. 3. *Logique possibiliste, modèles préférentiels et objets conditionnels* (p. 99-120)

MASSERON Marcel - Université de Paris-Nord. Chap. 4. *Une application de la logique linéaire : la planification déductive* (p. 121-164)

CROCCO Gabriella - IRIT - Université Paul Sabatier, Toulouse. Chap. 5. *Approches logiques du raisonnement taxinomique* (p. 165-194)

FARINAS DEL CERRO Luis et HERZIG Andreas - IRIT - Université Paul Sabatier, Toulouse
Chap. 6. *Changement minimaux constructifs* (p. 195-204)

Partie II. Logique et représentation

DUBUCS Jacques - Université de Paris I
Chap. 7. *Logique de la connaissance : un panorama* (p. 207-234)

MONGIN Philippe - CNRS
Chap. 8. *"Monde de croyance" et probabili-*

tés bivalentes. Une interprétation du théorème de Mertens et Zamir (p. 235-260)

LISMONT Luc - Université de Louvain.
Chap. 9. *Applications épistémiques de la théorie des ensembles anti-fondés* (p. 261-280)

GOCHET Paul - Université de Liège
Chap. 10. *Problème de logique de la connaissance* (p. 281-294)

VANDERVEKEN Daniel - Université du Québec. Chap. 11. *L'analyse de la forme logique des propositions en grammaire universelle* (p. 295-326)

Partie III. Logique et langage

LEPAGE François - Université de Montréal
Chap. 12. *Logique partielle et inférence valide* (p. 329-350)

LAPIERRE Serge - Université de Montréal
Chap. 13. *Conditionnels, quantificateurs et inférences* (p. 351-376)

USBERTI Gabriele - Université de Sienne
Chap. 14. *Logique constructive et donkey-anaphore* (p. 377-394)

MARQUIS Jean-Pierre - Université de Montréal. Chap. 15. *Approximations et négation* (p. 395-410)

DELAHAYE Jean-Paul - Université des Sciences et Technologies de Lille
Chap. 16. *Paradoxes sémantiques et programmation logique* (p. 411-426)

Introduction

Ce livre a pour objectif de donner un panorama des méthodes logiques applicables aux sciences cognitives. Il est le fruit de l'effort combiné de philosophes, d'informaticiens, d'économistes et de mathématiciens. Bien que l'intégration de leurs contributions respectives n'ait sans doute pas été poussée aussi loin qu'il eût été possible, leur entreprise est inspirée par une conviction commune, à savoir que la logique a un rôle essentiel à jouer dans le domaine multiforme de ce qu'il est aujourd'hui convenu d'appeler les "sciences cognitives". Une telle déclaration de principe est cependant trop vague pour être véritablement instructive, et la présente introduction est destinée à en préciser le contenu.

1. Logique, mathématique et psychologie

La logique fut longtemps conçue comme une branche de la psychologie. C'était avant Frege, en des temps où la logique stagnait, et que nous avons aujourd'hui oubliés. De la doctrine alors dominante, ne subsiste guère que la mémoire de son nom, qui la condamne : le *psychologisme*. La non-contradiction, par exemple, y était considérée non comme une incompatibilité logique (A et Non-A ne peuvent pas être simultanément vraies), mais comme une impossibilité psychique (le jugement que A et le jugement que Non-A ne peuvent coexister dans la même conscience). Le psychologisme est une impasse, la cause est aujourd'hui entendue, et nul ne songe plus, selon la cruelle expression de Frege, à laisser tremper la logique dans la "bassine psychologique" ¹. En revanche, il semble possible de conserver à la logique son caractère mathématique sans souscrire à la forme particulièrement radicale d'antipsychologisme que défendait Frege.

1. *Ecrits Logiques et Philosophiques*, Editions du Seuil, 1971, p. 144.

Pour Frege, les représentations mentales sont incapables d'assurer l'objectivité de la logique, car elles sont non seulement *privées* (autrui ne peut pas plus avoir mes représentations qu'il ne peut éprouver ma douleur), mais aussi *disparates* (elles varient sans loi d'un sujet à l'autre) et, en tout état de cause, *incommensurables* (elles demanderaient, pour être comparées, à être réunies dans une même conscience). Si la logique et les mathématiques avaient affaire à des constructions et à des représentations mentales, elles ne seraient donc pas des sciences, lesquelles sont des connaissances *publiques*, *communes* et *transmissibles*. Frege se proposait donc de séparer la logique de la psychologie et de la théorie de la connaissance : *commençons* par examiner la structure objective de ce qui est connu (les lois de l'"Être-vrai"), et nous verrons *ensuite* comment cette structure objective peut être appréhendée par l'esprit humain (les lois du "Tenir-pour-vrai")². Dans cette conception "absolutiste", la logique est donc concernée non par les représentations (*Vorstellungen*) mentales subjectives, mais par les Pensées (*Gedanken*) objectives, éternelles et indépendantes qui sont les porteurs des valeurs de vérité.

Une telle délimitation de la logique est assurément capable d'en assurer l'objectivité, mais elle rend la théorie de la connaissance tout à fait problématique : on ne voit pas en quoi pourrait consister le processus par lequel un individu en vient effectivement à "saisir" une Pensée, c'est-à-dire à se trouver en relation avec une entité abstraite, qui n'est ni un objet "réel" ni une représentation psychique. Il est certes bien clair que la logique telle que la conçoit Frege n'a rien à dire sur ce point, puisqu'une Pensée telle que celle que nous exprimons par la phrase "la Seine coule à Paris" doit justement pouvoir être considérée comme vraie même s'il n'existait aucun être humain capable de la reconnaître comme vraie, voire seulement capable de l'exprimer par une phrase. Mais la conception de la connaissance *implicitement* véhiculée par cette présentation de la logique ne semble guère en harmonie, *to say the least*, avec le projet même des "sciences cognitives". On est en somme devant la difficulté suivante : si nous voulons maintenir une conception "objective" de la *logique* (et nous y tenons, car il n'est pas question de renoncer à la rigueur mathématique qui lui a été pour la première fois conférée par les travaux de Frege), alors nous sommes apparemment contraints d'attribuer aux individus la capacité d'entrer en contact direct avec des entités non matérielles (ce qui n'est pas, loin s'en faut, la conception naturaliste que nous cherchons à favoriser en *psychologie*). Ceci définit deux stratégies possibles dans l'application des méthodes logiques aux "sciences cognitives".

1.1. La logique intensionnelle

La première stratégie s'accommode d'une psychologie non naturaliste, et traite donc la connaissance, la croyance et, de manière plus générale les *attitudes propositionnelles*, comme des relations entre agents et Pensées frégréennes. Dans ce

2. Cf. P.T. Geach, *Frege*, in G.E.M. Anscombe & P.T. Geach, *Three Philosophers*, Basic Blackwell, Oxford, 1961, p. 137.

cadre, la "saisie" de la Pensée que ϕ , c'est-à-dire la compréhension du sens de la phrase ' ϕ ', équivaut simplement à la connaissance de ses *conditions de vérité*, c'est-à-dire à la connaissance, parmi l'ensemble de toutes les situations possibles, du sous-ensemble $[[\phi]]$ (l'intension de ϕ) dans lesquelles ϕ est vraie. L'évaluation d'un énoncé comme 'X croit que ϕ ' ou 'X sait que ϕ ' met donc en jeu non seulement la valeur sémantique "actuelle" de ϕ (son *extension*), mais encore sa valeur dans une partie déterminée d'un certain espace d'indices (les *mondes possibles*). Une telle sémantique, qui reste évidemment *compositionnelle* au sens de Frege - l'intention d'une phrase complexe étant récursivement définissable à partir de l'intention de ses constituants immédiats en vertu de clauses comme $[[\phi \& \psi]] = [[\phi]] \cap [[\psi]]$ -, permet d'interpréter de manière particulièrement naturelle une multitude de systèmes formels (logiques épistémiques, doxastiques, etc) qui décrivent axiomatiquement les propriétés caractéristiques de divers types d'attitudes propositionnelles.

1.1.1. Une question controversée est celle de savoir si cette logique intensionnelle a quelque vocation à participer à une explication intégrée du comportement intelligent, ou bien si sa nature la destine à ne fournir qu'une sémantique formelle des langues naturelles, lesquelles comportent évidemment, de leur côté, des verbes d'attitudes propositionnelle. A une question sensiblement équivalente, Frege répondait lui-même qu'une Pensée peut agir, non pas certes causalement, mais "par cela même qu'elle est saisie et tenue pour vraie" ³. De fait, la sémantique des mondes possibles fournit un cadre tout à fait naturel pour l'étude de la prise de décision : une action peut être considérée comme une transformation de l'ensemble des mondes possibles, et une décision rationnelle comme le choix d'une action capable de transformer tout monde compatible avec les croyances de l'agent en un monde compatible avec ses désirs (tel est en effet le "noyau" de la théorie familière de la décision, qui s'obtient à partir de là par simple gradualisation de la notion de compatibilité avec les croyances (fonction de probabilité) et avec les désirs (fonction d'utilité)). L'intérêt des théoriciens des sciences sociales pour cette approche s'est grandement accru avec les développements les plus récents, qui concernent la possibilité de décrire rigoureusement les attitudes propositionnelles "croisées", du type "X croit que Y croit que ϕ ", lesquelles jouent un rôle tout à fait essentiel dans l'explication du comportement dans les *sociétés* d'agents. Comme le montre en effet le chapitre "*Mondes de croyance*" et *probabilités bivalentes*. Une interprétation logique du théorème de Mertens et Zamir, consacré par Ph. Mongin à un important résultat d'isomorphisme entre théorie des jeux et logique multi-épistémique, cette forme de logique intentionnelle constitue bel et bien un cadre approprié pour formaliser la délibération d'acteurs qui ont à considérer non seulement les effets potentiels de leurs choix, mais aussi les croyances entretenues par les autres au sujet de ces effets potentiels ⁴.

3. op. cit., p. 194.

4. On trouvera un panorama des résultats récents en ce domaine dans M. Bacharach & Ph. Mongin (eds.), *Logic and the Epistemic Foundations of Game Theory*, n° spécial de *Theory and Decision*, XXXVII-1 (1994). Cf. aussi B. Walliser, *Logique épistémique et théorie des jeux*, *Revue économique*, XLII-1991, p. 801-831.

1.1.2. La logique multi-modale en question trouve des terrains d'application dans des domaines d'une surprenante variété. Un exemple caractéristique est celui de l'informatique des systèmes distribués, dans lequel se pose crucialement le problème de la "cohérence interactive" ⁵, qui concerne la possibilité pour une collection de processeurs, dont certains sont éventuellement défectueux, de mener à bien une tâche collective comme la synchronisation. On a ici affaire à un cas typique dans lequel une certaine action ne peut être effectuée dans le groupe que si chacun possède non seulement une certaine connaissance relative à la Nature, mais encore une connaissance de la connaissance possédée par ses partenaires, et où le degré d'enchâssement des connaissances croît avec le nombre de cycles de messages échangés. C'est une observation de ce genre qui inspire la sémantique des "structures de connaissance" mise au point par Halpern, Fagin et Vardi dans les années 1980, et dont on trouvera une exposition détaillée dans le chapitre 8.

On sait par ailleurs, depuis l'ouvrage séminal de D. Lewis ⁶, que l'exécutabilité de certaines "actions collectives" (la possibilité pour un groupe d'atteindre une coordination satisfaisante) dépend de - et, dans certains cas paradigmatiques, se réduit à - la possibilité pour les agents du groupe d'atteindre un état de connaissance d'un type extrêmement particulier : la *connaissance commune*, c'est-à-dire l'état $C\phi$ dans lequel chacun sait que ϕ , chacun sait que chacun sait que ϕ , etc, avec un indice d'enchâssement *infini*. La littérature relative à cette notion défie désormais la recension, et s'étend à des domaines comme la philosophie politique (le propos initial de Lewis était de définir la notion de *convention*), la pragmatique linguistique (problème des conditions qui garantissent la convergence des locuteurs dans la référence des descriptions définies ⁷) et bien évidemment la théorie des jeux, dans laquelle la rationalité individuelle de chaque joueur est ordinairement supposée de connaissance commune dans le groupe. A côté de travaux relatifs aux conditions d'obtention d'un tel état de connaissance, et notamment aux limites éventuellement induites en ce domaine par la nature de la communication entre agents (l'acheminement des messages est-il garanti ? immédiat ? les messages sont-ils authentifiés ? oblitérés ?) ⁸, les propriétés proprement logiques de la nouvelle modalité C font aujourd'hui l'objet de recherches très actives. J. Barwise, par exemple, a suggéré ⁹ d'abandonner la définition itérative de C, qui fait appel à une conjonction *infinie*, au profit d'une définition *circulaire* : une situation dans laquelle ϕ est de connaissance commune est une situation dans laquelle chacun sait que ϕ , et dans laquelle chacun sait à quoi s'en tenir sur cette situation elle-même. Dans ces

5. Cf. par exemple M. Pease, R. Shostak & L. Lamport, *Reaching Agreement in Presence of Faults*, *Journal of the Association for Computing Machinery*, XXVII-2 (1980), p. 228-234.

6. *Convention*, Harvard U.P., 1968.

7. H.H. Clark & C.R. Marshall, *Definite Reference and Mutual Knowledge*, in A.K. Joshi, B.L. Webber & I.A. Sag (eds.), *Elements of Discourse Understanding*, Cambridge U.P., 1981

8. Cf. J.Y. Halpern & Y. Moses, *Knowledge and Common Knowledge in a Distributed Environment*, *Journal of the Association for Computing Machinery*, XXXVII-1990, p. 549-587.

9. J. Barwise, *On the Model Theory of Common Knowledge*, in *The Situation in Logic*, CSLI Lectures Notes, 17, Standford, 1988, p. 201-220.

conditions, $C\phi$ peut être défini comme un point fixe (en fait, le plus grand) solution de l'équation $X \equiv U(\phi \& X)$, où U est la modalité de connaissance universelle "chacun sait que ...". Dans le chapitre *Applications épistémiques de la théorie des ensembles anti-fondés*, L. Lismont montre que la manière la plus efficace d'interpréter une modalité ainsi définie est d'adopter une sémantique de type "situationnel", dans laquelle on renonce à l'axiome classique de fondation interdisant les cycles ou les descentes infinies dans la relation d'appartenance ensembliste.

1.2. Vers une logique hyperintensionnelle

Les considérations qui précèdent auront probablement convaincu le mathématicien ou l'économiste de tendance spéculative qu'il y a quelque intérêt à vivre dans l'air raréfié de la logique intensionnelle : la recherche des lois régissant une sociétés d'archanges capables de "saisir" au vol une Pensée frégeenne est un exercice subtil qui ne devrait pas exagérément les dépayser. En revanche, la lecture du chapitre consacré par D. Andler au thème général *Logique, raisonnement et psychologie* montre à l'évidence que les psychologues de terrain éprouveraient bien des difficultés à reconnaître leurs propres ouailles dans ces surhommes épistémiques. Des données expérimentales massivement attestées, comme celles relatives à la "tâche de Wason", montrent en effet que les compétences déductives des agents réels sont extrêmement modestes, tellement modestes, en fait, que la formalisation donnée par la logique intensionnelle apparaît comme une *idéalisation* outrancière, qui n'a rien de comparable avec le processus familier d'*abstraction* par lequel, par exemple, la mécanique élémentaire met entre parenthèses les forces de friction lorsqu'elle énonce les lois du mouvement.

Sans doute l'hypothèse d'une *omniscience logique* pleine et entière, dont le chapitre *Problèmes de logique de la connaissance*, rédigé par P. Gochet, détaille les aspects, n'est-elle pas impliquée par tous les systèmes de logique intensionnelle. Mais la sémantique topologique de Scott-Montague, qui est à cet égard le système le moins déraisonnable, engage elle-même à considérer comme substituables *salva veritate* deux énoncés quelconques de même *intension*. Or ce principe d'*équivalence intensionnelle* handicape à lui seul l'adéquation descriptive de ce type de logique, puisqu'il entraîne, par exemple, qu'un agent qui connaîtrait une seule vérité logique les connaîtrait toutes.

On touche ici un problème tout à fait central, dont la résolution correcte conditionne probablement la pertinence même de l'application des méthodes logiques en sciences cognitives : les contextes cognitifs sont *hyperintensionnels*, c'est-à-dire que les agents réellement observables sont sensibles à des différences entre énoncés que les systèmes raisonnant à une équivalence prouvable près sont pour leur part incapables d'enregistrer. Il y a là une difficulté à la fois monumentale et familière, puisqu'elle est, par exemple, substantiellement identique à celle que les théoriciens des sciences sociales entendent affronter dans leurs tentatives pour définir rigoureusement une notion de rationalité cognitive *limitée*, destinée à rendre compte de la relative inadéquation souvent observée entre les informations disponibles et les croyances entretenues. D. Vanderveken, auteur du chapitre consacré à *La quantification en logique des propositions généralisées*, enregistre

pour sa part en théoricien des actes de langage cette nécessité de mettre en œuvre une logique catégorisant les énoncés plus finement que ne le fait la logique intentionnelle, et met en avant une notion d'"implication cognitive forte", aux propriétés beaucoup plus strictes que l'implication logique usuelle.

En somme, les logiciens intensionnalistes sont à peu près unanimes à convenir que la description logique des opérations cognitives exige la mise au point d'instruments additionnels, capables d'expliquer que l'accès des agents aux Pensées frégeennes est médié par des représentations linguistiques qu'ils doivent identifier et sur lesquelles ils doivent calculer. Paradoxalement, ils rejoignent ainsi ce qui fait la préoccupation centrale d'un autre courant de recherches, dont le point de départ et la philosophie générale sont extrêmement différents.

1.3. Les logiques d'inspiration intuitionniste

L'intuitionnisme soutient que la logique n'a pas être objective au sens où Frege le voulait, c'est-à-dire qu'elle n'est pas destinée à décrire une réalité indépendante de l'esprit : elle a au contraire pour tâche de décrire une réalité pour laquelle le problème de l'accès ne se pose pas, à savoir les constructions mentales du sujet lui-même, en tant que ces constructions peuvent être objectivées, c'est-à-dire publiquement manifestées et communiquées sous forme de preuves. Comme le remarque J. Dubucs, dont le chapitre *Logique de la connaissance : un panorama* est consacré à une comparaison systématique des deux programmes, les phénomènes cognitifs ne sont donc plus abordés ici de l'extérieur, du point de vue de l'attributeur (observateur ou concepteur) qui *impute* une connaissance à un agent, mais de l'intérieur, du point de vue de l'agent qui *opère* les actions de vérification nécessaires à son obtention : renonçant à donner une définition *explicite* de la connaissance, on n'en propose plus qu'une caractérisation *inductive*, consistant à définir les actions qui vérifient un énoncé complexe sur la base de celles qui vérifient ses constituants.

La logique intuitionniste ne suppose pas que les agents soient toujours en mesure de justifier un énoncé ou sa négation, mais elle demeure toutefois une perspective idéalisée, qui admet leur capacité à de mener à bien toute procédure *effective*, sans limitation de temps ou de ressources physiques. On peut repérer les sources de cette idéalisation résiduelle en adoptant le format du calcul des séquents de Gentzen. L'intuitionnisme s'y obtient à partir de la logique classique par une certaine restriction dans les règles *structurales* de la logique, mais d'autres renoncements sont possibles. La logique linéaire, par exemple, accentue cette austérité en répudiant à la fois la règle d'affaiblissement et celle de contraction. En adoptant cette dernière restriction, c'est-à-dire en considérant que ce qui peut être obtenu avec deux occurrences d'une même formule ne peut l'être avec une seule, on donne un sens logique à l'idée de preuve ou de procédure (physiquement) *faisable*. De nouveaux connecteurs apparaissent, qui ne coïncident avec les connecteurs classiques qu'en présence de tout l'appareil des règles structurales dont on s'est défait. Dans le chapitre *Une application de la logique linéaire : la planification déductive*, M. Masseron utilise, si l'on peut dire, à rebours l'analogie entre les preuves et les actions exécutables, en mettant les connecteurs de la logique linéaire au service d'une analyse de l'action planifiée.

2. De la connaissance mathématique à la connaissance empirique

Une preuve mathématique est une justification catégorique et définitive : un théorème n'est pas *plus ou moins* prouvé, et lorsqu'il l'est *aucune* découverte ultérieure ne peut lui ôter ce statut. En dehors des mathématiques de telles justifications catégoriques et définitives n'existent pas. Par exemple, lorsque nous assertons, sur la base des informations disponibles, un certain énoncé universel H relatif à l'univers physique, nous sommes dans une situation qui peut être caractérisée de deux façons distinctes. D'une part en termes métriques de justification *partielle* : parmi les mondes possibles où sont réalisés les énoncés ε qui consistent les données, la proportion de ceux qui vérifient H reste inférieure à 1. D'autre part en termes non métriques de justification *défaissable* : ε peut être étendu à un ensemble $\varepsilon' \supset \varepsilon$ tel que ε' implique la négation de H . Ces deux manières de concevoir la connaissance empirique soulèvent des problèmes difficiles, en face desquels la logique classique, bâtie pour aborder les questions relatives aux fondements des mathématiques, nous laisse à peu près désarmés.

2.1. S'agissant de l'approche *métrique*, la logique probabiliste est une voie souvent explorée. En définissant une mesure de probabilités sur les modèles d'un ensemble d'énoncés (ou sur les atomes de l'algèbre engendrée par ces énoncés), on obtient une notion assez naturelle de "degré de confirmation" d'un énoncé par un autre : la probabilité conditionnelle. Mais cette analyse se heurte à des difficultés bien connues ¹⁰, comme celle qui concerne la nullité du degré de confirmation d'une hypothèse universelle par une conjonction quelconque de ses instances lorsque l'univers est infini. Par ailleurs, on peut difficilement se contenter d'une approche métrique de la connaissance dans le domaine empirique (certains énoncés doivent bien finir par être "acceptés" sur la base des données disponibles), mais il est malaisé de définir la relation "absolue" de confirmation $\Rightarrow$ dont la probabilité conditionnelle est censée mesurer le degré. Dans leur chapitre *Confirmation non probabiliste*, D. Zwirn et H. Zwirn abordent le problème à rebours, et justifient informellement un certain nombre de principes que $\Rightarrow$ devrait satisfaire. En raison de difficultés liées à la conjonction, ces principes ne peuvent être satisfaits dans une sémantique probabiliste ainsi construite. Ils peuvent l'être simultanément, en revanche, dans une sémantique reposant sur une distribution de *possibilités* du type exposé par D. Dubois et H. Prade dans le chapitre *Logique possibiliste, modèles préférentiels et objets conditionnels*.

En vérité, une telle situation n'a rien pour étonner, si l'on considère que la métrique de la certitude, de l'évidence et de la justification n'a aucune raison de coïncider avec la métrique probabiliste qui a fait ses preuves dans l'analyse des fréquences statistiques objectives : dans le domaine empirique, pas plus que dans le domaine mathématique, les lois du Tenir-pour-vrai ne sont celles de l'Etre-vrai. L'additivité des probabilités usuelles est l'exacte contrepartie du tiers-exclu rejeté par la logique intuitionniste : on suppose dans les deux cas que l'existence (ou le degré élevé) d'une justification de Non-A peut provenir de l'absence (ou du faible degré) de la justification de A. Cette critique est à l'origine d'une tradition

10. Cf. J. Dubucs, *Inductive Logic Revisited*, in *Philosophy of Probability*, Kluwer, 1993.

maintenant bien établie de probabilités "non standard", qui rejette le principe d'additivité. Les progrès les plus récents dans cette voie concernent la définition d'un nouvel objet logique dont ces probabilités déviantes Π donneraient la sémantique la plus naturelle. La caractérisation sémantique qui s'impose ($\phi \Rightarrow \psi$ exactement si $\Pi(\phi \& \psi) > \Pi(\phi \& \neg \psi)$) définit un conditionnel aux propriétés logiques remarquables, à l'aide duquel on peut envisager sur de nouvelles bases la question très controversée du raisonnement défaisable.

2.3. S'agissant de l'approche *non métrique*, l'idée séminale depuis les années 1980 est qu'à la différence du domaine mathématique, où ce qui compte est la conséquence classique (la conclusion doit être vraie dans *tous* les modèles qui avèrent les prémisses), la notion pertinente dans le domaine empirique est celle de conséquence "préférentielle" (nous sommes intéressés par les énoncés vrais dans les seuls modèles les plus "normaux" parmi ceux qui satisfont les prémisses, la normalité étant entendue par référence à la satisfaction de certaines "règles" expressives de généralisations plus ou moins stéréotypiques). Ici, le problème lancinant a été, dès le départ, d'ordre computationnel : les procédures proposées pour obtenir par calcul ces énoncés "préférentiellement valides" n'étaient pas même *effectives*, puisqu'elles faisaient appel à des tests de satisfaisabilité (leur effectivité eût supposé la décidabilité de la logique du premier ordre, dont on sait depuis 1936 qu'elle n'est pas réalisée). La syntaxe était au fond de la sémantique déguisée, puisque les preuves consistaient essentiellement à *scruter* les mondes préférés pour voir si telle ou telle formule y était avérée. Dans les axiomatisations les plus récentes (Gärdenfors et Makinson) de la notion de conséquence préférentielle, ces tests sont éliminés au profit de la prise en compte d'un *ordre* (par exemple l'"enracinement épistémique") sur l'ensemble des *formules*. Mais le gain en "réalisme" serait faible, s'il s'agissait seulement de "regarder" dans la relation d'ordre tout entière pour apercevoir si un couple donné de formules y figure ou pas. Dans le chapitre *Changements minimaux constructifs*, L. Farinas del Cerro et A. Herzig examinent justement les moyens de redonner une place au calcul, en s'abstenant de décrire *explicitement* cet ordre sur les formules.

G. Crocco, dont le chapitre *Approches logiques du raisonnement taxinomique* propose un panorama critique des solutions actuelles au problème des exceptions et des restrictions, porte sur la situation présente un diagnostic beaucoup plus réservé : la logique contemporaine s'est construite sur les ruines de distinctions anciennes, comme celle de la substance et de l'accident, distinctions qui, sous une forme ou sous une autre, sont requises par l'analyse du raisonnement taxinomique, mais dont la réappropriation exigerait de très profondes modifications dans la conception de la logique qui prévaut aujourd'hui.

3. Des langages formels aux langues naturelles

Prioritairement destinée à s'appliquer aux questions fondationnelles, la logique classique a en outre partie liée à des langages formels profondément différents des langues naturelles auxquelles ont affaire les agents humains dans l'essentiel de leurs activités cognitives. Cette situation, dont D. Andler montre à quel point elle est porteuse d'ambiguïté lorsqu'il s'agit d'examiner la compétence "logique" des individus - on peut toujours supposer que ces derniers maîtrisent d'*autres* notions

d'implication ou de conjonction que celles qui sont canonisées par la logique -, est en revanche une source d'inspiration réciproque pour la logique et la linguistique. On peut à cet égard discerner deux tendances dans les recherches actuelles sur la "logique des langues naturelles". La première fait fonds sur les régularités (ou les irrégularités) observées dans les langues d'usage, et tente d'élaborer les structures formelles qui pourraient les caractériser. A l'inverse, la seconde entreprend, à partir de langages formels extrêmement riches et généraux, au pouvoir expressif tout à fait considérable, de rendre compte de phénomènes linguistiques complexes : les langues naturelles deviennent alors de simples exemples d'implémentation de grammaires universelles.

3.1. Le chapitre consacré par J.-P. Delahaye à *Paradoxes sémantiques et programmation logique* est représentatif de la première démarche : sachant que les langues naturelles contiennent leur propre métalangage, et qu'il est donc possible d'engendrer en leur sein des paradoxes sémantiques - le plus connu touchant, bien sûr, à la valeur de vérité de la phrase "cette phrase n'est pas vraie" -, il examine s'il existe, dans le cadre des langages de programmation, des principes généraux permettant d'étendre des prédicats comme "être vrai" de manière maximale dans les limites de la non-contradiction.

Dans la même veine, J.P. Marquis remarque, dans le chapitre *Approximations et négations*, que la symétrie que l'on observe entre l'affirmation et la négation dans les langages logiques élémentaires est loin d'être vérifiée dans les langues naturelles (il y a plus de cas de fausseté pour la phrase "la masse de la terre est m" que pour sa négation). Ce phénomène peut être analysé de manière assez naturelle dans le cadre de la théorie des *topos*, et la théorie qui en résulte est applicable à la négation des universaux dans les langues naturelles.

J. von Benthem ¹¹ a noté l'existence d'une analogie entre d'une part les conditionnels comme *si ... alors ...*, qui expriment une relation entre l'ensemble des situations qui satisfont l'antécédent et l'ensemble de celles qui satisfont le conséquent, et d'autre part les quantificateurs dans les langues naturelles, lesquels expriment une relation binaire entre des sous-ensembles de situations. Le chapitre *Conditionnels, quantificateurs et inférences*, dû à S. Lapierre, développe cette analogie et présente une logique des quantificateurs *tout, tout ou au moins un et au moins la moitié*.

3.2. La tendance à appliquer une logique complexe à l'explication de la sémantique des langues naturelles est manifeste dans la contribution de F. Lepage, qui a en vue les phénomènes d'indéfinitude sémantique bien connus dans les langues naturelles (exemple : *l'actuel roi de France est chauve*). Le chapitre *Logique partielle et inférence valide* développe à cet effet une théorie des types propositionnels partiels, dans laquelle on construit une hiérarchie de fonctions croissantes par rapport à une relation d'ordre dont l'interprétation intuitive est d'être "plus défini", et dont la sémantique comporte deux notions de validité (qui coïncident en logique classique) : le *toujours vrai*, et le *jamais faux*.

11. *Essays in Logical Semantics*, Reidel, 1986.

Enfin le chapitre *Logique constructive et donkey-anaphore*, dû à G. Usberti, montre comment, dans le cadre de la logique intensionnelle de Montague, il est possible d'élaborer une logique d'inspiration intuitionniste qui permet de rendre compte de certaines infractions au principe de compositionnalité dans les langues naturelles (phénomènes de coréférence à l'extérieur de la portée d'un quantificateur).

4. Remerciements

Nos remerciements vont au programme Cognisciences du CNRS, au Conseil de recherches en sciences humaines du Canada, au Fonds pour la formation de chercheurs et l'aide à la recherche du gouvernement du Québec et au ministère français des Affaires étrangères, qui nous ont permis d'organiser en octobre 1992, à l'Université de Montréal, le colloque *Intelligence artificielle et logiques non classiques : le point de vue philosophique*, qui réunissait la plupart des membres du groupe *Philosophie, logique, cognition*, et qui a été le point de départ de ce livre.

PARTIE I

LOGIQUE ET RAISONNEMENT

Chapitre 1

Logique, raisonnement et psychologie

Les rapports entre logique et raisonnement ont longtemps été fort étroits. Ce commerce s'était établi à l'abri de la digue que Frege, Husserl et leurs successeurs avaient érigée pour protéger la logique de la psychologie : discipline normative ayant pour objet la manière dont l'homme doit raisonner, elle était en droit à l'abri de toute espèce d'enquête empirique sur la manière dont les humains raisonnent en fait, ou sur la manière dont ils apprennent à le faire. Pour faire bref, on pourrait parler d'une "ancienne alliance" entre logique, raisonnement et antipsychologisme.

Les psychologues, les pédagogues, les philosophes, les anthropologues et la sagesse commune, qui fournissaient mille témoignages sur la prévalence d'erreurs logiques et de modes de pensée illogiques ou alogiques, loin de faire peser une menace sur cette alliance, ne faisaient que renforcer sa légitimité. Elle était la gardienne d'un domaine particulier, celui de la rationalité.

Plus délicate était apparemment la situation créée par la nécessité, dans les sciences et dans certains domaines de la vie pratique, de recourir à des procédures inductives. Le calcul des probabilités et la recherche d'une logique inductive furent les principales et efficaces ripostes. Il y aurait beaucoup à dire sur les difficultés rencontrées d'une part pour fonder le calcul des probabilités, d'autre part pour découvrir une logique inductive – et il faudra y revenir même dans le cadre restreint du présent article. Mais il est évident que la psychologie "réelle" du savant, de l'assureur, du médecin ou du chasseur primitif n'était pour rien dans ces difficultés : l'"alliance" dont nous parlons n'avait rien à craindre de ce côté-là.

Tout a changé. Non pas précisément sous l'effet des premières recherches en intelligence artificielle et en psychologie cognitive, qui s'attachaient à l'explication et à la simulation des comportements humains en matière de résolution de problèmes : les heuristiques et les stratégies de l'IA première manière et du *human problem solving* dans le style de Herbert Simon et Allen Newell ne faisaient que reprendre, avec – du moins en principe – la rigueur expérimentale et la précision de la programmation en plus, les thèmes familiers de la découverte mathématique et scientifique et de la logique inductive. Mais c'est bien de la même région du savoir, à savoir les sciences cognitives au sens large que sont parties les attaques convergentes qui ont mis à bas l'ancienne alliance.

L'analyse complète de cette convergence nous entraînerait trop loin. Rappelons seulement ses facteurs les plus importants. *Primo* les difficultés rencontrées par l'intelligence artificielle première manière ont conduit de nombreux chercheurs du domaine à récuser en bloc la logique comme *organon* de la raison. *Secundo*, la prolifération de “logiques” développées pour la modélisation du raisonnement “naturel” ou scientifique, les difficultés rencontrées d'une part pour montrer qu'elles sont à la hauteur de la tâche, d'autre part pour leur donner une base formelle convenable, tout cela a conduit certains logiciens, dont Gilbert Harman, à mettre en cause, sur le plan proprement conceptuel, le lien entre logique et raisonnement, et à s'interroger avec une vigueur nouvelle sur la nature véritable et sur le statut cognitif de la logique. *Tertio*, l'émergence d'une solution de rechange possible à l'approche dite “symbolique” ou “formaliste” dans les sciences cognitives, solution fondée sur l'outil technique et conceptuel des réseaux de neurones formels, a donné du poids à l'idée que certains comportements peuvent sembler être déterminés par des règles alors qu'ils ne le sont pas : si la logique semble commander, en fait ou en droit, le raisonnement, il se pourrait donc qu'il n'en soit rien (ni en fait *ni en droit*). Enfin, un considérable ensemble de recherches relevant de la psychologie expérimentale et destinées à déterminer les aptitudes réelles de l'homme aux tâches déductives et inductives les plus élémentaires a mis en évidence des écarts systématiques entre les performances observées et les réponses dictées par les normes logiques. Ces résultats ont conduit certains auteurs à des conclusions très générales sur les aptitudes de l'homme en matière de rationalité. D'un autre côté, des théories ont été élaborées dans le but de rendre compte de ces résultats à partir des processus cognitifs qui sont effectivement mis en œuvre dans les tâches de raisonnement. Ces théories sont fortement divergentes et leur confrontation, sur le plan tant conceptuel qu'empirique, a donné naissance à un débat nourri. D'une complexité parfois déroutante, ce débat a un enjeu principal assez clair, qui n'est autre que le rôle que la logique joue dans le comportement humain.

Nous nous intéresserons dans le présent article principalement au dernier de ces quatre courants, même si tous sont au moins indirectement concernés ; le deuxième le sera plus particulièrement. Mais avant d'entreprendre l'examen de la psychologie du raisonnement, il peut être utile de la situer dans le contexte de ce qu'on pourrait appeler, par symétrie, la “nouvelle alliance”. Il s'agit du rapprochement de courants hétéroclites et encore mal soudés qui ont en commun de rejeter la logique comme ressource ou capacité cognitive fondamentale, soit dans les faits, soit même en droit. La “nouvelle alliance” est antilogiciste de deux manières : elle nie que la logique soit à la base d'une vaste famille de comportements, ceux que la tradition qualifie

de rationnels ; et elle nie que la logique soit *la norme* ou le *patron* d'une catégorie – ou d'une dimension – importante de ces comportements ; beaucoup plus modestement, la logique prendrait place dans un répertoire varié de techniques cognitives. “La” logique, à supposer qu’il convienne encore de s’y intéresser, serait d’ailleurs, de l’aveu même des logiciens vraiment modernes, mal nommée : il n’y a que *des* logiques, en quantité indéfinie, qui sont comme les outils d’une trousse de bricoleur (v. par exemple Gamut 1991). Encore n’est-on pas certain que ces logiques remplissent collectivement les fonctions dévolues, dans l’ancienne alliance, à la logique : nombre d’auteurs placent au cœur de la cognition des capacités telles que la manipulation de répertoires d’exemples, la reconnaissance de configurations, la construction d’images ou de modèles – bref, dans l’avant-garde de la nouvelle alliance on cherche à donner corps, de diverses manières, au rejet du *formel* : les mécanismes ou processus postulés ont en commun de faire intervenir, en un sens ou en un autre, mais toujours de manière décisive, un *contenu*¹.

Nous ne procéderons pas à l’évaluation systématique des positions de la nouvelle alliance, encore moins chercherons-nous à restaurer l’ancienne, puisque nous sommes acquis à certains des arguments qui ont été invoqués pour la discréditer. Notre tâche sera plus limitée, nous l’avons dit, et plus descriptive ; toutefois, nous serons amenés à insister sur ce qui nous apparaît comme des confusions de nature à compromettre une bonne partie des arguments de portée générale qu’on a cru pouvoir tirer de l’étude expérimentale du raisonnement quant aux aptitudes logiques et à la rationalité de l’homme. Cette attitude critique n’a rien d’original, et pour une bonne part le travail a été fait par d’autres. Notons cependant au passage qu’aucun des camps en présence n’a le monopole des confusions : ceux qui en dénoncent certaines à juste titre en commettent d’autres parfois aussi préoccupantes. C’est un véritable mystère que cette capacité que semble avoir la logique de faire tourner les têtes les plus solides. Puisse l’auteur n’avoir pas commis à son tour d’erreur plus grave que celles qu’il reproche aux autres !

Nous commencerons par examiner les résultats et les théories en présence dans le domaine de la déduction (§§1-2), en mettant l’accent sur la tâche de sélection de Wason. Puis nous procéderons à un examen critique de ces travaux, en nous en tenant en général à des arguments déjà présentés par divers auteurs ; cependant, nous nous appuierons sur un point de vue personnel sur les rapports entre l’implication logique, la conséquence logique et le “si” du langage naturel pour achever de rejeter complètement la thèse de l’incompétence logique des sujets naïfs (§3). Nous passerons alors à l’examen des expériences de Kahneman et Tversky sur le jugement en situation d’incertitude, paradigme de l’étude empirique du raisonnement inductif et nous évoquerons successivement les explications de Kahneman et Tversky fondées sur la notion d’heuristique, puis les critiques, par Gigerenzer et par Cohen, de leurs travaux (§4). En posant alors aux théories psychologiques du jugement en situation d’incertitude les questions que nous avons posées aux théories du raisonnement déductif, nous mettons en évidence d’abord les

1. Même des auteurs proches à beaucoup d’égards de positions classiques militent en faveur d’un certain retour au contenu ; v. Barwise 1986, 1987 et tout récemment Peacocke (1994).

insuffisances des premières, ensuite la douteuse pertinence de la théorie des probabilités pour l'explication du jugement sur l'incertain, enfin l'erreur qui consiste à mettre les deux sortes de phénomène sur le même plan, à l'image du couple philosophique déduction/induction : tout raisonnement fait intervenir, mais à des niveaux différents, inférences démonstratives et non démonstratives ; une perspective modulariste nous apparaît alors comme susceptible de fournir une bonne heuristique pour délimiter les aptitudes sous-tendant le raisonnement (§4). Nous concluons sur une réhabilitation de la logique classique qui laisse pourtant le champ libre à toutes sortes de processus non logiques.

1. Etude expérimentale et théories du raisonnement déductif élémentaire

Le paradigme de Wason

Depuis qu'un psychologue anglais, Philip C. Wason, se proposa il y a près de trente ans de revenir sur la question des aptitudes humaines fondamentales en matière de déduction, les anciennes certitudes en la matière vacillent. Des milliers d'études expérimentales ont été rapportées, des centaines d'articles et de livres s'interrogent sur les "comportements déductifs" de l'homme et sur leur explication en tant qu'ils sont le produit de certaines capacités cognitives. Or cette question était classique, et, semble-t-il, solidement prise en main, sinon complètement résolue, par l'école piagétienne. Mais Piaget est le représentant par excellence de l'ancienne alliance : il ne met pas en doute l'idée que la logique classique constitue la structure de tout raisonnement déductif correct, et que la psychologie n'a rien à y redire. Ce qu'il veut expliquer, c'est comment l'enfant en vient à *se saisir* de la logique. Wason est plus empiriste : il veut s'assurer que l'adulte possède bien les aptitudes déductives élémentaires que la tradition lui attribue, et pour cela il doit s'éloigner d'une autre tradition, celle de la psychologie expérimentale dans son ensemble qui, à l'instar de Piaget, tient pour acquise l'existence d'une aptitude logique fondamentale et se demande seulement comment elle est mise en œuvre dans la résolution de problèmes simples. Autrement dit, Wason porte un premier coup à l'ancienne alliance en osant descendre d'un cran dans l'élémentaire.

Pour cela, il imagine un dispositif qui ne tardera pas à constituer un paradigme, appelé la "tâche de sélection". L'expérience originale consiste à présenter aux sujets un jeu de cartes dont un côté porte une lettre et l'autre un chiffre, la seule contrainte étant que si une carte porte un A d'un côté, elle doit avoir un 4 de l'autre. Cette règle une fois expliquée aux sujets, on les place devant quatre cartes de ce jeu, posées à plat sur une table. Les symboles qui apparaissent sont A, D, 4, 7. On demande alors de déterminer par la pensée, et de dire, quelles sont les cartes qu'il est nécessaire de retourner pour s'assurer que le jeu obéit bien à la contrainte stipulée (si A d'un côté, 4 de l'autre). Wason pense déterminer ainsi le degré de maîtrise d'une aptitude logique fondamentale, comprendre l'implication logique

$P \supset Q$ ². Or il constate que loin d'obtenir à une très solide majorité la réponse correcte (à savoir "Il faut retourner les cartes portant le A et le 7, et celles-là seulement"), il ne trouve qu'un taux faible – de l'ordre de 10% – de bonnes réponses. L'expérience est exemplaire pour deux raisons : elle est robustement reproductible, donnant des résultats très voisins quelle que soit – dans certaines limites, nous y revenons dans un instant – la formulation exacte du test, et quels que soient les sujets, leur formation, leur âge, leur appartenance sociale, leur culture etc. ; et elle donne lieu à une répartition constante et caractéristique des erreurs (les deux réponses erronées "Carte A" et "Cartes A et 4" recueillant environ trois quarts des suffrages).

De nombreuses autres expériences ont été menées : vérification d'énoncés découlant déductivement de la description d'une situation ; construction ou vérification de tables de vérité ; jugement sur la validité d'arguments déductifs ; invitation à appliquer *modus ponens* ou *modus tollens*, ou encore le principe de transitivité, invitation à fournir la conclusion d'un syllogisme, etc. Il est remarquable cependant que, même si nul ne conteste l'importance qu'elles prennent dans leur ensemble, aucune d'elles n'ait attiré, en psychologie mais tout particulièrement en dehors d'elle, ne serait-ce qu'une fraction de l'attention réservée à la tâche de sélection. Les raisons de cette inégalité de traitement – abstraction faite d'effets sociologiques – sont intéressantes. La première est qu'aucun autre type d'expérience n'a paru longtemps autant que celle de Wason mesurer directement et, il faut y insister, d'une manière aussi robuste, une capacité logique absolument centrale. La deuxième est que la tâche de sélection donne lieu à une distribution caractéristique de réponses qui intrigue et appelle explication ; les tâches exécutées avec un taux de réussite très élevé intéressent peu les chercheurs, dans la mesure où elles ne semblent que confirmer une intuition commune (à savoir que les hommes sont parfois capables de raisonner logiquement), et celles qui ne sont accomplies correctement que par une très faible proportion des sujets sont manifestement inadaptées (puisqu'elles entrent en conflit avec l'intuition précédente). La troisième est que rapidement ont été découvertes des variantes de la tâche de Wason pour lesquelles les performances étaient très différentes, et sensiblement meilleures, ce qui a permis d'appliquer la méthode des différences pour faire, à partir de l'analyse des facteurs modulant la performance, des hypothèses sur les capacités cognitives sous-tendant le raisonnement "réel". Ce qui est instructif pour le non-spécialiste est que l'expérience de Wason soit apparue longtemps comme la seule à remplir ces conditions – est-il donc si difficile que cela de tester l'aptitude à la déduction simple ?

Il se trouve que – ironie de l'histoire – de nombreux auteurs contestent aujourd'hui que la tâche de sélection ait un rapport direct avec une capacité *logique* élémentaire quelle qu'elle soit, et certains doutent même qu'elle teste l'aptitude au

2. Il est vrai que cette expérience prend place dans une étude de la mise à l'épreuve d'hypothèses (*hypothesis testing*), ce qui n'est pas exactement la même chose que la possession d'un concept logique. Mais mettre à l'épreuve, cette possession, au niveau d'une maîtrise complète, revient à vérifier si le sujet est capable de s'assurer que le concept est instancié.

raisonnement ; ils insistent *a contrario* sur certaines expériences moins prisées, et soupçonnent un biais quasi idéologique³ dans le privilège accordé à la tâche de sélection. On assiste également à une remise en question des facteurs auxquels a été attribuée une influence cruciale sur les performances et qui ont nourri la réflexion sur la nature même du raisonnement. Mais ce retournement, du reste fort récent et qui est loin d'être unanime, ne nous dispense nullement d'examiner le paradigme de Wason et tout particulièrement les facteurs qui modifient la performance. En effet, l'état présent de la question est très largement déterminé par le débat suscité par ce paradigme ; que le ou les présumés qui lui ont donné vie connaissent un début de défaveur n'y change rien. D'ailleurs, bon nombre d'autres expériences sont encore interprétées comme allant dans le sens de ces conjectures – en particulier, elles semblent confirmer que, des quatre règles d'inférence simples auxquelles donne lieu l'implication, à savoir les deux règles valides, MP et MT et les deux règles invalides, AC (affirmation du conséquent) et NA (négation de l'antécédent), seule MP est traitée convenablement par une forte majorité des sujets. Bref, la tâche de Wason, même si elle est soupçonnée de particularités qui restreignent sa portée, demeure caractéristique de l'esprit des recherches actuelles.

Outre les médiocres performances obtenues sur la tâche originale, les résultats qui ont le plus marqué les spécialistes concernent l'effet bénéfique exercé sur les performances par certaines modifications de cette tâche. Ainsi, en lui donnant un certain type de contenu concret, ou en la situant dans un certain contexte, on obtient un spectaculaire renversement des proportions de bonnes et de mauvaises réponses. S'il s'agit par exemple de vérifier (le plus économiquement possible) que des règles telles que les suivantes sont respectées :

“Pour avoir le droit de boire de la bière, il faut avoir dix-huit ans ou plus”,
ou encore :

“Pour pouvoir être affranchi au tarif réduit, un pli ne doit pas être scellé”,
les sujets s'en sortent à plus de 60%. De même, si l'on donne au conséquent une forme négative :

“Une carte ayant un A d'un côté ne porte pas de 4 de l'autre”,

on améliore les résultats. Ainsi ont été postulés différents facteurs facilitants : le contenu, tout d'abord (le fait que les lettres et les chiffres soient remplacés par des propositions chargées de sens), puis plus particulièrement le caractère concret de la règle, la familiarité avec la situation ou avec les objets manipulés (“S'il pleut, je prends la voiture”), le contexte déontique (obligation/permission). On a aussi proposé l'existence d'une tendance, ou “biais”, qui pousse à choisir la réponse qui reflète les données de l'énoncé [*matching bias*] : un problème de la forme “Si P, alors Q” susciterait ainsi la (mauvaise) réponse “Carte P et carte Q”, alors que les problèmes de la forme “Si P, alors non-Q” susciteraient la *même* réponse, bonne

3. Déçus par ce que les premiers résultats expérimentaux (en matière de raisonnement tant déductif qu'inductif – v. §4 *infra*) semblait révéler sur les aptitudes véritables d'*Homo sapiens* en matière de raisonnement, et partant de rationalité, les psychologues auraient cultivé un style pessimiste et amer et donné la préférence aux expériences et interprétations qui les confortaient dans leur sombre vision. V. par exemple Lopes 1991.

cette fois, du moins si la négative apparaît de manière implicite. (La notion de biais joue un rôle central dans l'étude du raisonnement inductif et nous y reviendrons au §4.)

En fait, ces effets ne sont pas uniformément observés : ils peuvent se compenser mutuellement, ils peuvent disparaître dans certaines conditions, etc. Nous ne rentrerons pas dans ces considérations pour trois raisons : elles nous entraîneraient trop loin dans une voie dont il y a maintenant lieu de croire qu'elle est assez particulière ; elles ne sont pas nécessaires pour comprendre ce qui a motivé les théories du raisonnement qui ont été proposées à la lumière de l'ensemble des expériences ; enfin, nous sommes tentés de penser que l'explication générale de la tâche de sélection vient d'être fournie par Dan Sperber, Francesco Cara et Vittorio Girotto dans un texte non encore publié ; or cette explication, nous le verrons, fait appel à des idées assez différentes de celles qui viennent d'être évoquées.

Trois théories du premier ordre

A une théorie cognitive du raisonnement déductif, il revient d'expliquer (1) les compétences déductives de l'esprit humain, (2) les nombreuses erreurs qu'il commet, (3) la distribution de ces erreurs, qui n'est nullement uniforme ni, par exemple, le simple reflet de la complexité apparente de la tâche, et enfin (4) l'influence exercée par le contenu et par le contexte.

Par "théorie du premier ordre", nous désignerons ici une théorie cognitive du raisonnement déductif qui prétend rendre compte de cette capacité indépendamment de toute autre. Le terme, il faut l'avouer, est doublement trompeur, puisqu'il n'est pas habituel de l'appliquer, comme nous le ferons, à des théories du raisonnement, et que ces théories n'ont rien à voir avec ce qu'un logicien entend par là. Nous n'avons pas trouvé mieux pour distinguer ces théories de celles que nous appellerons "du second ordre" parce qu'elles font intervenir des considérations sur les conditions dans lesquelles s'exercent les capacités dont les théories du premier ordre veulent rendre compte, et sur le "matériau" sur lequel ces capacités opèrent. La distinction est du reste provisoire, elle est seulement utile pour se repérer dans le champ passablement chaotique ("fragmenté", comme le dit Jonathan Evans, l'un des spécialistes les plus lucides du raisonnement déductif) des recherches sur le raisonnement. A terme ne survivra, peut-on espérer, qu'une seule théorie qui intégrera les deux niveaux ou les deux sortes de mécanismes.

On distingue en général, pour les opposer, trois théories du premier ordre. (En fait, comme y a insisté Evans, elles ne sont pas rigoureusement comparables. Nous y reviendrons au § 3.)

a. La logique mentale.

Cette théorie attribue à l'esprit humain la possession de règles logiques de nature syntaxique, accompagnées le cas échéant de règles pour l'application de ces premières règles. Dans les exposés historiques du domaine, on fait remonter généralement à Piaget, voire à Boole la première idée de ce genre. Aujourd'hui, elle est défendue particulièrement, et testée expérimentalement, par Lance Rips et par Martin Braine et ses collaborateurs ; John Macnamara la développe (sous le nom de "psychologique") dans un contexte épistémologique très large où il tente de

redéfinir les rapports entre logique et psychologie. Actuellement, seule la partie propositionnelle de la théorie est pleinement développée.

Braine, auquel nous nous référons de préférence, a proposé un système de règles de type déduction naturelle, accompagné de règles de contrôle. Selon lui, ce qui explique notre capacité à effectuer des déductions est l'application de ce système logique interne à des représentations mentales des propositions à traiter. La logique mentale qu'il postule est non classique et incomplète : elle se veut "réaliste", c'est un "organe mental" de même nature que la grammaire universelle postulée par Chomsky. Donc même s'il s'agit d'une entité théorique qu'aucun microscope (ou "cérébroscopie") ne permettra jamais de voir, elle doit être conforme à ce que le jugement des psychologues tire de l'expérimentation et de la réflexion théorique d'ensemble. L'incomplétude n'est pas nécessairement un défaut, puisque c'est d'une capacité réelle, et non idéale, qu'il s'agit de rendre compte. Bref, la logique mentale est bien *une* logique, mais n'a pas exactement la même fonction que *la* logique des logiciens. A ce propos, deux importantes remarques s'imposent ici.

La première est que les règles de la logique mentale sont certes syntaxiques, donc ne prennent pas en compte le sens des propositions, mais seulement leur morphologie. Cependant, elles ne fonctionnent normalement que lorsque les représentations internes sont munies d'une interprétation. La déduction sur des propositions ininterprétées n'est pas de la compétence de ce système. Ce système, c'est la seconde remarque, doit selon Braine être compris comme une ressource *primaire* dont sont équipés tous les esprits normalement constitués, indépendamment de leur niveau d'instruction et notamment de leur compétence logique ou mathématique. Nous reviendrons au §3 sur certaines questions soulevées dans les deux dernières phrases, qui, loin d'aller de soi, résultent d'un effort de clarification qui s'est étendu sur une quinzaine d'années et n'est pas encore, loin s'en faut, compris et accepté par tous.

b. Les schémas pragmatiques.

Motivée par l'observation des effets facilitateurs des contextes déontiques sur la tâche de sélection, cette théorie rapporte nos capacités déductives à la présence dans le système cognitif de "schémas pragmatiques" qui se déclenchent, à la manière des systèmes de production de l'intelligence artificielle, dans certaines conditions et produisent une conclusion déterminée. Ce sont des *schémas* en un sens reçu depuis longtemps en psychologie cognitive : généralisations acquises par apprentissage, et dont la forme permet l'application à un nombre indéfini de cas nouveaux par instantiation de variables libres. Et ils sont *pragmatiques* parce qu'ils prescrivent la démarche appropriée dans les contextes créés par l'action.

Proposée dans le cadre de l'interprétation de la tâche de Wason par Patricia Cheng et Kenneth Holyoak, la théorie se ramène en pratique à la formulation des schémas dits de *permission* et d'*obligation*. On pourrait imaginer de l'étendre à d'autres situations au moyen d'autres schémas, mais pour diverses raisons, dont les faiblesses épistémologiques de la théorie sous sa forme présente, cette piste n'a pas

été poursuivie, du moins directement, par beaucoup de chercheurs⁴. Disons donc rapidement ce qu'est par exemple le schéma de permission. Il s'applique à toute situation obéissant à une règle conditionnelle de la forme "Pour qu'une certaine action puisse être accomplie, une certaine précondition doit être satisfaite" et se compose lui-même de quatre règles :

- (R1) Si l'action doit être accomplie, la précondition doit être satisfaite.
- (R2) Si l'action ne doit pas être accomplie, il n'est pas nécessaire que la précondition soit satisfaite.
- (R3) Si la précondition est satisfaite, l'action peut être accomplie.
- (R4) Si la précondition n'est pas satisfaite, l'action ne peut pas être accomplie.

Nous ne nous attarderons pas sur les obscurités de cette proposition. On voit bien que des clarifications sont nécessaires, ne serait-ce que pour éviter une régression à l'infini : les règles du schéma sont elles-mêmes, semble-t-il, des conditionnelles de la même nature que celle dont le schéma est censé dire comment elle est traitée. Contentons-nous de rapporter que l'on peut en donner une interprétation cohérente ; l'idée fondamentale étant que les règles du schéma ne sont pas des propositions, mais comme on l'a dit des règles d'action comparables aux règles de production au sens de Newell : elles sont immédiatement applicables ou non, selon les circonstances, et il n'y a rien à en déduire ou à en tirer lorsqu'elles ne le sont pas. Ce qui n'est pas le cas de la règle de départ, dont le schéma a précisément pour fonction de dicter les conséquences à tirer dans les différents cas de figure.

Insistons en revanche un instant sur ce qui sépare la théorie des schémas pragmatiques de celle de la logique mentale. D'abord elle postule non seulement que les mécanismes déductifs s'appliquent à des propositions interprétées (on a vu à l'instant que la logique mentale, en tout cas selon Braine, est dans le même cas), mais surtout qu'ils sont sensibles à l'interprétation ; en ce sens, la théorie de Cheng et Holyoak se veut non formelle, voire anti-formelle. En deuxième lieu, elle soutient que ces mécanismes sont activés par la reconnaissance de situations ayant une pertinence dans le contexte humain, plus particulièrement de situations à caractère social (permissions et obligations sont typiquement de nature collective). Enfin, elle suggère que ces mécanismes sont spécialisés : ils dépendent fortement du domaine auquel l'individu a affaire. Or ces trois idées exercent actuellement une très forte influence sur différentes écoles de pensée. La non-formalité inspire non seulement la théorie des modèles mentaux dont il va être question dans un instant, mais les travaux en sémantique et en philosophie du langage associés à la *situation theory* de Jon Barwise et John Perry, ainsi que tout ce qui se réclame, dans les sciences cognitives et l'intelligence artificielle, de la "cognition située". La sensibilité au contexte, et l'idée qu'un contexte est créé par une situation "écologiquement

4. On nous permettra à ce propos, dans un ouvrage dans lequel les psychologues interviennent peu et qui ne leur est pas particulièrement destiné, une remarque d'ordre sociologique : les psychologues sont une communauté plus nombreuse d'un ordre de grandeur au moins que celles des logiciens ou des philosophes, et ils produisent des publications à un rythme peut-être cent fois supérieur. Il est difficile dans ces conditions de se priver complètement, dans l'évaluation des différents programmes de recherche, de critères quantitatifs.

valide”, importante et familière pour l’organisme ou l’individu, autant de thèmes chers aux tenants du contextualisme, aux divers sens du terme, en particulier aux défenseurs de la “cognition sociale” ; nous en verrons un exemple au §3, celui de la théorie du contrat social de Lena Cosmides qui prolonge en un sens celle des schémas pragmatiques. Quant à la dépendance par rapport au domaine, sous l’étiquette de *domain specificity*, elle inspire un courant de recherche d’apparition récente mais qui s’enfle à vue d’œil (v. Hirschfeld & Gelman 1994).

c. Les modèles mentaux.

Le psychologue anglais Philip Johnson-Laird, élève de Wason et d’abord rallié, comme à l’époque la plupart des spécialistes du raisonnement, à l’idée de logique mentale, a proposé au début des années 80, pour expliquer le raisonnement syllogistique “réel”, une nouvelle théorie, dite des “modèles mentaux”, qui est désormais très largement majoritaire parmi les psychologues. Abondamment remaniée, raffinée, généralisée au fil des années, elle est donnée aujourd’hui par ses défenseurs comme la seule théorie qui soit à la fois conceptuellement complète et cohérente et capable de rendre compte de la plupart des résultats expérimentaux relatifs au domaine pris dans sa plus grande extension (qui comprend non seulement le raisonnement syllogistiques, pour lesquels la théorie a été d’abord proposée, mais tous les raisonnements relevant de la logique déductive, le raisonnement spatial, les raisonnements métallogiques impliqués dans les énigmes de type menteur, et même peut-être le raisonnement inductif au sens de notre §4 *infra*). Nous exprimerons au §3 quelques raisons de douter de ses qualités conceptuelles, mais comme dans le cas des schémas pragmatiques le plus important est d’abord d’en exposer les grandes lignes et surtout d’en saisir les motivations profondes.

Selon Johnson-Laird, le raisonnement ne fait pas intervenir de règles “logiques” : il ne consiste pas à appliquer des règles d’inférence formelles comme celles que postule la théorie de la logique mentale. Raisonner consiste au contraire à construire un modèle mental de la réalité décrite par l’énoncé du problème considéré et à “lire” sur ce modèle des informations nouvelles et utiles concernant la réalité en question. La construction du modèle procède elle-même en plusieurs étapes. Un premier modèle est construit à partir de la première indication de l’énoncé ; ce modèle est en général encore implicite à beaucoup d’égards : il reste “muet” sur bien des points ; ce sont les indications suivantes de l’énoncé (ou de son interprétation par le sujet) qui permettent de l’explicitier graduellement. Comme le processus consiste à combiner des données, il peut conduire à des incohérences. S’ouvre alors, du moins dans certains cas, une phase de vérification de la cohérence. Si une contradiction se révèle, une deuxième tentative est faite pour obtenir un modèle cohérent. D’ailleurs, au cours de la phase de construction, il a pu arriver que plusieurs possibilités apparaissent, donnant naissance à plusieurs modèles plus ou moins développés. Dans tous les cas, il vient un moment où le raisonneur a construit un ou plusieurs modèles dans lesquels il n’a pas détecté d’incohérence. La phase finale peut alors commencer. Elle consiste, s’il s’agit d’une épreuve de vérification de la validité d’une conclusion C, à “lire” sur le ou les modèles la réponse à la question “C est-elle vraie dans le modèle ?” ; et s’il s’agit d’une épreuve ouverte, dans laquelle on demande par exemple ce qui découle des prémisses, le sujet cherchera à deviner ce

qui est à la fois non trivial, non explicitement contenu dans les prémisses et vrai dans le ou les modèles construits.

Les tenants de cette théorie (évoquée ici de manière abstraite et générale et qu'il faudrait naturellement préciser et illustrer) lui voient deux avantages décisifs : d'une part elle rend raison de leur indéracinable conviction que le raisonnement naturel est un processus essentiellement sémantique ; d'autre part elle rend compte avec fidélité, selon eux, des différences de difficulté : les taux comparés d'erreurs reflètent le nombre de modèles dont la construction est nécessaire pour résoudre le problème.

Pour l'observateur critique en revanche, plusieurs questions restent entières, même après l'étude des exposés détaillés de la théorie : (1) En quel sens peut-on considérer que les opérations postulées sont élémentaires ? (2) En quel sens ces opérations sont-elles non logiques, sémantiques et informelles plutôt que logiques, syntaxiques et formelles ? (3) Quelle est la plausibilité de la théorie ? (4) Pourquoi faudrait-il qu'une théorie conçue pour rendre compte de la capacité à résoudre une famille de problèmes très artificiels (la construction ou la vérification de conclusions de syllogismes), quels que soient ses mérites en la matière, donne naissance par généralisation à une théorie du raisonnement déductif ?

2. Théories du deuxième ordre

Un bon nombre d'interrogations d'ordre conceptuel ou théorique ont été soulevées au passage. Pour commencer à y répondre, nous donnerons d'abord la parole à certains spécialistes du domaine qui, conscients des difficultés, ont voulu y parer en élaborant ce que nous avons choisi d'appeler, bien imparfaitement, des "théories du deuxième ordre". Ces théories sont d'ambition inégale et ne s'adressent pas toujours aux mêmes questions ; elles peuvent se compléter mutuellement ou se combiner à certaines variantes de certaines théories du premier ordre. Pour ne constituer qu'un échantillon, celles que nous allons évoquer sont représentatives, pensons-nous, des choix théoriques disponibles.

a. Théories de la rationalité limitée ; compétence et performance.

Deux voies s'ouvrent de prime abord devant celui qui rejette l'idée que les écarts observés entre la norme logique et les performances des sujets sont l'indice d'une déficience essentielle. La première est de rapporter ces écarts à une nécessité purement matérielle, celle qu'impose la limitation des ressources du système cognitif humain. Rien de déshonorant, plaide-t-il, à ne pouvoir satisfaire à des exigences illimitées, particulièrement lorsque les tâches considérées présentent une forte complexité algorithmique, ce qui est le cas, déjà, de la logique propositionnelle. Au contraire, les ruses déployées pour compenser ces limitations sont la marque de l'intelligence. Naturellement, il y a un prix à payer : il est précisément dans le risque d'erreur. Inversement, poursuit le tenant de la rationalité "limitée" ou "minimale", il y a de la part du théoricien une manière de stupidité à vouloir élaborer un modèle du raisonnement humain sans partir de cette donnée fondamentale qu'est la finitude : quel ingénieur ferait-il abstraction de la résistance de l'acier en concevant un avion, quel général oublierait-il dans son plan de bataille que ses hommes se déplacent à une vitesse finie, ont besoin de nourriture et de repos, etc. ? Herbert Simon, qui est,

on le sait, à la fois économiste et co-fondateur de l'intelligence artificielle et de la psychologie cognitives, a le premier fait valoir l'argument à propos de l'agent économique rationnel de la théorie économique classique. Le philosophe Christopher Cherniak, entre autres, a développé cette argumentation à propos de la cognition en général ; maints psychologues sont aujourd'hui tentés de l'appliquer spécifiquement au cas du raisonnement déductif⁵ – il leur semble évident que la logique classique ne peut tout simplement pas fournir un modèle même approché des capacités humaines en la matière, puisqu'elle exige de supposer des ressources calculatoires dont nous savons que l'esprit humain ne dispose pas.

L'argument ainsi présenté repose, semble-t-il, sur une confusion que nous dissiperons un peu plus bas. Tournons-nous alors vers la seconde possibilité.

Elle consiste à traiter la capacité de raisonnement à la manière dont Chomsky traite la capacité de langage : il y a un écart inéliminable entre le comportement linguistique observable d'un individu et le comportement théorique qui résulterait du déploiement de la base cognitive que le linguiste est fondé à postuler pour rendre compte de ce comportement. Pour le dire plus rapidement, la compétence du possesseur d'une langue ne détermine pas ni complètement ni exactement sa performance. Le philosophe Jonathan Cohen a le premier soutenu que les expériences de la psychologie du raisonnement ne prouvent, dans la mesure où elles sont significatives, rien de plus que l'existence d'un écart comparable à celui que décrit Chomsky ; et il tente de justifier cet écart par des considérations théoriques analogues. Macnamara et d'autres psychologues qui défendent la théorie de la logique mentale contre l'objection des erreurs systématiques adoptent la même stratégie. L'objection qu'ils doivent prévenir, tout comme Chomsky a dû le faire dans le cas du langage, est que cette stratégie risque d'être "immunisante" au sens de Popper : sans précautions théoriques, elle risque de paraître interdire toute invalidation empirique de la théorie de la compétence.

b. Evolution et contrat social.

On a évoqué plus haut les relativement bons résultats obtenus dans certaines situations – il s'agissait, dans la tâche de sélection de Wason, celles qui impliquent des dispositions telles que l'âge légal pour la d'alcool ou le droit au tarif réduit pour les plis postaux non cachetés. Pour expliquer ces phénomènes, Lena Cosmides a récemment proposé une théorie sociodarwinienne reposant sur trois idées principales : (i) pour l'animal social qu'est l'homme, les situations créées par l'existence d'un contrat social sont particulièrement importantes ; (ii) l'aptitude à gérer ces situations résulte d'un processus de sélection naturelle ; (iii) une telle aptitude repose en partie sur la capacité de faire respecter un contrat, et cette capacité repose à son tour sur l'aptitude à détecter les tricheurs. Sachant qu'un "contrat" a pour Cosmides la forme "Si l'on jouit d'un bien, il faut en payer le prix", le "tricheur" est pour elle celui qui jouit du bien sans en acquitter le prix. L'"algorithme de détection des tricheurs" dont nous serions équipés par la grâce de l'évolution permet donc, selon Cosmides, de repérer le jouisseur non payeur, c'est-

5. Voir par exemple Oaksford & Chater (1992).

à-dire de résoudre correctement la tâche de sélection correspondant aux situations de “contrat social”.

Laissant les spécialistes discuter de la validité empirique de la théorie, de ses rapports avec la théorie des schémas pragmatiques ou d'autres théories, de l'extension exacte de son domaine d'application, attardons-nous sur son originalité métathéorique. Voilà en effet une théorie qui premièrement, contrairement aux théories du premier ordre dont il a été question, n'a rien à dire sur la manière dont la tâche est accomplie *en général*, mais prétend expliquer *pourquoi* elle est exécutée correctement dans une famille spécifiée de cas et *comment* elle l'est ; et qui deuxièmement situe cette double explication en dehors du domaine logico-calculatoire : le *pourquoi* relève de la sociobiologie, le *comment* de la reconnaissance des formes. Bien entendu, le soubassement computationnel n'est pas nié : c'est bien un *algorithme* qui accomplit – qui explique que s'accomplisse – la tâche de détection du tricheur, c'est, on peut le présumer, un algorithme qui a préalablement permis d'identifier la situation contractuelle ; et c'est la présence effective, “implémentée” dans le système cognitif, de ces algorithmes dont rend compte l'explication darwinienne. Mais la fonction de l'algorithme n'est que de garantir la matérialité du processus : il n'y aurait pas un mot à changer dans la théorie si l'on découvrait demain que le rôle actuellement attribué aux algorithmes est en réalité tenu par le Saint Esprit⁶.

La théorie de Cosmides n'élimine pas seulement de l'explication d'une tâche de raisonnement la logique : elle élimine le raisonnement lui-même. Pour le dire de manière non paradoxale, elle est le premier exemple que nous rencontrons d'explication éliminative dans le domaine du raisonnement : ce qui *apparaît* comme un raisonnement élémentaire est *en réalité* autre chose. Voici du même coup posée la question de savoir si tout raisonnement élémentaire ne risque pas de subir un sort analogue.

c. La théorie heuristico-analytique.

Evans a la premier développé dans le présent contexte une idée essentielle : toute tâche de raisonnement se déroule en deux temps. Le sujet commence par sélectionner les aspects de la situation qui lui semblent mériter son attention ; cette première phase est dite *heuristique* parce qu'elle repose sur des mécanismes non logiques, préconscients, qui sont du même ordre que l'attention sélective ou qu'une forme de perception intellectuelle, et non de l'ordre de la délibération. Ayant ainsi formé sa représentation personnelle du problème, le sujet s'attache alors à le résoudre ; cette seconde phase est dite *analytique* parce qu'elle est de l'ordre du raisonnement logique : elle consiste à traiter, au niveau conscient et de manière cohérente, l'intégralité des éléments de la représentation obtenue à l'issue de la phase heuristique. Evans insiste sur le fait qu'il ne présuppose aucune théorie particulière de la phase analytique : sa conception biphasique lui paraît compatible

6. Je défends l'idée que les sciences cognitives jouissent d'une certaine autonomie à l'égard de l'hypothèse fonctionnaliste ou “computo-représentationnelle” (v. par exemple Andler 1994). On en trouve ici une illustration.

avec chacune des théories du premier ordre actuellement disponibles, comme avec diverses autres théories du deuxième ordre (par exemple celles qui veulent prendre en compte la complexité algorithmique de la tâche).

Le lecteur au fait de l'évolution de l'intelligence artificielle ou de celle de la linguistique s'étonnera peut-être que l'idée d'Evans n'aille pas de soi, ou du moins qu'il ait fallu longtemps pour qu'elle soit prise au sérieux. L'IA n'a-t-elle pas depuis longtemps mis au premier rang de ses préoccupations la "représentation des connaissances" comme facteur décisif dans l'exécution des tâches par l'application d'algorithmes ? Et la linguistique ne sait-elle pas au moins depuis Grice que la saisie d'un message met en jeu des processus pragmatiques qui interagissent avec les processus traditionnels du lexique, de la syntaxe, de la sémantique ? Or ce qu'Evans nomme heuristique, et qui conditionne le déploiement de l'analytique, n'est-il pas dans une large mesure l'homologue de la représentation des connaissances en IA et de la pragmatique en linguistique ? Peut-être les psychologues du raisonnement n'ont-ils tout d'abord pas songé qu'un tel facteur puisse entrer en jeu dans des tâches *élémentaires* : le passage de la présentation du problème qui est faite au sujet à la représentation qu'il en forme et sur laquelle il fait porter ses compétences proprement raisonnantes était vu sans doute comme une manière de copie. Si c'est le cas, ce n'est pas le seul piège que nous aura tendu l'idée d'élémentarité dans le domaine du raisonnement.

Quoi qu'il en soit, la conception d'Evans modifie considérablement la problématique. En premier lieu, elle jette un pont entre le raisonnement déductif et l'autre principale branche de la psychologie du raisonnement, appelée généralement raisonnement en situation d'incertitude et liée, comme nous le verrons au §4, au programme dit "heuristiques et biais" de Kahnemann et Tversky. L'heuristique d'Evans introduit – avec d'autant plus de force qu'elle est selon lui "préconsciente" – un élément d'alogicité, donc potentiellement d'irrationalité, dans le raisonnement déductif même (apparemment) élémentaire, et permet d'expliquer les erreurs systématiques de la même manière que dans le raisonnement incertain, c'est-à-dire comme des "biais" : par l'application inappropriée de stratégies de sélection des aspects pertinents du problème.

En deuxième lieu, la question de la nature de l'heuristique est posée : comment opère la sélection des aspects pertinents, et peut-on donner une définition de la pertinence qui ne soit ni triviale ni dogmatique, c'est-à-dire qui n'identifie la pertinence ni à ce en vertu de quoi une information est sélectionnée par les processus heuristiques, ni à ce en vertu de quoi la "bonne" réponse est obtenue ? Evans et d'autres ont proposé des réponses partielles à ces questions, mais on peut penser que les travaux de Sperber, Cara & Girotto rendent ces réponses largement caduques.

En troisième lieu, il revient à Evans le mérite supplémentaire d'avoir rendu possible de voir dans la tâche de sélection, longtemps considérée comme représentative du raisonnement déductif élémentaire, une exception ou un cas-limite : Evans conjecture en effet que cette tâche ne met en jeu que la phase heuristique. Il opère ainsi dans ce cas, comme Cosmides mais d'une autre manière, plus générale et peut-être mieux motivée, l'élimination du raisonnement d'une tâche de raisonnement.

d. L'approche pertinentielle de la tâche de sélection.

La phase heuristique d'Evans consiste, on l'a vu, en la formation par le sujet d'une représentation interne du problème. Dan Sperber, Francesco Cara et Vittorio Girotto ont appliqué la théorie de la pertinence de Sperber et Wilson à la description de ce processus dans le cas de la tâche de sélection. Cette théorie, on le sait, vise à rendre compte de la communication, plus précisément de la compréhension contextuelle des messages communiqués (principalement, mais non exclusivement, au moyen du langage). Trois objectifs sont atteints dans le travail de Sperber, Cara & Girotto : (i) Expliquer à partir de principes simples, clairement formulés et de portée générale l'ensemble proliférant des résultats empiriques obtenus dans la tâche de sélection, et réanalyser de manière lumineuse un grand nombre des explications "de second ordre" avancées jusqu'à présent. (ii) Prédire tout un ensemble de faits nouveaux, incompatibles avec les théories concurrentes, et les tester expérimentalement avec succès. (iii) Analyser le paradigme de Wason et la masse de ses variantes sur les plans logique, linguistique et cognitif avec une rigueur exemplaire, ce qui plaide, avec les succès expérimentaux mentionnés à l'instant, en faveur d'une extension de cette approche à d'autres problèmes que la tâche de sélection dont les auteurs soulignent, en accord avec Evans, le caractère très particulier.

L'idée qui préside à l'explication pertinentielle (au sens de Sperber et Wilson) de cette tâche est la suivante. La résolution du problème, comme l'a compris Evans, passe par la compréhension de la situation à l'aide de laquelle le problème est communiqué au sujet. Or cette situation est hautement complexe : elle comporte des aspects linguistiques, écrits ou oraux, mais aussi iconiques – par exemple les cartes, réelles ou représentées ; elle a surtout des composantes métalinguistiques (la structure du message complet – introduction, présentation des cartes, formulation de la question) et métathéoriques (cadre général d'une expérience de psychologie cognitive, rapports entre l'expérimentateur, omnipotent et omniscient, et le sujet, présumé apte à l'erreur, horizons culturels et épistémiques activés, ou non, au cours de la présentation de la tâche, stratégies savantes sollicitées ou non, stratégies spontanées encouragées ou découragées, etc.). L'interprétation par le sujet de cette situation, en particulier de ce que l'on attend de lui, est donc rien moins que triviale. Elle met en œuvre le "principe de pertinence" de Sperber et Wilson, principe qui comporte deux volets : $[\alpha]$ le sujet fait porter son attention sur les aspects de la situation qui sont pertinents, c'est-à-dire susceptibles de maximiser ses gains informationnels pour un coût de traitement minimal, et $[\beta]$ le sujet présume que l'information communiquée par l'expérimentateur est pertinente.

Nous ne rapporterons pas la manière dont cette idée est déployée dans l'analyse détaillée de la tâche de sélection, ni comment elle permet la reprise critique des explications antérieurement avancées. Disons seulement que selon Sperber, Cara & Girotto, les sujets infèrent de la règle, par ordre décroissant d'accessibilité, certaines de ses conséquences observables, et s'arrêtent lorsqu'ils jugent que les conséquences considérées suffisent à donner à la règle la pertinence qu'ils en attendent. Cela fait, ils choisissent les cartes susceptibles de mettre à l'épreuve ces conséquences. Comme l'accessibilité et la pertinence varient selon les versions, il n'est pas étonnant que les résultats puissent varier de manière spectaculaire de l'une à l'autre. Il faut en revanche insister sur la prédiction très générale à laquelle cette analyse conduit : toute expérience dans laquelle la conséquence "Il n'y a pas de cas

de **P-&-non Q**” est aussi accessible, et plus informative, que la conséquence “Il y a des cas de **P-&-Q**” conduira les sujets à choisir les cartes **P** et **non Q**. Cette prédiction prend la forme d’une “recette” pour fabriquer une version de la tâche de sélection sur laquelle on obtient d’excellents résultats :

- “- Choisir un couple de traits **P** et **Q** tels que le trait complexe **P-&-non Q** soit (ou puisse être rendu) plus facile à représenter que le trait complexe **P-&-Q**. ...”
- Créer un contexte dans lequel le fait de savoir qu’il existe des cas de **P-&-non Q** aurait de plus grands effets cognitifs que le fait de savoir qu’il existe des cas de **P-&-Q**.
- Présenter la règle “Si **P**, alors **Q**” sous une forme pragmatiquement adaptée.”⁷

La prédiction est corroborée par cinq séries d’expériences sur des versions inédites de la tâche de sélection. Elle permet également de réanalyser les explications de second ordre qui ont été avancées, soit pour les faire apparaître comme des cas particuliers, soit pour les rejeter : la théorie pertinentielle rendrait compte de la manière la plus générale de la phase “heuristique”, au sens d’Evans. En revanche, comme y insistent les auteurs, elle ne permet pas – du moins à elle seule – de départager les différentes théories du premier ordre, puisque la phase “analytique”, comme chez Evans, est selon la présente théorie réduite à néant. Ce n’est pas que le raisonnement ait entièrement disparu dans l’exécution de la tâche de sélection : il entre en jeu dans la mise au jour de conséquences (logiques) testables de la règle, et dans le processus de compréhension lui-même. Mais ce n’est pas la tâche de sélection, selon Sperber, Cara & Girotto, qui permet de trancher entre les différentes théories du raisonnement. La principale raison qu’ils avancent est que dans de nombreuses versions de la tâche, l’énoncé de la règle est pragmatiquement déviant : les mécanismes inférentiels à l’œuvre dans la compréhension sont alors pris en défaut et conduisent à des résultats sans grande portée (de même qu’il n’y a pas grand chose à tirer, quant au fonctionnement d’un moteur à explosion, de la constatation que l’ajout de sucre dans le réservoir à essence d’une voiture provoque une panne à retardement).

3. Questions théoriques

La situation dans laquelle nous nous trouvons au terme de ce parcours a quelque chose de singulier. Le principal effort des psychologues pour isoler une aptitude logique fondamentale, ou encore une capacité élémentaire de raisonnement déductif, semble avoir échoué. Les observateurs – philosophes, logiciens, informaticiens – pourraient donc estimer légitime de s’en tenir à leurs conceptions et préoccupations traditionnelles. Deux choses devraient les en dissuader : d’une part les psychologues poursuivent leur enquête empirique et les mettent au défi de juger, sur le plan

7. Traduit de l’article Sperber, Cara & Girotto (à paraître), fig. 6. L’expression “pragmatiquement adaptée” [*pragmatically felicitous*] renvoie à un ensemble de conditions portant sur les propriétés pragmatiques de l’énonciation linguistique en contexte de la règle.

théorique, du sens et de la cohérence de leurs conceptions du raisonnement ; d'autre part tout ce remue-ménage a obligé certains observateurs⁸ à reconsidérer et à reformuler des questions aussi anciennes que fondamentales : d'avoir labouré durement sans avoir – sans doute – exhumé le trésor espéré n'en a pas moins conduit les psychologues à modifier durablement le paysage conceptuel que nous contemplons tous. A ces deux raisons on ajoutera peut-être, à l'intention de ceux qu'elles laisseraient indifférents, la considération suivante : admettons que nous soyons justifiés dans notre tranquille assurance concernant la déduction ; sommes-nous aussi sereins pour ce qui est du ou des raisonnements non déductifs, de la ou des logiques non classiques ? et sommes-nous absolument certains que nos inquiétudes à cet endroit n'aient aucun retentissement sur nos conceptions de la déduction ? A tout le moins devons-nous admettre que l'étude (expérimentale ou théorique) de l'induction (nous prendrons ce terme en un sens très lâche, pour désigner de manière concise toute démarche, réelle ou idéale, faisant intervenir des inférences non démonstratives) prend appui sur nos théories de la déduction ; et plus celles-ci seront explicites et exhaustives, embrassant tout particulièrement les résultats et hypothèses des psychologues, mieux nous saurons mener notre enquête sur l'induction.

Quoi qu'il en soit, dès lors que nous acceptons d'examiner les travaux des psychologues et le débat qu'ils suscitent, nous ne pouvons éviter de poser à nouveaux frais tout un ensemble de questions épistémologiques des plus difficiles. Nous devons nous contenter ici de les effleurer, remettant à une occasion future leur examen approfondi. Elles peuvent être regroupées sous trois intitulés : logique et raisonnement ; bases cognitives du raisonnement ; logique, langage et pensée.

a. Logique et raisonnement

Comme on pouvait sans doute s'y attendre, la logique classique sort indemne de l'épreuve : aucun argument n'est avancé en faveur de son remplacement par un autre système. Ce qui est en revanche fortement mis en cause est son rôle dans la conduite du raisonnement. Pour le dire autrement, il n'existe pas d'autre candidat que la logique, classique à la dignité de théorie fondamentale universelle du raisonnement, mais les mérites de l'unique candidat sont sérieusement contestés.

En réalité, la déstabilisation de la logique entraîne fatalement celle du raisonnement : à partir du moment où le raisonnement n'est plus de la logique appliquée, diluée, contaminée par les imperfections du réel, qu'est-il donc ? Les réponses de Descartes et d'Arnault ne sont plus disponibles, pas plus que leur manière de faire coïncider, à la limite, description et prescription.

Néanmoins, dans la discussion présente, les participants ne mettent pas réellement en cause la notion de raisonnement : ils admettent qu'il existe, dans les cas simples de déduction qui sont ceux qui les intéressent, une notion claire et univoque de ce qu'est le bon raisonnement, la marche qu'il faut suivre ou en tout cas la conclusion correcte. Ils s'opposent en revanche sur le sens et les raisons de l'écart

8. Je pense notamment aux philosophes Jonathan Cohen, Alvin Goldman, Gilbert Harman ou Stephen Stich.

entre les réponses correctes et celles que fournissent les sujets d'expérience dans la résolution de problèmes ultra-simples de raisonnement déductif.

Harman, on l'a rappelé dans l'introduction, se place sur le terrain de l'analyse théorique. Le raisonnement, dit-il, n'est tout simplement pas l'application ou le reflet de la logique. Il va plus loin : pour lui la logique n'a pas de rapport privilégié avec le raisonnement – elle autorise des inférences inutiles et n'a pas les moyens d'imposer les inférences utiles ; pire, elle autorise des inférences que la raison peut préférer rejeter plutôt que d'accepter une conclusion inadmissible pour elle. Pour ces raisons, Harman n'est pas loin de voir dans la locution “raisonnement déductif” un oxymore⁹. En revanche, il considère avec sympathie l'hypothèse d'une “logique de base” qui serait la capacité psychologique en vertu de laquelle l'esprit manipule des pensées composées à l'aide des connecteurs logiques ; cette capacité reposerait sur la saisie des rapports d'implication immédiate et de contradiction immédiate. Il doute cependant que le genre de système de déduction naturelle qui peut être édifié sur cette base soit véritablement de nature logique au sens classique du terme : il n'est nullement évident, selon lui, que tout ce qui est immédiat en ce sens soit de nature logique au sens propre (il y a des immédiatetés spatiales, temporelles, etc. dont on ne sait si elles sont médiées par des axiomatiques spécifiques, ni même, estime-t-il, s'il est possible d'en décider).

Ce rappel des thèses de Harman permet de mesurer la distance qui sépare son point de vue de logicien-philosophe de celui d'un psychologue antilogiciste tel que Johnson-Laird. Les raisons que ce dernier avance pour contester à la logique un rôle, exclusif ou même partiel, dans le raisonnement, sont, on l'a vu, de nature empirique. Les réponses que donnent les êtres humains, assure-t-il, s'écartent très souvent des résultats dictés par la logique pour des raisons qui ne sont pas triviales (ce ne sont pas de simples erreurs d’“inattention”). Plus précisément, il rejette la solution théorique qui consiste à postuler d'abord un système logique de base, puis des mécanismes de performance qui viennent perturber les productions de ce système. Il estime avoir d'autant plus de raisons de rejeter une telle solution qu'il pense avoir identifié des processus non logiques qui permettent de rendre compte du raisonnement “réel”, ce qui à ses yeux prive la logique du dernier argument qu'elle pouvait évoquer, celui d'être le seul candidat en la matière.

Prises côte à côte, ou face à face, les deux positions que nous venons de caractériser nous mettent dans l'embarras. D'un côté, une analyse théorique solide mais dont les conséquences sur le plan psychologique restent entièrement dans l'ombre. De l'autre, une proposition apparemment assez claire sur le plan de la psychologie expérimentale, mais dont le statut théorique est loin d'être assuré. Elle semble en effet entrer en conflit direct avec deux évidences.

La première est que les processus “non logiques” postulés par le psychologue antilogiciste mettent en œuvre des aptitudes logiques complexes, plus complexes que la logique élémentaire, et qui semblent reposer sur elle. C'est criant dans le cas de la théorie des modèles mentaux, dont la plus sommaire description regorge littéralement de termes empruntés à la logique et qui désignent des sous-processus

9. “Obviously, there is deductive argument, but it is not similarly obvious that there is deductive reasoning.” Harman (1983), p. 6.

de nature au moins partiellement logique. La seconde est que l'être humain a bel et bien recours à la logique pour raisonner déductivement (ou, si l'on préfère avec Harman éviter l'expression, d'argumenter déductivement, c'est-à-dire de résoudre des problèmes nécessitant la recherche ou la vérification de preuves déductives) : c'est à quoi les mathématiciens, entre autres, consacrent une bonne partie de leur temps.

Ces deux conflits obligent à poser une série de questions, qui toutes visent à mieux délimiter ce sur quoi, sous le nom de "raisonnement", sont censées porter les explications en présence.

Première question : de quel sujet parlons-nous ?

La référence au mathématicien pourrait être considérée comme sophistique : n'est-il pas un expert en matière de logique, et ne pratique-t-il pas une forme de raisonnement tout à fait particulière ? Pour ces deux raisons, ses aptitudes ne se placent-elles pas nettement en dehors de la sphère examinée ?

Une première réponse consiste à contester les attendus de la question. D'abord il n'est juste de considérer le mathématicien comme un expert logique qu'au sens où on peut lui attribuer la logique infuse, c'est-à-dire une connaissance implicite de la logique. En effet, il ignore en général les règles d'inférence de la logique formelle ; néanmoins on peut plaider qu'il pourrait sans mal les retrouver, en tout cas qu'il en reconnaîtrait immédiatement la légitimité si on les lui présentait ; on peut dire aussi que la logique formelle contemporaine (pas celle d'Aristote) est avant toute chose une théorie du comportement déductif du mathématicien. Pour autant l'antilogiciste ne prétendra pas que le mathématicien n'emploie pas les ressources de la logique dans ses raisonnements (typiquement il procède par *modus ponens*, *modus tollens*, *reduction ad absurdum*, démonstration par cas – élimination de la disjonction –, pose puis congédiement d'hypothèses, etc. : bref, il déploie une série de procédés dont la déduction naturelle, au sens de la logique mathématique, a justement pour mission de fournir le modèle). Par conséquent l'antilogiciste doit exclure le mathématicien, non parce que celui-ci ne fait pas référence aux règles logiques, mais du seul fait que son comportement déductif est conforme à ces règles et que ces règles en rendent compte exhaustivement. Bien entendu, le mathématicien est également membre de l'espèce *homo sapiens*, donc relève des schémas explicatifs généraux relatifs aux capacités cognitives de l'espèce. Mais *en tant que* mathématicien, c'est-à-dire en tant qu'il se soumet, au prix d'un entraînement très particulier, à des règles de comportement déductif non naturelles, il est une exception que le psychologue peut estimer légitime d'écarter.

Cependant, le mathématicien est loin d'être le seul auquel il arrive d'appliquer des règles d'inférence logique. Il est caractéristique, au contraire, des comportements¹⁰ qualifiés de rationnels au sens le plus strict, d'être gouvernés par

10. Comportement intentionnel, s'entend, et décrit comme tel ("individué" en partie par les intentions qui le gouvernent). Il est clair qu'en tant qu'événement du monde physique, tout comportement admet une description à laquelle la logique s'applique sans faille : vivre deux vies est aussi bien vivre un nombre de vies égal à la racine carrée de 4 (je remercie J. Dubucs pour cette observation).

la logique, au sens non d'une référence explicite, mais d'un respect manifeste : le vigneron qui sait que pour qu'il devienne riche, ses vignes doivent être atteintes de pourriture noble et qui, constatant qu'elles ne le sont pas, en conclut qu'il restera pauvre, respecte *modus tollens* sans connaître le nom du principe qu'il respecte. La différence avec le mathématicien est qu'il n'a pas fait l'objet d'un entraînement susceptible de modifier ses aptitudes spontanées, et que, sous peine de frapper d'invalidité *a priori* la discussion dans son ensemble, on ne peut supposer que la logique rend compte "par construction" de son comportement. Le vigneron ne peut donc être écarté de la même manière que le mathématicien dans l'exercice de ses fonctions.

Nous devons donc envisager de restreindre la discussion à certaines aptitudes d'un sujet que nous appellerons *logiquement naïf*. Le sujet logiquement naïf (SLN) manifeste sa qualité par son incapacité à aborder un problème de nature déductive comme le ferait l'étudiant au cours de logique. Il doit cette qualité soit à une ignorance, implicite et pas seulement explicite, des règles de la logique, soit à une impossibilité provisoire de mettre en œuvre ses éventuelles compétences en la matière. Encore faut-il comprendre son "ignorance implicite" de la logique en sorte de ne pas préjuger de la question de savoir si son aptitude est ou non "fondée" sur la logique.

Il est essentiel d'observer que, pour logiquement naïf qu'il soit, notre sujet possède au moins une *certaine* compréhension des particules de la langue correspondant aux constantes logiques (nous serons amené à y revenir), et une *certaine* maîtrise de la notion de conséquence logique – faute desquelles ledit sujet ne serait pas à même de *comprendre* le problème posé. Il possède ainsi une aptitude qu'on pourrait être tenté de caractériser à l'aide du mot "logique", convenablement qualifié. Mais pour éviter de laisser croire que l'argument en faveur de cette aptitude rend oiseuses les questions que se posent les psychologues, ou y voir au contraire une pétition de principe, appelons-la "aptitude déductive primaire" pour l'opposer aux aptitudes secondaires qui résultent vraisemblablement de processus d'apprentissage culturels et individuels longs et complexes et sont fondées sur les capacités deductives primaires ainsi que sur d'autres ressources cognitives telle l'utilisation d'artefacts (écriture etc.)¹¹.

Deuxième question : quelle aptitude s'agit-il d'expliquer ?

L'objectif des psychologues qui s'inscrivent dans la tradition de Wason est de caractériser la capacité au raisonnement déductif du SLN, et leur méthode repose sur l'idée que les raisonnements étudiés en psychologie cognitive dans la tradition du *human problem solving* de Newell et Simon sont trop complexes pour permettre d'accéder aux bases de cette capacité. Ce qu'il faut d'abord comprendre, c'est le raisonnement élémentaire. Mais qu'entendre par là ? On a vu combien la tâche de sélection en est éloignée, contrairement aux espoirs de Wason. On pourrait examiner d'autres paradigmes – raisonnement conditionnel, complétion de tables de vérité, raisonnement disjonctif, syllogismes etc. Dispensons-nous en et allons droit au but :

11. On trouve dans Braine (1990) une distinction analogue entre "primary" et "secondary deductive skills".

les seuls candidats plausibles ne sont-ils pas les règles de la déduction naturelle ? Les tâches élémentaires sur lesquelles tester les sujets sont donc de la forme : “Vous savez que P implique Q. Maintenant, vous apprenez que P. Qu’en déduisez-vous ?”, ou encore “Vous savez que P et Q. Pensez-vous que P ?”

Or de deux choses l’une : ou bien les sujets réussissent massivement dans ces tests, ou bien ils échouent avec une fréquence non négligeable. Dans le premier cas, le psychologue antilogiciste n’a plus de grain à moudre ; dans le second, la première hypothèse à considérer est que les particules du langage naturel qui dénotent des constantes logiques ne prennent pas pour les sujets, dans le contexte de la tâche, le sens de ces constantes. Il se trouve que c’est, du moins pour la plupart des règles¹², la première hypothèse qui est la bonne, ce qui nous dispense d’envisager ici la seconde, mais nous y reviendrons. Répondons à une objection que le second de nos exemples pourrait susciter : il ne serait pas étonnant, dira-t-on, que plus d’un sujet refuse de répondre “oui” à la question posée. En effet : elle est stupide. Pour mettre à l’épreuve la capacité de nos sujets à effectuer une élimination de la conjonction, il faudrait imaginer une formulation pragmatiquement acceptable. On pourrait par exemple s’assurer que, sachant que P et Q, le sujet se comporte infailliblement, *ceteris paribus*, comme s’il savait que P (par exemple qu’il croque ce qu’il a de comestible dans la main chaque fois qu’il a envie de croquer un bonbon, et alors qu’il sait qu’il a dans la main un bonbon et un cure-dent).

Deux difficultés sont ainsi apparues à la faveur de la dernière partie de la discussion. La première est que la recherche du raisonnement élémentaire, dès lors qu’elle est guidée par la conception traditionnelle, donc essentiellement logique, du raisonnement, semble fatalement conduire aux règles de la déduction naturelle, et que ces dernières ont peu de chances de donner lieu à des données empiriques permettant d’asseoir une conception antilogiciste du raisonnement. Mais quel autre guide que la logique suivre pour découvrir le raisonnement élémentaire ? On est tenté à ce point de poser une question à la Cohen¹³ : la non-logicité du raisonnement élémentaire peut-elle être établie par l’expérience ?

La seconde difficulté vient de ce que la validité de toute expérience sur le raisonnement élémentaire repose sur la combinaison de deux exigences : la saisie consciente par le sujet du problème posé (et non d’un autre problème), et la mise en œuvre spontanée d’une stratégie de résolution qui ne peut être consciente en tant que telle, dès lors que l’élémentarité de la tâche interdit que cette stratégie consiste en l’application délibérée d’une suite d’étapes. Or ni l’une ni l’autre de ces exigences n’est facile à satisfaire, et leur conjonction encore moins. A l’appui de ces remarques, rappelons, pour la première, les défauts que l’analyse de Sperber, Cara & Girotto met en évidence dans tout le corpus expérimental post-wasonien, et pour la seconde l’obscurité dont s’enveloppe la théorie des modèles mentaux : il semble implausible que le processus de construction puis d’étude des modèles soit inconscient ; mais il n’est pas davantage plausible qu’il soit conscient (Johnson-

12. Comme on l’a dit à propos de la logique mentale proposée par Braine, la possibilité que la logique intuitive ou spontanée (à quelque niveau qu’on veuille la situer) soit plus faible que la logique classique ne pose pas de difficulté conceptuelle ; elle soulève au contraire des questions empiriques intéressantes.

13. Cf. le titre de Cohen (1981).

Laird n'étudie pas de protocoles de sujets, et n'apporte aucune autre donnée susceptible de soutenir l'hypothèse de processus conscients ; quant à l'introspection, elle plaiderait fortement, sauf peut-être chez un partisan convaincu de la théorie, contre l'interprétation consciente).

Troisième question : à quel niveau prétendons-nous rendre compte du comportement du SLN ?

Cette question peut se comprendre d'abord à la lumière du problème qui vient d'être soulevé, celui de la "saisie" de la tâche par le sujet et du caractère conscient de la stratégie de résolution. Mais elle est incontournable dans toute entreprise s'inscrivant dans le projet des sciences cognitives : dès lors qu'il est admis que la richesse des explications computationnelles ou informationnelles réside dans la possibilité de découpler au moins deux niveaux d'explication, et souvent en fait davantage, il est essentiel de dire à quel niveau se situe l'explication que l'on propose. Nous en arrivons ainsi à la deuxième partie de notre discussion.

b. Bases cognitives du raisonnement.

A ce point, le lecteur éprouve peut-être un sentiment de grande confusion. Il pouvait lui sembler au départ qu'il était placé devant une question bien posée, celle de départager le psychologue antilogiciste de son adversaire, partisan de la logique mentale. Abstraction faite des difficultés que nous avons cru détecter plus haut, nous pouvons formuler ainsi ce qui les sépare : selon le premier, le sujet résout le problème "à l'aide" de X, où X est (i) effectivement mis en œuvre, (ii) non consciemment invoqué, et (iii) tout différent de la logique ; le second affirme (i), il peut accepter (ii), mais il remplace (iii) par un (iii') de la forme "X est un certain système de déduction naturelle, éventuellement affaibli, et augmenté de règles procédurales". Et si la question peut maintenant paraître mal posée, cela tient, pour simplifier et comme on vient de voir, à deux raisons principales : l'une est que l'on n'a aucune vision claire de ce qu'il faudrait prendre comme "problème" à faire résoudre, l'autre que le statut de X est obscur.

L'enjeu est alors le suivant : on voudrait d'une part se faire une idée plus claire de la forme que pourrait ou devrait prendre une théorie cognitive du raisonnement ; on voudrait d'autre part expliquer les phénomènes mis au jour par les psychologues, et comprendre ce dont les théories qu'ils avancent sont des théories.

Le second objectif est un exemple de "recherche inverse" dans lequel, après avoir montré que la réponse proposée ne répond pas à la question posée, on cherche à déterminer la question à laquelle elle répond. L'élucidation par Sperber et collab. de la tâche de sélection en apporte une illustration. Il faudra étendre leur démarche à d'autres types d'expérience, ce qui dépasse largement et notre propos et nos compétences.

Quant au premier objectif, s'il reste hors de portée, il n'est pas interdit de tenter d'y contribuer en examinant la question du statut, en particulier du niveau, du fameux système X. Nous nous contenterons ici, à titre d'ébauche de ce programme ambitieux, de suggérer une distinction tripartite et d'en déduire une caractérisation très générale des difficultés rencontrées dans le domaine que nous avons examiné.

Personne ne peut raisonnablement nier que l'être humain mette en œuvre, dans toute démarche de raisonnement, une capacité fondamentale qui partage avec la logique au sens habituel certaines propriétés essentielles, à commencer par la généralité, dont, remarquons-le, la formalité n'est qu'une espèce particulière. On a vu, par exemple, que pour déclencher et mettre en œuvre la stratégie conjecturée dans la théorie des modèles mentaux, il fallait disposer de ressources logiques en ce sens. Pour prévenir certaines objections, précisons que nous refusons la qualité de raisonnement aux processus quasi perceptifs ou réflexes – ceux (par exemple certains partisans du connexionnisme) qui pensent que tous les processus cognitifs sont de cet ordre, ou s'y ramènent au niveau de description approprié, maintiendront leur désaccord ; mais les psychologues dont il a été question ici ne sont pas du nombre ; au contraire, ils sont tous acquis à la doctrine cognitiviste ; Johnson-Laird en particulier est l'auteur d'une introduction aux sciences cognitives¹⁴ qui est un véritable manifeste du programme en question (il tient en particulier que la cognition s'explique comme un ensemble de processus *formels*, ce qui peut, mais ne doit pas, étonner de la part d'un aussi ardent défenseur de l'idée que le raisonnement n'est *pas* formel, mais, comme il dit, *sémantique*).

Nous admettons donc en premier lieu, pour les besoins de la discussion, l'existence d'une capacité cognitive, ou psychologique, qui, étant à la fois à la base de toute autre capacité postulée et de nature au moins quasi logique, mérite d'être qualifiée d'*Urlogique* [UL]. C'est là que se situe en particulier ce qui permet au sujet logiquement naïf de saisir les notions de conséquence et de contradiction logique.

En deuxième lieu, nos psychologues étudient le comportement des sujets humains face à des problèmes de nature logique. Ce comportement est gouverné par une capacité cognitive que nous pouvons nommer par souci de brièveté *résolution logique* [RL], puisqu'elle est ce en vertu de quoi des problèmes logiques sont (plus ou moins bien) résolus, sans pour cela présupposer qu'elle possède elle-même un caractère logique, moins encore exclusivement logique.

Enfin, nous devons postuler une troisième capacité qui est celle dont est armé, comme nous l'avons vu, le sujet logiquement naïf et lui confère ce que nous avons appelé, nous inspirant de Braine, son aptitude déductive primaire. Cette capacité mérite donc l'appellation de capacité déductive de base [CDB]. Remarquons, sans nous y attarder, que CDB pourrait aussi être vue comme une "logique naïve", ou populaire, ou encore spontanée – ce qu'on nommerait en anglais *folk logic*. Elle concerne en effet l'attitude spontanée du sujet naïf devant les pensées comprenant des chevilles correspondant en un certain sens aux constantes logiques.

À l'aide de ces trois notions, voyons si nous pouvons proposer des formulations simples des positions en présence. Commençons par ce que les différentes écoles de psychologie cognitive ont en commun. Elles admettent toutes que RL est logiquement défectueuse ; elles présupposent toutes nécessairement UL, avons-nous

14. The Computer and the Mind, Johnson-Laird (1988).

soutenu, sans généralement le dire, ni pour certaines être disposées à l'admettre¹⁵ ; enfin, elles voient dans CDB l'enjeu empirique essentiel.

Venons-en aux différences. Les théories du premier ordre portent sur CDB et en donnent des descriptions différentes et, dans certains cas et sur certains aspects, incompatibles. Les théories du second ordre sont avancées dans le but de rendre compte des conditions dans lesquelles CDB est mobilisé par RL ; certaines sont de forme négative ou limitative – ainsi les considérations de complexité¹⁶, de facteurs de performance ; d'autres introduisent un processus X qui entre en interaction avec CDB pour produire RL – ainsi l'heuristique d'Evans ou l'interprétation pertinente de la tâche chez Sperber et collab. (il arrive pour certaines tâches, on l'a vu, que X fasse à lui seul tout le travail).

Sur le plan conceptuel, les psychologues poléminent et s'accusent de confusions¹⁷ ; les philosophes et les logiciens¹⁸ hésitent à entrer directement dans le débat, décidément trop confus à leurs yeux. Macnamara parmi les premiers, Harman parmi les seconds, font des contributions positives. Harman a réfléchi sur UL et fourni des raisons de douter que CDB ait la forme d'un système logique. Macnamara, partisan de la logique mentale (qu'il appelle psycho-logique), s'est interrogé sur son statut : si l'on postule en quelque sorte franchement une réalisation de la logique dans l'appareil cognitif, la distinction entre CDB et UL ne s'impose pas avec la même évidence que si l'on rejette un tel postulat ; mais la notion même de réalisation oblige à s'interroger sur le soubassement cognitif de la psycho-logique. Ce soubassement, Macnamara le situe dans un langage de la pensée comme le conçoit Fodor : c'est lui qui jouerait le rôle de UL, CDB étant la psycho-logique.

On commence ainsi peut-être à entrevoir la possibilité de restructurer le débat et de surmonter au moins en partie ce que Evans appelle la "fragmentation" du domaine. L'objectif commun serait une théorie intégrée des capacités UL, CDB et RL. La difficulté théorique est néanmoins considérable : il ne s'agit pas seulement de déterminer la structure de chacune de ces capacités ; il faut aussi, et même, logiquement parlant, avant tout déterminer leur *statut*. La difficulté empirique n'est pas moins redoutable : nous semblons confrontés à un cas grave de problème de

15. Le titre d'un article de Johnson-Laird est significatif : "Reasoning without logic" (1986).

16. Nous disons que ces considérations semblent reposer sur une confusion. Il est exact qu'un système computationnel donné se heurte à une barrière de complexité. Mais rien ne dit que les tâches trop complexes entrent dans son domaine d'application naturelle. Dans le cas présent, on ne voit pas pourquoi CDB devrait résoudre des problèmes difficiles (ces problèmes relèvent des *secondary skills* au sens de Braine, et l'homme semble plutôt *bien* équipé pour les résoudre ! La situation est toute différente dans le cas de la vision, par exemple : nous savons que des tâches de haute complexité incombent au système visuel primaire ; c'est bien pourquoi les descriptions qu'on en propose font appel au parallélisme, à des processeurs hautement spécialisés etc., aux antipodes de ce qui est envisagé pour CDB.

17. Voir en particulier Rips (1986) et le débat dans Johnson-Laird & Byrne (1993).

18. Le commentaire de W. Hodges, le logicien, dans Johnson-Laird & Byrne (1993), est caractéristique à cet égard.

Duhem-Quine – quelles parts reviennent respectivement aux capacités UL, CDB et autres dans RL ? Comment concevoir des expériences qui contournent l'obstacle ?

Rien ne dit, d'ailleurs, que RL puisse faire l'objet d'une description unifiée : d'une part les facteurs de performance peuvent s'y opposer pour des raisons analogues à celles que Chomsky invoque pour le langage ; d'autre part il n'est pas exclu que des mécanismes très différents rendent compte de différentes familles de comportements déductifs.

c. Logique, langage et pensée.¹⁹

Une question demeure, qui se rattache à deux des interrogations précédentes, sans se confondre avec elles. On s'est demandé, d'une part, comment il est possible de s'assurer que l'interprétation faite par le sujet de la tâche à accomplir est identique à celle que l'expérimentateur a en tête ; on s'est interrogé, d'autre part, sur le statut d'une *Urlogique*. L'interprétation des constantes logiques, plus exactement des mots ou locutions utilisés dans l'énoncé du problème, est au cœur de celle de l'énoncé, puisque c'est l'aptitude logique que l'on cherche à évaluer, non l'aptitude à saisir le sens des autres éléments du texte ; cette interprétation est aussi au cœur de la question de l'*Urlogique*, puisque celle-ci doit contenir au moins les prémices, les précurseurs (en un sens quasi biochimique) d'une compréhension des parties du discours dénotant les constantes logiques (nous les appellerons, pour faire bref, les *mots logiques*) : la question de l'aptitude logique ne se pose qu'à propos d'organismes ayant une manière d'accès aux constantes logiques.

L'essentiel des expériences (en particulier celle de Wason) portant sur la logique propositionnelle, nous nous restreindrons aux mots logiques correspondant (en quel sens exactement, c'est tout le problème) aux connecteurs. Aucune difficulté n'est soulevée (dans le contexte présent) par la négation ou la conjonction. C'est la disjonction et surtout l'implication qui font problème. (Cette partition n'est pas fortuite, mais nous ne nous y attarderons pas.) Les difficultés de la disjonction sont peu analysées ; en fait ce sont pour l'essentiel celles de l'implication, avec en plus l'ambiguïté entre les sens exclusif (dont on dit volontiers qu'il est celui du "ou" naturel) et inclusif. L'implication est en tout cas au centre de toute discussion du "sens" des connecteurs, et c'est à tenter de dissiper certaines des confusions auxquelles elle donne lieu que nous nous attacherons.

Nous partirons d'un constat : la plupart des psychologues (et ils ne sont pas les seuls) semblent tenir pour acquis que l'implication de la logique (l'implication dite matérielle, i.m.) n'a pas le sens du "si" naturel. Sur ce point, partisans et adversaires de la logique mentale s'accordent. C'est ainsi qu'ils s'interrogent sur les raisons que peuvent avoir les sujets d'interpréter le "si" naturel comme la bi-implication logique

19. Cette section présente pour la première fois quelques considérations qui vont à l'encontre d'opinions très largement partagées ; elles sont en ce sens originales et sujettes à caution, et j'aurais préféré m'en tenir à la discussion des positions internes à la problématique générale, comme je fais pour l'essentiel dans le reste de cet article. Mais cela m'a été impossible. Une présentation plus approfondie de mon argumentation devrait paraître sous la forme d'un article que je prépare.

(par exemple dans “Si tu es sage, je te donnerai un bonbon”). Et ils se fondent sur les fameux “paradoxes de l’implication matérielle”²⁰ pour arguer que l’i.m. ne saurait avoir le sens du “si” : on lit couramment, sous la plume d’auteurs habituellement lucides, des affirmations, présentées comme des évidences, telles que “dans le raisonnement ordinaire, du faux rien ne s’ensuit”, *par opposition* à (ce que dicterait) l’i.m. Ou encore, on demande à des populations de sujets de dire s’il est vrai que p entraîne q lorsque p est faux, et on s’interroge sur les réponses, qui vont d’un refus de se prononcer à la réponse “non”, en passant par divers intermédiaires. Tout cela est *totale*ment dénué de sens, malgré l’apparence de caution apportée par la tradition de logiciens qui de Lewis à Belnap ont prétendu que l’i.m. présentait de graves défauts qu’il fallait pallier. La position que nous prenons le risque de présenter ici sans la défendre s’articule autour des points suivants.

1. En tant que connecteur (donc dans son rôle logique au sens strict, notamment en tant qu’elle donne lieu à des tautologies jugées par certains “paradoxaes”), l’i.m. ne possède pas de *sens* au sens où les mots du langage naturel ont un sens : c’est une erreur de catégorie caractérisée que de lui en attribuer un, que ce soit à partir de son “rôle inférentiel” ou de sa table de vérité.

2. La relation de conséquence logique (il faudrait distinguer, même au niveau du sens commun, la conséquence prouvable et la conséquence tautologique, mais pour les besoins de la discussion présente nous les confondrons) appartient, elle, au métalangage, et possède bel et bien un sens strict auquel on peut envisager de comparer celui que lui attribue le sens commun. Cela suppose évidemment que le sens commun ait à son égard une opinion définie, et sur ce point tout dogmatisme serait déplacé. L’attitude qu’il convient d’adopter en la matière est d’ailleurs un problème en soi, qui a été longuement discuté, entre autres, par Cohen et par Stich. Contentons-nous ici de présenter une thèse :

(i) “A a pour conséquence logique B” signifie pour le sens commun que A est inférable de B *en toute circonstance*, ce que le philosophe peut vouloir interpréter comme “quel que soit le monde considéré” ;

(ii) le logicien comprend que cette condition *équival*ut à “en vertu des formes logiques de A et de B” et peut réaménager, compte tenu de cette équivalence, la distribution des sens techniques de l’une ou l’autre des manières classiques de le faire.

Cette thèse repose sur la notion d’*inféribilité* dont il est évidemment crucial de donner une définition non circulaire. Avant de le faire, précisons que l’important dans cette thèse ne réside pas dans la formulation de (ii), laissée dans le vague à dessein ; mais dans l’affirmation que le rapport entre la notion naturelle et la notion technique de conséquence logique est très exactement de l’ordre de celui qui s’établit entre n’importe quelle notion “informelle” (lâche et quotidienne, *loose*,

20. Ou d’autres, fabriqués pour la circonstance. Ainsi pour Jonathan Cohen (*in* Johnson-Laird & Byrne 1993, p. 341), la logique, et non le sens commun, autorise l’inférence de (a) *Si l’auto de Jean est une 2CV, Jean est pauvre, et si c’est une Rolls, il est riche* à (b) *Ou bien : si l’auto de Jean est une Rolls, il est pauvre, ou bien : si c’est une 2CV, il est riche*. Cet exemple est un concentré de perfidie.

everyday) et la notion savante (stricte et technique) qu'on lui a associée au terme d'un cheminement scientifique et philosophique. Ce rapport n'est ni de différence ni d'identité, mais de représentation ou d'approximation.

3. "Inférer B de A" signifie ici (ce n'est en aucune façon la description supposée d'un usage courant, ni le rappel d'un sens technique, mais une stipulation) quelque chose comme "passer de A à B", ou "importer B sous la condition que A", ou "acheter B au prix de A", ou encore "déclencher/activer B en présence de A" – bref, inférer est une *action*. D'autre part, cette action est primitivement épistémique et non linguistique : c'est d'un commerce de pensées ou de croyances qu'il s'agit fondamentalement.

4. Le "si" du langage naturel a un sens et un seul, qui est celui d'une invitation à inférer, d'une licence d'inférer, ou encore d'une obligation d'inférer. Penser "Si A, B" (ou "B si A"), c'est se proposer d'inférer B de A ; affirmer "Si A, B", c'est inviter l'auditeur à inférer B de A. Comme on vient de le dire – et c'est crucial –, A et B sont des pensées, et ce n'est que de manière dérivée qu'ils peuvent être soit des énoncés soit des faits.

Invitation, piège, licence... : "si" est accompagné d'une pragmatique précise et obligatoire (au sens où les indexicaux le sont, par opposition aux autres parties du discours qui charrient en général une pragmatique floue et optionnelle). "Si" renvoie contextuellement à l'*existence* d'une *famille de situations* qui constitue le domaine de validité de l'invitation (ou de la licence, le domaine de fonctionnement du piège). "Si A, B" *exprime* une invitation à inférer B de A et *indique* une famille $\mathcal{K}$ de situations (le philosophe peut vouloir parler de mondes possibles) dans lesquelles cette invitation est valable. La pragmatique de "si" impose à la famille $\mathcal{K}$ d'être non "A-triviale" c'est-à-dire de comporter au moins une situation dans laquelle A est vrai et une situation dans laquelle A est faux – en particulier, $\mathcal{K}$ n'est jamais réduite à 0 ou à 1 élément. En d'autres termes, l'emploi de "si" présuppose une *variation* non triviale.

La famille $\mathcal{K}$ peut être constituée de toutes les situations matériellement possibles dans notre monde physique, et A avoir pour contenu un événement qui peut ou non se produire ; dans ce cas, B sera un effet ayant A pour cause, et l'on pourra être tenté de parler d'un "si causal" ; mais ce n'est pas au sens de "si" que se rattache la notion de causalité, c'est à la famille $\mathcal{K}$ pragmatiquement associée à cet emploi de "si". De même, $\mathcal{K}$ peut être constituée de toutes les situations sans restriction, et A avoir pour contenu une pensée composée à l'aide de chevilles logiques ; B pourra alors être une conséquence logique de A, et le "si" ne sera "logique" qu'en vertu de k . Et ainsi de suite : "si" peut annoncer une connexion indicielle ("s'il y a de la fumée, il y a du feu"), une connexion conventionnelle ("si c'est février, il n'y a que 28 ou 29 jours dans le mois"), une connexion mathématique ("si n est pair, le carré de n est pair"), une connexion performative ("si vous signez, vous devenez propriétaire du château" ou "si tu bouges, tu es mort"), déontique ("si tu as péché, tu dois te confesser"), etc. Il peut même arriver que $\mathcal{K}$ comporte des situations impossibles : dans "si $2+2 = 4$, alors $(2+2)^2 = 4^2$ ", $\mathcal{K}$ comprend la situation dans laquelle $2+2$ est différent de 4. Bref, "si" annonce l'existence d'une connexion sans encoder la moindre information sur la nature de la connexion, ce qui explique que toute connexion puisse être en cause, et que les locuteurs du français puissent avoir une parfaite maîtrise du "si" sans avoir

nécessairement d'idée bien précise sur ce qu'est l'inférence logique, la cause, la convention etc. Même les usages tels que "si tu as soif, il y a de la bière dans le réfrigérateur" ou (message laissé sur un répondeur téléphonique) "si je suis bien chez M. X, qu'il me rappelle à tel numéro" qui paraissent très éloignés du "si" prototypique relèvent du même mécanisme. Dans le premier exemple, $\mathcal{K}$ est la famille des situations dans lesquelles tu as soif ou tu n'as pas soif, et la pensée "tu as soif", présente dans le premier genre de situation, doit (dans l'intention du locuteur) conduire l'auditeur à activer la pensée "il y a de la bière dans le réfrigérateur". De même dans le second exemple : $\mathcal{K}$ est l'ensemble des situations dans lesquelles l'abonné atteint est M. X et celles dans lesquelles il ne l'est pas, et la licence accordée permet à l'auditeur de passer de la pensée "c'est bien chez X" à la pensée "l'auteur du message souhaite que X le rappelle à tel numéro" (les paraphrases sont grossières, mais peu importe). Le traitement de ces exemples et d'autres du même genre, abondamment cités dans la littérature, repose sur l'idée que le "si" porte sur des pensées et non directement sur le contenu des pensées (dans les exemples prototypiques, la pensée et son contenu sont indiscernables pour tout autre que le philosophe, d'où l'apparente ambiguïté entre connexion des faits et connexions des pensées).

Toute la difficulté de "si" pour le théoricien provient de la nature très particulière du renvoi pragmatique à la famille $\mathcal{K}$, de son mode d'existence et de détermination, et de sa nature dans la variété indéfinie des emplois légitimes de "si". Quant aux prétendus paradoxes et aux erreurs de sujets mis en situation expérimentale, ils proviennent d'emplois illégitimes de "si", ou de montages pervers dans lesquels par exemple deux familles différentes de situations sont pragmatiquement mobilisées autour d'un seul "si", ou deux manières de quantifier sur les situations entrent en compétition²¹. C'est du moins ce que nous conjecturons, et qu'il faudra montrer de manière détaillée et systématique en un autre lieu.

5. S'il est vrai que l'i.m. n'a pas de "sens" au sens habituel, on ne peut se soustraire à l'obligation d'examiner ce qu'on pourrait peut-être appeler son "sens projeté" (projeté sur le langage naturel) et qui, selon la plupart des auteurs, serait différent du sens du "si" naturel. Ce sens projeté, attribuons-le par convention à un nouveau mot, "sim". C'est ce que certains des auteurs en question appellent le "si" de la logique. "Sim A, B" (ou "B sim A") est assertable, disent-ils, même lorsque A est faux, ou lorsque B est vrai "indépendamment" de A, etc., si bien que l'on peut dire "sim Paris est la capitale de la Prusse, ma tante est mon oncle", mais aussi "sim Paris est la capitale de la France, ma tante est mon oncle" et "sim Paris est la capitale de la Prusse, ma tante est ma tante", toutes choses inassertables avec "si" au

21. A titre d'illustration de ce dernier phénomène, je propose à la sagacité du lecteur l'explication du "paradoxe" obtenu en remplaçant dans la tautologie suivante :

$$[p \rightarrow (p \wedge q)] \vee [q \rightarrow (p \wedge q)],$$

la proposition p par "je suis riche" et la proposition q par "tu es riche". (*Indication* : la quantification [implicite, métalinguistique] à laquelle se prête la tautologie porte sur la formule entière, alors que la paraphrase en langue naturelle force une quantification sur chacun des membres de la disjonction ; le paradoxe se réduit à l'erreur classique de distribuer la quantification universelle sur une disjonction, et n'a *rien* à voir avec l'acceptabilité de la table de vérité de l'i.m.)

lieu de “sim”, sauf, en un sens ironique, la première (sur le patron de laquelle est construit le fameux “je veux bien être pendu sim Paris est la capitale de la Prusse”).

Qu’un tel mot n’existe pas dans la langue est évident ; ce qu’il faut expliquer, c’est d’une part pourquoi il n’existe pas, et d’autre part pourquoi la logique peut et doit attribuer à l’i.m. la table de vérité que l’on sait. Ce sont des questions difficiles, et nous devons nous contenter ici d’une indication. Si “sim” n’existe pas, c’est qu’il placerait l’auditeur d’une phrase le contenant dans une situation impossible – soit de rechercher une famille de variations tout en étant empêché de considérer toute autre situation que celle que la prémisse spécifie, soit de rechercher une connexion tout en étant obligé d’exclure qu’elle existe : il faudrait qu’à la fois on puisse savoir et ne pas savoir ce qu’il en est de A, ou savoir et ne pas savoir que A et B sont liés. Ce serait soit un “si” muni d’une pragmatique incohérente, soit un “si” arbitrairement privé de toute pragmatique, ce qui constituerait une incohérence au niveau supérieur.

Quant à la table de vérité de l’i.m., elle se comprend à la lumière de la distinction entre *fait* et *raison* : l’i.m. exprime *dans une situation entièrement fixée mais non nécessairement connue* le fait disjonctif “non-A ou B” ; la validité de cette disjonction *dans une variété logiquement définissable de situations* renvoie à une *raison* (nécessairement de nature logique) de cette invariance. Le “si” naturel indique de même l’*existence* d’une raison (non nécessairement logique) qui explique et garantit l’invariance relativement à une variété non nécessairement logiquement définissable de situations ; lorsque c’est un fait *singulier connu* qu’il s’agit d’exprimer, alors ce fait est soit non-A soit B, et il serait contraire aux règles de la pertinence d’appauvrir l’information “non-A” (respectivement “B”) en “non-A ou B”.

Quoiqu’il en soit, puisque “sim” n’existe pas, que peut faire le sujet d’expérience que l’on cherche à contraindre à interpréter “si” comme “sim” ? On le trompe, on se joue de lui : il n’est pas placé dans une situation normale de résolution de problème. Il doit faire face à une situation perverse, qu’il saisit mal, et ne peut que faire au mieux.

Que conclure de toute cette discussion, de manière plus générale, en ce qui concerne les expériences sur le raisonnement déductif ? C’est loin d’être clair. Trois difficultés semblent toutefois se dégager nettement. La première porte sur la distinction entre possession ou maîtrise des concepts correspondant aux mots logiques et aptitude déductive de base : quelle est au juste cette distinction, et comment en tenir compte dans l’investigation expérimentale du raisonnement élémentaire ? La seconde concerne la distinction entre conséquence logique et conséquence tout court : il est peu probable que la première de ces notions soit claire dans l’entendement naïf, a fortiori ne doit-on pas compter sur une perception claire de la différence. On ne peut donc mettre les capacités déductives à l’épreuve chez un sujet logiquement naïf en lui demandant de raisonner dans des situations dont la compréhension exige une maîtrise de cette différence (ce qui est le cas de la tâche de Wason). La troisième concerne le sens des mots logiques de la langue naturelle : s’il est très contestable que ces mots dénotent des constantes logiques non standard, il est non moins improbable qu’ils puissent indiquer par eux-mêmes la nature logique ou non logique des connexions qu’ils dénotent. On peut douter, par conséquent, qu’une formulation en langage naturel d’un problème de logique soit possible sans recours à une mise en scène explicitement logique, mise en scène qui ruine la possibilité de mettre à l’épreuve les capacités d’un sujet logiquement naïf !

4. Une catégorie de raisonnements inductifs : le jugement en situation d'incertitude

On oppose traditionnellement, chez les logiciens, le raisonnement déductif au raisonnement inductif. Harman, on l'a vu, fournit des raisons de douter, à supposer qu'il y ait là une distinction à faire, que ce soit une bonne manière de la formuler. Parlons donc plutôt de deux sortes d'arguments intervenant dans le raisonnement : ceux dans lesquels la conclusion suit avec certitude des prémisses (inférences *démonstratives*), et ceux dans lesquels elle ne suit qu'avec un degré d'incertitude (inférences *non démonstratives*).

Sous la dénomination de "jugement en situation d'incertitude"²², s'est développé, avec un retard de quelques années sur le paradigme de Wason, un programme de recherche en psychologie cognitive expérimentale visant à évaluer et à expliquer certaines capacités humaines en matière d'arguments de la seconde espèce²³. Ce programme (également appelé "*heuristics and biases*" du nom des deux principales notions mobilisées) a connu un développement quantitatif, mais surtout un retentissement, bien supérieurs à son homologue "déductif". Nous reviendrons dans un instant sur l'importance prise par ce courant de recherche en l'espace d'une vingtaine d'années, depuis les premières publications d'Amos Tversky et Daniel Kahneman, ses chefs de file longtemps incontestés. Regardons d'abord quelques-uns de ses résultats les plus caractéristiques.

Les expériences

Il y a deux façons d'évaluer une probabilité empirique : directement à partir de fréquences observées, ou indirectement à partir de données comparatives ou absolues, qualitatives ou quantitatives, sur des probabilités, et de relations entre les événements considérés. Toute détermination indirecte repose sur l'un ou l'autre des principes de la théorie des probabilités – on peut du moins le supposer.

22. Comme nous allons le voir, il s'agit essentiellement, mais pas exclusivement, d'expériences sur des raisonnements probabilistes ou statistiques. C'est du moins ainsi qu'elles sont généralement présentées.

23. Il ne sera question ici que d'un certain type d'argument non démonstratif ; les autres recherches dans le domaine portent principalement sur l'acquisition inductive de concepts, ou l'attribution de propriétés à partir d'informations finies, ou encore la saisie d'une règle à partir d'exemples. On peut observer à ce propos que la manière dont les psychologues découpent le domaine du raisonnement en sous-domaines est assez surprenante pour le logicien ; ainsi Garnham et Oakhill, dans leur manuel (1994), distinguent-ils entre "Raisonnement déductif" (chap. 5), "Raisonnement syllogistique" (chap. 6), "Induction" (chap. 7), "Mise à l'épreuve [*testing*] d'hypothèses" (chap. 8 – dans lequel est exposée notamment la tâche de sélection de Wason), "Raisonnement statistique" (chap. 9 – dans lequel il est question de ce qui nous occupe dans la présente section), et enfin "Prise de décision" (chap. 10 – de la réalité psychologique des principes de la théorie de la décision).

a. Observation directe de fréquences.

D'un ensemble assez complet d'expériences que nous laisserons en dehors de la discussion, il apparaît que de manière générale l'homme repère assez bien les fréquences d'événements identiques au sein d'une série d'épreuves (par exemple il évalue et classe sans erreurs importantes les proportions de boules de différentes couleurs tirées d'une urne sous ses yeux). On signale cependant une tendance à surestimer les fréquences très faibles. D'autre part, on a mis en évidence des perturbations plus graves lorsque les fréquences portent sur des événements véhiculant des informations significatives – une manifestation de plus de l'“effet de contenu”. Cette catégorie d'erreurs constitue l'un des premiers résultats de Tversky et Kahneman :

Expérience I. On fait défiler devant le sujet deux listes de personnalités.

La liste A comporte 19 hommes très célèbres, 20 femmes connues mais moins célèbres

La liste B comporte 19 femmes très célèbres, 20 hommes connus mais moins célèbres.

On demande au sujet de comparer, pour chaque liste, le nombre d'hommes et de femmes.

Résultat : 80% des sujets estiment qu'il y a plus d'hommes que de femmes dans A, et *vice-versa* dans B. (Kahneman & Tversky 1973)

Nous verrons un peu plus loin comment les auteurs de l'étude se proposent d'expliquer le phénomène. Mais leur explication s'applique à d'autres erreurs ; il sera plus économique de passer les exemples en revue avant de présenter les explications disponibles.

Les erreurs les plus significatives concernent, on ne s'en étonnera pas, les cas de détermination indirecte de fréquence. Toutes sortes de violations des principes probabilistes ont été mises en évidence. Nous présentons les principales, classées selon la difficulté apparente des principes probabilistes impliqués.

b. Sophisme de la conjonction.

Peu de principes probabilistes sont plus simples et apparemment naturels que le principe de la conjonction, selon lequel la probabilité d'une conjonction d'événements ne peut excéder les probabilités de chacun des événements conjoints. Or ce principe semble régulièrement violé :

Expérience II. Linda a 31 ans, elle est célibataire, elle a son franc-parler, elle est très intelligente. Elle a terminé des études de philosophie. Etudiante, elle était extrêmement sensible aux questions de discrimination²⁴ et de justice sociale ; elle a également pris part à des manifestations anti-nucléaires.

Veuillez noter les propositions suivantes selon leur probabilité, en donnant 1 à la plus probable et 8 à la moins probable :

24. Est-il besoin de préciser que ces expériences ont été menées en Amérique du Nord ?

- a. Linda est institutrice.
- b. Linda travaille dans une librairie et suit des cours de yoga.
- c. Linda milite dans le mouvement féministe.
- d. Linda est assistante sociale en milieu psychiatrique.
- e. Linda est membre de la Ligue des Electrices [*League of Women Voters*]
- f. Linda est employée de banque.
- g. Linda est représentante en assurances.
- h. Linda est employée de banque et milite dans le mouvement féministe.

Résultat : 89% des sujets placent (h) avant (f). (Tversky & Kahneman 1983)

Expérience IIbis. Même expérience, sauf qu'on ne propose que (f) et (h).

Résultat : 85% des sujets placent (h) avant (f).

Expérience III. Une enquête de santé publique a été effectuée en Colombie britannique²⁵ sur un échantillon d'hommes adultes de tous âges et de toutes professions.

Veuillez donner votre estimation des valeurs suivantes :

(a) Le pourcentage des hommes étudiés qui ont subi un infarctus ou plus.

(b) Le pourcentage des hommes étudiés qui ont 55 ans ou plus et ont subi un infarctus ou plus.

Résultat : 65% des sujets fournissent un pourcentage supérieur pour (b) que pour (a). (Tversky & Kahneman 1983)

Expérience IV. Un dé non pipé porte 4 faces vertes et 2 rouges. On va lancer le dé 20 fois et on enregistrera la suite des verts (V) et des rouges (R) qui sortent. On vous demande de choisir l'une des suites qui figurent parmi les trois ci-dessous ; vous recevrez 25\$ [de 1983 – environ 300 F de 1994] si une suite de jets consécutifs est identique à la suite que vous avez choisie. Cochez la suite sur laquelle vous souhaitez parier :

A. R V R R R

B. V R V R R R

C. V R R R R R

Résultat : Une majorité de sujets (260 étudiants de Stanford et de l'University of British Columbia) cochent la suite B. (Tversky & Kahneman 1983)

c. Sophisme du taux de base (non-prise en compte des chances a priori)

Dans de nombreux problèmes, la solution dictée par la théorie des probabilités est fonction de la composition de la population de référence (donnée par ce qu'on appelle les taux de base ou chances a priori). On sait depuis assez longtemps²⁶ que les sujets, placés devant ce genre de problème, se montrent généralement insensibles

25. Où Kahneman a enseigné de longues années.

26. Il semble que ce soit dans le contexte du diagnostic médical que l'effet dont il va être question a été mis pour la première fois en évidence par des psychologues (Meehl & Rosen 1955).

au rôle des taux de base. Kahneman et Tversky ont proposé une série d'expériences devenues célèbres pour mettre en évidence la robustesse de cette erreur.

Expérience V. On compte, dans une certaine ville, 85% de taxis verts et 15% de bleus. Un accident a lieu dans lequel un taxi est impliqué. Un témoin affirme que le taxi était bleu. Une série de tests montre que le témoin est fiable à 80% : il identifie correctement la couleur d'un taxi, qu'il soit vert ou bleu, dans 80% des cas. Quelle est la probabilité que le taxi impliqué soit effectivement bleu ?

Résultat : La grande majorité des sujets non formés en probabilités répond : environ 80%. [L'application du théorème de Bayes à partir du taux de base de 15% de taxis bleus fournit 41%.] (Kahneman & Tversky 1972a)

Expérience VI. Un test existe pour détecter une maladie rare. La maladie affecte 1 personne sur 1 000 dans une certaine population, et le test a un taux de faux positifs de 5%. On teste un patient et il est positif. Quelles sont les chances, en l'absence de toute autre information, que le patient ait effectivement la maladie en question ?

Résultat : Sur 60 étudiants et enseignants de la Harvard Medical School, près de la moitié estiment les chances à 95% ; la moyenne des réponses est 56%, la réponse 2% [correcte selon la formule de Bayes] est donnée par 18% des sujets. (Casscells, Schoenberger & Grayboys 1978 - repris par Tversky & Kahneman 1982a)

d. *“Loi des petits nombres” – signification et taille des échantillons.*

La loi des grands nombres, dont l'évidence est reconnue depuis l'aube du calcul des probabilités et la démonstration mathématique est un exercice standard, ne donne pas seulement aux philosophes du fil à retordre²⁷, comme semblent l'attester les expériences suivantes.

Expérience VII. Il y a deux maternités dans une certaine ville. Dans la première, environ 45 bébés naissent chaque jour ; dans la seconde, 15. Comme vous le savez, à peu près 50% des nouveaux-nés sont des garçons. Le pourcentage exact varie néanmoins d'un jour sur l'autre. Il peut être tantôt supérieur, tantôt inférieur. Au cours d'une année donnée, chaque maternité a compté le nombre de journées où ce pourcentage était supérieur à 60%. Selon vous, quelle est la maternité qui a compté le plus grand nombre de telles journées ?

Résultat : Une majorité des sujets estime qu'il n'y a pas de différence entre les deux maternités. Les autres sujets se partagent à peu près à égalité entre les deux autres réponses possibles. (Kahneman & Tversky 1972)

27. Karl Popper insiste sur l'impérieuse nécessité de rendre intelligible cette loi, que ses démonstrations dans un cadre axiomatique ne protègent pas du risque d'apparaître mystérieuse, voire miraculeuse. Voir Popper (1983).

Expérience VIII. On répartit 20 billes au hasard entre 5 enfants, André, Béatrice, Charles, Diane et Eléonore. On recommence un grand nombre de fois. Des deux répartitions suivantes, laquelle est la plus fréquente ?

Répartition I A=4, B=4, C=5, D=4, E=3

Répartition II A=4, B=4, C=4, D=4, E=4

Résultat : Une majorité des sujets estime que I est plus fréquente que II. (Kahneman & Tversky 1972)

Expérience IX. On considère une urne contenant des boules, dont deux tiers sont d'une couleur et un tiers d'une autre. Amédée tire 5 boules, dont 4 rouges et 1 blanche. Blandine tire 20 boules, dont 12 rouges et 8 blanches. Lequel des deux est en droit de se sentir le plus sûr de la répartition 2/3 de rouges et 1/3 de blanches, plutôt que l'inverse ?

Résultat : La majorité des sujets jugent qu'Amédée est plus justifié dans sa réponse que Blandine. [Selon le calcul des probabilités, Blandine est 2 fois plus justifiée qu'Amédée.] (Tversky & Kahneman 1974)

De ces expériences et d'autres se dégage l'hypothèse que les sujets ont du mal à saisir le rôle de la taille comparée des échantillons : ils respectent la "loi des petits nombres", affirment Tversky & Kahneman (1971), c'est-à-dire qu'ils tendent à surestimer le caractère significatif des petits échantillons. Le célèbre sophisme du joueur, dont l'expérience prouve abondamment que les sujets le commettent très souvent²⁸, relève de la même erreur.

e. Incompréhension de la variance.

Une distribution de probabilité ne détermine pas seulement l'espérance, mais aussi les moments d'ordre supérieur, en particulier la probabilité d'un écart donné par rapport à l'espérance. Les sujets semblent mal armés pour évaluer une telle probabilité, ce qui n'étonnera sans doute pas grand monde. Le sens d'une expérience telle que la suivante est de suggérer des stratégies susceptibles d'être mobilisées par les sujets dans des cas où la compétence probabiliste leur fait défaut.

Expérience X. Un lycée comprend deux filières. Dans la filière A, il y a 65% de garçons ; dans la filière B, il y en a 45%. Il y a le même nombre de classes dans chaque filière. Vous entrez dans une classe au hasard et vous comptez les garçons : il y en a 55%. Si vous deviez parier à quelle filière appartient la classe en question, choisiriez-vous A ou bien B ?

Résultat : La plupart des sujets parient A. [Or la loi binomiale a une variance plus forte pour $p = 0,45$ que pour $p = 0,65$, donc le bon pari est B, puisque l'écart est le même (0,1) par rapport à chacune des deux moyennes.] (Kahneman & Tversky 1972)

28. Gigerenzer *et al.* (1989) signalent même l'erreur dans le manuel d'un professeur de probabilités de l'Ecole polytechnique dans les années 1920 !

f. Evaluations non probabilistes.

Toujours dans l'espoir de mettre en évidence les stratégies d'estimation mises en œuvre dans le jugement incertain, diverses autres expériences ont été menées. En voici deux parmi les plus caractéristiques :

Expérience XI. Un groupe de 10 personnes doivent constituer des commissions. Combien de commissions différentes de 2 membres peuvent-elles être constituées ? Et combien de commissions de 6 membres ?

Résultat : La plupart des sujets donnent une réponse supérieure à la première question qu'à la seconde. [Au pays du triangle de Pascal, chacun sait naturellement que l'ordre est inverse.]

Expérience XII. a. Vous allez disposer d'exactly 5 secondes pour répondre à la question suivante : évaluer le produit $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$.

b. Vous allez disposer d'exactly 5 secondes pour répondre à la question suivante : évaluer le produit $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$.

Résultat : La médiane des réponses est de 512 pour *a* et de 2250 pour *b*. [La réponse exacte est 40320.]

Les explications de l'école "Heuristics and biases"

Ces expériences, et beaucoup d'autres menées dans le même esprit, prouvent selon Tversky et Kahneman que l'esprit humain est affecté, dans ses jugements en situation d'incertitude, par ce qu'ils appellent des "biais", comparables aux biais de certaines méthodes statistiques, ou de certains instruments, et responsables d'"illusions cognitives" qui seraient l'équivalent d'illusions d'optique, inévitables même en présence de données objectives qui devraient les dissiper.

Cette façon de voir oriente la recherche vers des explications de type fonctionnel : si le système cognitif responsable des jugements en situation d'incertitude est victime d'illusions systématiques, c'est qu'il fonctionne selon des principes non conformes à la norme rationnelle, mais qui répondent sans doute à des impératifs de *design* computationnel et qui fournissent néanmoins à l'organisme des réponses rapides et acceptables dans une bonne proportion des situations écologiques. Ces principes sont des "heuristiques" au sens que ce terme a pris en IA. Tversky et Kahneman ont cru détecter trois de ces heuristiques. La première, qui est aussi la plus clairement formulée et la mieux corroborée, est l'heuristique de *représentativité* ou de *similarité* ; Daniel Osherson (1990) propose une version simplifiée de ce principe :

Pour juger de la probabilité qu'un objet *O* appartient à une catégorie *C*, évaluer le degré de similarité de *O* à *C* : plus ce degré est élevé, plus la probabilité est grande.

C'est l'application de ce principe qui explique, selon Tversky et Kahneman (1974, etc.), les résultats observés dans les expériences II à X. En fait, il faut selon eux (1982) distinguer deux formes de l'heuristique de représentativité. Dans l'expérience II, une Linda employée de banque *et* féministe est plus représentative de la jeune étudiante qu'était la Linda d'autrefois (plus semblable à la catégorie des jeunes étudiantes en philosophie intelligentes et militantes) qu'une Linda *simple*

employée de banque : le jugement opère *par le biais* d'une comparaison de similarités perçues, puis conclut par application du principe simplifié d'Osherson. Dans l'expérience IX au contraire, l'échantillon d'Amédée (4+1) est jugé plus représentatif de la répartition 2/3 - 1/3 que l'échantillon de Blandine (12+8) : ici c'est la distance perçue (le degré de similarité) qui informe un jugement *sur* la représentativité, et la représentativité fixe le choix en vertu d'un principe apparemment trivial selon lequel plus un échantillon est représentatif, plus ses traits pertinents ont de chance d'être proches de ceux de la population entière.

L'heuristique de *disponibilité* expliquerait les expériences I et XI : les jugements comparatifs se basent sur la disponibilité d'exemples de chacune des catégories (dans I, pour la liste A les exemples masculins sont plus faciles à évoquer que les exemples féminins, et inversement pour la liste B ; dans XI, davantage de commissions distinctes de 2 membres sont facilement imaginables que de commissions de 6 membres, qui se recoupent fortement et se distinguent donc plus difficilement).

Enfin, l'heuristique d'*ancrage plus ajustement* expliquerait l'expérience XII : les sujets feraient une toute première estimation en effectuant les premières opérations (c'est l'ancrage), puis la corrigeraient pour tenir compte des opérations restantes qu'ils n'ont pas le temps d'effectuer ; or cet ajustement est en général insuffisant, les sujets répugnant à trop modifier leur première estimation (peut-être en raison du phénomène général de persistance des croyances initiales, illustré par la résistance au *debriefing*²⁹). On comprendrait ainsi la considérable différence entre les résultats des tests a et b.

Critiques de “Heuristics and biases”

L'énorme importance accordée aux thèses de Kahneman et Tversky explique le nombre considérable de travaux critiques qu'elles ont suscités. Souvent, il s'agit d'expériences tendant à montrer que certains effets disparaissent lorsqu'on fait varier les paramètres ou d'autres conditions d'expérience, ce qui permettrait de légèrement “remonter la note” de notre système de jugement spontané sous incertitude. Mais quelques auteurs sont bien plus sévères : ils contestent le bien-fondé de toute l'entreprise. Nous évoquerons ici les deux critiques les plus radicales, Gerd Gigerenzer et Jonathan Cohen.

Le premier (1991) déploie son attaque sur deux fronts. Fort de ses connaissances scientifiques et historiques de la théorie des probabilités et de la statistique, il conteste d'abord que dans son interprétation classique cette théorie ait quoi que ce

29. Il s'agit de situations, réelles ou expérimentales, dans lesquelles des sujets sont dans un premier temps induits en erreur et doivent exécuter diverses tâches sur la base de ces données trompeuses ; dans un deuxième temps, on leur explique la tromperie, ce qui ôte tout fondement à leurs croyances de la première phase ; on constate néanmoins que ces croyances persistent souvent, fût-ce sous une forme atténuée. Dans le cas qui nous occupe, la première estimation serait à la fois d'emblée tenue pour fausse, et prise, faute de mieux, comme (approximativement) juste au cours de la première étape ; le “*debriefing*” par le sujet lui-même ne suffirait pas pour le convaincre de modifier radicalement sa réponse. (Ce rapprochement est une spéculation de l'auteur du présent article.)

soit à dire concernant la plupart des problèmes chers à Tversky et Kahneman. Selon lui en effet, ces problèmes portent sur l'évaluation de la probabilité d'un événement unique, alors que l'interprétation classique, c'est-à-dire fréquentiste³⁰, des probabilités porte sur le comportement asymptotique de suites d'épreuves identiques. C'est donc à la norme invoquée qu'il s'en prend en premier lieu³¹. En second lieu, il montre qu'en reformulant les problèmes de manière à les situer clairement dans le champ d'application de la théorie fréquentiste, on peut obtenir des améliorations spectaculaires des performances – au point que les fameuses “illusions cognitives” se dissipent, non tant chez le sujet que chez le psychologue qui avait cru les déceler. Voici, à titre d'exemple, comment Cosmides & Tooby (1990), cités par Gigerenzer, reformulent le problème de l'expérience VI :

Expérience VI fréq. Un Américain sur 1 000 est atteint de la maladie X. Un test permettant de détecter la maladie chez un patient a été mis au point. Chaque fois que le test est administré à quelqu'un qui est atteint de X, le résultat est positif. Mais il arrive qu'il soit positif alors que la personne est en parfaite santé ; pour être précis, sur 1 000 personnes en bonne santé, 50 réagissent positivement au test.

Imaginez que nous ayons réuni un échantillon aléatoire de 1 000 Américains. Ils ont été choisis par tirage au sort. Ce tirage a été exécuté par des gens n'ayant aucune information sur l'état de santé des personnes concernées. Combien de personnes réagissant positivement au test sont effectivement atteintes de la maladie : sur ?

Résultat : 76% des sujets fournissent la réponse 2% (1 sur 50 ou 51).

Une version “iconique” du problème, dans laquelle on demande aux sujets de représenter les populations considérées sur une grille, fournie, de 10 sur 10, donne des résultats encore meilleurs (92%).

De la même manière, la version fréquentiste de l'expérience Ibis fait virtuellement “disparaître” le sophisme de la conjonction :

Expérience Ibis fréq. [Même description de Linda du temps de ses études.]

Il y a 100 personnes qui correspondent à cette description. Combien d'entre elles sont :

(a) des employées de banque ?

(b) des employées de banque qui militent dans le mouvement féministe ?

Résultat : A peine plus de 20% des sujets donnent une réponse supérieure pour (b) que pour (a).

30. Gigerenzer ne rejette pas l'interprétation subjectiviste, même si elle est minoritaire ; pour lui, le subjectiviste conséquent doit également rejeter les conclusions de Kahneman et Tversky, quoique pour d'autres raisons que le fréquentiste.

31. Il détecte aussi des erreurs de formulation ; par exemple, dans le problème VI dont il va être question, il manque le taux de faux négatifs, et surtout une spécification de la nature de l'échantillonnage : est-il représentatif de la population générale, ou d'une population “auto-sélectionnée” par les symptômes exhibés ?

Pour Gigerenzer, ces expériences, et d'autres du même genre, montrent que les sujets, contrairement aux psychologues qu'il critique, sont sensibles à la différence entre événement singulier et fréquences dans les séries longues.

Jonathan Cohen ne prétend pas, quant à lui, opposer aux expériences de "*heuristics and biases*" d'autres expériences. Philosophe, il se place sur le terrain de l'analyse théorique, et sa première interrogation porte sur ce que des expériences de psychologie *en général* sont susceptibles de révéler des facultés humaines de raisonnement (Cohen 1981). Son étude critique est au service d'un projet plus vaste, qui est de défendre la rationalité du sens commun – d'un sujet logiquement et probabilistiquement naïf ; cet objectif est lui-même subordonné pour Cohen à la volonté de justifier la philosophie analytique, qui n'est rien d'autre, selon lui, que l'étude *raisonnée* des raisons : la *raison* qui examine les raisons (dans tous les domaines et à tous les niveaux), cette raison-là est forcément "naïve" – qu'elle se révèle foncièrement et grossièrement défectueuse, et c'est toute l'entreprise de la philosophie analytique qui se trouverait privée de raison d'être (Cohen 1986).

La question de la rationalité – et du lien entre rationalité et raisonnement – n'entre pas dans le propos du présent article ; a fortiori n'aurons-nous rien à dire ici de la nature et du fondement de la philosophie analytique. En revanche, l'analyse que fait Cohen des travaux de Tversky et Kahneman, plus particulièrement des expériences II et VI dont il vient d'être question, retiendra un instant notre attention.

Le volet critique de cette analyse relève de la stratégie, également appliquée par Gigerenzer, de l'arroseur arrosé : Cohen – philosophe des méthodes inductives et stochastiques³² – met en doute la capacité des psychologues à faire ce dont ils jugent leurs sujets incapables, à savoir résoudre correctement les problèmes probabilistes posés. Plus précisément, il conteste que les solutions bayésiennes présentées par Tversky et Kahneman comme seules solutions correctes soient d'une part uniques, d'autre part optimales dans le contexte des expériences incriminées. Sa cible principale est précisément le taux de base que les psychologues accusent leurs sujets de négliger : Cohen soutient d'une part qu'il peut y avoir de bonnes raisons de le faire, et en présente certaines (ce qui lui vaut, à propos de l'expérience VI, l'accusation de ne devoir qu'au fait de n'être pas directeur d'hôpital de n'être pas non plus un assassin !³³) ; il rappelle d'autre part que dans d'autres expériences³⁴, les sujets se montrent parfaitement capable de prendre le taux de base en compte (il arrive même qu'ils en surestiment l'importance). Si donc les sujets négligent le taux de base dans certains cas, il se pourrait qu'ils le fassent délibérément, ce qu'ils sont a priori en droit de faire : c'est un choix qui relève de la sélection des données

32. Voir Cohen (1970, 1977, 1989).

33. Dans un commentaire de Cohen (1981).

34. Les expériences qu'il mentionne (Cohen 1986, p. 163, note 21) sont antérieures, pour la plupart, à la période faste de *heuristics and biases* ; Gigerenzer souligne de même que Tversky et Kahneman mettent un terme à une période *optimiste* au cours de laquelle avaient été réunis des témoignages expérimentaux encourageants quant aux aptitudes probabilistes des sujets naïfs. Ce mouvement pendulaire illustre les "biais" qui affectent le domaine (Lopes 1991).

pertinentes, laquelle est par excellence affaire d'intuition (et c'est pourquoi d'ailleurs la question intéresse tant Cohen).

Cependant, en l'absence d'un candidat plausible à la qualité de raison, pour les sujets, de ne pas tenir compte du taux de base, l'argument contre les psychologues qui estiment qu'ils le font sans raison (donc irrationnellement) demeure incertain. Reconnaisant cette faiblesse³⁵, Cohen fait pour y remédier un pas important dans une direction constructive, très différente de celle que suit Gigerenzer mais jouant un rôle dialectique analogue. Il existe selon Cohen une distinction capitale entre deux sortes de probabilités. Soit C une classe de référence (une population au sens statistique) et P une qualité à laquelle est attachée une probabilité p : cette probabilité est dite *contrefactualisable* si elle s'applique aussi bien aux non-membres qu'aux membres de C . Soit par exemple C l'ensemble des ouvriers d'usine d'amiante âgés de 50 ans – à une période et dans une région considérées –, P le fait de mourir avant 70 ans, et supposons que la probabilité p pour un ouvrier de l'amiante de 50 ans de mourir avant 70 ans soit de 0,8. Cette probabilité ne dépend pas des individus particuliers qui se trouvent composer la classe P : si Malcolm, philosophe d'Oxford, se trouvait avoir travaillé depuis toujours dans une usine d'amiante et être âgé de 50 ans, la probabilité qu'il meure avant 70 ans serait égale à 0,8 ; ainsi, la probabilité p est contrefactualisable. Au contraire, supposons que la probabilité pour un ouvrier de l'amiante de 50 ans, à l'époque et dans la région considérées, d'avoir les yeux bleus soit de $q = 0,6$; cette probabilité-là est (selon toute vraisemblance) non contrefactualisable : elle ne s'applique *pas* à Malcolm. Pour le dire en peu de mots, est non contrefactualisable une probabilité associée à une propriété *accidentelle* de la classe de référence.

Il échoit au philosophe de fonder cette distinction, et Cohen s'y emploie (*op.cit.* §18). Mais pour qu'elle soit pertinente sur le plan psychologique, il faut que le sujet naïf y soit sensible, et que cette sensibilité puisse expliquer la manière dont il résout les problèmes que lui soumet le psychologue. Cohen pense que c'est bien le cas. Selon lui, les sujets donnent spontanément la préférence aux explications causales ; or les probabilités contrefactualisables sont plus proches de la causalité que les probabilités non contrefactualisables. Les sujets attribuent à l'expérimentateur les mêmes préférences, et conjecturent par conséquent que ce qui l'intéresse, c'est leur capacité à construire des explications causales, ou à défaut, à évaluer des probabilités contrefactualisables, dans la mesure où les données du problème n'imposent pas clairement l'interprétation non contrefactualisable.

Or, dans le problème des taxis (expérience V) par exemple, le sujet n'a pas de raison de considérer la proportion de taxis bleus comme contrefactualisable, contrairement au taux de fiabilité du témoin. Refusant à juste titre de mélanger les deux sortes de probabilité, il interprète alors la question comme portant sur des probabilités contrefactualisables, avec absence d'information sur la probabilité contrefactualisable d'apparition d'un taxi bleu ; il adopte donc pour celle-ci la valeur 0,5, puis applique correctement le raisonnement probabiliste et obtient la réponse correcte dans l'interprétation considérée.

35. C'est un progrès sensible sur Cohen 1981, accompli dans Cohen 1986.

Cette “préférence causale” que Cohen attribue aux sujets lui permet d’expliquer d’autres “erreurs” : dans l’expérience VII, par exemple, il conjecture que les sujets pensent qu’on leur demande si la taille de l’hôpital a une influence causale sur la répartition en sexes des naissances ; de même, dans l’expérience II, c’est sur la cohérence causale d’un scénario que les sujets s’imagineraient être mis à l’épreuve. Et justement, lorsque aucune information interprétable causalement n’est disponible, les sujets semblent fort capables de respecter les principes probabilistes de la conjonction et de la disjonction³⁶.

Ces principes apparaissent du reste à Cohen si solidaires de la notion même de probabilité (de même d’ailleurs que la formule de Bayes) qu’en tout état de cause, si on lui prouvait que les sujets sont incapables d’appliquer de tels principes, il lui semblerait préférable de dire qu’ils ne possèdent pas le concept de probabilité, ou qu’ils ne saisissent pas que c’est de ce concept qu’il est question dans les problèmes qu’on leur soumet. On retrouve là une des difficultés centrales que soulèvent les expériences sur le raisonnement déductif³⁷.

Quelques enseignements tirés de leur traitement par la psychologie

On a déjà signalé le retentissement particulier qu’a connu le programme de *heuristics and biases*³⁸. Il s’explique d’au moins deux façons. Les examiner nous permettra de saisir très rapidement les différences profondes qui séparent les deux ordres de phénomènes, l’argument démonstratif et l’argument non démonstratif.

La première raison de la notoriété supérieure du programme de Tversky et Kahneman est qu’en matière de jugement en situation d’incertitude aucune norme aussi solide que la logique d’Aristote et de Frege ne vient baliser le terrain, que ce soit pour fixer un cadre de concepts et de termes, pour dicter les méthodes et déterminer les réponses, ou enfin pour orienter l’interprétation des expériences de psychologie. La logique s’impose doublement, en tout cas dans un premier temps : elle fournit l’interprétation privilégiée des mots logiques, elle est dotée de fondements solides ; elle occupe de la sorte une place pratiquement inexpugnable dans ce qu’on pourrait appeler l’“intuition instruite”. Aussi les résultats empiriques

36. Ici encore, ce sont des expériences datant de l’époque “optimiste” (1966) que Cohen invoque (*op. cit.* p. 186, note 43).

37. La discussion est en fait plus compliquée que cela : le non-respect par le sujet des lois d’usage d’un concept n’entraîne pas le verdict de non-possession par le sujet du concept, mais la disjonction de ce verdict et de celui de non-possession du concept *par le théoricien*. S’ouvre ici le débat sur l’“équilibre réflexif” entre théorie et intuitions préthéoriques, sur lequel Cohen fait fond plus particulièrement dans son article de 1981, et d’une manière que Stich (1990) conteste avec force.

38. Gigerenzer (1991) mentionne de vastes secteurs d’activité influencés par ce programme : droit, économie, médecine, gestion, etc. Le grand public a été informé également : Lopes (1991) fait référence par exemple à l’article vedette d’un numéro “récent” de *Newsweek*. En revanche, l’article de l’*Economist* du 4 juillet 1992 reflète une relative “remontée” des recherches sur le raisonnement déductif, présentées sur le même plan que les travaux sur le jugement incertain.

de la psychologie de la déduction se sont-ils heurtés à des convictions suffisamment fortes pour les empêcher d'être pris, sinon au sérieux, du moins pour argent comptant. Notons que la logique opère à deux niveaux : en tant que norme officielle du raisonnement déductif, puis en tant que "pilier" de notre régime cognitif "instruit". La psychologie de la déduction se trouve ainsi prise en tenaille, si l'on peut dire, par deux corps d'armée sous un même commandement.

La psychologie du jugement en situation d'incertitude n'a face à elle qu'une seule troupe : la théorie mathématique des probabilités en tant qu'interprète officiel des concepts et calculs relatifs à l'incertain. Rien dans l'"intuition instruite" ne vient en quelque sorte redoubler les certitudes offertes par la théorie mathématique : devant un problème de chances, même fort simple, que faisons-nous ? Le langage naturel ne nous a pas préparés à nous en saisir, aucune voie de résolution n'y est en quelque sorte frayée ; les mots autant que les procédures font totalement défaut – même la notion d'ordre de grandeur est pratiquement inexprimable, en tout cas elle n'est pas manipulable dans la langue ordinaire. Nous n'avons donc pas d'autre choix que d'appliquer le calcul des probabilités en priant le ciel de ne pas tomber dans l'un des pièges redoutables que cette théorie nous tend, comme nous le savons d'expérience.

Cette relative fragilité de la norme, dans un premier temps, a formidablement contribué au crédit accordé aux thèses de *heuristics and biases*. Lorsque le psychologue met en évidence un écart considérable et systématique entre les résultats prescrits par l'application de la théorie des probabilités et les performances des sujets, on n'a à lui opposer rien qui redouble, dans l'intuition, l'évidence mathématique : comment le psychologue n'aurait-il pas raison, et que pourraient montrer ses expériences, sinon que les gens sont incapables ne serait-ce que d'approcher la vérité en matière de jugements quantitatifs d'incertitude ? La situation ne se retourne que dans un deuxième temps, lorsque l'on s'avise enfin de s'interroger sur le statut de la "vérité" brandie par le psychologue, et que fournit selon lui la théorie mathématique. Mais les premières thèses auront occupé le devant de la scène près de vingt ans, et le problème désormais est posé en grand.

La seconde raison de la notoriété des travaux psychologiques sur l'incertain tient à l'importance accordée par beaucoup à la fonction cognitive qu'ils sont censés mesurer. Autant, en ces temps de "nouvelle alliance" antilogiciste, on peut être tenté d'estimer que la déduction est bien peu de chose, autant le rôle que jouent les jugements de probabilité dans la conduite rationnelle apparaît capital : tout agent économique, tout praticien des arts et des techniques, tout enquêteur, tout magistrat ou juré, tout responsable politique, etc. exerce à longueur de journée son jugement sur les chances comparées d'événements incertains. Peu d'informations sur les aptitudes cognitives réelles peuvent donc rivaliser en importance avec celles que le programme *heuristics and biases* prétend nous apporter. Nous ne serions pas plus étonnés et inquiets, sans doute, si les psychologues et neurophysiologistes de la perception et du mouvement venaient nous apprendre que l'homme est incapable d'estimer correctement la vitesse et la trajectoire d'objets tels que les automobiles.

De manière plus générale, les résultats du programme *heuristics and biases* paraissent affecter notre conception de la rationalité humaine réelle davantage que les recherches sur la déduction parce que, comme l'annonce son objet officiel (le jugement en situation d'incertitude), elles touchent au jugement, qu'une certaine

sagesse place depuis toujours³⁹ (bien antérieurement, donc, à la vague antilogiciste actuelle) très au-dessus de la déduction. Relayés et amplifiés par la psychologie sociale, les résultats négatifs de ce programme ont donné une nouvelle jeunesse, et une caution scientifique, au vieux pessimisme philosophique sur les limites de l'intelligence humaine.

Ces raisons contribuent sans doute à expliquer l'inégalité des traitements réservés aux deux grands courants de recherche sur le raisonnement. Mais elles s'alimentent elles-mêmes à une conception erronée des phénomènes étudiés.

Cette conception, nous l'appellerons, pour faire court, "symétrique" : déduction et induction, ou encore raisonnement déductif et raisonnement probabiliste (ou incertain) sont des facultés qui existent côte à côte, elles sont sur le même plan, comme labourage et pâturage sont les deux mamelles de la France. Ou encore, comme la logique (mathématique) et la théorie (mathématique) des probabilités sont deux branches des mathématiques.

Puisque dans les présentations par les psychologues tout nous y invite, et qu'a contrario rien ne nous en dissuade, faisons nôtre, provisoirement, cette conception, et posons, à propos du jugement incertain, les questions que nous avons posées à propos de la déduction (au §3).

En premier lieu, au rapport entre logique et raisonnement (déductif) doit correspondre le rapport entre probabilités et raisonnement en situation d'incertitude. Se pose alors la question du sujet dont les capacités en matière d'analogie sont l'objet des recherches : ce doit être l'analogue, pour l'induction, du "sujet logiquement naïf" (SLN), donc un "sujet probabilistiquement naïf" (SPN). Or, on s'en souvient, nous avons été amené à attribuer au SLN une maîtrise préthéorique de certaines notions logiques, à défaut desquelles il ne serait pas capable de saisir les problèmes logiques qu'on lui soumet. La question majeure à laquelle nous devons répondre est celle-ci : de quelles notions probabilistes pré-théoriques devons-nous créditer le SPN pour qu'il soit en mesure de saisir les questions que lui posent les psychologues ? Sans doute celle de "probabilité subjective" – mais qu'est-ce que cela ? Aucune réponse ne semble s'imposer. Mais poursuivons. En matière de déduction, nous avons tenté de préciser la nature de l'aptitude psychologique fondamentale qu'il s'agissait de caractériser. Il fallait en tout cas qu'elle soit élémentaire. Quelle est l'aptitude élémentaire que nous voulons cerner chez le SPN ?

Nous pourrions être tentés de répondre : la mesure de fréquences ou de corrélations à partir d'observations, mais nous avons vu que d'une part cette aptitude, à l'estime des psychologues, semble bien développée chez l'homme, et que d'autre part ce n'est pas elle qui est en question dans les expériences qui nous occupent. Nous devons donc nous tourner vers les *relations* entre fréquences (ou probabilités, ou chances, ou propensions, etc.), et à défaut d'une "théorie naïve des probabilités subjectives", si conjecturale qu'elle soit, nous n'avons d'autre choix qu'entre des fragments plus ou moins étendus de la théorie classique des

39. Georges Pompidou ne disait-il pas que la première qualité d'un homme d'Etat est le jugement ? On croit plus en Maigret qui y va au *jugé* qu'en ce logicien de Sherlock Holmes.

probabilités⁴⁰. Or quel fragment s'impose-t-il ? Rien n'indique même qu'il faille le prendre dans l'axiomatique de Kolmogorov. D'abord cette axiomatique n'a aucune prétention à être "naturelle" au sens de la déduction naturelle en logique mathématique, et il n'existe pas (on peut même douter qu'il existera jamais) de théorie mathématique "naturelle" des probabilités ; de plus, il semble qu'une notion que l'on aimerait pouvoir attribuer au SPN est celle de probabilité conditionnelle – or elle ne figure pas en tant que telle dans l'axiomatique de Kolmogorov ; enfin, le problème de l'omniscience logique se pose : quelles conséquences logiques de l'axiomatique "native" ou naïve doivent-elles être incluses dans l'aptitude étudiée ? Naturellement, c'est là, en un sens, toute la question *empirique et scientifique* qui est posée, et nous n'avons pas à nous étonner de l'absence d'une réponse toute faite. Ce qui est gênant, en revanche, c'est la difficulté qu'il y a à imaginer si vaguement que ce soit *ce que pourrait être* une réponse plausible, sachant que nous voulons cerner une aptitude *élémentaire*. Une tension semble exister entre les ressources disponibles (la théorie mathématique) et l'objectif (caractériser une aptitude cognitive élémentaire)⁴¹.

Venons-en à la question des niveaux : nous l'avons posée, pour la déduction, par le biais du "système X" que semblaient rechercher les psychologues, système dont la nature et la fonction demandaient à être précisées.

La question de la fonction se pose avec une acuité toute particulière pour le jugement incertain. La situation dialectique est en effet très différente dans ce cas. En matière déductive, les compétences naturelles de l'homme ne sont pas nulles, même si elles sont prises en défaut avec régularité ; au contraire, en matière d'évaluation de résultats incertains, il n'y a pour ainsi dire, dans le programme de Tversky et Kahneman, que des défauts. Ce serait donc à l'*absence* d'un système Y, postulé ou espéré, qu'il permettrait de conclure, et à la présence d'un système Z (les heuristiques) qui n'accomplit pas ce que Y devait accomplir, mais accomplit *autre chose*. Kahneman et Tversky – Lopes (*loc. cit.*) et Gigerenzer (*loc. cit.*) le soulignent – restent fort vagues sur ce Z (sur la fonction des heuristiques), nous sommes confrontés à un problème de recherche inverse caractérisé (pour la déduction, c'est l'épistémologue soupçonneux qui le pressent ; ici c'est patent). On pourrait objecter qu'il est hâtif de conclure des résultats négatifs de *heuristics and biases* qu'il n'y a rien dans le système cognitif humain qui fasse ce que le système Y devrait faire : Gigerenzer ne montre-t-il pas que lorsqu'on évite d'induire le sujet en erreur, et qu'on déclenche chez lui une démarche fréquentiste, il produit des réponses correctes ?

Sans doute, mais ce n'est pas le même genre de problème qui est posé. Il n'est pas nécessaire d'argumenter dans le détail sur les concepts engagés : il suffit de remarquer qu'il serait *absurde* de postuler une quasi-équivalence au niveau d'une

40. Je ne *crois* pas que la considération des théories non classiques actuellement disponibles modifie radicalement la situation que nous examinons ici, mais ce n'est peut-être que l'effet de mon ignorance en la matière.

41. Un rapprochement avec la théorie de la vision pourrait être instructif : dans ce contexte-là, on n'hésite pas à créditer des composantes du système nerveux de la capacité d'exécuter des fonctions mathématiques fort complexes.

“structure profonde” du problème de formulations délibérément différentes en “surface”, alors même que cette différence a des effets dans les réponses. En outre, il semble plausible que les énoncés de Gigerenzer déclenchent une réaction de type combinatoire qui n’a aucun rapport avec l’évaluation que cherchent à mesurer Kahneman et Tversky. (Le mathématicien serait enclin à estimer que Gigerenzer oblige ses sujets à faire des mathématiques élémentaires, et que le fait qu’elles aient pour objet des dénombrements de cas est sans pertinence ; autrement dit, il se pourrait que le système cognitif G – s’il existe – mis à l’épreuve par Gigerenzer soit celui – s’il existe – qui sous-tend l’aptitude mathématique *en général* !) Enfin, Osherson, par exemple, conjecture qu’il existe non pas un seul système, mais plusieurs, dont l’un serait un bayésianisme mental qu’il importerait d’encourager. Mais comment séparer la contribution des différents systèmes, à moins qu’ils ne forment collectivement un super-système, ou à moins que l’on puisse caractériser indépendamment les aptitudes qu’ils sous-tendent respectivement ? Bref, de quelque côté qu’on se tourne, le problème inverse se pose avec force.

Hasardons-nous alors à demander à quel niveau se situerait le système – Y, Z ou G – qu’il s’agit de cerner grâce aux expériences des psychologues. Si nous nous en tenons à la conception symétrique, il devrait prendre place entre une faculté primitive de saisie de l’incertain, homologue de l’*Urlogique*, et une aptitude de résolution de problèmes d’incertitude. Or que doit contenir cette faculté primitive ? D’abord, on l’a vu, un (ou des) concept(s) précurseur(s) de la probabilité, de l’incertitude, des chances, des propensions, etc. Ensuite, rien de moins que l’*Urlogique* elle-même. Comment, en effet, se passer du concept de la conjonction (ou du moins d’un précurseur), pour penser les chances d’une conjonction (que Linda soit employée de banque *et* militante), du concept de la négation pour penser les chances d’un non-événement, et surtout du concept de conséquence pour penser les relations logiques qui s’établissent entre relations probabilistes ?

A ce point de la discussion, il serait futile de maintenir la fiction de la symétrie. La logique, ou son homologue psychologique – une capacité déductive de base – est un préalable à tout raisonnement, qu’il comporte ou non des arguments non démonstratifs. Raisonner sur l’incertitude, en tant que c’est *raisonner*, n’est pas différent à cet égard de raisonner sur la tâche de Wason, sur les configurations spatiales, sur les relations temporelles, etc. (On opposerait la même objection, si c’était le lieu de le faire, à la manière dont Johnson-Laird traite le “raisonnement spatial” comme s’il était une espèce de raisonnement *à côté* et sur le même plan que le raisonnement propositionnel ou syllogistique.) Bref, il y a de la logique dans tout raisonnement, parce qu’il n’existe pas de raisonnement, quel qu’en soit le contenu, qui n’implique pas de manipulation de propositions complexes dont les articulations sont les chevilles logiques au sens classique.

Inversement, la plupart des raisonnements, y compris ceux qui ne concernent ni de près ni de loin des probabilités, des ordres de grandeur, ou toute espèce d’évaluation d’incertitude, reposent sur une capacité d’évaluation de données incertaines : la pertinence et la fiabilité des prémisses, les chances d’aboutir à une conclusion acceptable et utile, la fiabilité des inférences elles-mêmes, tout cela appelle des évaluations non démonstratives.

Prenons alors un peu de hauteur. Les sciences cognitives fournissent d’une part un cadre très général pour mener l’enquête sur les capacités de l’esprit, et d’autre

part une méthode qui, lorsqu'elle aboutit à la caractérisation d'une aptitude particulière, semble, en tout cas jusqu'à présent, devoir son succès au fait que cette aptitude possède une identité naturelle, qu'elle est naturellement isolée, ou encore *modulaire*, dans la terminologie de Fodor (1983). Ce n'est pas seulement qu'il se trouve que la liste de leurs succès est incluse dans la liste des aptitudes modulaires : en l'absence de modularité, il semble difficile aux sciences cognitives telles que nous les connaissons de même *poser des questions* susceptibles de recevoir une réponse : la modularité apparaît comme une condition de scientificité, un peu comme un critère de démarcation, au sein de la cognition, entre question scientifique et question qui ne l'est pas, ou pas encore. Adopter ce point de vue, ce n'est pas dire que d'autres questions soient dépourvues de sens ; ce n'est pas non plus se rallier à la caractérisation détaillée que Fodor propose des capacités modulaires, ni à sa conception de la distinction (exhaustive) entre processus modulaires et processus centraux⁴². C'est seulement (dans un premier temps du moins) se doter d'une heuristique. Sur le sujet qui nous occupe, cela consiste à se demander, parmi les diverses capacités évoquées dans la discussion, lesquelles ont des chances de donner prise à la méthode expérimentale, et à chercher des candidats parmi celles qui présentent de la manière la plus nette certains traits associés à la modularité, tels que le caractère irrépressible, automatique et rapide, l'isolement informationnel, l'ouverture limitée aux processus centraux, la spécialisation (*domain specificity*) ou encore la forme caractéristique des pannes.

Dans ce que ce crible retient, on ne distingue d'abord que l'*Urlogique* et peut-être la capacité logique de base (CDB). Il laisse filer, naturellement, résolution logique (RL) et résolution de problèmes d'incertitude. Reste la question de ce qui pourrait jouer, du côté de l'incertain, le rôle de UL et de CDB.

On peut penser en effet que des notions telles que le couple probable/improbable et certaines relations d'implication partielle ou d'ordre partiel entre hypothèses ou propositions (*assumptions*) plus ou moins probables font partie d'un équipement de base au même titre que les précurseurs des constantes logiques et de la relation de conséquence logique. Tout cela constituerait un module *Urprobabilité* UP (sans rapport direct, bien entendu, avec une théorie des probabilités même imparfaite ou incomplète) analogue de UL. Un deuxième module aurait pour fonction de proposer aux processus centraux certains résultats concernant des propositions probables particulières et des relations entre ces propositions : ces propositions et relations seraient fournies au module en question par UP, lequel aurait reçu en entrée quelque chose comme les données d'un problème de jugement en situation d'incertitude. (Évitons de spéculer sur la manière dont UL intervient : est-ce en parallèle avec UP, ou avant ? Il importe seulement de ne pas oublier qu'il intervient de quelque façon.) Ce deuxième module, appelons-le *évaluateur automatique de l'incertain* (EAI) – pour éviter quelque chose comme *capacité probabiliste de base* qui risquerait de nous faire pencher vers une théorie naïve des probabilités au sens de la théorie mathématique des probabilités, ce qu'il est méthodologiquement indispensable d'éviter. Qu'allons-nous mettre dans EAI ? Peut-être bien les heuristiques de

42. Sur cette question, voir Sperber (1994) ; Benny Shanon était allé dans le même sens (introduire de la modularité dans les processus centraux) dès 1985 environ.

Tversky et Kahneman, ou quelque chose d'approchant⁴³. Dans ce cas, la théorie des probabilités serait complètement écartée d'un domaine où elle n'aurait peut-être jamais dû occuper une place importante. Après tout, l'expérience XII (où il s'agit d'estimer la factorielle de 8) est peut-être moins marginale qu'il n'apparaît dans un contexte dans lequel on en est venu à parler de jugement probabiliste ou statistique pour désigner le domaine dans son ensemble. Si cette voie devait être poursuivie, elle ferait droit, d'une certaine manière, à la fois aux intuitions de Tversky et Kahneman et à celles de Gigerenzer et de Cohen.

* *

*

Celui qui attendait de la psychologie du raisonnement des bénéfices immédiats en matière de modélisation du raisonnement, au sens où on peut l'entendre en informatique et en intelligence artificielle, celui-là ne peut qu'être déçu. Il ne voit poindre, ni de bonnes raisons de se détourner résolument de la logique, ni les prémices d'une autre logique.

Au contraire, la logique classique semble plus indéradicable que jamais : elle intervient avant même que ne se déclenchent des stratégies de résolution, et elle conditionne ce déclenchement. Primaire, la logique est essentiellement *triviale*⁴⁴ : les deux vont ensemble (et contribuent sans doute conjointement à expliquer la difficulté d'une *théorie* de la logique). Une première leçon semble donc se dégager, purement négative : ne pas chercher à remplacer la logique classique.

Indispensable pour construire, ou valider, tout raisonnement quel qu'il soit, la logique ne suffit en général pas du tout à la tâche. Non seulement, comme on le sait depuis longtemps, parce qu'elle est permissive et doit donc être complétée par un dispositif de pilotage, mais, comme le montre Harman, parce qu'elle est indifférente au poids et à la pertinence des propositions, hypothèses ou conclusions virtuelles, entre lesquelles elle ne fait qu'établir un certain type de relations, relations qui peuvent entrer en conflit interne (contradiction logique ou externe (tension avec des croyances ou contraintes d'arrière-plan)). Une deuxième leçon, non moins négative que la première, serait qu'il est vain d'espérer modéliser le raisonnement, *même déductif*, à l'aide seulement de la logique classique et de stratégies de pilotage.

Enfin, puisque tout raisonnement fait appel à des arguments non démonstratifs, il est essentiel de comprendre ce qu'est un tel argument ; or, et c'est une troisième leçon négative, ce n'est pas la théorie des probabilités qui semble pouvoir nous y aider.

Mais il y a peut-être des bénéfices indirects à retirer de l'analyse des tentatives des psychologues. En matière de logique, d'abord, l'inamovibilité de la logique classique est compatible avec son enrichissement, et avec une multiplicité de tels

43. Il faudrait également examiner les "modèles mentaux probabilistes" de Gigerenzer *et al.* (1991), ce que je n'ai pas eu le loisir de faire.

44. C'est ce qu'avait fort bien vu Popper, redécouvrant en Nouvelle-Zélande, dans l'immédiat après-guerre, la déduction naturelle (*cf.* Popper 1948).

enrichissements, qui seraient autant de modules placés en aval de la logique primaire ou *Urlogique*. Cette disposition autorise notamment un dédoublement de niveaux, dans lequel l'*Urlogique* serait reflétée dans un module en aval, son image réfléchie entrant dans ce module en interaction avec d'autres primitives, propres au module. La question de savoir quel statut donner à ces autres primitives – sont-elles de nature logique, sont-elles de l'ordre d'axiomes spécifiques, sont-elles de l'ordre du "régime inférentiel" (au sens de Barwise 1986 et 1987) qui conditionnerait la nature du raisonnement ? – tout cela est ouvert : en plaçant la logique classique dans une sorte de sanctuaire, on gagne une liberté à peu près totale dans la conception de systèmes cognitifs de raisonnement, naturels ou artificiels – c'est un autre bénéfice indirect du travail des psychologues. De plus, c'est le module aval qui va déterminer la "coloration" du système de raisonnement dans son ensemble ; il peut aller jusqu'à masquer la nature logique du module amont. On peut donc à la fois faire droit au rôle de la logique et militer contre l'idée d'un raisonnement "essentiellement" logique. Enfin, rien ne s'oppose à l'idée d'une pluralité des raisonnements.

S'il y a une idée vers laquelle les efforts des psychologues, des informaticiens, des logiciens, des philosophes et des pragmaticiens semblent converger, c'est que nous sommes encore très ignorants sur la nature du raisonnement. En particulier, nous prenons conscience du rôle-clé de l'inférence non démonstrative, non seulement, en grand, comme une *sorte* ou un *domaine* de raisonnement, mais aussi, en petit, comme une *composante* intime de tout raisonnement. Or ce qu'est une inférence non démonstrative, nous n'en savons encore pas grand-chose.

Bibliographie

- Andler, D. (1994), "A quoi pensent les sciences cognitives ?", *Raison présente* 109, 29-50
- Bar-Hillel, M. (1980), "The base-rate fallacy in probability judgments", *Acta Psychologica* 44, 211-33.
- Barwise, J. (1986), "Information and circumstance", *Notre-Dame J. of Formal Logic*, 27, p. 324-338.
- Barwise, J. (1987), "Unburdening the language of thought", *Mind and Language* 2, 82-96.
- Barwise, J. & Perry, J. (1983), *Situations and Attitudes*, MIT Press, Cambridge, Mass.
- Braine, M.D.S. (1978), "On the relation between the natural logic of reasoning and standard logic", *Psychological Review*, 85, 1-21.
- Braine, M.D.S. (1990), "The «Natural Logic» approach to reasoning", in W.F. Overton, ed., *Reasoning, Necessity, and Logic : Developmental Perspectives*, Lawrence Erlbaum, Hillsdale, NJ.
- Braine, M.D.S. (1992), "Approches empiriques du langage de la pensée", in D. Andler *et al.*, dir., *Épistémologie et cognition*, P. Mardaga, Louvain.
- Braine, M.D.S., Reiser, B.J. & Rumain B. (1984), "Some empirical justifications for a theory of natural propositional logic", in G.H. Bower (ed.), *The Psychology of Learning and Motivation*, vol. 18, Academic Press, New York.

- Casscells, W., Schoenberger, A. & Grayboys, T. (1978), "Interpretation by physicians of clinical laboratory results", *New England Journal of Medicine* **299**, 299-300.
- Cheng, P. & Holyoak, K.J. (1985), "Pragmatic reasoning schemas", *Cognitive Psychology*, **17**, 391-416.
- Cherniak, Ch. (1986), *Minimal Rationality*, MIT Press, Cambridge, Mass.
- Cohen, L. J. (1970), *The Implications of Induction*, Methuen, Londres.
- Cohen, L. J. (1977), *The Probable and the Provable*, Clarendon Press, Oxford.
- Cohen, L. J. (1981), "Can human irrationality be experimentally demonstrated ?", with Open Peer Commentary, *Behavioral and Brain Sciences* **4**, 317-70.
- Cohen, L. J. (1982), "Are people programmed to commit fallacies ? Further thoughts about the interpretation of experimental data on probability judgements", *Journal for the Theory of Social Behaviour* **12**, 251-74.
- Cohen, L. J. (1986), *The Dialogue of Reason. An Analysis of Analytical Philosophy*, Clarendon Press, Oxford : §§ 16-19.
- Cohen, L.J. (1989), *An Introduction to the Philosophy of Induction and Probability*, Clarendon Press, Oxford.
- Cosmides, L. (1989), "The logic of social exchange : Has natural selection shaped how humans reason ?", *Cognition* **31**, 187-276.
- Cosmides, L. & Tooby, J. (1990), "Is the mind a frequentist ?", communication présentée au Second Annual Meeting of the Human Behavior and Evolution Society, Los Angeles
- Evans, J. St. B.T. (1989), *Bias in human reasoning : Causes and consequences*, Lawrence Erlbaum, Hillsdale, NJ.
- Evans, J. St B.T. (1991), "Theories of human reasoning : The fragmented state of the art", *Theory and Psychology* **1**, 83-105.
- Evans, J. St. B.T. (1994), "Bias and rationality", in Manktelow, K.I. & Over, D.E., eds., *Rationality*, Routledge, Londres.
- Evans, J. St. B.T. (1994a), "Bias in thinking and judgement", in Keane, M.T. & Gilhooly, K.J., eds., *Advances in the Psychology of Thinking (vol. 1)*, Harvester-Wheatsheaf.
- Evans, J. St. B.T., Newstead, S.E., & Byrne, R. M. J. (1993), *Human Reasoning*, Lawrence Erlbaum, Hillsdale, NJ.
- Fischhoff, B. (1988) "Judgment and decision making", in Sternberg, R. & Smith, E., eds., *The Psychology of Human Thought*, Cambridge University Press, Cambridge., 153-187.
- Fodor, Jerry A. (1983). *The Modularity of Mind*, MIT Press, Cambridge, Mass. Trad. fr. *La Modularité de l'esprit*, Minuit, Paris, 1986.
- Gamut, L.T.F. (1991), *Introduction to Logic* (vol. 1 : *Logic, Language, and Meaning*, vol. 2 : *Intensional Logic and Logical Grammar*), Chicago University Press, Chicago.
- Gärdenfors, P. (1988), *Knowledge in Flux*, MIT Press, Cambridge, Mass.
- Gigerenzer, Gerd (1991), "How to make cognitive illusions disappear : Beyond «Heuristics and Biases»", in Stroebe, W. & Hewstone M., *European Review of Social Psychology*, v. **2**, John Wiley & Sons, Chichester, 83-115.

- Gigerenzer, Gerd, Hoffrage, Ulrich & Kleinbölting, Heinz (1991), "Probabilistic mental models : A Brunswickian theory of confidence", *Psychological Review* 98, 506-528.
- Gigerenzer, G. & Hug, K. (1992), "Domain-specific reasoning : Social contracts, cheating, and perspective change", *Cognition* 43, 127-71.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989), *The Empire of Chance : How Probability Changed Science and Everyday Life*, Cambridge University Press, Cambridge
- Gilhooly, K.J., Keane, M.T.G., Logie, R.H., Erdos, G. (1990), *Lines of Thinking : Reflections on the Psychology of Thought*, vol 1 : *Representation, Reasoning, Analogy and Decision Making*, John Wiley & Sons, Chichester
- Giroto, V. (1991), "Reasoning on deontic rules : The pragmatic schemas approach", in Politzer (1991).
- Goldman, A. (1986), *Epistemology and Cognition*, Harvard University Press, Harvard.
- Grice, H.P. (1989), *Studies in the Way of Words*, Harvard University Press, Cambridge, Mass.
- Halpern, J., ed. (1986), *Theoretical Aspects of Reasoning About Knowledge*, Morgan Kaufmann, San Francisco.
- Harman, G. (1986), *Change in View*, MIT Press, Cambridge, Mass.
- Henle, M. (1962), "On the relation between logic and thinking", *Psychological Review* 69, 366-378.
- Hirschfeld, L.A. & Gelman, S.A., eds. (1994), *Mapping the Mind. Domain Specificity in Cognition and Culture*, Cambridge University Press, Cambridge.
- Holland, J.H., Holyoak K.J., Nisbett, R.E. & Thagard, P.R. (1986), *Induction : Processes of Inference, Learning, and Discovery*, MIT Press, Cambridge, Mass.
- Holyoak, K.J. (1990), "Problem solving", in Osherson & Smith (1990).
- Johnson-Laird, P. N. (1983), *Mental models : Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge.
- Johnson-Laird, P. N. (1986), "Reasoning without logic", in *Reasoning and Discourse Processes*, Academic Press, Londres
- Johnson-Laird, P. N. (1988), *The Computer and the Mind. An Introduction to Cognitive Science*, Harvard University Press, Cambridge, Mass.
- Johnson-Laird, P. N. & Byrne, R. M. J. (1991), *Deduction*, Lawrence Erlbaum, Hillsdale, NJ.
- Johnson-Laird, P. N. & Byrne, R. M. J. (1993), "Précis of *Deduction*", with Open Peer Commentary, *Behavioral and Brain Sciences* 16, 323-80.
- Kahneman, D. & Tversky, A. (1972), "Subjective probability : A judgment of representativeness", *Cognitive Psychology* 3, 430-54.
- Kahneman, D. & Tversky, A. (1972a), "On the psychology of prediction", *Oregon Research Institute Research Bulletin* 12.
- Kahneman, D. & Tversky, A. (1973), "On the psychology of prediction", *Psychological Review* 80, 237-51.
- Kahneman, D., Slovic, P. & Tversky, A., eds. (1982), *Judgment under Uncertainty : Heuristics and Biases*, Cambridge University Press, Cambridge.

- Lopes, Lola L. (1991), "The rhetoric of irrationality", *Theory and Psychology* 1, 65-82.
- Macnamara, J. (1986), *A Border Dispute : The Place of Logic in Psychology*, MIT Press, Cambridge, Mass.
- Manktelow, K.I. & Over, D.E. (1990), *Inference and Understanding : A Philosophical and Psychological Perspective*, Routledge, Londres.
- Manktelow, K.I. & Over, D.E., eds. (1993), *Rationality*, Routledge, Londres.
- Meehl, P. & Rosen, A. (1955), "Antecedent probability and the efficiency of psychometric sings, patterns, and cutting scores", *Psychological Bulletin* 52, 194-215.
- Newell, A. (1973), "Production systems : Models of control structures", in Chase, W.G., ed., *Visual Information Processing*, Academic Press, New York.
- Newell, A., & Simon, H.A. (1972), *Human Problem Solving*, Prentice Hall, Englewood Cliffs, NJ.
- Nisbett, R. & Ross, L. (1980), *Human Inference : Strategies and Shortcomings of Social Judgment*, Prentice-Hall, Englewood Cliffs, NJ.
- Noveck, I., Lea, R.B., Davidson, G.M., & O'Brien, D.P. (1991), "Human reasoning is both logical and pragmatic", in Politzer (1991).
- Oaksford, M. & Chater, N. (1992), "Bounded rationality in taking risks and drawing inferences", *Theory and Psychology* 2, 225-230.
- Osherson, D. (1975), "Logic and models of logical thinking", dans R.J. Falmagne (ed), *Reasoning : Representation and Process*, Wiley, New York.
- Osherson, D. (1990), "Judgment", in Osherson & Smith (1990), 55-87.
- Osherson, D. & Smith, E., eds. (1990), *Thinking*, MIT Press, Cambridge, Mass.
- Peacocke, C. (1994), "Content, computation and externalism", *Mind and Language* 9.
- Politzer, G, dir. (1991), Numéro spécial "Pragmatique et psychologie du raisonnement", *Intellectica* 11 [revue de l'ARC, Association pour la recherche cognitive].
- Popper, Karl R. (1948), "The trivialization of logic".
- Popper, Karl R. (1983), *Realism and the Aim of Science*, Hutchinson, Londres ; trad. fr. *Le réalisme et la science*, Hermann, Paris, 1990.
- Rips, L.J. (1986). "Mental Muddles", in M. Brand & M. Harnish, eds., *The Representation of Knowledge and Belief*, University of Arizona Press, Tucson, 258-86.
- Rips, L.J. (1988). "Deduction", in Sternberg, R. & Smith, E., eds., *The Psychology of Human Thought*, Cambridge University Press, Cambridge., 116-152.
- Sainsbury, M. (1991), *Logical Forms*, Blackwell, Oxford.
- Scholtz, R.W., ed. (1983), *Decision Making Under Uncertainty*, Elsevier, Amsterdam.
- Smith, E.E., Langston, Ch., & Nisbett, R. (1992), "The case for rules in reasoning", *Cognitive Science* 16, 1-40.
- Sperber, D. (1994), "The modularity of thought and the epidemiology of representations", in Hirschfeld & Gelman.
- Sperber, D., Cara, F., & Girotto, V. (à paraître), "Relevance Theory explains the selection task" (manuscrit, juin 1994).

- Sperber, D. & Wilson, D. (1986), *Relevance : Communication and Cognition*, Blackwell, Oxford ; tr. fr. *La pertinence : Communication et cognition*, Minuit, Paris, 1989.
- Stalnaker, R. (1984), *Inquiry*, MIT Press, Cambridge, Mass.
- Stich, S. (1990), *The Fragmentation of Reason*, MIT Press, Cambridge, Mass.
- Tversky, A. & Kahneman, D. (1971), "Belief in the law of small numbers", *Psychological Bulletin*, **76**, 105-10.
- Tversky, A. & Kahneman, D. (1973), "Availability : A heuristic for judging frequency and probability", *Cognitive Psychology* **5**, 207-32.
- Tversky, A. & Kahneman, D. (1974), "Judgment under Uncertainty : Heuristics and Biases", *Science* **125**, 1124-31.
- Tversky, A. & Kahneman, D. (1982), "Judgments of and by representativeness", in Kahneman, Slovic & Tversky 1982, 84-98.
- Tversky, A. & Kahneman, D. (1982a), "Evidential impact of base rates", in Kahneman, Slovic & Tversky 1982, 153-160.
- Tversky, A. & Kahneman, D. (1983), "Extensional *versus* intuitive reasoning : The conjunction fallacy", *Psychological Review* **90**, 292-315.
- Wason, P.C. (1966), "Reasoning", in B.M. Foss, ed., *New Horizons in Psychology*, Penguin, Hammondsworth.
- Wason, P.C. & Johnson-Laird, P.N. (1972), *Psychology of Reasoning : Structure and Content*, Batsford, Londres.

Daniel ANDLER

Chapitre 2

Confirmation non probabiliste

1. Introduction

Dans le langage courant, il est fréquent de dire ou de penser : "Cela *confirme* ce que je pensais". De même, dans la pratique scientifique, on peut commenter une expérience ayant donné des résultats prévus par une théorie T en affirmant : "Cette expérience confirme la théorie T". Dans les deux cas, on cherche à exprimer l'idée que certains faits renforcent notre croyance dans une hypothèse, voire rendent cette hypothèse très plausible. Par exemple, si je pense que les philosophes sont généralement myopes et qu'on me présente quelqu'un qui se dit philosophe et porte des lunettes, cela confirmera ma croyance en la myopie des philosophes (sans toutefois la transformer en certitude). La découverte de la planète Neptune par Le Verrier a conduit, selon un raisonnement de nature similaire au précédent, à confirmer les lois de la gravitation de Newton qui prédisaient son existence.

Cette notion de confirmation, apparemment très simple et facile à appliquer, est celle qui vient spontanément à l'esprit du philosophe des sciences moderne qui, au moins depuis Karl Popper [1934], a fait son deuil de toute idée de certitude scientifique. Les théories scientifiques, en effet, se présentent comme des hypothèses "universelles" : elles énoncent qu'un ensemble infini (ou extrêmement grand) d'objets possèdent les mêmes propriétés. Par exemple : "tous les objets physiques subissent l'action des forces gravitationnelles". Pour établir avec certitude la vérité de telles hypothèses, il faudrait soit un temps infini, soit un temps de toute façon trop grand pour être accessible à la communauté scientifique. La seule certitude qu'on puisse obtenir éventuellement au sujet des théories scientifiques, c'est celle de leur fausseté : si je trouve un objet physique qui ne subit pas l'action

des forces gravitationnelles, alors mon hypothèse est certainement fausse. Mais que peut-on dire des hypothèses qui n'ont pas été ainsi réfutées par une expérience ? Sont-elles toutes cognitivement équivalentes, dans la mesure où toutes sont possibles mais aucune certaine ? Une telle conclusion serait bien décevante : elle reviendrait à accorder le même statut à une nouvelle théorie qui n'a encore fait l'objet d'aucun test expérimental, et, par exemple, à la mécanique quantique, qui, bien que non certaine, a passé avec succès de très nombreux tests et a fait de nombreuses prédictions victorieuses. Il nous faut un concept pour rendre compte de cette différence : celui de confirmation. Ce qui distingue aux yeux de tous la mécanique quantique de notre théorie juvénile peut s'énoncer en disant que la mécanique quantique est fortement confirmée par l'expérience alors que la théorie nouvelle n'est qu'une conjecture qui demande à être confirmée.

La notion de confirmation est ainsi le substitut indispensable de la notion de vérification (entendue comme preuve de la vérité d'une théorie), dès qu'on a compris que celle-ci est un idéal inaccessible à la science. Ce substitut est particulièrement précieux pour les épistémologues, dont la tâche est d'analyser la nature de la connaissance scientifique. Mais pour qu'il puisse remplir efficacement sa mission de substitut, le concept de confirmation doit d'abord être défini clairement : on doit disposer d'un critère qui départage les cas authentiques de confirmation, des cas fallacieux. Ensuite, il serait souhaitable de formuler des règles de "raisonnement correct" lorsqu'on remplace la vérification par la confirmation. Par exemple, lorsqu'il est prouvé que deux énoncés p et q sont vrais, on sait que l'énoncé " p et q " est vrai. Mais " p et q " est-il nécessairement confirmé par un énoncé r , si p et q le sont séparément ? On verra que la réponse n'est pas évidente.

Autrement dit, la tâche de l'épistémologue est de *formaliser* le concept intuitif, apparemment simple, de confirmation, et d'essayer de construire une *logique de la confirmation*, à la fois cohérente et qui rende compte le mieux possible des intuitions attachées à ce concept.

Un des premiers auteurs à avoir tenté de formaliser rigoureusement la confirmation est Carl Hempel [1945]. Celui-ci a essayé d'exhiber un critère qualitatif reposant seulement sur la syntaxe du langage concerné et permettant de dire dans quels cas un énoncé donné confirmait un autre énoncé. Mais cette tentative était trop limitée et n'a pas réussi à modéliser correctement le concept, tout en étant en butte à certains paradoxes¹. Ce n'est que plus tard qu'on s'est rendu compte que la formalisation de la confirmation nécessitait une approche quantitative à travers la définition d'un degré de confirmation. De nombreuses propositions ont été avancées dans les années cinquante et le début des années soixante (Popper [1954, 1955-6], Good [1950, 1955], Carnap [1950, 1956-7], Levi [1962]) et ont conduit à des débats animés (tout spécialement le célèbre débat entre Carnap, Popper et Bar Hillel pour lequel on pourra consulter Bar Hillel [1974], Stopes-Roy [1968] ou Baillie [1969]). En fait les désaccords provenaient d'une insuffisante analyse du concept de confirmation.

1. Ces paradoxes consistaient en ceci que les critères retenus conduisaient à des cas de confirmation intuitivement fallacieux. Le plus célèbre est le paradoxe des corbeaux (Hempel [1945]).

La construction d'une théorie formelle de la confirmation nécessite en effet tout d'abord de clarifier la notion de confirmation, qui est ambiguë dans le langage courant. Ce problème a été quasiment résolu par Carnap [1962], bien que toute trace d'ambiguïté n'ait pas disparu dans certains débats plus récents sur la notion de confirmation. Lorsqu'on dit qu'une proposition en confirme une autre, on peut faire référence :

- a) à un concept *numérique* de confirmation : on désigne alors ainsi le degré de confirmation qu'une proposition peut conférer à une autre, degré "neutre" qui peut être faible ou élevé comme un degré de température.
- b) à un concept *comparatif* de confirmation : on désigne alors l'accroissement du degré de confirmation d'une proposition par une autre, ou encore le "soutien" qu'une proposition apporte à une autre dans une certaine métrique de confirmation. On peut parler dans ce cas de "confirmation relative" d'une proposition par une autre.
- c) à un concept *classificatoire* de confirmation : on désigne alors le fait qu'une proposition est "acceptable", sans être certaine, son degré de confirmation étant suffisamment élevé. On peut parla dans ce cas de "confirmation absolue" d'une proposition par une autre.

Une théorie complète de la confirmation devra rendre compte à la fois de la confirmation relative et de la confirmation absolue, en s'appuyant éventuellement (mais pas nécessairement) sur une métrique des degrés de confirmation.

En particulier, la question de l'acceptation (confirmation absolue) est fondamentale pour comprendre dans quelles circonstances les scientifiques peuvent travailler avec des hypothèses (pour en évaluer d'autres par exemple) sans que ces hypothèses soient jamais certaines. L'étude de la confirmation relative doit, elle, nous permettre de reconstruire la rationalité de l'apprentissage progressif par l'expérience, et d'éclairer les circonstances dans lesquelles on dit qu'un test est favorable à une hypothèse sans pour autant que cette hypothèse soit encore admise.

À la suite des éclaircissements de Carnap, le problème de construire un modèle formel adéquat pour la théorie de la confirmation a été abordé par de très nombreux épistémologues sous des angles divers mais fréquemment en proposant des modèles quantitatifs en termes de degré de confirmation. De plus, comme le calcul des probabilités était l'outil le plus commode à disposition, le degré de confirmation a été construit dans la plupart des cas, à partir des probabilités des énoncés. Cependant, on peut montra que le cadre des probabilités est insuffisant pour construire une formalisation adéquate (voir & 3 et Zwirn et Zwirn [à paraître]).

Certains auteurs ont depuis longtemps diagnostiqué cette insuffisance et proposé des modèles non probabilistes de la relation de confirmation. Ainsi, Levi [1967, 1986] a proposé d'interpréter certains aspects de la confirmation en termes de degrés de surprise potentielle de Shackle [1961] ou de fonctions de support de Shafer [1976]. Cohen [1966 a,b] a élaboré une fonction de confirmation dont les axiomes quantitatifs ressemblent formellement à ceux des nécessités et dont la contrepartie qualitative est une logique modale généralisée. Toutefois aucun de ces deux auteurs n'envisage le problème de la dualité de la notion de confirmation : les principes d'adéquation qu'ils retiennent ne concernent que la logique de l'acceptabilité.

En particulier, la question de l'acceptation (confirmation absolue) est fondamentale pour comprendre dans quelles circonstances les scientifiques peuvent travailler avec des hypothèses (pour en évaluer d'autres par exemple) sans que ces hypothèses soient jamais certaines. L'étude de la confirmation relative doit, elle, nous permettre de reconstruire la rationalité de l'apprentissage progressif par l'expérience, et d'éclairer les circonstances dans lesquelles on dit qu'un test est favorable à une hypothèse sans pour autant que cette hypothèse soit encore admise.

Par ailleurs, les travaux plus récents développés depuis une dizaine d'années en Intelligence Artificielle, pour modéliser les raisonnements non monotones et les raisonnements incertains, et pour montrer les rapprochements qui existent entre ces deux types de raisonnements, offrent un champ de modélisation nouveau pour la notion de confirmation. Ces travaux conduisent à de très nombreux résultats de convergence, par exemple entre différents modèles de logique non monotones (Gabbay [1985], Makinson [1989], Kraus & al. [1990]), entre ces systèmes de logique non monotones et la sémantique préférentielle de Shoham [1988], entre logiques non monotones et logique des conditionnels contrefactuels de D. Lewis [1973] (Delgrande [1988], Crocco et Lamarre [1991]), et, sur le plan numérique, entre certaines logiques non monotones et la théorie des possibilités de Dubois & Prade [1986] (Dubois & Prade [1991]).

Notre objectif est d'abord de nous placer dans le cadre de la théorie des possibilités de Dubois & Prade, qui présente l'avantage majeur de fournir une définition claire de la notion de possibilité *conditionnelle*, et de voir si elle fournit, contrairement aux probabilités, un modèle quantitatif adéquat à la fois pour la confirmation absolue et pour la confirmation relative. Pour cela, il nous faudra en premier lieu rappeler les principes d'adéquation traditionnellement retenus en théorie de la confirmation. Certains des résultats de convergence évoqués plus haut nous permettront de disposer d'une "palette interprétative" beaucoup plus riche, et notamment de relier la sémantique numérique de la confirmation à sa contrepartie qualitative (en termes de mondes préférentiels par exemple) et à sa contrepartie syntaxique (c'est-à-dire aux règles de raisonnement).

Il est important d'insister sur le fait que la notion de confirmation est d'une nature purement philosophique et que notre propos ici n'est pas d'apporter des résultats techniques nouveaux dans le domaine de la non-monotonie ou de l'incertitude, mais de montrer comment certains des progrès réalisés dans ce domaine peuvent être *utilisés* pour renouveler une problématique très ancienne en philosophie des sciences.

2. Principes d'adéquation

Tous les auteurs qui proposent une théorie de la confirmation se réfèrent, implicitement ou explicitement, à un certain nombre de *principes d'adéquation* qu'une "bonne" définition de la notion de confirmation est censée respecter.

Comme il est courant en philosophie, il n'existe pas de réel consensus sur ces principes d'adéquation eux-mêmes. Toutefois, il est possible de clarifier le paysage en procédant aux deux distinctions suivantes :

a) Il faut d'abord distinguer les principes qui s'appliquent exclusivement à la confirmation des hypothèses universelles par des énoncés singuliers, de ceux qui s'appliquent à la confirmation d'une proposition par une autre proposition.

b) Il faut ensuite, et crucialement, distinguer les principes d'adéquation pertinents pour la confirmation absolue de ceux qui sont pertinents pour la confirmation relative.

L'utilisation de ces deux distinctions permet de réduire la plupart des divergences entre les listes de principes rencontrées dans la littérature confirmationniste, soit à des différences de champ d'application, soit à des préférences pour l'importance philosophique de l'un des principes.

Dans la mesure où nous limitons pour le moment notre étude à la confirmation d'un énoncé quelconque par un autre énoncé quelconque, la première distinction permet de s'affranchir de principes du type "Instantial Comparability Condition" [Cohen 1966 b] : *une hypothèse universelle a un degré de confirmation supérieur à une autre si et seulement si toutes les instances de la première ont un degré de confirmation supérieur à toutes les instances de la seconde.*

On peut alors montrer, en se fondant à la fois sur des arguments syntaxiques liés à la compatibilité logique entre les principes possibles et sur des arguments épistémologiques liés à la pertinence des principes pour chacune des deux acceptions du terme "confirme", que les principes d'adéquation se répartissent en deux ensembles consistants, l'un pour la confirmation absolue, l'autre pour la confirmation relative, chacun d'entre eux étant maximum dans le sens qu'il est impossible de lui rajouter un principe non inclus appartenant à l'autre. Nous renvoyons à Zwirn & Zwirn [à paraître] sur ce point, ainsi que pour une justification du choix des principes d'adéquation pour la confirmation absolue et la confirmation relative.

2.1. Principes d'adéquation de la confirmation absolue

Les principes majeurs d'adéquation de la confirmation absolue sont principalement ceux définis par Hempel dans son travail sur la définition purement syntaxique de la confirmation (Hempel [1943]).

Nous utiliserons le symbole " $\rightarrow$ " pour désigner la relation de confirmation entre deux énoncés.

A chaque fois que cela sera possible, nous utiliserons l'appellation courante en logique non monotone pour désigner les règles formellement équivalentes aux principes énumérés ci-dessous ; entre crochets nous rappellerons les appellations utilisées par Hempel et plus connues en théorie de la confirmation.

P.1. SUPRA-CLASSICALITE [Condition d'implication]

$$\frac{A \rightarrow B}{A \rightarrow B}$$

P.2. CONSISTANCE [*conséquence n°2 de la Condition de Consistance Spéciale*]

$$\frac{A \rightarrow \perp}{A \equiv \perp}$$

P.3. AFFAIBLISSEMENT A DROITE [*Condition de Conséquence Spéciale*]

$$\frac{A \rightarrow B \quad \vdash \quad B \rightarrow C}{A \rightarrow C}$$

P.4. ET [*Condition de Conjonction*]

$$\frac{A \rightarrow B \quad A \rightarrow C}{A \rightarrow B.C}$$

De ces quatre Principes "majeurs", on peut bien entendu déduire de nombreuses autres propriétés dont certaines ont un intérêt philosophique particulier pour la notion de confirmation, dans la mesure où elles explicitent certaines intuitions fortes ou certains desiderata qui sont consubstantiels à cette notion. Par exemple, à partir de la Consistance et de la règle Et, on peut démontrer :

P.5. SELECTIVITE [Hempel 1943, *conséquence n°7 de la Condition de Consistance Spéciale*]

$$\frac{A \rightarrow B \quad A \rightarrow \bar{B}}{A \equiv \perp}$$

P.6. EQUIVALENCE A DROITE [Hempel 1943]

$$\frac{A \rightarrow B \quad B \equiv C}{A \rightarrow C}$$

De l'Affaiblissement à droite et la règle du Et on déduit les trois principes suivants dont on peut montrer qu'ils sont équivalents :

P.7. PRINCIPE DE LEVI [Levi 1986]

$$\frac{A \rightarrow B \quad A.B \equiv A.C}{A \rightarrow C}$$

P.8. AMPLIATIVITE [Popper & Miller 1983] :

$$\frac{A \rightarrow B}{A \rightarrow (A \rightarrow B)} \quad \text{et} \quad \frac{A \rightarrow (A \rightarrow B)}{A \rightarrow B}$$

P.9. REFERENTIALITE [Barwise & Cooper 1981, van Benthem 1984, Dubucs 1990]

$$\frac{A \leadsto A.B}{A \leadsto B} \quad \text{et} \quad \frac{A \leadsto B}{A \leadsto A.B}$$

Pour déterminer l'adéquation d'une définition de la confirmation absolue, les quatre principes majeurs énumérés sont suffisants puisqu'ils nous garantissent la réalisation des autres propriétés désirables.

2.2. Principes d'adéquation de la confirmation relative

Une définition de la confirmation relative est syntaxiquement adéquate si elle respecte un certain nombre de principes déjà énumérés pour la confirmation absolue. Ces principes communs sont les suivants :

P.1. SUPRA-CLASSICALITE

P.5. SELECTIVITE

P.6. EQUIVALENCE A DROITE²

On en déduit trivialement le principe de Consistance P.2.

En plus de ces principes communs, on attend en général d'une définition de la confirmation relative qu'elle respecte le principe suivant :

P.10. CONVERSION

$$\frac{A \vdash B}{B \leadsto A}$$

Ce principe exprime l'intuition, largement partagée en épistémologie, selon laquelle les succès prédictifs d'une hypothèse accroissent notre confiance dans cette hypothèse.

On a parfois considéré (Hempel [1943]) que ce principe, bien qu'il soit épistémologiquement désirable, était "de trop" puisqu'il est contradictoire avec le principe d'Affaiblissement à droite (sous couvert du principe de Sélectivité). Mais en fait cette incompatibilité logique n'existe plus dès lors qu'on distingue clairement la confirmation absolue, qui doit respecter l'Affaiblissement à droite, de la confirmation relative, qui doit respecter la Conversion.

Sans aucun principe de conversion, notre théorie de la confirmation relative serait très incomplète : comment peut-on accroître notre confiance dans une hypothèse, si ce n'est en vérifiant ses prédictions ?

2. Les principes P.5 et P.6 font ici partie des principes majeurs dans la mesure où le retrait de l'affaiblissement à droite et du ET les retire de la liste des principes démontrables.

On peut toutefois admettre, sans pour l'instant en préciser la forme syntaxique, qu'une version du principe de conversion réduite à *certaines* conséquences de l'énoncé confirmé serait acceptable, voire préférable : cette restriction devrait notamment permettre de limiter le soutien aux seuls cas de tests *sévères* d'un énoncé. Nous reviendrons sur ce point lorsque nous envisagerons la formulation d'un critère de soutien possibiliste.

3. Critères probabilistes

De très nombreuses mesures probabilistes du *degré de confirmation* ont été proposées. Le choix de la bonne mesure est à l'origine d'une controverse assez violente opposant Popper à Carnap et Bar-Hillel à la fin des années 50. En fait, une fois la distinction entre confirmation absolue et confirmation relative clairement établie, il s'avère qu'il n'y a que deux *définitions* intéressantes possibles de la relation de confirmation³:

a) Le critère du seuil : $a \leadsto b$ si et seulement si $P(b,a) > S > 0,5$.

b) Le soutien probabiliste : $a \leadsto b$ si et seulement si $P(b,a) - P(b) > 0$.

Le premier critère respecte les principes P.1, P.2, P.3, mais ne respecte pas le principe P.4 de conjonction (paradoxe de la loterie). (On peut démontrer d'ailleurs qu'aucun critère probabiliste raisonnable⁴ ne pourrait satisfaire ce principe (Zwirn & Zwirn [à paraître]). Il n'est donc pas totalement adéquat pour modéliser la confirmation absolue.

Le second critère respecte les principes P.1, P.5, P.6 et P.10. Il est donc adéquat pour modéliser la confirmation relative.

Par conséquent, le cadre probabiliste permet de représenter de manière satisfaisante la notion de confirmation relative, mais ne permet pas de donner une définition adéquate de la confirmation absolue.

Ce bilan apparemment mitigé est plus négatif qu'il n'y paraît : une théorie adéquate de l'acceptation sans théorie du soutien serait épistémologiquement incomplète mais constituerait néanmoins une construction fondamentale. Une théorie du soutien sans théorie de l'acceptation n'a par contre quasiment aucun intérêt : à quoi sert il de savoir quels arguments peuvent renforcer une hypothèse, si aucun ensemble d'arguments ne peut jamais être considéré comme suffisant pour adopter cette hypothèse ?

3. Deux mesures différentes du degré de confirmation peuvent aboutir à une définition équivalente de la relation de confirmation. Par exemple, si on mesure le degré de confirmation par le "weight of evidence" [Good 1950] $w(b,a) = \text{Log} [p(b,a) (1-p(b)) / p(b)(1-p(b,a))]$, et qu'on définit $a \leadsto b$ par $w(b,a) > 0$, on obtient un critère de confirmation relative strictement équivalent au soutien probabiliste simple : $p(b,a) - p(b) > 0$, qui résulte de la mesure directe du degré de confirmation par la probabilité conditionnelle.

4. Le terme "raisonnable" peut être précisé rigoureusement.

Il est donc nécessaire d'envisager la construction d'un critère de confirmation dans un cadre non probabiliste. C'est ce que nous allons faire en nous appuyant sur la théorie des possibilités-nécessités.

Préalablement, nous allons replacer les principes d'adéquation de la confirmation absolue dans le cadre des règles générales du raisonnement non monotone.

4. Confirmation absolue

4.1. Syntaxe : confirmation absolue et logique non monotone

La notion de confirmation absolue est intuitivement très voisine de la notion de conséquence non monotone : l'énoncé "h est acceptable en présence des faits e" pourrait être traduit "si e alors h (à moins que h', ce qui est peu plausible)". Le fait d'accepter une hypothèse empirique dans le corpus des théories scientifiques n'exclut jamais la possibilité d'une nouvelle expérience venant remettre en cause cette acceptation. La confirmation d'une hypothèse est donc toujours potentiellement sujette à révision. Mais elle implique qu'on croie peu à cette révision ultérieure.

Pour vérifier cette connexité, il faut s'assurer que les règles auxquelles obéit l'opérateur " $\sim>$ " dans les systèmes de logique non monotone connus assurent le respect des principes d'adéquation P.1 à P.4.

Pour cela, nous nous appuyerons sur la synthèse de Kraus B. et *al.* [1990], qui, à la suite des travaux de Gabbay [1985] et Makinson [1989], définissent une relation d'inférence non monotone " $\sim>$ " comme *cumulative* si elle respecte l'axiome suivant :

A.1. REFLEXIVITE : $A \sim> A$

et les règles suivantes :

R.1 . AFFAIBLISSEMENT A DROITE [cf. P.3]

R.2. ÉQUIVALENCE LOGIQUE A GAUCHE

$$\frac{A \sim> C \quad A \equiv B}{B \sim> C}$$

R.3. COUPURE

$$\frac{A.B \sim> C \quad A \sim> B}{A \sim> C}$$

R.4. MONOTONIE FAIBLE

$$\frac{A \leadsto B \quad A \leadsto C}{A.B \leadsto C}$$

Si on suppose que le principe de consistance (P.2) est respecté, on montre que toute relation d'inférence cumulative est adéquate du point de vue de la confirmation absolue. En effet :

Réflexivité (A.1) + Affaiblissement à droite (R.1) $\Rightarrow$ Supra-Classicalité (P.1)

$$\frac{A \leadsto A \text{ (A.1)} \quad \vdash \quad A \rightarrow B}{A \leadsto B} \text{ (R.1)}$$

Coupure (R.3) + Monotonie faible (R.4) + Supra-Classicalité (P.1) $\Rightarrow$ Et (P.4)

$$\frac{\frac{A \leadsto B \quad A \leadsto C}{A.B \leadsto C} \text{ (R.4)} \quad A.B.C \leadsto B.C \text{ (P.1)}}{A.B \leadsto B.C} \text{ (R.3)}$$

$$\frac{A.B \leadsto B.C \quad A \leadsto B}{A \leadsto B.C} \text{ (R.3)}$$

Par conséquent, toute relation d'inférence cumulative consistante est une relation de confirmation absolue adéquate du point de vue des principes P.1 à P.4.

L'objet "confirmation absolue" est donc effectivement un *cas particulier* de l'objet "inférence non monotone". D'un point de vue syntaxique, on peut remarquer (Crocco & Lamarre [1991]) que le principe de Consistance est une règle de type négatif qui n'est pas utilisée dans l'axiomatisation des systèmes d'inférence non monotone. Du point de vue de la philosophie des sciences, ce principe est essentiel car il nous garantit que nos tests ne vont pas confirmer n'importe quoi ! Il faudra donc simplement vérifier que la sémantique que nous choisirons pour interpréter " $\leadsto$ " respecte ce principe.

Il est encore possible d'enrichir le système de règles d'inférence non monotone précédent par la règle:

R.5 OU :

$$\frac{A \leadsto C \quad B \leadsto C}{A \vee B \leadsto C}$$

Dans ce cas (Kraus & al. [1990]) on obtient une *relation d'inférence préférentielle*. Le respect de la règle "Ou" n'est pas couramment revendiqué dans la littérature confirmationniste, toutefois, on peut admettre qu'il s'agit là d'une propriété supplémentaire intuitivement souhaitable pour la relation d'acceptation. Si on décide de rajouter ce réquisit aux principes d'adéquation P.1 à P.4, on conclura qu'une relation d'inférence non monotone est adéquate en tant que relation de confirmation

absolue s'il s'agit d'une *relation préférentielle consistante*. C'est le point de vue que nous retiendrons.

4.2. Sémantique : confirmation absolue et théorie des possibilités

Dans la mesure où toute relation préférentielle consistante est une relation adéquate de confirmation absolue, toute sémantique qui assure que l'opérateur " $\sim$ " est une relation préférentielle consistante (ou au minimum une relation cumulative consistante) est une interprétation adéquate de la notion de confirmation absolue.

C'est notamment le cas de la relation de conséquence préférentielle de Shoham [1988] comme l'ont montré Kraus & al. [1990], et de la relation de conséquence préférentielle *possibiliste* dont Dubois & Prade [1991] ont montré l'équivalence avec celle de Shoham. Rappelons brièvement sa construction :

On considère une algèbre booléenne $\mathcal{B}$ induite par un ensemble de propositions. Soit Ω l'ensemble des atomes (ou interprétations) de $\mathcal{B}$. On note $\omega \models e$ le fait que la proposition e soit vraie dans l'interprétation ω .

On se donne alors une fonction π de Ω dans $[0,1]$ qu'on appelle distribution de possibilités sur Ω et qui pour tout ω décrit son degré de possibilité. On définit alors la possibilité Π d'une proposition de $\mathcal{B}$ par :

$$\forall e \in \mathcal{B}, \Pi(e) = \text{Max} \{ \pi(\omega), \omega \in \Omega \text{ et } \omega \models e \}$$

et dualement sa nécessité par :

$$N(e) = 1 - \Pi(\bar{e}).$$

L'interprétation ω est un π -modèle préférentiel de la proposition e (noté $\omega \models_{\pi} e$) si et seulement si :

- a) C'est un modèle de e : $\omega \models e$
- b) Il n'existe pas de modèle de e dont la possibilité soit supérieure à celle de ω
- c) $\Pi(e) > 0$.

On dit alors que la proposition e implique π -préférentiellement la proposition h lorsque h est vraie dans tous les modèles π -préférentiels de e , ce qu'on peut noter :

$$e \models_{\pi} h.$$

On définit par ailleurs la possibilité conditionnelle par :

$$\Pi(h,e) = 1 \text{ si } \Pi(e) = \Pi(he)$$

$$\Pi(h,e) = \Pi(he) \text{ si } \Pi(e) > \Pi(he)$$

De manière duale la nécessité conditionnelle est définie par :

$$N(h,e) = 0 \text{ si } N(e \rightarrow h) = N(\bar{e})$$

$$N(h,e) = N(e \rightarrow h) \text{ si } N(e \rightarrow h) > N(\bar{e})$$

Dubois & Prade [1991] montrent alors que :

$$e \Vdash_{\pi} h \equiv N(h, e) > 0 \equiv \Pi(\bar{h}, e) < 1$$

Ils démontrent par ailleurs que lorsqu'on assimile l'inférence non monotone " $e \leadsto h$ " à $N(h, e) > 0$, toutes les propriétés des relations d'inférences préférentielles sont satisfaites.

Enfin, il est trivial de vérifier que cette interprétation respecte le principe de Consistance puisque : $\forall e \ N(\perp, e) = 0$.

Conformément à ce qui été vu plus haut $N(h, e) > 0$ est donc un critère adéquat de confirmation absolue.

Pierre d'achoppement pour les théoriciens probabilistes de la confirmation, le paradoxe de la loterie n'affecte pas le critère possibiliste d'inférence, ce qu'on peut vérifier directement dans la mesure où l'axiome fondamental des mesures de possibilités est aussi valide pour les possibilités conditionnelles (cf. Annexe). On peut en effet écrire :

$$\Pi(\bar{h}, e) < 1 \text{ (soit } e \leadsto h) \text{ et } \Pi(\bar{h}', e) < 1 \text{ (soit } e \leadsto h') \text{ implique :}$$

$$\Pi(\overline{hh'}, e) = \Pi(\bar{h} \vee \bar{h}', e) = \text{Max} [\Pi(\bar{h}, e), \Pi(\bar{h}', e)] < 1. \text{ Donc } e \leadsto h.h'.$$

Par conséquent, et contrairement à la théorie des probabilités : la *théorie des possibilités nous offre une sémantique adéquate pour la relation d'acceptation ou de confirmation absolue d'un énoncé en présence d'un autre*. Cette sémantique peut se lire de plusieurs manières :

a) Une hypothèse h est acceptable en présence d'un énoncé e si elle est vraie dans le(s) monde(s) le(s) plus possible(s) dans le(s)quel(s) e se réalise. Dans la mesure où il a été montré que la sémantique préférentielle présente de fortes similarités avec la sémantique des conditionnels contrefactuels, cette interprétation rapproche également la notion de confirmation de la notion de contrefactuel, ce qui est intuitivement satisfaisant. Les phrases "si cet animal avait des ailes, il pourrait voler" et "le fait que cet animal a des ailes *confirme* qu'il peut voler" ont selon toute apparence les mêmes conditions de vérité et les mêmes *défauts* (les mondes dans lesquels cet animal appartient à la classe des ratites, qui ne sont pas considérés comme les plus vraisemblables lorsqu'on sait qu'un animal a des ailes). Cette connexion entre confirmation et conditionnel contrefactuel a été par ailleurs anticipée par Goodman [1954] selon lequel les conditions de vérité d'un contrefactuel renvoient à l'acceptabilité d'un énoncé nomologique en vertu duquel le conséquent se déduit de l'antécédent et d'un ensemble de conditions pertinentes.

b) Numériquement, on accepte une hypothèse h en présence d'un énoncé e lorsque sa nécessité est positive, ou de manière équivalente lorsqu'elle est *plus possible que l'hypothèse contraire* $\bar{h}$. Dubois & Prade [1991] font remarquer que cette interprétation paraît plus satisfaisante que la sémantique de Adams [1975]: $\text{Prob}(h, e) > 1 - \varepsilon$ (ε infiniment petit), dont on démontre qu'elle satisfait également les règles d'inférence préférentielles (Pearl [1988]). En effet, la sémantique de Adams semble trop restrictive et peu utilisable puisque les exceptions possibles tolérées pour accepter une hypothèse devrait avoir une probabilité infiniment faible. Mais a contrario, on peut penser que la sémantique possibiliste est un peu laxiste, au moins dans une perspective d'acceptation scientifique : le fait qu'une théorie soit seulement

plus possible que sa négation semble être un argument de poids insuffisant pour admettre ladite théorie ! Cette question demeure ouverte pour les épistémologues, mais en tout état de cause il ne s'agit pas d'un argument décisif tant qu'on n'a pas exhibé le principe d'adéquation syntaxique que ce laxisme conduit à violer.

Tant qu'on en reste aux relations entre propositions la sémantique possibiliste semble donc assez correctement capturer les principales intuitions liées à la confirmation absolue. Nous allons voir que la question est un peu plus délicate en ce qui concerne l'idée de confirmation relative.

5. Théorie des possibilités et confirmation relative

La relation de confirmation relative (ou de "soutien") d'un énoncé par un autre a un statut très différent de la relation de confirmation absolue : elle ne peut pas être interprétée comme une relation *d'inférence*, puisqu'elle ne respecte ni l'Affaiblissement à droite, ni la Conjonction. Il n'est donc pas question de trouver un isomorphisme entre la confirmation relative et un quelconque système de logique non monotone. Ce qu'il serait par contre concevable de rechercher lorsqu'on dispose d'un système de logique non monotone adéquat, c'est la manière dont les arguments plaident *en faveur* d'une conclusion avant d'être suffisamment décisifs pour qu'on accepte cette conclusion, ou en renforçant des arguments déjà décisifs. C'est l'idée qui anime le fameux critère de Nicod :

le fait $Pa.Qa$ est un argument en faveur de l'hypothèse universelle $\forall x (Px \rightarrow Qx)$, bien qu'il ne soit pas un argument décisif.

Lorsqu'on dispose d'une définition quantitative adéquate de la confirmation absolue, liée à un couple (mesure d'un degré de confirmation, seuil à atteindre), il est très facile de déterminer la signification de la notion d'argument favorable que dans le cas purement qualitatif. Encore faut-il vérifier que la définition qu'on obtient est conforme à ce qu'on en attend, c'est-à-dire aux principes d'adéquation P.1, P.5, P.6 et P.10 (avec les réserves déjà mentionnées).

C'est ce que nous allons faire en partant du critère possibiliste de confirmation absolue. Pour cela, il est nécessaire d'explicitier la mesure du *degré de confirmation* qui est attachée à ce critère : nous dirons qu'il y a soutien, ou confirmation relative d'une proposition par une autre lorsque son degré de confirmation possibiliste augmente. Mais la notion de degré de confirmation est elle-même ambiguë. Expliquons-nous, en reprenant l'analogie thermique utilisée déjà par Carnap [1962] :

- Le degré de *température* est une mesure neutre par rapport aux idées de "froid" ou de "chaud". On peut définir l'idée de "chaud" à partir d'un seuil de température à dépasser (par exemple 50° centigrades) et l'idée de "chauffer" à partir de l'idée d'accroissement de la température. Mais on pourrait aussi se doter d'une mesure du degré de *chaleur*, qui ne soit plus neutre : un corps serait alors "chaud" si ce degré est positif, "froid" s'il est nul (les valeurs négatives n'étant pas admises) ; dans ce cas, "chauffer" signifierait encore accroître le degré de température mais ne serait plus défini pour les corps froids qui, bien que chauffés, n'atteindraient pas le seuil nécessaire pour être qualifiés de chauds.

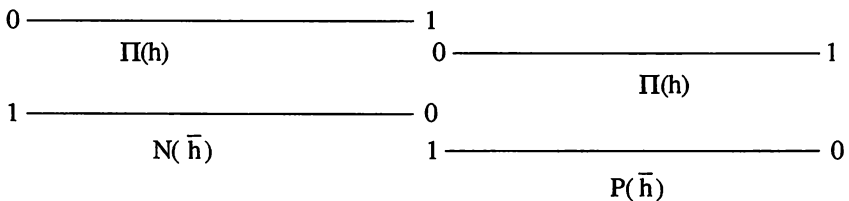
- De la même manière, le degré de *confirmation* peut être défini de manière neutre par rapport à l'idée d'acceptation, comme c'est le cas dans l'interprétation directe du degré de probabilité comme degré de confirmation (Baillie [1969]). Une proposition non confirmée absolument possède dans ce cas quand même un certain "degré de confirmation"; c'est quand ce degré est fort que la proposition est acceptée. Mais il peut aussi être défini de manière consubstantielle à l'idée d'acceptation, s'il n'est positif que lorsque la proposition est acceptée, et s'il est toujours nul autrement.

Si on retient $N(h,e)$ comme degré de confirmation, on obtient seulement un *degré d'acceptation*. En effet :

- a) Une proposition h est confirmée par une proposition e dès que $N(h,e) > 0$.
- b) Il est impossible que $N(h,e) < 0$.

Il n'y a pas de place pour un degré de confirmation "neutre" qui mesurerait (graduellement) le soutien qu'une proposition non encore acceptée peut recevoir. Pour obtenir une mesure de degré de confirmation *complète* en théorie des possibilités, il faut alors tenir compte de la double relation induite par les mesures de possibilités et de nécessités, qu'on peut visualiser de la manière suivante en remarquant que $N(h) > 0$ si

et seulement si $\Pi(h) = 1$:



En remarquant que $N(h) + \Pi(h) + N(\bar{h}) + \Pi(\bar{h}) = 2$, on peut voir que la fonction $C(h) = 1/2 [N(h) + \Pi(h)]$ satisfait une relation identique à la relation probabiliste classique : $C(\bar{h}) = 1 - C(h)$. Ceci incite à définir $C(h,e)$ comme degré de confirmation dans ce cadre.

On peut tout d'abord remarquer que ceci est cohérent avec notre définition de la confirmation absolue qui sera maintenant redéfinie par :

$e \rightarrow h$ (absolument) si et seulement si $C(e) > 1/2$

Il est en effet facile de voir que $C(h,e) > 1/2 \Leftrightarrow N(h,e) > 0$

La confirmation relative pouvant alors être définie par :

$e \rightarrow h$ (relativement) si et seulement si $C(h,e) > C(h)$

On peut alors dire qu'une proposition e *soutient* une proposition h si $C(h,e) > C(h)$ ce qui se produit quand $\Pi(h,e) > \Pi(h)$ ou $N(h,e) > N(h)$.

Réciproquement, on dira qu'une proposition e *contre-soutient* une proposition h si $C(h,e) < C(h)$ ce qui se produit quand $\Pi(h,e) < \Pi(h)$ ou $N(h,e) < N(h)$. On vérifie aisément que ces définitions sont cohérentes et qu'il est impossible qu'une proposition soutienne et contre-soutienne simultanément une autre proposition.

On remarque alors que le contre-soutien que e apporte à h est égal au soutien qu'il apporte à $\bar{h}$ de la même manière que dans le cadre probabiliste.

Pour étudier complètement les propriétés de la confirmation relative possibiliste, il faut alors distinguer quatre cas⁵ :

a) e soutient h car $N(h,e) > N(h)$ [ie. $\Pi(\bar{h},e) < \Pi(\bar{h})$]

$$\text{Ceci est équivalent à } \begin{cases} N(h,e) > 0 \\ N(\bar{e},\bar{h}) > 0 \end{cases}$$

En effet, dans le sens direct :

$N(h,e) > 0$ est évident.

$$[\Pi(\bar{h},e) < \Pi(\bar{h})] \Rightarrow \Pi(\bar{h},e) = \Pi(\bar{h}.e)$$

Donc on a :

$$[\Pi(\bar{h}.e) < \Pi(\bar{h})]$$

D'où :

$$\Pi(e,\bar{h}) < 1 \text{ et donc } N(\bar{e},\bar{h}) > 0$$

Réciproquement :

$$N(h,e) > 0 \Rightarrow P(\bar{h},e) < 1 \Leftrightarrow \begin{cases} \Pi(\bar{h},e) = \Pi(\bar{h}.e) \\ \Pi(\bar{h}.e) < \Pi(\bar{h}) \end{cases}$$

$$N(\bar{h},\bar{e}) > 0 \Rightarrow P(e,\bar{h}) < 1 \Leftrightarrow \begin{cases} \Pi(e,\bar{h}) = \Pi(\bar{h}.e) \\ \Pi(\bar{h}.e) < \Pi(\bar{h}) \end{cases}$$

D'où $\Pi(\bar{h}.e) < \Pi(\bar{h})$.

5. Nous ferons l'hypothèse que $\Pi(h) > 0$ et $N(h) < 1$ dans tous les cas : il n'est pas possible de soutenir ou de contre-soutenir une tautologie ou une contradiction.

En termes sémantiques ceci signifie :

h est vrai dans tous les π -modèles préférentiels de e (h est accepté en présence de e)

$\bar{e}$ est vrai dans tous les π -modèles préférentiels de $\bar{h}$ ($\bar{e}$ est accepté en présence de $\bar{h}$).

On peut remarquer qu'il n'y a aucune raison pour que ces deux conditions soient obligatoirement remplies lorsque e est une conséquence de h.

b) e soutient h car $\Pi(h, e) > \Pi(h)$ [ie. $N(\bar{h}, e) < N(\bar{h})$] :

Ceci est clairement équivalent à $\Pi(h) < 1$ et $\Pi(h, e) = 1$. L'implication directe est due à $\Pi(h, e) < 1 \Rightarrow \Pi(h, e) = \Pi(h, e) \leq \Pi(h)$. La réciproque est évidente. Ceci est équivalent à $N(\bar{h}) > 0$ et $N(\bar{h}, e) = 0$.

En termes sémantiques ceci signifie :

$\bar{h}$ est vrai dans tous les modèles les plus plausibles de Ω

$\bar{h}$ est faux dans certains des π -modèles préférentiels de e

Si $h \vdash e$, on a : $\Pi(h, e) = \Pi(h)$. Par conséquent, si $\Pi(e) = \Pi(h) < 1$, on a bien $\Pi(h, e) > \Pi(h)$ puisque dans ce cas $\Pi(h, e) = \Pi(e)$ et donc $\Pi(h, e) = 1$.

Tout se passe comme si, dans ce cas, le principe de conversion s'appliquait seulement aux tests les plus "sévères", représentés par les conséquences les moins possibles de h (celles dont la possibilité est égale à celle de h et inférieure à 1).

c) e contre-soutient h car $\Pi(h, e) < \Pi(h)$ [ie. $N(\bar{h}, e) > N(\bar{h})$] :

Ceci est équivalent à $N(h) > 0$ et $N(h, e) = 0$.

En effet $\Pi(\bar{h}, e) > \Pi(\bar{h}) \Rightarrow \Pi(\bar{h}, e) = 1$ (point b.) $\Rightarrow N(h, e) = 0$.

En termes sémantiques, ceci signifie :

h est vrai dans tous les modèles les plus possibles de Ω

h est faux dans certains des π -modèles préférentiels de e

Si $h \dashv e$, cela est impossible. En effet dans ce cas :

$N(h) > 0 \Rightarrow \Pi(h) = 1 \Rightarrow \Pi(h, e) = 1 \Rightarrow \Pi(e) = 1$. Par ailleurs :

$\Pi(\bar{h}, e) = 1 \Rightarrow \Pi(\bar{h}, e) = \Pi(e) = 1 \Rightarrow \Pi(\bar{h}) = 1$, ce qui est contradictoire avec $N(h) > 0$. Donc, si e est une conséquence de h, il ne peut pas diminuer le degré d'acceptation de h en diminuant sa nécessité.

d) e contre-soutient h car $\Pi(h, e) < \Pi(h)$ [ie. $N(\bar{h}, e) > N(\bar{h})$] :

Ceci est équivalent à $N(\bar{h}, e) > 0$ et $N(\bar{e}, h) > 0$ (démonstration identique au point a)

En termes sémantiques ceci signifie :

h n'est vrai dans aucun π -modèle préférentiel de e ($\bar{h}$ est accepté en présence de e).

$\bar{e}$ est vrai dans tous les π -modèles préférentiels de h ($\bar{e}$ est accepté en présence de h).

Si $h \vdash e$, cela est impossible puisque dans ce cas $\Pi(h.e) = \Pi(h)$. Donc, si e est une conséquence de h , il ne peut pas diminuer le degré d'acceptation de h en diminuant sa possibilité.

Par conséquent, une conséquence de h ne peut jamais contre-soutenir h , ni par le mécanisme c) ni par le mécanisme d), ce qui est intuitivement heureux.

En résumé e soutient h si :

*h est vrai dans tous les π -modèles préférentiels de e (h est accepté en présence de e)
et*

$\bar{e}$ est vrai dans tous les π -modèles préférentiels de $\bar{h}$ ($\bar{e}$ est accepté en présence de $\bar{h}$).

ou si :

*$\bar{h}$ est vrai dans tous les modèles les plus possibles de Ω
et*

*$\bar{h}$ est faux dans certains des π -modèles préférentiels de e
 e contre-soutient h si :*

*h est vrai dans tous les modèles les plus possibles de Ω
et*

h est faux dans certains des π -modèles préférentiels de e

ou si :

h est faux dans tous les π -modèles préférentiels de e ($\bar{h}$ est accepté en présence de e).

et

e est faux dans tous les π -modèle préférentiel de h ($\bar{e}$ est accepté en présence de h).

La construction d'un critère de confirmation relative à partir du critère d'acceptation possibiliste conduit donc au résultat suivant :

Le principe de conversion, souvent considéré comme un critère d'adéquation de la notion de soutien, n'est pas nécessairement respecté; rien ne garantit que toute conséquence de h accroisse son degré d'acceptabilité. On pourrait s'attendre à ce que les conséquences de h soient au minimum dans une situation plus favorable pour confirmer h que les autres propositions (c'est-à-dire que les conditions supplémentaires à remplir soient plus faibles), mais cela n'est même pas toujours le cas, ce qui est un résultat plutôt contre-intuitif, qui ne plaide pas en faveur de l'adéquation de la notion possibiliste de confirmation relative.

Cependant, la confirmation relative n'est pas totalement insensible à la relation d'antécédence entre e et h :

- Un test "sévère" de h par une de ses conséquences les moins possibles a priori (donc les plus riches en information), peut accroître la possibilité de h si celle-ci n'est pas déjà égale à 1.

- Une conséquence d'une hypothèse h ne peut jamais la contre-soutenir.

On obtient donc ce qu'on pourrait nommer un principe de conversion "faible", qui n'est pas dépourvu de pertinence épistémologique. En particulier, il évoque la thèse popperienne selon laquelle seuls les tests sévères sont à prendre en compte pour la "corroboration" d'une hypothèse. Les autres tests favorables des prédictions d'une hypothèse n'ont qu'une valeur épistémique négative : ils n'en diminuent pas le degré de corroboration.

Remarque : Il est facile de voir que la relation de soutien définie par les cas a) et b) ci-dessus respecte les principes d'adéquation P. 1, P.5 et P.6, requis au §3.2 (en plus du principe de conversion P.10) pour qu'une relation de confirmation relative soit adéquate.

On obtient donc une théorie de la confirmation fondée sur une métrique définie par :

$$C(h,e) = 1/2 [\Pi(h,e) + N(h,e)]$$

La confirmation relative étant définie par :

$$e \rightarrow h \text{ (relativement) ssi } C(h,e) > C(h)$$

La confirmation absolue étant définie par :

$$e \rightarrow h \text{ (absolument) ssi } C(h,e) > 1/2$$

Cette situation rappelle la situation probabiliste où la confirmation relative était définie par $p(h,e) > p(h)$ et la confirmation absolue par $p(h,e) > s > 1/2$. Cependant l'utilisation des nécessités/possibilités permet de disposer à la fois d'un critère absolu respectant le principe de conjonction et d'un critère relatif respectant un principe de conversion affaibli.

6. Conclusion

La théorie des possibilités de Dubois & Prade permet de construire un critère d'acceptation, ou de confirmation absolue, totalement adéquat par rapport aux principes syntaxiques traditionnellement exigés dans le cadre d'un langage propositionnel. Ces règles sont par ailleurs formellement les mêmes que celles qui caractérisent toute une classe de logiques non monotones, les logiques "préférentielles". Par conséquent, la relation de confirmation absolue hérite des interprétations non numériques qui sont compatibles avec ces règles, notamment la sémantique préférentielle de Shoham.

Un certain nombre de questions demeurent cependant ouvertes :

La notion de soutien, ou de confirmation relative, qu'on peut raisonnablement associer au critère possibiliste d'acceptation ne respecte pas strictement le Principe de Conversion, qui nous garantit que notre confiance dans nos conjectures croît avec leurs prédictions vérifiées. Il en respecte une version restrictive, dont les épistémologues devront dire si elle est plus ou moins pertinente que la version complète : après tout, il serait peut-être préférable de tenir compte de la sévérité différentielle des tests dans tous les cas. Or, si on définit un test sévère comme le test d'une conséquence dont la possibilité est la plus faible (c'est-à-dire est égale à la possibilité de l'hypothèse), la réussite d'un tel test n'entraîne obligatoirement un soutien de l'hypothèse que si la possibilité de celle-ci est inférieure à 1. Une hypothèse dont la possibilité est égale à 1 peut être, mais n'est pas forcément, soutenue par la réussite d'un test sévère. Ceci n'est pas très satisfaisant.

De plus, il serait nécessaire de préciser comment les critères de confirmation possibilistes permettent de formaliser de manière adéquate les relations de confirmation entre une hypothèse universelle de la forme $\forall x (Px \rightarrow Qx)$, et des énoncés singuliers qui résultent de tests empiriques de cette hypothèse. Dans ce contexte, il faudrait inclure dans la liste de nos principes d'adéquation des principes du type "Instantial Comparability Condition" ou du type du principe de Nicod, que nous avons délibérément exclus au début de la présente analyse.

Annexe

On peut montrer que les possibilités conditionnelles héritent de l'axiome des possibilités sur la disjonction :

En effet, ou bien $\Pi(h \vee h', e) = 1$ si $\Pi(e) = \Pi((h \vee h').e)$, ou bien $\Pi(h \vee h', e) = \Pi((h \vee h').e)$ si $\Pi(e) > \Pi((h \vee h').e)$. Étudions les deux cas :

$$\begin{aligned} \text{a) } \Pi(h \vee h', e) &= 1 && \Leftrightarrow \Pi(e) = \Pi((h \vee h').e) = \text{Max} [\Pi(h.e), \Pi(h'.e)] \\ &&& \Leftrightarrow \Pi(e) = \Pi(h.e) \text{ ou } \Pi(e) = \Pi(h'.e) \\ &&& \Leftrightarrow \Pi(h.e) = 1 \text{ ou } \Pi(h'.e) = 1 \\ &&& \Leftrightarrow \Pi(h \vee h', e) = \text{Max} [\Pi(h.e), \Pi(h'.e)] \end{aligned}$$

b) $\Pi(h \vee h', e) = \Pi((h \vee h').e) \Leftrightarrow \Pi(h \vee h', e) = \text{Max} [\Pi(h.e), \Pi(h'.e)]$. Dans ce cas on a $\Pi(e) > \Pi((h \vee h').e) = \text{Max} [\Pi(h.e), \Pi(h'.e)]$.

Donc $\Pi(h.e) = \Pi(h.e)$ et $\Pi(h'.e) = \Pi(h'.e)$.

D'où $\Pi(h \vee h', e) = \text{Max} [\Pi(h.e), \Pi(h'.e)]$

Ceci peut aussi se voir si on réinterprète les possibilités conditionnelles en termes de nouvelles distributions de possibilités sur les atomes. Considérons la possibilité conditionnelle $\Pi(., e)$. Soit alors $\Omega_e = \{\omega \in \Omega / \omega \Vdash e\}$. On redéfinit la distribution de possibilités de la manière suivante :

$$\forall \omega \in \Omega_e, \pi_e(\omega) = 1 \text{ ssi } \omega \Vdash e, \pi_e(\omega) = \pi(\omega) \text{ sinon.}$$

$$\forall \omega \in \Omega - \Omega_e, \pi_e(\omega) = 0.$$

On peut alors définir la possibilité conditionnelle $\Pi(h,e)$ par : $\Pi(h,e) = \Pi_e(h) = \text{Max} \{ \pi_e(\omega) / \omega \Vdash h \}$.

En effet : $\Pi_e(h) = 1 \Leftrightarrow \exists \omega / \omega \Vdash_\pi h \text{ et } \pi_e(\omega) = 1 \Leftrightarrow \exists \omega / \omega \Vdash h \text{ et } \omega \Vdash_\pi e$
 $\Leftrightarrow \Pi(he) = \Pi(e) \Leftrightarrow \Pi(h,e) = 1$

$\Pi_e(h) = \Pi(he) < 1 \Leftrightarrow \omega / \omega \Vdash h \text{ et } \omega \Vdash_\pi e \Leftrightarrow \Pi(he) < \Pi(e)$

$\Leftrightarrow \Pi(h,e) = \Pi(he) < 1$

La possibilité conditionnelle $\Pi(h,e)$ est alors interprétée comme une possibilité non conditionnelle, ce qui garantit l'application de l'axiome.

Bibliographie

- ADAMS E.W., (1975) *"The Logic of Conditionals"* Reidel, Dordrecht, The Netherlands.
- BAILLIE P., (1969) *"That Confirmation may yet be a Probability"*, B.J.P.S., Vol. 20, p. 41.
- BARWISE J. and COOPER R., (1981) *"Generalized Quantifiers and Natural Language"*, Linguistics and Philosophy, IV, p. 159-219.
- BENTHEM J. van, (1984) *"Foundations of Conditional Logic"*, Journal of Philosophical Logic, XIII, p. 303-349.
- CARNAP R., (1950 et 1962) *"Logical Foundations of Probability"*, Chicago, (2d ed 1962)
- COHEN L.J., (1966 a) *"What has Confirmation to do with Probabilities"*, Mind, Vol. 75, N° 300, p. 463.
- COHEN L.J., (1966 b) *"A Logic for Evidential Support"*, B.J.P.S., 17, p. 23 ; *"A Logic for Evidential Support II"*, B.J.P.S., 17, N° 2 p. 105-126.
- CROCCO G., LAMARRE P., (1991) *"Conditional Logic a Powerful Tool for Non Monotonicity"*, Tech. Rept IRIT/92-2-T.
- DELGRANDE J.P., (1988) *"An Approach to Default Reasoning based on First Order Conditional Logic"*, Artificial Intelligence 36, p. 63-90.
- DUBOIS D., PRADÉ H., (1991) *"Possibilistic Logic, Preferential Models, Non-Monotonicity and Related Issues"*, Proc. of the 12th Inter. Joint Conf. on Artificial Intelligence (IJCAI-91), Sydney, Aug 24-30, 1991.
- DUBUCS J., (1990) *"Carnapes ab omni naevo vindicatus"*, Les Cahiers du CREA N° 14, p. 97-120.
- GABBAY D.M., (1985) *"Theoretical Foundations for Non-monotonic Reasoning in Expert Systems"*, Proc. NATO Advanced Study Institute on Logics and Models of Concurrent Systems, pages 439-457, La Colle-sur-Loup, France (K.R. Apt, Ed), Springer Verlag, Berlin.
- GOOD I.J., (1950) *"Probability" and the Weighting of Evidence"*, New York and London.
- GOODMAN N., (1955) *"Fact, Fiction and Forecast"*, Harvard University Press.
- HEMPEL C.G., (1943) *"A Purely Syntactical Definition of Confirmation"*, The Journal of symbolic Logic, Vol. 8, N° 4, p. 121-143.

- KRAUS S., LEHMANN D., MAGIDOR M., (1990) *"Non-Monotonic Reasoning, Preferential Models and Cumulative Logics"*, Artificial Intelligence, 44(1-2) : 137-207.
- LAMARRE P., (1990) *"Modal Logics as Conditional Logics of Nonmonotonicity"*, Tech.Rept. IRIT/90-50-R, In-stitut de Recherche en Informatique de Toulouse (France).
- LEVI I., (1967) *"Gambling with truth"*, réédité par MIT Press, Cambridge Mass (1 973).
- LEVI I., (1986) *"Probabilistic Pettifoggery"*, Erkenntnis XXV, p. 1331 40.
- LEWIS D., (1973) *"Counterfactuals"*, Harvard University Press, Cambridge Mass.
- MAKINSON D., (1989) *"General Theory of Cumulative Inference"*, In Non-Monotonic Reasoning (2nd Inter. Workshop, Grassau, FRG, June 1988) (M. Reinfrank, J. De Kleer, M.L. Ginsberg, E. Sandewall, eds), Lecture Notes in Computer Science, Vol. 346, Springer Verlag, Berlin.
- PEARL J., (1988) *"ProbabilisVc Reasoning in Intelligent Systems : Networks of Plausible Inference "*, Morgan Kaufmann, Los Alamos, CA.
- POPPER K.R., (1934) *"Logik der Forschung"* Julius Springer, Vienne ; Traduction Anglaise (1959) *"The Logic of Scientific Discover"*, London.
- POPPER K.R. and MILLER D., (1983) *"A Proof of the Impossibility of Inductive Probability"*, Nature, 302, April 21, p. 687-688.
- SHACKLE G.L.F., (1961) *"Decision, Order and Time in Human Affairs"*, Cambridge University Press.
- SHAFFER G., (1976) *"A Mathematical Theory of Evidence"*, Princeton University Press.
- SHOHAM Y., (1988) *"Reasoning about Change - Time and Causation from the Standpoint of Artificial Intelligence"*, The MIT press, Cambridge, Mass.
- ZWIRN D. and ZWIRN H.P., (à paraître) *"Metaconfirmaton"*.

Denis ZWIRN, Hervé ZWIRN

Chapitre 3

Logique possibiliste, modèles préférentiels et objets conditionnels

Ce chapitre propose une synthèse de divers travaux portant sur le raisonnement plausible, qui tous cherchent à formaliser la capacité des individus à tirer des conclusions provisoires dans des situations où l'information disponible sur l'état du monde est incomplète. Ces travaux sont apparemment très disparates, mais il s'avère possible de les unifier. Ils concernent, d'une part la logique possibiliste qui est une extension de la logique classique où on range les formules par niveau de certitude, ce qui permet de raisonner de façon non triviale en présence de contradictions, et d'autre part la notion de relation de conséquence non monotone étudiée par D. Gabbay, D. Makinson, et D. Lehmann, notamment, qui affaiblit la notion classique de conséquence pour capturer l'inférence plausible. Enfin la notion d'objet conditionnel formalise l'idée de conditionnement indépendamment de la notion de probabilité, et permet de coder naturellement des connaissances génériques. On montre que l'inférence en logique possibiliste satisfait les propriétés des relations de conséquence non monotone, et que les objets conditionnels fournissent une sémantique naturelle pour ces dernières. On décrit un ordre naturel sur les objets conditionnels interprétés comme des règles par défaut ; on montre enfin qu'un ensemble ordonné d'objets conditionnels peut se coder naturellement en logique possibiliste dont la machinerie inférentielle fournit alors les conclusions plausibles attendues.

1. Introduction

On peut admettre que la logique classique est issue de tentatives anciennes pour formaliser le raisonnement humain, tentatives de philosophes principalement. Elle a été plus récemment formalisée par des mathématiciens dans la première moitié du 20^e siècle dans le but de servir de fondation au raisonnement mathématique. De nombreuses recherches en Intelligence Artificielle ont récemment mis en évidence ses insuffisances en tant qu'outil de représentation et d'utilisation des connaissances de sens commun. Quand il raisonne à partir d'informations incomplètes, incertaines, ou contradictoires, l'homme ne suit pas vraiment les lois de la logique classique, laquelle, en présence de telles informations, reste très prudente, et ne permet que rarement l'obtention de conclusions utiles.

On s'intéresse ici à un type de raisonnement de sens commun, que l'on peut appeler raisonnement plausible. Il tente de rendre compte des capacités de l'esprit humain à compléter les informations dont il dispose dans une situation donnée, grâce à l'emploi de connaissances "génériques", c'est-à-dire auxquelles on peut se fier au moins "en général". Les déductions obtenues ne seront souvent que plausibles au sens où l'arrivée de nouvelles informations peut les remettre en cause. Cette propriété des raisonnements plausibles est dite "non-monotonie". Elle est incompatible avec l'emploi de la logique classique dans son acception pure et dure. Aussi conteste-t-on parfois le nom de "logique" pour les formalismes non monotones. Ce problème du raisonnement plausible a en effet suscité un grand nombre de travaux très divers en apparence, parfois contestables et souvent contestés, depuis 1980. Une question importante qui émerge de ces travaux est : comment représenter une connaissance générique susceptible de présenter des exceptions, encore appelée "règle par défaut" ? Cette question n'est pas nouvelle au sens où elle était déjà au cœur de la problématique des systèmes experts dans les années 70. Les systèmes experts ont été beaucoup critiqués pour leur manque de fondement théorique, et parce que l'espérance qu'ils ont suscitée a été partiellement déçue. Malgré leurs imperfections et leurs limitations, ces systèmes ont établi, de par leur structure, une distinction importante entre connaissances factuelles et connaissances génériques (ce qu'on a appelé "base de faits" et "base de règles", respectivement). Malheureusement, les règles d'un système expert sont souvent entachées d'exceptions, et l'algorithme d'utilisation des connaissances, encore appelé "moteur d'inférences" (cf. Farrény, 1985), s'est avéré peu adapté à l'exploitation de telles règles par défaut.

Comment formaliser une connaissance générique conditionnelle incertaine, de la forme "si X est A alors "généralement" Y est B" ? Il n'est pas question de passer en revue ici les nombreux travaux sur cette question (voir, par exemple Léa Sombé, 1989, 1990). On se contentera de quelques remarques qui permettent de mieux situer les questions abordées dans cet article. Si on aborde la représentation des connaissances conditionnelles par la logique, on trouve naturel d'utiliser le connecteur d'implication, c'est-à-dire de représenter "si p alors q" par la formule logique $\neg p \vee q$ où p et q sont des variables propositionnelles, $\neg$ désigne la négation et $\vee$ la disjonction. Pour mettre en évidence le caractère générique d'une telle connaissance, on peut utiliser la logique du premier ordre avec $p = A(x)$, $q = B(x)$ et obtenir la formule clausale $\forall x \neg A(x) \vee B(x)$. C'est un peu ce qu'on fait en programmation logique où tout est formulé à l'aide de clauses (de Horn, notamment,

qui n'ont qu'un littéral positif). Si on veut prendre en compte l'incertitude, on peut vouloir l'aborder par la théorie des probabilités et la connaissance conditionnelle si p alors q se verra tout aussi naturellement attacher une probabilité conditionnelle $P(q \mid p)$. Une difficulté apparaît tout de suite : généralement on n'a pas l'égalité $P(q \mid p) = P(\neg p \vee q)$ (sauf si l'un des deux vaut 1). Néanmoins la probabilité conditionnelle semble très utile pour modéliser et quantifier la présence d'exceptions, tandis que la formulation clausale avec la présence du quantificateur universel semble exclure totalement l'idée d'exception.

Il y a une autre vision de l'incertitude, à savoir l'idée que certaines connaissances sont plus fiables que d'autres, au sens où, en présence de contradictions, on tiendra à garder certaines formules plutôt que d'autres, jugées moins "enracinées". Inversement, en l'absence d'informations suffisantes, on essaiera d'exploiter le maximum de connaissances cohérentes, même si elles sont peu sûres, pour pouvoir tirer des conclusions aussi précises que possible même si ces conclusions sont éminemment révisables et entachées d'incertitude. Cette idée conduit à la notion de bases de connaissances stratifiées en couches en fonction du degré de certitude des formules, une idée introduite par Rescher (1976).

Dans ce chapitre, nous tenterons de faire une synthèse entre ces deux approches apparemment très disjointes de l'incertain dans les bases de connaissances. Partant de la notion de base de connaissances stratifiée, nous montrons que de telles bases ont une interprétation naturelle dans le cadre de la théorie des possibilités. La sémantique d'une base de connaissance stratifiée peut alors s'exprimer par un ordre total sur les interprétations. L'inférence correspondante peut alors s'interpréter dans le cadre des logiques préférentielles de Shoham (1988). La notion de conséquence sous-jacente tolère des incohérences partielles, elle est non monotone et satisfait une propriété dite de "monotonie rationnelle".

Puis on aborde la représentation de règles génériques à l'aide d'objets conditionnels. On décrit la logique des objets conditionnels et on indique qu'elle offre une sémantique très simple au système de relations de conséquence non monotone introduit par Kraus, Lehmann et Magidor (1990), et, antérieurement par Adams (1975) dans un cadre probabiliste. Cette logique est plus prudente que la logique possibiliste. Elle est compatible avec la probabilité conditionnelle seulement si on utilise des probabilités conditionnelles infiniment proches de 1. On indique que la notion de possibilité conditionnelle permet de relier la logique possibiliste et les objets conditionnels par l'intermédiaire d'un ordre naturel sur les règles déjà proposé par Pearl dans son "System Z" (Pearl, 1990). On peut ainsi coder un ensemble de connaissances génériques en logique possibiliste à l'aide de clauses valuées où la pondération reflète le niveau de priorité de la règle générique correspondante.

2. Logique possibiliste et bases de connaissances stratifiées

Soit K un ensemble de formules de la logique propositionnelle, censées décrire des connaissances, des croyances, et représentant un état cognitif. En tant que telle cette représentation n'est pas complètement satisfaisante pour au moins une raison : nos croyances ne sont pas en général uniformes et nous tenons plus à certaines qu'à d'autres. Une façon simple de traduire cet état de fait consiste à équiper K d'une

relation d'ordre total que l'on peut traduire par une partition de K en n niveaux $K_1, K_2, \dots, K_n$ de certitude décroissante. K est alors vu comme un ensemble de croyances acceptées, mais on tient plus aux assertions contenues dans K_1 que dans celles contenues dans K_2 , etc. Une telle idée a été proposée dès 1976 par Rescher, qui attache un poids positif à chaque formule de K pour exprimer le degré de confiance correspondant. Cette idée a été reprise récemment par Brewka (1991). Notons α_i les poids attachés aux formules de K_i avec $\alpha_1 > \alpha_2 > \dots > \alpha_n > 0$. Par convention, $\alpha_1 = 1$ est réservé aux formules totalement certaines (par exemple les tautologies).

Comment raisonner avec un ensemble de formules propositionnelles pondérées $\tilde{K}$? il faut pour cela étendre la notion syntaxique de déduction, $K \vdash p$, où $\vdash$ symbolise cette déduction. Une façon simple d'étendre la définition classique est de chercher à prouver p à l'aide des formules les plus certaines ; en posant $K_0 = \emptyset$, on peut proposer

Définition 1. $\tilde{K} \vdash (p, \alpha_i)$ si et seulement si $K_1 \cup K_2 \cup \dots \cup K_i \vdash p$, mais on n'a pas $K_1 \cup K_2 \cup \dots \cup K_{i-1} \vdash p$.

Notons que si K est inconsistant, et $\perp$ dénote la contradiction, on aura $\tilde{K} \vdash (\perp, \alpha_i)$ pour une certaine valeur de i . On appellera niveau d'inconsistance de $\tilde{K}$ le nombre $\text{inc}(\tilde{K}) = \max\{\alpha_i, \tilde{K} \vdash (\perp, \alpha_i)\}$ avec la convention $\text{inc}(\tilde{K}) = 0$ si cet ensemble de poids est vide. On voit tout de suite que la stratification d'une base de connaissances permet de graduer la notion d'inconsistance.

En pratique, si on veut automatiser ce type de déduction, que Rescher (1976) nomme raisonnement plausible, il faut chercher à généraliser des règles d'inférence tels que le principe de résolution. Dans le cas propositionnel où nous nous plaçons, ce principe généralisé peut s'écrire

$$\begin{array}{l} (p \vee q, \alpha_i) \\ (\neg p \vee r, \alpha_j) \\ \hline (q \vee r, \min(\alpha_i, \alpha_j)). \end{array} \quad (R)$$

Cette règle, dans sa spécialisation au modus ponens, remonte à Théophraste, un élève d'Aristote, si l'on en croit Rescher. Elle exprime que la certitude d'une conclusion correspond à celle de la prémisse la moins certaine. Autrement dit, dans un raisonnement c'est le maillon le plus faible de la chaîne d'inférences qui détermine la force de la conclusion. Il est facile de montrer que la déduction à partir d'une base de connaissances stratifiée peut se calculer à partir du principe de résolution étendu et de la réfutation.

Proposition. $\tilde{K} \vdash (p, \alpha_i)$ si et seulement si on produit $(\perp, \alpha_i)$ en appliquant un nombre fini de fois la règle R à l'ensemble $\tilde{K} \cup \{(\neg p, 1)\}$.

Ce résultat est facile à établir ; voir Dubois et Prade (1987a) et Lang (1991), Dubois et col. (1994). Notons $\text{cons}(\tilde{K})$ la fermeture déductive de $\tilde{K}$. Il est aisé de voir que $\text{cons}(\tilde{K}) = \{(p, \alpha_i), \exists i, K_1 \cup K_2 \cup \dots \cup K_i \vdash p\}$, c'est-à-dire que la strate de niveau α_i est la fermeture déductive de $K_1 \cup K_2 \cup \dots \cup K_i$, pour tout $i = 1, n$.

On a introduit ci-dessus une syntaxe qui manipule des formules pondérées. Il est naturel de vouloir en trouver une sémantique, en interprétant le poids α attaché à une formule p comme une mesure de confiance attachée à l'événement correspondant à l'ensemble des interprétations de p .

Dans la suite on notera $[p]$ l'ensemble des interprétations de la formule p , que l'on identifiera à l'ensemble des situations où p est vrai, Ω désignant l'ensemble des interprétations de toutes les formules. On notera $\omega \models p$ quand $\omega \in [p]$. On dit alors que p est vrai dans l'interprétation ω ou encore que ω satisfait p . Soit $\tilde{K}$ une base de connaissance stratifiée *consistante*, c'est-à-dire $\text{inc}(\tilde{K}) = 0$. Notons $N([p]) = \max\{\alpha_i, \tilde{K} \vdash (p, \alpha_i)\}$ que l'on écrira $N(p)$ pour simplifier. Il est facile de vérifier que

$$N(\perp) = 0, N(\top) = 1 \quad (1)$$

$$N(p \wedge q) = \min(N(p), N(q)). \quad (2)$$

La formule (2) dit que si un état cognitif accepte p et q , il doit accepter leur conjonction. Ces propriétés sont caractéristiques d'une fonction d'ensemble appelé mesure de nécessité en théorie des possibilités (Dubois et Prade, 1987b). $N(p)$ est le degré de certitude de l'événement $[p]$ (donc de l'assertion p). $N(p) = 1$ signifie que p est certain, tandis que $N(p) = 0$ signifie seulement que p est totalement incertain. Il est facile de voir que $\min(N(p), N(\neg p)) = 0$, c'est-à-dire que sur la base d'informations consistantes (l'ensemble $\tilde{K}$), on ne peut être à la fois quelque peu certain de p et de son contraire. La fonction d'ensemble N est similaire à une fonction de probabilité sauf qu'elle obéit à des axiomes et des conventions différentes. En particulier la quantité $1 - N(\neg p) = \Pi(p)$ est différente de $N(p)$. Elle est appelée degré de possibilité de p , parce que si $\neg p$ est certain ($N(\neg p) = 1$) alors p est impossible ($\Pi(p) = 0$). La mesure de possibilité a été introduite par Zadeh (1978). Elle vérifie l'axiome caractéristique (dans le cas fini) : $\Pi(p \vee q) = \max(\Pi(p), \Pi(q))$ qui est dual de (2). Ce type de fonction a été aussi introduit par l'économiste Shackle (1961), et le philosophe L.J. Cohen (1973) sous des formes diverses et plus récemment par Gärdenfors (1988) sous le nom d' "enracinement épistémique", et encore Gärdenfors et Makinson (1994) sous le nom d' "expectation".

L'intérêt d'utiliser deux nombres $\Pi(p)$ et $N(p)$ pour représenter la croyance relative à p , c'est de pouvoir représenter l'ignorance ($\Pi(p) = 1, N(p) = 0$) de façon absolue, sans référence aux nombres de situations recouvertes par p et $\neg p$ respectivement. Cela n'est pas possible avec une fonction de probabilité unique. L'axiome des croyances acceptées (2) est bien sûr en contradiction avec la notion de probabilité.

La théorie des possibilités n'est pas étrangère aux logiques modales, puisqu'on manipule les concepts duaux de possibilité (ici entendue comme cohérence avec un

état de connaissance incomplet) et de nécessité (ici entendue comme certitude) (voir Fariñas del Cerro et Herzig (1991) sur ce point).

Etant donné une mesure de possibilité Π sur un ensemble Ω supposé fini, qui associe à chaque sous-ensemble A de Ω un degré de possibilité $\Pi(A)$, on peut définir complètement la fonction Π à l'aide d'une allocation de poids π aux éléments de Ω appelée distribution de possibilité. On peut en effet poser $\pi(\omega) = \Pi(\{\omega\})$ et on vérifie que

$$\Pi(A) = \max_{\omega \in A} \pi(\omega), \quad N(A) = \min_{\omega \notin A} 1 - \pi(\omega). \quad (3)$$

Un ensemble fini de formules propositionnelles K caractérise un sous-ensemble $[K]$ de Ω contenant toutes les interprétations ω pour lesquelles toutes les formules de K sont vraies. Ces interprétations sont appelées modèles de K , et existent dès que K est consistant.

De façon similaire une base stratifiée $\tilde{K}$ va caractériser un sous-ensemble flou (au sens de Zadeh (1965)) $[\tilde{K}]$ de modèles dont la fonction d'appartenance $\mu_{[\tilde{K}]}$ peut s'interpréter comme une distribution de possibilité sur Ω . On peut la calculer à partir des poids α_i attachés aux formules $p_i \in K$. Plus précisément la formule pondérée (p_i, α_i) correspond à l'inégalité

$$N(p_i) = \min_{\omega \notin [p_i]} 1 - \pi(\omega) \geq \alpha_i$$

soit $\pi(\omega) \leq 1 - \alpha_i$ si $\omega \notin [p_i]$, c'est-à-dire si $\omega \models \neg p_i$. S'il y a n formules pondérées dans $\tilde{K}$, on obtient n contraintes de ce type, soit

$$\forall \omega, \pi(\omega) \leq \min_{i: \omega \models \neg p_i} 1 - \alpha_i.$$

Une distribution de possibilité π est plus ou moins informative selon le nombre d'interprétations qu'elle suggère comme possibles. Plus l'ensemble d'interprétations possibles est grand, moins π est informative. Si $\pi < \pi'$, π est dite plus spécifique que π' car π est plus informative, moins imprécise que π' (Yager, 1983). En théorie des possibilités, le principe de minimum de spécificité (par exemple, Dubois et Prade (1991d)) stipule que la distribution de possibilité qui représente le plus fidèlement un ensemble de contraintes, est celle qui attribue à chaque interprétation le plus grand degré de possibilité tout en respectant ces contraintes. C'est un principe informationnel de moindre engagement. L'application de ce principe à notre problème conduit à définir la distribution de possibilité π induite par un ensemble de formules pondérées comme

$$\begin{aligned} \pi(\omega) &= \min_{i: \omega \models \neg p_i} 1 - \alpha_i \\ &= 1 \text{ si } \omega \models p_1 \wedge p_2 \wedge \dots \wedge p_n. \end{aligned}$$

Si $\tilde{K}$ est consistant, π est normalisée, c'est-à-dire vaut 1 pour au moins une interprétation. Si $\tilde{K}$ est inconsistant, π est sous-normalisée, mais reste définie et

permet de classer les interprétations, dès que la strate la plus certaine de $\tilde{K}$, K_1 , reste consistante (et ne contient pas que des tautologies). Les interprétations ω telles que $\pi(\omega) = \max \pi$ sont dites *modèles préférés* de $\tilde{K}$; lorsque $\tilde{K}$ est inconsistante, ces interprétations ne falsifient pas de formules p_i telles que $\alpha_i > 1 - \sup \pi$.

La notion de conséquence logique peut alors s'étendre au cas des bases stratifiées. On dira que (p, α) est conséquence logique de $\tilde{K}$ si et seulement si $\forall \omega \models \neg p, \pi(\omega) \leq 1 - \alpha$ et on notera $\tilde{K} \models (p, \alpha)$. Il est facile de voir que $\tilde{K} \models (p, \alpha) \Leftrightarrow N(p) \geq \alpha$ où N est construite à partir de π et K selon (3) quand K est consistant. Cette sémantique capture complètement l'inférence à partir des bases stratifiées au sens où on a pu montrer que la logique possibiliste est saine et complète par rapport à la technique de réfutation et résolution, soit $\tilde{K} \vdash (p, \alpha) \Leftrightarrow \tilde{K} \models (p, \alpha)$. Notons que ce dernier résultat inclut le cas où K est partiellement inconsistant (Lang, 1991 ; voir aussi Dubois et col., 1994) ; on a alors $\sup \pi = 1 - \text{inc}(\tilde{K}) < 1$.

3. L'inférence préférentielle et ses propriétés

Si la logique possibiliste permet de pallier quelque peu le problème de trivialisation des inférences due à l'inconsistance, grâce à la stratification, la notion d'inférence définie ci-dessus n'y échappe pas totalement. En effet si $\text{inc}(\tilde{K}) > 0$ on a bien $\tilde{K} \vdash (p, \text{inc}(\tilde{K}))$ pour toute formule p . On distingue alors deux types de conséquence de $\tilde{K}$: celles de la forme ci-dessus seront dites conséquences triviales car elles ne résultent que de la présence de l'inconsistance partielle. En revanche une conséquence de $\tilde{K}$ telle que $\tilde{K} \vdash (p, \alpha)$ avec $\alpha > \text{inc}(\tilde{K})$ sera dite conséquence non triviale de $\tilde{K}$; on notera $\sim_{\pi} p$ le fait que p est conséquence non triviale de $\tilde{K}$ associée à la distribution de possibilité π .

Au niveau syntaxique, l'ensemble des conséquences non triviales de $\tilde{K}$ s'obtient en calculant la fermeture déductive de l'ensemble $\tilde{K}^* = \{p_i, (p_i, \alpha_i) \in \tilde{K}, \alpha_i > \text{inc}(\tilde{K})\}$ c'est-à-dire le seul sous-ensemble consistant de formules $K_1 \cup K_2 \cup \dots \cup K_i$ tel que $K_1 \cup K_2 \cup \dots \cup K_{i+1}$ est inconsistant et $\alpha_{i+1} = \text{inc}(\tilde{K})$. Au niveau sémantique cela revient à calculer l'ensemble $C[\pi]$ des interprétations préférées au sens de π , soit $\pi(\omega) = \max \pi, \forall \omega \in C[\pi]$; on vérifie alors que $\sim_{\pi} p \Leftrightarrow C[\pi] \subseteq [p]$, c'est-à-dire que tous les modèles préférés de $\tilde{K}$ sont des modèles de p . On reconnaît ici l'inférence préférentielle proposée par Shoham (1988) pour caractériser de nombreux systèmes de logiques non monotones. En effet, on peut vérifier que les propriétés de la relation de conséquence $\sim_{\pi}$ sont plus faibles que celles de la notion usuelle de conséquence. Dans cette approche la distribution de possibilité π calculée à partir de $\tilde{K}$ représente à quel point les situations codées par des interprétations sont plausibles, normales. Elle est donc censée modéliser une connaissance générique.

On dira qu'une proposition p entraîne plausiblement q dans le contexte $\tilde{K}$, ce qu'on notera $p \sim_{\pi} q$, si et seulement si tous les modèles préférés de p sont des modèles de q , ou de façon équivalente $\tilde{K} \cup \{(p,1)\} \vdash (q, \alpha)$ avec $\alpha > \text{inc}(\tilde{K} \cup \{p\})$ (Dubois et Prade, 1991c). La relation $\sim_{\pi}$ n'est ni partout réflexive, ni transitive, ni monotone. C'est-à-dire par exemple, qu'il est faux que $p \sim_{\pi} q$, $q \sim_{\pi} r$ entraîne $p \sim_{\pi} r$, et que $p \sim_{\pi} r$ entraîne $p \wedge q \sim_{\pi} r$. Ce n'est pas surprenant. Une conclusion plausible est une conclusion à laquelle on parvient en supposant que la situation est normale, qu'on se trouve dans un des états possibles du monde les plus plausibles, c'est-à-dire sans vérifier toutes les conditions qui permettraient de pouvoir affirmer définitivement cette conclusion. En pratique, effectuer toutes ces vérifications peut s'avérer coûteux, et inutile sous l'hypothèse de "normalité" de la situation. Si tout ce qu'on sait d'un animal, c'est qu'il est un oiseau, on peut conclure provisoirement qu'il vole parce qu'on suppose qu'il s'agit d'un oiseau normal. Dans ce modèle, supposer la situation normale équivaut à se restreindre aux modèles préférés de $\tilde{K}$. Par contre, si l'on vient à apprendre que c'est un pingouin, on sort de la classe des oiseaux normaux et on ne peut plus conserver cette conclusion, d'où l'absence de monotonie de l'inférence plausible. L'absence de transitivité lui est très liée car ce n'est pas parce que les pingouins sont des oiseaux et que les oiseaux (normaux) volent qu'il faut en conclure que les pingouins volent.

D'un point de vue technique, la non-monotonie de $\sim_{\pi}$ s'explique très facilement. Pour vérifier $p \sim_{\pi} q$ on est amené à résoudre un problème d'optimisation, à savoir maximiser π sur l'ensemble $[p]$ d'interprétations. Dans le langage de l'optimisation, il est bien connu que l'ensemble des solutions qui maximisent une fonction sur un domaine D ne contient pas forcément l'ensemble des solutions qui maximisent cette fonction sur un sous-domaine D' de D (ici $D = [p]$, $D' = [p \wedge r]$). D'où la non-monotonie de l'inférence plausible. La non-transitivité peut s'expliquer de façon analogue.

Les propriétés de la relation de conséquence $\sim_{\pi}$ sont les suivantes (Dubois et Prade, 1991c ; Benferhat et col., 1992) :

Equivalence à gauche :	si $p \equiv q$ et $p \sim_{\pi} r$ alors $q \sim_{\pi} r$	(EG)
Affaiblissement à droite :	si $p \sim_{\pi} q$ et $q \models r$ alors $p \sim_{\pi} r$	(AD)
Coupure :	si $p \sim_{\pi} q$ et $p \wedge q \sim_{\pi} r$ alors $p \sim_{\pi} r$	(C)
Monotonie prudente :	si $p \sim_{\pi} q$ et $p \sim_{\pi} r$ alors $p \wedge q \sim_{\pi} r$	(MP)
Disjonction à gauche :	si $p \sim_{\pi} r$ et $q \sim_{\pi} r$ alors $p \vee q \sim_{\pi} r$	(DG)
Réflexivité restreinte :	si $p \neq \perp$ alors $p \sim_{\pi} p$	(RR)
Nihil ex absurdo :	$\text{non}(\perp \sim_{\pi} p)$	(NA)
Conjonction à droite :	si $p \sim_{\pi} q$ et $p \sim_{\pi} r$ alors $p \sim_{\pi} q \wedge r$	(CD)
Monotonie rationnelle :	si $p \sim_{\pi} r$ et $\text{non}(p \sim_{\pi} \neg q)$ alors $p \wedge q \sim_{\pi} r$	(MR)

Certaines de ces propriétés telles que MP, C et DG ont été introduites par Gabbay (1985) dans le cadre de la programmation logique, comme des propriétés

naturelles pour une logique non monotone. Elles ont été systématiquement étudiées par Makinson (1989), Lehmann (1989), et Kraus et col. (1990), sauf RR et NA que ces auteurs remplacent par la réflexivité $p \sim_{\pi} p$ et la supra-classicalité ($p = q$ implique $p \sim_{\pi} q$).

Discutons la pertinence de ces propriétés pour l'inférence plausible. EG n'est qu'une simple contrainte de cohérence avec le calcul propositionnel. CD est en accord avec l'axiome (2) des croyances acceptées. CD et AD assurent que l'ensemble des conclusions plausibles est clos par déduction. La règle de coupure C affaiblit la transitivité, en excluant le cas où les modèles de p sont des modèles anormaux de q par rapport à la propriété r . MP affaiblit la monotonie de la même façon. MR est généralement une propriété plus forte que MP, car il suffit qu'on ne puisse pas contredire q dans les situations où p est normal pour pouvoir faire hériter de la propriété r la sous-classe $p \wedge q$ de situations où p est vrai. MR permet de capturer une certaine absence de lien entre p et la propriété q , et produit une inférence plus aventureuse que MP. DG permet d'inférer une conclusion quand cette conclusion peut être obtenue de deux façons sans avoir à disposer de l'information suffisante pour emprunter l'un des chemins de preuve correspondants. RR et NA permettent d'assurer qu'un ensemble de conséquences plausibles sera toujours consistant. En effet en cas d'inconsistance de la prémisse on préfère ne rien inférer plausiblement plutôt que d'inférer n'importe quoi.

On a démontré (Benferhat et col., 1992) qu'inversement, toute relation de conséquence non monotone $\sim$ sur un langage fini, qui satisfait RR, NA, EG, AD, DG, CD, MR est représentable par une relation d'inférence plausible possibiliste. Plus précisément, il existe π telle que $\sim = \sim_{\pi}$. Un résultat analogue, utilisant la supra-classicalité, est prouvé par Makinson et Gärdenfors (1994). Pour construire π , quand on connaît $\sim$, relation binaire sur le langage, on procède comme suit :

On construit sur l'ensemble des formules du langage la relation $>_{\Pi}$ telle que $p >_{\Pi} q \Leftrightarrow p \vee q \sim \neg q$. On peut vérifier que la relation complémentaire $\geq_{\Pi}$ (telle que $p \geq_{\Pi} q \Leftrightarrow \text{non}(q >_{\Pi} p)$) est complète et transitive, et qu'elle satisfait l'axiome des relations de possibilités comparatives (Lewis, 1973)

$$p \geq_{\Pi} q \Rightarrow p \vee r \geq_{\Pi} q \vee r.$$

Dubois (1986) a montré qu'une fonction Π qui associe à toute formule un nombre entre 0 et 1 représente une relation de possibilité comparative (c'est-à-dire $p \geq_{\Pi} q \Leftrightarrow \Pi(p) \geq \Pi(q)$) si et seulement si Π est une mesure de possibilité. Il existe donc une mesure de possibilité Π telle que $\Pi(p) > \Pi(q) \Leftrightarrow p \vee q \sim \neg q$. On peut alors poser $\pi(\omega) = \Pi(\{\omega\})$, $\forall \omega \in \Omega$.

4. Représentation de connaissances génériques

L'approche présentée ci-dessus n'est pas totalement satisfaisante pour le raisonnement plausible. On voit bien que l'ensemble $\tilde{K}$ joue le rôle de connaissances génériques qui, appliquées au fait p , permettent de sauter à la conclusion q si $p \vdash_{\pi} q$. Trois motifs d'insatisfaction peuvent être au moins mentionnés :

1) quand on vérifie si $p \vdash_{\pi} q$ tient, on ajoute le fait p à $\tilde{K}$. Dans $\tilde{K} \cup \{(p,1)\}$ plus rien ne permet de distinguer ce fait des connaissances génériques contenues dans $\tilde{K}$. Le calcul des propositions ne permet pas d'exprimer la distinction entre connaissance générique (par exemple, le savoir d'un médecin) et connaissances factuelles (les résultats de l'observation d'un patient) ;

2) $\tilde{K}$ est une base de connaissances stratifiée, mais en faisant cette supposition on n'a rien dit sur la façon dont on construit cette stratification ;

3) les connaissances génériques s'expriment souvent sous forme de règles "si... alors..." entachées d'exceptions, sans pour autant que l'on puisse toujours quantifier ces exceptions. La logique possibiliste ne permet pas de formaliser ces assertions conditionnelles. En fait on voudrait disposer d'un langage où on puisse exprimer que p_i est une conséquence plausible de q_i , et ce pour un certain nombre de couples de formules (p_i, q_i) , et pouvoir en conclure si, pour une autre paire (p, q) , q est oui ou non conséquence plausible de p . Cela impose d'introduire le symbole $\sim$ dans le langage objet.

La logique possibiliste semble avoir un pouvoir d'expression limité parce qu'une grande part de nos connaissances de sens commun sont de type conditionnel, et expriment directement que certaines conclusions peuvent être plausiblement déduites dans certaines situations génériques. Par exemple, "les oiseaux volent" signifie que lorsqu'on sait seulement qu'un individu est un oiseau, on peut plausiblement inférer qu'il vole. La logique possibiliste permet ce type d'inférence mais non d'exprimer de telles règles par défaut directement en tant que formules dans la syntaxe de cette logique. Il est facile de voir en effet que l'implication matérielle ne rend pas bien compte des mots "si... alors..." dans ces éléments de connaissance.

Une tentation naturelle est de vouloir donner une sémantique probabiliste aux connaissances conditionnelles de sens commun, "si p alors plausiblement q " se traduisant par l'inégalité $P(q | p) \geq \alpha$ avec $\alpha > 0.5$. Malheureusement cette traduction est trop riche pour exprimer fidèlement l'idée d'inférence plausible puisqu'on évalue implicitement la proportion d'exceptions. De plus, l'inférence $\sim$ obtenue viole toutes les propriétés évoquées ci-dessus, sauf si on prend pour α un infinitésimal de la forme $1 - \varepsilon$ où ε est positif infiniment proche de 0. Dans ce cas seule la monotonie rationnelle MR est violée. Cette notion d'inférence plausible à base de probabilités infinitésimales, plus faible que celle calculée par la logique possibiliste a été formalisée par Adams (1975) dans sa logique des conditionnels basée sur les probabilités infinitésimales, et reprise par Pearl (1988, 1990) (qui la note $\vdash_{\phi}$) dans l'optique du raisonnement par défaut.

Soit $\Delta = \{(p_i, q_i), i = 1, n\}$ un ensemble de paires de formules telles que l'on ait $P(q_i | p_i) \geq 1 - \varepsilon, \forall i = 1, n$. On note $\mathcal{P}(\Delta) = \{P | P(q_i | p_i) \geq 1 - \varepsilon, i = 1, n\}$. Adams et Pearl définissent la relation d'inférence (dite p-conséquence) de (p, q) à partir de Δ , notée $\Delta \vdash_0 (p, q)$ par $\forall P \in \mathcal{P}(\Delta), P(q | p) \geq 1 - 0(\varepsilon)$. Cette relation d'inférence vérifie les propriétés EG, AD, C, MP, DG, RR, NA, CD, mais pas MR.

Avec les deux types d'inférence $\vdash_\pi$ et $\vdash_0$, on est capable de gérer dans une large mesure les phénomènes d'exception (les A sont des B, x_0 est un A et n'est pas un B). Mais seule l'inférence possibiliste peut absorber les informations dites non-pertinentes grâce à la propriété MR, (par exemple de conclure que si un oiseau vole, un oiseau rouge vole aussi pour autant qu'on n'ait rien dit sur l'influence de la couleur sur la capacité de voler ; c'est ce qu'exprime la monotonie rationnelle). La relation d'inférence $\vdash_0$ est très prudente à cet égard car elle ne conclut rien sur les oiseaux rouges (car si $P(q | p) > 1 - \varepsilon$ on ne sait rien sur $P(q | p \wedge r)$).

Pearl (1988, 1990) a beau défendre une vision probabiliste de l'inférence plausible, il est difficile d'admettre que "les oiseaux volent" puisse signifier que la probabilité de voler pour un oiseau pris au hasard (dans quoi ?) est infiniment proche de 1. Pour mieux apprécier cette interprétation probabiliste extrême de la règle par défaut, il convient peut être de renvoyer à la notion de croyance acceptée. On sait que croire quelque chose veut dire s'engager à agir comme si ce quelque chose était vrai. Ce que la logique d'Adams nous suggère, c'est qu'on peut rendre compte de cette notion de croyance en augmentant arbitrairement la probabilité du fait cru sans pour autant la mettre à 1 (ce qui permet les remises en cause en évitant que la logique des croyances acceptées coïncide avec la logique classique car $P(q | p) = 1 \Rightarrow P(\neg p \vee q) = 1$; l'équivalence tient si $P(p) > 0$). Ainsi la probabilité infinitésimale $1 - \varepsilon$ n'est pas la probabilité réelle qu'un oiseau vole (si tant est qu'elle existe), mais sa valeur provisoire pour mieux avancer dans un raisonnement plausible.

En fait, ce dont on a surtout besoin pour coder les connaissances génériques, c'est de placer l'inférence plausible $p \sim q$ dans le langage objet, et de lui définir une sémantique plus simple et plus crédible que celle d'Adams. Kraus, Lehmann et Magidor (1990) sont parvenus à donner à l'objet $p \sim q$ une sémantique basée sur des relations de préférence plus générales que celle obtenue à partir d'une distribution de possibilité. Mais leur construction sémantique est complexe et ne cherche pas vraiment à formaliser une intuition. Une autre voie est de considérer $p \sim q$ comme un *objet conditionnel* (encore appelé événement conditionnel) plus à même de coder des connaissances génériques entachées d'exception que des clauses universellement quantifiées de la logique du premier ordre, et moins gourmand en données que des valeurs numériques de probabilités conditionnelles. Cette notion d'objet conditionnel est étroitement liée à un problème qui a fait couler beaucoup d'encre dans les années 70 (voir la célèbre controverse entre Stalnaker et Lewis dans Harper et col. (1989)) pour répondre à la question : de quelle formule logique la probabilité conditionnelle est-elle la probabilité ?

Ce problème a été résolu récemment par Calabrese (1987), Goodman, Nguyen et Walker (1991). L'événement conditionnel $b | a$ tel que $P("b | a") = P(b | a)$ où a et b sont éléments d'une algèbre de Boole $\mathfrak{B}$ (induite par un ensemble d'événements) peut se définir à l'aide de l'intervalle de $\mathfrak{B}$ défini par $\{x \in \mathfrak{B}, a \wedge b \leq x \leq \neg a \vee b\}$ où

$\leq$ est l'ordre partiel défini sur $\mathfrak{B}$ par l'implication logique (ou en termes ensemblistes, l'inclusion). $b \mid a$ est borné inférieurement par la conjonction $a \wedge b$ et supérieurement par l'implication matérielle $\neg a \vee b$. De fait, on peut vérifier facilement que $P(b \mid a)$ est une fonction de $P(a \wedge b)$ et $P(\neg a \vee b)$ à savoir

$$P(b \mid a) = \frac{P(a \wedge b)}{P(a)} = \frac{P(a \wedge b)}{P(a \wedge b) + 1 - P(\neg a \vee b)},$$

et que $P(a \wedge b) \leq P(b \mid a) \leq P(\neg a \vee b)$.

D'un point de vue logique cette représentation d'un événement conditionnel conduit à suggérer une sémantique trivaluée à la règle générique $p \sim q$, définie par (voir Dubois et Prade, 1987b, 1991a, 1994)

- $v(p \sim q) = \text{VRAI}(V)$ dans l'interprétation ω si $\omega \models p \wedge q$ (règle vérifiée)
- $v(p \sim q) = \text{FAUX}(F)$ dans l'interprétation ω si $\omega \models p \wedge \neg q$ (règle falsifiée)
- $v(p \sim q) = \text{INAPPLICABLE}(I)$ sinon.

Cette sémantique a été proposée il y a plus de cinquante ans pour les événements conditionnels par De Finetti (1936, 1937) qui l'a cependant assez peu développée. Les premiers travaux de Calabrese, Goodman et Nguyen, Dubois et Prade sur les objets conditionnels ignoraient ce travail de De Finetti. La sémantique ci-dessus revient à considérer $p \sim q$ comme un événement conditionnel $[q] \mid [p]$ où $[p]$ représente l'ensemble des modèles de $[p]$, etc. $[p \wedge q]$ représente l'ensemble des exemples de la règle, et $[p \wedge \neg q]$ l'ensemble de ses exceptions, dont le complément est justement $[\neg p \vee q]$. $\mathfrak{B}$ représente alors l'ensemble des sous-ensembles d'interprétations.

Une syntaxe pour les règles génériques de la forme $p \sim q$ a été étudiée par Kraus et col. (1990). Etant donné une base d'assertions conditionnelles $K_c = \{p_i \sim q_i \mid i = 1, n\}$. La déduction de $p \sim q$ à partir de K_c , notée $K_c \vdash p \sim q$ s'effectue en utilisant comme règles d'inférence les propriétés de l'inférence non monotone listées à la section précédente sauf MR d'une part, et d'autre part NA et RR remplacées par un axiome de réflexivité $p \sim p$. Ce système formel est appelé P par Kraus et col. (1990), et s'avère très proche de la logique conditionnelle d'Adams.

La sémantique de $p \sim q$ en tant qu'événement conditionnel $[q] \mid [p]$ doit être complétée en précisant la notion de conséquence logique entre événements conditionnels, ainsi que leur conjonction. On dira que $p \sim q$ implique sémantiquement $p' \sim q'$ (noté $p \sim q \models p' \sim q'$) si et seulement si $p' \sim q'$ possède plus d'exemples et moins d'exceptions que $p \sim q$, ce qui s'écrit $p \wedge q \models p' \wedge q'$ et $p' \wedge \neg q' \models p \wedge \neg q$. En termes de valeurs de vérité, cela revient à interpréter la troisième valeur de vérité comme "inconnue" plutôt qu'inapplicable et à ordonner $\{F, I, V\}$ du faux au vrai ($F < I < V$). On a alors (Dubois et Prade, 1991a)

$$p \sim q \models p' \sim q' \Leftrightarrow v(p \sim q) \leq v(p' \sim q').$$

Il faut maintenant étendre la notion de conséquence logique à un ensemble d'assertions conditionnelles. Pour cela on utilise une quasi-conjonction introduite par Adams (1975) et retrouvée par Calabrese (1987). Elle construit une assertion conditionnelle $p \sim q$ à partir d'objets conditionnels $p_1 \sim q_1$ et $p_2 \sim q_2$ formant une base $\{p_1 \sim q_1, p_2 \sim q_2\} = K_c$. De manière générale une base de règles K_c est dite *vérifiée par une interprétation* si l'une de ses règles est vérifiée sans qu'aucune autre ne soit falsifiée ; elle est dite falsifiée si l'une de ses règles est falsifiée, et inapplicable si aucune règle n'est applicable, ce qui suggère l'opération suivante

$$(p_1 \sim q_1) \& (p_2 \sim q_2) = (p_1 \vee p_2) \sim (\neg p_1 \vee q_1) \wedge (\neg p_2 \vee q_2).$$

Cette quasi-conjonction exprime que, dans un contexte où soit p_1 est vrai, soit p_2 est vrai soit les deux (d'où $p_1 \vee p_2$ en partie condition de l'assertion résultante), on dispose des deux implications matérielles $p_1 \rightarrow q_1$ et $p_2 \rightarrow q_2$. Sémantiquement, cela revient à considérer un autre ordre sur $\{V, I, F\}$ à savoir $F < V < I$, et à poser

$$v((p_1 \sim q_1) \& (p_2 \sim q_2)) = \min(v(p_1 \sim q_1), v(p_2 \sim q_2))$$

au sens de cet ordre (Dubois et Prade, 1991a). Cela permet de ne pas faire intervenir les règles inapplicables dans la conjonction. Cette quasi-conjonction est donc associative. Il est à noter que De Finetti (1936) propose comme conjonction celle (préférée également par Goodman et *al.* (1991)) compatible avec l'ordre $F < I < V$ pour décrire "un pari dont la validité est subordonnée à des conditions qui doivent être vérifiées", et non dans une perspective de raisonnement plausible. La logique des objets (ou événements) conditionnels repose donc sur une structure trivaluée qui n'est pas sans rappeler les bi-treillis au sens de Ginsberg (1988) puisque basée sur deux ordres.

Si S est un ensemble d'assertions conditionnelles, notons $C(S)$ leur quasi-conjonction. On peut vérifier que $C(S)$ se déduit de S au niveau syntaxique, par les règles du système P , soit $S \vdash C(S)$ comme l'a remarqué Adams (1975). On définit par ailleurs la notion de conséquence logique $K_c \models p \sim q$ comme suit

$$K_c \models p \sim q \Leftrightarrow \exists S \subseteq K_c, C(S) \models p \sim q$$

au sens de la notion de conséquence logique entre objets conditionnels introduite plus haut. Par convention, si $S = \emptyset$ alors $K \models p \sim q \Leftrightarrow p \models q$. On a pu démontrer que $K \vdash p \sim q$ et $K \models p \sim q$ sont équivalents, c'est-à-dire que la logique des objets conditionnels capture le raisonnement plausible au sens du Système P (Dubois et Prade, 1994).

La notion d'inconsistance dans K_c est plus complexe que dans le cas propositionnel car elle correspond à l'existence d'une sous-base $S \subseteq K_c$ telle que $C(S)$ n'est jamais vérifiée. Notons qu'on peut avoir $S \subset K_c$, $C(S)$ jamais vérifiée, mais $C(K_c)$ vérifiable : contrairement au cas classique, une interprétation qui vérifie toutes les règles de $K_c - S$, et rend inapplicables les règles de S , est une interprétation qui vérifie K_c . La vérification de cohérence est donc plus complexe

qu'en logique classique, mais elle se ramène bien à des tests classiques de consistance. En effet, si $S = \{p_i \sim q_i, i = 1, k\}$, $C(S)$ non vérifiable signifie

$$\forall i = 1, k, \{p_i \wedge q_i\} \cup \{\neg p_1 \vee q_1, \dots, \neg p_k \vee q_k\} \text{ est inconsistent}$$

c'est-à-dire que pour tout i les exemples de $p_i \sim q_i$ sont des exceptions pour une autre règle de S . Une règle $p_i \sim q_i$ dont les exemples ne sont exceptions d'aucune règle de S est dite *tolérée* par S (Pearl, 1990). Cette notion de tolérance va induire un ordre entre les règles par défaut. On va ainsi pouvoir faire le lien entre les connaissances génériques codées sous la forme d'une base d'assertions conditionnelles et la logique possibiliste où ces connaissances sont codées comme des clauses pondérées, ce qui permet de se ramener à la logique classique moyennant une stratification.

5. Des objets conditionnels à la logique possibiliste

Considérons à nouveau l'exemple classique des pingouins (Pin) qui ne volent pas et des oiseaux (Oi) qui volent (Vo). Dans la logique des assertions conditionnelles, on codera l'exemple sous la forme de trois assertions : $K_c = \{Oi \sim Vo, Pin \sim Oi, Pin \sim \neg Vo\}$. Dans cet exemple on constate que les exemples de la règle $Pin \sim \neg Vo$ sont des exceptions soit de la règle $Pin \sim Oi$, soit de la règle $Oi \sim Vo$. La règle $Pin \sim Oi$ n'est pas tolérée par les deux autres. En revanche $Oi \sim Vo$ est bien tolérée par les deux autres puisque cette règle a des exemples qui ne sont pas des exceptions des deux autres règles (il suffit de considérer les non-pingouins). Cela signifie qu'appliquer la règle $Oi \sim Vo$, plus générale, n'engendre pas de conflit avec les autres règles. En revanche appliquer une règle non tolérée va créer des incohérences avec les autres règles. Pour éviter cela, Pearl (1990) a suggéré de donner la priorité aux règles non tolérées sur les règles tolérées. Ici on enlèvera donc la règle $Oi \sim Vo$ pour la placer sur une strate de moindre priorité, la strate prioritaire contenant les deux autres règles puisqu'elles se tolèrent mutuellement.

On peut donc définir une procédure de stratification d'une base d'assertions conditionnelles en procédant comme suit (Pearl, 1990) : $K_c = K_{c_1} \cup K_{c_2} \cup \dots \cup K_{c_k}$ où K_{c_k} contient les règles tolérées par $K_{c_1} \cup \dots \cup K_{c_{k-1}}$, $K_{c_{k-1}}$ les règles tolérées par $K_{c_1} \cup \dots \cup K_{c_{k-2}}$ etc. L'ordre ainsi construit sur les règles correspond à donner la priorité aux règles les plus spécifiques, c'est-à-dire à celles qui s'appliquent aux plus petits sous-ensembles d'interprétations, aux classes de références les plus restreintes. Pearl code en nombres entiers les niveaux de priorité et en déduit un ordre sur les interprétations du langage. Dans Benferhat et col. (1992) nous avons montré que cet ordre coïncide avec celui qui est induit par la distribution de possibilité associée à la base de connaissances propositionnelles stratifiée $\tilde{K}$ construite à partir de K_c en remplaçant chaque règle générique $p \sim q \in K_{c_i}$ par la clause pondérée $(\neg p \vee q, \alpha_i)$ où les α_i sont choisis arbitrairement dans $[0,1]$ en satisfaisant les inégalités $\alpha_1 > \alpha_2 > \dots > \alpha_k$. Les déductions effectuées à partir de la

base stratifiée d'assertions conditionnelles $K_{c_1} \cup K_{c_2} \cup \dots \cup K_{c_k}$ où on tient compte de la priorité sont les mêmes que celles qui seront effectuées à partir de la base de connaissance possibiliste $K_1 \cup K_2 \cup \dots \cup K_k$. Plus précisément on démontre que

1) Les déductions effectuées à partir de $K_{c_1} \cup \dots \cup K_{c_2}$ contiennent celles obtenues à partir des propriétés de la logique conditionnelle P de Kraus et col. (1990), et vérifient de plus la propriété de monotonie rationnelle MR, étudiée par Lehmann (1989). Plus précisément on calcule $\text{cons}(K_c) = \{p \sim q, K_c \vdash p \sim q\}$ dans le Système P. L'application systématique à $\text{cons}(K_c)$ de la règle MR de monotonie rationnelle, quand elle est possible, définit une extension dite rationnelle (fermée pour la règle MR) de K_c . Il peut exister plusieurs extensions rationnelles de K_c . Lehmann (1989) a proposé une méthode pour sélectionner une unique extension rationnelle, appelée "fermeture rationnelle", au moyen d'une fonction à valeurs sur les entiers positifs qui ordonne les propositions selon leur "exceptionnalité". Notons $R(K_c)$ la fermeture rationnelle de K_c , et $\vdash_R$ la relation d'inférence associée.

2) Si π^* est la distribution de possibilité la moins spécifique induite par la stratification de K, alors on démontre que :

$$K_c \vdash_R p \sim q \text{ si et seulement si } \Pi^*(p \wedge q) > \Pi^*(p \wedge \neg q)$$

où Π^* est la mesure de possibilité construite sur π^* . Ce résultat se déduit facilement des travaux de Pearl (1990) qui construit la distribution de possibilité la moins spécifique (sous le nom de "compact ranking") sur les interprétations, directement à partir d'un ensemble de règles par défaut. Il la code sur les entiers et il suffit d'un simple changement d'échelle pour se ramener à la théorie des possibilités (Benferhat et col., 1992). Notons que l'inégalité $\Pi^*(p \wedge q) > \Pi^*(p \wedge \neg q)$ exprime qu'il est strictement plus possible, étant donné $\tilde{K}$, que $p \wedge q$ soit vrai plutôt que $p \wedge \neg q$, ce qui exprime que quand p est vrai il est davantage possible que q soit vrai plutôt que le contraire.

En conséquence, la mise en œuvre de la déduction $\vdash_R$ peut se faire grâce à la logique possibiliste en codant la base de connaissances génériques K_c sous la forme $\tilde{K}$ ci-dessus. On a alors

$$K_c \vdash_R p \sim q \text{ si et seulement si } p \sim_{\pi} q$$

c'est-à-dire si $p \neq \perp$, $\tilde{K} \cup (p, 1) \vdash (q, \alpha)$ avec $\alpha > \text{inc}(\tilde{K} \cup (p, 1))$. On voit que dans cette approche du raisonnement plausible on doit mettre en œuvre deux représentations : une représentation conditionnelle où la distinction entre connaissances factuelles et génériques apparaît clairement (si F est la base de faits, K_c la base de règles et q une conclusion à prouver, on cherche à montrer que $K_c \vdash_R F \sim q$) ; une représentation classique stratifiée qui ne sert que pour le calcul effectif de l'inférence, et où la distinction factuel/générique disparaît. La nécessité de deux langages, l'un pour la représentation l'autre pour l'inférence, a été mise en évidence par Lamarre (1992) dans un formalisme logique modal conditionnel, plus général que celui qui est présenté ici, car acceptant l'emboîtement des (objets) conditionnels.

Si on reprend l'exemple des pingouins, on définira $\tilde{K} = \{(\neg Oi \vee Vo, \alpha_2), (\neg Pin \vee Oi, \alpha_1), (\neg Pin \vee \neg Vo, \alpha_1)\}$ avec $\alpha_1 > \alpha_2$. Si Rou désigne "rouge", on n'a pas $K_c \vdash Oi \wedge Rou \sim Vo$ mais bien $K_c \vdash_R Oi \wedge Rou \sim Vo$, de plus on a vu que $\tilde{K} \cup \{(Oi, 1)\} \vdash (Vo, \alpha_2)$, $\tilde{K} \cup \{(Oi, 1), (Pin, 1)\} \vdash (\neg Vo, \alpha_1)$, ces inférences étant non triviales. Cette coïncidence entre inférence possibiliste et fermeture rationnelle ne doit pas surprendre au vu de la section 3.

Pour terminer, deux remarques semblent dignes d'intérêt.

• La logique conditionnelle décrite à la section précédente peut s'interpréter, comme il a été dit, en termes probabilistes : si $p \sim q$ s'entend comme $P(q | p) \geq 1 - \varepsilon$ où ε est un infinitésimal, on obtient une autre sémantique pour le système P. C'est d'ailleurs Adams (1975) qui le premier a offert une version du système P avec cette sémantique. Néanmoins si ε n'est plus infinitésimal, on perd les propriétés du système P, qui ne subsistent que sous une forme dégénérée et peu puissante ; voir (Dubois et Prade, 1991a, 1994). Dans le cas de l'inférence rationnelle ou possibiliste, $p \sim q$ peut s'entendre comme $N(q | p) > 0$ où $N(q | p)$ est une mesure de nécessité conditionnelle. Elle est définie en écrivant une version qualitative du théorème de Bayes pour les mesures de possibilité (Dubois et Prade, 1987b)

$$\Pi(p \wedge q) = \min(\Pi(q | p), \Pi(p)).$$

$\Pi(q | p)$ se calcule à partir de $\Pi(p \wedge q)$ et de $\Pi(p)$ en appliquant le principe de spécificité minimale à l'équation ci-dessus. On a alors si $\Pi(p) > 0$:

$$\begin{aligned} \Pi(q | p) &= 1 \text{ si } \Pi(p) = \Pi(p \wedge q) \\ &= \Pi(p \wedge q) \text{ sinon} \end{aligned}$$

$$\text{et } N(q | p) = 1 - \Pi(\neg q | p).$$

On a vu que $p \sim_\pi q$ équivaut à $\Pi(q \wedge p) > \Pi(\neg q \wedge p)$. Il est facile de voir que cette inégalité implique $\Pi(p) = \Pi(q \wedge p)$ (car $\Pi(p) = \max(\Pi(q \wedge p), \Pi(\neg q \wedge p))$) par application de l'axiome caractéristique des mesures de possibilité, et donc $\Pi(q | p) = 1$, $\Pi(\neg q | p) = \Pi(\neg p \wedge q) < 1$ ce qui équivaut à $N(q | p) > 0$.

De fait, on peut représenter directement un ensemble de connaissances génériques $K = \{p_i \sim q_i, i = 1, n\}$ par des inégalités de la forme $N(q_i | p_i) > 0$. Si on cherche la distribution de possibilité la moins spécifique qui respecte les inégalités $\Pi(q_i \wedge p_i) > \Pi(\neg q_i \wedge p_i)$, $\forall i = 1, n$, on retrouve bien l'ordre calculé par la procédure de Pearl sur les objets conditionnels (voir Benferhat et col., 1992). Par ailleurs, si au lieu de sélectionner une distribution de possibilité particulière, on les considère toutes, on peut définir une relation de conséquence plus prudente :

$\tilde{K} \vdash_\vee p \sim q$ si et seulement si pour tout Π satisfaisant les contraintes $\Pi(p_i \wedge q_i) > \Pi(p_i \wedge \neg q_i)$, $i = 1, n$, on a $\Pi(p \wedge q) > \Pi(p \wedge \neg q)$.

Cette relation de conséquence a été initialement étudiée par Fariñas, Herzig et Lang (1994). Il a été établi qu'elle ne satisfait pas la monotonie rationnelle et qu'elle correspond exactement à la relation de conséquence définie par Lehmann (1989) à partir du système P (ou encore à l'inférence $\vdash_0$ de Pearl (1990), voir (Dubois et Prade, 1995). Il y a donc une approche "purement possibiliste" du raisonnement plausible, depuis l'expression de connaissances génériques jusqu'à l'inférence en passant par le calcul de la stratification.

- L'approche conditionnelle met en lumière deux types de révision dans les bases de connaissances : la révision factuelle (ou focalisation) et la révision générique. La révision factuelle consiste à enrichir une base de faits F en y ajoutant une nouvelle information p sans toucher à la base de règles K_c ; dans ce cas la prise en compte de p se fait par simple ajout de p à la base de connaissances propositionnelles stratifiée $\tilde{K}$ construite à partir de K_c , l'effet de non-monotonie ayant lieu si $\text{inc}(\tilde{K} \cup \{p\}) > \text{inc}(\tilde{K})$. La révision générique consiste à modifier la base de connaissances génériques K_c en une nouvelle base de règles K'_c (par exemple en y ajoutant une nouvelle règle). Dans ce cas il faut recalculer l'ordre induit par la nouvelle base générique modifiée, c'est-à-dire modifier en conséquence la stratification de la base propositionnelle associée $\tilde{K}'$ avant de procéder à l'inférence.

Il y a là une grosse différence avec l'approche de Gärdenfors (1988) qui n'envisage la révision d'une base de connaissances que comme factuelle (c'est-à-dire due à l'ajout d'une proposition inconsistante avec un ensemble de propositions clos par déduction) et la stratification (appelée par lui enracinement épistémique) que comme une façon de coder l'opération de révision. Dans l'optique du raisonnement plausible adoptée ici, la stratification est calculée à partir des connaissances génériques, la révision à la Gärdenfors ne sert que pour la révision factuelle, tandis que la révision générique modifie la stratification elle-même. Cela répond à une question qu'ont posée de nombreux chercheurs à l'approche de Gärdenfors : comment modifie-t-on l'enracinement épistémique ? La réponse est : cela correspond à une révision de connaissance générique et peut être réalisé dans le cadre de la logique des événements (ou objets) conditionnels, par le recalcul de la stratification induite par l'adjonction de la nouvelle connaissance générique.

6. Conclusion

L'émergence sous diverses formes, dont seules quelques-unes ont été évoquées ici, d'un concept d'objet conditionnel à partir des divers travaux existants sur la formalisation des raisonnements de sens commun, par opposition à la notion habituelle de formule logique ne devrait pas surprendre. Elle ne fait que souligner l'existence dans ces raisonnements de deux types de connaissances : celles que l'on peut qualifier de génériques et celles que l'on peut qualifier de factuelles. Cette distinction est à la base de la structuration des systèmes experts en bases de faits et bases de règles. Le recours à la seule logique des propositions ne permet pas de faire cette distinction au niveau du langage. L'utilisation de la logique du premier ordre permet d'exprimer cette distinction (des faits instanciés, et des règles quantifiées

universellement), mais se heurte au traitement des exceptions. La théorie des probabilités gère cette distinction depuis longtemps, mais propose un arsenal beaucoup trop riche pour formaliser les connaissances de sens commun dans leur pauvreté (on ne quantifie pas facilement les taux d'exceptions et on ne cherche pas toujours à le faire). Une logique des objets conditionnels semble à même de capturer la distinction entre connaissances génériques et factuelles et d'offrir un modèle symbolique digne d'intérêt pour certains raisonnements de sens commun.

L'intérêt potentiel des objets conditionnels ne remet pas pour autant en cause le bien-fondé et l'utilité de la logique classique. En effet l'inférence plausible, comme on l'a suggéré ci-dessus, peut se calculer à l'aide de procédures classiques sur des ensembles stratifiés de formules propositionnelles. Par ailleurs, l'approche décrite ici n'est pas sans défaut. Elle ne règle pas de façon satisfaisante tous les problèmes de raisonnement plausible. Notamment il faut remarquer que l'ordre induit sur les défauts par la relation de tolérance ne permet pas bien de gérer les processus d'héritage entre classes comportant des exceptions. Par exemple, si à la base de connaissances sur les oiseaux on ajoute que les oiseaux ont des pattes ($O_i \sim Pa$) alors on ne peut pas conclure que les pingouins (qui ne volent pas) ont des pattes. C'est-à-dire si $K_c = \{Pin \sim \neg Vo, O_i \sim Vo, O_i \sim Pa, Pin \sim O_i\}$, on n'a ni $K \vdash Pin \sim Pa$ ni $K \vdash_R Pin \sim Pa$. En effet la règle $O_i \sim Pa$ est jugée moins prioritaire que $Pin \sim O_i$ et $Pin \sim Vo$ et sera inhibée par ces règles. En termes de logique possibiliste cela se traduit par le fait que $\{Pin\} \cup \tilde{K} \models (Pa, \alpha)$ est une inférence triviale (effet de noyade dans le niveau d'inconsistance α). Pourtant on peut souhaiter que les pingouins héritent de toutes les propriétés de la classe oiseau sauf de celles pour lesquelles on sait que la classe pingouin est exceptionnelle.

Un certain nombre de propositions ont été émises pour traiter ce problème (Geffner, 1992 ; Cayrol et col., 1992 ; Benferhat et col., 1992, 1993). Elles consistent à altérer la relation d'inférence de manière à éviter l'inhibition des règles de faible priorité qui ne participent pas au conflit engendré par le phénomène d'exception. Par exemple, en logique possibiliste on peut considérer comme non triviale une inférence $\tilde{K} \models (p, \alpha)$ avec $\alpha = inc(\tilde{K})$ s'il existe un chemin de preuve de la proposition p n'utilisant aucune formule qui cause l'inconsistance. De manière plus fondamentale, il semble qu'une solution puisse être apportée à ce problème en introduisant une notion d'indépendance qualitative. En effet si on souhaite voir les pingouins hériter de la propriété d'"avoir des pattes", du fait qu'ils sont des oiseaux, quoique ce soient des oiseaux qui ne volent pas, c'est qu'on considère intuitivement que, dans le contexte 'oiseau', 'avoir des pattes' et 'ne pas voler' sont des choses indépendantes. Une notion qualitative d'indépendance entre événements a été ainsi formalisée dans le cadre de la théorie des possibilités. Dire que q est indépendant de p dans le contexte c , revient à écrire que si $N(q \mid c) > 0$ on doit aussi avoir $N(q \mid c \wedge p) > 0$. On est ainsi conduit à ajouter à la base de connaissances de nouvelles règles par défaut (ici $c \wedge p \sim q$ si $c \sim q$ y figure déjà et qu'on juge q indépendant de p dans le contexte c) qui permettent alors d'obtenir les conclusions attendues ; voir (Benferhat et col. 1994 a), (Dubois et col. 1994). Néanmoins le travail doit encore se poursuivre sur ce thème avant qu'une approche totalement satisfaisante puisse être proposée.

Pour terminer, on peut s'interroger sur le statut épistémologique des recherches présentées ci-dessus. Nous considérons qu'elles relèvent des sciences cognitives et

de ce qu'on appelle « Intelligence Artificielle ». En effet on s'intéresse à un mode de raisonnement "naturel" que l'homme pratique quotidiennement et on cherche à en donner un modèle formel ; on peut considérer qu'une telle tentative est du ressort des sciences cognitives. Par ailleurs on cherche également une formalisation qui soit compatible avec une mécanisation de ce type de raisonnement sur un ordinateur ; en ce sens, ces recherches participent pleinement à ce qu'on appelle Intelligence Artificielle. Néanmoins, contrairement à certaines démarches rencontrées en sciences cognitives, l'approche présentée ici pour le raisonnement plausible est plus normative que descriptive. On peut voir les propriétés de l'inférence plausible définies en section 3 comme des postulats de rationalité au même sens que les postulats proposés par Savage (1972) pour le choix dans l'incertain. En ce sens nous observons une convergence des méthodologies, et des préoccupations entre les travaux qui cherchent à formaliser les raisonnements de sens commun et la théorie de la décision. L'analyse des ressemblances et des différences entre ces deux domaines nous paraît prometteuse pour mieux comprendre le statut de l'Intelligence Artificielle en tant que discipline scientifique.

Bibliographie

- Adams E.W. (1975) *The Logic of Conditionals*. D. Reidel, Dordrecht, Pays-Bas.
- Benferhat S., Cayrol C., Dubois D., Lang J., Prade H. (1993) Inconsistency management and prioritized syntax-based entailment. *Proc. of the 13th Inter. Joint Conf. on Artificial Intelligence (IJCAI'93)*, Chambéry, Aug. 28-Sept. 3, 640-645.
- Benferhat S., Dubois D., Prade H. (1992) Representing default rules in possibilistic logic. *Proc. of the 3rd Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'92)*, Cambridge, MA, USA, Oct. 26-29, 673-684.
- Benferhat S., Dubois D., Prade H. (1993) Argumentative inference in uncertain and inconsistent knowledge bases. *Proc. of the 9th Conf. on Uncertainty in Artificial Intelligence*, Washington, D.C., Morgan-Kaufmann, 411-419.
- Benferhat S., Dubois D., Prade H. (1994) Expressing independence in a possibilistic framework and its application to default reasoning. *Proc. 11th European Conference on Artificial Intelligence (ECAI'94)* Edited by A. Cohn, John Wiley & Sons, 150-154.
- Brewka G. (1991) *Nonmonotonic reasoning : logical foundations of commonsense*. Cambridge Tracts in Theoretical Computer Science, 12, Cambridge Univ. Press, UK.
- Calabrese P. (1987) An algebraic synthesis of the foundations of logic and probability. *Information Sciences*, 42, 187-237.
- Cayrol C., Royer V., Saurel C. (1992) Management of preferences in assumption-based reasoning. *Proc. 4th Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'92)*, 221-224. Version étendue in *Tech. Rep. IRIT/92-13/R*.
- Cohen L.J. (1973) A note on inductive logic. *The J. of Philosophy*, LXX, 27-40.
- De Finetti B. (1936) *La logique de la probabilité*. Actes du Congrès Inter. de Philosophie Scientifique, Paris, 1935, Hermann et Cie, 565-573.

- De Finetti B. (1937) La prévision, ses lois logiques et ses sources subjectives. *Ann. Inst. H. Poincaré*, 7, 1-68.
- Dubois D. (1986) Belief structures, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence (Bratislava)*, 5(5), 403-416.
- Dubois D., Fariñas del Cerro L., Herzig A., Prade H. (1994a) An ordinal view of independence with application to plausible reasoning. *Proc. 10th Conf. Uncertainty in Artificial Intelligence*, Seattle, July 29-31, Morgan-Kaufmann Publ., 195-203.
- Dubois D., Lang J., Prade H. (1994) Possibilistic logic. In : *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 3 (D.M. Gabbay et col., eds.), Oxford University Press, UK, 439-513.
- Dubois D., Prade H. (1987a) Necessity measures and the resolution principle. *IEEE Trans. on Systems, Man and Cybernetics*, 17(1), 474-476.
- Dubois D., Prade H. (avec la collaboration de Farreny H., Martin-Clouaire R., Testemale C.) (1987b) *Théorie des Possibilités — Applications à la Représentation des Connaissances en Informatique*. Deuxième édition, Masson, Paris.
- Dubois D., Prade H. (1991a) Conditional objects and non-monotonic reasoning. *Proc. of the 2nd Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'91)* (J. Allen, R. Fikes, E. Sandewall, eds.), Cambridge, MA, April 22-25, 1991, Morgan & Kaufmann, San Mateo, CA, USA, 175-185.
- Dubois D., Prade H. (1991b) Epistemic entrenchment and possibilistic logic. *Artificial Intelligence*, 50, 223-239.
- Dubois D., Prade H. (1991c) Possibilistic logic, preferential models, non-monotonicity and related issues. *Proc. of the 12th Inter. Joint Conf. on Artificial Intelligence (IJCAI'91)*, Sydney, Australie, Aug. 24-30, 419-424.
- Dubois D., Prade H. (1991d) Fuzzy sets in approximate reasoning — Part 1 : Inference with possibility distributions. *Fuzzy Sets and Systems*, 40, 143-202.
- Dubois D., Prade H. (1994) Conditional objects as non-monotonic consequence relationships. *IEEE Trans. on Systems, Man and Cybernetics*, 24 (12) 1724-1740. Une version abrégée est dans les *Proc. 4 th Inter. Conf. on Principles of Knowledge Representation and Reasoning*, Bonn, Allemagne, 24-27 Mai, 1994, 170-177.
- Dubois D., Prade H. (1995) Conditional objects, possibility theory and default rules. To appear in "Conditionals : From Philosophy to Computer Sciences". (Crocco G., Fariñas del Cerro L., Herzig A., eds.), Oxford University Press, 311-346.
- Fariñas del Cerro L., Herzig A. (1991) A modal analysis of possibility theory. *Proc. of the Inter. Workshop on Fundamentals of Artificial Intelligence Research (FAIR'91)* (P. Jorrand, J. Kelemen, eds.), Smolenice Castle, Tchécoslovaquie, Sept. 8-12, 1991, Springer Verlag, Berlin, Allemagne, 11-18.
- Fariñas del Cerro L., Herzig A., Lang J. (1994) From ordering-based nonmonotonic reasoning to conditional logics. *Proc. of the 10th Europ. Conf. on Artificial Intelligence (ECAI'92)* (B. Neumann, ed.), Vienna, Australia, Aug. 3-7, 1992, Wiley, New York, 314-318. Version étendue in *Artificial Intelligence*, 66, 375-390.
- Farrény H. (1985) *Les Systèmes Experts*. Cepadues-Editions, Toulouse.

- Gabbay D.M. (1985) Theoretical foundations for non-monotonic reasoning in expert systems. In : *Logics and Models of Concurrent Systems* (K.R. Apt, ed.), Springer Verlag, Berlin, Allemagne, 439-457.
- Gärdenfors P. (1988) *Knowledge in Flux — Modeling the Dynamics of Epistemic States*. The MIT Press, Cambridge, MA, USA.
- Gärdenfors P., Makinson D. (1994). Non-monotonic inference based on expectation ordering. *Artificial Intelligence*, 65, 197-245.
- Geffner H. (1992) *Default Reasoning : Causal and Conditional Theories*. The MIT Press, Cambridge, MA, USA.
- Ginsberg M. (1988) Multivalued logics — A uniform approach to reasoning in artificial intelligence. *Computational Intelligence*, 4, 265-316.
- Goodman I.R., Nguyen H.T., Walker E.A. (1991) *Conditional Inference and Logic for Intelligent Systems : a Theory of Measure-Free Conditioning*. North-Holland, Amsterdam, Pays-Bas.
- Harper W. Stalnaker R., Pearce G. (Eds.) (1981) *Ifs*. D. Reidel, Dordrecht, Pays-Bas.
- Kraus S., Lehmann D., Magidor M. (1990) Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44, 167-207.
- Lamarre P. (1992) *Etude des raisonnements non monotones : apport des logiques des conditionnels et des logiques modales*. Thèse de Doctorat, IRIT, Université Paul Sabatier, Toulouse.
- Lang J. (1991) *Logique possibiliste : aspects formels, déduction automatique, et applications*. Thèse de Doctorat, Université Paul Sabatier, Toulouse.
- Léa Sombé (Besnard P., Cordier M.O., Dubois D., Fariñas del Cerro L., Froidevaux C., Moinard Y., Prade H., Schwind C., Siegel P.) (1989) *Raisonnements sur des Informations Incomplètes en Intelligence Artificielle*. Teknea, Marseille ; *Revue d'Intelligence Artificielle* (Hermès), 2(3-4), 1989, 9-210.
- Léa Sombé (Besnard P., Cordier M.O., Dubois D., Fariñas del Cerro L., Froidevaux C., Moinard Y., Prade H., Schwind C., Siegel P.) (1990) *Reasoning under Incomplete Information in Artificial Intelligence — A comparison of Formalisms Using a Single Example*. Wiley, New York, USA.
- Lehmann D. (1989) What does a conditional knowledge base entail ? *Proc. of the 1st Inter. Conf. on Principles of Knowledge Representation and Reasoning* (R.J. Brachman, H.J. Levesque, R. Reiter, eds.), Toronto, Ont., May 15-18, 212-222.
- Lewis D. (1973) *Counterfactuals*. Basil Blackwell, Oxford, UK.
- Makinson D. (1989) General theory of cumulative inference. In : *Non-Monotonic Reasoning* (Proc. of the 2nd Inter. Workshop, Grassau, FRG, June 1988) (M. Reinfrank, J. De Kleer, M.L. Ginsberg, E. Sandewall, eds.), *Lecture Notes on Artificial Intelligence*, Vol. 346, Springer Verlag, Berlin, Allemagne, 1-18.
- Pearl J. (1988) *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan & Kaufmann, San Mateo, CA, USA.
- Pearl J. (1990). *System Z : a natural ordering of defaults with tractable applications to default reasoning*. *Proc. of the 3rd. Conf. on Theoretical Aspects of Reasoning About Knowledge* (M. Vardi, ed.), Morgan & Kaufman, San Mateo, CA, USA, 121-135.

Rescher N. (1976) Plausible Reasoning. Van Gorcum, Assen, Pays-Bas.

Savage L.J. (1972) The Foundations of Statistics. Dover, New York.

Shackle G.L.S. (1961) Decision, Order and Time in Human Affairs. Deuxième édition, Cambridge University Press, Cambridge, MA, USA.

Shoham Y. (1988). Reasoning About Change – Time and Causation from the Standpoint of Artificial Intelligence. The MIT Press, Cambridge, MA, USA.

Yager R.R. (1983) An introduction to applications of possibility theory. Human Systems Management, 3, 246-269.

Zadeh L.A. (1965) Fuzzy sets. Information and Control, 8, 338-353.

Zadeh L.A. (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-28.

Didier DUBOIS et Henri PRADE

Chapitre 4

Une application de la logique linéaire : la planification déductive

Tout dans ce théâtre était encore
dans le même état, si ce n'est que
le rideau y tombait en lambeaux.

D'après Oscar WILDE.

1. Généralités

Un système est un assemblage cohérent, disposant de degrés de liberté qui lui permettent de se présenter sous des états divers. Agir consiste à modifier l'état d'un système tout en préservant sa cohérence ; planifier consiste à organiser des actions simples en actions complexes, dans un but déterminé.

Ce premier chapitre expose, de façon naïve, quelques-unes des notions relatives au problème général de la planification : *systèmes, ressources, actions*. Tous nos systèmes sont supposés finis et en particulier discrets ; chacun de nos exemples doit être regardé comme un jouet dont la seule fonction est d'illustrer un propos particulier mais qui n'a aucune prétention au réalisme : nous ne dirons rien par exemple du déménagement d'un piano ou de l'organisation d'un chantier de construction. De même, nous supposerons que tout fonctionne à merveille : l'éventualité d'un échec ou d'un incident d'exécution, dû par exemple à une intervention extérieure, est une donnée importante de la planification réaliste mais nous ne l'envisagerons pas ici.

1.1. La planification en général

Quelle que soit la méthode par laquelle on se propose de traiter la question, une première étape dans l'analyse d'une action complexe est le choix d'un langage capable de caractériser le système sur lequel on souhaite agir. Ce langage devra pouvoir désigner chacun des éléments constitutifs du système et décrire les propriétés que ces éléments peuvent posséder ainsi que les relations mutuelles que ces éléments peuvent entretenir. Supposons fixé un tel langage.

La description d'un état particulier E du système est l'inventaire des propriétés effectivement vérifiées et des relations effectivement entretenues par les éléments du système lorsque celui-ci est effectivement dans l'état E. Les objets de cet inventaire seront appelés *les ressources* de E.

Une action D appliquée à l'état E peut alors se décrire comme une transformation qui aboutit à un autre état E' du même système. Lors de l'exécution de D, certaines ressources de E sont *consommées* et de nouvelles ressources sont *produites* pour participer au nouvel état E' ; par contre les autres ressources de E restent intactes et doivent se retrouver dans E'. Ce phénomène de changement d'état doit être pris en compte très soigneusement.

De même qu'il était naturel de décomposer un système en ses éléments constitutifs, il est naturel de déterminer un ensemble d'*actions élémentaires*, à partir desquelles il est possible de déduire toutes les autres.

Ce travail de spécification est du ressort d'un expert du domaine considéré. En particulier, l'épithète "élémentaire" souvent utilisée signifie que l'expert n'a pas voulu, ou n'a pas pu pousser son analyse plus loin ; les éléments du système et les actions élémentaires nous sont donc opaques et, par exemple, il pourra être difficile de comparer les ressources consommées aux ressources produites par l'exécution d'une action élémentaire.

1.2. La planification déductive

Dans cet article, nous nous intéressons à un mode de traitement logique, ou plus précisément, déductif, et nous supposons donc que notre expert s'est exprimé dans un langage des prédicats.

Confondant le problème initial et la spécification qui nous en est donnée, nous qualifions tout ce qui s'y rattache de "concret" par opposition avec les objets logiques et déductifs qui sont introduits par la suite que nous qualifierons de "formels". Comme il se doit, le domaine d'interprétation des objets formels est constitué des objets concrets et le résultat principal qui sera exposé est l'expression d'une bonne relation entre le problème concret et sa représentation formelle (Proposition 1).

1.2.1. Des ressources

La difficulté essentielle à résoudre, déjà évoquée et dont nous reparlerons souvent, concerne la gestion des ressources. Considérons une action D transformant l'état E en l'état E' : une ressource consommée par D doit se trouver initialement dans E mais peut ne pas se retrouver dans E' ; de même D peut produire des ressources entièrement nouvelles. Or la logique, disons traditionnelle, est conçue pour exprimer des vérités éternelles : l'utilisation réitérée d'un théorème, n'appauvrit pas les ressources de celui-ci ; par contre, si vous disposez d'une pièce de 10 francs et que vous la jetez dans la manche d'un mendiant, elle ne sera plus jamais dans votre poche. Une logique qui vaut pour le théorème ne vaut plus pour la pièce.

Familles de ressources

Nous avons dit qu'un état est un inventaire de ressources $E = R_1, \dots, R_m$ mais, il est clair que E peut comporter des répétitions : imaginez que votre poche contienne une certaine somme d'argent sous la forme de pièces de 10 francs ; il est certainement souhaitable de représenter "la" pièce de 10 francs par un prédicat P signifiant "ceci est une pièce de 10 francs", mais il sera tout aussi souhaitable de faire figurer dans votre inventaire le nombre d'occurrences de P adapté à la somme que vous possédez. Ecrire E comme un ensemble conduirait à identifier toutes les pièces et donc à réduire la somme à 10 francs.

Une méthode bien connue pour rendre compte de l'éventualité de répétitions consiste à indexer les constituants de E , comme nous l'avons fait plus haut. Il y a cependant une petite nuance à observer : il ne faut pas considérer que l'indexation se fait par les entiers $1 < \dots < m$ dans cet ordre, mais simplement par l'ensemble non ordonné $\{1, \dots, m\}$. En effet, si un tel ordre devait jouer un rôle dans la description d'un état d'un système, il y figurerait sous la forme de ressources explicites !

L'indexation par des ensembles est la solution la plus satisfaisante : nous utilisons l'expression de famille dans ce sens et une notation sans parenthèse comme ci-dessus. Les quelques précisions qui sont données en annexe sont utiles pour l'exacte compréhension de l'article : elles sont élémentaires mais un peu didactiques et leur lecture n'est certainement pas à faire dans un premier temps. Nous n'en reparlerons pas dans l'article pour ne pas indisposer les lecteurs trop sensibles mais les autres pourront s'y référer à partir de la section 2.1.1. Si vous ne souhaitez pas lire l'annexe, vous pourrez entendre l'expression de "famille" dans le sens qui vous conviendra le mieux, par exemple, celui de l'une des expressions plus courantes *ensemble d'occurrences* ou *multi-ensemble*, vous ferez de même pour les notions qui s'y rattachent, par exemple, pour les ensembles d'occurrences : ensemble vide ($\emptyset$), inclusion ($\subseteq$), union ($\cup$), intersection ($\cap$), différence ($X-Y$ est l'ensemble des éléments de X qui n'appartiennent pas à Y) et somme ($X+Y$ est l'union disjointe de X et Y).

Le temps et la durée

Le temps est une ressource précieuse mais d'une nature très spéciale : on peut la partager. Ce partage peut être exploité avantageusement comme chacun sait. Si on introduit un paramètre "date" pour tenir compte de la durée d'exécution d'une action, on est obligé de mettre à jour la date de chacun des constituants du système, même

pour ceux qui n'ont visiblement rien à voir avec l'action en cours : le fait qu'un cube du jeu dont nous allons bientôt parler ne soit pas manipulé ne l'empêche pas de vieillir ! ceci est connu sous le nom de *problème du cadre* (appelé aussi *problème du décor*).

L'une des vertus du traitement que nous exposons dans cet article est, au prix d'un petit sacrifice, d'éviter ce grave problème : *ceci veut dire que le problème du cadre est résolu par le traitement qui est exposé ici* (pour les problèmes traités). La victime en est la durée, c'est-à-dire, la mesure de l'écoulement du temps. Par contre, comme nous le verrons au chapitre 3 (proposition 2), le système déductif qui est adopté ici prend tout naturellement en compte le temps, disons plus simplement l'ordre chronologique : la tâche de compléter cette information par la connaissance des durées mises en cause dans son système reviendra à l'utilisateur.

1.3. Le contenu de cet article

L'objet principal de cet article est l'application de la logique linéaire à l'étude d'un cas particulier de planification déterministe : la logique linéaire a été présentée dès le début comme *une logique consciente des ressources*, il était donc naturel de tenter son application aux problèmes de planification.

Le chapitre 2 est une introduction aux problèmes de planification qui nous intéressent. Les actions y sont présentées concrètement, mais dans la perspective d'une représentation formelle ultérieure.

Le chapitre 3 introduit alors, d'une façon qui nous semble très naturelle le système déductif adapté à leur étude : c'est un petit fragment du calcul des séquents linéaire de Girard. Toute action concrète est représentable par une preuve et, réciproquement, toute preuve est interprétable par une action concrète. C'est dans cette réciproque que la "linéarité" de la logique est essentielle. Nous profitons de la simplicité du fragment introduit dans ce chapitre pour en faire une étude un peu détaillée, destinée à familiariser le lecteur avec son fonctionnement et à le sensibiliser à ses propriétés générales.

Le chapitre 4 présente l'intégralité du calcul des séquents de Girard à votre curiosité : plus précisément, le calcul dont il sera question est l'analogue linéaire de la logique intuitionniste et non pas de la logique classique. Il faut bien insister sur le fait que notre traitement d'un cas déterministe ne prend tout son intérêt que dans la mesure où il s'insère dans un système général parfaitement cohérent. Nous ne pourrions pas entrer dans les détails mais nous espérons qu'un lecteur ayant saisi le caractère exemplaire du fragment étudié dans le chapitre 3 abordera ce système sans trop d'appréhension, saura y trouver son profit et, pourquoi pas, de l'agrément.

Nous ne tirerons pas d'autre conclusion et ne nous livrerons à aucune comparaison avec d'autres méthodes : l'objet de l'article est bien une application de la logique linéaire et non pas une étude panoramique de la planification ...

2. Planification déterministe

Voici deux exemples exprimés en termes généraux : ils sont assez simples pour que tout un chacun soit expert dans ces domaines.

Un jeu de cubes

Ce jeu est très connu des spécialistes de la planification. On considère des cubes posés en piles sur une table ; à l'aide d'une seule main qui ne peut tenir qu'un seul cube à la fois, on peut saisir l'un de ceux qui se trouvent en haut d'une pile pour le poser sur une autre ou directement sur la table (ce qui donne l'élément de base d'une nouvelle pile). La figure 1 montre un état possible du système que constitue ce jeu (c'est un robot qui va jouer !).

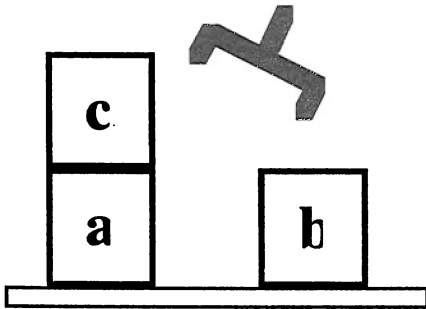


Figure 1.

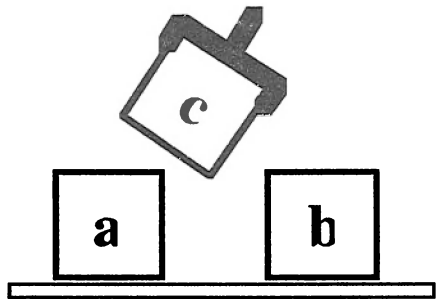


Figure 2.

Le cube **c** posé sur **a**, est en haut de sa pile ; la main du robot est vide. Le robot peut donc saisir **c** : lorsqu'il l'a fait, **c** n'est plus sur **a** qui est maintenant en haut de sa pile ; la main du robot contient **c** et n'est donc plus vide. La figure 2 représente l'état dans lequel notre système se trouve après l'exécution de cette action.

Un système de réécriture.

On considère deux types de suites finies d'objets de base désignés par **a**, **b** et **c** :

- 1) les suites non vides entre parenthèses, par exemple : (**a b a**)
- 2) les suites de longueur au plus un, entre crochets, par exemple : [**c**] et [].

Les états du système que nous voulons définir et que nous appellerons "expressions", sont des familles de suites des deux types précédents ; nous noterons ces expressions par simple juxtaposition (dans un ordre quelconque), par exemple : (**a c**)(**b**)[] et (**a c**)[](**b**) ne sont que deux façons différentes de représenter la même

expression, par contre $(a\ c)(b)(a\ c)(b)[\][\]$ désigne une expression différente de la précédente.

Les modifications autorisées sont décrites par les règles obtenues en substituant, dans les schémas suivants

$$\mathbf{pr}(x) : (x)[\] \Rightarrow [x]$$

$$\mathbf{po}(x) : [x] \Rightarrow (x)[\]$$

$$\mathbf{ot}(x,y) : y\ x[\] \Rightarrow y)[x]$$

$$\mathbf{me}(x,y) : y)[x] \Rightarrow y\ x)[\]$$

des objets de base aux indéterminées : une modification est le remplacement d'un motif s'identifiant au premier membre d'une règle par le deuxième membre de cette même règle. Par exemple, on peut appliquer **ot**(c, a) à l'expression $(a\ c)(b)[\]$ et cette opération produit l'expression $(a)(b)[c]$.

Vous aurez reconnu un système de réécriture^{*} dans ce deuxième exemple.

2.1. Planification déterministe "concrète"

Pour spécifier un problème de planification nous supposons que nous disposons des données suivantes :

- un ensemble de constantes d'individus et de symboles fonctionnels,
- un ensemble de prédicats.

Aucun symbole fonctionnel n'intervient dans nos exemples, aussi utiliserons nous toujours le mot constante à la place du mot terme, mais ce n'est pas un fait essentiel du point de vue qui nous intéresse ici.

Définitions et remarques

- Les constantes et les prédicats dont il est question ici sont réputés concrets, par opposition aux constantes et aux prédicats formels qui seront introduits par la suite. Cependant, afin de ne pas multiplier les notations, nous utiliserons les mêmes symboles dans les deux circonstances : le contexte permettra toujours de faire la distinction.

- Un prédicat P à n places est une relation n-aire, qu'il sera commode d'écrire sous la forme $P(x_1, \dots, x_n)$ où les x_i sont des indéterminées.

Un énoncé atomique clos est obtenu en instanciant les indéterminées d'un prédicat c'est-à-dire en substituant une constante à chacune de ses indéterminées :

* dans un sens un peu généralisé puisque l'on considère ici des ensembles d'occurrences de suites et pas seulement des suites.

- pour $n = 0$, P est clos par nature et exprime une propriété qui peut ne dépendre d'aucun objet ou au contraire, une propriété exprimée "réflexivement" par l'un des constituants du système ;

- pour $n = 1$, $P(a_1)$ exprime une propriété de a_1 ;
- pour $n > 1$, $P(a_1, \dots, a_n)$ exprime une relation entre $a_1, \dots, a_n$.

Ceci posé, les notions de base sont les suivantes :

- Une ressource est un énoncé atomique clos.
- Un état est une famille de ressources.
- Une action élémentaire est un couple (I, F) de familles de ressources : on la représentera sous la forme $I \Rightarrow F$.
- Une action élémentaire $I \Rightarrow F$ peut se lire "I devient F" comme il est précisé par la suite ; I et F ne sont pas nécessairement disjointes.
- Dans la pratique, il peut être commode d'écrire les actions élémentaires sous forme schématique. Un schéma, souvent appelé "règle de transition" s'exprime par le moyen d'énoncés non clos : on obtient une action élémentaire en substituant des constantes aux indéterminées d'une règle de transition. Illustrons les définitions sur nos exemples.

Le jeu de cubes

Les constantes a, b, c sont les noms de nos cubes ; les prédicats peuvent être les suivants :

- $B(x)$: x est en bas de sa pile
- $H(x)$: x est en haut de sa pile
- $S(x, y)$: x est posé sur y
- $T(x)$: x est tenu par le robot
- V : la main du robot est vide.

L'état du système représenté par la figure 1 est $E = B(a), S(c, a), H(c), B(b), H(b), V$

Les règles de transition suivantes permettent de décrire les actions élémentaires évoquées dans l'énoncé de notre exemple :

- Prendre** (x) : $V, H(x), B(x) \Rightarrow T(x)$
- Poser** (x) : $T(x) \Rightarrow V, H(x), B(x)$
- Oter** (x, y) : $V, H(x), S(x, y) \Rightarrow T(x), H(y)$
- Mettre** (x, y) : $T(x), H(y) \Rightarrow V, H(x), S(x, y)$

Le système de réécriture

Le langage précédent est encore parfaitement adapté à cet exemple, chacun des prédicats représentant la configuration indiquée entre accolades :

$$B(x) \{ \{x\}, H(x) \{x\}, S(x,y) \{y \ x\}, T(x) \{[x]\}, V \{[]\}.$$

L'état correspondant à l'expression

$$X = (a \ c)(b)(a \ c)(b)[[]]$$

s'écrit :

$$E = B(a), B(a), S(c, a), S(c, a), H(c), H(c), B(b), B(b), H(b), H(b), V, V.$$

X et E sont deux familles, mais les membres de E sont plus simples que ceux de X et, par exemple, la sous famille $I = S(c, a), S(c, a)$ de E est difficile à traduire dans X.

Les règles données pour les cubes traduisent aussi nos schémas de réécriture.

Remarques

- Un état est censé décrire complètement le système à l'instant considéré. Remarquez bien le caractère purement affirmatif de cette description : aucune négation n'est utilisée dans le cours de cet article ! Bien entendu, on peut déjouer une telle contrainte en introduisant des prédicats à signification négative : dans un jeu à trois cubes **a**, **b**, **c** et une seule main, les ressources **V**, **T(a)**, **T(b)** et **T(c)** sont mutuellement incompatibles et la présence de l'une d'elles dans un état particulier du système implique l'impossibilité de la présence des autres.
- Dans le deuxième exemple, les constantes **a**, **b** et **c** interviennent aussi par leurs occurrences ! Tout ce que nous avons dit pour les prédicats devrait donc être redit ici. Même dans le petit monde des cubes, ce phénomène peut se présenter : imaginez qu'une constante ne désigne plus un cube particulier mais sa couleur ; un beau jeu pourra comporter deux cubes verts, etc... et notre joueur pourra enfin se servir efficacement de ses deux mains.

2.1.1. Systèmes d'action déterministes

Avant de définir précisément la notion de système que nous adopterons, souvenons nous que l'exécution d'une action concrète $D : I \Rightarrow F$ consiste à consommer les ressources de **I**, à produire celles de **F**, mais à n'altérer aucune des autres. Vous n'aurez pas de mal à reconnaître dans ce qui suit la description d'un système de réécriture dont les expressions sont des familles de ressources : ceci est typique des systèmes que nous étudions dans ce chapitre (nous aurons l'occasion d'en reparler à la fin du chapitre 3). Soit E un état :

Exécutabilité : pour que **D** soit exécutable sur E, il est nécessaire que les ressources que **D** doit consommer soient disponibles dans E c'est-à-dire que $I \subseteq E$: nous supposons que cette condition est aussi suffisante.

Après consommation, on peut imaginer le système dans l'état transitoire $E - I$ puisque :

- les ressources consommées ont disparu,
- les autres ressources sont restées intactes.

Exécution : D produit alors les ressources de F qui viennent s'ajouter à celles de l'état transitoire : $E - I$ et F sont donc nécessairement disjoints et l'état qui résulte de cette adjonction est $E' = (E - I) \cup F$. Ainsi, lorsqu'une action est exécutable, son exécution est déterminée de façon unique : ceci justifie l'épithète "déterministe" qui est employée dans le cas présent (nous ferons une classification plus précise des actions dans le dernier chapitre).

Enfin, il faut se rappeler qu'une action modifie l'état d'un système, mais pas le système lui-même : l'emploi d'un même langage pour décrire les états et les actions assure une certaine cohérence, mais ce n'est pas suffisant.

Définition

Un ensemble Σ d'états et un ensemble S d'actions élémentaires, écrits dans le même langage, sont dits compatibles ssi pour toute $E \in \Sigma$ et toute $D : I \Rightarrow F$ de S tels que $I \subseteq E$, on a : $E \cap F \subseteq I$ et $(E - I) \cup F \in \Sigma$.

Nous appelons *système d'action déterministe* un couple (Σ, S) formé d'un ensemble d'états et d'un ensemble d'actions élémentaires, compatibles au sens de la définition précédente.

En général, S contient toutes les actions élémentaires que l'on peut obtenir par instanciation à partir d'un ensemble de règles de transition et de toutes les constantes du langage.

Définition

Dans un système, une action élémentaire $D : I \Rightarrow F$ est exécutable sur l'état E ssi $I \subseteq E$: dans ce cas, l'exécution de D transforme E en $E' = (E - I) \cup F$.

Nous noterons l'exécution de D sur E par $D : E \Rightarrow E'$.

Commentaires

- Par la suite, nous nous placerons toujours (plus ou moins explicitement) dans un système. La compatibilité d'un ensemble d'états et d'un ensemble d'actions élémentaires est un problème théorique très délicat qui ne sera pas abordé. Citons tout de même l'une des questions intéressantes liées à ce problème. Soit S un ensemble d'actions élémentaires et E un état réputé initial ; l'ensemble des états accessibles à partir de E est l'ensemble $\Sigma = S^*(E)$ des états obtenus de proche en proche à partir de E par l'exécution des actions de S : (Σ, S) est un système d'action. La question qui s'impose alors est au cœur de la planification : existe-t-il un état $E' \in \Sigma$ satisfaisant des contraintes données (présence de certaines ressources, absence de certaines autres ressources, ...) ?

Notre article présente des outils qui sont utiles pour répondre à cette question mais il ne l'abordera pas.

• *Actions élémentaires concurrentes.* Soient $\mathbf{D} : I \Rightarrow F$ et $\mathbf{E} : J \Rightarrow G$ deux actions élémentaires vérifiant $I \cap J = \emptyset$. Si elles sont exécutables sur le même état E , c'est-à-dire si $I \cup J \subseteq E$, les définitions de l'exécution impliquent que $F \cap G = F \cap G = I \cap G = \emptyset$ (I, J, F et G sont donc disjointes deux à deux). En fait, cette propriété exprime une condition nécessaire de compatibilité entre S et Σ .

• *STRIPS*, le premier traitement effectif des actions déterministes, et qui reste la grande référence en matière de planification, introduit des notions qui pour $\mathbf{D} : I \Rightarrow F$ sont à peu près les suivantes :

- $\text{Pre}(\mathbf{D}) = I$
- $\text{Del}(\mathbf{D}) = I - \text{Inv}(\mathbf{D})$
- $\text{Add}(\mathbf{D}) = F - \text{Inv}(\mathbf{D})$

où $\text{Inv}(\mathbf{D}) = I \cap F$ est la famille des ressources mises en cause dans l'exécution de $\mathbf{D}$ mais qui sont restées invariantes, et donne la définition :

$\mathbf{D}$ est exécutable sur un état E ssi $\text{Pre}(\mathbf{D}) \subseteq E$ (vérification des "préconditions")
l'exécution de $\mathbf{D}$ consiste alors à effacer $\text{Del}(\mathbf{D})$ de E puis à y ajouter $\text{Add}(\mathbf{D})$.

Dans la pratique $\text{Inv}(\mathbf{D})$ est évidemment très significative ; par contre, elle ne jouera pas de rôle dans la formalisation que nous présentons. En fait, ce qui se passe pendant l'exécution d'une action élémentaire ne nous est pas connu et il ne serait donc pas paradoxal de supposer $\text{Inv}(\mathbf{D}) = \emptyset$ pour toute action élémentaire.

2.2. Définition des actions concrètes d'un système

Les actions concrètes sont obtenues par "composition" d'actions élémentaires. Notons encore $\mathbf{D} : E \Rightarrow E'$ l'exécution d'une telle action sur un état E :

- on devra pouvoir disposer des actions $\mathbf{D} : E \Rightarrow E'$ qui laissent toutes les ressources intactes ;
- lorsque $\mathbf{D} : E \Rightarrow E'$ et $\mathbf{E} : E' \Rightarrow E''$, on devra pouvoir disposer de l'action composée $\mathbf{D} \circ \mathbf{E} : E \Rightarrow E''$ résultant de l'exécution successive de $\mathbf{D}$ et de $\mathbf{E}$ (nous écrivons la composition dans l'ordre qui nous semble le plus naturel ici) ;
- lorsque $\mathbf{D} : E \Rightarrow F$ et $\mathbf{E} : E \Rightarrow G$ consomment des familles disjointes de ressources, on devra pouvoir disposer d'une action $\mathbf{D} \parallel \mathbf{E} : E \Rightarrow E'$ résultant de l'exécution simultanée de $\mathbf{D}$ et $\mathbf{E}$.

Ces exigences semblent mettre tout l'état E en cause, alors que seules certaines de ses ressources sont concernées. C'est pourquoi il est nécessaire de donner les définitions suivantes, indépendantes de tout état : les actions intéressantes sont celles qui sont exécutables sur un état du système.

Construction des actions concrètes d'un système

Les actions concrètes sont construites inductivement par les clauses suivantes :

- *L'action identique* $\text{Id} : \emptyset \Rightarrow \emptyset$ est une action concrète.
- *L'action triviale* : pour chaque ressource A , $\text{T}_A : A \Rightarrow A$ est une action concrète.
- *Les actions élémentaires* sont des actions concrètes.
- *Composition (séquentielle)* : Deux actions concrètes $\text{D} : I \Rightarrow F$ et $\text{E} : J \Rightarrow G$ sont dites composables si et seulement si $J = F$; lorsque cette condition est vérifiée, la composée $\text{DoE} : I \Rightarrow G$ est une action concrète.
- *Composition en concurrence* : Deux actions concrètes $\text{D} : I \Rightarrow F$ et $\text{E} : J \Rightarrow G$ sont dites concurrentes ssi $I \cap J = \emptyset$; lorsque cette condition est vérifiée, la composée en concurrence $\text{D} \parallel \text{E} : I \cup J \Rightarrow F + G$ (où $+$ est l'union disjointe) est une action concrète.
- *Il n'y a pas d'autre action concrète* : autrement dit, l'action identique, les actions triviales et les actions élémentaires sont des actions concrètes et, toute autre action concrète s'obtient par composition, séquentielle ou en concurrence, d'actions concrètes convenables déjà connues.

Remarque fondamentale

Ce qui nous importe dans une action concrète $\text{D} : I \Rightarrow F$ n'est pas seulement la connaissance de I et de F mais la façon dont elle a été construite. D'un autre côté, deux constructions différentes peuvent produire des actions équivalentes. La définition de l'équivalence entre deux actions n'est donnée qu'à la fin du chapitre 3 mais nous donnerons des indications dans ce sens au fur et à mesure qu'elles se présenteront. Pour l'instant, on peut déjà signaler la compatibilité de l'équivalence avec les compositions ainsi que l'associativité des compositions, c'est-à-dire :

- $(\text{DoE}) \circ \text{F}$ et $\text{Do}(\text{E} \circ \text{F})$ sont équivalentes ;
- $(\text{D} \parallel \text{E}) \parallel \text{F}$ et $\text{D} \parallel (\text{E} \parallel \text{F})$ sont équivalentes.

Propriété

Soit (Σ, S) un système. Pour tout état $E \in \Sigma$ et toute action concrète $\text{D} : I \Rightarrow F$ vérifiant $I \subseteq E$ on a $E \cap F \subseteq I$ et $(E - I) \cup F \in \Sigma$.

Cette propriété permet d'étendre à toutes les actions concrètes les notions d'exécutabilité et d'exécution :

- D est exécutable sur E ssi $I \subseteq E$;
- l'exécution de D transforme E en $E' = (E - I) \cup F$.

Nous écrirons encore $\text{D} : E \Rightarrow E'$ pour signifier que D est exécutable sur E et transforme celui-ci en E' .

La démonstration de cette propriété se fait par induction : elle est un peu fastidieuse mais sans surprise et nous la laissons aux courageux. Voici cependant quelques observations et remarques utiles.

- **Id** : $\emptyset \Rightarrow \emptyset$ est exécutable sur n'importe quel état. C'est une merveille d'inaction, analogue au mouvement de vitesse nulle qui s'exécute de la façon suivante **Id** : $E \Rightarrow E$.

- De même, **T_A** : $A \Rightarrow A$ est exécutable sur tout état E disposant de la ressource A et l'on a alors **T_A** : $E \Rightarrow E$.

- Plus généralement, une action concrète construite sans utiliser aucune action élémentaire est de la forme : **T_X** : $X \Rightarrow X$ où $X = A_n, \dots, A_n$ est une famille de ressources, et est exécutable sur tout état pour lequel $X \subseteq E$ par **T_X** : $E \Rightarrow E$. Ces actions seront aussi qualifiées de *triviales*. Les ressources de X sont ici explicitement et intentionnellement laissées intactes par l'exécution de **T_X** qui, en somme, se contente de contrôler la disponibilité dans E des ressources de X. Ceci est à distinguer d'une autre forme d'invariance : une action concrète peut fort bien agir de la façon suivante **D** : $E \Rightarrow E$ sans pour cela être équivalente à une **T_X** ; par exemple, dans un jeu de cubes, **D** = **Oter**(c,a) o **Mettre**(c,a) est dans ce cas. Ainsi, tourner en rond ne signifie-t-il pas que l'on reste immobile* : l'action **D** n'est en fait équivalente à aucune action triviale. Plus généralement, une action concrète dans la construction de laquelle intervient au moins une action élémentaire n'est équivalente à aucune action triviale.

- Pour une famille donnée X de ressources, éventuellement vide, toutes les actions triviales de la forme **T** : $X \Rightarrow X$ sont équivalentes. A une équivalence près, une action triviale peut se construire sans le concours de la composition séquentielle. En effet, si l'on pose génériquement $X_n = A_n, \dots, A_n$, la récurrence suivante :

$$- \mathbf{T}_{X_0} = \mathbf{T}_{\emptyset} = \mathbf{Id} ;$$

$$- \mathbf{T}_{X_{n+1}} = \mathbf{T}_{X_n} \parallel \mathbf{T}_{A_{n+1}} ;$$

permet de construire toute action triviale.

- Les actions triviales jouent un rôle dans la construction des actions concrètes.

Encadrement : Soit **D** : $I \Rightarrow F$ une action concrète et X une famille de ressources vérifiant $I \cap X = \emptyset$, alors l'encadrement de **D** par X est **D** || **T_X** : $I \cup X \Rightarrow F + X$. Cette opération étend le contrôle de la disponibilité des ressources mais ne modifie pas l'action elle-même.

Composition : Considérons deux actions concrètes **D** : $I \Rightarrow F$ et **E** : $J \Rightarrow G$. Lorsqu'elles sont composables, c'est-à-dire lorsque $J = F$, alors si **D** : $E \Rightarrow E'$ on a **E** : $E' \Rightarrow E''$ et **DoE** : $E \Rightarrow E''$. Il est bien naturel de considérer que **D** et **E** sont composables lorsqu'il existe un état E tel que **D** : $E \Rightarrow E'$ et **E** : $E' \Rightarrow E''$ mais, la condition $J = F$ n'a alors aucune raison d'être satisfaite : en fait, il est

* je veux dire qu'une montre arrêtée ne rend pas les mêmes services qu'une autre, qui fonctionne !

possible de construire deux actions composables D' et E' par encadrement de D et E . En effet E est exécutable sur $E' = (E-I) \cup F$ on a $J \subseteq E'$ et donc $I \cap J \subseteq F$.

Cette condition est suffisante pour définir de bons encadrements : si elle est satisfaite, les familles $X = I \cap F$, $Y = J - X$ et $Z = F - X$ sont disjointes, $J = X \cup Y$, $F = X \cup Z$, $I \cap Y = \emptyset$ et $J \cap Z = \emptyset$ et on peut effectuer les encadrements :

$$D' = D \parallel T_Y: I \cup Y \Rightarrow X \cup Y \cup Z, \quad E' = E \parallel T_Z: X \cup Y \cup Z \Rightarrow G + Z.$$

On peut maintenant poser : $D \circ E = D' \circ E': I \cup Y \Rightarrow G + Z$.

Nous utiliserons librement cette notion généralisée qui est naturelle au point que c'est à elle que nous avons pensé avant de définir les actions concrètes.

- Si D et E sont concurrentes et exécutables sur le même état alors I , J , F et G sont disjointes deux à deux, comme dans le cas des actions élémentaires : on peut donc, lorsque $D \parallel E$ est exécutable, remplacer $+$ par $\cup$ dans sa définition (ceci signifie que l'on n'est pas obligé de réindexer les familles avant d'en faire l'union) ; cette remarque vaut aussi pour la composition généralisée.

- La composition en concurrence ne construit pas d'action concrète exécutable qui ne soit constructible par encadrement et composition séquentielle :

- Si $D \circ E$ est exécutable sur E et si $F \cap J = \emptyset$ alors D et E sont concurrentes et exécutables sur E .

- Réciproquement, si D et E sont concurrentes et exécutables sur un même E alors $D \parallel E$ est définie et exécutable ; de plus :

- 1) $I \cap G = F \cap J = \emptyset$,

- 2) $E \circ D$ et $D \circ E$ sont définies et transforment un état de la même manière que $D \parallel E$.

- Dans les circonstances précédentes, les trois actions concrètes sont considérées comme étant équivalentes. Sur cet exemple il est clair que les modes d'exécution de deux actions équivalentes peuvent différer sensiblement : la composition séquentielle impose des exécutions successives alors que la composition en concurrence permet des exécutions simultanées ; le choix du mode d'exécution est affaire de besoins et de moyens mais, la composition en concurrence est plus performante car elle utilise un partage du temps.

Un exemple

Considérons l'action concrète représentée par la figure 3 : nous avons mis des accents pour distinguer et suivre différentes occurrences mais ceci va à l'encontre du fait que notre expert a choisi une seule constante pour désigner éventuellement plusieurs objets (qu'il ne peut pas ou ne veut pas distinguer). Il s'agit de la composition séquentielle des actions élémentaires $1 \circ \dots \circ 8$. En y regardant de plus près, on voit que

- 1 et 2 sont concurrentes et donc permutable ;

- 2 et $1 \circ 3$ sont concurrentes ;

la composée $1 \circ 2 \circ 3$ est donc équivalente à $D = (1 \circ 3) \parallel 2$. De même $4 \circ \dots \circ 8$ est équivalente à $E = (5 \circ 7) \parallel (4 \circ 6 \circ 8)$.

Finalement $1o \dots o8$ est équivalente à **DoE**.

$$E_0 = (a\ c)(b)(a'\ c')(b')[[\]']$$

$$\begin{aligned} \Rightarrow & (a)(b)(a'\ c')(b')[c][\]' & 1 : ot(c, a) \\ \Rightarrow & (a)(b)(a')(b')[c][c']' & 2 : ot(c', a') \\ \Rightarrow & (a)(b)(a')(b')(c)[\][c']' & 3 : po(c) \\ \Rightarrow & (a)(b)(a')(b')(c\ c')[\][\]' & 4 : me(c', c) \\ \Rightarrow & (a)(b)(b')(c\ c')[\ a'][\]' & 5 : pr(a') \\ \Rightarrow & (a)(b)(c\ c')[\ a'][\ b']' & 6 : pr(b') \\ \Rightarrow & (a\ a')(b)(c\ c')[\][\ b']' & 7 : me(a', a) \\ \Rightarrow & (a\ a')(b\ b')(c\ c')[\][\]' & 8 : me(b', b) \\ = & E_g \end{aligned}$$

Figure 3.

3. Planification déductive

Dans cette section, nous formalisons ce qui précède en termes logiques et traduisons les actions concrètes en preuves. Pour simplifier les notations, nous avons convenu d'utiliser comme langage formel celui qui a déjà servi à faire la description concrète : la distinction sera cependant toujours possible car le contexte n'est jamais ambigu. Nous parlerons de la *représentation* d'un objet concret par un objet formel et de l'*interprétation* d'un objet formel par un objet concret.

La représentation traditionnelle, c'est-à-dire immédiate, d'une action concrète

$$A_1, \dots, A_m \Rightarrow B_1, \dots, B_n$$

est une preuve de l'implication

$$A_1 \wedge \dots \wedge A_m \rightarrow B_1 \wedge \dots \wedge B_n$$

dans la théorie **T** dont les axiomes propres sont les représentations des actions élémentaires du système étudié. Les questions qui se posent sont les suivantes :

- 1) Toute action concrète est elle représentable dans **T**?
- 2) Réciproquement, une preuve de $A_1 \wedge \dots \wedge A_m \rightarrow B_1 \wedge \dots \wedge B_n$ est elle interprétable par une action concrète $A_1, \dots, A_m \Rightarrow B_1, \dots, B_n$?

• Notre première tentative, basée sur une logique traditionnelle (à savoir la logique intuitionniste) débouche sur une impasse : la réponse à la première question est bien entendu "oui" mais pour la deuxième, de loin la plus importante, la réponse est "NON".

- Les qualités propres au calcul des séquents permettent de repérer où git le problème. La logique linéaire de Girard, dans sa version intuitionniste, devient alors une excellente candidate pour une deuxième tentative qui, elle, répond affirmativement aux deux questions.

Le calcul des séquents, introduit par G. Gentzen en 1934 est le système déductif le plus proche de la description des actions que nous avons faite et le meilleur outil théorique pour analyser les preuves. Le fait qu'il ne travaille pas seulement sur des formules mais aussi sur des formules décomposées en séquents (voir ci-dessous) lui donne un aspect très "massif" mais le rend plus souple et plus analytique que les systèmes déductifs couramment utilisés. En particulier, chaque opération logique est traitée isolément et pour elle-même, de façon symétrique (règles droites ou d'introduction, règles gauches ou d'élimination).

3.1. Tentative dans un cadre traditionnel

Les formules que nous considérons dans cette section sont des conjonctions de formules atomiques du langage, la conjonction vide étant représentée par la constante de prédicat $\mathbf{v}$ (identiquement vraie), c'est-à-dire :

- $\mathbf{v}$ est une formule ; toute formule atomique est une formule ;
- si A et B sont des formules alors $A \wedge B$ est une formule.

Un *séquent* intuitionniste est une expression symbolique $\Gamma \vdash B$ (lire " Γ thèse B ") dans laquelle Γ est une famille (finie) de formules et B est une formule.

Une formule est identifiée à une famille dont c'est le seul membre ; si Γ et Δ sont des familles de formules, Γ, Δ est une abréviation pour $\Gamma + \Delta$; enfin, la famille vide est signalée par l'absence de toute notation.

La signification intuitive de $A_1, \dots, A_m \vdash B$ est "si A_1 et... et A_m alors B " : autant dire que le signe $\vdash$ a un sens implicatif et que les virgules qui interviennent dans le membre de gauche d'un séquent ont un sens conjonctif. Si l'on voulait utiliser des séquents classiques, c'est-à-dire admettre une famille quelconque de formules dans le membre de droite

- dans le cas de plusieurs formules à droite, les virgules prendraient un sens disjonctif ;
 - l'absence de formule à droite prendrait le sens d'une proposition "absurde" c'est-à-dire identiquement fausse, correspondant à une impasse ;
- et ceci ne saurait convenir à notre problème.

Les axiomes et les règles du calcul des séquents qui nous intéressent ici sont les suivantes :

Groupe de l'identité

(axiome) $A \vdash A$ (pour A atomique)

(coupure)
$$\frac{\Gamma \vdash A \quad A, \Delta \vdash B}{\Gamma, \Delta \vdash B}$$

Groupe de la conjonction

($\wedge$ -d)
$$\frac{\Gamma \vdash A \quad \Delta \vdash B}{\Gamma, \Delta \vdash (A \wedge B)}$$

($\wedge$ -g)
$$\frac{A, B, \Gamma \vdash C}{(A \wedge B), \Gamma \vdash C}$$

($\vee$ -d)
$$\vdash \vee$$

($\vee$ -g)
$$\frac{\Gamma \vdash C}{\vee, \Gamma \vdash C}$$

Groupe structurel

(Affaiblissement)
$$\frac{\Gamma \vdash B}{A, \Gamma \vdash B}$$

(Contraction)
$$\frac{A, A, \Gamma \vdash B}{A, \Gamma \vdash B}$$

Dans ce calcul, une preuve d'un séquent S est un arbre dont les nœuds sont des séquents :

- les feuilles sont des axiomes (axiomes d'identité ou $\vee$ -d) ;
- la racine est S lui-même ;

et dont chaque embranchement est justifié par une règle du tableau précédent.

Plus précisément, si l'on note $\mathbf{D} :: S$ (lire " $\mathbf{D}$ montre S") le fait que $\mathbf{D}$ soit une preuve de S, la construction d'une preuve $\mathbf{D}$ et le calcul de sa hauteur $h(\mathbf{D})$ se font par l'induction suivante :

- Un axiome S (axiome d'identité ou $\vee$ -d) est une preuve $\mathbf{D} :: S$ et $h(\mathbf{D}) = 0$;

- si $E :: T$ et si S provient de T par application d'une règle unaire, alors

$$E :: T$$

$$S$$

est une preuve D de S et $h(D) = h(E) + 1$;

- si $E :: T$ et $F :: U$, et si S provient de T et U par application d'une règle binaire, alors

$$E :: T \quad F :: U$$

$$S$$

est une preuve D de S et $h(D) = h(E) + h(F) + 1$.

Dans chaque règle, nous avons souligné les *formules actives* c'est-à-dire celles qui sont vraiment en cause dans la règle en question (cette pratique ne sera pas poursuivie). La formule active dans la règle de coupure est appelée la formule de coupure. Il faut *a priori* bien se méfier d'une formule de coupure car elle disparaît sans laisser de trace alors qu'elle peut être beaucoup plus complexe que toutes les autres formules présentes dans la règle ; en fait, le théorème d'élimination des coupures qui est valable ici mais que nous ne détaillerons que dans la section suivante nous aidera à apaiser nos craintes.

3.1.1. Quelques exemples

Nous allons construire quelques preuves utiles : nous espérons que ces exemples permettront aussi aux lecteurs non familiers avec le calcul des séquents, de comprendre son fonctionnement.

" $A \vdash B$ est prouvable" est une relation réflexive et transitive.

- La réflexivité signifie que $A \vdash A$ est prouvable pour toute formule A . En effet,

- si A est une formule atomique, l'axiome d'identité pour A est une preuve de $A \vdash A$;

- si $A = v$, l'application de $v.g$ à l'axiome $v.d$

$$\vdash v$$

$$v \vdash v$$

est une preuve de $A \vdash A$;

- si $A = (B \wedge C)$ et si nous avons $E :: B \vdash B$ et $F :: C \vdash C$ alors

$$E :: B \vdash B \quad F :: C \vdash C$$

$$B, C \vdash (B \wedge C)$$

$$(B \wedge C) \vdash (B \wedge C)$$

est une preuve de $A \vdash A$.

Dans les preuves qui suivent, nous utiliserons librement les séquents de la forme $A \vdash A$ comme s'il s'agissait d'axiomes : il suffit de remplacer chacun de ces séquents par sa preuve, telle qu'elle vient d'être construite.

- La transitivité provient de l'application de la règle de coupure :

$$E :: A \vdash B \quad F :: B \vdash C$$

$$A \vdash C$$

Par la suite, lorsque nous évoquerons l'équivalence de deux formules A et B , nous voudrions dire que $A \vdash B$ et $B \vdash A$ sont prouvables tous les deux.

La conjonction $\wedge$ est commutative. Puisque le membre gauche d'un séquent est une famille de formules, la règle $\wedge$ -g pourrait s'écrire de bien des façons, par exemple :

$$A, B, \Gamma \vdash C$$

$$(B \wedge A), \Gamma \vdash C$$

et on peut donc considérer la preuve

$$A \vdash A \quad B \vdash B$$

$$A, B \vdash (A \wedge B)$$

$$(B \wedge A) \vdash (A \wedge B)$$

La conjonction $\wedge$ est associative.

Voici une preuve de $(A \wedge (B \wedge C)) \vdash ((A \wedge B) \wedge C)$:

$$\begin{array}{c}
 A \vdash A \quad B \vdash B \\
 \hline
 A, B \vdash (A \wedge B) \quad C \vdash C \\
 \hline
 A, B, C \vdash ((A \wedge B) \wedge C) \\
 \hline
 A, (B \wedge C) \vdash ((A \wedge B) \wedge C) \\
 \hline
 (A \wedge (B \wedge C)) \vdash ((A \wedge B) \wedge C)
 \end{array}$$

De même, $((A \wedge B) \wedge C) \vdash (A \wedge (B \wedge C))$ est prouvable.

Nous utiliserons librement cette propriété en oubliant d'écrire les parenthèses.

v est neutre pour $\wedge$. v représente la conjonction de la famille vide de formule, ce qui est bien le fait de l'élément neutre. La démonstration en est donnée par les preuves :

$$\begin{array}{c}
 A \vdash A \\
 \hline
 v, A \vdash A \\
 \hline
 v \wedge A \vdash A \\
 \vdash v \quad A \vdash A \\
 \hline
 A \vdash v \wedge A
 \end{array}$$

Chacun des séquents $A, B, \Gamma \vdash C$ et $A \wedge B, \Gamma \vdash C$ est prouvable à partir de l'autre. L'une des deux affirmations est simplement l'application de la règle $\wedge_g$; l'autre s'obtient en appliquant une coupure :

$$\begin{array}{c}
 A \vdash A \quad B \vdash B \\
 \hline
 A, B \vdash A \wedge B \quad A \wedge B, \Gamma \vdash C \\
 \hline
 A, B, \Gamma \vdash C
 \end{array}$$

La virgule joue donc un rôle "interchangeable" avec celui de la conjonction.

Rôle des règles structurelles. La propriété "si A et B alors A" est dans la nature de la conjonction traditionnelle et il faut donc que le séquent $A \wedge B \vdash A$ soit prouvable ; c'est le rôle de la règle d'affaiblissement de rendre cela possible :

$$\begin{array}{c} A \vdash A \\ \hline A, B \vdash A \\ \hline A \wedge B \vdash A \end{array}$$

De même la propriété "si A alors A et A", qui est naturelle à la conjonction traditionnelle, exige la prouvabilité de $A \vdash A \wedge A$; c'est le rôle de la règle de contraction de rendre cela possible :

$$\begin{array}{c} A \vdash A \quad A \vdash A \\ \hline A, A \vdash A \wedge A \\ \hline A \vdash A \wedge A \end{array}$$

3.1.2. Toute action concrète est représentable par une preuve

Ceci est la réponse positive à la première question. Rappelons que nous nous plaçons dans la théorie T qui admet le séquent $A_1, \dots, A_m \vdash B_1 \wedge \dots \wedge B_n$ ($A_1, \dots, A_m \vdash \vee$ si $n = 0$) comme axiome propre pour chacune des actions élémentaires $A_1, \dots, A_m \Rightarrow B_1, \dots, B_n$. Ceci veut dire que, pour la construction d'une preuve dans T, on admet les séquents précédents comme axiomes au même titre que les axiomes d'identité et v.d. La démonstration se fait par une induction qui suit point par point la construction des actions concrètes. Chaque action concrète **D** est représentable par la preuve **D** suivante :

L'action identique est représentée par l'axiome $\vee.d \vdash \vee$.

L'action triviale $\mathbf{T}_A : A \Rightarrow A$ est représentée par l'axiome d'identité correspondant $A \vdash A$.

Une action élémentaire est représentée par l'axiome propre correspondant.

La composition séquentielle. Soient $\mathbf{E} :: A_1, \dots, A_m \vdash B$ où $B = B_1 \wedge \dots \wedge B_n$ et $\mathbf{F} :: B_1, \dots, B_n \vdash C$ les représentations respectives de **E** et **F**; une application itérée de la règle $\wedge.g$ à partir de **F** produit une preuve $\mathbf{F}' :: B \vdash C$. **D** est alors obtenue par la coupure

$$\begin{array}{c} \mathbf{E} :: A_1, \dots, A_m \vdash B \quad \mathbf{F}' :: B \vdash C \\ \hline A_1, \dots, A_m \vdash C \end{array}$$

(le cas $n = 0$ utilise $\vee.g$ au lieu de $\wedge.g$).

La composition en concurrence est une application de la règle $\wedge.d$

$$\frac{\mathbf{E} :: A_1, \dots, A_m \vdash B \quad \mathbf{F} :: C_1, \dots, C_p \vdash D}{A_1, \dots, A_m, C_1, \dots, C_p \vdash B \wedge D}$$

Comme il se doit, les actions triviales, au sens large, sont représentées par des preuves dans lesquelles n'intervient aucun axiome propre.

3.1.3. Une preuve n'est pas toujours interprétable par une action concrète

La construction que nous venons de faire est sans problème et montre surtout que les séquents et leur calcul sont bien adaptés à la représentation des actions concrètes. Cependant, il faut remarquer le fait que nous n'y avons jamais utilisé les règles structurelles (affaiblissement et contraction) : lorsque nous sommes en présence d'une preuve, nous ne contrôlons pas l'usage qu'elle a fait de ces règles et, c'est peut être une catastrophe !

Supposons notre jeu de cubes dans l'état $E = B(a), S(c, a), H(c), B(b), H(b), V$.

- La preuve

$$\frac{\frac{V \vdash V \quad V \vdash V}{V, V \vdash V \wedge V}}{V \vdash V \wedge V}$$

qui utilise la règle de contraction, est parfaitement constituée mais, son interprétation

$$V \Rightarrow V, V$$

ressemble fort à un miracle de multiplication de ressources.

- La preuve

$$\frac{V \vdash V}{H(c), V \vdash V}$$

qui est une simple application de la règle d'affaiblissement est tout à fait correcte mais, son interprétation $H(c), V \Rightarrow V$ ne saurait être une action concrète car c'est la disparition en pure perte de la ressource $H(c)$: l'état du système serait

$$E' = B(a), S(c, a), B(b), H(b), V$$

après exécution de cette "mauvaise action". Bien que n'étant pas recouvert par un autre cube, c qui reste posé sur a n'est plus accessible :

- l'action élémentaire $Oter(c, a)$ n'est plus exécutable ;
- il ne sera plus jamais possible de poser quoi que ce soit sur c .

Remarque

On peut imaginer un système où une ressource, disons V , est duplicable : ceci se traduit par la présence explicite de l'action élémentaire $V \Rightarrow V, V$ dans la définition dudit système. De même, le fonctionnement d'un système peut fort bien comporter la possibilité de faire disparaître une ressource H lorsque qu'une autre ressource V est disponible : ceci se traduit par la présence explicite de l'action élémentaire $H, V \Rightarrow V$. Même pour ces systèmes, les preuves construites ci-dessus ne seraient pas admissibles car elles ne mettent en cause aucun axiome propre !

On peut résumer ces observations en disant que *les règles structurelles de la logique traditionnelle ne permettent pas de gérer les ressources correctement*.

3.2. Planification en logique linéaire

La caractéristique de la logique linéaire qui est essentielle à notre propos est qu'elle permet de considérer la contraction et l'affaiblissement comme des opérations logiques et non plus comme des opérations structurelles (cf. l'opérateur ! dans la dernière section) ; en ce qui nous concerne, ceci revient à exclure purement et simplement toute règle structurelle. La nature de la conjonction en est profondément bouleversée ; disons tout de suite que la "linéarisation" entraîne un scindage de la conjonction traditionnelle en deux conjonctions : l'une, dite multiplicative et notée $\otimes$, est la conjonction utile dans le présent chapitre ; l'autre, dite additive et notée $\&$, a des propriétés très différentes qui seront brièvement évoquées dans le dernier chapitre. Remarquez qu'en nous plaçant dans un cadre intuitionniste, nous avons déjà choisi d'interdire l'application de règles structurelles sur la partie droite de nos séquents puisque celle-ci est toujours constituée d'une unique formule.

Les formules linéaires que nous considérerons maintenant sont construites à partir des formules atomiques du langage et de la constante de prédicat **1**, par conjonction multiplicative, c'est-à-dire :

- **1** est une formule ;
- toute formule atomique est une formule ;
- si A et B sont des formules alors $(A \otimes B)$ est une formule.

Un séquent linéaire intuitionniste se définit comme précédemment, avec cette fois des formules linéaires. La signification intuitive de $A_1, \dots, A_m \vdash B$ est

" A_1 et ... et A_m deviennent B "

où "et" est à prendre dans le sens d'une juxtaposition. Les axiomes et les règles du calcul des séquents qui nous intéressent ici sont les suivantes :

Groupe de l'identité(axiome) $A \vdash A$ (pour A atomique)

$$\begin{array}{c}
 \Gamma \vdash A \quad A, \Delta \vdash B \\
 \text{(coupure)} \quad \hline
 \Gamma, \Delta \vdash B
 \end{array}$$

Groupe de la conjonction multiplicative

$$\begin{array}{ccc}
 \text{(\otimes\_d)} \quad \frac{\Gamma \vdash A \quad \Delta \vdash B}{\Gamma, \Delta \vdash (A \otimes B)} & & \frac{A, B, \Gamma \vdash C}{(A \otimes B), \Gamma \vdash C} \quad \text{(\otimes\_g)} \\
 \text{(1\_d)} \quad \vdash 1 & & \frac{\Gamma \vdash C}{1, \Gamma \vdash C} \quad \text{(1\_g)}
 \end{array}$$

*Notez bien l'absence de toute règle structurelle.**Définitions et propriétés*

- Les définitions d'une preuve et de sa hauteur s'adaptent sans problème du cas traditionnel.

- " $A \vdash B$ est prouvable" est une relation réflexive et transitive ; $\otimes$ est associative, commutative et admet **1** comme élément neutre : les preuves faites plus haut s'adaptent sans difficulté.

- Par contre, les séquents $A \otimes B \vdash A$ et $A \vdash A \otimes A$ ne sont pas prouvables en général. En effet, supposons A et B atomiques et tentons de prouver ces séquents sans aucune application de la règle de coupure :

- une telle preuve du séquent $A \otimes B \vdash A$ ne peut se terminer que par

$$\begin{array}{c}
 A, B \vdash A \\
 \hline
 A \otimes B \vdash A
 \end{array}$$

mais, aucune règle n'admet $A, B \vdash A$ comme conclusion lorsque A et B sont atomiques !

- de même, une preuve sans coupure de $A \vdash A \otimes A$ ne peut se terminer que par

$$\frac{A \vdash A \quad \vdash A}{A \vdash A \otimes A} \quad \text{ou par} \quad \frac{\vdash A \quad A \vdash A}{A \vdash A \otimes A}$$

or visiblement, $\vdash A$ n'admet aucune preuve sans coupure !

La démonstration de notre affirmation sera complète lorsque l'on aura vérifié la propriété suivante.

Théorème d'élimination des coupures

Toute preuve D peut se transformer en une preuve D^* du même séquent qui n'utilise pas la règle de coupure.

Nous nous limitons évidemment aux preuves construites dans notre petit calcul linéaire mais la possibilité d'éliminer les coupures est beaucoup plus générale, c'est même la propriété essentielle des systèmes déductifs à la Gentzen. La démonstration du théorème est un raisonnement par induction basé sur le lemme suivant :

Lemme principal

Soit $E :: \Gamma \vdash A$ et $F :: A, \Delta \vdash B$ sans coupure, alors il existe $D :: \Gamma, \Delta \vdash B$ sans coupure telle que $h(D) \leq h(E) + h(F) + 1$.

Ce lemme se démontre par induction sur $h(E) + h(F)$: ceci conduit à passer en revue toutes les façons que E et F ont de se terminer. Nous ne regarderons que deux cas typiques : le premier utilise le fait que certaines règles sont permutable entre elles, le second la symétrie des règles relatives à un même connecteur.

- Cas de permutation de règles. A n'est pas active dans la dernière règle de E . Si cette règle est par exemple $\otimes\text{-g}$, E se présente ainsi

$$\frac{E' :: C, D, \Gamma \vdash A}{C \otimes D, \Gamma \vdash A}$$

l'hypothèse d'induction est applicable à E' , F et produit une preuve $D' :: C, D, \Gamma, \Delta \vdash B$; il suffit maintenant d'appliquer la règle $\otimes\text{-g}$ à D' pour obtenir une preuve satisfaisante $D :: C \otimes D, \Gamma', \Delta \vdash B$.

- Cas d'élimination de règles. A est active dans la dernière règle de $\mathbf{E}$ et dans la dernière règle de $\mathbf{F}$, par exemple $A = A' \otimes A''$. Ces preuves se présentent donc ainsi

$$\frac{\mathbf{E} :: \Gamma \vdash A' \quad \mathbf{E}'' :: \Gamma' \vdash A''}{\Gamma, \Gamma' \vdash A' \otimes A''} \qquad \frac{\mathbf{F} :: A', A'', \Delta \vdash B}{A' \otimes A'', \Delta \vdash B}$$

l'application de l'hypothèse d'induction à $\mathbf{E}''$, $\mathbf{F}$ et A'' produit la preuve sans coupure $\mathbf{D}' :: \Gamma'', A', \Delta \vdash B$. La hauteur de cette preuve ne dépassant pas $h(\mathbf{E}'') + h(\mathbf{F}) + 1$ on peut encore appliquer l'hypothèse d'induction à $\mathbf{E}'$, $\mathbf{D}'$ et A' pour obtenir une preuve satisfaisante $\mathbf{D} :: \Gamma, \Gamma', \Delta \vdash B$.

La démonstration complète du lemme est fastidieuse mais ne ménagera plus aucune surprise à un lecteur courageux. Maintenant, l'élimination des coupures d'une preuve $\mathbf{D}$ se fait par l'induction suivante :

- Si $\mathbf{D}$ est un axiome, elle est sans coupure et $\mathbf{D}^* = \mathbf{D}$;
- si $\mathbf{D}$ provient de $\mathbf{E}$ par application d'une règle unaire : $\mathbf{D}^*$ s'obtient en appliquant cette même règle à la preuve sans coupure $\mathbf{E}^*$, provenant de $\mathbf{E}$ par l'hypothèse d'induction ;
- le cas où la dernière de $\mathbf{D}$ est binaire, autre que la coupure se traite de façon analogue ;
- reste le cas où la dernière règle de $\mathbf{D}$ est une coupure entre $\mathbf{E}$ et $\mathbf{F}$:
 - l'application de l'hypothèse d'induction à $\mathbf{E}$ et $\mathbf{F}$ produit deux preuves sans coupure $\mathbf{E}^*$ et $\mathbf{F}^*$,
 - l'application du lemme principal à ces deux preuves permet de construire $\mathbf{D}^*$ satisfaisant les conditions du théorème.

Le lecteur courageux dont nous supposons l'existence il y a un instant vérifiera que l'on a $h(\mathbf{D}^*) \leq h(\mathbf{D})$.

3.2.1. Théories associées à un système d'action

Une théorie linéaire se définit comme dans le cas traditionnel en adjoignant de nouveaux axiomes, réputés propres à la théorie, au système déductif : ceci veut dire que pour la construction d'une preuve dans une théorie, on admet comme axiomes les axiomes propres au même titre que les axiomes d'identité et **1**.d.

Nous nous placerons principalement dans la théorie **T** dont les axiomes propres, appelés *axiomes de transition*, sont la représentation des actions élémentaires en séquents linéaires : à chaque action élémentaire $A_1, \dots, A_m \Rightarrow B_1, \dots, B_n$ de notre système correspond donc un axiome de transition $A_1, \dots, A_m \vdash B_1 \otimes \dots \otimes B_n$ de la théorie **T**. Afin de bien marquer le parallèle que nous allons établir entre les *actions concrètes* et les preuves dans la théorie **T**, nous appellerons celles-ci des *actions formelles*. Rappelons que pour simplifier l'écriture, les deux langages en cause,

formel et concret, sont identiques, mais qu'il convient tout de même de faire la distinction entre les deux registres.

Lorsqu'un état E du système est précisé, on peut enrichir T en une théorie $T(E)$ par adjonction d'une famille d'axiomes propres, appelés *axiomes d'état* qui sont la représentation des ressources de E : à chaque membre A de E correspond un axiome d'état $\vdash A$ de $T(E)$; comme il se doit en matière de ressources, la famille des axiomes d'état de $T(E)$ est en bijection avec E .

Elimination des coupures

Qualifions de "propre" une coupure sur une formule d'un axiome propre : la permutation et l'élimination de règles qui nous ont permis d'éliminer les coupures ne peuvent rien contre une coupure propre ! Pour une théorie, on ne peut donc énoncer qu'un théorème d'élimination partielle des coupures pour une théorie :

Toute preuve dans une théorie T peut se transformer en une preuve du même séquent qui ne comporte que des coupures propres.

Le traitement des coupures propres a été revisité de nombreuses fois ; actuellement il est très utilisé, en association avec divers algorithmes, sous le nom de "méthode de résolution" : un lecteur nanti d'une bonne résolution devrait tenter de traduire ce qui va suivre dans ces termes. Venons en au résultat principal de cet article.

PROPOSITION 1

Toute action concrète est représentable par une action formelle et, réciproquement, toute action formelle est interprétable par une action concrète.

La première partie de cette proposition a déjà été établie dans le cadre traditionnel : en effet, la représentation des actions concrètes que nous avons donnée alors n'utilisait pas de règle structurelle et se traduit donc immédiatement dans le cadre linéaire. Il nous reste à vérifier la réciproque.

3.2.2. *Toute action formelle est interprétable par une action concrète*

L'interprétation A d'une formule A correspond exactement à la famille de ses constituants atomiques. Plus précisément, une interprétation d'une formule est une famille de ressources définie par :

- $\underline{1} = \emptyset$;
- $\underline{A} = A$ si A est atomique ;
- $\underline{A \otimes B} = \underline{A} + \underline{B}$, c'est-à-dire $\underline{A \cup B}$ à la condition que les interprétations $\underline{A}$ et $\underline{B}$ satisfassent $\underline{A} \cap \underline{B} = \emptyset$.
- De même, pour $\Gamma = A_1, \dots, A_m$ on pose $\underline{\Gamma} = \underline{A_1} + \dots + \underline{A_m}$, c'est-à-dire $\underline{A_1 \cup \dots \cup A_m}$ à la condition que $\underline{A_1}, \dots, \underline{A_m}$ soient disjointes deux à deux.

Nous allons démontrer la propriété suivante, pour tout état E du système :

Toute action formelle $D :: \Gamma \vdash B$ telle que $\underline{\Gamma} \subseteq E$ est interprétable par une action concrète $\underline{D} : \underline{\Gamma} \Rightarrow \underline{B}$.

Cette construction inductive suit point par point celle des actions formelles :

- $D : \vdash 1$ est 1.d : $\underline{D}$ est l'action identique $\emptyset \Rightarrow \emptyset$;
- $D : A \vdash A$ est l'axiome d'identité : $\underline{D}$ est l'action triviale $T_A : A \Rightarrow A$;
- $D : \Gamma \vdash B$ est un axiome de transition : $\underline{D}$ est l'action élémentaire $\underline{\Gamma} \Rightarrow \underline{B}$;
- D provient de $D' :: \Gamma \vdash B$ par application de 1.g ou $\otimes$.g, on a alors $\underline{\Gamma}' = \underline{\Gamma}$ et $\underline{D}$ est l'action concrète $\underline{D}'$ produite par l'hypothèse d'induction ;
- D provient de $D_1 :: \Gamma_1 \vdash C$ et $D_2 :: \Gamma_2, C \vdash B$ par application d'une coupure sur C , avec $\Gamma = \Gamma_1, \Gamma_2$. Par notre hypothèse sur D , $\underline{\Gamma} \subseteq E$ donc $\underline{\Gamma}_1$ et $\underline{\Gamma}_2$ sont disjointes et sont des sous familles de E :
- par l'hypothèse d'induction D_1 est interprétable par une action concrète $\underline{D}_1 : \underline{\Gamma}_1 \Rightarrow \underline{C}$ exécutable sur E : on a donc $E \cap \underline{C} \subseteq \underline{\Gamma}_1$ et en particulier, $\underline{\Gamma}_2 \cap \underline{C} = \emptyset$,
- si l'on pose $\underline{D}_1 : E \Rightarrow E_1$, on constate que $\underline{\Gamma}_2 \cup \underline{C} \subseteq E_1$ et on peut appliquer l'hypothèse d'induction à D_2 pour obtenir une action concrète $\underline{D}_2$ exécutable sur E_1 .

Finalement, $\underline{D} = \underline{D}_1 \circ \underline{D}_2$ est l'action concrète qui interprète D .

- D provient de $D_1 :: \Gamma_1 \vdash B_1$ et $D_2 :: \Gamma_2 \vdash B_2$ par application d'une règle $\otimes$.d avec $\Gamma = \Gamma_1, \Gamma_2$ et $B = B_1 \otimes B_2$. Comme ci-dessus, $\underline{\Gamma} \subseteq E$ donc $\underline{\Gamma}_1$ et $\underline{\Gamma}_2$ sont disjointes et sont des sous familles de E ; soient $\underline{D}_1 : \underline{\Gamma}_1 \Rightarrow \underline{B}_1$ et $\underline{D}_2 : \underline{\Gamma}_2 \Rightarrow \underline{B}_2$ les actions concrètes produites par l'hypothèse d'induction appliquée respectivement à D_1 et D_2 : elles sont concurrentes et on peut donc interpréter D par $\underline{D} = \underline{D}_1 \mid \underline{D}_2$.

- IL N'Y A PAS D'AUTRE CAS.

Remarques

- La démonstration précédente conduit à un résultat satisfaisant parce qu'elle sait s'arrêter à temps : le fait qu'il n'y ait pas d'autre cas que ceux que nous avons envisagés est essentiel. Lorsque nous avons tenté de faire une interprétation dans le cadre traditionnel, nous avons considéré des cas où D se terminait par une règle structurelle mais, nous avons constaté qu'alors, les ressources se trouvaient mal gérées et nous avons dû conclure négativement.

- Les actions formelles triviales, c'est-à-dire celles qui ne comportent pas d'axiome propre, sont représentées par des actions concrètes triviales. Le fait que toutes les coupures d'une action formelle triviale soient éliminables est le correspondant formel de la réduction que nous avons faite dans la section 2.2 de toute action triviale T à la forme T_X . La question des ressources laissées intactes par une action sera traitée dans la section 3.3 consacrée à la géométrie de l'action.

Un exemple d'action formelle

Nous allons représenter le début de l'action concrète de la figure 3, à savoir $\mathbf{D} = (1 \circ 3) \mid \mid 2$. On pourra se persuader du caractère massif des preuves dans un calcul des séquents en représentant complètement l'action $\mathbf{D} \circ \mathbf{E}$.

Représentation de $1 \circ 3 = \mathbf{ot}(c, a) \circ \mathbf{po}(c)$: les axiomes de transition représentant $\mathbf{ot}(c, a)$ et $\mathbf{po}(c)$ sont respectivement : $V, H(c), S(c, a) \vdash T(c) \otimes H(a)$ et $T(c) \vdash V \otimes H(c) \otimes B(c)$; l'action formelle cherchée est donc :

$$\begin{array}{c}
 T(c) \vdash V \otimes H(c) \otimes B(c) \quad H(a) \vdash H(a) \\
 \hline
 T(c), H(a) \vdash V \otimes H(c) \otimes B(c) \otimes H(a) \\
 \hline
 V, H(c), S(c, a) \vdash T(c) \otimes H(a) \quad T(c) \otimes H(a) \vdash V \otimes H(c) \otimes B(c) \otimes H(a) \\
 \hline
 V, H(c), S(c, a) \vdash V \otimes H(c) \otimes B(c) \otimes H(a)
 \end{array}$$

La construction de l'action formelle $\mathbf{D}$ représentant

$$\mathbf{D} = (1 \circ 3) \mid \mid 2 = (\mathbf{ot}(c, a) \circ \mathbf{po}(c)) \mid \mid \mathbf{ot}(c, a)$$

s'obtient en appliquant $\otimes_d$ à la preuve précédente et à l'axiome de transition représentant $\mathbf{ot}(c, a)$: le séquent ainsi prouvé est

$$V, V, H(c), H(c), S(c, a), S(c, a) \vdash V \otimes H(c) \otimes B(c) \otimes T(c) \otimes H(a) \otimes H(a)$$

Il resterait à construire la représentation $\mathbf{E}$ de $\mathbf{E} = (5 \circ 7) \mid \mid (4 \circ 6 \circ 8)$; puis celle de $\mathbf{D} \circ \mathbf{E}$, qui s'obtiendrait à partir de $\mathbf{D}$ et $\mathbf{E}$ par application d'une coupure, après quelques applications de $\otimes_d$ et $\otimes_g$: la chose est pénible et on a vraiment envie de laisser ce travail à une machine.

3.2.3. Consommation et production formelles de ressources

Dans la démonstration précédente, l'état $\mathbf{E}$ du système joue son rôle à l'extérieur de la scène purement formelle de la théorie $\mathbf{T}$ mais il ne peut plus en être de même si l'on travaille dans la théorie $\mathbf{T}(\mathbf{E})$: nous allons constater dans ce qui suit que la logique linéaire est capable d'agir réellement, c'est-à-dire de consommer des ressources et d'en produire des nouvelles.

Consommation

Une preuve dans $\mathbf{T}(\mathbf{E})$ est *parcimonieuse* lorsqu'elle consomme chaque axiome d'état au plus une fois, plus précisément, une preuve parcimonieuse $\mathbf{D}$ et la famille $\mathbf{AC}(\mathbf{D})$ des axiomes d'état qu'elle consomme sont simultanément définis par les clauses inductives suivantes :

- $\mathbf{D}$ est un axiome d'identité, 1_d ou un axiome de transition : $\mathbf{D}$ est parcimonieuse et $AC(\mathbf{D}) = \emptyset$;
- $\mathbf{D}$ est un axiome d'état $\vdash A$: $\mathbf{D}$ est parcimonieuse et $AC(\mathbf{D}) = A$;
- $\mathbf{D}$ provient de la preuve parcimonieuse $\mathbf{E}$ par application de 1_g ou $\otimes\_g$: $\mathbf{D}$ est parcimonieuse et $AC(\mathbf{D}) = AC(\mathbf{E})$;
- $\mathbf{D}$ provient des preuves parcimonieuses $\mathbf{E}$ et $\mathbf{F}$ par application d'une coupure ou d'une règle $\otimes\_d$ alors : si $AC(\mathbf{E}) \cap AC(\mathbf{F}) = \emptyset$, $\mathbf{D}$ est parcimonieuse et $AC(\mathbf{D}) = AC(\mathbf{E}) \cup AC(\mathbf{F})$.

Dans $\mathbf{T}(\mathbf{E})$ les axiomes d'état devront être soigneusement distingués des axiomes de transition car l'usage de ces derniers n'est pas limité *a priori* mais rien n'interdit à un axiome de transition d'avoir la forme $\vdash A$.

J'insiste : la logique linéaire permet de donner un sens à une expression comme "une preuve n'utilise certains axiomes qu'au plus une fois" (ce qui n'est certainement pas le cas d'une logique traditionnelle !) mais en aucun cas elle n'oblige à le faire. Si par exemple l'une des ressources A de votre système est inusable, cela signifie que sa production ne consomme rien et donc que votre expert a traduit celle-ci par un axiome de transition $\vdash A$.

Production

Cette notion est immédiate : une preuve $\mathbf{D} :: \Gamma \vdash B_1 \otimes \dots \otimes B_n$ "produit" les nouveaux axiomes d'état $\vdash B_1, \dots, \vdash B_n$.

Remarque

L'exécution de l'action concrète $\underline{\mathbf{D}}$ qui interprète l'action formelle $\mathbf{D} :: \Gamma \vdash B$ consomme les ressources de $\underline{\Gamma}$ et produit les ressources de $\underline{B}$: c'est exactement ce que ferait une preuve parcimonieuse $\mathbf{D}' :: \vdash B$ telle que $AC(\mathbf{D}') = \underline{\Gamma}$. On peut paraphraser cette remarque en disant que $\mathbf{D}'$ transforme $\mathbf{T}(\mathbf{E})$ en $\mathbf{T}(\underline{\mathbf{E}}')$ où $\mathbf{E}' = (\mathbf{E} - \underline{\Gamma}) \cup B$, ce qui est l'expression formelle de $\underline{\mathbf{D}} : \mathbf{E} \Rightarrow \mathbf{E}'$.

Lemme

Toute action formelle $\mathbf{D} :: \Gamma \vdash B$ peut se transformer en une preuve parcimonieuse $\mathbf{D}' :: \vdash B$ telle que $AC(\mathbf{D}') = \underline{\Gamma}$ et réciproquement.

D'après la remarque, ce lemme est très apparenté à la preuve de la proposition 1. Nous donnons au lecteur le moyen de s'en convaincre en lui laissant le soin de faire lui-même la démonstration. En tout cas, ce lemme met en valeur un caractère dynamique de la logique linéaire.

3.3. Une géométrie de l'action

La construction d'une action concrète ou formelle contient beaucoup d'opérations purement "administratives" liées à l'encadrement. Nous allons maintenant définir un mode de représentation très économique des preuves qui substitue des opérations géométriques aux opérations logiques. Dans un premier temps nous définissons un

type de graphes orientés adapté à la théorie **T** de notre système (les pseudo-plans), puis nous associons un graphe de ce type à chaque action formelle (son plan), enfin nous caractérisons les plans parmi les pseudo-plans par une propriété intrinsèque de nature purement géométrique.

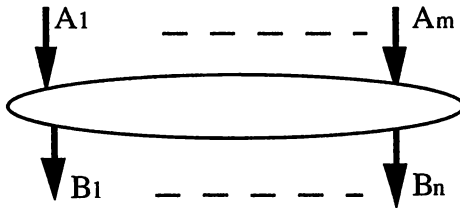
Les pseudo-plans

Un pseudo-plan est un graphe composé de nœuds et de liens orientés :

- un nœud, modelé sur un axiome de transition

$$A_1, \dots, A_m \vdash B_1 \otimes \dots \otimes B_n$$

est équipé des entrées $eA_1, \dots, eA_m$ et des sorties $sB_1, \dots, sB_n$;



- un lien orienté admet une sortie sA comme origine et une entrée eA de même type comme extrémité.

Une entrée ou une sortie n'est l'extrémité ou l'origine que d'un lien au plus, celles qui ne le sont pas sont appelées les entrées ou les sorties du pseudo-plan : elles constituent évidemment des familles.

L'orientation d'un pseudo-plan est la clôture transitive de la relation définie sur les nœuds par :

$E \angle F$ si et seulement si il existe un lien dont l'origine est attachée à E et l'extrémité à F .

Les plans

On tire un pseudo-plan **D** de toute action formelle **D** de la façon suivante :

- **D** est un axiome d'identité ou $1.d$: **D** est vide ;
- **D** est un axiome de transition : **D** est réduit à un nœud modelé sur **D** ;
- **D** provient de **E** par application de $\otimes.g$ ou $1.g$: **D** = **E** ;
- **D** provient de **E** et **F** par application de $\otimes.d$: **D** est la réunion de **E** et **F** ;
- **D** provient de **E** et **F** par application d'une coupure : **D** s'obtient à partir de la réunion de **E** et **F** par adjonction d'un lien entre sC_k et eC_k , pour chaque constituant atomique C_k de la formule de coupure tel que sC_k soit une sortie de **E** et eC_k une entrée de **F**.

Un plan est le pseudo-plan tiré d'une action formelle par les clauses précédentes.

Remarques

- Grâce à la proposition 1, on peut aussi tirer directement le plan d'une action concrète : le passage par l'action formelle n'est pas indispensable. Toutefois, il faut bien réaliser que les plans (et plus généralement les réseaux de preuves) sont encore des preuves mais exprimées de façon géométrique (cf. la "déduction naturelle" de G. Gentzen lui-même).

- Le plan tiré d'une action formelle ne comportant pas d'axiome de transition, est vide : en effet un plan néglige de noter les ressources qui restent intactes.

- Plus généralement, deux actions distinctes peuvent avoir le même plan. Nous pouvons enfin donner la définition de l'équivalence entre actions :

Deux actions $I \Rightarrow F$ sont équivalentes ssi on en tire le même plan.

- Dans le cas d'une coupure, il se peut qu'aucun lien ne soit ajouté à la réunion de E et F : lorsque cela est possible, les plans transforment automatiquement une composition séquentielle en une composition en concurrence (ce phénomène peut aussi s'interpréter comme l'élimination d'une coupure).

- L'orientation d'un plan ne comporte pas de cycle (c'est donc un ordre partiel strict), comme on peut le constater sur la construction.

PROPOSITION 2

Soit D un pseudo-plan dont la famille des entrées est incluse dans un état E . Alors, D est un plan si et seulement si son orientation n'admet pas de cycle.

Nous allons démontrer la propriété suivante par induction sur les pseudo-plans, ordonnés par la relation d'inclusion :

Tout pseudo-plan sans cycle D dont la famille d'entrées est $eA_1, \dots, eA_m$ et dont la famille des sorties est $sB_1, \dots, sB_n$, est le plan tiré d'une action formelle $D :: A_1, \dots, A_m \vdash B_1 \otimes \dots \otimes B_n$.

- D est vide : $D = 1.d$;

- D est réduit à un seul nœud : D est l'axiome de transition sur lequel D a pris modèle ;

- D est la réunion disjointe de deux pseudo-plans non vides E et F : l'hypothèse d'induction produit des actions formelles E et F desquelles E et F sont respectivement tirés, D s'obtient en leur appliquant $\otimes.d$;

- dans les autres cas, choisissons un nœud A qui est maximal dans D pour la relation $\angle$ (un tel nœud existe puisqu'il n'y a pas de cycle). Aucune sortie de A n'est l'origine d'un lien, donc si l'on rompt les liens qui aboutissent à des entrées de A , on sépare ce nœud de D ; appelons E ce qui reste de D après cette opération et livrons nous à un inventaire :

- entrées de E : $eA_1, \dots, eA_q$

- sorties de E : $sC_1, \dots, sC_p, sB_1, \dots, sB_r$
- entrées de A : $eA_{q+1}, \dots, eA_m, eC_1, \dots, eC_p$
- sorties de A : $sB_{r+1}, \dots, sB_n$

l'hypothèse d'induction produit alors deux actions formelles

$$E :: A_1, \dots, A_q \vdash C_1 \otimes \dots \otimes C_p \otimes B_1 \otimes \dots \otimes B_r$$

$$A :: C_1, \dots, C_p, A_{q+1}, \dots, A_m \vdash B_{r+1} \otimes \dots \otimes B_n$$

d'où il est facile, par une coupure sur $C_1 \otimes \dots \otimes C_p \otimes B_1 \otimes \dots \otimes B_r$ de déduire l'action formelle cherchée

$$D :: A_1, \dots, A_m \vdash B_1 \otimes \dots \otimes B_n.$$

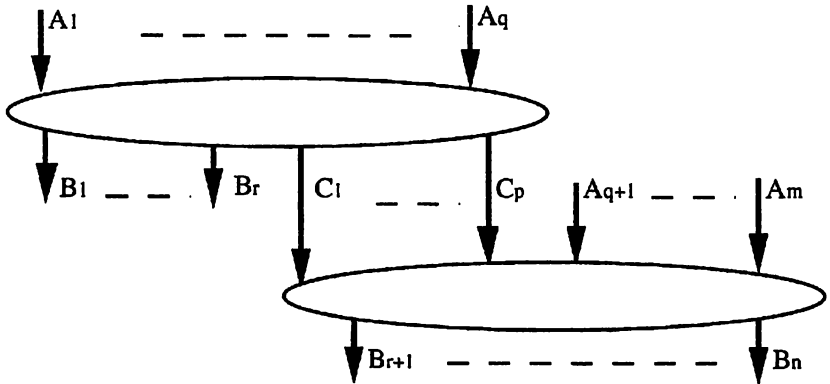


Figure 4.

Exemple

La figure 5 représente le plan tiré de l'action de la figure 3, que nous avons représentée en partie dans une précédente section.

Cette figure, dans son laconisme même, représente bien un plan :

- Les axiomes de transition sont identifiables à la lecture des entrées et des sorties : 1, 3, 5 et 7 dans la colonne de gauche et 2, 4, 6 et 8 dans celle de droite.
- Les occurrences des constantes ne sont pas distinguées explicitement et on ne peut donc plus les suivre au passage d'un nœud. Par contre, deux entrées sont nécessairement des membres distincts de la famille (de même que les sorties) et la bonne gestion des ressources est donc assurée.

Remarques

- La démonstration de la proposition 2 est basée sur un découpage méthodique : en fait, tout découpage "raisonnable" d'un pseudo-plan sans cycle est utilisable. Par exemple, dans la figure 5, la rupture des liens qui isole les trois nœuds du haut (1, 2 et 3) est raisonnable et produit la description donnée à la fin de la section 2.2 :

$$((1\text{ }03) \mid \mid 2)\alpha((5\text{ }07) \mid \mid (4\text{ }06\text{ }08)).$$

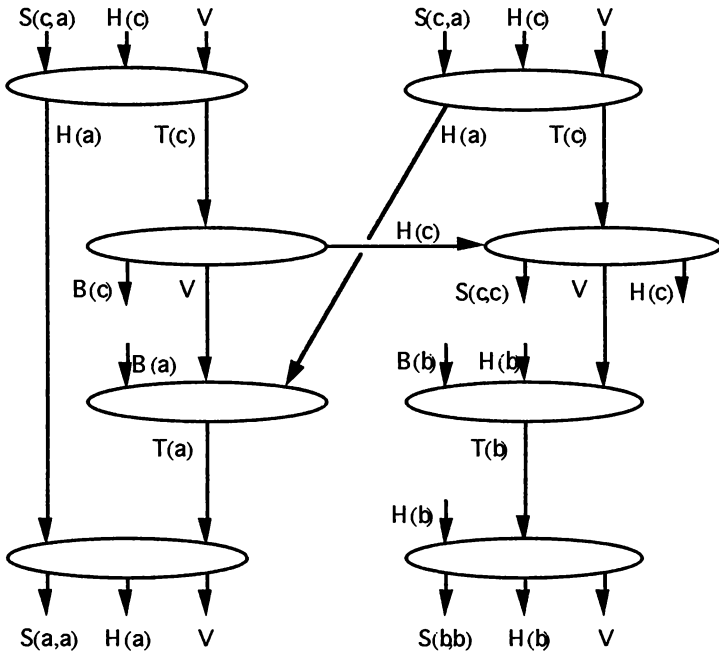


Figure 5.

- Il peut être intéressant d'interpréter un plan comme un schéma de réécriture : nous avons déjà signalé la possibilité de ce point de vue lors de la description des actions concrètes. La notion de plan a cependant un avantage théorique et pratique sur celle de la réécriture : elle ne prend en compte que les ressources qui sont effectivement en cause, à l'exclusion de tout élément du "décor", autrement dit : *le problème du cadre ne se pose pas* lorsque l'on utilise la logique linéaire comme système déductif et les plans qui lui correspondent. Il reste que la réécriture est un domaine riche en problèmes intéressants du point de vue de la planification ; par ailleurs beaucoup de résultats relatifs, par exemple, à la complexité, sont directement transposables à la planification.

- *Le temps et la durée.* L'orientation d'un plan a une signification immédiate :

$A \angle B$ si et seulement si A doit être exécutée avant B .

Il faut donc la comprendre comme la contrainte minimale sur l'ordre chronologique d'exécution. Lorsque l'on cherche à optimiser le mode d'exécution d'un plan en fonction de la durée des actions élémentaires, on peut être conduit à modifier $\angle$ en l'étendant en un ordre (partiel) strict plus contraignant : par exemple, un ordre linéaire... si l'on ne peut pas exécuter des actions en parallèle. Mais ceci n'est plus de la logique.

4. Calcul des séquents linéaire

Dans ce chapitre, nous présentons brièvement et par petits morceaux, l'intégralité du calcul des séquents linéaire de Girard dans sa version intuitionniste (une seule formule à droite des séquents).

Les opérateurs logiques (connecteurs et constantes de prédicat) sont introduits par le seul énoncé des règles et des axiomes qui définissent leur emploi dans une preuve : toutefois, certains commentaires relatifs à des opérateurs particuliers tiennent compte du calcul tout entier. En dehors des observations ci-dessous, nous ne tenterons aucune théorie générale à leur sujet : simplement, nous en illustrerons l'interprétation possible par des exemples.

Les théories linéaires

Les définitions de base du chapitre 3 s'adaptent plus ou moins immédiatement aux cas qui seront envisagés maintenant.

Une théorie **T**, représentant un système d'action dans un sens généralisé, est la donnée d'un ensemble d'axiomes propres, dits *axiomes de transition*. Nous appellerons encore *action formelle* toute preuve dans une théorie **T** du type précédent.

- Dans l'exposé des chapitres précédents, nous avons toujours travaillé avec des formules closes. Vous avez certainement compris que cette restriction n'avait rien d'essentiel : au contraire, il est intéressant de n'instancier que le plus tard possible. C'est pour la commodité de l'exposé et pour faciliter une première lecture que nous avons procédé ainsi ; nous vous invitons, si vous désirez faire une étude plus approfondie, à revoir au moins le chapitre 3 dans cette perspective.

- Les action formelles qui ne comportent aucun axiome propre sont des réaménagements plus ou moins "actifs" des ressources : ceci n'est pas sensible dans le chapitre 3 (voir la remarque qui suit) mais est important à remarquer pour les cas que nous considérerons par la suite.

- Lorsqu'un état E , c'est-à-dire une famille de formules, est donné, la théorie $T(E)$ s'obtient en adjoignant une famille d'*axiomes d'état* à **T** c'est-à-dire $\vdash A$ pour chaque membre A de E . Les preuves "parcimonieuses" de $T(E)$ sont celles qui utilisent correctement les axiomes d'état (la définition n'est pas donnée ici).

- La leçon générale à tirer de l'étude des plans faite précédemment est qu'il est intéressant de noter une ressource $A \otimes B$ sous la forme A, B (c'est-à-dire, de ne pas marquer $\otimes$ lorsqu'il est opérateur principal). Il est clair que, dès que l'on sort du cas purement conjonctif, il n'est plus possible de concevoir des plans dont toutes les entrées et les sorties sont atomiques. A notre connaissance, il n'existe que des résultats partiels à ce sujet, dont nous donnerons seulement une illustration, sans justification.

- L'élimination des coupures logiques est encore possible et nous avons donc la propriété :

Toute preuve dans une théorie peut se transformer en une preuve du même séquent qui ne comporte que des coupures propres à cette théorie.

où une coupure est propre si elle porte sur une formule d'un axiome propre.

Notion de fragment

Un fragment est la donnée d'un ensemble $\mathbf{Fr}$ d'opérateurs logiques : un séquent (une preuve, une théorie) du fragment $\mathbf{Fr}$ ne contient que des opérateurs qui sont dans $\mathbf{Fr}$.

Pour construire une preuve $\mathbf{D}$ d'un séquent S du fragment $\mathbf{Fr}$ dans une théorie du même fragment, on n'utilise que les règles et axiomes du groupe de l'identité et ceux qui sont relatifs aux éléments de $\mathbf{Fr}$:

non seulement ceci est naturel mais encore est-ce correct !

En effet, si un opérateur qui n'est pas dans $\mathbf{Fr}$ intervient dans $\mathbf{D}$, il doit disparaître ultérieurement puisqu'il ne se retrouve pas dans S , or :

- seule l'application d'une coupure est capable de faire disparaître une formule,
- cette coupure ne porte pas sur une formule d'un axiome propre de la théorie puisque celle-ci est écrite dans le fragment $\mathbf{Fr}$; on peut l'éliminer.

On obtient une preuve de S entièrement écrite dans $\mathbf{Fr}$ en éliminant toute coupure de ce genre.

4.1. Le groupe de l'identité

C'est le noyau minimal de toute logique : il définit le fragment $\mathbf{Fr} = \emptyset$.

(axiome) $A \vdash A$

(coupure)
$$\frac{\Gamma \vdash A \quad A, \Delta \vdash B}{\Gamma, \Delta \vdash B}$$

Un axiome de transition d'une théorie $\mathbf{T}$ de ce fragment est un séquent constitué uniquement de formules atomiques $A_1, \dots, A_m \vdash B$: c'est donc un programme Prolog (la clause Prolog correspondant à l'axiome précédent est $B :- A_1, \dots, A_m$) où les clauses d'état sont *rétractées* après usage.

4.2 . Le groupe multiplicatif

Ce groupe définit le fragment $\{\otimes, \mathbf{1}, \multimap\}$: il est caractérisé par une gestion des contextes par juxtaposition.

$$\begin{array}{ll}
 (\otimes.d) \quad \frac{\Gamma \vdash A \quad \Delta \vdash B}{\Gamma, \Delta \vdash (A \otimes B)} & (\otimes.g) \quad \frac{A, B, \Gamma \vdash C}{(A \otimes B), \Gamma \vdash C} \\
 (\mathbf{1}.d) \quad \vdash \mathbf{1} & (\mathbf{1}.g) \quad \frac{\Gamma \vdash A}{\Gamma, \mathbf{1} \vdash A} \\
 (O.d) \quad \frac{\Gamma, A \vdash B}{\Gamma \vdash (A \multimap B)} & ((.g) \quad \frac{\Gamma \vdash A \quad B, \Delta \vdash C}{(A \multimap B), \Gamma, \Delta \vdash C}
 \end{array}$$

La planification qui interprète ce groupe est dite *déterministe* :

- Le déterminisme correspondant au fragment $\{\otimes, \mathbf{1}\}$ est qualifié de *conjonctif* : c'est le cas qui nous a occupés dans cet article. Les fragments que l'on est conduit à considérer contiennent généralement $\otimes$.

- Lorsque $\multimap \in \mathbf{Fr}$, la planification dans $\mathbf{Fr}$ est qualifiée de *conditionnelle*.

Un exemple de déterminisme conditionnel

On posait jadis une devinette dans le genre de celle-ci :

Un clochard dispose de deux mégots mais il lui en faut absolument trois pour se rouler une cigarette. Comment va-t-il faire pour fumer* ?

Il emprunte un mégot à l'un de ses compagnons de dessous les ponts, roule une cigarette et la fume : il ne lui reste plus qu'à donner le mégot qu'il vient de produire à son coopérant ami.

Représentons cette action de la façon suivante en prenant les prédicats : M pour un mégot, C pour une cigarette et F pour la confiance d'un ami qui possède un mégot. Les axiomes de transition sont :

$F \vdash M \otimes (M \multimap F)$: emprunter un mégot,

$M, M, M \vdash C$: rouler une cigarette,

$C \vdash M$: fumer.

L'état du système au début du drame est $E = M, M, F$.

* Selon la loi numéro 91-32 : fumer nuit gravement à la santé.

L'action formelle qui suit représente la composée des actions élémentaires emprunter ◦ rouler ◦ fumer :

$$\begin{array}{c}
 M \vdash M \quad F \vdash F \\
 \hline
 \\
 \begin{array}{cc}
 M, M, M \vdash C & C \vdash M \\
 \hline
 M, M, M \vdash M & \\
 \hline
 \end{array}
 \quad
 \begin{array}{c}
 M, M \rightarrow F \vdash F \\
 \hline
 M \rightarrow F \vdash M \rightarrow F \\
 \hline
 \end{array}
 \\
 \hline
 M, M, M, M \rightarrow F \vdash M \otimes (M \rightarrow F) \\
 \hline
 F \vdash M \otimes (M \rightarrow F) \quad M, M, M \otimes (M \rightarrow F) \vdash M \otimes (M \rightarrow F) \\
 \hline
 M, M, F \vdash M \otimes (M \rightarrow F)
 \end{array}$$

Notre clochard dispose maintenant des ressources $E' = M, M \rightarrow F$: pour retrouver la confiance de son ami, il doit lui rendre le mégot (qu'il vient de produire !) en réaménageant ses ressources par l'exécution de l'action formelle suivante :

$$\begin{array}{c}
 M \vdash M \quad F \vdash F \\
 \hline
 M, M \rightarrow F \vdash F
 \end{array}$$

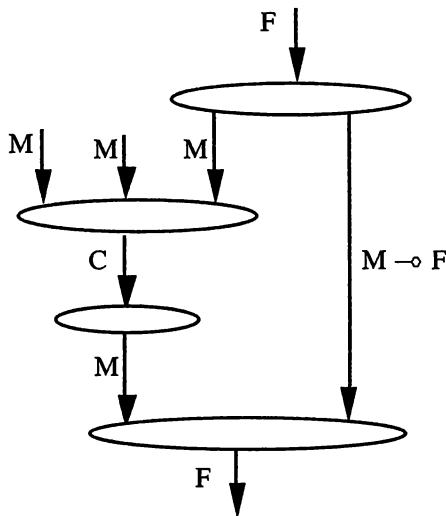


Figure 6.

qui n'utilise aucun axiome de transition. Ce réaménagement de ressources est certainement le plus "actif" de tous ceux qui peuvent se présenter, en effet, chacun des séquents $\vdash M \rightarrow F$ et $M \vdash F$ est prouvable à partir de l'autre :

l'axiome d'état $\vdash M \rightarrow F$ se comporte donc comme un "axiome de transition" qui ne peut servir qu'une fois.

La figure 6 est le plan tiré de l'action formelle que nous venons de construire : sa signification est claire sur cet exemple simple.

4.3. Le groupe additif

Ce groupe est caractérisé par une gestion des contextes par identification : il introduit deux formes de non déterminisme.

Le groupe additif : les opérateurs conjonctifs

$$(\&_d) \frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash (A \& B)}$$

$$\frac{A, \Gamma \vdash C}{(A \& B), \Gamma \vdash C} (\&_g1)$$

$$(T_d) \quad \Gamma \vdash T$$

$$\frac{B, \Gamma \vdash C}{(A \& B), \Gamma \vdash C} (\&_g2)$$

$$(\wedge_d) \frac{\Gamma \vdash A}{\Gamma \vdash \wedge_x A}$$

$$\frac{A[u], \Gamma \vdash B}{\wedge_x A[x], \Gamma \vdash B} (\wedge_g)$$

Condition : x n'est pas libre dans le Γ de la règle $\wedge_d$ pour le quantificateur universel.

L'opérateur *avec* $\&$ introduit un non déterminisme externe : l'exécution d'une action représentée par $\Gamma \vdash A \& B$ dépend d'un choix entre A et B .

Le seul exemple que nous donnerons ici est celui du fameux choix que l'on a à faire dans le menu de maint restaurant vers la fin du repas : "fromage ou dessert" : les deux possibilités sont bien offertes mais, si l'on mange du fromage, on ne peut plus avoir de dessert (sauf à payer un supplément !).

La constante T correspond à une absence de choix ; le quantificateur universel $\wedge$ est simplement un $\&$ généralisé.

Le groupe additif : les opérateurs disjonctifs

$$\begin{array}{c}
 \text{(\oplus\_d1)} \quad \frac{\Gamma \vdash A}{\Gamma \vdash (A \oplus B)} \qquad \frac{A, \Gamma \vdash C \quad B, \Gamma \vdash C}{(A \oplus B), \Gamma \vdash C} \qquad \text{(\oplus\_g)}
 \end{array}$$

$$\begin{array}{c}
 \text{(\oplus\_d2)} \quad \frac{\Gamma \vdash B}{\Gamma \vdash (A \oplus B)} \qquad \text{(\mathbf{0\_g})} \quad \mathbf{0}, \Gamma \vdash A
 \end{array}$$

$$\begin{array}{c}
 \text{(\vee\_d)} \quad \frac{\Gamma \vdash A[u]}{\Gamma \vdash \vee_x A[x]} \qquad \frac{A, \Gamma \vdash B}{\vee_x A, \Gamma \vdash B} \qquad \text{(\vee\_g)}
 \end{array}$$

Condition : x n'est pas libre dans le Γ de la règle $\vee\_g$ pour le quantificateur existentiel.

L'opérateur $\oplus$ introduit un non déterminisme interne : l'exécution d'une action représentée par $\Gamma \vdash A \oplus B$ offre les alternatives A ou B : une seule de ces alternatives se trouve réalisée après l'exécution mais ce n'est plus l'utilisateur lui-même qui choisit.

Voici un exemple, lui aussi très classique, dans le fragment $\{\otimes, \oplus\}$. On dispose de chaussettes noires ou blanches dans une boîte fermée : le jeu consiste à extraire des chaussettes jusqu'à obtenir une bonne paire, or chacun sait que trois tirages sont toujours suffisants. Représentons cette action de la façon suivante en utilisant les prédicats B pour une chaussette blanche, N pour une chaussette noire, Ok pour une bonne paire. Les axiomes de transition sont :

$B, B \vdash Ok$: paire blanche,
 $N, N \vdash Ok$: paire noire,
 $Ok, B \vdash Ok$: merci, on est déjà satisfait,
 $Ok, N \vdash Ok$: merci, on est déjà satisfait.

L'état initial du système $E = B \oplus N, B \oplus N, B \oplus N, \dots$ contient au moins trois chaussettes.

L'action commence par un réaménagement des ressources de E qui est représenté par une action formelle n'utilisant pas d'axiome propre :

$\mathbf{D} : B \oplus N, B \oplus N, B \oplus N \vdash ((B \otimes B) \oplus (N \otimes N)) \otimes (B \oplus N).$

cette preuve est la simple expression de la distributivité de $\otimes$ par rapport à $\oplus$ et de l'idempotence de $\oplus$ (cette dernière propriété signifie que $A \oplus A \vdash A$ et $A \vdash A \oplus A$ sont prouvables). Maintenant on a l'action formelle $\mathbf{E}$

$$\begin{array}{c}
 \begin{array}{cc} B, B \vdash \text{Ok} & N, N \vdash \text{Ok} \end{array} \qquad \begin{array}{cc} B \vdash B & N \vdash N \end{array} \\
 \hline
 \begin{array}{cc} B \otimes B \vdash \text{Ok} & N \otimes N \vdash \text{Ok} \end{array} \qquad \begin{array}{cc} B \vdash B \oplus N & N \vdash B \oplus N \end{array} \\
 \hline
 \begin{array}{cc} (B \otimes B) \oplus (N \otimes N) \vdash \text{Ok} & B \oplus N \vdash B \oplus N \end{array} \\
 \hline
 ((B \otimes B) \oplus (N \otimes N)), (B \oplus N) \vdash \text{Ok} \otimes (B \oplus N) \\
 \hline
 ((B \otimes B) \oplus (N \otimes N)) \otimes (B \oplus N) \vdash \text{Ok} \otimes (B \oplus N)
 \end{array}$$

et l'action formelle **F**

$$\begin{array}{c}
 \text{Ok}, B \vdash \text{Ok} \qquad \text{Ok}, N \vdash \text{Ok} \\
 \hline
 \text{Ok}, B \oplus N \vdash \text{Ok} \\
 \hline
 \text{Ok} \otimes (B \oplus N) \vdash \text{Ok}
 \end{array}$$

Des coupures appliquées à **D** et **E** puis à **F** produisent l'action formelle cherchée, c'est-à-dire une preuve de $B \oplus N, B \oplus N, B \oplus N \vdash \text{Ok}$. L'action **D** s'est montrée très déterminante bien qu'elle n'utilise aucun axiome de transition.

4.4 . Le groupe de l'exponentielle

Pour clore ce bref exposé sur la logique linéaire intuitionniste, voici le groupe de l'exponentielle *bien sûr* qui réintroduit la possibilité de faire des contractions et des affaiblissements sur les formules $!A$. Son nom vient de ce qu'il permet le passage du fragment $\{\&, T\}$ au fragment $\{\otimes, 1\}$, par exemple, les séquents

$$!(A \& B) \vdash !A \otimes !B \text{ et } !A \otimes !B \vdash !(A \& B)$$

sont prouvables.

$$\begin{array}{c}
 \Gamma \vdash B \\
 \hline
 !A, \Gamma \vdash B \qquad (A) \\
 \\
 !A, !A, \Gamma \vdash B \\
 \hline
 !A, \Gamma \vdash B \qquad (C)
 \end{array}$$

$$\frac{A, \Gamma \vdash B}{!A, \Gamma \vdash B} \quad (D)$$

$$\frac{! \Gamma \vdash A}{! \Gamma \vdash !A} \quad (!)$$

En plus des règles correspondant à l'affaiblissement (A) et à la contraction (C), on a une règle à droite (!) (qui n'est pas très belle : ! Γ signifie que ! s'applique à chaque membre de la famille Γ) et la règle à gauche, dite de déréluction (D). Pour ce qui nous concerne, cet opérateur donne la possibilité, par exemple, de considérer la production de ressources qui ne sont pas périssables : c'est ce que fait, par exemple, un axiome de transition de la forme $\Gamma \vdash A \otimes B$

Bibliographie

Pour un traitement abordable mais un peu ancien du calcul des séquents consultez [1], vous trouverez tout sur ce sujet, avant l'introduction de la logique linéaire dans [2] :

- [1] S.C. Kleene.- Logique mathématique (traduit de l'anglais), Armand Colin 1971, réédité chez J. Gabbay, 1987.
- [2] J.-Y. Girard.- Proof theory and logical complexity, Bibliopolis (Napoli), 1987.
- L'article [3] très simple et intéressant est idéal pour faire ses premiers pas en logique linéaire, [4] est l'article fondateur dont [5] est bon complément.
- [3] J.-Y. Girard.- La logique linéaire, *Pour la Science* **150** (1990), 74-85.
- [4] J.-Y. Girard.- Linear logic, *Theoretical Computer Science* **50** (1987), 1-102.
- [5] J.-Y. Girard, Y. Lafont, P. Taylor.- Proofs and types, *Cambridge Tracts in Computer Science*, Cambridge U. Press, Cambridge, 1989.

Vous pouvez consulter les actes de l'IJCAI et de l'AAAI (chez Morgan Kaufmann et aux M.I.T. Press) pour avoir un bon panorama sur la planification ; [8] fait en français, un bilan de la planification, régulièrement mis à jour :

- [6] R.E. Fikes, N.J. Nilsson.- STRIPS : A new approach to the application of theorem proving to problem solving, *Artificial Intelligence* **2** (1971), 189-208.
- [7] M.P. Georgeff, A.L. Lansky eds.- Reasoning about actions and plans (Proc. of the 1986 Workshop), Morgan Kaufmann, 1987.
- [8] P. Regnier.- Planification : historique, principes, problèmes et méthodes (de GPS à ABTWEAK), rapport IRIT 90/22, Université Paul Sabatier, Toulouse.

Une forme de plans conjonctifs, avant la logique linéaire, est envisagée dans [9] ; [12] est une version abrégée et très accessible de [10] et [11] :

- [9] W. Bibel.- A deductive solution for plan generation, *New generation Computing* **4** (1986), 115-132.
- [10] M. Masseron, C. Tollu, J. Vauzeilles.- Generating plans in linear logic I : Actions as proofs, *Theoretical Computer Science* **113** (1993), 349-370.
- [11] M. Masseron.- Generating plans in linear logic II : A geometry of actions, *Theoretical Computer Science* **113** (1993), 371-375.
- [12] M. Masseron, C. Tollu, J. Vauzeilles.- Planification et logique linéaire, *Revue d'intelligence artificielle* **8** (1992), 285-311.

Annexe des familles

Une suite $X_1 \dots X_m$ d'éléments d'un ensemble X n'est qu'une notation commode pour l'application $f : \{1, \dots, m\} \rightarrow X$ définie par $f(i) = X_i$ pour chaque $i \in \{1, \dots, m\}$.

Dans ce cas, les entiers sont pris dans l'ordre linéaire qui leur est naturel $1 < \dots < m$ mais, lorsque les "indices" sont les éléments d'un ensemble quelconque, on obtient une famille, une liste linéaire n'est plus adaptée pour la représenter (bien qu'on le fasse encore souvent, par abus de notation) : les familles sont plus simples que les suites, puisque la notion d'ensemble est plus primitive* que celle d'ensemble ordonné, mais la représentation en est plus délicate.

Nous allons rappeler succinctement quelques unes des opérations qui sont possibles sur les familles (en indiquant l'opération analogue pour les suites, qui est normalement bien connue).

Une famille $X = (X_x)_{x \in L}$ d'éléments de X indexée par l'ensemble L est une application $f : L \rightarrow X$ pour laquelle X_x est une notation de $f(x)$ pour chaque $x \in L$.

Le cas où f n'est pas injective rend compte des éventuelles répétitions que l'on veut représenter. Pour fixer des notations, nous considérerons les familles

$$W = (W_w)_{w \in K}, X = (X_x)_{x \in L}, Y = (Y_y)_{y \in M} \text{ et } Z = (Z_z)_{z \in N}.$$

- $X = \emptyset$ signifie $L = \emptyset$ (suite de longueur nulle).
- $X \subseteq W$ signifie $L \subseteq K$ et $X_x = W_w$ pour tout $x \in L$ (sous suite).
- si $X \subseteq W$ et $Y \subseteq W$, on peut définir $X \cap Y$, $X \cup Y$ et $X - Y$, par exemple $Z = X \cap Y$ est indexée par $N = L \cap M$ et $Z_z = X_z = Y_z$ pour tout $z \in N$.
- La nature particulière des éléments d'un ensemble d'indices est de peu d'importance et l'on est souvent conduit à "réindexer" une famille : si $\varphi : K \rightarrow L$ est une bijection, $W_w = X_{\varphi(w)}$ définit W comme le résultat d'une réindexation de X ; on dit alors que W est équivalente à X . Dans cette circonstance, il est

* si je ne me trompe !

courant d'employer l'abus de notation ** qui consiste à désigner W simplement par X : ceci donne des énoncés plus simples mais un peu trompeurs, comme nous le verrons dans un instant.

- La somme ou union disjointe $Z = X+Y$ de X et Y (concaténation) est la famille indexée par $N = L+M = (L \times \{0\}) \cup (M \times \{1\})$ et définie par

$$Z_{(x,0)} = X_x \text{ pour } x \in L \text{ et } Z_{(y,1)} = Y_y \text{ pour } y \in M.$$

Lorsque $X \subseteq W$ et $Y \subseteq W$ sont disjointes, c'est-à-dire lorsque $X \cap Y = \emptyset$, $X+Y$ est égale à $X \cup Y$.

Application : exégèse d'une définition

Il s'agit de la définition de la section 2.1.1 dont les données sont :

- un couple de familles (I, F) : on peut théoriquement supposer que $I = (I_x)_{x \in L}$ et $F = (F_y)_{y \in M}$ sont disjointes mais, ceci peut ne pas correspondre à la situation concrète que l'on a en vue et l'on doit les considérer comme des sous familles de $W = (W_w)_{w \in L \cup M}$.

- une famille $E = (E_z)_{z \in N}$.

- La condition $I \subseteq E$ doit se comprendre de la façon suivante : il est possible d'unifier I à une sous famille de E , c'est-à-dire, il existe une injection $f : L \rightarrow N$ telle que $E_{f(x)} = I_x$ pour tout $x \in L$. Ceci implique aussi la réindexation d'une partie de F .

- La condition $E \cap F \subseteq I$ signifie qu'il est alors encore possible de mettre $M-L$ en bijection avec un M' vérifiant $N \cap M' = \emptyset$ (ce qui est peu contraignant).

- Après les réindexations précédentes, on doit pouvoir unifier la famille $(E-I) \cup F$ à un élément de Σ .

La majeure partie du chapitre 2 demande une interprétation de ce type.

Marcel MASSERON

** très utilisé par les physiciens, selon lequel, par exemple, les fonctions représentant l'entropie sont toujours désignées par la même lettre, indépendamment des paramètres utilisés pour l'exprimer.

Chapitre 5

Approches logiques du raisonnement taxinomique

Introduction

Appelons raisonnement taxinomique l'ensemble de procédures qui gouvernent l'application correcte d'une taxinomie, c'est-à-dire d'une structure hiérarchique analogue à celles d'usage courant en botanique ou en zoologie. De telles procédures nous permettent d'une part de désigner le genre ou l'espèce d'un individu donné à partir de l'examen de ses caractéristiques telles que les révèle la seule observation, et d'autre part, de prédire pour une espèce ou un genre donné, les propriétés qui sont communes aux individus qu'ils désignent. L'inaptitude de la logique classique à formaliser ces types de procédures a été souvent évoquée.

Les problèmes soulevés par le raisonnement taxinomique renvoient à deux questions plus générales concernant : 1) le rapport entre termes théoriques et termes observationnels et 2) le rôle que pourrait jouer la logique dans l'examen de ces rapports. La philosophie de ce siècle nous montre clairement que la tentative d'appliquer la logique classique à la formalisation de sciences autres que les mathématiques, qui doivent donc intégrer dans un corps théorique des données observationnelles, n'a jamais été accomplie de façon satisfaisante. Le raisonnement taxinomique n'est alors qu'un exemple particulièrement représentatif de cet échec. Toutefois, si une spécificité doit lui être attribuée elle vient du fait que les taxinomies ne portent que sur des connaissances qualitatives¹. Dans un tel cas aucune métrique ne guide dans l'identification de ce qui doit être tenu pour un individu et de ce qui doit être tenu pour une propriété le caractérisant. La nature qualitative des connaissances auxquelles s'appliquent les taxinomies ne fait que compliquer l'analyse et à fortiori la formalisation.

1. En suivant un usage courant nous entendons par connaissance qualitative, toute connaissance non susceptible de se prêter à l'attribution de valeurs numériques de mesure.

Ce constat, relatif à l'échec de la logique classique, n'exclut pas la possibilité que d'autres formalismes logiques puissent remplir cette tâche. C'est explicitement ce que l'intelligence artificielle se propose, non seulement pour la formalisation du raisonnement taxinomique, mais aussi pour celle d'un ensemble de notions problématiques telles que celles d'action, de cause, d'obligation, de loi.

En informatique, les problèmes liés à la formalisation du raisonnement taxinomique sont connus sous le nom de "problèmes d'héritage". Ils concernent, outre les applications de l'intelligence artificielle et en particulier les réseaux sémantiques, des domaines tels que le génie logiciel, les langages de programmation de type "orientés objet" et les bases de données déductives. Les difficultés soulevées dans ces domaines par les taxinomies ne concernent pas seulement leur formalisation logique, mais aussi tout simplement leur interprétation. Elles constituent ainsi un cas typique des problèmes propres à l'intelligence artificielle : l'interprétation d'un procédé technique de notation conduit à s'interroger sur la nature des problèmes soulevés et sur la possibilité d'avoir recours à la logique pour les résoudre. La recherche de nouvelles logiques non classiques ne constitue qu'une hypothèse de recherche afin de donner une réponse positive à cette dernière interrogation.

Nous proposons, dans la première section de ce chapitre, une analyse générale de l'outil taxinomique. En particulier, nous essayons de montrer que les liens entre les concepts intervenant dans une taxinomie sont de deux types. L'un que l'on pourra appeler provisoirement 'nominal', relatif aux termes abstraits en tant que tels, l'autre 'objectif' et relatif à des individus concrets. Les difficultés que l'on rencontre dans la formalisation des taxinomies sont ainsi réductibles à l'ambivalence constitutive de cette structure conceptuelle et le problème des exceptions, si central pour les applications en intelligence artificielle, n'est qu'une conséquence de cette ambivalence fondamentale.

Dans la deuxième section le problème de l'héritage est exposé dans une de ses versions usuelles en intelligence artificielle. Nous nous proposons essentiellement d'indiquer la structure générale d'une hiérarchie d'héritage et les aspects qui soulèvent des problèmes quant à la formalisation logique. Les sections 2.2 et 2.3 considèrent deux stratégies de résolution du problème des exceptions. L'une est basée sur la logique non monotone, l'autre sur la logique linéaire. Il s'agit de deux formalismes opposés tant par l'approche de fond des questions de l'informatique que par les moyens techniques employés. Leurs résultats et leurs limites sont brièvement comparés dans le but d'indiquer les différents mécanismes logiques mis en œuvre en vue de la résolution du problème.

Sur la base de ces conclusions, nous montrons à quel point les problèmes posés par les hiérarchies d'héritage sont déterminants dans les rapports entre logique et intelligence artificielle : la solution de ces problèmes, qui ne peut pas être simplement réduite au choix d'une logique non classique particulière, semble au contraire appeler une analyse théorique plus poussée des notions mises en jeu.

1. Le raisonnement taxinomique

1.1. La taxinomie en tant qu'outil conceptuel

Deux aspects semblent coexister dans les taxinomies : d'un côté un aspect d'organisation rationnelle qui touche à la *hiérarchie* des concepts fondamentaux d'une discipline, de l'autre côté un aspect *classificatoire* qui concerne le regroupement des individus en classes par le biais de leurs caractéristiques. Selon le premier point de vue une taxinomie range des termes théoriques par leur spécificité : elle organise les éléments du lexique scientifique d'un domaine en les subsumant les uns sous les autres. Selon le second, elle met en relation l'abstrait (l'espèce biologique) et le concret (l'individu réel). Classer signifie créer des classes disjointes et il est nécessaire à cette fin de disposer d'une combinatoire de traits pertinents, tenus pour primitifs (avoir des vertèbres, allaiter ses petits, etc.). L'absence ou la présence d'un trait pertinent (avoir ou ne pas avoir de vertèbres) permet la constitution de classes distinctes de même niveau (vertébrés et invertébrés), et l'inclusion des traits pertinents permet d'ordonner une classe à l'autres (mammifères sous vertébrés).

Une taxinomie concrète est ainsi le résultat d'une tension entre une théorie rationnelle, organisant des termes abstraits selon leur spécificité, et un ensemble d'individus particuliers qu'il s'agit de classer et qui manifestent des propriétés saillantes entre lesquelles on devra choisir : doit-on fonder la classification sur les analogies morphologiques (la similarité structurelle des organes), les homologies morphologiques (la similarité fonctionnelle, mais non nécessairement structurelle), les caractéristiques phylogénétiques ou bien ontogénétiques des individus ? Cette tension est la manifestation de l'ajustement dynamique entre d'une part les exigences de rationalité et de symétrie que l'organisation hiérarchique d'un ensemble de termes théoriques est censée respecter, par le biais de la notion *qualitative* de spécificité, et d'autre part la nécessité de produire un outil pratique qui permette de réduire la variété des individus objets de l'expérience, à l'unité de la classification.

Ainsi une taxinomie, en tant que théorie rationnelle, ne s'intéresse pas directement à désigner des inclusions de classes d'individus. La relation de spécificité entre termes, qui fonde la structure hiérarchique, peut entrer en conflit avec la relation d'inclusion relative aux traits pertinents caractérisant les individus : un terme A peut être plus spécifique qu'un terme B, sans que les traits pertinents caractérisant A soient inclus dans ceux qui caractérisent B (un monotrème est un mammifère même s'il ne partage pas certains des caractéristiques des mammifères). En tant qu'instrument classificatoire, par contre, une taxinomie peut manquer de fondement dans ses grandes dichotomies et dans le choix des traits pertinents, sans comporter d'inconvénients majeurs sur le plan de l'utilisation pratique, qui consiste à ranger des individus de manière univoque².

2. L'histoire des sciences montre que le projet taxinomique n'a jamais cessé d'osciller entre ces deux pôles rationnel et classificatoire. La taxinomie aristotélicienne est une *théorie rationnelle* du vivant reliée à sa métaphysique et à sa cosmologie. A l'opposé la classification du *Systema Naturae* de Linné est un *outil classificatoire* permettant d'identifier et de nommer de façon univoque les animaux par leur genre et leur espèce. La théorie évolutionniste

Cette double nature théorique et classificatoire de la taxinomie se manifeste dans les deux caractéristiques les plus importantes de cet outil : *les restrictions* et *les exceptions*.

a) Le mécanisme des restrictions est interne à la taxinomie. Il permet d'harmoniser la structure hiérarchique des termes avec l'ordre d'inclusion des traits pertinents imposés par les classes d'individus, moyennant des restrictions sur les inclusions des propriétés. A la limite, A peut être considéré comme une spécialisation de B, même si aucun des attributs décrivant B ne se trouve dans la liste des attributs décrivant A : les A sont donc des B moyennant des restrictions sur leurs propriétés respectives. C'est à travers le mécanisme des restrictions que l'outil classificatoire parvient à s'articuler à l'aspect théorique de hiérarchie des concepts : nous pouvons avoir de bonnes raisons théoriques pour affirmer que les A sont un cas spécifique des B même si leurs attributs ne sont pas inclus les uns dans les autres. Ainsi Aristote mettait en relation les quadrupèdes vivipares avec les cétacés, bien que ces derniers ne soient évidemment pas des quadrupèdes. De même la classification des mammifères en trois catégories (les ordinaires, les marsupiaux et les monotrèmes) porte trace du fait que l'Australie a été découverte dans un temps relativement récent, mais elle nous permet de garder la structure originarie de la taxinomie, antérieure à cette découverte.

b) Les exceptions, en revanche, concernent le rapport avec les données qui sont externes au système taxinomique. Il est nécessaire de distinguer deux types d'exceptions.

i) D'une part une exception peut être la manifestation d'une méprise sur les conditions d'utilisation de la taxinomie. Quand nous parlons de classes d'individus rangés selon leurs traits caractéristiques, nous nous référons à un ensemble idéal, constitué par tous les individus présents, passés et futurs, constituant une classe d'équivalence. Cet ensemble n'est pas donné à la classification, et seule une idéalisation de celui-ci est effectivement prise en compte. En réalité une taxinomie ne fait que décrire les caractéristiques communes à des individus, ce qui les caractérise en tant qu'appartenant à une espèce, en tant que rangés sous un abstrait. Elle est censée décrire des individus dans des conditions normales d'observation (une théorie biologique ne considère pas les individus malades, ou ceux qui ont contracté dans leur existence des mutilations ou les mutations dues à leur environnement). Ces conditions sont dans la plupart des cas présumées sans être explicitées dans la taxinomie. On considère ainsi comme exceptionnel un individu qui ne correspond pas aux conditions normales implicites qui gouvernent l'application d'une taxinomie. Un oiseau qui aurait été amputé d'une aile constitue une exception dans ce sens restreint.

ii) D'autre part l'exception peut être la manifestation d'un véritable problème classificatoire : il existe un individu, observé dans des conditions jugées normales, qui ne correspond pas à la description prévue par la taxinomie. Il s'agit dans ce cas d'une véritable exception, d'un contre-exemple "falsifiant" la taxinomie.

darwinienne, au contraire, se veut une *classification scientifique* reposant sur la théorie de l'évolution et sur la notion de phylogénèse.

Ces deux mécanismes distincts, restriction et exception, peuvent se combiner afin de stabiliser la taxinomie : tout ensemble d'exceptions qui constitue une menace potentielle pour la taxinomie peut être réuni dans un nouvel abstrait classificatoire, moyennant le mécanisme des restrictions. Mais une différence essentielle demeure : les restrictions, étant internes à l'organisation taxinomique, sont connues à l'avance, au moment de l'application d'une taxinomie à un individu. Mais il n'en va pas de même pour les exceptions. D'une part parce que, la notion même de *conditions normales d'application* ou d'*individu normal* est par sa nature même incomplète et problématique. D'autre part parce que, bien évidemment, nous ne pouvons pas prévoir à l'avance de quelle manière "le monde" sera réticent ou conciliant vis-à-vis de nos schèmes conceptuels.

La thèse que je soutiens est que la double fonctionnalité, théorique et classificatoire, de la taxinomie, se manifeste par la nature fondamentalement ambiguë du lien taxinomique, tantôt exprimant la spécificité des termes abstraits, tantôt exprimant l'inclusion classificatoire. Les formes de prédication exprimées par le lien taxinomique sont multiples et seulement partiellement réductibles à celles de la logique. C'est bien ceci qui provoque les difficultés de formalisation dans la logique classique, et c'est ce lien que nous examinons de plus près dans la section suivante.

1.2. Trois formes de prédication

La question qui sera traitée dans cette section est préliminaire à l'analyse logique des structures d'héritage. Elle concerne la nature du lien taxinomique exprimé par le verbe "être" dans les deux expressions : les *A* sont des *B* et *a* est un *B*, où *A* et *B* sont des abstraits, ou termes généraux, et *a* est un individu rangé sous un abstrait.

Nous nous intéressons en fait à la nature du décalage entre ces expressions et les formes de prédication exprimables par la logique. Celle-ci enrégimente les expressions contenant le verbe "être" en deux formes qui du point de vue de la théorie des ensembles correspondent l'une à l'appartenance d'un individu à un ensemble et l'autre à l'inclusion d'un ensemble dans un autre. La première forme de prédication s'exprime donc par le fait qu'une constante est argument d'un prédicat (par exemple la constante *a* et le prédicat *P* dans *P(a)*) ; l'autre par une formule complexe contenant un conditionnel (symbolisé par $\rightarrow$) et un quantificateur universel (symbolisé par $\forall$) : $\forall x(P(x) \rightarrow Q(x))$.

La logique classique naît du rejet des termes abstraits ou génériques³, précisément les termes auxquels nous nous intéressons ici. Rien ne nous oblige à nous conformer aux formes de prédication qu'elle propose, en particulier quand la

3. Frege présente comme un des objectifs fondamentaux de son Idéographie celui de se libérer de l'ambiguïté du langage ordinaire. Pour évoquer cette ambiguïté il donne comme exemple le nom cheval, tantôt utilisé comme désignant un individu, tantôt comme désignant l'espèce naturelle, (dans le cas de "le cheval est herbivore"), tantôt dit-il pour désigner le concept (dans le cas de "ceci est un cheval").

taxinomie est utilisée dans le raisonnement dit de sens commun, pour ranger des connaissances pré-scientifiques parfois incomplètes qui concernent un domaine d'individus non préalablement fixé. Mais avant de chercher de nouveaux formalismes qui s'écarteraient plus au moins de la logique classique, il nous faut parvenir à des structures plus élémentaires de prédication, auxquelles la formalisation doit s'attacher.

Dans son article [Vuillemin 84]⁴ Jules Vuillemin, s'inspirant des travaux de Benveniste et Meillet, distingue parmi les formes fondamentales de la prédication trois types qui semblent appropriés pour notre analyse : la prédication nominale, la prédication substantielle et la prédication accidentelle.

La première est illustrée par la phrase :

1) "L'humilité (est) une vertu"

Les parenthèses autour du verbe "est" indiquent le fait qu'aucune détermination temporelle ou modale n'affecte la simple subsumption du premier terme au second. Un énoncé mettant en jeu une prédication nominale est en fait "un énoncé composé par deux noms : le nom de l'universel et le nom de l'individu, qui peut être et est généralement un abstrait. En latin *humilitas virtus*. Cette phrase s'oppose à l'énoncé avec copule : *humilitas est virtus* qui signifierait plutôt : *il y a une vertu d'humilité*, où les mots *il y a* ont un sens existentiel et non seulement classificatoire". L'affirmation de 1) ne dit rien sur le monde, elle ne désigne pas une classe d'individus, elle ne porte pas sur les objets du monde, mais plutôt sur la structure des termes abstraits du lexique que nous utilisons pour accéder au monde. On pourrait la paraphraser par : 'L'abstrait humilité est une spécification de l'abstrait vertu'.

La logique classique, qui n'admet que les deux type de prédications $P(a)$ et $\forall x(P(x) \rightarrow Q(x))$, traduit l'énoncé 1) par une expression du type :

2) "Tous ceux qui sont humbles sont vertueux",

mais cette traduction apporte évidemment un glissement des termes abstraits vers les individus. Au contraire l'énoncé 1) ne porte pas sur des individus concrets et le mot abstrait 'vertu' ne correspond pas au prédicat $V(x)$. Par conséquent l'affirmation de 1) ne contraint nullement à l'affirmation de 2), le premier énoncé pouvant être considéré comme vrai, même si tous les individus humbles dans notre univers ne sont pas vertueux. Si 1) n'exprime rien sur les individus et ne détermine pas une partition sur l'ensemble des objets dont il est question, celui qui prononce 1) ne s'engage pas à sa traduction extensionnelle. Il s'engage plutôt à pouvoir spécifier, *en ligne de principe*, les conditions de détermination des individus qui dès lors qu'ils sont humbles sont aussi vertueux.

La deuxième caractéristique de la prédication nominale, est que sa négation ne supporte pas de contraposition, car la négation d'un mot, à la différence de la négation d'un prédicat, n'est pas son complément. La donnée même du code linguistique spécifie quels types d'objets peuvent être considérés comme satisfaisant

4. La même classification est présentée dans [Vuillemin 1986].

la négation : la clôture de l'univers du discours est telle qu'un nombre ne peut pas être compté parmi les non-vertus : ce n'est que par la connaissance du code que nous apprenons à clore correctement l'univers du discours et donc à désigner correctement ce qui appartient à la négation d'un mot.

Si la phrase nominale véhicule un message purement conceptuel, la prédication substantielle et la prédication accidentelle, formes objectives de la prédication, concernent en revanche les individus du monde, et elles présupposent leur identification et re-identification dans l'espace et le temps. La prédication substantielle qualifie les propriétés d'un individu qui permettent son identification dans le temps : "Socrate est un homme", qualifie Socrate comme individu, tout au long de son existence. Les propriétés accidentelles, par contre sont soumises à la vérification temporelle : "Socrate court" est tantôt vrai tantôt faux. Les propriétés accidentelles ne caractérisent pas l'objet, mais certains "accidents" qui lui sont attachés.

On peut maintenant analyser le lien taxinomique à partir de ces formes de la prédication.

i) D'abord, une taxinomie est une *théorie rationnelle* organisant les abstraits du lexique par le lien de spécificité. (A est B au sens de la prédication nominale).

ii) En second lieu, une taxinomie est un *outil de classification* rangeant des individus sous un abstrait. (L'individu *a* est B au sens de la prédication substantielle)⁵.

Toutefois, comme nous l'avons vu lorsque nous avons discuté de la tension entre nature rationnelle et classificatoire de la taxinomie, la correspondance entre la hiérarchie des abstraits et l'ordre d'inclusion des classes peut ne pas être totale. Le mécanisme des restrictions intervient pour harmoniser ces éventuelles différences.

iii) En troisième lieu, la donnée des propriétés substantielles (ou encore prototypiques, ou normales) présuppose que l'on a fait abstraction des propriétés accidentelles des individus. Ces dernières peuvent être en contradiction avec certaines des propriétés substantielles prévues par la taxinomie. Un individu *a* peut être substantiellement un B, tout en n'ayant pas l'un des traits pertinents définissant les B, car *a* peut ne pas être un individu normal. Un tel cas ne compromet pas la cohérence du raisonnement taxinomique, car les propriétés substantielles et les propriétés accidentelles se situent sur des plans différents et non conflictuels.

La référence aux propriétés accidentelles et substantielles ne doit pas être considérée ici comme hors de propos. Etant donné que la taxinomie appartient à une phase de développement qualitatif d'une science il est clair que l'utilisation de formes de prédication différentes de celles utilisées par les mathématiques s'impose. A la différence des concepts empiriques de la science, les concepts empiriques naturels ne peuvent pas être institués par une métrique. Il faut donc s'interroger sur

5. L'ensemble des *a* qui sont des *B* constitue donc un ensemble d'individus qui partagent les mêmes caractéristiques substantielles. En ce sens la prédication substantielle indique des véritables inclusions de classes d'individus.

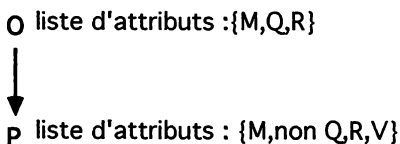
les moyens qui, dans le cas des concepts qualitatifs, peuvent remplir le rôle joué par les théories de la mesure dans le cas des concepts quantitatifs. La logique ne peut s'appliquer au raisonnement taxinomique qu'au prix de la reintroduction de la distinction entre substance et accident ou d'une distinction analogues.

Quand nous rangeons un individu dans une taxinomie nous cherchons d'abord l'abstrait dont les traits pertinents lui correspondent, en parcourant ainsi la taxinomie du bas vers le haut et du haut vers le bas. S'il y a décalage entre l'individu et les espèces classées par leur traits pertinents nous en cherchons une justification. Nous tâchons de considérer comme accidentels les traits qui ne correspondent pas aux traits prescrits par la taxinomie. Cet ajustement progressif de l'individu à son image normale, et cette appréciation de ce qui concerne son individualité et de ce qui relève de sa généralité, est le type de savoir intelligent que demande l'application d'une taxinomie. L'analyse de ce processus est non seulement souhaitable mais nécessaire quand on se place, comme le fait l'intelligence artificielle, dans le cadre des changements des théories, si important dans l'étude de ce que l'on y appelle apprentissage, et qui est tenu pour être le cœur des comportements intelligents dans les systèmes artificielles. L'interprétation des notions de spécificité, d'individu normal, d'exception, de terme générique constitue le vrai problème philosophique soulevé par les taxinomies, dont les problèmes de formalisation, compliqués par la forme particulière qu'elles assument dans le cadre des hiérarchies d'héritage en intelligence artificielle, ne sont qu'une conséquence.

1.3. Les taxinomies et la logique

La distinction proposée par Vuillemin constitue un essai de classification des différentes formes que prend la prédication dans la langue. Elle ne constitue pas une analyse de leurs propriétés logiques. Toutefois le hiatus existant entre ces formes de la prédication et celles que reconnaît la logique classique en sort confirmé. L'enrégimentation induite par cette dernière oriente l'espace de recherche vers des solutions basées sur des logiques non classiques. Dès lors qu'elles retiennent les formes de la prédication propres à la logique classique, toutes les éventuelles modifications des mécanismes déductifs susceptibles d'être offertes par les logiques non classiques seront inadaptées au traitement de notre question.

Considérons deux termes abstraits, désignés par P et O, et supposons qu'une prédication nominale soit établie entre eux : "P est O" au sens nominal. Soient ensuite M, Q et R les traits pertinents caractérisant O ; soit V le trait pertinent caractérisant proprement P et soit Q une restriction dans P des propriétés héritées de O. Les P sont donc des O ayant le trait caractéristique V et pour lesquels la propriété Q est restreinte. La situation peut être résumée par le schéma suivant.



Soit b un individu particulier, rangé sous P , mais qui fait exception pour la propriété R . Une formalisation classique, représentée par les axiomes 1, 2 et 3 suivants rend trivial cet ensemble d'informations.

- (1) $O(x) \vdash M(x) \wedge Q(x) \wedge R(x)$
- (2) $P(x) \vdash O(x) \wedge M(x) \wedge \neg Q(x) \wedge V(x)$
- (3) $\vdash P(b) \wedge \neg R(b)$

En fait nous pouvons déduire l'absurde à cause de la présence des restrictions :

- (4) $O(x) \vdash Q(x)$ {de 1 par définition de $\wedge$ }
- (5) $P(x) \vdash \neg Q(x)$ {de 2 par définition de $\wedge$ }
- (6) $P(x) \vdash O(x)$ {de 2 par définition de $\wedge$ }
- (7) $P(x) \vdash Q(x)$ {de 6 et 4 par transitivité de $\vdash$ }
- (8) $P(x) \vdash \perp$ {de 5 et 7 par déf. de $\perp$ }

et à cause de la présence des exceptions :

- (9) $O(b) \vdash R(b)$ {de 1 par instanciation et définition de $\wedge$ }
- (10) $P(b) \vdash O(b)$ {de 2 par instanciation et définition de $\wedge$ }
- (11) $P(b) \vdash R(b)$ {de 9 et 10 par transitivité de $\vdash$ }
- (11) $P(b) \vdash \perp$ {de 3 et 11 par définition de $\wedge$ et par transitivité de O }

D'un point de vue formel la déduction de l'absurde est entraînée par les propriétés de la conjonction, par la transitivité de la relation de déduction classique et par la définition de la négation. Du point de vue de l'analyse du problème, toutefois, nous avons formalisé le lien nominal entre les deux abstraits P et O de la même manière que le lien substantiel entre un abstrait et les traits pertinents le caractérisant. La caractérisation d'un objet b par une propriété R , qu'elle soit substantielle ou accidentelle, est aussi exprimée de la même manière. Les restrictions et les exceptions sont traitées symétriquement par le biais d'une négation classique. Etant donnée la configuration logique du problème, deux solutions ont été envisagées : l'une visant à la modification des propriétés de l'implication, l'autre à la modification de la négation et, éventuellement, de la conjonction.

Selon la première stratégie, le conditionnel classique et le conditionnel taxinomique se différencient par la notion d'exception. Par exception on entend ici les exceptions entre classes (que nous avons dénommées restrictions), et les véritables exceptions qui peuvent être reconnues entre un individu et une classe. Ainsi l'on recherche un conditionnel intentionnel capable de formaliser le lien taxinomique et interprété comme un générique du type :

Les individus typiques A sont des B , ou bien

Les A sont des B , sauf exceptions, ou encore

Normalement les A sont des B .

Un tel conditionnel devrait exprimer la notion de spécificité, propre à la prédication nominale, la notion d'instanciation à un individu concret et à ses propriétés substantielles, propre à la prédication substantielle et il devrait supporter à la fois restrictions et exceptions. Une fois identifiées les prédications substantielles, nominales et accidentelles, le maintien de la cohérence devient le problème majeur des formalismes non classiques, alors qu'il est secondaire du point de vue de l'analyse du problème (cf. section 1.2, point iii).

Nous présentons dans la section 2.2 deux formalismes logiques qui se proposent de formaliser des conditionnels génériques. Le premier fait appel à des règles introduisant directement des tests de consistance. Le deuxième définit des logiques des conditionnels (du type [Lewis 75]), dont le pouvoir déductif est extrêmement pauvre et qui constituent à leur tour une base pour des extensions obtenues par des tests de consistance.

La deuxième stratégie qui a été envisagée considère que les restrictions et les exceptions, ne doivent pas être traduites par des négations. Le fait qu'une propriété soit absente relativement à un individu ou à une classe ne doit pas être traduit par la négation logique de telle propriété. Nous traiterons une solution de ce type dans la section 2.3.

2. La formalisation logique des hiérarchies d'héritage

Nous pouvons désigner par hiérarchies d'héritage les taxinomies qui sont utilisées en intelligence artificielle. Elles sont définies en termes de graphes et constituent des procédés techniques de stockage d'information. Leur construction, et donc leur interprétation, ne semblent pas être régies par des principes généraux, mais leur structure n'est pas neutre par rapports aux questions que nous avons posées dans le paragraphe précédent. Cette structure élimine la distinction entre abstrait et propriétés d'un abstrait : être un oiseau et avoir la propriété de voler sont traités comme nœuds de la hiérarchie. Elle élimine aussi la distinction entre restriction et exception et confond le lien entre un individu et une espèce et le rapport de spécificité entre espèces. Ainsi, le problème de la formalisation logique des hiérarchies d'héritage paraît mal engagé dès le départ, car l'objet même de la formalisation ne semble pas bien défini. Soit la formalisation a comme objet la représentation de toutes les stratégies possibles de parcours d'un graphe, soit elle impose une interprétation particulière difficile à évaluer en l'absence d'une analyse claire de ce qu'une hiérarchie représente. Néanmoins les solutions logiques à l'interprétation des hiérarchies d'héritage demeurent intéressantes pour deux raisons. En premier lieu les limites des solutions logiques peuvent nous montrer les raisons d'une surestimation fréquente des tâches qui peuvent être menées à bien par la logique. Elles nous fournissent des éléments d'analyse sur les limites et sur la nature des logiques non classiques. En second lieu toute formalisation, bonne ou mauvaise, entraîne un choix interprétatif des liens taxinomiques et donc une sélection parmi les parcours possibles de la hiérarchie. Ainsi les alternatives des formalisations possibles nous parlent à leur tour de la nature de l'instrument taxinomique.

La distinction entre une bonne et une mauvaise formalisation, que nous venons de mentionner, appelle quelques mises au point sur les critères généraux auxquels il faudrait s'en tenir par la suite. Sur la base du simple bon sens, nous considérons qu'une bonne formalisation devrait constituer un cadre d'interprétation général, simple et logique du problème des hiérarchies. Le cadre devrait être "général" au sens où la solution n'avantage pas un type particulier de hiérarchie sans justification ; il devrait être "simple" de telle sorte que les notions d'individu et d'espèce et de lien conservent un certain caractère intuitif, la formalisation logique ne devant pas les compliquer *ad hoc* ; il devrait être "logique" enfin, dans le sens où la formalisation fait effectivement usage des moyens déductifs de la logique. Tout recours à des moyens de calcul autres, faisant intervenir des instructions extra-logiques, devrait être considéré comme non acceptable.

2.1. Les hiérarchies d'héritage

Une hiérarchie d'héritage est un graphe orienté acyclique ayant comme nœuds un ensemble d'objets. Les objets sont de deux types : les individus, ne pouvant constituer que les feuilles du graphe, notés par les lettres a,b,c, etc. et les espèces, correspondant à des abstraits et notés par les lettres p,q,r, etc. Nous noterons x,v,w,y,z les variables qui varient sur l'ensemble des objets. Les nœuds du graphe sont reliés par deux types de liens : les liens positifs ($x \rightarrow y$) et les liens négatifs ($x \nrightarrow y$) où y doit être une espèce. Les hiérarchies, à la différence de la logique, ne distinguent pas la prédication d'une espèce à un individu de la prédication d'un individu à une espèce sinon par le fait qu'un individu ne peut pas être lié à un autre individu. Deux liens seront dit conflictuels s'ils ont la forme $a \rightarrow p$ et $a \nrightarrow p$ ou $p \rightarrow q$ et $p \nrightarrow q$.

Parmi les suites possibles de liens, les chemins constituent des parcours admissibles dans le graphe. Par définition tout lien constitue un chemin, et si σ est un chemin comprenant uniquement des liens positifs alors $\sigma \rightarrow y$ et $\sigma \nrightarrow y$ sont aussi des chemins.

Il s'ensuit qu'un individu ne peut figurer dans un chemin qu'en tant qu'objet de départ ; et par ailleurs qu'un lien négatif ne peut figurer dans un chemin qu'à la fin.

Par analogie avec les liens, deux chemins seront dits conflictuels s'ils ont la forme : $x \rightarrow \sigma \rightarrow y$ et $x \rightarrow \tau \nrightarrow y$.

Un chemin du graphe correspond à une chaîne de raisonnement. De manière intuitive on peut dire que l'affirmation "x est un y" est *supportée* par une hiérarchie s'il existe un chemin dans la hiérarchie de x à y. En présence de conflits de hiérarchie peut toutefois supporter des affirmations contradictoires. Il est alors nécessaire de mettre en œuvre une stratégie. Nous traiterons la question plus loin.

On peut distinguer deux alternatives fondamentales en ce qui concerne le traitement des théories d'héritage.

La première concerne l'interprétation de la hiérarchie.

Une première approche traduit le graphe en un formalisme logique intermédiaire dans lequel les liens objet-espèce sont traduits par des prédicats s'appliquant à des constantes, et les liens espèce-espèce sont traduits par un conditionnel d'une logique particulière. Les conditionnels ne sont pas préfixés par des quantificateurs, car le calcul sur la hiérarchie étant toujours relatif à un objet donné, le fragment propositionnel suffit. L'ensemble ou les ensembles d'assertions supportées correspond ainsi à l'ensemble des propositions déductibles par la logique [Etherington 87], [Boutilier 90], [Fouquerre et al. 93].

Une deuxième approche conserve par contre le langage de la théorie des graphes, sans le traduire par un formalisme logique intermédiaire, l'ensemble des assertions supportées par la hiérarchie étant l'ensemble des chemins admissibles pour une stratégie donnée. Il y a une évidente analogie avec la notion de déduction, mais la théorie est extraite directement en termes de liens et chemins [Touretzky 86], [Horthy et al. 87]

La deuxième alternative à propos du traitement des hiérarchies, concerne le type de stratégie de parcours du graphe, en présence de chemins conflictuels. Une stratégie dite *crédule* associe à la même hiérarchie différentes extensions, chacune représentant un ensemble maximal de chemins non conflictuels [Touresky 86], [Etherington 88], [Fouquerre et al. 93]. Tout ensemble de chemins non conflictuels est considéré aussi plausible qu'un autre. L'approche *crédule* retient toutes les alternative plausibles, c'est-à-dire supportées par la hiérarchie.

Un approche *sceptique* par contre associe à une hiérarchie une seule extension, choisissant un sous-ensemble de chemins ne contenant pas de chemins conflictuels. Les chemins conflictuels sont rejetés si on ne dispose pas d'information supplémentaire, susceptible de résoudre le conflit en faveur de l'un d'entre eux. Les stratégies sceptiques les plus sophistiquées ont été définies par [Horthy et al. 87] [Stein 89], et sont basées sur des définitions récursives de l'ensemble des chemins admis. Ces stratégies mettent en œuvre de mécanismes différentes de propager l'attitude sceptique (le refus des chemins conflictuels), dans le parcours de la hiérarchie.

Différents mécanismes de résolution de conflits peuvent être mis en œuvre. L'un des plus connus est le mécanisme dit de la spécificité. Etant donné deux chemins contradictoires un critère de préférence basé sur des considérations relatives à la forme de la hiérarchie, résout le conflit en faveur de l'un d'eux. Intuitivement le critère de préférence peut être énoncé comme suit : soit y et z les nœuds de départ des derniers liens de deux chemins contradictoires, partant de x. S'il est possible d'établir que l'objet y est plus spécifique que l'objet z, la contradiction est levée en faveur du chemin passant par y. De manière intuitive, un objet y est plus spécifique qu'un objet z par rapport à x, s'il existe un chemin admis menant de x à z qui passe par y.

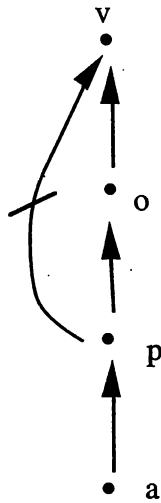


Figure 1.

Dans la figure 1, l'objet "p" est plus spécifique que l'objet "o", le chemin $a \rightarrow p \rightarrow v$ sera préféré au chemin $a \rightarrow o \rightarrow v$.

Le mécanisme de spécificité peut s'appliquer à une stratégie crédule ou sceptique. Les stratégies proprement sceptiques mettent en œuvre des mécanismes de résolution de conflits basés essentiellement sur la forme des chemins.

D'autres alternatives concernent la définition de la notion d'exception, les types de parcours (chaînage avant et chaînage arrière des liens), la présence ou à l'absence de liens redondants, la présence d'autres types de liens dans la hiérarchie (liens stricts et liens exceptionnels). L'importance de leur nombre et de leur variété a conduit certains auteurs [Touretzky et al. 87] à parler d'un conflit d'interprétation que la simple intuition sur la signification des liens, ne peut pas résoudre. Il semble évident que les hiérarchies d'héritage en tant que telles sont des objets non interprétés, s'appuyant sur une ambiguïté constitutive du lien taxinomique. Les formalisations que nous considérons par la suite en proposent des interprétations, qui plutôt que relever d'une analyse générale semblent être liées aux moyens expressifs de ces logiques. Ainsi, d'une part le rêve d'une logique neutre par rapport aux différences de construction et de stratégie semble s'éloigner, d'autre part aucune interprétation ou restriction fournie par les logiques non classiques, ne semble être plus justifiée qu'une autre, face à ces objets définis seulement partiellement.

La première des formalisations que nous examinons par la suite, est basée sur les formalismes non monotones et s'attache à exprimer un conditionnel générique, capable de supporter des exceptions (2.2). La deuxième formalisation est basée sur la logique linéaire et s'attache à exprimer les restrictions et les exceptions par des constructions du langage différentes de la négation (2.3).

2.2. Les solutions non monotones

Les logiques non monotones ont été développées dans le but de formaliser le raisonnement dit révisable, s'appuyant sur des informations incomplètes. Leur nom vient du fait que la relation de conséquence logique qui les définit ne possède pas la propriété de monotonie. Cette dernière peut être énoncée comme suit : soient T et T' deux ensembles de formules, et soient $Cn(T)$, $Cn(T')$ les ensembles respectifs de leurs conséquences logiques si $T \subseteq T'$ alors $Cn(T) \subseteq Cn(T')$. La propriété de monotonie est souvent confondue avec une propriété syntaxique concernant la relation de déduction. Il s'agit de l'affaiblissement (parfois aussi appelé monotonie), règle qui permet d'augmenter les prémisses :

$$\frac{X, Y \vdash Z}{X, A, Y \vdash Z} \text{ ou les conclusions : } \frac{X \vdash Y, Z}{X \vdash Y, A, Z} \text{ d'une déduction.}$$

La confusion est regrettable, car si, en général, les relations de déduction des logiques non monotones ne sont pas fermées sous la règle d'affaiblissement, c'est aussi le cas de bon nombre de relations de déduction monotones [Anderson Belnap 75].

Deux types de systèmes non monotones peuvent être définis. Le premier contient la logique des défauts [Reiter 80], les logiques modales non monotones [McDermott et al. 80], les logiques autoépistémiques [Moore 84], la circonscription [McCarthy, 80]. Ces systèmes intègrent directement dans le calcul syntaxique la notion de consistance sémantique. Le deuxième est celui des systèmes d'inférence non monotones, équivalents à des sous-systèmes de logiques des conditionnels [Lewis 75], [Stalnaker 68]. Il s'agit donc de logiques modales ayant une implication stricte non transitive.

2.2.1. Les défauts

L'approche basée sur la logique des défauts [Etherington et al. 83] [Etherington 87], [Etherington 88] est peut être la plus générale et la plus utilisée dans la formalisation des systèmes d'héritage. Nous présentons la version standard utilisée pour la formalisation des stratégies crédules et les problèmes liés à leur extension au principe de spécificité.

Un défaut est une règle de la forme $\frac{p:q}{r}$ où p est dite pré-condition, q justification et r conséquent. La lecture intuitive d'un défaut est du type : si p alors r , si q est justifié, la justification de q étant sa consistance avec la théorie courante.

Soit Γ une hiérarchie d'héritage telle qu'elle a été définie plus haut, (section 2.1), $U \vdash \Gamma$ l'ensemble d'objets qui sont des individus dans Γ et $\rightarrow$, $\nrightarrow$ respectivement les liens positifs et négatifs de la hiérarchie. $T(\Gamma)$, la théorie des défauts correspondant à la hiérarchie, est un couple : $T(\Gamma) = \langle W(\Gamma), D(\Gamma) \rangle$.

$W(\Gamma) = \{p(a) : a \rightarrow p \in \Gamma\} \cup \{\neg p(a) : a \rightarrow \neg p \in \Gamma\}$, et

$D(\Gamma) = \left\{ \frac{p(a) : q(a)}{q(a)} : p \rightarrow q \in \Gamma \text{ et } a \in U_\Gamma \right\} \cup \left\{ \frac{p(a) : \neg q(a)}{\neg q(a)} : p \rightarrow \neg q \in \Gamma \text{ et } a \in U_\Gamma \right\}$.

Une théorie de défaut ainsi définie donne lieu à plusieurs extensions, c'est-à-dire à plusieurs ensembles consistants de conclusions obtenues par application des défauts.

Une extension peut être ainsi définie :

E est une extension d'une théorie $T(\Gamma)$ ssi $\Delta(E) = E$ où $\Delta(E)$ satisfait les conditions suivantes:

1) $W(\Gamma) \subseteq \Delta(E)$ 2) $\Delta(E)$ est fermé par déduction

3) si $\frac{p : q}{r} \in D(\Gamma)$ et $p \in \Delta(E)$ et $\neg q \notin \Delta(E)$ alors $r \in \Delta(E)$.

La clause 3 contient le test de consistance. Si q est consistant avec la théorie E (ce qui correspond à la condition $\neg q \notin E$) alors la règle par défaut correspondante peut être déclenchée. Il est bien connu que le problème de la satisfiabilité d'une formule dans la logique du premier ordre est indécidable. L'ensemble des non-théorèmes n'est pas récursivement énumérable. Toutefois le langage utilisé pour les théories d'héritage est celui de la logique des propositions, et donc les théories d'héritage ne contiennent pas des lois générales mais exclusivement des instances de celles-ci relativement à un nombre fini d'individus⁶.

Ce type de théorie s'adapte bien aux hiérarchies d'héritage contenant des inconsistances directes, celles qui sont appelées dans la littérature "le diamant de Nixon", illustré par la figure 2. La théorie correspondant à cette hiérarchie est constituée par :

$W(\Gamma) = \{r(a), q(a)\}$, et

$D(\Gamma) = \left\{ \frac{q(a) : p(a)}{p(a)}, \frac{r(a) : \neg p(a)}{\neg p(a)} \right\}$. Les deux extensions requises sont

$E = \{q(a), r(a), \neg p(a)\}$, et $\{q(a), r(a), p(a)\}$.

6. Il est clair que l'utilisation de tests de satisfiabilité pour le calcul des propositions demeure problématique comme l'atteste le fait que le problème de la satisfiabilité en calcul des propositions est un problème NP-complet.

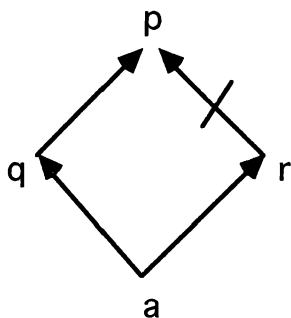


Figure 2.

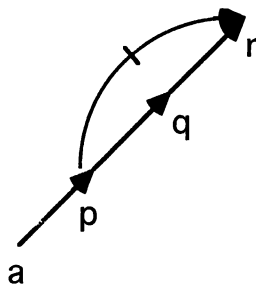


Figure 3.

Comme nous l'avons annoncé, les théories des défauts des hiérarchies ne sont pas adaptées aux stratégies sceptiques ni au principe de spécification.

La figure 3 montre un exemple de hiérarchie dans laquelle le chemin $a \rightarrow p \rightarrow r$ est préféré au chemin $a \rightarrow p \rightarrow q \rightarrow r$ car p est plus spécifique que q , mais l'ensemble des extensions obtenues par la théorie des défauts contient deux éléments :

$E' = \{p(a), q(a), \neg r(a)\}$, $E'' = \{p(a), q(a), r(a)\}$. Pour résoudre cette difficulté, Etherington a proposé des défauts semi-normaux, c'est-à-dire des règles de défauts dans lesquelles la justification n'est pas égale à la conclusion, mais l'implique logiquement.

La hiérarchie de la figure 3 serait donc formalisée par la théorie suivante : $W(\Gamma) = \{p(a)\}$, et

$$D(\Gamma) = \left\{ \frac{p(a) : \neg r(a)}{\neg r(a)}, \frac{q(a) : r(a) \wedge \neg p(a)}{r(a)} \right\} \text{ et la seule extension est alors :}$$

$$E = \{p(a), \neg r(a), q(a)\}.$$

La règle $\frac{q(a) : r(a) \wedge \neg p(a)}{r(a)}$ formalise directement le principe de spécificité, mais cette solution ne peut pas être étendue à toutes les hiérarchies. Le problème naît du fait qu'un nœud peut être plus spécifique qu'un autre dans une extension, sans l'être dans toutes les extensions. Considérons l'exemple illustré par la figure 4 [Thomason et al. 88].

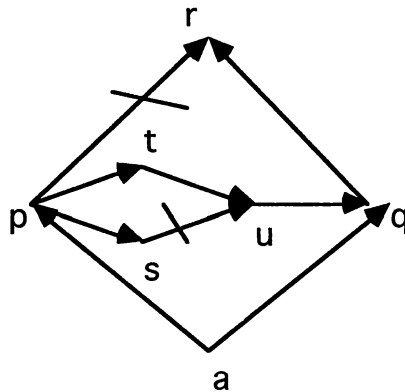


Figure 4.

Le chemin $a \rightarrow p \rightarrow r$ sera préféré au chemin $a \rightarrow q \rightarrow r$ dans l'extension contenant le chemin $a \rightarrow p \rightarrow t \rightarrow u \rightarrow q$. Mais il existe une extension contenant le chemin $a \rightarrow p \rightarrow s \rightarrow u$ dans laquelle p n'est pas plus spécifique que q car elle ne contient pas de chemins menant à q et passant par p .

La formalisation correcte de cette hiérarchie, par la logique des défauts, ne contient donc pas la règle :

$$\text{pas la règle : } \frac{q(a) : r(a) \wedge \neg p(a)}{r(a)}, \text{ mais plutôt la règle : } \frac{q(a) : r(a) \wedge (u(a) \rightarrow \neg p(a))}{r(a)}.$$

Toutefois pour formaliser la hiérarchie il a fallu connaître à l'avance les extensions possibles en termes de chemins du graphe, ce qui représente une forte restriction à la prétention qu'une formalisation interprète directement une hiérarchie d'héritage.

2.2.2. Systèmes d'inférence non monotone et logiques des conditionnels

La difficulté de comparer le pouvoir déductif des différentes logiques non monotones, d'en étudier les propriétés et de donner une sémantique claire au raisonnement dit révisable, a conduit autour des années 80 à la définition des systèmes d'inférence non monotones [Gabbay 85], [Makinson 88]. Il existe une traduction simple [Crocco 93] de ces systèmes dans le fragment non imbriqué de certaines logiques des conditionnels, une famille de logiques modales, dont le langage est constitué par le langage de la logique classique, augmenté d'une implication stricte (que nous symbolisons par la suite par ' $\Rightarrow$ ') et pouvant s'interpréter dans le cadre de la sémantique des mondes possibles. La transitivité, $((A \Rightarrow B) \wedge (B \Rightarrow C)) \rightarrow (A \Rightarrow C)$, le Modus Ponens, l'augmentation de l'antécédent, $(A \Rightarrow B) \rightarrow (A \wedge C \Rightarrow B)$, les lois de contraposition, $((\neg B \Rightarrow \neg A) \rightarrow (A \Rightarrow B))$ et $(A \Rightarrow B) \rightarrow (\neg B \Rightarrow \neg A)$, ne sont pas des inférences valides pour ces conditionnels. Par conséquent ces logiques, et les systèmes d'inférence non monotones auxquels elles sont équivalents, s'avèrent déductivement très pauvres. C'est ainsi que les

logiques des conditionnels, qui sont monotones, sont utilisées comme base de formalismes non monotones, c'est-à-dire comme base pour la construction d'extensions obtenues faisant intervenir des tests de consistance [Delgrande 87,88], [Pearl 90], [Lehman 89]. De manière générale les tests de consistance permettent d'obtenir un ensemble de conclusions plus grand que celui qui est déductible par la logique conditionnelle. Ils permettent de considérer le conditionnel " $\Rightarrow$ " comme transitif dans tout les cas où ceci n'entraîne pas d'incohérences [Lamarre 92], [Pearl 90], [Lehman 89].

Nous ne considérerons pas de traductions complètes et directes dans ces formalismes des hiérarchies d'héritage. De nombreuses difficultés techniques sont soulevées par le choix d'un langage logique suffisamment restreint pour ne générer que l'information taxinomique attendue [Boutilier 90]. Nous nous limitons à indiquer quelques problèmes de nature moins technique soulevés par l'interprétation des taxinomies à l'aide de ces logiques.

La logique des conditionnels à laquelle nous nous référons est apparentée avec le système NP, défini par Delgrande, et avec ce qu'on appelle dans la littérature le "système d'inférence rationnel". Elle est définie par les axiomes et les règles d'inférence de la logique classique plus les axiomes suivant :

$$\text{RCK} : \frac{B1 \wedge \dots \wedge Bn \rightarrow B}{(A \Rightarrow B1 \wedge \dots \wedge A \Rightarrow Bn) \rightarrow (A \Rightarrow B)} \quad n \geq 0$$

$$\text{ASC} : ((A \Rightarrow B) \wedge (A \Rightarrow C)) \rightarrow (A \wedge B \Rightarrow C)$$

$$\text{RT} : ((A \Rightarrow B) \wedge (A \wedge B \Rightarrow C)) \rightarrow (A \Rightarrow C)$$

$$\text{CA} : ((A \Rightarrow B) \wedge (C \Rightarrow B)) \rightarrow (A \vee C \Rightarrow B)$$

$$\text{T} : (\neg A \Rightarrow A) \rightarrow A$$

$$\text{CV} : ((A \Rightarrow C) \wedge \neg(A \Rightarrow \neg B)) \rightarrow (A \wedge B \Rightarrow C).$$

La sémantique de cette logique peut être définie en termes de mondes possibles et d'une relation d'accessibilité entre ensembles de mondes, qui, entre autres, est réflexive, transitive et connexe.

Les liens positifs $p \rightarrow q$ d'une hiérarchie sont traduits par la formule conditionnelle du langage $p \Rightarrow q$; les liens négatifs $p \nrightarrow q$ par la formule $p \Rightarrow \neg q$.

L'axiome RT exprime directement le principe de spécificité.

L'axiome ASC est fondamental pour pouvoir donner une sémantique en termes de relation d'accessibilité entre ensembles de mondes, mais il impose une propriété particulière de stabilité aux hiérarchies interprétés.

Considérons l'exemple des deux hiérarchies représentées par les figures 5 et 6.

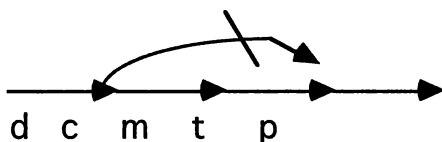


Figure 5.

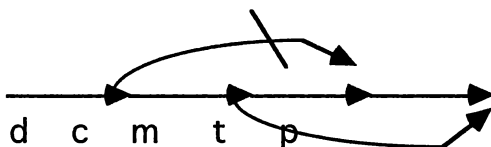


Figure 6.

Selon le principe de spécificité que nous avons illustré à la section 2.1, dans la hiérarchie de la figure 5, le chemin $d \rightarrow c \rightarrow t$ est préféré au chemin $d \rightarrow c \rightarrow m \rightarrow t$. Ainsi l'affirmation $d \rightarrow p$ n'est pas supportée, malgré le fait que le soient les affirmations $d \rightarrow m$ et $m \rightarrow p$. Quand par contre nous ajoutons explicitement le lien $m \rightarrow p$ alors le chemin $d \rightarrow c \rightarrow m \rightarrow p$ est admis, et l'affirmation $d \rightarrow p$ supportée⁷ par la hiérarchie. Cette sensibilité à la forme de la hiérarchie a été justifiée intuitivement par le fait que l'ajout explicite d'un lien modifie la hiérarchie, même si l'affirmation correspondante était déjà supportée par la hiérarchie⁸. Interprétons c comme cachalot, m comme mammifère, t comme animal terrestre, p comme ayant des poumons. Dans la hiérarchie de la figure 6, le lien direct $m \rightarrow p$ indiquerait le fait que les mammifères ont des poumons indépendamment du fait d'être terrestres. Ainsi d, qui est bien un mammifère, mais qui par le fait qu'il est un cachalot n'est pas terrestre, hérite sans exceptions la propriété d'avoir des poumons. Au contraire, dans la hiérarchie de la figure 5, l'absence de lien direct entre m et p, ne nous permet pas de tirer cette conclusion.

Dans la formalisation proposée par la logique des conditionnels, l'axiome ASC par contre, impose que si la formule $m \Rightarrow p$ appartient aux conséquences logiques d'une théorie T, son ajout en tant qu'axiome explicite ne modifie pas ce qui est déjà déductible de la théorie. En fait ASC interdit de différencier le cas dans lequel la formule $m \Rightarrow p$ appartient à une extension par application des règles et des axiomes du systèmes, du cas dans lequel $m \Rightarrow p$ est obtenue par traduction directe de la hiérarchie, et donc représente un lien de la hiérarchie.

7. La notion d'affirmation supportée, a été introduite au début de la section 2.1. Dans une hiérarchie d'héritage une affirmation $x \rightarrow y$ est supportée si et seulement si il existe dans la hiérarchie un chemin admis de x à y.

8. Voir § 2.1 pour la notion d'affirmation supportée.

Sans cet axiome nous ne pouvons pas définir une relation d'accessibilité sur les mondes, il est, donc, techniquement nécessaire dans la formalisation. Toutefois l'interprétation de la redondance imposée par cet axiome, ne semble pas justifiée du point de vue de la hiérarchie, pas plus que l'interprétation contraire basée sur le principe de spécificité⁹.

En conclusion les formalismes non monotones s'attachent à formaliser un conditionnel générique, si p alors normalement q , susceptible d'être soumis aux restrictions et aux exceptions.

La logique des défauts l'exprime par une règle $\frac{p : q}{r}$ intégrant directement un test de consistance. La théorie qui en résulte n'est pas sensible aux formes de résolution des conflits, propres des stratégies sceptiques (cf. 2.1), basées sur la définition récursive des chemins admis par la hiérarchie.

Les logiques des conditionnels, traduction des systèmes d'inférence non monotones, utilisent un conditionnel modal $p \Rightarrow q$. On obtient par traduction des liens de la hiérarchie d'héritage, une théorie dans le langage de la logique des conditionnels. Des tests de consistance sont ensuite appliqués à cette théorie, afin de raisonner dans un modèle particulier et de dépasser ainsi la pauvreté déductive de ces logiques. Elles imposent, par la structure de la logique conditionnelle sous-jacente, une structure particulière à la théorie supportée par la hiérarchie d'héritage.

2.3. Les solutions dans le cadre de la logique linéaire

Nous pouvons définir la relation de déduction de la logique linéaire dans le cadre des séquents de Gentzen [Girard 87], comme une relation binaire sur des ensembles d'occurrences de formules. Cette relation est définie par un ensemble de règles structurelles, c'est-à-dire des règles manipulant des séquents sans utiliser les connecteurs du langage objet.

Nous distinguons deux types de *règles structurelles* :

a) les règles de raisonnement concernant les propriétés de la relation de déduction. Ce sont :

identité $A \vdash A$

coupure
$$\frac{X_1 \vdash Y_1, A \quad X_2, A \vdash Y_2}{X_2, X_1 \vdash Y_2, Y_1}$$

9. Dans [Boutilier 90] une justification de la rationalité d'une telle approche est fournie. Elle ne semble pas toutefois avoir une portée plus grande que la simple défense de son propre formalisme.

b) les *règles des données* concernant les propriétés des ensembles d'occurrences de formules reliées par la relation de déduction. Dans le cadre de la logique linéaire les règles des données de la logique classique :

affaiblissement	$\frac{X \vdash Y}{X, A \vdash Y}$	$\frac{X \vdash Y}{X \vdash Y, B}$
contraction	$\frac{X, A, A \vdash Y}{X, A \vdash Y}$	$\frac{X \vdash Y, A, A}{X \vdash Y, A}$
expansion	$\frac{X, A \vdash Y}{X, A, A \vdash Y}$	$\frac{X \vdash Y, A}{X \vdash Y, A, A}$

sont rejetées.

De façon générale les *règles logiques*, règles d'introduction des connecteurs, permettent de manipuler les séquents en utilisant les opérateurs logiques du langage objet [Dosen 87], [Avron 90].

Le connecteur conditionnel de la logique linéaire est défini par les règles :

$$(\rightarrow g) \frac{X, A \vdash Y, B}{X \vdash Y, A \rightarrow B} \quad (\rightarrow d) \frac{X \vdash Y, A \quad Z, B \vdash W}{X, A \rightarrow B, Z \vdash Y, W}$$

Certaines manipulations sur les séquents qui dans le contexte classique sont exprimées par un même connecteur, ne sont plus identifiables dans un contexte réduit. La règle d'introduction de la conjonction classique se retrouve redoublée dans le contexte des règles structurelles de la relation de déduction linéaire.

On définit ainsi une conjonction multiplicative¹⁰ (symbolisée par $\bullet$) :

$$(\bullet g) \frac{X_1 \vdash A, Y_1 \quad X_2 \vdash B, Y_2}{X_1, X_2 \vdash A \bullet B, Y_1, Y_2} \quad (\bullet d) \frac{X, A, B \vdash Y}{X, A \bullet B \vdash Y}$$

et une conjonction additive¹¹ (symbolisée par $\wedge$) :

$$(\wedge g) \frac{X \vdash A, Y \quad X \vdash B, Y}{X \vdash Y, A \wedge B} \quad (\wedge d) \frac{X, A \vdash Y}{X, A \wedge B \vdash Y}$$

Certaines des propriétés de la conjonction classique sont héritées par la conjonction additive mais non par celle qui est multiplicative : $(A \wedge B) \rightarrow A$ est un théorème linéaire mais pas $A \bullet B \rightarrow A$ et la conjonction multiplicative n'est pas idempotente.

10. Nous utilisons la notation de la logique de la pertinence, comme suggéré dans [Dosen 87] [Avron 90]

11. [Dosen 87] [Avron 90]

Nous retrouvons la conjonction classique, grâce à la présence des connecteurs exponentiels. Il s'agit des modalités du type S.4. Les règles d'introduction du carré ou "bien sûr" comme l'appelle Girard, sont les suivantes :

$$(\Box g) \frac{X, A \vdash Y}{X, \Box A \vdash Y}$$

$$(\Box d) \frac{X \vdash Y, B}{X \Box \vdash Y, \Box B}$$

$$(\Box w) \frac{X \vdash Y}{X, \Box A \vdash Y}$$

$$(\Box c) \frac{X, \Box A, \Box A \vdash Y}{X, \Box A \vdash Y}$$

Nous pouvons avec ces règles démontrer l'équivalence : $\Box A \bullet \Box B \vdash \Box A \wedge \Box B$

Dans la suite nous utiliserons un élément neutre multiplicatif défini par les règles :

$$(1g) \frac{X \vdash Y}{X, 1 \vdash Y}$$

$$(1d) 1$$

On peut démontrer : $A \bullet 1 \leftrightarrow A$ et $A \wedge 1 \rightarrow A$ et $A \wedge 1 \rightarrow 1$

La première solution proposée ici prend en compte les seules restrictions. Plus qu'une solution aux problèmes de formalisation des hiérarchies d'héritage elle se veut une énonciation des limites d'une telle entreprise. La stratégie de la logique linéaire consiste à formaliser les listes d'attributs (les traits pertinents) de façon telle que seules les propriétés intuitivement attendues dans le contexte du raisonnement taxinomique soient dérivables. En particulier l'information négative contenue par une taxinomie n'affirmerait pas, selon Girard, la fausseté de l'attribut Vol pour l'abstrait P. Elle affirmerait simplement qu'on a de bonnes raisons de 'penser' un pingouin comme un oiseau pour lequel on a masqué la propriété voler. De la même façon pour des raisons liées à la taxinomie animale on a préféré ranger les monotrèmes sous l'espèce des mammifères plutôt que de revoir une classification antérieure à l'exploration de l'Australie [Girard 92]. S'appuyant sur le fait que le raisonnement taxinomique n'a pas une rationalité "objective mais subjective", la logique linéaire propose comme solution le rejet des connecteurs classiques. En particulier on n'utilisera pas une négation classique ni une conjonction classique, mais les connecteurs multiplicatifs et additifs de la logique linéaire.

Relativement à chaque terme abstrait les informations taxinomiques, concernant les liens de la hiérarchie et les listes d'attributs, sont exprimées par un axiome. Nous utilisons le pouvoir déductif de la logique linéaire pour extraire les informations relatives à la taxinomie, à partir de l'ensemble d'axiomes qui la décrit. Les informations linéaires peuvent être traduites successivement en informations classiques, grâce aux connecteurs exponentiels (les modalités S4). Par le mécanisme déductif linéaire, seules les informations classiques attendues seront déductibles, évitant entre autre la production d'incohérences.

Considérons à nouveau l'exemple présenté au paragraphe 2, en ce qui concerne les restrictions relativement à l'individu b. Nous pouvons les formaliser avec ces deux axiomes de la logique linéaire :

$$I) O(b) \vdash (M(b) \wedge 1) \bullet (Q(b) \wedge 1) \bullet (R(b) \wedge 1)$$

$$II) P(b) \vdash (O(b) \wedge 1) \bullet (V(b) \wedge 1) \bullet (Q(b) \rightarrow 1)$$

Les *attributs positifs* (ceux qui ne sont pas préfixés par un "non" dans la liste) sont représentés par des conjonctions additives $M(b) \wedge 1$, $R(b) \wedge 1$ etc. On représente de même la relation taxinomique indiquant que P est plus spécifique de O.

L'attribut négatif "non Q" est représenté par la formule $Q(b) \rightarrow 1$ exprimant le fait que l'attribut Q est *masqué*, selon l'expression de Girard, pour P. Le masquage ne correspond pas à une négation, mais à une sorte d'instruction, que l'on pourrait appeler *instruction de masquage*, $Q(b) \rightarrow 1$, qui prescrit de ne pas considérer l'attribut Q pour l'objet b.

Les attributs positifs et les instructions de masquage sont liées entre elles, dans chaque axiome, par une conjonction multiplicative. Cette conjonction nous permet de considérer ces informations comme un tout, dont on ne peut pas négliger des éléments car en général : $A \bullet B \nrightarrow A$.

Toutefois, dans le cas du premier axiome, le fait d'avoir une description en termes des seuls attributs positifs, nous permet de "briser" le lien multiplicatif. En fait en utilisant le fait que : $M(b) \wedge 1 \vdash M(b)$, $Q(b) \wedge 1 \vdash 1$ et $R(b) \wedge 1 \vdash 1$ nous pouvons déduire par applications de coupure de I) que $O(b) \vdash M(b) \bullet 1 \bullet 1$. De ceci grâce à la neutralité de 1 nous avons que : $O(b) \vdash M(b)$. Avec ce procédé nous obtenons l'information classique $\Box O(b) \vdash \Box M(b)$, $\Box O(b) \vdash \Box Q(b)$, $\Box \vdash (b) \vdash \Box R(b)$

Par contre nous ne pouvons pas appliquer directement la même démarche à l'axiome II) à cause de la présence de l'instruction de masquage : $Q(b) \rightarrow 1$. En fait $Q(b) \rightarrow 1 \nvdash 1$ mais $Q(b) \bullet (Q(b) \rightarrow 1) \vdash 1$. La formule $Q(b) \rightarrow 1$ correspond à une instruction du type : "masquer l'attribut Q". Il faudrait d'abord accomplir cette instruction avant de pouvoir extraire l'information positive contenue dans l'axiome II). Nous devons donc trouver une occurrence de $Q(b)$ avec laquelle exécuter le masquage. Cette information $Q(b)$, à masquer, est contenue dans la définition des $O(b)$. Nous obtenons à partir de II) et I), par introduction de 1 à gauche dans I, par conjonction additive et coupure :

$$\text{III) } P(b) \vdash (M(b) \wedge 1) \bullet (Q(b) \rightarrow 1) \bullet (V(b) \wedge 1) \bullet (Q(b) \wedge 1) \bullet (R(b) \wedge 1)$$

De III en utilisant que $Q(b) \wedge 1 \vdash Q(b)$ et $Q(b) \bullet (Q(b) \rightarrow 1) \vdash 1$, et la neutralité de 1, nous obtenons :

$$\text{IV) } P(b) \vdash (M(b) \wedge 1) \bullet (R(b) \wedge 1)$$

qui contient tous les attributs (directs ou hérités) positifs concernant les P. Avec le même procédé appliqué pour I) nous obtenons $\Box P(b) \vdash \Box M(b)$, $\Box P(b) \vdash \Box R(b)$.

Par cette formalisation nous avons donc évité la production d'inconsistances et en particulier de $P(b) \vdash Q(b)$. Les autres informations déductibles correspondent-elles effectivement à nos attentes ?

Nous ne pourrions pas déduire la relation taxinomique : $P(b) \vdash O(b)$, mais en revanche nous aurons $P(b) \vdash O(b) \bullet (Q(b) \rightarrow 1)$ et $P(b) \bullet Q(b) \vdash O(b)$, qui peuvent être interprétés comme énonçant que les P sont des O sans la propriété Q, ou que si l'on restaure la propriété Q pour la classe des P on obtient tout court celle des O.

Au-delà des difficultés de détail, la caractéristique de cette solution est de ne prendre en compte que les restrictions. Girard dit explicitement¹² (page 6) :

"le mécanisme de la logique, ..., fonctionne mais ne fait pas de miracles. En particulier *nous devons savoir à l'avance ce qu'il faut effacer*, et combien de fois il faut le faire dans le cas d'occurrences multiples : par exemple nous ne pouvons pas dire "effacer A, partout où il apparaît". D'ailleurs aucun *système logique ne peut le faire* : tout système logique raisonnable devrait permettre une certaine forme de déduction taxinomique – autrement le système ne serait pas efficace –... Le seul moyen est de connaître à l'avance combien d'occurrences de A doivent être détruites et la particularité des connecteurs de la logique linéaire est que ce nombre est significatif".

Cette déclaration d'intention n'a pas été respectée par les tentatives successives de solution aux questions des hiérarchies dans le cadre de la logique linéaire. Une stratégie qui prend en compte des véritables exceptions, sans la connaissance préalable de leur nombre, a été définie dans [Fouquerre et al. 93]. Ce résultat est obtenu essentiellement par trois mécanismes. Le premier modifie l'instruction de masquage en la préfixant par le connecteur exponentiel $\square$; de cette manière, ce connecteur étant fermé par la règle $(\square w)$ toutes les exceptions relatives à un chemin peuvent être masquées par la même instruction : le contrôle sur ce qui est déductible ou pas dans le raisonnement taxinomique ne repose plus sur le nombre d'occurrences des instructions de masquage. Le deuxième mécanisme sert donc à contrôler les informations contradictoires qui sont exprimées par une modification du langage : il y a trois types d'atomes dans le système : les atomes non signés p, q, r , les atomes négatifs $p-, q-, r-$ et les atomes positifs $p+, q+, r+$. La présence d'information contradictoire est détectée par la présence d'atomes de signe opposé ($p+, p-$) dans un séquent. Le troisième mécanisme se base sur l'imposition d'un ordre dans l'application des règles du fragment intuitionniste de la logique linéaire. Le libre jeu d'application des règles est interdit. On impose un ordre afin d'obtenir une classe particulière de séquents, les séquents simples, qui ne contiennent pas d'atomes opposés, et qui correspondent aux extensions possibles d'une hiérarchie.

La formalisation proposée par [Fouquerre et al. 93], concerne une stratégie crédule avec principe de spécificité, appliquée toutefois à un type particulier de hiérarchie d'héritage, les hiérarchies taxinomiques avec exceptions¹³. Par rapport aux hiérarchies que nous avons présenté nous pouvons distinguer trois différences fondamentales :

- a) Il n'y a que des liens positifs, exprimant des propriétés typiques des objets. Ces liens donnent lieu à un graphe acyclique
- b) il y a des liens exceptionnels, qui constituent une inhibition directe des liens positifs, et qui peuvent donner lieu à des cycles
- c) les liens exceptionnels ne peuvent pas partir des racines.

12. C'est nous qui soulignons et traduisons

13. Les auteurs mentionnent des extensions de leur formalisation à des contre-exceptions.

Comme il n'y pas de véritable lien négatif il n'est pas possible d'exprimer en forme canonique la structure de type 'diamant de Nixon'. On peut toutefois obtenir une structure similaire en dédoublant les objets : la figure 7 montre comment on pourrait exprimer le correspondant de la hiérarchie de la figure 2 dans les hiérarchies taxinomiques avec exceptions :

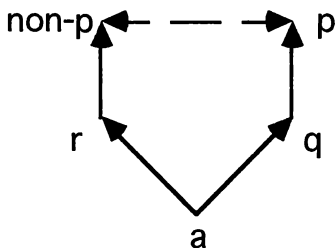


Figure 7.

La double flèche en traits pointillés représente deux liens exceptionnels entre p et $\text{non } p$, qui sont donc entre eux directement inhibés.

Si N est l'ensemble des objets de la hiérarchie, dans lequel on ne distingue pas entre individus et espèces, pour tout $p \in N$ nous définissons un axiome dans la logique intuitionniste linéaire comme suit :

$$p \vdash p^+ \bullet q_1 \bullet \dots \bullet q_n \bullet r_1 \bullet \dots \bullet r_m \bullet \Box(r_1 \rightarrow 1) \bullet \dots \bullet r_m \bullet \Box(r_m \rightarrow 1)$$

où q_i , $1 \leq i \leq n$ est l'ensemble des objets dans N qui sont liés à p par des liens positifs ; r_j , $1 \leq j \leq m$ est l'ensemble des objets liés à p par des liens exceptionnels.

A partir de cet ensemble d'axiomes propres, on peut obtenir les séquents simples, par application des règles de la logique linéaire. Un séquent simple est un séquent de la forme : $q_1, \dots, q_k \vdash X$ où q_i , $1 \leq i \leq k$ sont des atomes non signés correspondant aux feuilles de la hiérarchie, X est une conjonction multiplicative d'atomes signés, ne comptant pas un même atome avec signe opposé.

Un séquent simple est obtenu par des *preuves standard*. Une preuve standard est une preuve dans le fragment intuitionniste de la logique linéaire dans laquelle toutes les prémisses gauches de l'application de $\Box$ -g sont des axiomes d'identité, toutes les introductions de $\bullet$ -g, $\rightarrow$ -g et $\Box$ -g précèdent les applications de la règle de coupure et il n'y a que des règles de coupure ayant comme prémisses gauches des axiomes propres.

L'exemple de la figure 7 peut être axiomatisé par les cinq axiomes suivants :

$$n(a) \vdash n(a)^+ \bullet r(a) \bullet q(a)$$

$$q(a) \vdash q(a)^+ \bullet p(a)$$

$$r(a) \vdash r(a)^+ \bullet \text{non } p(a)$$

$$p(a) \vdash p(a)^+ \bullet \text{non } p(a)^- \bullet \Box(\text{non } p(a) \rightarrow 1)$$

$$\text{non } p(a) \vdash \text{non } p(a)^+ \bullet p(a)^- \bullet \Box(p(a) \rightarrow 1).$$

Les deux séquents simples sont :

$$n(a) \vdash n(a)^+ \bullet q(a)^+ \bullet r(a)^+ \bullet p(a)^+ \bullet \text{non } p(a)^-$$

$$n(a) \vdash n(a)^+ \bullet q(a)^+ \bullet r(a)^+ \bullet p(a)^+ \bullet \text{non } p(a)^-$$

Ils correspondent donc aux deux extensions maximales de la hiérarchie.

La première formalisation que nous avons examinée dans le cadre de la logique linéaire n'a pas recours à instructions extra-logiques, elle n'accorde pas de priorité à la notion de consistance car les restrictions ne sont pas formalisées par la négation du langage, elle est simple et intuitive, mais ne s'applique qu'aux restrictions.

L'extension de cet approche aux véritables exceptions rend nécessaire le recours à des instructions extra-logiques (l'ordre d'application des règles) et au contrôle de la génération des incohérences par les étiquettes + et -, qui dénaturent la notion logique de formule atomique.

3. Conclusion

L'analyse que nous avons menée ne se prétend pas exhaustive, mais permet toutefois de tirer quelques conclusions.

1) Les tentatives de formalisation examinées montrent qu'il n'existe pas à l'état actuel une "bonne formalisation" (au sens de la section 2.1) neutre par rapport aux différentes stratégies qui peuvent être appliquées à une hiérarchie. La logique des défauts formalise aisément les stratégies "crédules"¹⁴ mais soulève des problèmes quant au principe de spécificité et aux stratégies sceptiques. Les logiques des conditionnels, bases des formalismes non monotones, imposent une analyse figée de la notion de redondance et ne sont pas adaptées aux stratégies sceptiques à la [Horty et al. 87] et [Stein 88]. Quant à la logique linéaire, la stratégie de [Girard 92] ne s'applique qu'aux restrictions, dont nous devons connaître le nombre à l'avance et si le travail présenté en [Fouquerre et al. 93] prend en compte les exceptions il ne s'applique qu'à un type particulier de hiérarchie d'héritage.

14. La notion d'extension dans la logique des défauts, correspond exactement à la notion imal de chemins non conflictuels d'une stratégie crédule. (cf. § 2.1 et § 2.2.1)

2) Ce constat soulève une question de possibilité concernant l'existence d'une telle logique, applicable à toutes hiérarchies et capable d'exprimer n'importe quel stratégie.

Ce qu'en principe on attend d'une formalisation logique c'est qu'elle fournisse une traduction dans laquelle chaque lien est traduit de manière homogène et compositionnelle, la traduction d'un lien ne devant pas dépendre de celle des autres liens, ou de sa position dans la hiérarchie, mais exclusivement du type de lien (positif ou négatif) et du type d'objets connectés (individu-espèce ou espèce-espèce). On demande aussi que la traduction soit modulaire, c'est-à-dire que la modification ou l'ajout d'un lien dans la hiérarchie, ne doivent pas se traduire par la modification de l'entière théorie logique correspondante.

L'analyse que nous avons conduite dans la section 1, nous amène par contre à relever la nature fondamentalement ambiguë du lien taxinomique exprimant des formes de la prédication différentes. Dans le cadre des hiérarchies d'héritage, nous avons vu que, dans le sillage de la logique classique, toutes les formes de prédication étaient réduites à une seule. En conséquence toute désambiguïsation ne peut être que locale, relative à un certain configuration de la hiérarchie. Le parcours de la hiérarchie impose, surtout dans les stratégies sceptiques à la Horty et Stein, l'analyse de la forme du graphe afin de résoudre les conflits. Or un des principes de base de la logique est le principe de compositionnalité. Les stratégies de tests de consistance, ainsi que celles qui imposent un ordre à l'application des règles, constituent toutes des tentatives de dépasser la nature compositionnelle de la logique. Ce caractère holistique de la définition des hiérarchies est évident dans la nature récursive des définitions des stratégies sceptiques. Ce n'est probablement pas le simple apanage des hiérarchies d'héritage mais cela semble être propre aux taxinomies en tant que telles. La prédication nominale, la notion de spécificité, la notion d'individu normal pourraient rendre nécessaire une telle approche holistique mais seule l'analyse des propriétés logiques des formes de la prédication permettra d'éclaircir ce point. La difficulté de principe d'exprimer à l'aide d'une technique essentiellement compositionnelle des connaissances holistiques indique que nous devons faire appel à des formes multiples de prédication. Les tentatives faites par rapport aux problèmes de révision par [Fariñas Herzig 92] constituent un premier pas dans cette direction. La notion de dépendance qu'ils introduisent est définie par une relation logique très pauvre, symétrique et intransitive mais qui exprime le lien cognitif minimal entre abstraits. La prédication nominale pourrait être le correspondant linguistique de cette relation.

3) Y-a-t-il vraiment besoin d'une traduction logique pour interpréter les hiérarchies d'héritage?

Ceux qui soutiennent cette nécessité prétendent que ce n'est que par la traduction logique que l'on peut donner une sémantique claire aux hiérarchies d'héritage, ce qui revient à en donner une interprétation rationnelle et définitive. Mais ce raisonnement semble surestimer le rôle de la logique en tant qu'outil de formalisation.

Nous sommes en présence d'un procédé technique, nous ne savons pas quel est le modèle correct de son fonctionnement, ni comment interpréter sa structure. La simple formalisation logique ne peut pas nous fournir le moyen de décider entre différents choix interprétatifs. Ce que l'intuition n'arrive pas à démêler, ne peut pas

être résolu par la logique sans une vraie analyse préalable de ce qu'on est censé formaliser. En effet, dans l'analyse des formes de la prédication impliquées par le raisonnement taxinomique, les informaticiens semblent se contenter souvent de ce que l'introspection, notion par ailleurs assez douteuse, suggère, sans considérer que le poids de la tradition logique, pourrait avoir "corrompu" nos capacités introspectives. La réduction fréquente de problèmes de nature très différente à la matrice commune du maintien de la cohérence devrait nous mettre en garde à ce propos.

Quant à la question de la nécessité d'une formalisation logique, si on se restreint au simple problème technique de représenter différentes stratégies de parcours d'un graphe, il semble bien que les approches non logiques soient suffisantes et claires, pour rendre compte de la notion d'affirmation supportée par une hiérarchie.

Si en revanche, on interprète les hiérarchies d'héritage comme des outils de mise en œuvre d'un fragment de comportement intelligent, alors la vraie question repose sur l'analyse du problème et, comme dans le cas de la notion de normalité, on est bien loin de disposer à ce jour d'une analyse suffisante.

Bibliographie

- [Anderson et Belnap 75] A.R. Anderson and N.D. Belnap jr. "Entailment", vol. 1 Princeton University Press (1975).
- [Avron 87] : A. Avron, "Simple consequence relations", *Information and Computation* 92, p. 105-139 (1987).
- [Boutilier 90] C. Boutilier "A semantical approach to stable inheritance reasoning" Dans : *Proceedings of IJCAI 89*, Morgan Kaufman (1989), pages 1134-1139.
- [Crocco 93] : G. Crocco "Fondements logiques du raisonnement contextuel" Thèse de l'université Paul Sabatier, juillet 1993, Toulouse.
- [Delgrande 87] J.P. Delgrande A first order conditional logic for prototypical properties" *Artificial Intelligence* n. 36, pages 63-90 1987.
- [Delgrande 88] J. Delgrande "An approach to default Reasoning Based on a First-Order Conditional logic : Revised report" *Artificial Intelligence* n. 36 (1988) pages 63-90.
- [Etherington et al. 83] D. Etherington, R. Reiter "On inheritance hierarchies with exceptions", In *Proceedings of AAAI-83* Morgan Kaufmann pages 104-108 (1983).
- [Etherington 87] D.W. Etherington "Formalizing nonmonotonic reasoning systems" *Artificial Intelligence* vol. 31 pages 41-85 (1987).
- [Etherington 88] D.W. Etherington "Reasoning with incomplete information" *Research notes in Artificial Intelligence*, Pitman London, Morgan Kaufmann Los Altos 1988.
- [Fariñas, Herzig 92] : L. Fariñas, A. Herzig "Belief Change and Revision. Preliminary report, Université Paul Sabatier.
- [Fouquerre et al 93] C. Fouquerre, J. Vauzeilles "Taxonomic Linear Theories" *Proceeding of ESQUARQ* pages 121-128. Spain.
- [Gabbay 85] : D.M. Gabbay "Theoretical Foundations for Non-Monotonic Reasoning in Expert Systems" in *Logics and Models of Concurrent Systems*, publié par, K.R. Apt, Springer Verlag éditeur, 1 pages. 439-457 (1985).

- [Girard 87] : J.Y. Girard, Linear logic, Theoretical Computer Science 50 (1987) p. 1-102.
- [Girard92] : J.Y. Girard "Logic and exceptions : a few remarks", Journal of Logic and Computation n° 2, pages 11-118 (1992).
- [Horty Al. 87] J. F. Horty, R.H. Thomason, D. Touretzky "A skeptical theory of inheritance in non monotonic semantic networks" Proceedings of AAAI pages 358-363 Seattle, july 1987.
- [Lamarre 92] "A promenade from monotonicity to non monotonicity via a theorem prover K. R 92, pages 572-580.
- [Lewis 75] : Lewis D., *Counterfactuals*. Basil Blackwell Ltd, 1986 (first published 1975).
- [Lehmann 89] Lehmann D. What does a conditional base entails in Proceedings of KR.89, Edité par R.J. Brachman, H.J. Levesque, R. Reiter, Morgan Kaufmann pages 212-222.
- [McCarthy 80] J. MacCarthy Circumscription A form of non-monotonic reasoning Artificial intelligence n. 13, p. 27-39 (1980).
- [Mc Dermott Doyle 80] : D. Mcdermott, J. Doyle "Non-monotonic logic I" *Artificial intelligence*, vol. 13 p. 41-47 (1980).
- [Makinson 88] D. makinson "A general theory of cumulative Inference, Non monotonic Reasoning M. Reinfrank, J. de Kleer., M.L. Ginsberg, E.Sandewall (eds), Springer Verlag n° 346, (1988) pages 1-18.
- [Moore 84] : R.C. Moore "Possible world semantics for Autoeistemic logic" dans *Proc. de AAAI workshop on non monotonic reasoning*. pages 664-654 (1984).
- [Pearl 90] J. Pearl System Z a natural Ordering of defaults with tractable Applications to non monotonic reasoning" in Theoretical aspects of Reasoning about knowledge (Tark III) Morgan Kaufmann 1990 pages 121-135.
- [Reiter 80] : R. Reiter "A logic for default reasoning" *Artificial Intelligence* vol. 13 pages 81-132 (1980).
- [Stalnaker 68] R. Stalnaker "A theory of conditionals" in Studies in Logical Theory, N. Rescher ed. Basil Blackwell 1968.
- [Stein 89] L.A. Stein "Skeptical inheritance : computing the intersection of credulous extensions" Proocceeding of IJCAI 89, pages 1153-1158, Detroit 1989.
- [Touretzky 86] D. Touretzky "The mathematics of inheritance systems" Research notes in Artificial intelligence, Pitman, Morgan Kaufmann Los Altos 1988.
- [Touretzky et al.87] D. Touretzky, J. Horty, R. Thomason "A clash of intuitions : the courant state of Nonmonotonic multiple Inheritance systems". Procceeding of IJCAI 87. pages 476-482.
- [Thomason et al. 88] R. Thomason J. Horty "Logics for inheritance theory" Non monotonic Reasoning M.Reinfrank,J.de Kleer., M.L. Ginsberg, E.Sandewall (eds), Springer Verlag n° 346, (1988) pages 220-237.
- [Vuillemin 86] : J. Vuillemin "What are philosphical systems" Cambridge University Press (1986).
- [Vuillemin 84] J. Vuillemin, "Les formes fondamentales de la prédication : un essai de classification" Recherches sur la philosophie et le langage. Langage et philosophie de sciences, Cahier n. 4, DREUG, Grenoble, Vrin Paris (1984).

Gabriella CROCCO

Chapitre 6

Changements minimaux constructifs

Nous présentons une logique de conditionnels, dont l'opérateur conditionnel est interprété en utilisant une opération de révision. Le sens associé à l'expression conditionnelle $A \Rightarrow B$, relativement à une base de connaissances KB, est obtenu en évaluant l'expression B sur la base de connaissances KB' obtenue à partir de KB par la mise à jour de B. Cette interprétation constructive de l'expression conditionnelle, à notre point de vue, ouvre une voie réaliste à l'application et au traitement informatique des conditionnels.

1. Introduction

Apprendre un fait nouveau nous oblige à modifier l'organisation de nos connaissances. Si nous représentons nos connaissances par une théorie (ou une base de connaissance) KB, un fait nouveau par une formule A (la formule de changement) et l'opération d'apprentissage par l'opérateur *, ce qui nous intéresse est le résultat de l'opération $KB * A$. Plus précisément, nous voulons calculer ce qui peut être déduit de $KB * A$. Nous distinguons trois aspects :

- L'aspect syntaxique : clarifier la relation entre le changement et les connecteurs.

- L'aspect sémantique : refléter une opération de changement par des opérations sur les modèles.

• L'aspect computationnel : calculer de manière effective ce qui peut être déduit de la base de connaissance modifiée par le changement.

L'objectif final est de donner une axiomatique avec une sémantique naturelle pour laquelle une méthode de déduction raisonnable peut être définie.

La plupart des approches développées à ce jour se sont concentrées sur la clarification de la relation entre les connecteurs et le changement. La contribution la plus importante aux discussions concernant les fondements logiques du changement est celui apporté par les postulats AGM (Alchurron et al. 1985, Gärdenfors 1988) — qui sont en fait une axiomatisation (dans le métalangage) de l'opération de changement — lesquels sont censés, selon les auteurs, réglementer toute opération de révision de croyance rationnelle. Ces postulats sont présentés comme un cadre à l'intérieur duquel toute opération de révision pourraient se retrouver. Ils nous donnent cependant trop peu d'information concernant le résultat d'un changement. En particulier, les axiomes de révision de base sont satisfaits par l'opération de changement trivial, $*_{\text{trivial}}$, définie de la façon suivante :

$$\begin{aligned} \text{KB } *_{\text{trivial}} A &= \text{Cn}(\text{KB} \cup \{A\}) & \text{si } \text{KB} \vdash A \\ \text{KB } *_{\text{trivial}} A &= \text{Cn}(\{A\}) & \text{si non}(\text{KB} \vdash A) \end{aligned}$$

où Cn est l'opérateur de conséquence Tarskien (Gärdenfors 1988). Nous aimerions au moins que toutes les formules B qui n'ont "rien à voir" avec la formule de changement ne soient pas influencées par l'opération de changement, c.à.d.

$$\text{KB} \vdash B \text{ ssi } \text{KB} * A \vdash B \quad \text{si } A \text{ et } B \text{ n'interfèrent pas}$$

Nous appelons ceci le *postulat de non-interférence*. (Nous devons préciser ce que nous entendons par "interférence").

Nous pensons que, jusqu'ici, les aspects syntaxiques, sémantiques et computationnels n'ont pas été considérés de façon équilibrée. En particulier, l'aspect computationnel du changement n'est pas suffisamment pris en considération. Nous essayons ici d'avancer dans les trois directions simultanément. À notre avis, pour ne violer aucune des exigences formelles ou computationnelles lors de l'application des opérateurs de changement, il faudrait commencer par restreindre les formules de changement qu'il nous est permis d'utiliser. Dans un travail précédent (Fariñas Herzig 1988), nous avons défini une logique conditionnelle appelée logique ASSUME où les formules de changement étaient limitées aux *littéraux*. Une sémantique a été donnée en termes de proximité entre les interprétations. Dans cet article, nous généralisons cette approche.

Tout d'abord, nous introduisons le langage, la sémantique, et l'axiomatique de notre logique des conditionnels que nous appelons la logique ASSUME. Ensuite, nous introduisons une forme normale à partir de laquelle nous montrons dans les sections suivantes que — à l'opposé des logiques conditionnelles standard — une preuve de complétude simple et une procédure de décision peuvent être données.

2. Langage et sémantique

Soit un langage construit à partir d'un ensemble fini de variables propositionnelles PVAR avec les connecteurs classiques plus un opérateur conditionnel $>$. Dans la formule

$A > C$, A est appelé la *formule de changement*. Notez que les emboîtements d'opérateurs conditionnels sont autorisés : $A > (B > C)$ est une formule de notre langage. Comme d'habitude, les *littéraux* sont des variables propositionnelles ou des négations de variables propositionnelles, et les *clauses* sont des disjonctions d'éléments littéraux. Nous utiliserons une fonction *pvar* qui associe à toutes les formules l'ensemble des variables propositionnelles qui y apparaissent. Par exemple, $\text{pvar}(p > (q \wedge r)) = \{p, q, r\}$

Habituellement, la sémantique de la logique conditionnelle est donnée en termes de modèles utilisant des mondes possibles. Ici, nous n'avons besoin de nous référer qu'à une seule interprétation classique (que nous supposons être un sous-ensemble de l'ensemble des variables propositionnelles PVAR). Cette façon de faire peut sembler bizarre : l'interprétation habituelle des opérateurs conditionnels réfère à d'autres interprétations classiques (à travers la notion de monde possible). En fait, l'effet des axiomes de non-interférence est à ce point draconien que, étant donné une interprétation classique I et une formule de changement A , nous sommes en mesure de construire entièrement l'ensemble des interprétations classiques qui en résultent. Nous avons ainsi réduit l'opération de changement à une opération de changement sur des bases de connaissance complète (c'est-à-dire des interprétations classiques). Dans ce qui suit, une interprétation qui satisfait A sera appelée une *A-interprétation*.
Définition. Soit I et J deux interprétations données, la *distance* entre I et J est définie comme étant la différence symétrique $I \dot{-} J = I \setminus J \cup J \setminus I$.

En conséquence, la distance entre deux interprétations est l'ensemble des variables propositionnelles pour lesquelles elles diffèrent. Notre notion de distance est celle de (Winslett 1988). Nous caractérisons d'abord ce qu'est l'opération de changement sur les bases de connaissances complètes : l'ensemble I^*A est l'ensemble des A -interprétations dont la distance à I est minimale. Nous pouvons alors définir la relation de satisfaction.

Définition. Soit I une interprétation classique et A une formule. I^*A est défini par :

$$I^*A = \{J : J \text{ est une } A\text{-interprétation et il n'y a pas de } A\text{-interprétation } J' \text{ telle que } I \dot{-} J \subset I \dot{-} J'\}$$

La relation de satisfaction pour l'opérateur conditionnel sera :

$$I \models A > C \text{ si pour tout } J \in I^*A, J \models C$$

Propriétés. Soit I une interprétation et A une formule.

- (1) Si A est un littéral, alors I^*A interprète comme I toutes les variables propositionnelles sauf celles de $\text{pvar}(A)$.
- (2) Si A est une clause et $I \models A$, alors $I^*A = \{I\}$.
- (3) Si A est une clause de la forme $B \vee C$ et $I \not\models B \vee C$, alors
 $I^*B \vee C = I^*B \cup I^*C$.

Remarquons que chaque interprétation classique étant une théorie complète, l'opération sémantique $*$ est définie pour les théories complètes. Elle vérifie autant les postulats de révision que ceux de mise à jour. Ceci justifie l'utilisation de $*$ à un niveau sémantique.

3. Axiomatique

Suivant Chellas (Chellas 1975), voici les règles d'inférence et les schèmes d'axiomes pour la logique conditionnelle ASSUME, construits à partir de l'axiomatisation de la logique propositionnelle classique.

RCEA	$\frac{A \leftrightarrow B}{A > C \leftrightarrow B > C}$
RCK	$\frac{B_1 \wedge B_2 \wedge \dots \wedge B_n \rightarrow C}{A > B_1 \wedge A > B_2 \wedge \dots \wedge A > B_n \rightarrow A > C}$
ID	$A > A$
CA	$A > C \wedge B > C \rightarrow (A \vee B) > C$
MP	$A > C \rightarrow A \rightarrow C$
CS	$A \wedge C \rightarrow A > C$
consistance de littéraux	$\neg(A > \neg A) \quad \text{si } A \text{ est un élément littéral}$
non-interférence	$C \rightarrow A > C \quad \text{si } A \text{ et } C \text{ n'interfèrent pas}$
non-interférence disjonctive	$(A \vee B) > C \rightarrow A \vee B > C \quad \text{si } A \text{ et } B \text{ n'interfèrent pas}$

Nous disons que deux formules A et B *interfèrent* (Nute 1980) si elles ont certaines variables propositionnelles en commun, c'est-à-dire que

$$\text{pvar}(A) \cap \text{pvar}(B) \neq \emptyset.$$

Tous les axiomes sont standard sauf les deux derniers. Donc la logique ASSUME est construite sur le système CK (Nute 1980, Chellas 1975) qui ne contient que les règles d'inférences (RCEA) et (RCK). Les deux derniers axiomes sont supposés donner une approximation des exigences du postulat de non-interférence. La (consistance de littéraux) signifie que les éléments littéraux correspondent à des énoncés consistants. (Remarquez que le faux n'est pas une

formule propositionnelle.) La lecture à faire de la (non-interférence) est claire. Elle formalise une notion de consommation qui est essentielle dans plusieurs domaines de l'informatique, la génération de plan par exemple (Bibel et al. 1988).

Cette notion est à la base d'une nouvelle approche de la question "qu'est-ce que la logique ?" Nous faisons ici référence à la logique linéaire (Girard 1986), aux logiques substructurales (Dosen 1989), et aux relations de conséquences basées sur des structures (Fariñas et Crocco 1992).

L'axiome de (non-interférence disjonctive) peut être lu de la façon suivante : si A n'appartient pas à la base de connaissance de départ et que A et B n'interfèrent pas, le changement par B a, au plus, les conséquences du changement par $A \vee B$.

Ainsi, notre axiomatique est donnée pour des formules de changement de tout genre. Elle sera correcte, mais nous montrerons qu'elle n'est complète que pour les clauses, c'est-à-dire pour les disjonctions de littéraux. Cela est dû au fait que la (non-interférence disjonctive) est apparemment trop faible, et qu'il n'y a pas d'axiomes pour les changements conjonctifs.

Propriété. Les formules suivantes sont dérivables dans la logique ASSUME.

(1)	$A > (C \wedge D) \rightarrow A > C \wedge A > D$	
(2)	$A > C \wedge A > D \rightarrow A > (C \wedge D)$	
(3)	$A > \text{vrai}$	
(4)	$\text{faux} > A$	
(5)	$A \rightarrow C \leftrightarrow A > C$	
(6)	$A > (B \wedge C) \leftrightarrow A > B \wedge A > C$	
(6)	$(A \vee B) > C \rightarrow A > C \vee B > C$	si A et B n'interfèrent pas
(7)	$(A \rightarrow B) > C \rightarrow A \rightarrow (B > C)$	si A et B n'interfèrent pas
(8)	$\neg A \wedge \neg B \rightarrow (A \vee B) > C \leftrightarrow (A > C \wedge B > C)$	si A et B n'interfèrent pas
(9)	$(A \vee B) > C \rightarrow A \vee B \vee (A > C \wedge B > C)$	si A et B n'interfèrent pas

4. Mise en forme normale

Ce qui suit sera utile pour mettre des formules en forme normale. De telles formes normales pourront toujours être obtenues quand les formules de changement sont des clauses.

Propriété. Soit une clause $A \vee B$. Alors les équivalences suivantes sont dérivables dans ASSUME.

$$(A \vee B) > C \leftrightarrow ((A \vee B) \wedge C) \vee (A > C \wedge B > C)$$

Cette propriété permet de réduire les formules de changement clausales en formules de changement littérales. En conséquence, nous pouvons maintenant supposer que les formules sont des formes normales négatives (c'est-à-dire que les négations s'appliquent seulement aux variables propositionnelles). La prochaine propriété nous permettra de distribuer les changements d'éléments littéraux sur des

conjonctions et des disjonctions. Ainsi nous pouvons produire des formules où $>$ ne s'applique plus aux conjonctions et aux disjonctions.

Propriété. Soit L un élément littéral. Alors les équivalences suivantes sont dérivables dans la logique ASSUME.

$$(1) \quad L > (B \wedge C) \leftrightarrow L > B \wedge L > C$$

$$(2) \quad L > (B \vee C) \leftrightarrow L > B \vee L > C$$

Preuve. La première formule est une conséquence de (RCK). La seconde est prouvée par induction sur les formules B et C en utilisant la (non-interférence) et le raisonnement au cas par cas. (Fariñas Herzig 1988). Le prochain théorème nous permettra d'éliminer entièrement l'opérateur conditionnel des formules.

Propriété. Soit L un élément littéral. Alors les équivalences suivantes sont dérivables dans la logique ASSUME.

$$(1) \quad L > L' \leftrightarrow \text{vrai} \quad \text{si } L = L'$$

$$(2) \quad L > L' \leftrightarrow \text{faux} \quad \text{si } L = \neg L' \text{ ou } L' = \neg L$$

$$(3) \quad L > L' \leftrightarrow L' \quad \text{si } \text{pvar}(L) \neq \text{pvar}(L')$$

Preuve. Il s'agit de conséquences immédiates de (ID), soit, respectivement, la (consistance atomique) et la (non-interférence) (Fariñas Herzig 1988). Toutes ces propriétés nous conduisent à l'important résultat suivant :

Théorème de normalisation. Dans ASSUME, toute formule de changement clausale est équivalente à une formule classique.

5. Correction et complétude

Il est facile de vérifier que les axiomes de ASSUME sont valides et que les règles d'inférence préservent la validité. Pour les formules de changement clausales, la complétude peut être prouvée en utilisant le fait que toutes les formules de ASSUME sont équivalentes à une formule classique.

Remarque. Il est tentant de formuler la non-interférence de deux formules A et C comme dans la théorie des probabilités par " $A \wedge C, A \wedge \neg C, \neg A \wedge C$ et $\neg A \wedge \neg C$ sont consistants" comme ce fut suggéré par D. Dubois. Mais il y a le contre-exemple suivant : soit C la formule $p \wedge q \vee \neg p \wedge r$, et soit A la formule $\neg q$. C et A n'interfèrent pas, $(p \wedge q) \vee (\neg p \wedge r) \rightarrow \neg q \vee ((p \wedge q) \vee (\neg p \wedge r))$ serait un théorème. Mais cette formule ne serait pas valide : soit I une interprétation telle que $I \models p$ et $I \models q$. (Donc $I \models C$). Si nous changeons I par $\neg q$ nous aurons J tel que $J \models p$. Donc $I \not\models C$ et $I \not\models C \rightarrow A > C$.

6. Dédution automatique dans ASSUME

Le théorème de normalisation fournit un outil pour prouver la complétude mais aussi pour faire de la déduction automatique dans la logique ASSUME. Un éventail de possibilités apparaît ici, selon la façon dont est utilisé le théorème de normalisation.

Il peut être utilisé globalement : dans un premier temps, nous transformons la formule $KB \rightarrow A > B$ en une formule classique équivalente et ensuite une méthode classique de preuve peut être utilisée. Comme la normalisation est exponentielle, ceci n'est pas très efficace.

À l'autre extrémité, nous utilisons une normalisation indolente, en d'autres mots nous utilisons le théorème de normalisation seulement si cela est nécessaire pour poursuivre la déduction.

7. La représentation du changement dans ASSUME

Étant donné une base de connaissance KB et une formule de changement clausale A , nous pouvons maintenant poser le problème de l'appartenance :

Une formule C est-elle dans la théorie révisée $KB * A$?
en problème de déduction dans notre logique ASSUME.

Puisque notre logique est propositionnelle et possède un nombre fini de variables propositionnelles, toute base de connaissance complète KB (qui est en fait une interprétation) peut être représentée par une conjonction d'éléments littéraux dans KB . Ainsi, toute base de connaissance peut être représentée par la disjonction de ses interprétations. En utilisant la complétude, nous obtenons la propriété principale suivante.

Propriété. Soit KB une base de connaissance et A une formule de changement. Alors

$$KB * A = \{C : \vdash \text{ASSUME } KB \rightarrow A > C\}$$

De cette façon, nous avons défini une opération de changement $*$ à partir d'une logique conditionnelle comme le fait le test de Ramsey. Nous devons maintenant vérifier quelles sont les propriétés de l'opération de changement que nous avons ainsi définie. Il est facile de voir que parmi les postulats de révision, c'est le postulat de consistance qui n'est pas satisfait⁰. Les postulats de mise à jour sont satisfaits et la plupart d'entre eux peuvent être établis directement.

Exemple. Dans (Winslett 1988) nous trouvons l'exemple suivant. Soit deux bases de connaissance $KB1 = \{p, q\}$ et $KB2 = \{p, p \rightarrow q\}$. $KB1$ et $KB2$ sont équivalents. Donc la mise à jour par $\neg q$ devrait avoir le même effet sur $KB1$ que sur $KB2$. C'est effectivement le cas: nous avons $KB1 \vdash [\neg q](p \wedge \neg q)$ aussi bien que $KB2 \vdash [\neg q](p \wedge \neg q)$ ¹.

0. Comme nous l'avons noté, $*$ satisfait les postulats de révision dans les cas de bases de connaissance complètes.

1. En fait, l'exemple est utilisé dans (Winslett 1988) comme contre-exemple des "updates" de bases de formules comme dans (Ginsberg Smith 1988), où les "update" de $KB1$ et $KB2$ sont différents.

Nous avons montré que la logique ASSUME peut exprimer les changements clausaux. Cependant notre approche peut être étendue dans le but d'exprimer des formules générales de changement : pour toute formule générale de changement A , il y a une formule équivalente en forme normale disjonctive canonique, où chaque disjonction de la formule est associée à un modèle minimal de A . Soit $B \wedge C$ qui est de forme normale et I une interprétation, alors nous avons $I * B \wedge C = I * B * C \cup I * C * B$.

En conséquence, nous devons ajouter à notre axiomatique l'axiome

$$(A \wedge B) > C \leftrightarrow A > (B > C) \wedge B > (A > C)$$

si $A \wedge B$ est une forme normale disjonctive canonique

8. Discussion

Nous avons défini une logique conditionnelle ASSUME pour représenter le changement et nous avons fourni une procédure de preuve.

Le fragment des formules de changement que nous avons considérées ici respecte presque toutes les notions de minimalité proposées dans la littérature, par exemple (Dalal 1988, Hegner 1987, Satoh 1988, Winslett 1990, Jackson Pais 1992). Notre notion de minimalité est un cas particulier de celle de (Katsuno Mendelzon 1991b, Grahne 1991), où la notion de minimalité a été généralisée en introduisant des ordres sur les mondes. G. Grahne, à la manière de Lewis, introduit un opérateur conditionnel supplémentaire dans le langage qui permet de structurer les connaissances. Cette structure, en termes d'ordre total, est employée pour calculer le résultat de la mise à jour. Nous pensons que d'un point de vue computationnel, tout comme dans l'approche du raisonnement non monotonique de type basé sur l'expectative (Gärdenfors Madinson 1992, Fariñas et al. 1992), le problème majeur de cette approche est que l'ordre complet doit être décrit de façon explicite.

Une de nos hypothèses principales jusqu'à maintenant a concerné le fait que des variables propositionnelles différentes n'interfèrent pas. L'exemple suivant, dû à F. Veltman, montre que ceci n'est pas réaliste : soit p et q des variables propositionnelles se lisant respectivement "le feu de circulation est au rouge" et "le feu de circulation est au vert". Comme p et q sont des variables propositionnelles différentes, il s'ensuit du postulat de non-interférence que $p \rightarrow q > (p \wedge q)$ est un théorème de la logique ASSUME. Mais il est absurde de penser que, après que le feu de circulation soit devenu vert, il continue d'être rouge ! Nous avons besoin d'informations supplémentaires capables de nous dire, en dépit du fait que les variables propositionnelles sont différentes, que p et q n'interfèrent pas.

Deux voies sont possibles : la première approche a été suivie dans (Ginsberg Smith 1988) ou dans (Segin 1992), où la notion de contrainte a été employée. Nous travaillons présentement dans une autre direction, où la notion d'interférence est prise en un sens élargi : nous supposons donnée une relation de similarité entre les formules, relation réflexive et symétrique entre les variables propositionnelles. Nous disons que deux formules interfèrent si elles sont similaires. Ainsi nous tenons compte d'une information métalinguistique concernant des formules qui sont reliées entre elles. Formellement, nous supposons donné une relation " $\sim$ " entre des formules qui est réflexive et symétrique (mais non transitive). De plus, nous

supposons que toute relation sur des formules peut être construite de ses parties qui ne concernent que des variables propositionnelles de la façon suivante :

$A \sim B$ ssi il y a un $p \in \text{pvar}(A)$ et un $q \in \text{pvar}(B)$ tel que $p \sim q$

L'axiome de (non-interférence) devient

$C \rightarrow A > C$ si non $(A \sim C)$

Avec cette nouvelle axiomatisation, la formule $p \rightarrow q > (p \wedge q)$ n'est plus un théorème, en autant que nous avons $p \sim q$.

REMERCIEMENTS

Nous remercions Gabriella Crocco, Didier Dubois et Alberto O. Mendelzon pour leurs précieux commentaires.

Bibliographie

- Alchourron et al. 85 C. E. Alchourron, P. Gärdenfors and D. Makinson. *On the Logic of Theory Change: Partial Meet Contractions and Revision Functions*. The Journal of Symbolic Logic. 50. 2. p. 510-530, 1985.
- Bibel et al. 89 W. Bibel, L. Fariñas del Cerro, B. Fronhöfer and A. Herzig. *Plan Generation by Linear Proofs : On Semantics*. German Workshop on AI (GWAI 89). Springer Verlag 1989.
- Brewka Hertzberg 91 G. Brewka, J. Hertzberg. *How do things with worlds : on formalizing actions and plans*. Internal report GMD, 1991.
- Chellas 75 B.F. Chellas. *Basic conditional logic*. J. of Philos. Logic, 4, 1975, p. 133-53.
- Chou Winslett 91 T.S. Chou, M. Winslett, *Immortal : A model-based belief revision system*. In : Proc. of KR'91.
- Crocco, Farinas 94. G. Crocco, L. Farinas del Cerro. *Structure, Consequence relation and logic : What is a logical system ?* Ed. D. Gabbay. Oxford university Press, 1994.
- Dalal 88 M. Dalal. *Investigations into Theory of Knowledge Bases Revision : Preliminary Report*. Proc. of the Seventh Nat. Conf. on AI, Minneapolis p. 475-479. 1988.
- Del Val 92 A. Del Val, *Computing knowledge base updates*. Proc. of KR'92.
- Dubois Prade 91 *Epistemic entrenchment and possibilistic logic*. Journal of AI 50 : 223-239.
- Eiter Gottlob 92 Th. Eiter, G. Gottlob, *On the complexity of propositional knowledge base revision, updates, and counterfactuals*. Journal of AI 57 (1992), p. 227-270.
- Fariñas Herzig 88 L. Fariñas del Cerro and A. Herzig. *An automated modal logic for elementary changes*. Non-Standard Logics for Automated Reasoning (ed. P Smets, A, Mandani, D. Dubois and H. Prade). Academic Press, 1988, p. 63-79.
- Fariñas et al. 92 L. Fariñas del Cerro, A. Herzig, J. Lang. *Ordering-based nonmonotonic reasoning..* Artificial Intelligence Journal. 66, 375-393, 1994
- Gärdenfors 88 P. Gärdenfors. *Knowledge in Flux*. MIT Press, 1988.

- Ginsberg Smith 88 M. L. Ginsberg, D.E. Smith. *Reasoning about action I : A possible world approach*. Readings in Nonmonotonic Reasoning, M.L. Ginsberg ed., Morgan Kaufmann, 1987.
- Girard 86 J.-Y. Girard, *Linear logic*. Theoretical Computer Science 50 (1987) p. 1-102.
- Grahne 91 C. Grahne. *Updates and Contrafactuals*. Principles of Knowledge Representation and Reasoning. Morgan and Kaufmann, 1991.
- Grove 86 A. Grove. *Two modelings for Theory Change*, Auckland Philosophy Papers 13, 1986.
- Hegner 87 S. Hegner. *Specification and Implementation of Programs for Updating Incomplete Information Databases*. Proc. of the Sixth Symp. on Principles of Database Systems. San Diego, p. 146-158, 1987.
- Katsuno Mendelzon 91a H. Katsuno and A.O. Mendelzon. *Propositional knowledge base revision and minimal change*. AI Journal 52, 263-294, 1991.
- Katsuno Mendelzon 91b H. Katsuno and A.O. Mendelzon. *On the Difference between Updating a Knowledge Base and Revising it*. Principles of Knowledge Representation and Reasoning, Morgan Kaufmann 1991.
- Lewis 73 D. K. Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.
- Nute 80 Donald Nute. *Topics in Conditional Logic*. D. Reidel Publishing Company, 1980.
- Satoh 88 K. Satoh. *Nonmonotonic Reasoning by Minimal Belief Revision*. Proc. of the Int. Conf. on Fifth Generation Computer Systems. 1988 ICOT, p. 455-462.
- Segerberg 86 Krister Segerberg. *On the logic of small changes in theories, I*. Auckland Philos. Papers, 1986.
- Seguin 92 Christel Seguin. *De l'action à l'intention : Une caractérisation formelle des agents*. Phd. thesis, Universié Paul Sabatier, Toulouse, march 1992.
- Winslett 88 M. Winslett. *Reasoning about actions*. Proc. AAAI 1988, p. 89-93.
- Winslett 90 M. Winslett. *Updating Logical Databases*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press 1990.

Luis FARIÑAS DEL CERRO

Andreas HERZIG

PARTIE II

LOGIQUE ET REPRÉSENTATION

Chapitre 7

Logique de la connaissance : un panorama

Bien que les théorèmes de la logique classique ne contiennent pas les expressions "vrai" ou "faux", cette logique est une *logique de la vérité*. Mais elle l'est *implicitement*, en ce qu'elle ne distingue pas les énoncés de même valeur de vérité, et qu'elle ne répertorie que les inférences valides, c'est-à-dire celles qui transfèrent à leur conclusion la vérité éventuelle de leurs prémisses. Peut-on construire une *logique de la connaissance* obéissant au même schéma ? La connaissance est essentiellement véridique (si ϕ est sue, alors elle est vraie), mais la réciproque ne vaut pas : il existe des vérités inconnues. Une logique de la connaissance doit donc être une logique non classique : certains principes traditionnels, comme le tiers exclu $\phi \vee \neg\phi$, qui sont indiscutablement corrects dans l'interprétation "ontique" (ϕ est vraie, ou bien $\neg\phi$ est vraie), deviennent inacceptables dans l'interprétation "épistémique" (ϕ est connue, ou bien $\neg\phi$ est connue). Sauf à décrire le savoir d'individus omniscients, pour lesquels connaissance et vérité s'égalertaient, il faut donc consentir à des modifications dans la logique usuelle.

Ces modifications peuvent être de deux ordres :

(i) Ou bien l'on veut *préserver* la sémantique vériconditionnelle et les principes fondamentaux qui définissent la logique classique, et il convient alors d'en *étendre* le langage de base en y ajoutant un opérateur épistémique *explicite* (la signification attendue de la formule $K\phi$ étant : ϕ est connue). Le tiers-exclu, par exemple, continue à s'appliquer à la logique ainsi étendue sous la forme non problématique $K\phi \vee \neg K\phi$ (ϕ est connue, ou bien inconnue).

(i) Ou bien l'on cherche à interpréter *implicitement* le langage de base lui-même en termes épistémiques, et à *remplacer* la question de savoir à quelles conditions un énoncé est vrai par celle de savoir à quelles conditions son assertion est justifiée. Il convient alors de *réviser* la logique : puisque l'assertion de $\phi \vee \psi$ n'est justifiée que si ϕ est justifiée ou si ψ est justifiée, le principe du tiers-exclu doit être abandonné (un agent peut être à la fois incapable de justifier ϕ et incapable de justifier $\neg\phi$).

L'alternative précédente sépare deux programmes de recherche distincts en logique de la connaissance. Le premier, inauguré par Hintikka au début des années 1960¹, est celui d'une logique *explicitement épistémique* (la *logique épistémique* tout court), qui définit la connaissance *statiquement*, par l'état d'information en lequel elle consiste, et dans lequel un agent donné peut ou non se trouver. Le second programme, qui remonte à Brouwer et Heyting dans les années 1930, est celui d'une logique *implicitement épistémique*, qui définit la connaissance *dynamiquement*, en la caractérisant de manière inductive par l'ensemble des actions (preuves ou vérifications) qu'il incombe à tout agent d'effectuer pour y parvenir².

Ces deux traditions interfèrent peu, et se sont développées dans une ignorance et une indifférence mutuelles à peu près totales. La logique épistémique a été créée par des philosophes avant tout soucieux de définir explicitement les conditions de vérité de "x sait que ϕ ", et peu intéressés par le détail syntaxique des processus de preuve. La logique intuitionniste, de son côté, s'est si fructueusement appliquée à l'informatique théorique que les préoccupations philosophiques des fondateurs ont été quelque peu perdues de vue. Les pages qui suivent ont pour objectif de donner une analyse comparative des deux programmes³ et une présentation de leurs développements les plus récents.

1. Connaissance, sens et référence

En logique classique, la valeur sémantique d'une expression est sa pure et simple référence (pour les désignateurs, l'objet désigné, et pour les énoncés, la valeur de vérité). Cette logique est *compositionnelle* : chaque fois qu'une règle morphologique permet de construire une expression $* \langle x_1, \dots, x_n \rangle$ à partir d'expressions $x_1, \dots, x_n$, il y a une règle permettant de déterminer la référence de l'expression

1. *Knowledge and Belief. An Introduction to the Logic of the Two Notions*, Cornell U.P., 1962.

2. Heyting caractérise comme suit la logique intuitionniste (qu'il nomme "logique de la connaissance" par opposition à la logique classique, qualifiée quant à elle de "logique de l'existence") : les théorèmes de cette logique expriment "le fait que, si nous connaissons une preuve de certains théorèmes, alors nous connaissons aussi la preuve d'un autre théorème." (*Intuitionism in Mathematics*, in R. Klibansky (ed.), *Philosophy in Mid-Century. A Survey*, La Nuova Italia Editrice, Florence, 1958, I, p. 107).

3. On trouvera une discussion pénétrante de l'ensemble du problème dans J. van Benthem, *Reflections on Epistemic Logic, Logique et Analyse*, XXXIV-1991, p. 5-14.

complexe sur la base de la référence de ses constituants. C'est le principe d'extensionnalité

$$(Ext) \quad \text{Ref} (* <x_1, \dots, x_n>) = f_* (\text{Ref}(x_1), \dots, \text{Ref}(x_n))$$

(la règle $f_{\&}$ pour la conjonction, par exemple, est donnée par la table de vérité familière de $\&$).

Les choses ne peuvent en aller ainsi pour la logique de la connaissance, et la forme que prend cette altération dépend pour l'essentiel du point suivant : veut-on s'en tenir à une caractérisation *interne* de la connaissance, ou bien estime-t-on au contraire possible ou nécessaire d'expliquer ou de représenter le fait que la connaissance est une expérience *au sujet du monde extérieur* ?

1.1. *L'intuitionnisme, une logique du sens*

L'intuitionnisme développe la première branche de l'alternative précédente. Sa logique est de part en part épistémique, et ne conserve aucun fragment "ontique" dans lequel on pourrait parler de la "réalité", connue ou non, fût-ce pour dire que c'est d'elle que la connaissance est vraie :

"Le programme de Brouwer (...) consiste à explorer les constructions mathématiques mentales en tant que telles, sans référence aux questions qui touchent à la nature des objets construits, telle que celle de savoir si ces objets existent indépendamment de la connaissance que nous avons d'eux. (...) Les mathématiques, du point de vue intuitionniste, sont l'étude de certaines fonctions de l'esprit humain." ⁴

La référence est donc radicalement mise entre parenthèses, et il ne subsiste, au titre de valeur sémantique, que la manière dont elle se donne, c'est-à-dire le sens dans l'acception frégréenne du terme (*die Art des Gegenbenseins* ⁵) : dans le cas d'un énoncé, c'est la procédure par laquelle son éventuelle vérité est reconnue, en bref sa justification. L'intuitionnisme ne laisse donc aucune place à l'idée de valeurs de vérité que les agents pourraient être éventuellement incapables de déterminer : du point de vue philosophique, c'est un antiréalisme au sens que Dummett donne à ce mot ⁶.

Le sens doit être compositionnel, ce qui suppose, par analogie avec le principe (Ext) qui régit la référence, que soit défini ce qui compte comme une justification de $* <x_1, \dots, x_n>$ en fonction de ce qui constitue une justification de $x_1, \dots$, et de x_n .

4. A. Heyting, *Intuitionism. An Introduction*, North-Holland Publ., 3^e édit., 1976, p. 1 et 10.

5. Frege, *Sinn und Bedeutung*, in *Kleine Schriften* (I. Angelelli ed.), G. Olms Verlag, 2^e édit., 1990, p. 144.

6. Cf. M.A.E. Dummett, *The Philosophical Basis of Intuitionistic Logic* (1973), repris dans *Truth and Other Enigmas*, Duckworth, 1978, p. 215-247, trad. française dans *Philosophie de la logique*, Editions de Minuit, 1991, p. 79-131.

Dans la tradition de l'interprétation BHK (Brouwer-Heyting-Kolmogoroff), on donnera les clauses suivantes ⁷ :

- une preuve de $\phi \ \& \ \psi$ est la donnée d'une preuve pour chacun des énoncés ϕ et ψ
- une preuve de $\phi \vee \psi$ est la donnée d'une preuve pour l'un au moins des énoncés ϕ et ψ
- une preuve de $\phi \rightarrow \psi$ est une construction qui transforme toute preuve de ϕ en une preuve de ψ
- une preuve de $\neg\phi$ est une preuve de $\phi \rightarrow \perp$, $\perp$ étant une absurdité quelconque du type $0=1$.

C'est cette définition des connecteurs de base qui assure la portée épistémique de la logique.

1.2. La logique épistémique, une logique de la référence indexée

Contrairement à l'intuitionnisme, la logique épistémique conserve une dimension ontique, c'est-à-dire que l'on peut y parler *aussi* d'une réalité extra-mentale⁸. La difficulté essentielle d'une "Grande Logique" capable de cela concerne l'articulation des deux fragments, ontique et épistémique. La logique épistémique résout ce problème de modularité en *adjoignant* simplement une superstructure épistémique à la logique classique. Elle doit donc exprimer en termes non épistémiques les conditions de vérité des phrases épistémiques, et doit par conséquent résoudre le problème suivant.

1.2.1. Le problème de l'opacité référentielle des contextes épistémiques

Puisque ϕ peut être vrai sans que $K\phi$ le soit aussi, il n'existe aucune fonction K_* permettant de définir $\text{Ref}(K\phi)$ à partir de la seule donnée de $\text{Ref}(\phi)$. Le contexte $K[\dots]$ est donc opaque : on ne peut pas y remplacer l'un par l'autre deux énoncés de même valeur de vérité. Un autre aspect de cette opacité est le défaut de substituableté des termes codésignatifs, c'est-à-dire l'impossibilité bien connue de conclure de $K\phi[\dots v \dots]$ et de $v = f(u)$ à $K\phi[\dots f(u) \dots]$ (Oedipe sait qu'il épouse Jocaste, mais ignore qu'il épouse sa mère). La référence des phrases épistémiques ne dépend pas uniquement de la référence de leurs constituants.

7. Heyting, op. cit., chap VII.

8. En logique intuitionniste, on ne peut même pas *dire* qu'il existe une preuve non-constructive d'un certain théorème mathématique.

1.2.2. La "solution" de Quine

Quine propose sur ce point une solution radicale, qui consiste, d'une certaine manière, à résorber l'épistémique dans l'ontique, et qui est donc à l'exact opposé de ce qu'entreprend l'intuitionnisme. Il suggère en effet⁹ d'écarter les apparences grammaticales, et de traiter les phrases épistémiques comme des phrases atomiques du langage de base. Dire de Julie qu'elle sait que la Seine coule à Paris consisterait donc, en réalité, à lui attribuer une propriété indécomposable, celle de savoir-que-la-Seine-coule-à-Paris, la proposition subordonnée introduite par "que" ne figurant alors dans la phrase complète que comme "chat" figure dans "achat", c'est-à-dire en vertu d'un pur accident orthographique. Mais le prix à payer pour ce traitement de l'opacité référentielle est manifestement trop élevé : d'une part un langage ainsi construit n'aurait aucune chance de pouvoir être appris, d'autre part les phrases enchâssées dans les contextes épistémiques ne deviendraient plus accessibles à *aucune* transformation logique, en sorte qu'un tel système décrirait des agents aux capacités déductives *nulles*, dont les connaissances ne pourraient être explicitées que par des listes de phrases essentiellement inabrégeables. La solution ne peut pas être là.

1.2.3. L'idée de référence indexée

La solution de Quine écartée, la définition des conditions de vérité de $K\phi$ doit mettre en jeu *deux* composantes (l'opacité référentielle indique seulement l'insuffisance de la première) : d'une part la satisfaction de ϕ dans la "réalité objective", d'autre part l'occurrence dans l'esprit de l'agent décrit d'un certain état mental, lequel consiste, sommairement dit, à considérer ϕ comme correct, ou à ne pas envisager que ϕ ne soit pas le cas. Tel est le point de départ de la logique épistémique de Hintikka : de tous les scénarios logiquement possibles de la réalité, l'agent ne considère que ceux qui sont compatibles avec les informations qu'il détient, en d'autres termes ceux-là seuls qui pourraient être, pour autant qu'il sache, le scénario correct. Dire qu'il sait que ϕ est donc à dire à la fois que ϕ est vrai dans la "réalité" et que A est vrai dans *chacun* de ces scénarios alternatifs. La première clause pouvant aisément être absorbée dans la seconde, l'opérateur K équivaut en définitive à une quantification restreinte sur une classe W d'indices ou d'entités (les "mondes possibles"¹⁰) :

9. *Word and Object*, M.I.T. Press, Cambridge, 1960, p. 215-216.

10. La notion de monde possible, qui a évidemment son origine dans la philosophie de Leibniz, a été remise en honneur par Kripke en 1959 (*A Completeness Theorem in Modal Logic*, *Journal of Symbolic Logic*, XXIV-1959, p. 1-14), article consacré à la sémantique de (S5). Les "descriptions d'état" utilisées par Carnap dans *Meaning and Necessity* (Chicago U.P., 1947) sont les ancêtres directs des mondes possibles. On trouvera, dans l'introduction de Fr. Rivenc à la traduction française du livre de Carnap (Gallimard, à paraître) une comparaison des sémantiques de Carnap et de Kripke pour la logique modale.

(*) $\models_w K\phi$ si et seulement si $\models_{w'} \phi$ pour chaque $w' \in W$ tel que wRw'

(la relation R sur W , dite d'"accessibilité", a la signification intuitive suivante : w' est relié à w exactement si l'agent, dans le monde w , ne possède aucune information lui permettant d'exclure que w' soit le monde où il "vit").

Contrairement à l'interprétation BHK de la logique intuitionniste, qui est une sémantique du sens, la sémantique des mondes possibles reste, *lato sensu*, une sémantique référentielle, dans laquelle la référence des expressions est simplement *relativisée* aux points d'un certain ensemble : la particularité des énoncés épistémiques tient au seul fait que leur valeur de vérité en un point est susceptible de dépendre de la valeur de vérité de leurs constituants ailleurs qu'en ce point. Comme l'a écrit Hintikka,

"L'opacité référentielle n'est pas due ici à quelque chose d'étrange qui se produirait dans la manière dont nos termes singuliers réfèrent aux objets, ni à quelque chose d'inhabituel dans les objets auxquels ces termes sont censés référer. (...) Ce à quoi nous avons ici affaire n'est pas tant un défaut de référentialité qu'une sorte de *référentialité multiple*".¹¹

En d'autres termes, le principe d'extensionnalité est ici simplement *libéralisé* sous la forme

(Ext_M)

$$\text{Ref}_w (* \langle x_1, \dots, x_n \rangle) = f_* \left[w, \underset{w \in W}{\text{Ref}} (x_1), \dots, \underset{w \in W}{\text{Ref}} (x_n) \right]$$

1.2.4. Systèmes de logique épistémique

L'appellation "logique épistémique" est, d'une certaine manière, insidieuse et équivoque. D'une part elle peut convenir à tout système de logique comportant un opérateur de modalité doté de propriétés plus ou moins semblables à celles de l'expression "X sait que..." de la langue naturelle, si bien que c'est plutôt de logique modale interprétable en termes épistémiques qu'il conviendrait de parler. D'autre part l'analyse informelle du concept de connaissance, qui suffit à attester la correction intuitive du principe de véridicité

(T) $K\phi \rightarrow \phi$,

est en revanche moins concluante à l'égard de schémas comme

(4) $K\phi \rightarrow KK\phi$ (introspection positive)

ou, *a fortiori*,

(5) $\neg K\phi \rightarrow K\neg K\phi$ (introspection négative),

11. op. cit., p. 140.

en sorte que plusieurs systèmes de logique épistémique peuvent être en principe défendus. La sémantique des mondes possibles, qui est évidemment incapable de trancher entre les diverses familles de propriétés auxquelles la discussion philosophique est susceptible d'aboutir, peut en revanche donner une systématisation rigoureuse de la plupart d'entre elles. En effet la relation d'accessibilité peut se voir attribuer diverses propriétés algébriques, répondant terme à terme aux propriétés que l'on souhaite voir respecter par les opérateurs que l'on a en vue¹². En l'absence de toute stipulation supplémentaire sur R , l'ensemble des formules valides est axiomatisé par le système $\mathbf{K}$, qui contient toutes les tautologies propositionnelles, le schéma

$$(K) \quad K(\phi \rightarrow \psi) \rightarrow (K\phi \rightarrow K\psi)$$

et qui est clos pour les règles de substitution uniforme, de *modus ponens* et de nécessairement ($\phi / K\phi$).

L'adjonction à $\mathbf{K}$ des schémas (T), (4) et (5) a pour contrepartie sémantique la définition de conditions supplémentaires sur R (respectivement la réflexivité, la transitivité et l'eulidianité)¹³. Les systèmes $\mathbf{KT4}$ et $\mathbf{KT5}$ ((S4) et (S5) dans la terminologie traditionnelle de Lewis et Langford) jouent un rôle particulièrement important. Le premier est notamment lié à l'interprétation de la connaissance en termes de prouvabilité (cf. ci-dessous § 1.3.2.), et le second (dans lequel la relation d'accessibilité est donc une relation d'équivalence sur W), est couramment considéré comme le système modal canonique dans le domaine de la vérification des protocoles de communication.

12. La première mention explicite de la relation d'accessibilité remonte à l'article de Kripke de 1963 *Semantical Analysis of Modal Logic I, Normal Propositional Calculi*, *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, IX-1963, p. 67-96.

13. Ces exemples semblent accréditer d'une part l'idée d'une correspondance absolument générale entre schémas modaux et formules de la logique du premier ordre, et d'autre part le projet de décrire la connaissance dans la logique usuelle, en éliminant les modalités au profit des quantificateurs. Mais l'idée est erronée : certaines conditions du premier ordre, comme l'irréflexivité $\forall w \neg wRw$, n'ont aucune contrepartie modale, et inversement certains schémas modaux, comme $\Box \Diamond A \rightarrow \Diamond \Box A$, ne sont traductibles par aucune formule du premier ordre (la logique modale propositionnelle est en fait un fragment de la logique du second ordre (cf S.K. Thomason, *Reduction of Second-Order Logic to Modal Logic*, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, XXI-1975, p. 107-114)). Quant au projet, il est, sous sa forme naïve, inapproprié (cf. D. Scott, *Advice on Modal Logic*, in K. Lambert (ed.), *Philosophical Problems in Logic. Some Recent Developments*, D. Reidel Publ., 1970, p. 143), mais une forme plus sophistiquée ("fonctionnelle") de traduction est adéquate: on y remplace la donnée de R par la donnée d'une famille F d'applications de W dans lui-même, KA étant alors vrai en w exactement si A est vrai en $f(w)$ pour chaque $f \in F$ (cf. H.J. Ohlbach, *Semantics-Based Translation Methods for Modal Logics*, *Journal of Logic and Computation*, 1-1991, p. 691-746).

1.3. Sémantiques de la référence indexée : convergences et controverses

1.3.1. Une sémantique vériconditionnelle indexicale pour la logique intuitionniste

La logique intuitionniste peut être elle-même interprétée selon le principe (Ext_M), car elle possède de son côté une sémantique de type kripkéen¹⁴. On peut en effet concevoir que la connaissance du mathématicien idéal (le "sujet créatif", dans les mots de Brouwer) s'étend au fil du temps : ses états épistémiques successifs ne sont que partiellement ordonnés (ce qu'il sait aujourd'hui ne détermine pas totalement ce qu'il saura demain), mais sa connaissance est cumulative (il ne sera jamais contraint de réviser ce qu'il sait). Ceci revient à évaluer les énoncés en chaque point d'un ensemble W muni d'une relation réflexive et transitive $\leq$, en respectant la condition de *persistance*

(P) Si $\models_w \phi$, et si $w \leq w'$, alors $\models_{w'} \phi$

et en se conformant à des clauses récursives dont certaines font dépendre la vérité d'un énoncé complexe en un point de la vérité de ses constituants en d'autres points :

$\models_w \phi \ \& \ \psi$ si et seulement si $\models_w \phi$ et $\models_w \psi$

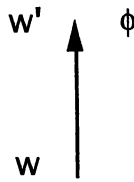
$\models_w \phi \ \vee \ \psi$ si et seulement si $\models_w \phi$ ou $\models_w \psi$

$\models_w \phi \rightarrow \psi$ si et seulement si $\models_{w'} \phi$ entraîne $\models_{w'} \psi$ en chaque w' tel que $w \leq w'$

$\models_w \neg \phi$ si et seulement si $\not\models_{w'} \phi$ en chaque w' tel que $w \leq w'$

La dernière clause, par exemple, est clairement une contrepartie de l'interprétation fonctionnelle de la négation donnée au § 1.1. : de même que $\neg \phi$ n'est pas justifié par la seule absence de justification de ϕ , mais par l'existence d'une construction capable de montrer que toute justification putative de ϕ pourrait être transformée en justification d'une absurdité, la satisfaction de $\neg \phi$ dans un état épistémique donné suppose non seulement l'absence de satisfaction de ϕ dans l'état considéré, mais aussi l'absence de cette satisfaction dans tout état ultérieur. Du reste il découle immédiatement de cette clause un contre-exemple au principe du tiers exclu :

14. Cf. S. Kripke, *Semantical Analysis of Intuitionistic Logic*, in J. Crossley & M.A.E. Dummett (eds.), *Formal Systems and Recursive Functions* (North-Holland Publ., 1965, p. 92-130) et M.C. Fitting, *Intuitionistic Logic, Model Theory and Forcing* (North-Holland Publ., 1969).



(on n'a ni $\models_w \phi$ ni $\models_w \neg\phi$).

1.3.2. Référence indexée, réalisme et antiréalisme

Dummett considère l'admission illimitée du principe de bivalence (*un énoncé est vrai, ou bien il est faux*) comme une caractéristique fondamentale du réalisme. Or en chaque nœud d'un arbre de Kripke pour la logique intuitionniste tout énoncé est déterminément vrai ou faux, et la référence à la vérité *en un point* ne semble guère éloignée de la référence à la vérité *simpliciter*. On peut donc se demander si l'existence d'une sémantique vériconditionnelle "indexicale" pour la logique intuitionniste n'est pas de nature à estomper la différence entre le réalisme, qui accepte l'idée de conditions de vérité nues, c'est-à-dire éventuellement capables de transcender les capacités recognitionnelles des agents, et l'antiréalisme, qui pour sa part s'y refuse. En fait il n'en est rien. Comme le remarque en effet Dummett ¹⁵, l'interprétation réaliste d'une sémantique de ce genre ne s'impose que lorsque le mode de détermination de la valeur de vérité de la phrase en un certain indice peut être réduit au mode de détermination usuel (vériconditionnel pur et simple) par le simple effet d'un ajustement de peu d'ampleur dans l'allure grammaticale de la phrase. C'est le cas, par exemple, avec la logique temporelle (où les indices sont conçus comme des instants déterminés), puisque l'évaluation à l'instant *t* d'un énoncé rédigé au présent intemporel est évidemment réductible à l'évaluation "absolue" du même énoncé mis au temps convenable et dans lequel le verbe principal se voit attacher la mention explicite de l'instant considéré. Il en va différemment dans le cas de la logique intuitionniste : à la différence d'expressions comme 'devant la mairie d'Hendaye' ou 'le 2 décembre 1994 à minuit', l'expression 'dans le monde possible *w*' n'est pas authentiquement référentielle, c'est-à-dire qu'elle ne peut pas être déplacée vers l'intérieur de l'énoncé qu'elle préfixe pour en faire un énoncé vrai ou faux en vertu de la seule constitution objective de la réalité actuelle¹⁶. La sémantique vériconditionnelle indexicale commune aux deux programmes de logique de la

15. *Realism* (1982), repris dans *The Seas of Language*, Oxford U.P., 1993, p. 268.

16. Il est vrai que D. Lewis (*On the Plurality of Worlds*, Basil Blackwell, 1986) soutient une variété de réalisme modal ("les mondes possibles sont ce qu'ils sont") dans laquelle les expressions 'dans le monde possible *w*' et 'sur Alpha du Centaure' possèdent justement des statuts désignatifs similaires. Mais cette thèse - qui soulève en elle-même bien des objections - concerne plutôt les mondes possibles dans l'interprétation métaphysique (logique modale "classique") que dans l'interprétation épistémique qui nous occupe ici.

connaissance est donc philosophiquement neutre à l'égard de la controverse entre l'antiréalisme et le réalisme. Cet œcuménisme, on va le voir, fait à la fois son intérêt et sa fragilité.

1.3.3. La traduction modale de Gödel : une voie de passage d'un programme à l'autre ?

L'existence d'une sémantique de Kripke pour la logique intuitionniste n'est pas véritablement surprenante : dans un article important¹⁷ de 1933, Gödel avait déjà montré l'existence d'une traduction modale du calcul intuitionniste des propositions **IPC**, c'est-à-dire la possibilité d'associer à toute formule propositionnelle ϕ une formule modalisée $\phi^\Box$ vérifiant

$$\vdash \text{IPC } \phi \text{ si et seulement si } \vdash (\text{S4}) \phi^\Box \quad 18$$

L'existence d'une sémantique de Kripke pour l'intuitionnisme découle donc immédiatement de l'existence d'une sémantique du même genre pour (S4).

La logique intuitionniste peut donc, en un certain sens, être plongée dans la logique épistémique. L'importance de ce "passage" entre l'intuitionnisme et la logique modale (et donc, potentiellement, la logique épistémique) ne doit cependant pas être surestimée, et il convient d'examiner avec circonspection la possibilité d'unifier, ou même seulement d'articuler *sur cette base* les deux programmes de logique de la connaissance apparemment incommensurables que nous avons distingués :

1) Bien que les deux types de logique puissent être interprétés dans la sémantique des mondes possibles, la signification intuitive de la relation d'accessibilité y est fort différente, puisqu'elle représente dans un cas la *compatibilité* avec l'information possédée par l'agent et dans l'autre la *croissance potentielle* de cette information.

2) La traduction de Gödel est à sens unique, et aucune route n'est connue pour revenir de la logique modale à l'intuitionnisme. En d'autres termes, s'il est toujours possible d'*expliciter* la portée épistémique éventuelle des connecteurs de base, on ignore toujours, en revanche, comment "*impliciter*" dans une "logique de base" un quelconque système de principes épistémiques explicites.

17. *Eine Interpretation des intuitionistischen Aussagenkalküls, Ergebnisse eines mathematischen Kolloquiums* 4, p. 34, repris in *Collected Works*, I, Oxford U.P., p. 300-302.

18. En vérité, Gödel n'avait établi que l'implication directe, et seulement conjecturé l'implication converse. On connaît aujourd'hui beaucoup de traductions de ce type, dont la plus simple est sans doute celle qui consiste à définir $\phi^\Box$ comme la formule qui résulte de ϕ en préfixant $\Box$ à chacune de ses sous-formules.

1.3.4. Le programme de Shapiro

Pour unilatérale qu'elle soit, la traduction de Gödel offre cependant au logicien classique la possibilité de représenter les notions intuitionnistes dans (une extension de) son propre langage, et fait donc apparaître la logique épistémique comme une sorte de *lingua franca* dans laquelle l'intuitionnisme pourrait être compris sans malentendu. C'est une situation prometteuse si l'on est convaincu que la perspective intuitionniste sur la connaissance est correcte, mais que l'on a également le souci de conserver en logique un fragment purement descriptif ou "ontique" (cf. § 1.2. ci-dessus) : il suffit alors de projeter la logique intuitionniste dans une extension épistémique du langage classique. Tel est, sommairement décrit, l'objectif du programme de "mathématique intensionnelle" de Shapiro, qui a déjà inspiré une extension de la traduction modale de Gödel à l'arithmétique intuitionniste¹⁹ et à la théorie des types intuitionniste (entendue au sens de Ramsey et non de Martin-Löf)²⁰. L'idée directrice est ici d'identifier la valeur sémantique que l'intuitionniste a en vue (la "vérité" au sens intuitionniste, si l'on veut²¹) à la pure et simple *prouvabilité absolue* ou "*en principe*", c'est-à-dire à la possibilité objective d'être justifié par des moyens corrects quelconques (par opposition à la *démontrabilité formelle* dans une théorie donnée²²). Ainsi, le système **EA** (*arithmétique épistémique*) de Shapiro contient un fragment épistémique, dont l'opérateur **K** est destiné à exprimer les propriétés de la prouvabilité absolue (**K** a les caractéristiques habituelles du $\Box$ dans (**S4**)), et dans lequel est interprété la logique intuitionniste : la disjonction intuitionniste $\phi \vee_{\text{int}} \psi$, par exemple, se traduit dans ce fragment par la disjonction classique $K\phi \vee_{\text{class}} K\psi$, et l'on obtient de cette manière une contrepartie de la "propriété de disjonction" bien connue en intuitionnisme : si $\vdash_{\text{EA}} K(\phi \vee \psi)$, alors $\vdash_{\text{EA}} K\phi$ ou $\vdash_{\text{EA}} K\psi$.

Le programme de Shapiro postule donc que la logique intuitionniste ne diffère pas *significativement* de la logique classique, pour autant que les énoncés connaissables (c'est-à-dire prouvables ou réfutables) en principe soient seuls concernés. En somme l'intuitionnisme serait compatible avec l'idée traditionnelle de vérité comme conformité à des faits, dès lors que les faits en question seraient des faits immanents, en principe cognitivement accessibles de par leur nature même : ϕ est "intuitionnistiquement" vrai exactement s'il existe une preuve de ϕ par des

19. N.D. Goodman, Epistemic Arithmetic is a Conservative Extension of Intuitionistic Arithmetic, *Journal of Symbolic Logic*, XLIX-1984, p. 192-203.

20. A. Scedrov, Extending Gödel's Modal Interpretation to Type Theory and Set Theory, in St. Shapiro (ed.), op. cit., p. 81-119.

21. Sur le malaise des intuitionnistes à l'égard du mot "vrai", cf. par exemple D. Prawitz: "Les philosophes intuitionnistes utilisent quelques fois vrai comme synonyme de la vérité en tant que connue, mais il s'agit manifestement d'un usage étrange et malencontreux" (Intuitionistic Logic: a Philosophical Challenge, in G.H. von Wright (ed.), *Logic and Philosophy*, Martinus Nijhoff Publ., 1980, p. 8).

22. Cf. sur ce point J. Myhill, Some Remarks on the Notion of Proof, *Journal of Philosophy*, LVII-1960, p. 463-471.

moyens corrects quelconques. On peut douter qu'une telle réinterprétation, qui constitue une sorte de "platonisme des preuves", puisse parvenir à rendre justice à l'aspect *constructif* de la logique intuitionniste. Cette dernière est destinée à expliquer comment les preuves se construisent, non comment elles sont "découvertes"...

1.4. Critiques des sémantiques de la référence indexée

1.4.1. La critique opérationnaliste

La sémantique de Kripke pour l'intuitionnisme est mathématiquement irréprochable, et elle fournit effectivement les résultats de complétude que l'on attendait d'elle. Mais il existe une manière d'introduire l'intuitionnisme comme logique de la connaissance à laquelle cette sémantique est loin de rendre pleinement justice. Selon Prawitz²³, la question à laquelle l'intuitionnisme entend répondre est avant tout celle des *comportements* et des *actions* par lesquels la connaissance (qui est, en elle-même, une entité "essentiellement inobservable") peut être *communiquée* et *publiquement manifestée*²⁴. De ce point de vue, savoir que ϕ consiste pour l'essentiel à pouvoir en donner une justification, et comprendre ' ϕ ' consiste, de même, à pouvoir reconnaître dans quelles conditions l'assertion de ϕ serait justifiée, et non pas simplement à pouvoir répéter homophoniquement les conditions de vérité de ϕ . La sémantique "fonctionnelle" BHK répond à cette question, puisqu'elle caractérise la signification d'un connecteur logique quelconque $*$ en indiquant le type d'opération capable de justifier un énoncé dont $*$ est le connecteur principal, en fonction du type d'opération capable de justifier les composants immédiats de cet énoncé. Mais cet aspect *opérateur* est absent de la sémantique de Kripke, et un logicien sensible aux arguments philosophiques de Prawitz ne donc peut guère lui reconnaître qu'un rôle *heuristique* dans l'analyse de la connaissance.

1.4.2. La critique représentationnaliste

A l'autre extrémité du spectre, c'est-à-dire dans le camp réaliste — mais aussi chez les *utilisateurs* potentiels, agnostiques par profession, de la logique de la

23. *Meaning and Proofs : On the Conflict between Classical and Intuitionistic Logic, Theoria*, XLIII-1977, p. 2-40 (spécialement § 2 : *On the Explanation of Knowing*).

24. Il convient de noter que les bases philosophiques sur lesquelles Prawitz (ou Dummett) entendent construire l'intuitionnisme sont diamétralement opposées aux conceptions de Brouwer lui-même, selon qui : "Faute de pluralité de l'esprit, il n'y a pas non plus d'échange de pensées. Les pensées sont inséparablement liées au sujet (...) Dans ce qu'on appelle échange de pensées avec autrui, le sujet ne fait que toucher la paroi d'un automate." (*Consciousness, Philosophy, and Mathematics* (1948), in A. Heyting (ed.), *Collected Works I. Philosophy and Foundations of Mathematics*, North-Holland Publ., 1975, p. 485).

connaissance²⁵ — la sémantique des mondes possibles est tout aussi contestée. On l'accuse ici non pas d'être trop proche de la sémantique vériconditionnelle ordinaire, mais au contraire d'en être trop éloignée. Car les conditions de vérité proposées par la logique épistémique pour l'énoncé 'X sait que ϕ ', à savoir la vérité de ϕ dans tout monde compatible avec ce que sait X, ne font aucune mention de l'élément de réalité capable d'avérer cet énoncé (son *true-maker*, si l'on veut), à savoir que X traverse actuellement un certain épisode mental du type "acquiescement à ϕ " (je passe désormais sous silence l'autre "élément de réalité" en jeu, à savoir ϕ lui-même). Or c'est bien cet état mental actuel, *en vertu duquel* 'X sait que ϕ ' est vrai, qui doit jouer un rôle dans l'*explication causale* du comportement de X (par exemple X est disposé à agir de telle et telle façon sur la base de sa connaissance de ϕ). La classe des mondes possibles "sélectionnés" par la connaissance de X — celle des mondes dans lesquels ϕ est vérifiée — n'a, quant à elle, aucun rôle à y jouer : les mondes possibles n'ont aucune espèce d'influence causale sur l'individu en question. Bien peu de psychologues, à cet égard, seraient tentés de suivre Hintikka lorsqu'il écrit que la connaissance doit être *littéralement* comprise comme une relation directe ("intentionnelle") entre un agent et une proposition ou un ensemble d'entités abstraites :

"Savoir, pour autant que l'on ne sait pas tout, c'est « avoir l'œil sur plus d'un monde possible » (*to keep an eye on more than a "possible world"*)"²⁶.

La plupart d'entre eux considèrent qu'une telle analyse est en attente d'une réduction "scientifique", et préfèrent mettre en avant une conception *représentationnaliste* des attitudes propositionnelles, qui en fait plutôt des relations entre l'agent et des *noms* de propositions. Comme l'écrit en effet Fodor,

On ne voit pas du tout "comment un organisme pourrait être dans une relation épistémique intéressante avec une proposition, si ce n'est en étant dans une relation fonctionnelle/causale avec une énonciation (*token*) de la formule qui exprime la proposition"²⁷.

La connaissance ainsi "naturalisée" devrait donc "se rapporter à quelque chose qui ressemble plus aux énoncés entre guillemets qu'aux propositions abstraites"²⁸. Dans une logique de la connaissance construite selon les principes représentationnalistes, l'expression 'X sait que ...' devra donc être traitée non comme un *opérateur propositionnel*, mais comme un *prédicat métalinguistique* pouvant être ou non satisfait par le code d'un énoncé (on retrouve par ce biais le traitement de la modalité recommandé par Quine dans un texte bien connu²⁹).

25. Cf. par exemple sur ce point B. Hall-Pardee, *Possible-Worlds Semantics and Linguistic Theory*, *The Monist*, XL-1977, p. 303-326, et notamment p. 305 sq.

26. *Knowledge and Belief*, Cornell U.P., 1962, p. 138 sq.

27. *Representations*, M.I.T. Press, 1981.

28. *Computation and Cognition*, Bradford, Cambridge, 1984.

29. Selon Quine, les contextes de citation sont *par excellence* les contextes dans lesquels la transparence référentielle est en défaut, et le mieux que puisse faire l'analyse des contextes modaux est encore de se régler sur l'analyse du discours citationnel. Cf. *Three*

1.5. Les paradoxes du représentationnalisme

Le programme représentationnaliste que l'on vient de décrire se heurte à une difficulté logique sur laquelle Gödel avait attiré l'attention dans les dernières lignes de son article de 1933 sur la traduction modale de l'intuitionnisme, et qui a été ultérieurement précisée par un résultat de Montague³⁰. La théorie **T** d'un opérateur épistémique à la Hintikka satisfaisant au moins aux schémas de véridicité et d'introspection positive vérifie certainement les principes suivants :

$$\vdash_{\mathbf{T}} K\phi \rightarrow \phi \quad (\text{ce qui est su est vrai})$$

$$\vdash_{\mathbf{T}} K(K\phi \rightarrow \phi) \quad (\text{la propriété précédente est elle-même connue})$$

$$\vdash_{\mathbf{T}} K(\phi \rightarrow \psi) \rightarrow (K\phi \rightarrow K\psi) \quad (\text{le } \textit{modus ponens} \text{ est maîtrisé})$$

$$\text{Si } \phi \text{ est une tautologie, alors } \vdash_{\mathbf{T}} K\phi$$

Nous pouvons ajouter sans contradiction au système **T** ainsi obtenu la conjonction **Q** de la classe finie des axiomes de l'arithmétique de Robinson. Mais si nous décidons de traiter l'opérateur **K** comme un prédicat attaché au nom (au "numéro de Gödel") des énoncés, nous obtenons un système **T'** vérifiant :

$$(I) \quad \vdash_{\mathbf{T}'} Q$$

$$(II) \quad \vdash_{\mathbf{T}'} K(\ulcorner \phi \urcorner) \rightarrow \phi$$

$$(III) \quad \vdash_{\mathbf{T}'} K(\ulcorner K(\ulcorner \phi \urcorner) \rightarrow \phi \urcorner)$$

$$(IV) \quad \vdash_{\mathbf{T}'} K(\ulcorner \phi \rightarrow \psi \urcorner) \rightarrow (K(\ulcorner \phi \urcorner) \rightarrow K(\ulcorner \psi \urcorner))$$

$$(V) \quad \text{Si } \phi \text{ est une tautologie, alors } \vdash_{\mathbf{T}'} K(\ulcorner \phi \urcorner)$$

Or un tel système est incohérent : le prédicat **K** ainsi défini possède toutes les propriétés du prédicat de prouvabilité **Bew** de Gödel, et le "principe de réflexion" $K(\ulcorner \phi \urcorner) \rightarrow \phi$ ne devrait donc pas être prouvable dans **T'**.

Grades of Modal Involvement (1953), repris dans *The Ways of Paradox and other Essays*, 1966, Harvard U.P., p. 158-178.

30. R. Montague, *Syntactical Treatment of Modality, with Corollaries on Reflection Principles and Finite Axiomatizability* (1963), repris in R. Thomason (ed.), *Formal Philosophy. Selected Papers of Richard Montague*, Yale U.P., 1974, p. 286-302. Le résultat discuté ici est le lemme 3.

En effet si $K(\lceil 0 = 1 \rceil) \rightarrow 0 = 1$ était un théorème de T' , $\neg K(\lceil 0 = 1 \rceil)$, qui exprime la cohérence de T' , en serait également un, et ceci n'est possible, compte tenu du second théorème d'incomplétude de Gödel, qu'au prix de l'incohérence de T' elle-même. Thomason a montré depuis³¹ que cette impasse dans la logique de la connaissance d'inspiration représentationnaliste affectait également la théorie des attitudes "solipsistes", comme la croyance, qui n'entraînent pas la vérité de leur corrélat.

S'agissant de la logique modale "pure", des résultats de ce type semblent donc établir, comme le remarque Montague³², que

"Si la nécessité doit être traitée syntaxiquement, c'est-à-dire comme un prédicat d'énoncés, conformément aux recommandations de Carnap et de Quine, alors c'est virtuellement toute la logique modale qu'il faut sacrifier".

(à quoi Quine répond sobrement : "tant pis pour elle... !").

S'agissant de la logique de la connaissance, le résultat de Montague constitue le "paradoxe du connaissant" (*Knower's Paradox*) : si la conception représentationnaliste est correcte, alors à tout agent dont les représentations sont structurées, qui connaît au moins les lois de la logique (principe V) et dont le savoir s'enchaîne aux structures de ses représentations suivant les lois intuitivement vraisemblables (II)-(IV), nous sommes capables, moyennant un peu d'arithmétique, d'attribuer deux connaissances contradictoires. Les recherches actuelles³³ pour construire une logique des attitudes propositionnelles conforme aux principes représentationnalistes n'ont pas encore trouvé de parade vraiment convaincante à ce genre de difficulté.

2. Perspectives actuelles en logique de la connaissance

Pour une vue cavalière, le paysage en logique de la connaissance n'est pas aujourd'hui très significativement différent de celui qui pouvait être observé dans les années 1960 : il demeure deux programmes de recherche dominants — l'un consacré à une analyse explicite de la connaissance et utilisant des méthodes essentiellement sémantiques, l'autre d'orientation intuitionniste et massivement tourné vers la syntaxe et la théorie de la démonstration —, dont aucun n'a supplanté l'autre, mais qui persistent à coexister sans interférence. A y regarder de plus près, chacun de ces programmes a pourtant connu un infléchissement significatif dans les années 1980, infléchissement qui l'a conduit, chacun pour sa part, à explorer certaines caractéristiques de la connaissance tout à fait négligées par l'autre :

31. A Note on Syntactical Treatments of Modality, *Synthese*, XLIV-1980, p. 391-395.

32. op. cit., p. 294.

33. Cf. par exemple D. Perlis, Languages with Self-Reference, I : Foundations, II. Knowledge, Belief and Modality, *Artificial Intelligence*, resp. XXV-1985, p. 301-322 et XXXIV-1988, p. 179-212.

1) la logique épistémique issue des travaux de Hintikka s'est appliquée avec succès à la description formelle de la connaissance dans les systèmes d'agents (la logique d'inspiration intuitionniste, reste, quant à elle, dénuée de toute dimension collective, et l'imagerie brouwérienne du "sujet créatif" solitaire et solipsiste³⁴ en demeure un assez bon symbole).

2) la logique intuitionniste apparaît aujourd'hui comme un élément d'un vaste ensemble de logiques "structurales" qui semblent capables d'encoder les mécanismes inférentiels d'agents dotés de ressources déductives sévèrement limitées (par contraste, l'hypothèse de l'"omniscience logique", qui impute aux agents décrits une compétence déductive sans bornes, est profondément enracinée dans les principes fondamentaux de la logique épistémique usuelle).

Cette évolution laisse entrevoir l'idée d'une *spécialisation* des deux programmes : à la logique "philosophique" reviendrait la systématisation du discours "intentionnel" des sciences humaines et des sciences sociales, à la logique intuitionniste et néo-intuitionniste l'analyse des procédures effectives à l'œuvre dans les systèmes de traitement de l'information. Bien que cette division du travail doive probablement être d'ores et déjà considérée comme une réalité — y compris, dans nombre de cas, comme une réalité institutionnelle et sociologique, voire, si l'on ose dire, mentale —, il n'est pas certain que sa perpétuation soit une chose que l'on pourrait souhaiter : elle explique peut-être en partie, par exemple, la stagnation dans la mise au point d'une définition satisfaisante de la rationalité limitée dans le domaine des sciences économiques³⁵.

2.1. La connaissance en société

La vie sociale repose sur la capacité générique de chacun à former et à utiliser une représentation interne de soi-même et des autres. Une part cruciale de la connaissance *détenue* par un individu socialisé consiste donc en connaissances relatives à ses partenaires et aux connaissances qu'ils détiennent eux-mêmes. Ainsi, l'analyse de la vie sociale suppose une théorie correcte des connaissances enchâssées (ou d'ordre supérieur). Les travaux conduits en logique "multi-épistémique" (dont le langage comporte des modalités indexées K_i (*l'agent i sait que ...*)) depuis le début des années 1980 fournissent le cadre général d'une telle théorie³⁶. On se contentera ici d'attirer l'attention du lecteur des deux chapitres suivants, qui leur sont expressément consacrés, sur la richesse des interactions mathématiques que ces travaux ont suscitées : mise au point de nouvelles sémantiques pour la logique modale, théorèmes d'isomorphisme avec des structures hiérarchiques déjà par ailleurs connues

34. Cf. ci-dessus, note 24.

35. Cf. H.A. Simon, *Models of Bounded Rationality*, M.I.T. Press, 1982.

36. Pour une introduction rudimentaire, on peut se reporter à J. Dubucs, *The Logical Way of Describing Societies*, *Revue Internationale de Systémique*, VIII-1, 1994, p. 123-132.

en théorie des jeux, applications de la théorie des ensembles non fondés dans l'étude des points fixes des opérateurs épistémiques, etc.

2.2. Rationalité limitée : vers la fin des logiques immaculées

On aura déjà noté que la logique de la connaissance est d'autant moins classique que les agents qu'elle décrit en savent moins. Une logique capable de décrire les agents ordinaires, qui en savent ordinairement peu, a donc toute chance de différer assez radicalement des systèmes que la tradition a pu nous léguer.

2.2.1. L'impasse de la logique épistémique comme discipline descriptive

Le savoir — et donc l'ignorance — a deux composantes : empirique, et inférentielle. La logique épistémique traite adéquatement de l'ignorance empirique : le spectre est largement ouvert, depuis celui qui sait toute chose, et qui n'accède donc qu'à la réalité, jusqu'à celui qui ne sait rien, et à qui tout est donc accessible. Elle est, en revanche, mal équipée pour traiter de l'ignorance inférentielle : comme l'expose en détail le chapitre 10. de ce livre, l'hypothèse de l'"omniscience logique" est profondément enracinée dans ses principes fondamentaux. Cette situation ne disqualifie pas totalement la logique épistémique : nous avons certainement besoin aussi d'une approche normative de la rationalité, capable par exemple de nous indiquer ce qu'un individu *devrait* entreprendre, compte tenu des informations dont il dispose et des objectifs qu'il poursuit. Mais elle ne la qualifie pas pour proposer une approche descriptive d'agents qui se comportent comme des systèmes de traitement de l'information : le calcul est exactement le genre d'activité dont des individus logiquement omniscients pourraient se dispenser sans inconvénient majeur³⁷.

Selon une opinion répandue, on pourrait obtenir une logique épistémique descriptivement adéquate à partir de la logique épistémique usuelle en *soustrayant*, si l'on ose dire, l'omniscience logique de cette dernière. Il y a des manières plus ou moins heureuses de procéder. En remplaçant la sémantique de Kripke par la sémantique topologique de Scott-Montague³⁸, qui est, en un sens, beaucoup plus naturelle, on parvient à éliminer certaines formes de l'omniscience logique, mais les énoncés "intensionnellement" équivalents restent encore substituables dans les contextes épistémiques (un agent qui sait que ϕ est supposé savoir que ψ , pour chaque énoncé ψ logiquement équivalent à ϕ). Aller au-delà, et adopter, sous une forme ou sous une autre, une sémantique des mondes possibles "impossibles" ³⁹ permet sans doute de décrire des comportements déductifs arbitrairement déviants,

37. Cf. sur ce point J. Dubucs, *On Logical Omniscience*, in *Logique et Analyse*, 133-134 (1991), p. 41-55.

38. Cette sémantique est exposée p. ? du chapitre suivant.

39. L'article de base en ce domaine est celui de V. Rantala, *Impossible Worlds Semantics and Logical Omniscience*, in I. Niiniluoto & E. Saarinen (eds.), *Intensional Logic: Theory and Applications*, Acta Philosophica Fennica, 1982, p. 1-9.

mais autorise aussi un comportement erratique des connecteurs de base, dont la signification n'est donc plus tenue rigoureusement fixe⁴⁰.

2.2.2. De la sémantique vers la syntaxe

Pourquoi cette impasse ? Afin de rendre la logique épistémique descriptivement adéquate, on cherche à forger une sémantique qui ne validerait, parmi les énoncés attribuant une certaine compétence inférentielle aux agents, que les énoncés corrects eu égard à la compétence réellement manifestée par les agents de la classe observée. Tout repose ici sur l'ingéniosité du sémanticien, car il n'y a aucun lien *interne* entre les concepts de la sémantique et les capacités cognitives d'une quelconque classe d'agents. La situation est différente avec des concepts syntaxiques comme ceux de preuve ou de déductibilité, qui sont, d'une certaine manière, *intrinsèquement épistémiques*. Les conditions auxquelles ψ est déductible de ϕ diffèrent des conditions auxquelles ψ est conséquence logique de ϕ , car il est de l'essence des premières, mais non des secondes, de pouvoir être reconnues lorsqu'elles sont réalisées : comme le remarque en substance Dummett⁴¹, il serait en effet *absurde* que la condition même à laquelle nous reconnaissons qu'un énoncé découle logiquement de certaines hypothèses ne soit pas elle-même reconnaissable lorsqu'elle est satisfaite. La relation de déductibilité *ne* peut donc *pas* être définie indépendamment du type de procédures qu'une certaine classe d'agents est en principe capable de mettre en œuvre : sa définition est assujettie à respecter à la fois une exigence logique de *correction* (les déductions doivent transmettre à leur conclusion la vérité éventuelle de leurs prémisses, en sorte que ce qui est déductible de ϕ soit aussi conséquence de ψ) mais aussi une exigence épistémique de *faisabilité* (s'il existe une déduction Π de ψ à partir de ϕ , les agents de la classe considérée doivent être en mesure de produire un exemplaire de Π par le seul jeu de leurs ressources cognitives propres). Cette exigence épistémique se manifeste dans la définition des constantes logiques en termes de conditions d'assertabilité (cf ci-dessus, § 1.1.). Laquelle définition assure, de proche en proche, que ψ ne pourra jamais être déduit de ϕ sans que les constructions mentales qui justifieraient ϕ puissent être transformées en des constructions mentales qui justifieraient ψ : les logiques d'inspiration intuitionniste s'efforcent *dès le départ* de construire un système de règles relativement auquel les agents *pourraient* être considérés comme logiquement omniscients.

2.2.3. Règles effectives versus règles effectuelles

A l'inverse de la condition de correction qui est, en un certain sens, "absolue", la condition épistémique de faisabilité demeure indéterminée tant que n'a pas été strictement définie la classe d'agents (ou de capacités recognitionnelles) qui sert de

40. Cf. sur ce point J. Dubucs, Hintikka et le problème de l'omniscience logique, à paraître.

41. Truth and Other Enigmas, Duckworth, 1978, p. 226-227.

référence. Le choix de cette classe de référence peut varier entre une borne inférieure définie par l'anthropologisme strict, qui prend en compte les seules capacités manifestées *hic et nunc* par les agents humains, et une borne supérieure représentée par un ultra-libéralisme volontiers enclin à se référer à d'hypothétiques agents dont les ressources cognitives seraient une extension arbitrairement prolongée de celles dont nous faisons nous-mêmes preuve, et qui seraient, par exemple, capables de tirer une conclusion dépendant d'une infinité de prémisses. Bien que certains logiciens aient été tentés par cette option ultra-libérale, qui a évidemment le mérite de résoudre la tension en éliminant en quelque sorte l'un des termes en conflit, c'est-à-dire en faisant coïncider la notion de déductibilité avec celle de conséquence logique⁴², la tradition s'est massivement ralliée à une position médiane, selon laquelle les déductions doivent simplement être *effectives*, c'est-à-dire capables d'être effectuées en un nombre fini d'étapes de nature élémentaire : une déduction de ψ à partir de Γ doit alors être considérée comme une suite finie d'énoncés-types dont le dernier est ψ , et dont chacun est un membre de Γ ou résulte immédiatement d'éléments précédents de la suite par application d'une règle d'inférence, les règles étant elles-mêmes en nombre fini. Cette condition d'effectivité est évidemment *nécessaire*, serait-ce seulement pour assurer le contrôle et la communication des assertions dans une collectivité linguistique : comme le remarque Church⁴³, en l'absence de procédure routinière permettant de décider à coup sûr si une suite d'énoncés est ou non une déduction, un lecteur pourrait toujours se déclarer non convaincu par une preuve et en exiger une autre. En revanche il est plus malaisé de dire pourquoi cette condition d'effectivité serait *suffisante* pour ce qu'on attend d'elle, à savoir la garantie que les agents de référence sont en mesure de reconnaître que les conditions de déductibilité de ψ sont satisfaites lorsqu'elles le sont : un agent donné peut fort bien être *pratiquement* incapable de conduire jusqu'à son terme une procédure pour laquelle il dispose cependant de consignes algorithmiques. C'est une platitude, allègue-t-on souvent, qu'une machine de Turing peut toujours être réalisée de mille et une manières, et notamment d'au moins une, en sorte que les règles effectives sont, *ipso facto*, des règles physiquement exécutables. Mais quelle que soit la faveur dont jouit aujourd'hui cette dernière affirmation, elle n'est ni plate ni même strictement vraie.

2.2.3.1. L'irréalisme des procédures exponentielles

Le fait que toute règle effective puisse sûrement être exécutée *au moins une fois* par un dispositif matériel de type approprié n'assure nullement que le dispositif en question soit physiquement capable d'appliquer la règle *suffisamment de fois* pour

42. Un exemple assez typique est celui de Zermelo, qui attribuait les résultats d'incomplétude de Gödel à une notion de déduction foncièrement pervertie par le "préjugé finitiste" et qui plaiderait pour une logique infinitaire très éloignée des méthodes "imparfaites" et "trompeuses" dont use notre entendement fini, et dans laquelle tout énoncé vrai serait également prouvable (*Über Stufen der Quantifikation und die Logik des Unendlichen, Jahresbericht der Deutschen Mathematiker-Vereinigung*, XLI-1932, p. 85-88).

43. *Introduction to Mathematical Logic*, Princeton U.P., 1956, 6^e édit., 1970, p. 53.

parvenir à la réalisation de l'objectif pour lequel il a été conçu. Par exemple, une procédure effective qui exige 2^n étapes pour traiter un élément de taille n ne peut certainement pas être conduite à son terme si on l'applique à un élément de taille supérieure à 100 : même si l'effectuation d'une étape ne demandait que le temps mis par la lumière pour traverser un proton, l'ensemble de la procédure exigerait plus de temps que l'âge actuel de l'univers. En particulier, le problème de reconnaître si ψ est ou non une conséquence propositionnelle de Γ (c'est-à-dire de reconnaître si l'ensemble $\Gamma \cup \{\neg\psi\}$ est ou non insatisfaisable au sens du calcul des propositions) est typiquement un problème dont la solution est possible "en principe" sans l'être toujours "en pratique", puisque la procédure requise demande 2^n étapes si Γ et ψ comportent n lettres propositionnelles distinctes. Bien que chaque étape de cette procédure soit de nature effective, l'idéalisation selon laquelle un agent cognitif pourrait venir à bout d'un tel travail pour des données Γ et ψ suffisamment vastes n'est donc pas une idéalisation significativement moins radicale que l'idéalisation selon laquelle l'agent considéré pourrait accomplir des tâches non effectives.

2.2.3.2. Un exemple de règle effective *spécialement* problématique du point de vue de l'effectuabilité

Bien que toute règle effective puisse être au moins une fois exécutée par un dispositif matériel, certaines ne peuvent l'être que de manière *essentiellement transitoire*, au sens où il est de leur nature que leur exécution ne puisse être *indéfiniment itérée* par aucun dispositif. Considérons l'exemple d'une photocopieuse, supposée réaliser une machine de Turing qui duplique la suite de symboles écrites sur son ruban. La copieuse est décrite comme capable, *dans les circonstances appropriées* (il faut, par exemple, que le bac à papier soit approvisionné), de passer régulièrement d'un état initial dans lequel elle est en attente d'un document à un état final dans lequel une copie du document figure dans le bac de sortie. Mais il est clair, toutes choses égales par ailleurs, c'est-à-dire *compte non tenu de l'inévitable usure des composants permanents de la machine*, que chaque exécution du processus tend à rapprocher le dispositif de la situation dans laquelle les circonstances *ne* se prêteront *plus* à l'exécution : le bac à papier une fois totalement déchargé, l'opération de copiage devient impossible. Or des limitations du même type, c'est-à-dire des limitations *intrinsèquement liées* à l'exécution des consignes, s'appliquent assurément à certains des mouvements inférentiels permis par les règles logiques usuelles, comme l'inférence de ϕ à $\phi \& \phi$: aucune architecture matérielle n'est capable d'être indéfiniment le support d'une transition causale d'une occurrence de « ϕ » à une occurrence de « $\phi \& \phi$ ». Pour autant que l'application concrète d'une règle d'inférence doive se traduire par une transition causale d'une occurrence des prémisses à une occurrence de la conclusion, l'effectivité des règles ne garantit donc pas leur effectuabilité physique.

2.2.4. La logique de la connaissance entre l'hyperidéalisation et l'ad hocité

On est en somme devant le dilemme suivant :

- Ou bien on libéralise la notion de déduction au point d'admettre qu'un agent puisse ne manifester sa capacité à reconnaître les liens déductifs que de façon purement *inchoative*, par son simple *engagement* dans la construction d'une dérivation qu'il sera peut-être constitutivement incapable de mener jusqu'au bout s'il n'a pas la *chance* de trouver d'emblée les mouvements inférentiels les mieux appropriés à l'objet de sa recherche. Mais s'il est, dans le cas le plus défavorable, tout à fait exclu que l'agent parvienne à achever la construction qui lui permettrait de reconnaître que ψ est conséquence de Γ , il n'y a plus grand sens à maintenir que cette construction constitue précisément pour lui le moyen de s'assurer que ψ est bien une telle conséquence.

- Ou bien, et c'est l'autre branche du dilemme, on abandonne toute forme d'idéalisation, et l'on s'engage dans des considérations de détail sur le fonctionnement cognitif des agents de la classe concernée, dans le but plus ou moins avoué de formuler un ensemble de règles inférentielles qui reflèteraient ce fonctionnement au plus près. Mais c'est cher payer l'immanence des liens déductifs, puisque cette démarche conduit à peu près inmanquablement à polluer la logique en sacrifiant la simplicité et la généralité de ses règles.

La question est en somme de concevoir un système de règles qui soit *suffisamment contraint* pour éviter la transcendance des conditions de déductibilité tout en restant *assez général* pour ne pas se réduire à la codification pure et simple des façons d'opérer d'une classe d'agents qui ne comptent pas nécessairement parmi les plus efficaces. Un compromis probablement acceptable entre les deux inconvénients consiste à s'en tenir aux contraintes *les plus générales* susceptibles de borner les performances cognitives d'un agent *physiquement possible*, en d'autres termes d'admettre comme légitimes certaines déductions que les agents de la classe de référence ne pourraient pratiquement effectuer dans tous les cas que s'ils possédaient des ressources cognitives supérieures aux leurs, mais en subordonnant cette admission à la condition expresse que ce surcroît hypothétique de ressources puisse leur être attribué *sans* altérer radicalement les lois physiques les plus fondamentales (par exemple les principes de la thermodynamique)⁴⁴.

2.2.5. Effectuabilité explicite versus effectuabilité implicite

De ce qu'une action est permise ne découle généralement pas qu'elle soit possible : la légitimité d'une règle ne saurait être sérieusement menacée par son inapplicabilité dans tel ou tel cas. En revanche, une règle *constitutivement* inefficace, c'est-à-dire une règle dont l'application répétée conduirait inmanquablement un individu à ne plus être, s'il l'avait jamais été, en position de

44. C. Wright discute une suggestion de cet ordre dans son article *Strict Finitism*, repris dans *Realism, Meaning and Truth*, Blackwell, 1987, p. 124.

l'appliquer, une telle règle est appelée à tomber en désuétude : la maintenir en vigueur au motif qu'elle est supposée ne s'appliquer au commun des mortels qu'au travers d'une clause *ceteribus paribus* généreusement formulée reviendrait à considérer qu'elle n'est au fond édictée que pour une classe raréfiée d'*Übermenschen* épistémiques. Le logicien traditionnel court exactement le même risque en proposant des règles d'inférence effectives mais ineffectuables : il forge un concept de déduction approprié aux capacités d'agents supra-physiques, auprès desquels les agents ordinaires font inévitablement figure d'agents infra-rationnels⁴⁵. Or ces derniers *sont* de toute évidence rationnels, puisqu'ils le sont autant qu'il est en leur pouvoir de l'être.

Une fois reconnue sur ce point la nécessité d'une *réforme* des règles logiques, un obstacle de taille se présente : toute règle effective est matériellement exécutable dans le cas d'une application singulière, en sorte que ce qui doit être modifié dans la règle ne concerne que l'itération éventuelle de son application. Autant dire que la réforme ne saurait consister à modifier la règle elle-même, mais seulement à lui adjoindre un décret d'application, et que ce décret d'application ne saurait à son tour consister qu'en une unique clause, à savoir la clause déclarant que la règle *ne doit pas* être appliquée dans les conditions où elle serait inapplicable *en fait*. Appelons "explicitement effectuelles" de telles règles, qui définissent expressément une borne aux ressources requises par leur application⁴⁶. Ce qui doit leur être reproché est moins de rapprocher le droit du fait — puisque tel est, au fond, leur objectif — que de se *contenter* de le faire et d'être de pures et simples légalisations de la manière d'opérer des agents : on leur préférerait des dispositions qui obtiennent leur effet sous la forme de la *généralité*, c'est-à-dire sans prendre expressément en compte les caractéristiques matérielles des agents auxquels elles s'appliquent, en un mot des règles "implicitement effectuelles" auxquelles les agents pourraient se conformer sans que cette conformité soit stipulée dans la règle.

45. Le problème de l'"omniscience logique" n'a pas d'autre origine : faute d'avoir défini *en logique* une notion de déduction faisable, on retrouve *en psychologie et en méthodologie* le problème d'expliquer pourquoi les capacités déductives des individus s'avèrent si faibles. Contrairement à ce qu'écrit, par exemple, Ch. Cherniak (*Minimal Rationality*, Bradford, M.I.T., 1986, p. 12), l'idéalisation *fondamentale* n'est donc pas celle qui consiste à affirmer que l'ensemble des croyances des agents est déductivement clos, mais celle qui consiste à proposer un concept de déduction relativement auquel cet ensemble n'a aucune espèce de chance de jamais l'être.

46. Si l'on veut, l'effectualité explicite est à l'effectivité ce qu'est la calculabilité par automates linéaires bornés à la calculabilité au sens de Turing, les fonctions ALB-calculables étant définies comme celles des fonctions Turing-calculables dont le calcul peut être effectué sans occuper *trop* de place sur le ruban (précisément : le calcul de la valeur ne doit pas occuper plus de place que l'argument lui-même). Cf J.E. Hopcroft & J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley, 1979, p. 225.

2.2.7. La forme des règles implicitement effectuelles

De telles règles ne peuvent certainement pas être énoncées dans le format traditionnel de la déduction entendue comme suite finie d'énoncés-types, qui cache jusqu'à l'identité des règles appliquées, puisqu'il permet de distinguer entre l'inférence de $\phi \& \psi$ à ϕ et l'inférence de $\phi \& \psi$ à ψ , mais non entre l'inférence de $\phi \& \phi$ à ϕ qui conserve l'occurrence de gauche et celle qui conserve l'occurrence de droite. Une réforme minimale voudrait donc que l'on garde la trace des règles en vertu desquelles les énoncés ont été insérés dans la suite, et que l'on adopte le format de la déduction annotée, du type

(1)	$\phi \& \phi$	(hypothèse)		
(2)	$\phi \& \psi \rightarrow \phi$	(règle &-gauche ⁴⁷)		
(3)	$\phi \& \phi \rightarrow \phi$	(2)	+	sub ψ/ϕ
(4)	ϕ	(1), (3)	+	<i>modus ponens</i>

Mais le résultat obtenu est un mélange chimérique de langage-objet et de métalangage qui ne peut recevoir aucun statut logique. Comme l'écrit encore Church⁴⁸, qui assume sur ce point la position traditionnelle de manière parfaitement conséquente :

« Notre programme est d'exprimer les preuves (aussi bien que les théorèmes) dans un langage-objet pleinement formalisé, et aussi longtemps qu'une part de la preuve reste écrite dans un métalangage non formalisé, l'analyse logique doit être tenue pour incomplète. S'il n'est pas superflu et s'il ne peut donc être simplement omis, un énoncé du méta-langage, par exemple l'énoncé selon lequel une formule est immédiatement inférée de formules particulières qui la précèdent, cet énoncé doit toujours être remplacé d'une certaine manière par un ou plusieurs énoncés du langage-objet. (...) Le langage-objet doit être un langage indépendant capable d'exprimer les choses pour lesquelles il a été conçu, et cela sans être continuellement aidé ou complété par le métalangage. »

La notion traditionnelle de déduction ne peut donc laisser apparaître que le *résultat* de l'application des règles. Comme leurs *modalités d'application* restent nécessairement cachées, aucune règle (implicitement) effectuelle ne peut être définie dans ce cadre.

47. Pour plus de simplicité dans l'exposé, je considère les axiomes comme des règles sans prémisses.

48. op. cit., p. 53 n. 121.

2.2.8. Les modalités d'application des règles logiques

Les calculs de "déduction naturelle" mis au point par Gentzen livrent, quant à eux, assez d'information sur la structure des preuves pour faire apparaître des choix possibles entre diverses manières d'appliquer une règle. Une dérivation y est conçue non comme une suite linéaire d'énoncés-types, mais comme une structure arborescente dans laquelle certaines hypothèses peuvent être actives ou pas (une hypothèse est dite active si la conclusion en dépend). En vertu de cette structure arborescente, les dérivations peuvent contenir plus d'une occurrence d'une hypothèse donnée, et la question se pose alors de savoir si toutes les occurrences actives doivent être déchargées à la fois. La règle d'introduction de $\rightarrow$ du système NJ pour la déduction naturelle intuitionniste reste muette à ce sujet, et peut donc être appliquée de différentes manières⁴⁹. En l'absence de toute consigne d'application, on obtient la pleine logique intuitionniste, qui est une logique de l'effectivité. Mais on peut ajouter de telles consignes, c'est-à-dire spécifier les mouvements inférentiels licites de manière encore plus rigide, en ne les autorisant jamais dans une situation où il serait *cognitivement* ou *physiquement* aberrant pour un agent de les effectuer systématiquement. L'objectif est de garantir que les ressources, cognitives ou physiques, qui ont permis à un agent de s'assurer des prémisses de l'inférence vont aussi lui *suffire* dans tous les cas pour s'assurer de la conclusion. L'inférence doit transmettre *deux* legs, la vérité et l'exécutabilité. Les règles se chargent de la première tâche, car elles sont précisément rédigées pour cela, mais il n'entre pas dans leurs attributions de veiller à la seconde : si elles sont *effectives*, alors elles ont déjà donné, en ce domaine, tout ce que l'on peut attendre d'elles. Le reste est l'affaire des consignes d'application, comme on va le voir sur l'exemple suivant.

2.2.9. Un exemple d'"effectuabilisation": le fragment implicatif de la logique intuitionniste

Il existe dans la littérature de nombreux systèmes de notation pour relier chaque occurrence d'une hypothèse active à l'inférence qui la décharge⁵⁰. Ici on en utilisera deux à la fois : d'une part celui de Gentzen-Prawitz, où l'on décore par un même entier l'occurrence et l'inférence, d'autre part la notation séquentielle du type $\Gamma \vdash \phi$, où Γ contient toutes les occurrences encore actives au nœud courant. On se limitera à l'examen de deux consignes de décharge : l'obligation et l'unicité.

49. Elle est "impropre" au sens que D. Prawitz donne à ce mot (*Natural Deduction. A Proof-Theoretical Study*, Acta Universitatis Stockholmiensis. Stockholm Studies in Philosophy 3, Almqvist & Wiksell, 1965, p. 22-23).

50. Un tel système n'est rien d'autre qu'une représentation de la "fonction de décharge" définie par Prawitz, op. cit.

2.2.9.1. La décharge obligatoire

En l'absence de toute directive supplémentaire, la décharge d'une occurrence de l'hypothèse est simplement facultative lors de l'introduction de $\rightarrow$. En d'autres termes, l'hypothèse peut être déchargée "à blanc", comme c'est le cas pour ψ dans la déduction suivante :

$$\frac{\phi \quad 1}{\psi \rightarrow \phi} \quad 1$$

Ou encore, dans la notation séquentielle :

$$\frac{\Gamma = x:\phi}{\Gamma \vdash \phi} \quad (*)$$

$$\frac{\Gamma, \psi \vdash \phi}{\Gamma \vdash \psi \rightarrow \phi}$$

$$\frac{\vdots}{\vdash \phi \rightarrow (\psi \rightarrow \phi)}$$

Cette licence est régulièrement critiquée par les théoriciens de l'implication "pertinente" ou "analytique"⁵¹, qui déplorent que $\phi \rightarrow (\psi \rightarrow \phi)$ soit accepté en l'absence de toute véritable *connexion déductive* entre ϕ et ψ , puisque ψ n'est pas supposé figurer dans la déduction de ϕ .

Mais elle peut être également mise en cause dans une perspective d'économie, cognitive ou physique. Comme l'action de quantifier une variable dans une expression où elle ne figure pas libre, l'introduction de ψ comme antécédent de ϕ est un acte gratuit qui n'est rétribué par rien. Corrélativement, en admettant une telle licence, on ouvre l'angle entre ce qui se déduit *logiquement* et ce qui peut

51. On peut citer Anderson et Belnap (*Entailment, I*, Princeton U.P., 1975, p. 12-13) ainsi que, dans une moindre mesure, Tennant (*Anti-Realism and Logic, I*, Clarendon Press, Oxford, 1987, p. 194 sq).

cognitivement être inféré d'un énoncé donné⁵². La consigne de décharge obligatoire, qui s'interdit cette licence, définit pour l'essentiel la logique de la "pertinence".

Par ailleurs, la licence de décharge à blanc a pour contrepartie la règle structurale d'affaiblissement dans l'antécédent du calcul des séquents, qui apparaît en (*) :

$$\frac{\Gamma \vdash \Delta}{\Gamma, \phi \vdash \Delta}$$

Rendre plus strictes les modalités d'application des règles de la déduction naturelle équivaut donc à limiter les règles structurales du calcul des séquents. Ce sont ces règles structurales, et non les règles logiques proprement dites, qui contiennent toutes les idéalizations de la logique.

2.2.9.2. La décharge singulière

Que la décharge soit ou non obligatoire, le choix demeure possible entre la stratégie de décharge totale et diverses stratégies de décharge partielle. L'exemple crucial est donné par la dérivation suivante, où les deux occurrences actives de ϕ sont simultanément déchargées :

$$\frac{\frac{\frac{\phi_1 \quad \phi \rightarrow (\phi \rightarrow \psi) \quad 2}{\phi \rightarrow \psi}}{\phi \quad 3} \quad \psi}{\phi \rightarrow \psi} \quad 1,3$$

$$\frac{\phi \rightarrow \psi}{[\phi \rightarrow (\phi \rightarrow \psi)] \rightarrow (\phi \rightarrow \psi)} \quad 2$$

52. On se reportera sur ce point au chapitre 11 de ce livre, p. 301, où il est suggéré que l'antécédent d'une implication "cognitivement effectuable" devrait rendre son conséquent "présent à l'esprit".

Ou encore, dans la notation séquentielle :

$$\Gamma = x:\phi, y:\phi \rightarrow (\phi \rightarrow \psi), z:\phi$$

$$\Gamma \vdash \phi \quad \Gamma \vdash \phi \rightarrow (\phi \rightarrow \psi)$$

$$\Gamma \vdash \phi \quad \Gamma \vdash \phi \rightarrow \psi$$

$$\Gamma \vdash \psi$$

(**)

$$x:\phi, y:\phi \rightarrow (\phi \rightarrow \psi) \vdash \psi$$

$$y:\phi \rightarrow (\phi \rightarrow \psi) \vdash \phi \rightarrow \psi$$

$$\vdash [\phi \rightarrow (\phi \rightarrow \psi)] \rightarrow (\phi \rightarrow \psi)$$

Cette licence de décharge obligatoire a donc pour contrepartie la règle structurale de contraction, qui apparaît en (**):

$$\Gamma, \phi, \phi \vdash \Delta$$

$$\Gamma, \phi \vdash \Delta$$

L'énoncé $[\phi \rightarrow (\phi \rightarrow \psi)] \rightarrow (\phi \rightarrow \psi)$ — ou, plus lisiblement : $[(\phi \& \phi) \rightarrow \psi] \rightarrow (\phi \rightarrow \psi)$ —, pas plus que la règle de contraction qui lui répond, n'ont attiré la réprobation des logiciens "pertinentistes". La raison en est claire : ces auteurs, tout comme Tennant, ont à cœur d'exprimer l'idée d'une connexion *conceptuelle* et non pas *causale* entre les prémisses et la conclusion d'une inférence. La règle de contraction est à cet égard tout à fait irréprochable : l'énoncé ϕ , considéré comme type, met en jeu exactement les mêmes concepts que ceux qui sont impliqués dans $\phi \& \phi$. Si $\phi \& \phi$ est conceptuellement connecté à ψ , ϕ l'est donc aussi. En revanche il en va très différemment dans la perspective de l'exécutabilité physique des règles: il n'y a aucune raison de supposer que l'effet qui peut être obtenu par deux exemplaires du même type de cause puisse être obtenu avec un seul⁵³.

53. Cf. J.Y. Girard, *Linear Logic, Theoretical Computer Science*, **50** (1987), p. 1-102.

La décharge exacte, qui est à la fois la décharge unique et la décharge obligatoire, et qui consiste à désactiver une et une seule occurrence d'une hypothèse à chaque application d'une règle d'introduction de $\rightarrow$, est donc celle qui correspond à une ontologie logique d'objets concrets. En l'exigeant, on abandonne simultanément l'affaiblissement et la contraction. On connaît depuis longtemps les propriétés remarquables qui découlent du rejet de ces règles (rappelons seulement, par exemple, que la logique du premier ordre devient *décidable* lorsque l'on écarte la règle de contraction⁵⁴). D'autres consignes d'application sont encore concevables, et donc d'autres amputations dans les règles structurales⁵⁵. On peut mentionner la consigne de décharge ordonnée, qui demande que les occurrences actives soient déchargées dans l'ordre où elles apparaissent dans la dérivation, et à laquelle répond l'abandon de la règle d'échange⁵⁶.

Ayant stipulé toutes ces consignes, ou renoncé à toutes ces licences, on obtient une logique véritablement minimale, à partir de laquelle il est peut-être possible de bâtir une logique de la connaissance ayant quelque chance de ressembler à la condition cognitive qui nous est malheureusement la plus familière.

Jacques DUBUCS

54. Selon un résultat de H. Wang, *A Survey of Mathematical Logic*, North-Holland, Amsterdam, 1963, p. 228.

55. En fait les choix stratégiques sont extrêmement variés. Cf. sur ce point D.M. Gabbay & R.J.B.G. de Queiroz, *Extending the Curry-Howard Interpretation to Linear, Relevant and Other Resource Logics*, *Journal of Symbolic Logic*, LVII-4, 1992, p. 1319-1365.

56. L'article précurseur est celui de J. Lambek (*The Mathematics of Sentence Structure*, *American Mathematical Monthly*, LXV-1958, p. 154-170), qui étudie, à des fins linguistiques, un système gentzien pour la logique intuitionniste dans lequel on écarte également la règle structurale d'échange.

Chapitre 8

“Mondes de croyance” et probabilités bivalentes

Une interprétation logique du théorème de Mertens et Zamir

1. Introduction

Depuis les travaux d'Harsanyi (1967-1968), la théorie des jeux en information incomplète s'est fondée sur l'idée de *type*. Cette notion est censée rassembler toute l'information relative au joueur considéré : comme l'écrit Harsanyi, “elle résume certains paramètres cruciaux de la fonction de paiement U_i propre au joueur i ainsi que les principaux paramètres qui décrivent ses croyances” (1967, p. 171). On postule que i connaît son propre type et qu'il est incertain de celui des autres, sur lequel il forme une distribution de probabilité subjective, puisqu'il est un décideur bayésien. La méthode consiste donc à placer avant le déroulement du jeu proprement dit un coup initial d'un joueur supplémentaire : la nature, qui détermine le type de chaque individu. Le résultat de ce coup est partiellement inconnu à i . Suivant la terminologie traditionnelle, on a ramené les jeux à information “incomplète” au cas — plus facile à maîtriser — des jeux à information “imparfaite”.

Cette réduction paraît se heurter à une difficulté logique préalable. En effet, quelle que soit la définition proposée pour le “type du joueur i ”, celle-ci manquera son but, qui est de synthétiser *toute* l'information relative à ce joueur, si elle ne tient pas compte de la manière dont i se représente le type de j , dont il se représente la manière dont j se représente son propre type, et ainsi de suite. Harsanyi parle à ce

sujet d'une "régression infinie des anticipations réciproques". Voici une illustration, qui semblera peut-être éclairante, de la difficulté ainsi indiquée de prime abord. On aimerait faire servir la théorie des jeux à une modélisation élémentaire de la Guerre Froide. Comme celle-ci découlait d'une incertitude fondamentale des deux joueurs, le Soviétique et l'Américain, sur les intentions de l'autre, l'appareillage — bien maîtrisé — des jeux en information complète ne peut malheureusement pas s'appliquer. Mais comment décrire la Guerre Froide à la façon d'un jeu en information incomplète ?

Suivant une première modélisation, naïve, on supposera un ensemble T_1 de types susceptibles d'influencer les fonctions de paiement, soit t_1 , le type "agressif", et t_1^a , le type "pacifique" ; et l'on prêterà à chaque joueur une probabilité subjective, soit, respectivement, p_1^s et p_1^a , sur le type de l'autre. Ce premier essai n'apparaît pas du tout satisfaisant. Car les dispositions agressives ou pacifiques du joueur dépendront non seulement de la possession, par ce joueur, du trait t_1 ou t_1^a , mais encore de l'anticipation qu'il peut avoir de la possession de ce trait par l'autre. Les stratégies de la Guerre Froide envisageaient, à juste titre, que l'U.R.S.S. ne fût pas dominée par un expansionnisme naturel, imposé, voire irréflecti, mais plutôt par la crainte de voir les États-Unis s'en prendre à leurs intérêts vitaux. A la thèse d'une agressivité *intrinsèque*, liée, par exemple, aux considérations de politique intérieure et au souvenir des conquêtes tsaristes, on opposait donc, valablement, celle d'une agressivité *extrinsèque*, déterminée par un système de représentations portant plus particulièrement sur l'adversaire. On tiendra compte de ces considérations en incluant les probabilités subjectives dans la définition du type. Formellement parlant, l'ensemble des types que l'Américain peut attribuer au Soviétique devient T_2^s , dont les éléments sont de la forme (t_1, p_1^s) ou (t_1^a, p_1^s) , la probabilité p_1^s étant a priori quelconque. On construira, symétriquement, l'ensemble T_2^a des types que le Soviétique peut attribuer à l'Américain. Mais le deuxième essai n'est pas moins naïf que le premier. Les dispositions du joueur apparaissent également influencées par son anticipation de l'anticipation que l'autre a pu faire de la possession, par lui-même, de certains traits. Cette considération n'échappait pas non plus aux stratégies de la Guerre Froide : l'U.R.S.S. pouvait n'être dominée ni par une forme d'expansionnisme naturel, ni par la vision naïve d'une Amérique intrinsèquement aggressive, mais bien par la représentation, plus subtile, d'une Amérique inquiète, spéculant sur la menace soviétique et, de ce fait, dangereusement tentée par une frappe nucléaire préventive. Il faudra donc faire intervenir les probabilités de second ordre dans la notion de type. Comme la critique peut se reporter d'étape en étape, quoiqu'elle perde progressivement de sa force intuitive, on construira finalement, pour chaque joueur, une suite infinie d'ensembles de types qu'il peut attribuer à l'autre, soit $(T_1, T_2^s, \dots, T_n^s, \dots)$ et $(T_1, T_2^a, \dots, T_n^a, \dots)$. Aucun des ensembles indicés n'épuise la notion pertinente de type stratégique : telle est l'objection allusivement indiquée par Harsanyi, celle dont l'examen préalable conditionne l'existence même d'une théorie des jeux en information incomplète.

Il faut distinguer l'objection faite ici d'une autre, plus archaïque, envisagée dès les débuts de la théorie des jeux. Même si l'on se place dans le contexte plus simple des jeux à information complète, comme le faisaient von Neumann et Morgenstern, on n'échappe pas à l'examen d'une difficulté intuitive de régression infinie. Mais

celle-ci porte alors sur les *actions* des joueurs ; elle s'enchaîne automatiquement dès lors qu'on rend dépendants les choix stratégiques effectués par le joueur et son anticipation des choix stratégiques effectués par l'adversaire. Il importe de voir que la régression infinie envisagée par Harsanyi ne concerne pas les actions, mais, plus profondément, les facteurs qui les déterminent : la construction, pour l'instant inaboutie, de la notion synthétique de type, vise à assurer la définition correcte des fonctions de paiement (ou d'utilité) ; celles-ci sont prises comme *données* dans la théorie des jeux à information complète. Pour achever de cerner la nature particulière de la difficulté ici discutée, il faut peut-être ajouter qu'elle ne dépend pas essentiellement du choix de l'ensemble tout d'abord considéré, T_1 . En particulier, il n'est pas nécessaire que celui-ci contienne des traits qui influencent réellement la décision stratégique : il peut suffire que les joueurs *croient* que ces traits sont pertinents. Si les stratèges de deux pays considèrent, par exemple, que le niveau des dépenses consacrées à la recherche fondamentale est un indicateur significatif d'"agressivité", ce paramètre figurera dans l'ensemble T_1 , et la régression infinie s'enchaînera tout aussi bien que si l'on avait fait figurer un paramètre choisi pour des raisons d'apparence plus objective, comme le niveau des dépenses militaires. Cette remarque justifiera que l'on examine les propriétés de la régression infinie des types sans qu'on précise concrètement la description naïve des types qui lui donne naissance.

Après avoir signalé une difficulté apparemment très profonde pour la théorie qu'il veut entreprendre, Harsanyi conclut étrangement qu'elle se résout d'elle-même : "de telles séquences d'anticipations réciproques d'ordre croissant ne se présenteront tout simplement pas" (1967, p. 165). Cette indication paraît énigmatique. S'agissant de la théorie des jeux en information complète, on considère généralement que la définition de concepts d'équilibre appropriés permettra de parer à l'objection de régression infinie. Mais on a vu que celle dont il s'agissait ici n'était pas du même ordre ; elle appelle une résolution séparée. Il reste à se convaincre qu'il existe un ensemble convenable de types, c'est-à-dire, compte tenu de l'analyse précédente, un ensemble dont chaque élément *inclurait* la régression infinie des anticipations réciproques. Mais quel sens technique faut-il donner ici au verbe "inclure" ? Si importants qu'ils fussent, les articles d'Harsanyi en 1967-68 laissaient subsister un doute sur la nature (et l'existence même) d'une construction mathématique susceptible de répondre au problème posé.

Le doute s'est dissipé depuis que, dans un travail remarqué et difficile (1985), Mertens et Zamir ont produit une notion mathématique de type qui correspond précisément aux intentions d'Harsanyi. Ce faisant, ils ont établi par le meilleur moyen possible, c'est-à-dire constructivement, la cohérence interne de la théorie des jeux à information incomplète. Harsanyi avait déjà suggéré d'identifier un individu i à une liste d'*attributs objectifs et subjectifs* : la première catégorie se résume en une variable θ^i qui détermine partiellement la fonction de paiement de i ; dans la seconde catégorie devraient figurer non seulement les représentations de i au premier ordre, mais ses représentations à un ordre n quelconque. Systématisant l'indication, Mertens et Zamir introduisent des *hiérarchies infinies de croyances probabilistes* $f = (f_1, \dots, f_n, \dots)$. La probabilité f_1 porte sur l'ensemble Θ des vecteurs $\theta = (\theta^i)_{1 \leq i \leq I}$ d'attributs objectifs (un pour chaque joueur) ; f_2 sur

l'ensemble des vecteurs $(f_1^i)_{1 \leq i \leq I}$ de croyances, ou d'attributs subjectifs, du premier ordre (une croyance pour chaque joueur) ; et ainsi de suite. Dans cette formalisation, chaque joueur se représente non seulement les attributs des autres joueurs, mais les siens propres. Une hiérarchie f obéit à des conditions de cohérence, d'allure naturelle, qui relient les niveaux consécutifs f_n et f_{n+1} . Soit T l'ensemble des hiérarchies bien formées : *c'est l'ensemble des types que l'on cherchait à construire*. Cette affirmation repose sur le théorème fondamental que démontrent Mertens et Zamir : l'ensemble des probabilités sur $\Theta \times T^{I-1}$ est homéomorphe à T lui-même. L'identification obtenue est de nature topologique, mais il en découle un isomorphisme de structure mesurable qui est sans doute plus important, conceptuellement parlant, que l'homéomorphisme ; car il s'agit finalement d'identifier les deux espaces au point de vue de la formation des croyances probabilistes. La bijection bimesurable de Mertens et Zamir répond exactement à la question initiale. On a dit que la régression infinie des anticipations paraissait menacer la notion de type ; de fait, elle détruit la conception naïve qui assimilerait le vecteur des types et celui des attributs objectifs θ^i , et d'ailleurs toute autre élaboration finitaire de la notion. Mais la construction infinitaire du type chez Mertens et Zamir stabilise, en quelque sorte, la régression. L'isomorphisme de structure mesurable indique que l'on n'apprend rien à remonter au-delà de la première hiérarchie : *on peut se passer d'une hiérarchie construite sur l'espace des hiérarchies*.

Le résultat initial de Mertens et Zamir dépend d'hypothèses mathématiques qui ne sont pas insignifiantes : Θ est un espace topologique compact, les ensembles de probabilité successivement construits sont dotés de la topologie faible*, à chaque étape la structure mesurable est celle des boréliens de la topologie considérée. Du point de vue de la théorie des jeux, ces hypothèses ne sont peut-être "pas trop contraignantes", ainsi que le notent les auteurs (1985, p. 3) : on peut représenter l'ensemble des jeux possibles comme un compact.¹ Sans modifier les hypothèses de départ, Mertens, Sorin et Zamir (1990) proposent une variante de la construction de 1985, qui en facilite considérablement l'accès et simplifie la démonstration de l'isomorphisme. Ils indiquent aussi qu'il existe une variante *mesurable* de la construction et de l'isomorphisme.² Ce résultat est bienvenu, car l'accent mis sur les hypothèses topologiques, sans doute naturel du point de vue de la théorie des jeux, a certainement fait obstacle à l'exploitation, par les logiciens et les informaticiens, d'une construction et d'un résultat qui devaient les intéresser vivement. De fait, ceux-ci n'ont su, jusqu'à présent, ni tirer parti du travail de Mertens et Zamir, ni

1. Il suffit de fixer une fois pour toute l'espace des actions et de prendre la fonction d'utilité bornée (ce qui est classique en théorie de l'utilité espérée).

2. Elle consiste à prendre $(\Theta, \mathcal{A}(\Theta))$ borélien standard, c'est-à-dire isomorphe, au point de vue mesurable, à une partie mesurable de $([0,1], \mathcal{B}[0,1])$. Cette hypothèse est légèrement plus restrictive que celle des sections 2 et 3, qui prend $(\Theta, \mathcal{A}(\Theta))$ séparable, c'est-à-dire isomorphe à une partie quelconque de $([0,1], \mathcal{B}[0,1])$. L'auteur doit cette précision technique, ainsi que d'autres figurant dans le texte, à J.-F. Mertens. Parmi les variantes topologiques du théorème, on peut signaler celle de Brandenburger et Dekel (1985), qui prennent Θ métrique séparable au lieu de topologique compact.

même le retraduire dans leur propre formalisme. Les organisateurs des colloques *Theoretical Aspects of Reasoning About Knowledge* (1986, 1988, 1990) ont perçu que la construction hiérarchique de Mertens et Zamir était proche de celles des *structures de connaissances* et des *mondes de croyance* que les informaticiens avaient conçus dans le contexte fort différent des systèmes distribués et de l'intelligence artificielle (voir Fagin, Halpern et Vardi, 1984, pour la première construction et Vardi, 1986, pour la seconde). Mais, à ce jour, l'analogie est restée relativement inactive.³ Il aurait été souhaitable de chercher à retranscrire l'isomorphisme de Mertens et Zamir comme un isomorphisme approprié de structure logique : cette tâche n'a pas été entreprise. La mener à bien permettrait d'affiner la compréhension que l'on a des structures de connaissance et des mondes de croyance, peut-être même d'élargir la palette des sémantiques actuellement disponibles en logique épistémique.⁴

En retour, si l'on ose dire, la théorie des jeux peut bénéficier des investigations du logicien. Les conditions de cohérence qui règlent les hiérarchies probabilistes ont généralement une interprétation épistémique déterminée, mais celle-ci apparaît plus clairement lorsqu'on introduit des opérateurs de croyance discrets. En outre, et plus gravement, les interprétations épistémiques restent flottantes et discutables aussi longtemps que l'on a pas mis en œuvre la traditionnelle *méthode axiomatique*. Il s'agirait, en l'occurrence, de se donner une langue formelle assez pauvre pour rester maîtrisable, mais assez riche, tout de même, pour avoir prise sur des hiérarchies intéressantes la théorie des jeux, de formuler un système d'axiomes dans cette langue, enfin d'en vérifier l'adéquation et la complétude relativement à la classe des hiérarchies considérées. Les tentatives faites récemment en marge de la théorie des jeux pour formaliser des opérateurs de croyance ne sont presque jamais des axiomatisations en ce sens technique.⁵ Elles peuvent être éclairantes ; mais ici comme ailleurs, les procédés du logicien se recommandent par la considération ordinaire : l'usage d'une langue formelle, contrainte et relativement facile à dominer, comme le sont les calculs de premier ordre, présente des garanties par rapport à la langue informelle des mathématiques ordinaires. En outre, ces garanties s'énoncent dans des théorèmes d'adéquation et de complétude auxquels on peut souvent trouver des conséquences en matière de décidabilité des propositions, conséquences qui peuvent retenir l'attention d'une théorie qui se préoccupe davantage que par le passé de la valeur cognitive de ses hypothèses.⁶

3. Elle a fait néanmoins l'objet de commentaires d'ordre général, par exemple Halpern (1986) et Fagin et alii (1992).

4. Les structures de connaissance (*knowledge structure*) de Fagin, Halpern et Vardi (1984), qui se retraduisent en structures de Kripke, apparaissent bien cernées maintenant. Mais les mondes de croyances (*belief worlds*) introduits par Vardi (1986), et reliés aux structures de voisinages, sont moins connus. Ils appellent encore des éclaircissements techniques ; voir la note 13 ci-dessous.

5. Geanakoplos (1990) fait une revue de cette littérature.

6. Lorsqu'on peut démontrer qu'un système a la propriété de modèle fini, c'est-à-dire que chacune des formules autres que les théorèmes est réfutable sur un modèle fini au moins, il en résulte que le système est décidable (l'ensemble de ses théorèmes est récursif). Cf. Chellas,

Ce travail tentera de jeter une passerelle entre deux domaines qui, au-delà de la contribution de Mertens et Zamir, gagneraient à interagir plus systématiquement. Il entreprend d'axiomatiser la construction, ou plutôt une variante particulière de celle-ci, à l'aide d'un langage simple, la logique modale propositionnelle, qui bénéficie aujourd'hui, grâce à ses excellentes propriétés, de la faveur de nombreux logiciens et informaticiens. Les structures que l'on mettra en relation seront, d'une part, des structures ad hoc, dérivées du travail de Mertens et Zamir, et qui s'avèrent très proches de mondes de croyance introduits par Vardi (1986) ; d'autre part, les classiques structures de voisinage définies par Montague et Scott. La principale difficulté que l'on ait rencontrée tenait au choix des premières. Il convenait de garder l'esprit de la construction initiale tout en la transformant en une *sémantique* de calculs propositionnels modaux et, bien sûr, tout en préservant l'isomorphisme caractéristique. De toute évidence, il fallait essayer d'éviter les hypothèses topologiques, et donc d'extraire de l'article de 1985 les considérations qui ressortent à la théorie de la mesure. Mais il fallait aussi appauvrir la structure probabiliste elle-même, puisque la logique modale traditionnelle a un pouvoir expressif bien moindre que celui des probabilités. On aurait pu imaginer d'enrichir simultanément le langage modal.⁷ La solution qu'on a retenue ici est plus simple, puisqu'elle repose sur des axiomes modaux traditionnels ; elle semble avoir une sorte de priorité heuristique et peut préparer des réinterprétations plus subtiles. Elle découle de l'observation suivante : les probabilités bivalentes (valant 0 ou 1) sont un cas particulier, simultanément, des structures de voisinage et du calcul des probabilités.

La première partie de ce travail exploite cette observation et donne les rappels de logique modale dont le lecteur pourra avoir besoin. La seconde partie reproduit la construction de Mertens et Zamir dans le contexte appauvri des probabilités bivalentes. La troisième partie interprète la hiérarchie ainsi construite comme une structure logique, fournissant une sémantique à un système modal particulier, qu'on appelle *SP*, et qui se trouve proche du *S5* de la logique modale traditionnelle. On établit l'adéquation et la complétude de *SP* par rapport à la classe des hiérarchies de probabilités bivalentes (*HPB*), et l'on signale la parenté de ce résultat avec ceux de Vardi (1986). La conclusion discute brièvement les perfectionnements possibles de l'approche. On a découpé les démonstrations en une suite de remarques et de lemmes dont la vérification ne présente pas de difficultés techniques.

2. Probabilités à deux valeurs et structures de voisinage

Soit $(\Theta, \mathcal{A}(\Theta))$ un espace mesurable quelconque. On considère E_1 l'ensemble des $e_1 \in 2^{\mathcal{A}(\Theta)}$ qui vérifient :

1980, p. 62. Sans être universelle, la propriété de modèle fini s'obtient facilement en logique modale propositionnelle.

7. L'enrichissement devait porter soit sur l'objet syntaxique de la croyance, qui pourrait alors revêtir la forme d'un énoncé probabiliste, soit sur l'opérateur B_a , lequel deviendrait multivalent, soit enfin sur les deux aspects à la fois. Par exemple, Gärdenfors (1975) va dans la première direction, Lismont (1992), dans la seconde, et Fagin et Halpern (1986) discutent une approche synthétique.

1. la fermeture par surensembles de $A(\Theta)$
2. la non-vacuité (ou, compte tenu de 1. : $\Theta \in e_1$)
3. la fermeture par intersections dénombrables
4. $\emptyset \notin e_1$
5. $\forall X \in A(\Theta), X \in e_1$ ou $X^c \in e_1$.

On notera $E_1 = F(\Theta)$, ou éventuellement $E_1 = F_{A(\Theta)}(\Theta)$ s'il est nécessaire de préciser la tribu.

L'élément e_1 est un *ultrafiltre* de parties prises dans $A(\Theta)$; mais la condition 3. est plus forte que la condition habituelle de clôture par intersection finie. Les conditions 1. à 5. sont redondantes. La littérature sur les ultrafiltres peut justement suggérer des simplifications. Mais on verra que chacune des conditions précédentes peut recevoir une interprétation logique.

L'ensemble E_1 peut également recevoir une interprétation probabiliste :

Remarque 1. E_1 s'identifie à l'ensemble des probabilités à deux valeurs (valant 0 ou 1) sur $(\Theta, A(\Theta))$. S'il faut distinguer e_1 de la probabilité qui lui est bijectivement associée, on notera celle-ci P_{e_1} : $P_{e_1}(B) = 1$ ssi $B \in e_1$, $\forall B \in A(\Theta)$.

Les probabilités à deux valeurs sont un cas singulièrement dégénéré de la théorie des probabilités. Toutefois, le choix de la structure mesurable $(\Theta, A(\Theta))$ détermine la simplicité plus ou moins grande d'un tel objet. Rappelons (p. ex. Dudley, (1989), ch. 3) qu'un atome d'une probabilité π (on dit aussi : d'un espace probabilisé $(\Theta, A(\Theta), \pi)$) est un $B \in A(\Theta)$ et non vide tel que $\forall C \in A(\Theta), C \subseteq B \Rightarrow \pi(C) = 0$ ou $\pi(C) = \pi(B)$. La probabilité π est atomique si elle a un atome. Une probabilité π à deux valeurs est trivialement une probabilité atomique. Un atome d'une tribu $A(\Theta)$ (ou d'un espace mesurable $(\Theta, A(\Theta))$) est un mesurable minimal non vide, c'est-à-dire un $A \in A(\Theta) \setminus \{\emptyset\}$ tel que $\forall C \in A(\Theta), C \subseteq A \Rightarrow C = \emptyset$ ou $C = A$ (de façon équivalente : $\forall C \in A(\Theta)$, ou bien $A \subseteq C$ ou bien $A \cap C = \emptyset$). Soit $At(\Theta)$ l'ensemble des atomes de la tribu $A(\Theta)$. On appellera *atomique* une tribu telle que $At(\Theta) \neq \emptyset$, et *absolument atomique* une tribu vérifiant : $\forall C \in A(\Theta) \setminus \{\emptyset\}, \exists B \in At(\Theta)$ tel que $B \subseteq C$. On aimerait parvenir à l'identification commode entre atomes de la tribu et probabilités bivalentes : $At(\Theta) \cong F(\Theta)$ (avec éventuellement $\emptyset$ de part et d'autre de $\cong$). Même si on laisse de côté le cas où $At(\Theta) = \emptyset$, des difficultés peuvent survenir. Il est immédiat que, $A \in At(\Theta)$ étant donné, on définit une probabilité bivalente en posant : $\forall B \in A(\Theta), \pi_A(B) = 1$ ssi $A \subseteq B$. L'application $At(\Theta) \rightarrow F(\Theta) : A \mapsto \pi_A$ est bien sûr injective, Mais elle n'est pas nécessairement surjective, comme le montre l'exemple suivant.

Exemple 1. Soit Θ infini non dénombrable et $A(\Theta) =$ la tribu des parties dénombrables (i.e. finies ou infinies dénombrables) ou codénombrables (i.e. de complémentaire fini ou infini dénombrable). Cette tribu est absolument atomique puisque les points sont mesurables. $F(\Theta)$ contient toutes les masses de Dirac δ_θ . Mais il contient aussi une probabilité bivalente qui n'est pas du type δ_θ : posons $\pi(B) = 0$ si B dénombrable, $\pi(B) = 1$ si B codénombrable ; on vérifie que π est une

probabilité. On peut remarquer que π admet une infinité non dénombrable de mesurables B de mesure 1 et que $\bigcap_{B \in \pi^{-1}(1)} B = \emptyset$.

L'exemple 1 montre qu'il ne suffit pas de prendre $A(\Theta)$ atomique, ou même absolument atomique, pour parvenir à l'identification $At(\Theta) \equiv F(\Theta)$. La difficulté vient de ce qu'il existe des π bivalentes telles que $\pi(A) = 0, \forall A \in At(\Theta)$. Cette pathologie est bien sûr absente du cas où $(\Theta, A(\Theta)) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dans ce cas, les seules $\pi \in F(\Theta)$ sont les masses de Dirac.⁸ On cherche ici à éviter les hypothèses topologiques, dont l'interprétation en logique épistémique est peu claire. La bonne hypothèse consiste alors à prendre $A(\Theta)$ *séparable*, c'est-à-dire engendrée par une famille dénombrable de parties. On vérifie :

Remarque 2. Si $(\Theta, A(\Theta))$ est séparable, on a $At(\Theta) \equiv F(\Theta)$.

Voici un exemple de tribu séparable qui est particulièrement utile pour les applications logiques. Soit VP l'ensemble, a priori quelconque, des variables propositionnelles et $L(VP)$ l'ensemble des formules bien formées, c'est-à-dire le plus petit ensemble qui contienne VP et soit clos pour la formation de formules à partir des connecteurs ordinaires $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$. Une valuation est une application $v : VP \rightarrow \{0,1\}$. On sait que v a une extension unique à $L(VP)$. On s'intéresse à la tribu A engendrée par les V^p , où $p \in VP$ et $V^p = \{v \in \{0,1\}^{VP} \mid v(p) = 1\}$. Les théorèmes de complétude et d'adéquation du calcul propositionnel affirment en substance que A est aussi la tribu engendrée par les V^φ , où $\varphi \in L(VP)$ et $V^\varphi = \{v \in \{0,1\}^{L(VP)} \mid v(\varphi) = 1\}$.⁹ Les singletons de A coïncident avec les $\bigcap_{p \in VP} V^{\pm p}$ le symbole

$\pm p$ désignant p ou $\neg p$ suivant le cas. Si l'on prend VP dénombrable, une hypothèse qu'il n'est pas rare de faire en logique, A est séparable mais n'est pas engendrée par une *partition* dénombrable ; les singletons sont les atomes, de sorte que, par la remarque 2, les probabilités bivalentes s'identifient aux masses de Dirac. Si l'on prend VP fini, A est engendrée par une partition finie.

Soit E_1 l'ensemble des probabilités à deux valeurs sur $(\Theta, A(\Theta))$; on s'intéressera maintenant aux probabilités à deux valeurs sur E_1 . La tribu $A(E_1)$ que l'on considérera exclusivement est celle qui est engendrée par la famille $A^-(E_1)$ des

$$\{e_1 \in E_1 \mid B \in e_1\}, \forall B \in A(\Theta).$$

Ce choix de structure mesurable est naturel étant donné le propos, qui est de construire des hiérarchies infinies de croyance. On peut aussi noter une analogie utile : si l'on avait pris Θ topologique, $A(\Theta)$ = la tribu borélienne de Θ , et que l'on ait défini E_1 comme $\Delta(\Theta)$, l'ensemble des probabilités sur Θ , alors $A^-(E_1)$ s'identifierait (au sens de la remarque 1) à une famille d'ouverts élémentaires de la topologie faible* sur E_1 .

8. Plus généralement, si Θ est métrique séparable et $A(\Theta)$ la tribu borélienne, les probabilités bivalentes sont exactement les masses de Dirac.

9. Cet énoncé découle de la version algébrique du théorème d'adéquation et de complétude (p. ex. Halmos, 1963).

Le choix que l'on a fait ici de considérer des probabilités bivalentes seulement a une conséquence importante :

Remarque 3. $A^-(E_1) = A(E_1)$. En outre, si $A(\Theta)$ est séparable, $A^-(E_1)$ est en bijection avec $A(\Theta)$.

Le reste de cette section justifie le choix d'un formalisme qui, aux yeux du théoricien des probabilités, est évidemment dégénéré. Ce faisant, on introduira des notions logiques qui serviront tout au long de ce travail. Soit VP l'ensemble des variables propositionnelles (sur lequel on ne fait pas d'hypothèses de cardinalité à ce stade). Les opérateurs sont les connecteurs propositionnels ordinaires et, pour chaque $a \in A$, un opérateur unaire non propositionnel B_a destiné à signifier "l'agent a croit que". L'ensemble A des *agents* est, comme à l'ordinaire en logique épistémique ou en théorie des jeux, supposé fini. L'ensemble des *formules bien formées*, $L(VP)$, est le plus petit ensemble contenant VP et clos pour la formation des formules à l'aide des connecteurs $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$ et des opérateurs B_a . Tous les systèmes d'axiomes envisagés contiendront, par hypothèse, les axiomes et règles ordinaires du calcul propositionnel, que l'on n'explicitera pas. Le premier système, appelé d'habitude *système minimal* (SM), comporte pour seule clause spécifique : si $\varphi, \psi \in L(VP)$,

$$(R_a) \quad \frac{\varphi \leftrightarrow \psi}{B_a \varphi \leftrightarrow B_a \psi}, \forall a \in A.$$

Chaque a -règle (R_a) formalise une notion affaiblie d'omniscience logique (les croyances de a respectent l'équivalence logique). Soit $\vdash_{SM}$ la relation d'inférence formelle relative au système minimal. Comme il est bien connu en logique modale, (SM) constitue une axiomatisation adéquate et complète de l'ensemble des *structures de voisinage*, dont on restituera brièvement la définition (pour plus de détails, voir Chellas, 1980, ou Hughes et Cresswell, 1984).

Une structure de *voisinage* (ou de *Scott*) est un uplet :

$$m = \langle \Omega, (N_a)_{a \in A}, \nu \rangle$$

ainsi composé :

- Ω est un ensemble quelconque non vide, appelé *univers* de la structure,
- pour chaque $a \in A$, N_a est une application $\Omega \rightarrow 2^{2^\Omega}$ (elle assigne à chaque point w un système $N_a(w)$ de parties de Ω appelées *voisinages*),
- ν est une application : $\Omega \times VP \rightarrow \{0, 1\}$ appelée *valuation*.

Soit S l'ensemble des structures de Scott. Etant donné $m \in S$ d'univers Ω , et $w \in \Omega$, on définit la relation de validation sémantique $m \models \varphi$ par une récurrence sur la complexité de $\varphi \in L(VP)$. La clause principale¹⁰ de cette définition est :

10. Les autres clauses de la récurrence sont purement propositionnelles :

· si $\varphi \in VP$, $m \stackrel{w}{\models} \varphi$ ssi $\nu(w, \varphi) = 1$

• si $\varphi = B_a \psi$, $m \models^w \varphi$ ssi $[|\psi|]_m \stackrel{\text{def}}{=} \{w' \mid m \models^{w'} \psi\} \in N_a(w)$.

$m \models \varphi$ signifie $m \models^w \varphi$, $\forall w \in \Omega$, et $S \models \varphi$ signifie $m \models \varphi$, $\forall m \in S$. On peut maintenant énoncer le théorème d'adéquation : $\vdash \varphi \Rightarrow S \models \varphi$, et celui de complétude : $S \models \varphi \Rightarrow \vdash \varphi$.

A bien des égards, le formalisme des structures de voisinage est plus naturel, du point de vue épistémique, que celui des structures de Kripke, plus courantes cependant en logique modale. Dans la clause rappelée ci-dessus, $[|\psi|]_m$ est l'ensemble de vérité de ψ dans m , c'est-à-dire la proposition ψ considérée de façon extensionnelle. La famille de parties $N_a(w)$ contient toutes les propositions — en ce sens — auxquelles l'agent adhère. Dans ces conditions, on peut s'attendre que les analogies entre syntaxe et sémantique soient particulièrement lisibles. De fait, différents théorèmes d'adéquation et de complétude rattachent les axiomes syntaxiquement naturels qui viennent renforcer SM à des contraintes ensemblistes, elles aussi naturelles, sur les $N_a(w)$. On s'intéressera ici seulement aux axiomes dont la traduction sur les $N_a(w)$ est interprétable en calcul des probabilités. C'est pour cette raison que l'on a commencé par rappeler les propriétés des probabilités à deux valeurs. Celles-ci paraissent constituer l'articulation la plus élémentaire entre le formalisme des structures de Scott (et plus généralement celui de la logique modale) et une formalisation mesurable des croyances. Il est curieux que cette voie de passage entre les deux approches ait été si rarement empruntée.

On appellera SP le système qui comporte les règles et axiomes suivants : $\forall a \in A$,

(RM_a)	$\frac{\varphi \rightarrow \psi}{B_a \varphi \rightarrow B_a \psi}$
(N_a)	$B_a \top$
(C_a)	$B_a \varphi \wedge B_a \psi \rightarrow B_a(\varphi \wedge \psi)$
(P_a)	$\neg B_a \perp$
(IPD)	$B_a B_a \varphi \rightarrow B_a \varphi$
(IPA)	$B_a \varphi \rightarrow B_a B_a \varphi$
(CP_a)	$\neg B_a \varphi \rightarrow B_a \neg \varphi$

• si $\varphi = \neg \psi$,	$m \models^w \varphi$ ssi	(non $m \models^{w'} \psi$) (noté : $m \not\models^{w'} \psi$)
• si $\varphi = \psi_1 \wedge \psi_2$,	$m \models^w \varphi$ ssi	$m \models^w \psi_1$ et $m \models^w \psi_2$
• si $\varphi = \psi_1 \vee \psi_2$,	$m \models^w \varphi$ ssi	$m \models^w \psi_1$ ou $m \models^w \psi_2$
• si $\varphi = \psi_1 \rightarrow \psi_2$,	$m \models^w \varphi$ ssi	$m \models^w \psi_2$ lorsque $m \models^w \psi_1$

La première règle — de monotonie — renforce l'unique règle (R_a) du système minimal SM . Les trois premières clauses forment le système K , l'un des plus banals de la logique modale. Il est contenu dans tous les systèmes dit *normaux*, c'est-à-dire ceux qui s'interprètent à l'aide de structures de Kripke. L'adjonction de (P_a) donne un système appelé parfois (KD). Les axiomes (IPD) et (IPA) — introspection positive descendante et ascendante — sont moins courants. Ils formalisent l'hypothèse généralement retenue dans les hiérarchies de croyances probabilistes suivant laquelle l'agent croit qu'il croit en un énoncé si et seulement si il croit en cet énoncé. L'adjonction de (CP_a) — axiome de complémentarité — s'impose dès lors qu'on veut retrouver les propriétés des probabilités à deux valeurs dans la sémantique. Cet axiome figure rarement dans les systèmes de logique modale, sans doute parce qu'il est extrêmement contraignant, comme on va le voir facilement. Il rend d'ailleurs redondants certains axiomes qui le précèdent dans la liste.

Remarque 4.

- (i) Le système composé des axiomes (RM_a), (N_a) et (C_a) — c'est-à-dire le système K — comprend le théorème :

$$(M_a) \quad B_a(\varphi \wedge \psi) \rightarrow B_a\varphi \wedge B_a\psi$$

$$\text{et les règles :} \quad \frac{B_a\varphi}{\varphi} \text{ et } \frac{\neg B_a\neg\varphi}{\varphi}$$

Il admet comme équivalent à (P_a) :

$$(P'_a) \quad B_a\varphi \rightarrow \neg B_a\neg\varphi$$

(P_a) et (CP_a) rendent redondant l'axiome (N_a).

- (ii) Le système composé des axiomes (RM_a) à (IPA) — qu'on appellera système BB — a pour théorème :

$$(IND) \quad B_a\neg B_a\varphi \rightarrow \neg B_a\varphi$$

Dans ce système, (CP_a) est un théorème ssi :

$$(CP'_a) \quad B_a(\varphi \vee \psi) \rightarrow B_a\varphi \vee B_a\psi \quad \text{l'est.}$$

- (iii) Le système total — SP — comprend :

$$(INA) \quad \neg B_a\varphi \rightarrow B_a\neg B_a\varphi.$$

Cette remarque se démontre par des moyens purement syntaxiques. On voit notamment que SP implique les deux axiomes d'introspection *négative* descendante et ascendante. Les utilisateurs de la logique modale épistémique regardent d'ordinaire ces axiomes comme plus litigieux, d'un point de vue empirique et même normatif, que leurs homologues *positifs* (IPD) et (IPA).¹¹ A ce point, il faut

11. Ces axiomes sont discutés tout au long des recueils *Theoretical Aspects of Reasoning About Knowledge*.

souligner que l'on s'est beaucoup rapproché du système $S5$, l'un des plus forts parmi ceux qu'envisage couramment la logique modale : celui-ci contient K — c'est-à-dire (RM_a) , (N_a) et (C_a) — ainsi qu' (IPA) , (INA) et l'axiome :

$$(T_a). \quad B_a \varphi \rightarrow \varphi$$

Dans le système SP , (T_a) est *affaibli* en (P_a) (ou B_a^*) et (IPD) . C'est toute la différence principale avec le très classique $S5$, et elle s'explique en partie par la différence entre deux objectifs de formalisation : la croyance, dans un cas ; la connaissance (qui, suivant la définition ordinaire, porte sur le vrai), dans l'autre cas.

Il serait intéressant d'introduire la notion de structure de voisinage, ou de Scott, *mesurable* : tout $m = \langle \Omega, (A_a(\Omega))_{a \in A}, (N_a)_{a \in A}, \nu \rangle$ tel que, pour tout $a \in A$, $(\Omega, A_a(\Omega))$ soit un espace mesurable, et N_a soit $\Omega \rightarrow 2^{A_a(\Omega)}$. Cette notion généraliserait celle de structure de voisinage, dans laquelle on prend $A_a(\Omega) = 2^\Omega$, $\forall a$, en introduisant des *possibilités de discrimination* variables d'agent à agent. Soit S_M la classe de telles structures. Une structure de voisinage *probabilisée* — au sens des propriétés bivalentes — serait alors toute $m \in S_M$ telle que : $\forall a \in A, \forall w \in \Omega$, $N_a(w)$ vérifie les règles de clôture 1. à 5. Soit S_P l'ensemble des structures de voisinages probabilisées. Il n'est pas certain qu'on puisse caractériser axiomatiquement S_P . Mais on peut du moins obtenir la correspondance suivante entre syntaxe et sémantique :

Proposition 1. *Le système (SP) est une axiomatisation adéquate et complète de l'ensemble S_P^* des $m \in S_P$ de la forme :*

$$m = \langle \Omega, 2^\Omega, (N_a)_{a \in A}, \nu \rangle$$

qui vérifient en outre :

$$\forall w \in \Omega, \forall a \in A, \forall B \in 2^\Omega, B \in N_a(w) \text{ ssi } \{ w' \in \Omega \mid B \in N_a(w') \} \in N_a(w).$$

Cette proposition, qui n'est pas exactement surprenante, indique en substance que (SP) axiomatise la notion de probabilité bivalente surdéterminée par la contrainte : B est de probabilité 1 si et seulement si l'événement " B est de probabilité 1" est lui-même de probabilité 1. Cette dernière contrainte se lit dans la dernière ligne de la proposition et correspond syntaxiquement à (IPA) et (IPD) . Il n'est pas difficile de construire une démonstration en adaptant des faits connus sur les structures de voisinage.

Après avoir rapproché les deux approches, logique et probabiliste, dans le contexte simple des structures de voisinage, il reste maintenant à comparer leurs analyses respectives de la régression infinie des croyances.

3. Les hiérarchies de probabilité bivalentes - un analogue simplifié du théorème des Mertens et Zamir

La construction présentée dans cette section est un analogue grandement simplifié de la première construction effectuée par Mertens et Zamir (1985). Comme celle-ci, et contrairement à la variante plus élégante et ramassée que proposent

Mertens, Sorin et Zamir (1990)¹², elle est *récapitulative à tous les stades* : chaque niveau fini k contient les $k - 1$ niveaux précédents. Il a semblé préférable de procéder ainsi pour gagner en explication et faciliter les interprétations logiques. On se place dans le contexte où les espaces de croyances successivement définis sont du type $F(E)$ — ensemble des probabilités bivalentes — et non pas $\Delta(E)$ — ensemble des probabilités. Il faut alors procéder à une vérification directe de l'isomorphisme. Sous l'hypothèse de tribu séparable, les identifications signalées dans la première section rendent cette tâche élémentaire. La question demeure de savoir si la simplification massive que représente le passage de $\Delta(E)$ à $F(E)$ ne suffit pas à garantir l'isomorphisme dans des cas plus généraux encore que celui des tribus séparables. La logique épistémique peut suggérer des généralisations dans une autre direction : par affaiblissement des règles de clôture 1. à 5. Afin de préserver ces différentes possibilités, on emploiera deux notations successives : la première est d'apparence générale, la seconde provient des identifications que permet l'hypothèse de tribu séparable.

$(\Theta, A(\Theta))$ est l'espace mesurable séparable qui décrit le paramètre de base (ou les attributs objectifs, dans la terminologie de Harsanyi). Soit W_1 l'ensemble des suites unaires (θ) pour $\theta \in \Theta$; il s'identifie évidemment à Θ , mais il est utile d'avoir deux écritures. On pose $E_1 = F(\Theta)$, $A(E_1)$ comme dans la section 1 (donc, $A(E_1) = A^-(E_1)$ suivant la remarque 3) et $F_1 = E_1^A$. Pour alléger la notation, on se placera désormais dans le cas de deux joueurs, $A = \{a, b\}$. *Chaque fois qu'il apparaîtra un produit d'espaces mesurables, il sera automatiquement muni de la tribu-produit.* Donc, si $W_2 = W_1 \times F_1$,

$$A(W_2) = A(W_1) \otimes A(F_1) = A(W_1) \otimes A(E_1) \otimes A(E_1).$$

Soit $E_2 = F(W_2)$, $A(E_2) = A^-(E_2)$ et $F_2 = E_2 \times E_2$. On va s'intéresser aux suites $(\theta, f_1, f_2) \in W_2 \times F_2$ qui vérifient une *condition de cohérence à définir maintenant*.

L'application $cp(\cdot)$ — pour “coupure” — part d'un produit fini à $i \geq 2$ composantes et enlève la dernière composante :

$$(g_0, \dots, g_i) \xrightarrow{cp} (g_0, \dots, g_{i-1})$$

Soit φ_1 , définie sur E_2 , telle que :

$$\varphi_1(e_2) = \{cp(X) \mid X \in e_2\}$$

et $\bar{\varphi}_1$, la fonction sur F_2 consistant à prendre φ_1 composante par composante.

12. La variante de 1990 est isomorphe à la construction initiale, mais n'explicite pas la croyance sur la croyance propre, et rassemble toute l'information de niveau k dans une seule fonction f_k plutôt que dans une suite partielle $(f_1, \dots, f_k)$.

Remarque 5. φ_1 va dans E_1 ; elle est surjective et mesurable ; elle consiste à prendre la loi-image de la propriété bivalente e_2 par $cp(\cdot)$; l'ensemble des suites (θ, f_1, f_2) telles que $f_1 = \overline{\varphi_1}(f_2)$ est mesurable dans $W_1 \times F_1 \times F_2$.

On appellera *2-cohérentes* les suites (θ, f_1, f_2) telles que :

$$i) \quad f_1 = \overline{\varphi_1}(f_2)$$

$$ii) \quad \forall a \in A, X \in f_1(a) \text{ ssi } \{(\tilde{\theta}, \tilde{f}_1) \mid X \in \tilde{f}_1(a)\} \in f_2(a).$$

Soit W_3 l'ensemble de telles suites.

Remarque 6. (ii) est équivalente à la condition (ii') :

$$\forall a \in A, \{(\tilde{\theta}, \tilde{f}_1) \mid \tilde{f}_1(a) = f_1(a)\} \in f_2(a).$$

En outre, $W_3 \neq \emptyset$ et il est mesurable dans $\Theta \times F_1 \times F_2$.

$A(W_3)$ désignera la trace de $A(W_2) \otimes A(F_2)$ sur W_3 , et l'on peut poursuivre la construction.

Supposons définis les ensembles $E_1, \dots, E_k, F_1, \dots, F_k, W_1, \dots, W_{k+1}, A(E_1), \dots, A(E_k), A(F_1), \dots, A(F_k), A(W_1), \dots, A(W_{k+1})$, et les fonctions $\varphi_1, \dots, \varphi_{k-1}, \overline{\varphi_1}, \dots, \overline{\varphi_k}$, pour $k \geq 2$. On pose :

$$- E_{k+1} = F(W_{k+1})$$

$$- F_{k+1} = E_{k+1} \times E_{k+1}$$

- φ_k l'application définie sur E_{k+1} par :

$$\varphi_k(e_{k+1}) = \{cp(Y) \mid Y \in e_{k+1}\},$$

- $\overline{\varphi_k}$ l'application sur F_{k+1} qui prend φ_k composante par composante,

- $W_{k+2} \subset W_{k+1} \times F_{k+1}$ l'ensemble des suites $(\theta, f_1, \dots, f_{k+1})$ qu'on appellera *k + 1-cohérentes*, telles que : $(\theta, f_1, \dots, f_k) \in W_{k+1}$ et :

$$(i) \quad f_k = \overline{\varphi_k}(f_{k+1})$$

$$(ii) \quad \forall a \in A, X \in f_k(a) \text{ ssi } \{(\tilde{\theta}, \tilde{f}_1, \dots, \tilde{f}_k) \in W_{k+1} \mid X \in \tilde{f}_k(a)\} \in f_{k+1}(a)$$

- enfin : $A(W_{k+2}) = \text{la trace de } A(W_{k+1}) \otimes A(F_{k+1}) \text{ sur } W_{k+2}$.

Cette construction par récurrence suppose que l'on ait établi pour chaque niveau les faits signalés dans les remarques 5 et 6 :

Remarque 7. Pour tout $k \geq 2$, φ_k va dans E_k ; elle est surjective et mesurable ; elle consiste à prendre la loi-image de la probabilité bivalente e_{k+1} par $cp(\cdot)$, l'ensemble des suites $(\theta, f_1, \dots, f_{k-1}, f_k)$ telles que :

$$(\theta, f_1, \dots, f_{k-1}) \in W_k \text{ et } f_{k-1} = \bar{\varphi}_{k-1}(f_k)$$

est un mesurable de $W_1 \times F_1 \times \dots \times F_k$; W_{k+1} est non vide et également mesurable.

En résumé :

W_1	$\xleftarrow{cp}$	W_2	$\xleftarrow{cp}$	W_3
$A(W_1)$		$A(W_2)$		$A(W_3)$
$E_1 = F(W_1)$	$\xleftarrow{\frac{\varphi_1}{\varphi_1}}$	$E_2 = F(W_2)$	$\xleftarrow{\frac{\varphi_2}{\varphi_2}}$	$E_3 = F(W_3) \dots$
$F_1 = E_1 \times E_1$	$\xleftarrow{\varphi_1}$	$F_2 = E_2 \times E_2$	$\xleftarrow{\varphi_2}$	$F_3 = E_3 \times$
$W_2 = \Theta \times F_1$		$W_3 \subset W_2 \times F_2$		$W_4 \subset W_3 \times F_3$
		ensemble des suites 2-cohérentes		ensemble des suites 3-cohérentes

Soit $F \subset \prod_{k \in \mathbb{N}} F_k$ l'ensemble des *suites infinies cohérentes*, c'est-à-dire des

$(f_1, \dots, f_k, \dots)$ telles que : $\forall k \geq 1, (f_1, \dots, f_{k+1}) \in W_{k+2}$ (la condition de cohérence finie ne joue qu'à partir de W_3 , i.e. il n'y a pas de contrainte sur les croyances de premier niveau). F est contenu dans l'ensemble des $(f_1, \dots, f_k, \dots)$ telles que : $\forall k \geq 1, f_k = \bar{\varphi}_k(f_{k+1})$, autrement dit dans $\varprojlim_{\leftarrow \varphi_k} F_k$ la limite projective des F_k pour

les "projections" $\bar{\varphi}_k$.

Il faut vérifier que $F \neq \emptyset$. On peut le voir sur l'exemple du *prolongement sans information*, qui peut également servir à établir que $W_2 \neq \emptyset, \forall k \geq 3$. Supposons que, pour un $k \geq 1, (f_1, \dots, f_{k+1})$ soit $k+1$ -cohérente. Le prolongement sans information à l'ordre $k+2$ est défini par : $\forall a \in A$,

$$Y \in f_{k+2}(a) \text{ ssi } \exists X \in f_{k+1}(a) \text{ tq } Y = \{(\tilde{\theta} \tilde{f}_1, \dots, \tilde{f}_{k+1}) \in W_{k+2} \mid X \in \tilde{f}_{k+1}(a)\}.$$

Cette condition est manifestement un renforcement de la moitié (ii) de la condition de cohérence entre $k+1$ et $k+2$. On peut vérifier par récurrence qu'elle implique également (i). Elle dit en substance qu'à partir du niveau $k+2$, l'agent ne fait plus d'hypothèse particulière sur la croyance de l'autre. En termes lâches, les conditions de cohérence pour toute suite $(f_1, \dots, f_k, \dots)$ déterminent *entièrement* le niveau inférieur à partir du supérieur, et *partiellement* le niveau supérieur à partir de l'inférieur. Une condition spécifique comme celle de prolongement sans information achève la détermination d'un niveau par l'autre.

Comme les produits précédemment introduits, $\prod_k F_k$ est muni de la tribu produit $\otimes A(F_k)$. Elle est engendrée par les cylindres :

$$A_1 \times \dots \times A_k \times \dots$$

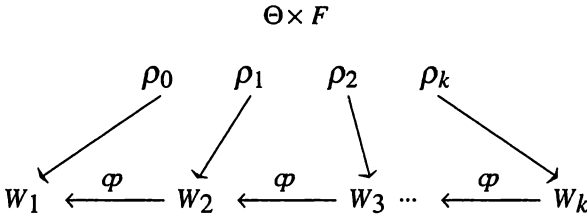
(c'est-à-dire les pavés de $\prod_k F_k$ tels que $A_k \in A(F_k)$, $\forall k$, et que $A_k = F_k$ sauf pour un nombre fini de k).

Remarque 8. L'ensemble $\Theta \times F$ est un mesurable de $\Theta \times \prod_k F_k$.

On s'intéressera aux projections ρ_k de $\Theta \times F$ sur Θ et les produits finis successifs :

$$\rho_0 : \Theta \times F \rightarrow \Theta \text{ et } k \geq 1$$

$$\rho_k : \Theta \times F \rightarrow \prod_{1 \leq i \leq k} F_i$$



Ces fonctions, qui sont des restrictions des projections naturelles, sont mesurables en vertu de la remarque 8. Soit $E^* = F(\Theta \times F)$ et $A(E^*)$ définis comme précédemment pour un tel ensemble. Donc, $A(E^*) = A^-(E^*)$, la famille des

$$\{e^* \in E^* \mid B \in e^*\} \text{ pour } B \text{ parcourant } A(\Theta \times F),$$

la trace de la tribu-produit sur $\Theta \times F$. On pose aussi $F^* = E^* \times E^*$. On veut démontrer qu'un élément f^* de F^* est essentiellement identique à une suite $f = (f_1, \dots, f_k, \dots)$ de F et que la bijection préserve la structure mesurable. Ce résultat peut se voir à l'aide des applications suivantes :

- pour tout $a \in A$, soit $\zeta^a : F \rightarrow 2^{2^{\Theta \times F}}$ telle que :

$$B \in \zeta^a(f) \text{ ssi } B \in A(\Theta \times F) \text{ et } \forall k \geq 1, \rho_{k-1}(B) \in f_k(a)$$

et ζ la fonction définie sur F qui applique ζ^a composante par composante ;

- soit $\eta : F^* \rightarrow \prod_k F_k$ définie par $\eta = (\eta_1, \dots, \eta_k, \dots)$ et :

$$\forall a \in A, \forall k \geq 1, \forall X \in A(W_{k-1}), X \in \eta_k(f^*)(a) \text{ ssi } \exists Y \in f^*(a) \text{ tq. } Y = \rho_{k-1}(X).$$

Remarque 9. ζ^a va dans E^* et η va dans F . Les fonctions ζ et η sont inverses l'une de l'autre. Elles sont bimesurables.

Cette dernière remarque livre l'analogie suivant du théorème de Mertens et Zamir :

Proposition 2. La construction implique l'isomorphisme suivant :

$$F^*, A(F^*) \cong F, A(F).$$

La notation qui précède ne se justifie, encore une fois, que par des généralisations potentielles et pour éclairer la nature abstraite de la démarche. Car, pour traiter le cas où $(\Theta, A(\Theta))$ est séparable, il vaut mieux identifier les $e_1 \in E_1$ à des masses de Dirac. Cette identification résulte elle-même des deux identifications suivantes : e_1 s'identifie à son atome caractéristique dans $At(\Theta)$, et les θ contenus dans un même atome s'identifient entre eux (aucun mesurable ne les distingue). On écrira donc e_1 comme un point de Θ . L'identification se propage aux niveaux supérieurs, de sorte que chaque e_k ou f_k peut s'écrire comme une suite convenable d'éléments de Θ . L'application récursive des conditions (i) et (ii) permet des simplifications supplémentaires. Ainsi,

$$\cdot f_1 \in F_1 \Rightarrow f_1(a), f_1(b) \in \Theta$$

$$\begin{aligned} \cdot f_2 \in F_2 &\Rightarrow f_2(a) = (\theta, (g_1(a), g_1(b))) \text{ avec } g_1(a), g_1(b) \in \Theta \\ f_2(b) &= (\theta', (g'_1(a), g'_1(b))) \text{ avec } g'_1(a), g'_1(b) \in \Theta \end{aligned}$$

$$\cdot \text{si } (f_1, f_2) \in W_2,$$

$$(i) \Rightarrow \theta = f_1(a) \text{ et } \theta' = f'_1(b)$$

$$(ii) \Rightarrow \theta = g_1(a) \text{ et } \theta' = g'_1(b)$$

de sorte que f_2 s'identifie maintenant à $(g_1(b), g'_1(a))$

Il est fastidieux, mais facile, de vérifier plus généralement :

Remarque 10. Une suite $k + 1$ -cohérente $(f_1, \dots, f_{k+1})$ s'identifie à un élément de Θ^{2k+2} . Une suite $f = (f_1, \dots, f_k, \dots) \in F$ s'identifie à un élément de $\Theta^{\mathbb{N}} \times \Theta^{\mathbb{N}}$.

Comme une probabilité bivalente e^* sur F s'identifie à un élément de $\Theta^{\mathbb{N}}$, l'isomorphisme $f^* \cong f$ n'est plus qu'une trivialité. Il est utile de noter ainsi l'identification des suites cohérentes :

$$f \cong (\theta^a, \theta^b, \theta^{ba}, \theta^{ab}, \theta^{aba}, \theta^{bab}, \theta^{abab}, \dots)$$

Le niveau $k + 1$ n'ajoute au niveau K que deux informations : la croyance que a forme de la k -croyance de b , et la croyance que b forme de la k -croyance de a .

4. Une interprétation logique des hiérarchie de probabilités bivalentes

Cette section se propose d'interpréter logiquement la variante simplifiée de la construction de Mertens et Zamir : il s'agira d'axiomatiser les hiérarchies de probabilités bivalentes (H.P.B.) induites par des espaces mesurables $(\Theta, \mathcal{A}(\Theta))$ séparables. Compte tenu des identifications signalées à la fin de la section 3, il suffira d'axiomatiser les H.P.B. construites sur des espaces mesurables *discrets* $(\Theta, 2^\Theta)$, avec $|\Theta| \leq 2^N$.

On se donne un ensemble dénombrable VP de variables propositionnelles et on note $\Theta = 2^{VP}$ l'ensemble des valuations. La morphologie et la syntaxe du calcul sont définies comme dans la première section : soit $L(VP)$ l'ensemble des formules bien formées ; (SP) est le système d'axiomes et de règles choisi.

La sémantique du calcul est fournie par la construction de la seconde section, qui se simplifie grandement, puisque les aspects mesurables en disparaissent. Une *k-structure* est une suite finie $(\theta, f_1, \dots, f_{k-1})$ telle que $\theta \in \Theta = 2^{VP}$ et que $(f_1, \dots, f_{k-1})$ soit $(k-1)$ -cohérente. Soit W_k l'ensemble des *k-structures*. Une *structure* est une suite infinie $(\theta, f_1, \dots, f_{k-1}, \dots)$ telle que $\theta \in \Theta = 2^{VP}$ et que $(f_1, \dots, f_{k-1}, \dots)$ soit cohérente. Soit $\Theta \times F$ l'ensemble des structures. Il reste à définir les conditions de validation ; on le fera en suivant Vardi (1986).

Celui-ci introduit tout d'abord la validation d'une formule φ par une *k-structure* :

$\forall k \geq 0$ [on pose $f_0 = \theta$]

- si $\varphi \in VP$, $w_k = \langle \theta, f_1, \dots, f_{k-1} \rangle \models \varphi$ ssi $\theta(\varphi) = 1$,
 - si $\varphi = \neg\psi$, $w_k = \langle \theta, f_1, \dots, f_{k-1} \rangle \models \varphi$ ssi $w_k \not\models \psi$,
- et ainsi de suite pour les autres clauses propositionnelles,
- si $\varphi = B_a \psi$, $w_k = \langle \theta, f_1, \dots, f_{k-1} \rangle \models \varphi$ ssi $k \geq 1$

et $[|\psi|]_{k-1} \stackrel{\text{def}}{=} \{w_{k-1} \in W_{k-1} \mid w_{k-1} \models \psi\} \in f_{k-1}(a)$

Avant de définir la validation par une *structure* de $\Theta \times F$, Vardi introduit la notion auxiliaire de *profondeur* d'une formule φ . Celle-ci est notée $\text{prof}\varphi$ et s'énonce dans la récurrence suivante :

- $\text{prof}\varphi = 0$ ssi $\varphi \in VP$
- $\text{prof}\neg\varphi = \text{prof}\varphi$
- $\text{prof}(\varphi \wedge \psi) = \max\{\text{prof}\varphi, \text{prof}\psi\}$

et de même pour les autres connecteurs binaires

- $B_a \varphi = 1 + \text{prof}\varphi$.

La profondeur est une notion morphologique, naturellement différente de celle de complexité (qui intervenait dans la construction de $L(VP)$).

La validation d'une formule φ par une structure peut enfin s'écrire :

si $\text{prof } \varphi = k$,

$$w = \langle \theta, f_1, \dots, f_k, \dots \rangle \models \varphi \text{ ssi } w|_k \stackrel{\text{def}}{=} \langle \theta, f_1, \dots, f_k \rangle \models \varphi.$$

Cette définition paraît se heurter au problème suivant. Soit φ_1 et φ_2 de profondeurs respectives k_1 et k_2 , avec $k_1 < k_2$. On a :

$$w \models \varphi_1 \text{ ssi } w|_{k_1} \models \varphi_1 \text{ et } w \models \varphi_2 \text{ ssi } w|_{k_2} \models \varphi_2.$$

Mais l'expression $w|_{k_2} \models \varphi_1$ a également du sens, puisque la validation *finitaire* ne fait pas référence à la profondeur. On peut donc se trouver dans le cas :

$$w \models \varphi_1 \vee \varphi_2 \text{ ssi } w|_{k_2} \models \varphi_1 \text{ ou } w|_{k_2} \models \varphi_2.$$

Il serait très inopportun que l'on ait simultanément $w|_{k_2} \not\models \varphi_1$ et $w|_{k_1} \models \varphi_1$, puisque l'on pourrait avoir à conclure : $w \not\models \varphi_1 \vee \varphi_2$ et $w \models \varphi_1$. Le problème hypothétiquement soulevé par cet exemple ne se pose en fait pas, comme le montre la remarque suivante :¹³

Remarque 11. Soit $k_1 = \text{prof } \varphi$, $k_2 > k_1$ et $w_{k_2+1} = \langle \theta, f_1, \dots, f_{k_2} \rangle \in W_{k_2+1}$. Alors

$$w_{k_2+1} \models \varphi \text{ ssi } \langle \theta, f_1, \dots, f_{k_1} \rangle \models \varphi$$

On va montrer :

Proposition 3. *Le système (SP) est une axiomatisation adéquate et complète de la classe des structures $\Theta \times F$.*

Comme à l'accoutumée, la démonstration d'adéquation est facile :

Lemme 3.1. *Pour tout $\varphi \in L(VP)$,*

$$\frac{}{SM} \varphi \Rightarrow \models_{\Theta \times F} \varphi.$$

La démonstration de complétude repose sur la technique ordinaire des *ensembles maximaux de formules*. Cette technique demande à être adaptée au cas des structures itératives $(\theta, f_1, \dots, f_k, \dots)$, ce qui entraîne dans des démonstrations par récurrence assez lourdes. Pour simplifier l'exposition, on se restreindra tout d'abord aux

13. La remarque 11 est identique au lemme 1 de Vardi (1986, p. 298). Malheureusement, l'énoncé du lemme est faux dans la définition très générale que cet auteur donne des *k-structures* (les *finitary belief worlds*). Il est en revanche facile à vérifier sous les conditions beaucoup plus fortes que l'on impose ici sur les *k-structures*. Vardi est revenu sur la sémantique des mondes de croyances dans un article de 1989, mais il n'y signale pas cette difficulté technique, pourtant sérieuse : elle affecte sa démonstration de complétude au moins dans le cas plus général des mondes de croyances (celui qui correspond, syntaxiquement, au système minimal de la logique modale propositionnelle).

systèmes et aux structures obtenus à partir d'un ensemble VP fini de variables propositionnelles. On indiquera ensuite comment on peut se débarrasser de cette limitation pour obtenir le théorème général de complétude.

Soit $VP = \{p_1, \dots, p_q\}$ et I l'ensemble au maximaux consistants associés au système (SP) . On définit :

$$[\varphi]_I = \{\Delta \in I \mid \varphi \in \Delta\}$$

et, pour tout $a \in A$, la fonction $N_a : I \rightarrow 2^{2^I}$ telle que :

$$\forall \Delta \in I, N_a(\Delta) = \{P \in 2^I \mid \exists \varphi \text{ tq. } B_a \varphi \in \Delta \text{ et } [\varphi]_I \subseteq P\}$$

$N_a(\Delta)$ s'appellera *système de voisinages* de l'élément Δ pour $a \in A$. La terminologie évoque à dessein le formalisme des structures de voisinage, que l'on a utilisé dans la première section.

Lemme 3.2. Pour tout $a \in A$ et $\Delta \in I$, $N_a(\Delta)$ vérifie :

- 0. $\forall \varphi \in L(VP)$, $[\varphi]_I \in N_a(\Delta)$ ssi $B_a \varphi \in \Delta$.
- 1. – 5. $N_a(\Delta)$ est un ultrafiltre de parties de I .
- 6. $P \in N_a(\Delta) \Leftrightarrow \{\Delta \in I \mid P \in N_a(\Delta)\} \in N_a(\Delta)$.

Pour tout $\Delta \in I$, on pose :

$$\theta^\Delta = \Delta \cap VP \text{ et } \langle \theta^\Delta \rangle = w_1^\Delta.$$

Soit W_1^I l'ensemble des w_1^Δ , pour Δ parcourant I . On définit ensuite :

$$f_1^\Delta : 2^{2^{W_1^I}} \text{ par}$$

$$f_1^\Delta(a) = \{\{w_1^\Delta, \Delta \in y\}, y \in N_a(\Delta)\}.$$

Soit F_1^I l'ensemble des f_1^Δ et W_2^I l'ensemble des $\langle \theta^\Delta, f_1^\Delta \rangle$ pour Δ parcourant I .

Supposons construits $\langle \theta^\Delta, f_1^\Delta, \dots, f_k^\Delta \rangle$, ainsi que $F_1^I, \dots, F_k^I$ et $W_1^I, \dots, W_{k+1}^I$.

On définit :

$$f_{k+1}^\Delta : A \rightarrow 2^{2^{W_{k+1}^I}}$$

$$f_{k+1}^\Delta(a) = \{\{w_{k+1}^\Delta, \Delta \in y\}, y \in N_a(\Delta)\}.$$

F_{k+1}^I et W_{k+2}^I sont les ensembles, respectivement, des f_{k+1}^Δ et des w_{k+2}^Δ , $\Delta \in I$.

On note enfin F^I l'ensemble des suites indicées $\langle \theta^\Delta, f_1^\Delta, \dots, f_k^\Delta, \dots \rangle$, $\Delta \in I$.

Il faut vérifier que les suites ainsi construites vérifient les propriétés des hiérarchies de probabilités bivalentes :

Lemme 3.3. $F^I \subseteq F$.

La démonstration fait bien ressortir le rôle joué par l'hypothèse de finitude que, pour l'instant, on impose sur VP . A chaque étape k , on n'a jamais considéré qu'un nombre *fini* de propositions, ou d'ensembles de propositions. Le caractère infinitaire de la construction apparaît uniquement dans la récurrence sur k . Ce point est important pour l'établissement du lemme.

Il reste maintenant à établir la propriété dite de *canonicité* des éléments de $\Theta \times F^I$:

Lemme 3.4. Pour tout φ et tout $\Delta \in I$,

$$(\theta^\Delta, \theta_1^\Delta, \dots, \theta_k^\Delta, \dots) \models \varphi \text{ ssi } \varphi \in \Delta.$$

Cette propriété est caractéristique des classes de modèles construits à partir des maximaux consistants (p. ex. Chellas, 1986, p. 60). Elle permet de conclure à la complétude par le raisonnement classique : supposons $\models_{\Theta \times F} \varphi$, alors φ est vraie dans la sous-classe $\Theta^I \times F^I$ (lemme 3.3) et φ appartient à tout maximal consistant Δ (lemme 3.4), c'est-à-dire $\vdash_{SM} \varphi$.

L'extension au cas où VP est infini demande quelques préalables algébriques. Une formule $\varphi \in L(VP)$ étant donnée, on isole les variables propositionnelles $p_{i_1}, \dots, p_{i_k}$ — nécessairement en nombre fini — qui interviennent dans cette formule. Soit $VP_\varphi = \{p_{i_1}, \dots, p_{i_k}\}$ et $L(VP_\varphi)$ l'ensemble correspondant de formules bien formées. Les notions de structure, de validation sémantique et de tautologie sont, ici comme ailleurs en théorie des modèles, relatives au langage considéré. Il s'agit alors de montrer que, si φ est une tautologie pour la sémantique relative à VP , elle est une tautologie pour la sémantique relative au sous-ensemble *fini* VP_φ :

Lemme 3.5. Pour tout $\varphi \in L(VP)$,

$$\models_{VP} \varphi \Rightarrow \models_{VP_\varphi} \varphi$$

Une fois acquis ce lemme, on dispose, quel que soit le (VP) considéré, d'une démonstration de complétude *formule par formule*. Cette manière non uniforme d'établir la complétude d'un système n'est pas rare en logique modale. Elle s'apparente aux procédés de "filtrages" discutés par Chellas (1980, p. 41-45). Lismont (1992) a suivi cette voie pour établir la complétude de systèmes modaux "multi-agents", enrichis d'un opérateur de "connaissance commune". On peut

vérifier que, contrairement à la démonstration classique, menée indépendamment du $\varnothing$ considéré, une démonstration administrée formule par formule interdit de conclure de la complétude du système à la propriété (sémantique) de *compacité*. Compte tenu de cette perte d'information, le schéma de raisonnement proposé ici pour dériver la Proposition 3 ne dispense pas d'examiner si une démonstration classique — éventuellement plus complexe à mettre en œuvre — n'aboutirait pas au même résultat.

5. Conclusion

Lorsqu'elle entreprend de définir une notion de *type* stratégique, la théorie des jeux se heurte, de prime abord, à une difficulté logique : la “régression infinie des anticipations mutuelles”. En suivant les indications d'Harsanyi, Mertens et Zamir ont proposé une construction probabiliste qui répond à cette difficulté. Leur méthode consiste, en quelque sorte, à transformer l'itération qui fait initialement problème — la régression infinie — en un objet mathématique ; une fois soumis à des restrictions techniquement appropriées, cet objet mathématique peut s'interpréter comme la notion de type que l'on recherchait. Les restrictions techniques s'expriment en langage probabiliste et apparaissent interprétables, d'un point de vue épistémique plus général.

En revenant sur le moment constitutif de la théorie des jeux en information incomplète, on ne se proposait pas de réexaminer les fondements, désormais bien compris, de cette approche, mais plutôt d'en faire ressortir, pour l'analyser ensuite, la parenté inattendue avec les approches logiques de l'Intelligence Artificielle et de l'Informatique des Systèmes Répartis. Il s'agit de logiques modales propositionnelles à opérateurs multiples ; les opérateurs, qui sont dits “de connaissance” ou “de croyance”, renvoient chacun à un individu ou à un “processeur” distinct. Plus précisément, on s'est laissé guider par l'analogie frappante que la construction de Mertens et Zamir présente avec celle de l'informaticien Vardi (1986). Sous le nom de *monde de croyance* (*belief world*), celui-ci propose une sémantique itérative nouvelle pour la logique modale épistémique. La construction de Vardi peut s'interpréter comme une autre formalisation — exempliste et non pas probabiliste, donc plus pauvre, mais aussi plus générale — de la “régression infinie des anticipations mutuelles” qui préoccupait les théoriciens des jeux.

Pour établir une relation techniquement précise entre les deux démarches dont on percevait la proximité, on s'est efforcé d'isoler une classe d'objets ensemblistes — de structures — répondant à une double contrainte a priori.

(1) Ces structures devaient formaliser la “régression infinie des croyances mutuelles” à la manière de Mertens et Zamir, c'est-à-dire en préservant les *propriétés d'isomorphisme* qui caractérisent leur construction. Ce sont en effet ces propriétés qui permettent d'aboutir à une notion logiquement satisfaisante de type stratégique. Elles indiquent en substance que la régression infinie épuise le contenu que l'on peut donner à la notion de type. Plus techniquement, celle-ci se définit comme une hiérarchie de croyances. Le théorème d'isomorphisme ajoute cette

précision importante : il n'est pas nécessaire de considérer des hiérarchies d'ordre supérieur (c'est-à-dire, la hiérarchie infinie des croyances portant sur les hiérarchies tout d'abord considérées, et ainsi de suite). Indépendamment de l'objet particulier de la théorie des jeux, on peut admettre que la "stabilisation" ainsi obtenue constitue une propriété souhaitable de toute hiérarchie épistémique. On s'est donc imposé la contrainte de MAINTENIR LES PROPRIÉTÉS D'ISOMORPHISME DE LA CONSTRUCTION DE MERTENS-ZAMIR.

(2) Les structures recherchées devaient, comme celles de Vardi, constituer une sémantique de la logique modale épistémique. Il fallait, en substance, passer d'une construction probabiliste, trop riche pour être logiquement interprétable, à une construction, purement ensembliste, qui serait, quant à elle, logiquement interprétable. Plus précisément, l'appauvrissement visait à remplir la deuxième condition posée a priori : PARVENIR À UNE CLASSE DE STRUCTURES QU'ON PUISSE AXIOMATISER PAR LA LOGIQUE MODALE ÉPISTÉMIQUE. L'axiomatisation s'entend ici au sens ordinaire — l'obtention d'un théorème d'adéquation et de complétude.

En réponse à cette double exigence, on a proposé la classe des *hiérarchies de probabilités bivalentes*, construites sur des tribus séparables. Une telle résolution n'est pas très satisfaisante, parce que les probabilités à deux valeurs sont un cas-limite dégénéré. Syntaxiquement, la contrainte de bivalence s'exprime sous la forme d'un axiome de "complémentarité", noté (CP_a) , dont la logique modale épistémique ne fait pas couramment usage. La dégénérescence du système modal peut se voir ainsi : en présence du schéma traditionnel (P_a) , qui est inévitable ici, (CP_a) conduit à une équivalence de l'opérateur de croyance B_a et de l'opérateur dual $\neg B_a \neg$. Cette équivalence n'est évidemment pas souhaitable au point de vue épistémique. Il faut donc regarder les résultats du présent travail comme une première approximation, à perfectionner rapidement.

La classe de structures plus complexes dont on poursuit actuellement l'examen comporte des hiérarchies possibles sans restriction sur le nombre de valeurs permises : on revient donc à l'opération $\Delta(E)$ — prendre l'ensemble de toutes les probabilités sur E — comme chez Mertens et Zamir. Mais on définit la validation sémantique de telle façon que seuls les ensembles de mesure 1 interviennent : on dira que $B_a \varphi$ est vrai si et seulement si la proposition $[|\varphi|]$ reçoit une valeur 1 par une probabilité choisie à un niveau convenable de la hiérarchie. Il n'y a pas de difficultés à préciser les conditions de cohérence imposées à la hiérarchie qui permettront d'introduire formellement une définition de la validité sémantique. En outre, celle-ci est naturelle, et même banale, d'un point de vue épistémique. Le problème est — encore une fois — d'assurer la réalisation simultanée des contraintes a priori (1) et (2). S'agissant de (1), il faudra spécifier les conditions de mesurabilité sur l'espace initial Θ qui assureront indirectement l'existence de l'isomorphisme. Il n'est pas sûr que ces conditions techniques soient logiquement interprétables, ce qui reporte la difficulté sur (2).

Bibliographie

I. Travaux cités dans l'article

- Brandenburger, A. et E. Dekel (1985), "Hierarchies of Beliefs and Common Knowledge," Research Paper N° 841, Graduate School of Business, Stanford University.
- Chellas, B.F. (1980), *Modal Logic*, Cambridge, Cambridge University Press.
- Dudley, R.M. (1989), *Real Analysis and Probability*, Pacific Grove, Ca., Wadsworth & Brooks.
- Fagin, R., J. Geanakoplos, J.Y. Halpern et M.Y. Vardi (1992), "The Expressive Power of the Hierarchical Approach to Modeling Knowledge and Common Knowledge," in Moses (1992), p. 229-244.
- Fagin, R. et J.Y. Halpern (1986), "Reasoning About Knowledge and Probability : Preliminary Report," in Halpern (1986), p. 277-293.
- Fagin, R., J.Y. Halpern et M.Y. Vardi (1984), "A Model-Theoretic Analysis of Knowledge," *Proc. 25th IEEE Symp. on the Foundations of Computer Science*, West Palm Beach, p. 268-278. Version révisée dans *Journal of the Association for Computing Machinery*, 1991, 38, p. 382-428.
- Gärdenfors, P. (1975), "Qualitative Probability as an Intensional Logic," *Journal of Philosophical Logic*, 4, p. 171-185.
- Geanakoplos, J.D. (1992), "Common Knowledge," *Journal of Economic Perspectives*, 6, p. 53-82.
- Halmos, P.R. (1974), *Lectures on Boolean Algebras*, New York, Springer.
- Halpern, J.Y. (1986), "Reasoning about Knowledge : An Overview," in Halpern (1986), p. 1-29.
- Halpern, J.Y. (dir.) (1986), *Second Conference on Theoretical Aspects of Reasoning About Knowledge*, Los Altos, Ca., Morgan Kaufmann.
- Harsanyi, J.C. (1967-1968), "Games with Incomplete Information Played by 'Bayesian' Players, I-III," *Management Science*, 14, p. 159-182, p. 320-334, p. 486-502.
- Hughes, G.E. et M.J. Cresswell (1984), *A Companion to Modal Logic*, Londres, Methuen.
- Lismont, L. (1992), *La connaissance commune. Approches modale, ensembliste et probabiliste*, Thèse de Doctorat en Sciences, Université Catholique de Louvain.
- Mertens, J.-F., S. Sorin et S. Zamir (1990), "The Belief Spaces," Chapitre d'un ouvrage en préparation, Université Catholique de Louvain.
- Mertens, J.-F. et S. Zamir (1985), "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, p. 1-29.
- Moses, Y. (dir.) (1992), *Fourth Conference on Theoretical Aspects of Reasoning About Knowledge*, Los Altos, Ca., Morgan Kaufmann.
- Vardi, M.Y. (1986), "On Epistemic Logic and Logical Omniscience," in Halpern (1986), p. 293-305.
- Vardi, M.Y. (1989), "On the Complexity of Epistemic Reasoning," in *Proceedings Fourth Annual Symposium on Logic in Computer Science*, p. 243-252.

II. Travaux récents liés au sujet de cet article

- Aumann, R. et D. Samet (1993), "A Syntactic Model of Interactive Beliefs," manuscrit, Universités de Jérusalem et de Tel-Aviv.
- Bacharach, M. et P. Mongin (1994), "Epistemic Logic and the Foundations of Game Theory," Introduction au numéro spécial *Logic and the Epistemic Foundations of Game Theory, Theory and Decision*, 37(1), p. 1-6.
- Bradenburger, A. et E. Dekel (1993), "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, p. 189-198.
- Fagin, R. (1994), "A Quantitative Analysis of Modal Logic," *Journal of Symbolic Logic*, 59, p. 209-252.
- Hammond, P. (1994), "Information Partitions Over Multiperson Hierarchies of Partial Information," article présenté au Colloque : "Epistemic Logic and the Theory of Games and Decisions," Marseille, Centre International de Recherches Mathématiques (24-28 janvier 1994).
- Heifetz, A. (1993), "The Bayesian Formulation of Incomplete Information. The Noncompact Case," *International Journal of Game Theory*, 21, p. 329-338.
- Heifetz, A. (1994), *Transfinite Levels of Mutual Knowledge and Belief in a Group of Players*, Thèse de Doctorat, Faculté des Sciences Exactes, Département de mathématiques, Université de Tel-Aviv.
- Heifetz, A. et D. Samet (1993), "Universal Partition Structures," Working Paper N° 26/93, Faculté de gestion, Université de Tel-Aviv.
- Lismont, L. (1993a), "La connaissance commune en logique modale," *Mathematical Logic Quarterly*, 39, p. 115-130.
- Lismont, L. (1993b), "Common Knowledge. Relating Anti-Founded Situation Semantics to Neighbourhood Semantics," Prépublication du Département de Mathématiques, Université Catholique de Louvain.
- Lismont, L. et P. Mongin (1994a), "On the Logic of Common Belief and Common Knowledge," *Theory and Decision*, 37, p. 75-106.
- Lismont, L. et P. Mongin (1994b), "Belief Closure : A Semantics of Common Knowledge for Modal Propositional Logic," CORE Discussion Paper N° 9339, à paraître dans *Mathematical Social Sciences*.
- Mertens, J.-F., S. Sorin et S. Zamir (1994), "Repeated Games. Part A : Background Material," CORE Discussion Paper N° 9420, Université Catholique de Louvain

Philippe MONGIN

Ce texte a été rédigé, pour l'essentiel, en 1990-1991. Il a fait l'objet de légères modifications avant d'être présenté au Colloque : "Intelligence artificielle et logiques non classiques" (Montréal, 17-20 octobre 1992). Au moment de le livrer tardivement à la publication, l'auteur a préféré ne pas entreprendre de révisions supplémentaires. L'évolution de son propre travail, mais, surtout, l'apparition ultérieure d'une littérature spécialisée qui vise à relier théorie des jeux et logique épistémique, incitaient sans doute à reprendre systématiquement le projet de rapprochement qui fait l'objet de cet article. Mais celui-ci a le mérite, aussi bien que les défauts, de son antériorité relative. C'est pourquoi il a semblé préférable de laisser le texte en l'état, tout en étoffant la bibliographie de références techniques sélectionnées dans la littérature récente.

Chapitre 9

Applications épistémiques de la théorie des ensembles anti-fondés

Le but de cet article est de montrer l'importance de la théorie ZFC-AFA des ensembles anti-fondés pour ses applications épistémiques. L'intérêt de cette théorie pour développer des sémantiques dans le cas des modélisations de situations où intervient une circularité est déjà bien connu. Ceci a été mis en évidence par J. Barwise notamment par une analyse du paradoxe du menteur ([5]) et une modélisation de la connaissance commune en logique épistémique ([4]). Nous montrons une nouvelle situation où cette théorie permet une modélisation aisée : les espaces universels de croyances introduits par Mertens et Zamir ([15]). De plus, utilisant ZFC-AFA comme méta-théorie, nous développons un langage modal anti-fondé autorisant, par exemple, des formules circulaires telles que celle définie par $\phi = \neg\phi$. Nous proposons un calcul de séquents anti-fondé, contre partie syntaxique de modèles anti-fondés. La sémantique proposée est une extension naturelle et conservatrice de la sémantique usuelle des voisinages clos par sur-ensembles. Nous énonçons des théorèmes d'adéquation et de complétude pour ce calcul de séquents. Nous envisageons les applications d'une telle logique, naturellement paraconsistante, pour la modélisation des croyances.

1. Introduction

Notre propos concerne l'utilisation épistémique de la théorie ZFC-AFA des ensembles anti-fondés. Introduite par M. Forti ([6]), cette théorie fut popularisée par P. Aczel ([1]) et par J. Barwise dans son analyse du paradoxe du menteur ([5]). Nous pensons que cette théorie pourra se révéler très féconde dans le domaine de la logique épistémique. L'objectif de cet article est d'essayer de le montrer. Pour cela, nous présenterons de manière relativement brève, sans démonstration, deux nouvelles pistes que nous avons explorées : une version anti-fondée d'un théorème important

en théorie des jeux et une extension anti-fondée de la logique modale monotone. Nous reviendrons également, encore plus brièvement, sur le travail de Barwise dans son analyse de la connaissance commune ([4]).

Avant de caractériser brièvement la théorie ZFC-AFA, nous allons passer en revue les deux contextes épistémiques dans lesquels nous l'utiliserons : les *espaces universels de croyance* et la *connaissance commune*.

La construction des *espaces universels de croyance* est liée au fondement de la modélisation bayésienne des jeux à information non complète. Dans ce cadre, on suppose qu'avant même le début du jeu, il y a des informations *privées*, qui ne sont pas *connaissance commune*. Nous reviendrons plus loin sur la notion de connaissance commune. Tout ce qu'il nous faut savoir ici, c'est qu'il y a une incertitude des agents (ou des joueurs) quant à la connaissance des autres agents. On appellera *type* d'un agent l'ensemble des informations privées dont il dispose au départ. Naturellement, ce type comporte non seulement l'information de l'agent par rapport à la réalité objective mais aussi son information quant à l'information dont dispose les autres agents. Les incertitudes éventuelles à propos de ces informations seront évaluées au moyen de probabilités dites *subjectives*.

La construction de Mertens et Zamir ([15]) ou de Mertens, Sorin et Zamir ([16]) répond à une objection contre une telle modélisation des jeux à information incomplète. Pour présenter cette objection, considérons deux agents Albert et Bernadette.

- Supposons que T soit l'ensemble de tous les types possibles ;
- $\Lambda(T)$ est l'ensemble de toutes les probabilités sur T ;
- La croyance d'Albert au sujet du type de Bernadette est donnée par $\pi \in \Lambda(T)$;
- La croyance de Bernadette au sujet de π est un paramètre qui devrait compter dans la croyance d'Albert au sujet du type de Bernadette ;
- Donc π ne capture pas toute l'incertitude d'Albert au sujet du type de Bernadette et T ne peut pas être l'ensemble de tous les types possibles ;
- Supposons alors que $T \times \Lambda(T)$ soit l'ensemble de tous les types possibles ;
- Ceci ne suffit pas puisqu'alors la croyance d'Albert au sujet de la croyance de Bernadette au sujet de la croyance d'Albert capturé par π devient alors un élément dont Bernadette devrait tenir compte. Donc le type de Bernadette doit contenir un élément de $\Lambda(T \times \Lambda(T))$.
- Etc.

Comment dès lors construire un ensemble suffisamment grand qui prenne en compte cette hiérarchie infinie de probabilités subjectives ? C'est la question que résout la construction de Mertens et de Zamir. Nous ne reprendrons pas en détail la construction telle que ces auteurs l'ont présentée. Nous nous contenterons de proposer une autre construction, semblable, mais techniquement plus aisée qui se fait dans le cadre de la théorie des ensembles anti-fondés ZFC-AFA. Nous établirons également un lien entre l'espace obtenu par cette nouvelle construction et l'espace obtenu par Mertens et Zamir.

La deuxième problématique que nous aborderons est liée à celle de la *connaissance commune*. Cette notion fut introduite par Lewis en 1968 ([10]). Son importance en théorie des jeux est apparue clairement à la suite d'un article fameux de R. Aumann ([2]). Pour présenter cette notion, voici une petite histoire :

Un groupe de cinquante hommes à l'habitude de se réunir tous les midis sur la place d'un petit village proche de Bagdad. Tous ces hommes sont cocus mais ils ne le savent pas. Par contre ce que chacun sait pertinemment bien, c'est que tous les autres membres du groupe sont cocus, et pour cause ! La coutume veut que quand quelqu'un sait qu'il est cocu, il n'a qu'une attitude possible : il tue sa femme et va aussitôt dans le désert pour s'en repentir. Un jour, un missionnaire européen vint à passer par là. Ayant confessé en secret un des "coupables", il ne put s'empêcher de s'écrier devant tout le groupe réuni : "Vous rendez-vous compte qu'il y a un cocu parmi vous ?". La réaction se fit attendre... 49 jours plus tard tous les hommes tuèrent leur femme et s'enfuirent dans le désert.

Pour expliquer ce qui s'est passé, prenons la même histoire avec 2 hommes seulement. Après l'annonce publique du missionnaire, le raisonnement de chacun est le suivant :

- Le jour même : comme je ne suis pas cocu, ce que l'autre sait bien, il sait maintenant qu'il est cocu.

- Le lendemain : l'autre est toujours là. Incroyable. S'il est toujours là, c'est qu'il ne sait pas que je ne suis pas cocu, c'est donc qu'il sait que je le suis, c'est donc que je le suis. Rentrant chez lui, il fait ce que la morale lui réclame.

L'autre personne tient évidemment exactement le même raisonnement. Que se passe-t-il s'il y a 3 personnes ? Chacun croit qu'il n'est pas cocu. Il se dit donc que les deux autres tiendront le raisonnement que nous venons de faire pour deux personnes. Il s'attend donc à ne plus voir les deux autres le deuxième jour. Comme ils sont encore là, c'est qu'il est cocu. En prolongeant ce raisonnement, on voit que s'il sont cinquante, chacun s'attend à ce que les 49 autres soient absents le 49ème jour. Comme ce n'est pas le cas, chacun se dit qu'il est cocu.

On peut analyser cette situation en remarquant qu'au départ chaque personne savait individuellement qu'il y avait un cocu ; cette connaissance était en un sens partagée par tout le monde mais restait une *information privée*. Au moment où il fut publiquement dit qu'il y avait un cocu parmi le groupe, cette information a pris le statut de *connaissance commune*. Ceci a permis à chacun de faire un certain nombre d'inférences logiques aboutissant à la conclusion qu'il était lui-même cocu.

De manière plus formelle, cette notion de connaissance commune peut être approchée de deux manière différentes. Une première approche consiste à dire qu'un événement ou une proposition ϕ est connaissance commune de deux agents A et B si la suite infinie des affirmations suivantes est vérifiée :

A sait que ϕ • B sait que ϕ
 A sait que B sait que ϕ • B sait que A sait que ϕ
 A sait que B sait que A sait que ϕ • B sait que A sait que B sait que ϕ
 ... • ...

Ceci est vraiment la façon classique de voir les choses (voir par exemple [2], [3], [8], [17], [18] ou [19]). On peut la formaliser entièrement dans un cadre non probabiliste au moyen de la logique modale. C'est ce que nous avons fait dans [14].

Un autre point de vue consiste à considérer ce que représente réellement une situation où il y a connaissance commune d'un certain fait σ : ce qui fait partie intégrante de cette situation, c'est que les acteurs connaissent σ et qu'ils connaissent cette situation. Nous avons ici une situation circulaire où l'un des constituants de la situation est la connaissance qu'en ont les acteurs. On peut la résumer en une phase :

**La situation de connaissance commune de σ
est une situation dans laquelle
chaque acteur connaît σ
et dans laquelle
chaque acteur connaît cette situation dans laquelle σ est
connaissance commune.**

On peut appeler cette façon de voir *approche du point fixe*. C'est elle qui est la base du développement de Barwise dans [4].

Dans [13] nous montrons comment mettre en relation une des approches modales que nous avons développées dans [14] avec cette approche par point fixe, dans ZFC-AFA, que propose Barwise dans [4]. Ce rapprochement nous a donné l'idée de prolonger le langage modal usuel en un langage modal *anti-fondé* de manière à y faire apparaître explicitement la circularité de l'approche par point fixe de la connaissance commune. Ceci permet de traiter la question de la connaissance commune en logique modale tout en prenant explicitement en considération l'approche par point fixe.

Il est temps maintenant de présenter plus avant cette théorie des ensembles anti-fondés. L'idée générale est de chercher une alternative à l'axiome de fondement qui élimine de la théorie usuelle des ensembles la circularité ou les descentes infinies pour la relation d'appartenance.

Pour présenter l'axiome d'anti-fondement qui, dans ZFC-AFA, remplace l'axiome de fondement, il est nécessaire d'introduire un certain nombre de considérations. Si l'on prend la théorie usuelle des ensembles, ZFC, *chaque ensemble peut se représenter par un graphe pointé et accessible*.

Un *graphe pointé* est un ensemble de la forme $\langle g, f, s \rangle$ où g est un ensemble de nœuds, f un ensemble de flèches, c'est-à-dire une relation binaire sur g , et s le sommet du graphe. Un *chemin* est une suite, finie ou infinie, de flèches $n_1 \rightarrow n_2 \rightarrow \dots$, où nous avons noté $n_1 \rightarrow n_2$ pour $\langle n_1, n_2 \rangle \in f$. Un graphe pointé est *accessible* si tout nœud du graphe est relié au sommet par un chemin.

Un ensemble a est *représenté* par un graphe $\langle g, f, s \rangle$ lorsqu'il y a une application d de l'ensemble g dans l'univers des ensembles telle que

- $d(s) = a$;
- $\forall n \in g, d(n) = \{d(n') \mid n \rightarrow n'\}$.

Une telle application d est appelé *décoration* du graphe $\langle g, f, s \rangle$.

On dira qu'un graphe est *bien fondé* s'il est pointé, accessible et qu'il n'a pas de chemin infini. Il est à noter que ceci supprime en même temps les chemins circulaires. On montre facilement que dans ZFC,

- *Les ensembles sont représentés par des graphes bien fondés.*
- *Tout graphe bien fondé possède une décoration unique.*

Ceci permet d'introduire l'axiome d'anti-fondement :

AFA. *Tout graphe pointé et accessible possède une décoration unique.*

Notons que cet axiome comporte deux parties ; la première, disant que chaque graphe pointé et accessible possède une décoration, entraîne l'existence d'ensembles non bien fondés. La deuxième, affirmant l'unicité de la décoration, est relativement restrictive. Alors que la première nous dit par exemple l'existence d'un ensemble égal à son propre singleton.

$$x = \{x\},$$

la deuxième nous dit qu'il n'y a qu'un seul ensemble possédant cette forme.

Dans ce qui suit, nous utiliserons principalement deux résultats de cette théorie dont on trouvera une démonstration dans [1].

Le premier de ces résultats est peut-être ce qui rend cette théorie efficace dans le traitement de la circularité. C'est un résultat général concernant les points fixes d'opérateurs de classes. Nous n'en énonçons ici qu'une partie, celle qui sera concrètement utilisée dans notre propos. Un opérateur de classe Φ est dit *continu* si

- il est monotone : si $X \subseteq Y \Rightarrow \Phi(X) \subseteq \Phi(Y)$;
- il est *localement petit*¹ : si $x \in \Phi(X)$ alors il y a un ensemble $e \subseteq X$ tel que $x \in \Phi(e)$.

POINTS FIXES. *Soit Φ un opérateur de classe continu. La classe $U\{c \in V \mid c \subseteq \Phi(c)\}$ où V est la classe de tous les ensembles est le plus grand point fixe de l'opérateur Φ .*

Le deuxième résultat permet de "construire" ou de caractériser des ensembles à partir de systèmes d'équations. Ceci est un outil pratique pour manipuler des objets dont il est parfois difficile de cerner la forme autrement que par un graphe.

1. Nous traduisons ainsi l'expression "set-based".

LEMME DE SOLUTION. *On étend l'univers des ensembles en considérant une classe X d'atomes². Si pour chaque $x \in X$, a_x est un ensemble construit sur X , alors le système d'équations $x = a_x (x \in X)$ possède une solution unique dans l'univers des ensembles purs.*

2. Une classe universelle anti-fondée pour les croyances

Le nœud central de la construction de Mertens et de Zamir [15] ou de Mertens, Sorin et Zamir [16] concerne donc la théorie des jeux à information non complète. La modélisation souhaitée doit prendre en compte un ensemble de joueurs, un ensemble d'actions possibles pour chaque joueur, un type pour chaque joueur, ce type représente l'information privée dont il dispose avant le début du jeu, et une fonction de probabilité pour chaque joueur, cette fonction indiquant la probabilité subjective que chaque joueur assigne aux types possibles pour les autres joueurs, en supposant a priori que chaque joueur soit certain de son propre type.

Nous avons vu dans l'introduction que la situation peut devenir extrêmement complexe. Chaque joueur a une probabilité subjective sur le type des autres joueurs. Il peut être important, suivant le jeu, que chaque joueur puisse évaluer cette probabilité subjective des autres joueurs quant à son propre type, et donc puisse assigner une nouvelle probabilité subjective à l'ensemble des probabilités que chaque autre joueur peut avoir assigné à son type. A nouveau, il peut être important de pouvoir évaluer ces nouvelles probabilités ainsi définies par d'autres probabilités subjectives, etc. Ceci nous mène à une hiérarchie infinie de probabilités subjectives.

Partant de l'ensemble des agents $A = \{1, \dots, n\}$, d'un espace compact (séparé) de paramètres "objectifs" Θ et de ces hiérarchies infinies de croyances, Mertens et Zamir ont proposé, par exemple dans [15], la construction d'espaces Y et T tels que

$Y \simeq \Theta \times T^n$ et $T \simeq \Delta(\Theta \times T^{n-1})$, où $\simeq$ signifie "est homéomorphe à" et $\Delta(X)$ est l'espace des probabilités boréliennes sur X , muni de la topologie $*$ faible. L'espace Y est appelé espace universel des croyances. Mertens et Zamir montrent ainsi qu'il existe un objet mathématique permettant de modéliser cette situation et où les hiérarchie infinies se ramènent à la circularité manifestée par cet objet.

Les éléments de l'univers Y sont de la forme $\langle \theta, t_1, \dots, t_n \rangle$ avec $t_a \in T$ ($a = 1, \dots, n$). On notera h_0 l'homéomorphisme de T sur $\Delta(\Theta \times T^{n-1})$ et $T_1, \dots, T_n$ n copies de T . Quel que soit $t_a \in T_a$, $h_0(t_a)$ est une probabilité sur $K \times T^{n-1}$. Cette probabilité se prolonge trivialement en une probabilité $h(t_a)$ sur $K \times T^n$ de support dans $K \times T^{n-1} \times \{t_a\} \times T^{n-a}$.

2. Nous n'avons pas présenté ici l'extension de ZFC-AFA lorsque l'on travaille avec des atomes. Une telle extension n'est pas obligatoire mais permet de faciliter l'expression des choses. Une telle extension n'a donc ici qu'un rôle simplificateur. La démonstration du lemme de solution peut se faire à l'intérieur de l'univers des "ensembles purs" comme le montre l'argument utilisé par Aczel.

Un *sous-espace de croyance* B de Y est une partie mesurable de Y tel que $\forall (k, t_1, \dots, t_n) \in B, \forall a \leq n, h(t_a)(B) = 1$.

En regardant cet univers Y , on est frappé par la circularité que l'on y trouve : les éléments de cet espace sont constitués de probabilités sur ce même espace. C'est cette circularité qui nous a amenés à proposer une construction différente de l'univers des croyances. Cette construction, comme celle de Barwise pour construire des classes de situations et d'infons dans [4], se fait dans le cadre de la théorie des ensembles anti-fondés ZFC-AFA.

Soit une classe C . On notera $\Psi(C)$ la classe de toutes les probabilités sur C , c'est-à-dire des applications $\pi : S \rightarrow [0, 1]$ où S est une σ -algèbre définie sur un ensemble quelconque $s \subseteq C$, $\pi(s) = 1$ et quelle que soit la suite dénombrable $(P_i)_{i \in I}$ d'éléments disjoints deux à deux de S , $\pi(\bigcup_{i \in I} P_i) = \sum_{i \in I} \pi(P_i)$. Nous appelons *support* d'une telle probabilité tout ensemble de probabilité 1.

Soit donc $A = \{1, \dots, n\}$ l'ensemble des agents et un ensemble non vide Θ . Pour $a = 1, \dots, n$, on pose $\Psi_a(C)$ la classe de toutes les probabilités π de $\Psi(\Theta \times C^n)$ telles qu'un support de π soit une partie de $\Theta \times C^{a-1} \times \{\pi\} \times C^{n-a}$. L'aspect circulaire de cette définition est clair. Dans un univers d'ensembles bien fondés, $\Psi_a(C)$ serait vide. On dira que les probabilités de Ψ_a sont *a-cohérentes*. L'intuition sous-jacente est que l'agent a est certain de ce qu'il croit³.

On définit à partir des Ψ_a l'opérateur Ψ_* :

$$\Psi_*(C) = \Theta \times \Psi_A(C),$$

où $\Psi_A(C)$ est l'ensemble de toutes les suites $\langle \pi_a \rangle_{a \in A} \in \prod_{1 \leq a \leq n} \Psi_a(C)$ où toutes les π_a sont définies sur la même σ -algèbre d'un même ensemble. Autrement dit ce sont les suites $\langle \pi_a \rangle_{a \in A}$ telles qu'il existe un ensemble $c \subseteq C$ et une σ -algèbre S sur c tels que $\forall a \in A, \pi_a$ est définie sur S .

On vérifie que cet opérateur est continu. Il est d'abord trivialement monotone. Ensuite, il faut vérifier que Ψ_* est localement petit, c'est-à-dire que si $\omega \in \Psi_*(C)$, il existe un ensemble $c \subseteq C$ tel que $\omega \in \Psi(c)$. Soit donc un tel ω qui est de la forme $\langle \theta, \pi_i \rangle_{i \in A}$. Il suffit de prendre l'ensemble c sur la σ -algèbre duquel sont définies les π_a pour avoir le résultat.

3. Ceci n'est pas du tout indispensable pour la construction elle-même. Nous n'avons considéré des probabilités *a-cohérentes* que pour correspondre entièrement à la construction de Mertens et Zamir, où chaque agent est certain de son propre type.

Soit Ω le plus grand point fixe de l'opérateur Ψ_* . On a donc

$$\Omega = \Theta \times \Psi_A(\Omega).$$

Nous appellerons Ω classe universelle des croyances.

Nous pouvons vérifier que, quel que soit Θ non vide, Ω est non vide. Comme $\Omega = \bigcup \{c \in V \mid c \subseteq \Psi_*(c)\}$, il suffit de montrer un tel c pour avoir le résultat. Soit $\theta \in \Theta$. On pose $\omega = \langle \theta, r_1, \dots, r_n \rangle$ avec $\forall a \leq n, r_a = \{\langle \emptyset, 0 \rangle, \langle \{\omega\}, 1 \rangle\}$. Par le lemme de solution, un tel ensemble ω existe (et est unique). En posant $c = \{\omega\}$, on voit que $c \subseteq \Psi_*(c)$.

Revenons aux espaces Y et T_a ($a \in A$) construits par Mertens et Zamir à partir de l'espace de paramètre Θ et soit un sous-espace de croyance B dans Y . Nous allons voir qu'un tel sous-espace B peut se plonger dans Ω , la classe universelle des croyances construite sur le même ensemble Θ .

On définit d'abord un langage modal qui pourra être interprété dans un tel sous-espace de croyance et qui permettra de capturer les informations souhaitées. On prend comme ensemble VP de variables propositionnelles l'ensemble des mesurables de Θ . On construit l'ensemble Φ des formules de ce langage en utilisant les connecteurs logiques usuels et des opérateurs modaux $B_{a,r}$ quels que soient l'agent a et $r \in [0, 1]$. On définit par induction sur la complexité des formules une relation de satisfaction $\models$:

- Si $p \in VP$ et $\alpha = \langle \theta, \dots \rangle$, alors $B \stackrel{\alpha}{\models} p \Leftrightarrow \theta \in p$;
- ...
- $B \stackrel{\alpha}{\models} B_{a,r} \phi \Leftrightarrow h(t_a)(\llbracket \phi \rrbracket) \geq r$ où $\alpha = \langle \theta, t_1, \dots, t_n \rangle$ et $\llbracket \phi \rrbracket = \{\beta \in B \mid B \stackrel{\beta}{\models} \phi\}$.

Ceci nous indique que $B_{a,r} \phi$ peut être interprété par *l'agent a croit ϕ avec une probabilité (un coefficient de certitude) au moins égale à r* .

Pour voir que cette définition a du sens, il faut vérifier que $\forall \phi \in \Phi$, $\llbracket \phi \rrbracket$ est un ensemble mesurable. Ceci se fait aisément par induction sur la complexité des formules en sachant que si p est un mesurable de Θ , alors $(p \times \prod_{a \in A} T_a) \cap B = \llbracket p \rrbracket$ est un mesurable de B et que si P est un mesurable de B , $r \in [0, 1]$ et $M_a = \{t_a \in T_a \mid h(t_a)(P) \geq r\}$, alors

$(\Theta \times \prod_{b < a} T_b \times M_a \times \prod_{b > a} T_b) \cap B$ est un mesurable de B .

Soit $\alpha, \beta \in B$. On dira que α et β contiennent la même information objective et épistémique ssi

$$\forall \phi \in \Phi \quad (B \stackrel{\alpha}{\models} \phi \Leftrightarrow B \stackrel{\beta}{\models} \phi)$$

On le notera $\alpha \equiv \beta$. Cette relation est trivialement une équivalence.

Exemple 1. Soit $n = 2$, $B = \{\alpha, \beta\}$

$$\begin{cases} \alpha \simeq \langle \theta, \langle 1/3, 2/3 \rangle, \langle 1, 0 \rangle \rangle \\ \beta \simeq \langle \theta, \langle 1/3, 2/3 \rangle, \langle 0, 1 \rangle \rangle \end{cases}$$

où, si $\alpha = \langle \theta, t_1, t_2 \rangle$, on peut interpréter par exemple $\langle 1, 0 \rangle$ comme $h(t_1)(\{\alpha\}) = 1$ et $h(t_2)(\{\beta\}) = 0$. On voit facilement que

$$\{\llbracket \phi \rrbracket \mid \phi \in \Phi\} = \{\emptyset, B\} \text{ et que } \alpha \equiv \beta.$$

Exemple 2. Soit $n = 2$, $B = \{\alpha^1, \alpha^2, \alpha^3, \alpha^4\}$,

$$\begin{cases} \alpha^1 \simeq \langle \theta_1, \langle 1/2, 1/2, 0, 0 \rangle, \langle 1/3, 0, 2/3, 0 \rangle \rangle \\ \alpha^2 \simeq \langle \theta_2, \langle 1/2, 1/2, 0, 0 \rangle, \langle 0, 1/3, 0, 2/3 \rangle \rangle \\ \alpha^3 \simeq \langle \theta_1, \langle 0, 0, 1/2, 1/2 \rangle, \langle 1/3, 0, 2/3, 0 \rangle \rangle \\ \alpha^4 \simeq \langle \theta_2, \langle 0, 0, 1/2, 1/2 \rangle, \langle 0, 1/3, 0, 2/3 \rangle \rangle \end{cases}$$

On a $VP = \{\theta, \{\theta_1\}, \{\theta_2\}, \{\theta_1, \theta_2\}\}$. On montre par induction sur la complexité des formules que $\{\llbracket \phi \rrbracket \mid \phi \text{ est une formule sur } VP\} = \{\emptyset, \{\alpha^1, \alpha^3\}, \{\alpha^2, \alpha^4\}, B\}$. Le cas des variables propositionnelles est évident. Le cas des formules $\neg\phi$, $\phi \wedge \psi$... l'est aussi puisque l'ensemble considéré est une algèbre. Le cas des formules $B_a, r\phi$ demande un certain travail. Si on note $\alpha^i = \langle \theta_i, t_a^i \rangle_{a=1,2}$ on voit que

- $\{\alpha^i \in B \mid h(t_1^i)\{\alpha^1, \alpha^3\} \geq r\} \in \{\emptyset, I\}$;
- $\{\alpha^i \in B \mid h(t_2^i)\{\alpha^1, \alpha^3\} \geq r\} \in \{\{\alpha^1, \alpha^3\}, I\}$;
- $\{\alpha^i \in B \mid h(t_1^i)\{\alpha^2, \alpha^4\} \geq r\} \in \{\emptyset, I\}$;
- $\{\alpha^i \in B \mid h(t_2^i)\{\alpha^2, \alpha^4\} \geq r\} \in \{\{\alpha^2, \alpha^4\}, I\}$

Moyennant l'hypothèse d'induction, et en voyant que dans le cas des événements $\emptyset$ et I c'est évident, nous obtenons le résultat attendu. On en déduit que $\alpha^1 \equiv \alpha^3$ et $\alpha^2 \equiv \alpha^4$.

Théorème 1. *Pour chaque sous-espace de croyance B de Y , on peut construire une application $*$: $B \rightarrow \Omega$: $\alpha \rightarrow \alpha_*$, telle que $\forall \alpha, \beta \in B$, $\alpha_* = \beta_* \Leftrightarrow \alpha \equiv \beta$.*

La preuve complète de ce théorème peut être trouvée dans [12]. Dans ce qui suit nous donnons la démarche générale de cette preuve que nous illustrerons sur les deux exemples précédents.

Soit donc B un sous-espace de croyance de Y , X un ensemble d'atomes de même cardinal que B et d une bijection entre B et X qui associe une inconnue x à chaque élément α de B . L'ensemble $V = \{ \llbracket \phi \rrbracket \mid \phi \in \Phi \}$ est une algèbre sur B contenue dans la σ -algèbre des boréliens sur B . La σ -algèbre F engendrée par V est donc contenue dans celle des boréliens.

Quel que soit $P \in F$, posons

$$P_X = \{ d(\alpha) \mid \alpha \in P \}.$$

Pour $t_a \in T_a$,

$$h(t_a) \subseteq \{ \langle P, r \rangle \mid P \subseteq \Theta \times \prod_{a \in A} T_a, r \in [0, 1] \}.$$

On pose alors

$$t_a^X = \{ \langle P_X, r \rangle \mid P \in F, \langle P, r \rangle \in h(t_a) \}.$$

Pour $\alpha = \langle \theta, t_1, \dots, t_n \rangle \in B$ et $x = d(\alpha)$, on définit $\alpha_x = \langle \theta, t_1^X, \dots, t_n^X \rangle$.
Le système

$$x = \alpha_x \quad (\alpha \in B \text{ et } x = d(\alpha))$$

admet une solution unique $(\alpha_*)_{\alpha \in B}$. Si on note $\alpha_* = \langle \theta; t_1^*, \dots, t_n^* \rangle$, on a

$$t_a^* = \{ \langle P_*, \tau \rangle \mid P \in F, \langle P, r \rangle \in h(t_a) \} \text{ où } P_* = \{ \alpha_* \mid \alpha \in P \}.$$

Ceci nous donne donc la construction de l'application $*$. La suite de la preuve consiste à montrer que, quel que soit α dans B , α_* est dans Ω et que $\alpha_* = \beta_*$ ssi $\alpha \equiv \beta$.

Regardons ce que cela donne pour les deux exemples précédents.

Exemple 1 (suite). On a

$$\begin{cases} x_1 = \langle \theta, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_2\}, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_2\}, 1 \rangle \} \rangle \\ x_2 = \langle \theta, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_2\}, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_2\}, 1 \rangle \} \rangle \end{cases}$$

comme système d'équations. L'unique solution est $\alpha_* = \beta_* = \omega$ avec

$$\omega = \langle \emptyset, \{ \langle \emptyset, 0 \rangle, \langle \{ \omega \}, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{ \omega \}, 1 \rangle \} \rangle$$

Exemple 2 (suite). On a 4 inconnues x_1, x_2, x_3, x_4 . $B_X = \{x_1, x_2, x_3, x_4\}$. Le système d'équations est

$$\begin{cases} x_1 = \langle \emptyset, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_3\}, 1/2 \rangle, \langle \{x_2, x_4\}, 1/2 \rangle, \langle B_X, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_3\}, 1 \rangle, \langle \{x_2, x_4\}, 0 \rangle, \langle B_X, 1 \rangle \} \rangle \\ x_2 = \langle \emptyset, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_3\}, 1/2 \rangle, \langle \{x_2, x_4\}, 1/2 \rangle, \langle B_X, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_3\}, 0 \rangle, \langle \{x_2, x_4\}, 1 \rangle, \langle B_X, 1 \rangle \} \rangle \\ x_3 = \langle \emptyset, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_3\}, 1/2 \rangle, \langle \{x_2, x_4\}, 1/2 \rangle, \langle B_X, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_3\}, 1 \rangle, \langle \{x_2, x_4\}, 0 \rangle, \langle B_X, 1 \rangle \} \rangle \\ x_4 = \langle \emptyset, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_3\}, 1/2 \rangle, \langle \{x_2, x_4\}, 1/2 \rangle, \langle B_X, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{x_1, x_3\}, 0 \rangle, \langle \{x_2, x_4\}, 1 \rangle, \langle B_X, 1 \rangle \} \rangle \end{cases}$$

L'unique solution nous donne $B_* = \{\alpha_*^1, \alpha_*^2\}$ ($\alpha_*^3 = \alpha_*^1$ et $\alpha_*^4 = \alpha_*^2$) où

$$\begin{cases} \alpha_*^1 = \langle \emptyset, \{ \langle \emptyset, 0 \rangle, \langle \{ \alpha_*^1 \}, 1/2 \rangle, \langle \{ \alpha_*^2 \}, 1/2 \rangle, \langle B_*, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{ \alpha_*^1 \}, 1 \rangle, \langle \{ \alpha_*^2 \}, 0 \rangle, \langle B_*, 1 \rangle \} \rangle \\ \alpha_*^2 = \langle \emptyset, \{ \langle \emptyset, 0 \rangle, \langle \{ \alpha_*^1 \}, 1/2 \rangle, \langle \{ \alpha_*^2 \}, 1/2 \rangle, \langle B_*, 1 \rangle \}, \{ \langle \emptyset, 0 \rangle, \langle \{ \alpha_*^1 \}, 0 \rangle, \langle \{ \alpha_*^2 \}, 1 \rangle, \langle B_*, 1 \rangle \} \rangle \end{cases}$$

3. Logique modale anti-fondée

Le deuxième type d'utilisation de la théorie des ensembles anti-fondés que nous présentons est la construction d'un langage modal circulaire. Nous en donnons une présentation assez générale⁴.

Soit un ensemble fini A . L'ensemble des variables propositionnelles, noté VP , est un ensemble d'atomes. On considère également des atomes distincts des variables propositionnelles : $\neg$, $\wedge$ et K_a ($a \in A$).

La classe des formules⁵, notée F , sera la plus grande classe telle que si ϕ est dans F , alors ϕ est soit une variable propositionnelle soit un ensemble de la forme⁶ :

- $\langle \neg, \phi \rangle$ avec ϕ dans F ;
- $\langle \wedge, \Sigma \rangle$ avec Σ un ensemble fini de formules de F ;
- $\langle K_a, \phi \rangle$ avec ϕ dans F .

4. Nous ne donnons aucune démonstration. Celles-ci peuvent être trouvées dans [11].

5. Le type de définition que nous utilisons est appelé co-inductive. Dans une définition inductive nous cherchons la *plus petite* classe telle que si un objet vérifie certaines conditions, cet objet se trouve dans cette classe. Dans une définition co-inductive, on considère la *plus grande* classe telle que si un objet se trouve dans cette classe alors il vérifie certaines conditions.

6. Nous avons repris à [9] l'idée de considérer les formules comme des couples.

La justification d'une telle définition se fait en considérant le plus grand point fixe d'un opérateur de classe continu. Nous pouvons définir également la classe Bf-F des formules bien fondées comme étant la plus petite classe satisfaisant les conditions réciproques de celles indiquées (définition inductive). On peut voir que Bf-F est le plus petit point fixe de l'opérateur dont F est le plus grand point fixe. On vérifie facilement que F est un ensemble.

On introduit les abréviations suivantes :

• $\neg\phi \equiv \langle \neg, \phi \rangle$;	• $\phi \vee \psi \equiv \bigvee \{ \phi, \psi \}$;
• $\bigwedge \Sigma \equiv \langle \bigwedge, \Sigma \rangle$;	• $\phi \rightarrow \psi \equiv (\neg\phi) \vee \psi$;
• $K_a\phi \equiv \langle K_a, \phi \rangle$;	• $\phi \leftrightarrow \psi \equiv (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$;
• $\Sigma^- \equiv \{ \neg\phi \mid \phi \in \Sigma \}$;	• $\top \equiv \bigwedge \emptyset$;
• $\bigvee \Sigma \equiv \neg \bigwedge \Sigma^-$;	• $\perp \equiv \bigvee \emptyset$;
• $\phi \wedge \psi \equiv \bigwedge \{ \phi, \psi \}$;	• $\Box\phi \equiv \langle \bigwedge, \{ K_a\phi \mid a \in A \} \rangle^a$

$^a K_a\phi$ sera évidemment interprétée par *a sait (ou croit) ϕ* ; $\Box\phi$ sera donc interprétée par *tous les agents savent (ou croient) ϕ* .

Exemple. On considère l'équation $\psi = \neg\psi$. Cette équation admet une et une seule solution. On notera cette formule entre crochets $[\psi = \neg\psi]$, pour signifier qu'elle est solution de cette équation. De la même façon, à partir d'une formule quelconque ϕ_0 , on peut construire la formule $[\phi = \Box(\phi \wedge \phi_0)]$ dont nous reparlerons plus loin.

Une *interprétation anti-fondée de Scott*⁷ M est une suite $\langle I, V, U_a \rangle_{a \in A}$ où I est un ensemble de mondes possibles, V une valuation, c'est-à-dire une application de $I \times VP$ dans 2 et pour chaque $a \in A$, U_a un système de voisinages défini sur I , c'est-à-dire une application de I dans $P(P(I))$. Étant donné une interprétation M , on définit les relations $M \models^+ \phi$ et $M \models^- \phi$ comme les plus grandes parties de $I \times F$ satisfaisant les conditions suivantes :

- si $p \in VP$ alors $M \models^+ p \Rightarrow V(\alpha, p) = 1$;
- si $p \in VP$ alors $M \models^- p \Rightarrow V(\alpha, p) = 0$;
- $M \models^+ \neg\phi \Rightarrow M \models^- \phi$;

7. Nous parlerons simplement d'interprétations.

- $M \models^{\alpha} \neg \phi \Rightarrow M \models^{+ \alpha} \phi$;
- $M \models^{+ \alpha} \wedge \Sigma \Rightarrow \forall \phi \in \Sigma, M \models^{+ \alpha} \phi$;
- $M \models^{- \alpha} \wedge \Sigma \Rightarrow \exists \phi \in \Sigma, M \models^{- \alpha} \phi$;
- $M \models^{+ \alpha} K_a \phi \Rightarrow \llbracket \phi \rrbracket^+ \in U_a(\alpha)$ où $\llbracket \phi \rrbracket^+ = \{\beta \in I \mid M \models^{+ \beta} \phi\}$;
- $M \models^{- \alpha} K_a \phi \Rightarrow \llbracket \phi \rrbracket^- \in U_a^-(\alpha)$ où
 - $\llbracket \phi \rrbracket^- = \{\beta \in I \mid M \models^{- \beta} \phi\}$
 - $U_a^-(\alpha) = \{P \subseteq I \mid I \setminus P \notin U_a(\alpha)\}$

Il est à remarquer que dans la condition définissant ces deux relations⁸, on a des implications et non des équivalences. Comme nous considérons les plus grandes parties satisfaisant les conditions indiquées, on peut en fait remplacer ces implications par des équivalences.

Voici comment cela se traduit pour les formules $\forall \Sigma$ et $\phi \rightarrow \psi$:

$$\begin{array}{lll}
 M \models^{+ \alpha} \forall \Sigma & \text{ssi} & \exists \phi \in \Sigma, M \models^{+ \alpha} \phi. \\
 M \models^{- \alpha} \forall \Sigma & \text{ssi} & \forall \phi \in \Sigma, M \models^{- \alpha} \phi. \\
 M \models^{+ \alpha} \phi \rightarrow \psi & \text{ssi} & M \models^{- \alpha} \phi \text{ ou } M \models^{+ \alpha} \psi. \\
 M \models^{- \alpha} \phi \rightarrow \psi & \text{ssi} & M \models^{+ \alpha} \phi \text{ et } M \models^{- \alpha} \psi.
 \end{array}$$

Nous avons ainsi défini une relation de satisfaction "double" : $M \models^{+ \alpha} \phi$ pourra se lire ϕ est vraie dans le monde α de l'interprétation M et $M \models^{- \alpha} \phi$ se lira ϕ est fausse dans le monde α de l'interprétation M . Il y a des formules qui sont à la fois vraie et fausse :

Lemme 2. Soit M une interprétation et $\alpha \in I$. Il y a au moins une formule ϕ de F telle que $M \models^{+ \alpha} \phi$ et $M \models^{- \alpha} \phi$.

Preuve. Il suffit de prendre $[\psi = \neg \psi]$ $\square$

8. La justification complète d'une telle définition de même que celle de la classe des formules, se fait au moyen d'un opérateur de classe continu.

On appellera une formule ϕ telle que $M \models^{\alpha} \phi$ et $M \models^{\alpha} \neg \phi$ une formule $\langle M, \alpha \rangle$ -ambiguë. On dira que ϕ est ambiguë s'il y a une interprétation M et $\alpha \in I$ tel que ϕ soit $\langle M, \alpha \rangle$ -ambiguë.

On dira que ϕ est *vraiment vraie* dans le monde α de M , ce que l'on notera $M \models^{\alpha} \phi$, si $M \models^{\alpha} \phi$ et $M \not\models^{\alpha} \neg \phi$. On dira que ϕ est *vraiment vraie* dans l'interprétation M , ou encore que $M \models \phi$, si $\forall \alpha \in I, M \models^{\alpha} \phi$. On dira que ϕ est *valide* dans la classe des interprétations anti-fondées de Scott si elle est vraie dans toutes les interprétations de cette classe, et on le notera $\models \phi$.

Exemple. Soit la formule $[\psi = \psi \rightarrow p] = [\psi = (\neg \psi) \vee p]$ et une interprétation M . Il y a deux cas possibles :

• $V(\alpha, p) = 1$:

- $M \models^{\alpha} \phi \Leftrightarrow M \models^{\alpha} \neg \phi$ ou $M \models^{\alpha} p$, ce qui est vrai.
- $M \models^{\alpha} \neg \phi \Leftrightarrow M \models^{\alpha} \neg \neg \phi$ ou $M \models^{\alpha} p$, ce qui est faux.

On a donc $M \models^{\alpha} \phi$.

• $V(\alpha, p) = 0$:

- $M \models^{\alpha} \phi \Leftrightarrow M \models^{\alpha} \neg \phi$ ou $M \models^{\alpha} p$. Ce sera le cas si $M \models^{\alpha} \neg \phi$.
- $M \models^{\alpha} \neg \phi \Leftrightarrow M \models^{\alpha} \neg \neg \phi$ ou $M \models^{\alpha} p$. Ce sera le cas si $M \models^{\alpha} \phi$.

On peut ajouter $\langle \alpha, \phi \rangle$ à $M \models^{\alpha} \phi$ et à $M \models^{\alpha} \neg \phi$; les conditions sur ces relations sont vérifiées. C'est donc que $\langle \alpha, \phi \rangle$ s'y trouvait déjà. Donc ϕ est $\langle M, \alpha \rangle$ -ambiguë.

La formule $[\psi = \psi \rightarrow p]$ est donc un exemple de formule vraiment vraie dans certains mondes de certaines interprétations et ambiguë dans d'autres.

On peut voir immédiatement que la formule $p \rightarrow [\psi = \psi \rightarrow p]$ est valide. En effet $M \models^{\alpha} p$ ou $M \models^{\alpha} [\psi = \psi \rightarrow p]$ par l'exemple précédent. Toujours par le même exemple, on n'a pas $M \models^{\alpha} p$ et $M \models^{\alpha} \neg [\psi = \psi \rightarrow p]$. Ceci montre bien l'existence de formules non bien fondées valides (et donc non ambiguës).

Théorème 3. Soit une interprétation M et une formule ϕ

$$M \models^{\alpha} \phi \text{ ou } M \models^{\alpha} \neg \phi$$

Corollaire 4.

$$M \models^{\alpha} \phi \Leftrightarrow M \not\models^{\alpha} \neg \phi$$

$$M \models^{\alpha} \phi \rightarrow \psi \Leftrightarrow (M \models^{\alpha} \phi \Rightarrow M \models^{\alpha} \psi).$$

Lemme 5. Soit $\phi \in Bf\text{-}F$ et une interprétation M .

$$M \models^{\alpha} \phi \Leftrightarrow M \models^{+} \phi.$$

Nous pouvons définir un calcul de séquents à la Gentzen ([7]) pour cette logique modale anti-fondée. Un séquent est une expression $\Gamma \vdash \Delta$ où Γ et Δ sont des ensembles (finis ou infinis) de formules. Dans ces séquents, on notera Γ, Σ pour $\Gamma \cup \Sigma$ et $\Gamma, \emptyset$ pour $\Gamma \cup \{\emptyset\}$.

Les séquents axiomatiques sont les séquents $\Gamma \vdash \Delta$ tels que $\Gamma \cap \Delta \cap VP \neq \emptyset$.

Les séquents théorèmes seront les séquents obtenus à partir de séquents axiomatiques et d'un nombre fini d'applications des règles suivantes :

$$\wedge_G \frac{\Gamma, \Sigma \vdash \Delta}{\Gamma, \wedge \Sigma \vdash \Delta}$$

$$\wedge_D \frac{\forall \phi \in \Sigma (\Gamma \vdash \Delta, \phi)}{\Gamma \vdash \Delta, \wedge \Sigma}$$

où Σ est un ensemble fini de formules ;

$$\neg_G \frac{\Gamma \vdash \Delta, \Sigma}{\Gamma, \Sigma^- \vdash \Delta}$$

$$\neg_D \frac{\Gamma, \Sigma \vdash \Delta}{\Gamma \vdash \Delta, \Sigma^-}$$

$$RM \frac{\phi \vdash \psi}{K_a \phi \vdash K_a \psi}$$

$$RA \frac{\Gamma \vdash \Delta}{\Gamma, \Sigma_1 \vdash \Delta, \Sigma_2}$$

RA est appelée règle d'affaiblissement ou d'atténuation.

Une preuve d'un séquent théorème est donc un arbre fini dont la racine est le séquent en question, les sommets de l'arbre sont des séquents axiomatiques et le passage de nœuds à un autre nœud se fait au moyen d'une de ces règles. On appellera longueur de la preuve la longueur de la plus grande branche de l'arbre au dessus du segment théorème dont il est question. La preuve des séquents axiomatiques est donc une preuve de longueur nulle.

Soit une interprétation $M = \langle I, V, U_{\frac{a}{M}} \rangle_{a \in A}$. On dira que le séquent $\Gamma \vdash \Delta$ est vrai dans l'interprétation M ou encore $\Gamma \models \Delta$ ssi quel que soit $\alpha \in I$,

$$(\forall \phi \in \Gamma, M \models^{+} \phi) \Rightarrow (\exists \psi \in \Delta, M \models^{\alpha} \psi).$$

On notera $\Gamma \models \Delta$ si le séquent $\Gamma \vdash \Delta$ est vrai dans toutes les interprétations anti-fondées. On trouvera dans [11] la preuve de l'adéquation et de la complétude du calcul de séquents pour la classe des interprétations anti-fondées de Scott :

Théorème 6.

$\Gamma \vdash \Delta$ est un séquent théorème ssi $\Gamma \models \Delta$.

4. Logique modale anti-fondée et connaissance commune

On peut utiliser un tel langage pour analyser la question de la connaissance commune. Le seul point que nous aborderons ici est la relation complète que nous pouvons établir entre la modélisation de la connaissance commune faite par Barwise dans le cadre des situations anti-fondées et l'analyse de cette connaissance commune en logique modale épistémique.

APPROCHE DE BARWISE ([4]). Utilisant également ZFC-AFA comme métathéorie, Barwise construit la classe *INFON* des infons comme la plus grande classe telle que si θ est dans *INFON* alors θ est soit de la forme

$$\bullet \langle H, a, c \rangle \text{ où } a \in A \text{ et } c \in C ;$$

soit la forme

$$\bullet \langle K, a, s \rangle \text{ où } a \in A \text{ et } s \subseteq \text{INFON}, s \text{ étant un ensemble.}$$

Les infons du type $\langle H, a, c \rangle$ sont appelés *infons de base*.

La classe des situations, *SIT*, peut alors être définie comme suit :

$$s \in \text{SIT} \text{ ssi } s \subseteq \text{INFON} \text{ et } s \text{ est un ensemble.}$$

Les infons de base expriment un fait objectif. Les infons $\langle K, a, s \rangle$ expriment que l'agent a voit ou connaît la situation s .

Dans ce contexte, la situation de connaissance commune d'une situation s_0 peut se représenter par une situation circulaire :

$$s = \{ \langle K, a, s \cup s_0 \rangle \mid a \in A \}.$$

Nous noterons une telle situation représentant la connaissance commune de la situation s_0 par $s_c(s_0)$.

Dans ce contexte, Barwise définit une relation $\models$ de satisfaction entre situations et infons. On pourra dire que dans la situation s , il y a connaissance commune de la situation s_0 ssi

$$s \models s_c(s_0),$$

c'est-à-dire que

$$\forall \theta \in s_c(s_0), s \models \theta.$$

APPROCHE EN LOGIQUE MODALE BIEN FONDÉE ([14]). Utilisant un langage modal bien fondé ayant des opérateurs modaux K_a pour la connaissance individuelle, $\Box$ pour la connaissance partagée par tous les agents et $\Box$ pour la connaissance commune, on modélise la connaissance commune en sémantique des voisinages en construisant, dans une interprétation M , à partir des systèmes de voisinages individuels U_a fermés par sur-ensembles et exprimant les propositions que connaissent individuellement les agents, les systèmes de voisinages $U_\Box$ et $U_\Box$ reprenant respectivement les propositions que tous les agents connaissent et celles qui sont connaissance commune. Dans la définition de la relation de satisfaction, nous aurons bien sûr les étapes suivantes :

$$\bullet M \models^\alpha K_a \phi \Leftrightarrow \llbracket \phi \rrbracket \in U_a(\alpha) ;$$

$$\bullet M \models^\alpha \Box \phi \Leftrightarrow \llbracket \phi \rrbracket \in U_\Box(\alpha) ;$$

$$\bullet M \models^\alpha \Box \phi \Leftrightarrow \llbracket \phi \rrbracket \in U_\Box(\alpha) ;$$

Dans ce contexte sémantique, la connaissance commune n'est pas définie explicitement comme point fixe mais par une itération infinie de connaissances croisées des agents.

RELATION ENTRE LES DEUX APPROCHES DANS LE CONTEXTE D'UN LANGAGE MODAL ANTI-FONDÉ. Barwise ne travaillant qu'avec des infons positifs non disjonctifs, nous nous restreignons pour la relation que nous voulons faire entre les deux approches à un langage modal anti-fondé positif non disjonctif. Nous pouvons considérer la formule

$$[\phi = \Box (\phi \wedge \phi_0)]$$

qui permet d'exprimer, directement comme point fixe, la connaissance commune de la formule ϕ_0 .

Dans le théorème qui suit, on suppose avoir construit $U_\Box$ exactement de la même manière que dans le contexte bien fondé de [14]. On en trouvera une preuve dans [11].

Théorème 7.1. *Sémantiquement, la formule $[\phi = \Box (\phi \wedge \phi_0)]$ représente bien la connaissance commune de la manière dont elle a été définie dans un contexte modal bien fondé :*

$$M \models^\alpha [\phi = \Box (\phi \wedge \phi_0)] \Leftrightarrow \llbracket \phi \rrbracket \in U_\Box(\alpha).$$

2. Nous pouvons faire correspondre à chaque situation une formule anti-fondée de telle manière que

- les situations de connaissance commune correspondent aux formules circulaires définissant la connaissance commune par point fixe.
- les relations de satisfaction définies dans chacun des contextes correspondent entièrement ;

Ceci revient à dire que l'on peut construire

(a) une application

$$\phi : SIT \rightarrow F : s \rightarrow \phi(s) ;$$

(b) à partir d'un ensemble de situations S possédant une stabilité pour les sous-situations, une interprétation de Scott M telle que l'ensemble des mondes possibles soit S ,

de telle manière que

- $\phi(s_c(s_0)) = [\psi = \Box (\psi \wedge \phi(s_0))]$;
- $\forall s, s' \in S, s \models s' \Leftrightarrow M \stackrel{s}{\models} \phi(s')$.

5. Conclusion et perspectives

Nous avons utilisé la théorie ZFC-AFA dans deux contextes différents. Dans le premier, il s'agit de modéliser un espace dans lequel se manifeste une circularité. Dans le deuxième, il s'agit de permettre l'écriture d'une circularité au niveau même du langage que l'on utilise. Nous devons mettre en évidence les limites d'une telle approche. La construction de Mertens et Zamir se fait dans la théorie "classique" des ensembles. Nous proposons une autre construction qui n'apporte pas de grandes nouveautés. Ce qui nous paraît important dans cette construction anti-fondée est sa simplicité conceptuelle. Nous n'avons pas besoin d'hypothèses topologiques pour construire la classe universelle des croyances et la circularité existant au niveau de cette classe est présente directement dans sa construction. Il nous faut encore remarquer que, même si l'univers de Mertens et Zamir se plonge dans la classe anti-fondée des croyances, une telle modélisation n'est pas une généralisation de la construction de Mertens et Zamir. Elle ne permet en effet pas de retrouver le théorème de Mertens et Zamir comme cas particulier.

De la même manière, pour étudier la connaissance commune en logique modale, il n'est pas nécessaire de disposer d'un langage anti-fondé. Nous pouvons même établir une relation entre la vision par point fixe de Barwise et l'approche modale de la connaissance commune à l'intérieur d'un langage modal bien fondé ([13]), même si cette relation est moins complète que celle que nous avons esquissée ici. Ce qu'apporte principalement ce langage anti-fondé est à nouveau l'apparition immédiate de la circularité présente de la notion de connaissance commune au niveau de son écriture.

Dans le cadre de l'étude de la connaissance commune, un élément nouveau serait de poursuivre l'analyse de la connaissance commune vue comme point fixe sans se restreindre aux formules positives non disjonctives. Nous avons commencé une telle analyse dans [11] et l'on voit qu'il y a encore des éléments à découvrir dans une telle perspective, notamment sur la relation entre connaissance commune comme point fixe et connaissance commune comme itération infinie de connaissances croisées.

Pour terminer, nous voudrions souligner un aspect très prometteur des résultats que nous avons présentés ici : il nous semble qu'en mettant ensemble ces deux utilisations de la théorie des ensembles anti-fondés, il devrait être possible de construire une sémantique et un langage permettant d'avoir un cadre général unifiant diverses modélisations de la connaissance, à savoir la modélisation probabiliste en théorie des jeux, la sémantique des situations telle que présentée ici et les interprétations de Scott en logique modale.

Enfin, il nous paraît important de noter que l'intérêt majeur de la théorie ZFC-AFA réside dans l'existence d'un théorème de point fixe extrêmement général. Ce résultat permet de manipuler les expressions circulaires avec une facilité déconcertante lorsque l'on est parvenu à dépasser des objections conceptuelles pouvant apparaître en ce qui concerne l'existence d'ensembles non bien fondés. Cette facilité se révèle alors liée à une certaine esthétique qui n'est pas non plus pour déplaire.

Bibliographie

- [1] Aczel P., *Non-well-founded sets*, Center for de Study of Language and Information, Stanford, 1988.
- [2] Aumann R., *Agreeing to Disagree*, The Annals of Statistics 4, 1236-1239, 1976.
- [3] Bacharach M., *Some Extensions of a Claim of Aumann in an Axiomatic Model of Knowledge*, Journal of Economic Theory 37, 167-190, 1985.
- [4] Barwise J., *The situation in Logic*, Center for de Study of Language and Information, Stanford, 1989.
- [5] Barwise J., Etchemendy J., *The liar, an essay on truth and circularity*, Oxford University Press, 1987.
- [6] Forti M., Honsell F., *Set Theory with Free Construction Principles*, Annali Scuola Normale Sup. di Pisa, Classe di Sc, Series IV, X(3) : 493-522, 1983.
- [7] Gentzen R., *Untersuchungen über das logische Schliessen*, Mathematische Zeitschrift, 39 : 176-210, 405-431, 1935. Traduction française par Feys R. et Ladrière J., *Recherches sur la déduction logique*, Presses Universitaires de France, 1955.
- [8] Halpern J.Y., *Knowledge and Common Knowledge in a Distributed Environment*, mimeo, 1988 (version étendue et revue d'un article du même nom paru dans Proceedings of the 3rd ACM Conference on Principles of Distributed Computing, 1984).
- [9] Keisler H.J., *Model Theory for Infinitary Logic*, North-Holland Publishing Company, Amsterdam, 1971.
- [10] Lewis D.K., *Convention. A Philosophical Study*, Harvard University Press, 1969.

- [11] Lismont L., *La Connaissance Commune : approches modale, ensembliste et probabiliste*, dissertation doctorale, Département de Mathématique, Université Catholique de Louvain, 1992.
- [12] Lismont L., *Une classe universelle anti-fondée pour les croyances*, Compte Rendus de l'Académie des Sciences de Paris, t. 315, Série I, p. 117-1120, 1992.
- [13] Lismont L., *Common Knowledge : Relating Anti-Founded Situation Semantics to Modal Logic Neighbourhood Semantics*, Rapport 221, Département de Mathématique, Université Catholique de Louvain, 1992.
- [14] Lismont L., *La connaissance commune en logique modale*, Mathematical Logic Quarterly (formerly Zeitschrift für Mathematische Logik und Grundlagen der Mathematik), vol. 39, p. 115-130, 1993.
- [15] Mertens J-F., Zamir S., *Formulation of bayesian analysis for games with incomplete information*, International Journal of Game Theory, Vol. 14, Issue 1, p. 1029, 1985.
- [16] Mertens J-F., Sorin S., Zamir S., *Repeated Games*, manuscrit à paraître, 1990.
- [17] Milgrom P., *an axiomatic characterization of common knowledge*, Econometrica, vol. 49-1, 219-222, 1981.
- [18] Vassilakis S., Zamir S., *Common Beliefs and Common Knowledge*, University of Pittsburgh, Working Paper N° 259, 1990.
- [19] Walliser B., *Logique épistémique et théorie des jeux*, Revu Économique, Paris, 1991.

Luc LISMONT

Les résultats présentés dans ce texte ont été obtenus alors que l'auteur était assistant à l'Université Catholique de Louvain. L'auteur remercie tout particulièrement Thierry Lucas pour son aide dans l'élaboration de ces résultats et ses commentaires à propos d'une première version de ce texte.

Chapitre 10

Problèmes de logique de la connaissance

1. Introduction

La logique de la connaissance et de la croyance inaugurée par Hintikka en 1962 a été présentée par son inventeur comme une version idéalisée de la connaissance et de la croyance. Cette idéalisation se manifeste dans l'omniscience logique imputée à l'agent dont on représente la connaissance ou la croyance. La logique est construite de manière telle que l'agent - c'est-à-dire le sujet connaissant - se voit attribuer la connaissance des conséquences de ce qu'il connaît. La logique de la croyance qui est devenue classique utilise le système modal bien connu KD45 et lui adjoint une sémantique des mondes possibles qui a été développée par Kripke en 1963. Levesque a conservé la connaissance dans sa version idéalisée, mais a introduit, à côté, un concept plus réaliste de connaissance (distinction entre connaissance implicite et explicite). La sémantique de cette logique fait appel à la notion de situation partielle et de situation incohérente. Le concept de situation incohérente fait problème. L'opinion reçue est que l'incohérence est dans notre pensée ou dans notre discours, mais pas dans les choses. Pour atteindre l'objectif visé par Levesque sans recourir à des situations incohérentes, plusieurs auteurs ont fait appel à une notion nouvelle, la notion exprimée en anglais par le terme de "*awareness*" qui peut prendre plusieurs sens et que nous traduirons momentanément par "prise de conscience". La connaissance (ou la croyance) explicite serait dans cette optique une combinaison de la connaissance implicite et de la prise de conscience. Notre traduction est provisoire. On souhaite, en effet, pouvoir appliquer l'opposition entre connaissance implicite ou explicite en intelligence artificielle où le mot "conscience" n'est qu'une métaphore. C'est à l'étude de cette attitude propositionnelle et au problème qu'elle a

pour mission de résoudre que cet article est consacré. E. Thijssen a souligné l'importance d'une "logique de la prise de conscience qui soit à la fois correcte et complète, mais aussi descriptivement adéquate à la croyance humaine": "En dehors de son intérêt académique, cette entreprise est importante aussi pour les systèmes de communication de l'avenir auxquels il sera demandé d'agir comme s'ils comprenaient comment les êtres humains pensent (Thijssen, 1991, 2)".

Nous commencerons par dresser l'inventaire des formes de l'omniscience logique qu'une logique de la prise de conscience doit éviter pour être descriptivement adéquate et nous les commenterons succinctement. Nous rappellerons ensuite les deux solutions que Halpern et Fagin ont proposées pour éviter les "paradoxes" de l'omniscience logique. Nous évoquerons les observations et les critiques qu'elles ont suscitées. Nous exposerons enfin les résultats de notre collaborateur E. Gillet, présentés récemment dans une dissertation doctorale encore inédite.

2. Les formes du problème de l'omniscience logique

2.1. La fermeture sur l'implication matérielle

Dans le relevé de W. van der Hoek et J.J. Ch. Meyer, la première forme de l'omniscience logique mentionnée est la fermeture de la croyance sur l'implication matérielle. De "A croit que p" et de "A croit que p implique q", découle logiquement "A croit que q". Cette forme de l'omniscience provient de l'adoption de l'axiome K (l'axiome de normalité commun à toutes les logiques modales normales). Il est clair que les êtres humains ne vérifient pas cet axiome. Aucun homme ne tire toutes les conséquences logiques de ce qu'il sait ou croit.

Dretske avance un argument plaisant pour justifier l'abandon de cet axiome. Considérons le cas suivant : je sais que les animaux du zoo sont des zèbres. Je sais que si les animaux du zoo sont des zèbres, ils ne sont pas des mules habilement déguisées en zèbres par les employés du zoo. De cela, il ne découle pas que je sache que les animaux du zoo ne sont pas des mules habilement déguisées en zèbres par les employés du zoo. L'idée de Dretske, c'est que pour savoir que les animaux du zoo sont des zèbres avec un degré raisonnable de certitude, il me suffit d'inspecter ces animaux à la lumière du jour alors que pour savoir que ce ne sont pas des mules habilement déguisées, je dois avoir procédé à une enquête sur la moralité des employés du zoo (Dretske 1970). On voit ici que Dretske comprend par connaissance la croyance vraie et justifiée et que pour lui deux connaissances qui ont même objet, mais qui reposent sur des justifications différentes, diffèrent. Le fait d'introduire dans le concept de connaissance non seulement celui de croyance vraie, mais aussi celui de croyance justifiée, n'est pas une raison suffisante pour motiver le rejet de l'axiome qui affirme la fermeture de la connaissance (ou de la croyance) sur l'implication matérielle. F. Voorbraak a d'ailleurs présenté un système axiomatique de la connaissance justifiée qui concède cette fermeture sur l'implication matérielle (F. Voorbraak, 1992).

2.2. La fermeture sur l'implication valide

$\models$ Rome est vaste $\supset$ Rome est vaste $\vee$ Washington est une capitale

$\models$ César croyait que Rome est vaste $\supset$ César croyait que Rome est vaste $\vee$ que Washington est une capitale

2.3. La fermeture sur la règle de nécessité épistémique

$\vdash$ Si p alors si q alors p

A croit que si p alors si q alors p

Ces deux fermetures sont évidemment contre-intuitives. Il ne suffit évidemment pas qu'une proposition soit un théorème pour que nous la connaissions et nous ne connaissons certainement pas toutes les conséquences logiques des théorèmes ou des axiomes que nous connaissons. La fermeture mentionnée en 2.2 dans l'inventaire que nous utilisons est une règle dérivée qui découle de l'axiome de normalité et de la règle de nécessité qu'il est bien difficile d'abandonner si l'on veut formaliser la logique épistémique dans le cadre d'un système de logique modale. La solution sera de limiter ces règles par une condition de "prise de conscience".

2.4. La fermeture sur la conjonction

Cette fermeture n'est pas réaliste. De ce qu'un agent A croit p et croit q, il ne découle pas qu'il croie p&q. Comme le note Hintikka, "Un génie pourrait être prêt à voir tout de suite les conséquences éloignées de ce qu'il croit, alors qu'un autre homme peut parfois être presque littéralement incapable de mettre ensemble deux et deux (Hintikka, 1970)". Nous verrons que Fagin et Halpern ont construit une logique du raisonnement local qui permet de rendre compte de cette absence de fermeture.

2.5. L'atténuation de la croyance : $Bp \Rightarrow B(p \vee q)$

2.6. La cohérence de la croyance : $Bp \Rightarrow \neg B\neg p$

Si l'on rejette la fermeture sur la conjonction, on devrait rejeter aussi 2.6. En effet, un sujet connaissant qui n'est pas capable de mettre ensemble deux et deux, pour reprendre l'exemple de Hintikka, pourrait très bien abriter dans son esprit des croyances incompatibles dont il n'a jamais vu qu'elles étaient incompatibles parce qu'ils ne les a jamais rapprochées l'une de l'autre.

2.7. Bp implique p. Ceci ne fait pas difficulté dans le système KD45 qui ne contient pas l'axiome T dont 2.7 est une instance

2.8. $B(Bp \Rightarrow p)$

2.9. B *true*. Le sujet croit toutes tautologies

2.10. Un plus grand nombre d'alternatives épistémiques fournit un plus petit nombre de croyances (principe de l'antimonotonie des alternatives épistémiques)

Nous ajouterons à l'inventaire précité le cas suivant d'omniscience qui mérite une attention particulière.

2.11. Si $p \Leftrightarrow q$, alors $Bp \Leftrightarrow Bq$

Dans la sémantique des mondes possibles, deux énoncés expriment la même proposition s'ils ont les mêmes conditions de vérité, c'est-à-dire si la classe des mondes possibles dans lesquels le premier est vrai est la même que celle dans lesquelles le deuxième est vrai. Or deux énoncés logiquement équivalents sont vrais et faux respectivement dans les mêmes mondes possibles. Il s'ensuit que si le premier est cru le second l'est aussi et réciproquement. Comme le note Stalnaker, "Ceci, en soi n'est pas l'omniscience déductive, mais c'est un principe très fort qui ne serait pas plausible dans une sémantique réaliste de la croyance (Stalnaker, 1991, 432)". La solution généralement proposée, c'est de distinguer la croyance braquée sur des états de chose de l'assertion qui a pour objets ces états de choses médiatisés par le langage. Pour ces derniers, un critère d'individuation plus fin est de rigueur. Le problème qui se pose alors est de conjurer le danger opposé, celui d'avoir un critère d'individuation si fin qu'il distingue les phrases qui expriment en d'autres mots le même contenu. Nous n'avons cependant pas à traiter ce dernier problème si notre objectif est de rendre compte de la compétence déductive plutôt que de la compétence sémantique (capacité de reconnaître les synonymes). Il reste que nous ne pouvons imputer à un agent humain la loi 2.11. En effet, cela reviendrait à lui imputer la connaissance de toutes les tautologies dès le moment où il connaîtrait l'une d'elles.

3. La logique générale de la conscience et la logique du raisonnement local de Fagin et Halpern

3.1. La logique générale de la conscience

Fagin et Halpern introduisent un opérateur syntaxique de conscience où "i est conscient de la formule F" peut recevoir diverses interprétations parmi lesquelles "i est capable de calculer la vérité de p en un temps t". On se donne une structure de Kripke qui vérifie les axiomes de KD45 et on définit pour chaque agent et pour chaque monde possible auquel il a accès une classe de formules dont il est conscient dans le monde possible (dans l'alternative doxastique) considéré. Aucune restriction n'est imposée à cet ensemble de formules : "En particulier, il est possible pour F et -F de figurer dans l'ensemble A(i,s) (de telle sorte que l'ensemble des formules dont l'agent est conscient peut être incohérent. Il est possible aussi que l'une seulement des deux formules F et -F figure dans l'ensemble ou encore qu'aucune des deux ne s'y trouve. Il est possible aussi, par exemple, que l'ensemble en question contienne

F&G, mais pas G&F. Les formules de l'ensemble de conscience sont les formules dont l'agent "est conscient", et ces formules ne sont pas nécessairement les formules qu'il croit (Fagin et Halpern, 1988, 53)".

Dans cette logique, on a $Bp \Leftrightarrow Lp \& Ap$ où p est une formule quelconque (pas nécessairement une formule atomique). Les formes les plus gênantes d'omniscience logique (2.1-2.5 et 2.11) sont écartées. Par exemple, la croyance cesse d'être fermée sur l'implication matérielle. On peut, en effet, construire un modèle qui vérifie $Bp, Bp \Rightarrow q$ et $\neg Bq$ puisque l'agent i peut ne pas être conscient de q bien qu'il soit conscient de p et de $p \Rightarrow q$. Dans cette logique, on peut non seulement être conscient des parties sans être conscient du tout, mais on peut même, comme le montre l'exemple qui vient d'être présenté, être conscient du tout sans être conscient des parties. Cette logique n'entraîne plus l'omniscience logique, mais ne tombe-t-elle pas dans l'erreur opposée ? Ne traite-t-elle pas les sujets connaissant comme des agents totalement dépourvus de capacité logique, y compris de la capacité élémentaire de pouvoir analyser une formule en ses constituants ? Pour remédier à ce défaut, on pourrait postuler que la propriété de conscience est fermée sur l'opération de décomposition d'une formule en ses sous-formules. Mais, comme le remarquent Fagin et Halpern, cette supposition apparemment inoffensive a un effet inattendu : elle restaure la fermeture de la croyance sur l'implication matérielle.

Dans la logique générale de la conscience, la loi de nécessité n'est pas supprimée, elle n'est que restreinte. Pour qu'une formule soit crue, il ne suffit pas qu'elle soit un théorème, il faut que l'agent ait conscience de la dite formule. La croyance explicite, qui est la croyance implicite conjuguée à la prise de conscience, garde - Fagin et Halpern en conviennent - "beaucoup des propriétés de la croyance implicite une fois qu'on relativise celle-ci à la prise de conscience (Fagin et Halpern, Op. cit., 55)". Si l'on peut construire un modèle qui vérifie $Bp, B(p \Rightarrow q)$ et $\neg Bq$, en revanche, on ne peut construire un modèle qui vérifie $Bp, B(p \Rightarrow q)$ et $B\neg q$ (W. van der Hoek et J.J. Ch. Meyer, 1988, 6). La différence radicale que la logique générale de la conscience institue entre ces deux cas paraît assez artificielle. Elle disparaîtra dans la logique de E. Gillet.

La logique générale de la conscience prête le flanc à une autre objection, celle-ci dirimante : elle attribue à l'agent une perfection logique que l'on pourrait qualifier de négative : l'immunité vis-à-vis des croyances incohérentes. Si on peut construire un modèle dans lequel un agent ne croit pas une tautologie, on ne peut, en revanche, construire un modèle où il croit une contradiction. Certes, l'agent peut être conscient à la fois de p et de $\neg p$, mais il ne peut croire $p \& \neg p$. Comme l'a observé E. Gillet, cette impossibilité est liée à la définition de la croyance explicite comme équivalente à la conjonction de la conscience et de la croyance implicite. Le raisonnement est immédiat : si "i croit explicitement p " est logiquement équivalent à "i croit implicitement p et est conscient de p " (où p est une formule quelconque), alors "i croit explicitement p " implique "i croit implicitement p ". Or la croyance implicite étant tenue pour logiquement parfaite, i ne peut croire implicitement une contradiction, donc - par *modus tollens* - il ne peut croire explicitement une contradiction. Il suffit d'un seul contre-exemple vraiment rebelle pour réfuter une théorie. C'est parce que la logique générale de la conscience achoppait sur cette facette de l'omniscience logique que Fagin et Halpern ont élaboré une solution de rechange, la logique du raisonnement local.

3.2. La logique du raisonnement local

Dans la logique générale de la croyance, l'agent i dans le monde s croit la proposition p si p est vraie dans l'ensemble des mondes que i croit possibles dans s . Dans la logique du raisonnement local, "il n'y a pas nécessairement un ensemble unique d'états ("mondes" dans la terminologie courante) qu'un agent pense être possibles, mais plutôt une pluralité d'ensembles, chacun correspondant à un amas différent de croyances (Ibid.58)". La croyance implicite s'obtient en regroupant toute l'information contenue dans chacun de ces amas de croyances qui constituent autant d'"états d'esprit" (*frames of mind*).

Les auteurs formulent alors une définition récursive de la vérité pour cette nouvelle structure. Le regroupement de l'information se traduit par l'action de prendre l'intersection des amas de croyances. L'omniscience logique est évitée par d'autres voies que dans la logique générale de la conscience : "La formule $Bi\ p \ \& \ Bi(p \Rightarrow q)$ & $-Bi\ q$ est vérifiable dans un modèle simplement parce que dans un état d'esprit l'agent i pourrait croire p et qu'il pourrait dans un autre état d'esprit croire $p \Rightarrow q$, sans être jamais dans un état d'esprit où il mettrait ensemble ces faits pour conclure q (Ibid. 59)". On trouve ici une modélisation parfaite et rigoureuse de la situation évoquée au début dans le passage de Hintikka.

L'atout décisif de cette nouvelle sémantique réside dans sa capacité de rendre compte de la croyance incohérente : " $Bi\&Bi\text{-}p$ est *satisfiable* puisque dans un état d'esprit l'agent i pourrait croire p , tandis que dans un autre, il pourrait croire $\neg p$ (Ibid.59-60)". Cette sémantique cesse de vérifier le système axiomatique KD45, mais on peut l'enrichir en lui imposant des contraintes. On peut également "ajouter une fonction de conscience à une structure conçue pour (représenter) le raisonnement local". On peut d'ailleurs, comme l'a montré E. Thijssse dans la monographie mentionnée plus haut, *réduire* les structures à base d'amas à des structures beaucoup plus classiques, les modèles de voisinage de Montague-Scott.

Mais la possibilité de cette réduction a une conséquence inattendue que E. Thijssse exprime en ces termes : "les modèles de voisinage reposent sur l'idée simple qu'une proposition correspond à l'ensemble des mondes possibles dans lesquels elle vraie. Cette sémantique ne peut donc distinguer les unes des autres les propositions logiquement équivalentes, ce qui est inacceptable pour les logiques de la conscience (E. Thijssse, 1991, 24)". La leçon que l'on doit tirer du résultat profond obtenu par E. Thijssse est claire. Malgré sa grande souplesse et sa puissance explicative, la sémantique du raisonnement local et l'axiomatique correcte et complète qui lui est associée ne réussissent pas à exorciser le paradoxe de l'omniscience logique.

4. La contribution d'E. Gillet

4.1. Les degrés de conscience

La logique du raisonnement local permet de rendre compte du cas où un agent entretient des croyances mutuellement incompatibles, ce qui formellement se note par $Bi\&Bi\text{-}p$, mais cette logique exclut comme impossible qu'un agent croie la

formule contradictoire p&-p. On doit certes accorder à Fagin et Halpern qu'aucun agent n'est assez illogique pour acquiescer à une formule contradictoire dont l'incohérence est aussi flagrante. Mais on trouvera sans peine une formule singulière qui est tout aussi contradictoire, mais dont le caractère contradictoire n'apparaît pas à la conscience et qui, pour cette raison, sera crue par certains. La formule que l'on obtient quand on réunit par une conjonction les prémisses des célèbres dilemmes d'Eulathus et de Protagoras en est un bon exemple. Eulathus avait convenu avec son maître de rhétorique, le sophiste Protagoras, de payer celui-ci quand il aurait gagné son premier procès au tribunal. Eulathus remettant sans cesse à plus tard le moment de payer son cachet à son professeur Protagoras, ce dernier l'assigna en justice et lui tint le raisonnement suivant : "Si tu perds, tu dois me payer en vertu de la loi et si tu gagnes tu dois me payer en vertu de notre convention. Donc tu dois me payer. A quoi Eulathus répondit : "Si je gagne, je ne dois pas te payer en vertu de la loi et si je perds, je ne dois pas te payer en vertu de notre convention. Donc je ne dois pas te payer". La conjonction des prémisses de ce raisonnement a la forme suivante :

$$(p \Rightarrow q) \& (r \Rightarrow q) \& (p \vee r) \& (r \Rightarrow \neg q) \& (p \Rightarrow \neg q) \& (p \vee r)$$

Or cette forme est contradictoire, mais son caractère contradictoire n'apparaît pas à la conscience.

Prenons maintenant le cas des tautologies, on verra que le caractère tautologique n'apparaît que lorsqu'on analyse la formule à un certain niveau de profondeur. Considérons par exemple la tautologie suivante :

$$(p \Rightarrow (q \Rightarrow r)) \Rightarrow (r \vee (p \Rightarrow (q \Rightarrow r)))$$

Cette tautologie est une instance du schéma formée par la variable propositionnelle A. C'est aussi une instance du schéma $A \Rightarrow B$.

C'est également une instance du schéma $C \Rightarrow (D \vee C)$.

C'est encore une instance du schéma $(E \Rightarrow F) \Rightarrow (G \Rightarrow (E \Rightarrow F))$.

Enfin elle est une instance d'elle-même. Il y a donc cinq niveaux d'analyse. Au premier niveau d'analyse, la formule est prise comme un bloc atomique. Son caractère tautologique n'apparaît pas. Au deuxième niveau de décomposition en constituants, on n'a pas encore dégagé assez de structure pour voir qu'on a affaire à une tautologie. Au troisième niveau d'analyse, en revanche, le but est atteint. Le caractère tautologique est devenu flagrant. Comme la substitution préserve la tautologie et que les analyses plus fines ne sont pas autre chose que des schémas obtenus par substitution à partir d'un schéma valide, ces analyses préservent également la tautologie atteinte au niveau 3, niveau que l'on pourrait appeler pour cette raison "profondeur critique" de la formule.

Si on donne au concept de "prise de conscience" dont il a été question jusqu'ici le sens de profondeur d'analyse, on remplace une notion qui obéissait à la loi du tout ou rien par une notion qui admet des degrés et qui peut même être traitée comme une grandeur extensive. C'est dans cette voie que s'est engagé E. Gillet. Il a formulé d'abord une définition récursive de la profondeur d'une formule qui prend en compte les emboîtements de connecteurs et ensuite une définition récursive d'une fonction d'analyse qui satisfait un certain nombre de conditions d'adéquation. Celles-ci nous

assurent notamment "que la chaîne des analyses se termine bien à la formule voulue et qu'il s'agit donc bien de l'analyse de cette formule-là et d'aucune autre (E. Gillet, 1992, 57)".

Chez Fagin et Halpern, les ensembles de formules associés à un agent et à un monde dans la logique générale de la conscience sont dépourvus de structure. L'ensemble des phrases dont l'agent est conscient est totalement arbitraire comme Konolige le remarque (Konolige 1986, 245). Il arrivera, pour cette raison, que l'on doive attribuer à un agent la croyance à la conséquence de la conséquence d'une certaine formule, et à lui dénier la croyance à la conséquence intermédiaire (la conséquence tirée de la formule initiale). L'innovation d'E. Gillet réside dans la spécification réursive de niveaux de conscience, ou, pour reprendre sa terminologie, de niveaux d'analyse. Ceux-ci vont conférer une structure aux ensembles mentionnés plus haut.

4.2. L'enchâssement de l'analyse dans la conscience implicite

La définition de la croyance explicite comme conjonction de la croyance implicite et de la conscience repose sur une conception de la conscience vue comme un *réservoir de phrases*. Stalnaker critique, à juste titre, cette conception de la croyance : "Notre connaissance ordinaire est une *capacité* et notre croyance ordinaire, écrit-il, est une *disposition*. En raison des limites de nos ressources calculatoires, nous pouvons avoir la capacité constituée par la connaissance que P ou la disposition constituée par la croyance que P tout en étant dépourvus de la capacité ou de la disposition que nous aurions si nous étions en possession de la connaissance ou de la croyance de certaines conséquences déductive de P. Mais qu'est-ce que la capacité ou disposition conçue comme capacité ou disposition à faire quelque chose ? Le modèle du réservoir n'a rien à dire à ce sujet et offre peu d'espoir de clarifier le problème de l'omniscience logique (Stalnaker, Op. cit. 436)". Que manque-t-il au modèle du réservoir de phrases ? La connaissance et la croyance conçues respectivement comme capacité et disposition "supposent la capacité d'avoir accès à l'information et pas simplement d'emmagasiner celle-ci dans un réservoir, pour pouvoir distinguer ce que nous connaissons et croyons réellement, au sens ordinaire de ces termes de ce que nous connaissons ou croyons seulement de manière implicite (Ibid. 438)".

La définition de la connaissance explicite en termes de connaissance implicite et de profondeur d'analyse que propose E. Gillet échappe à la critique de Stalnaker. Chez le premier, la croyance explicite n'est pas obtenue par la conjonction de la croyance implicite et de la conscience, mais par l'application de la croyance implicite à un objet qui n'est pas une proposition, mais une proposition analysée à un niveau de profondeur déterminé. La différence entre ces deux conceptions se traduit formellement de la manière suivante : dans la conception de la croyance explicite comme *conjonction* de la croyance implicite et de la conscience, l'ensemble des phrases crues explicitement dans un monde s'obtient en prenant l'intersection des phrases crues implicitement et des phrases dont l'agent est conscient dans ce monde. La croyance explicite B_p est identifiée à la conjonction de L_p et de A_p . Dans la conception d'E. Gillet, au contraire, la proposition analysée à un certain niveau est

enchâssée dans l'opérateur de croyance implicite. En d'autres termes, l'argument de cet opérateur n'est pas une proposition pure et simple, mais une proposition analysée à un certain niveau de profondeur : "Si l'analyse est suffisamment fine, c'est la formule dans toute sa structure qui est crue explicitement, sinon c'est une autre proposition... (E. Gillet Op. cit. 45)", à savoir une proposition à partir de laquelle on pourra accéder à p au prix d'une analyse structurelle plus fine.

Dans la conception courante de la croyance explicite comme conjonction de la croyance explicite et de la conscience, la collection de propositions dont l'agent est conscient n'est pas la même dans toutes les alternatives doxastiques. Comme l'observe E. Gillet, "... la condition supplémentaire Ap ne dépend pas de la structure doxastique, (c'est-à-dire de la relation d'accessibilité qui modélise la croyance implicite), mais seulement de la phrase p et de l'agent qui la croit (Gillet, Op. cit. 45)". Il s'ensuit que "l'idéalisation de la croyance implicite rejaillit directement sur la croyance explicite". Comme on l'a vu plus haut, dès que la clause de prise de conscience est satisfaite, l'agent redevient *parfait raisonneur*. Fagin et Halpern sont les premiers à le reconnaître : "Si vous croyez explicitement p et que vous croyez explicitement $p \Rightarrow q$, alors vous croyez explicitement q pour autant que vous soyez conscient de q (Fagin et Halpern, Op. cit. 55)". Ils affaiblissent la règle de nécessité en introduisant dans la conclusion la condition de "conscience". Leur règle de nécessité doxastique se présente ainsi :

De p , inférez $Ap \Rightarrow Bp$

Si, comme c'est le cas dans la logique de E. Gillet, la proposition analysée à un niveau de profondeur donné tombe dans le champ de l'opérateur d'accessibilité doxastique qui représente la croyance implicite, la conscience n'est pas simplement *juxtaposée* à la croyance implicite comme dans l'analyse précédente. Elle est l'objet auquel s'applique la croyance implicite. On ne passe plus brusquement de la croyance implicite à la croyance explicite, mais on passe de l'une à l'autre de manière graduelle. Il y aura des degrés dans la croyance explicite comme il y a des degrés dans la prise de conscience ou, plus exactement, dans la finesse de l'analyse :

$Bp \Leftrightarrow L(Ap)$

Ce qui se lit : si p , analysé à un degré de finesse i est cru implicitement, alors p est cru explicitement au degré i . La règle de nécessité se formule comme suit :

Ap

Bp

On pourrait juger contre-intuitive une règle qui permet d'inférer que p est cru à partir du fait que p est analysé. C'est oublier que dans la règle de nécessité, ici comme chez Fagin et Halpern, l'input de la règle doit être un axiome ou un théorème. L'indice attaché à "A" et à "B" précise le degré de rationalité de l'agent, rationalité conçue comme capacité d'analyser une proposition.

4.3. La mesure de la profondeur

La profondeur d'une formule atomique est 0. La profondeur d'une formule $\neg p$ est égale à la profondeur de la formule p augmentée d'une unité. La profondeur d'une formule formée par la combinaison de deux formules au moyen d'un des connecteurs binaires suivants ($\&, \vee, \Rightarrow, \Leftrightarrow, \supset$) est égale à la profondeur de la plus profonde des deux accrue d'une unité.

4.4. La fonction d'analyse

Si la profondeur d'une formule est inférieure ou égale au degré de profondeur auquel on l'analyse, toute la structure de la formule analysée apparaît.

Si la profondeur de la formule est plus grande que la profondeur d'analyse, on définit la fonction d'analyse à l'aide d'une définition récursive dont le cas de base est le cas où le degré de finesse de l'analyse est 0. (Dans le contexte de cette définition, nous utilisons des lettres grecques pour représenter des formules, alors que précédemment nous utilisions p, q, r dans ce rôle.)

Dans le cas de base, $A^0(p) = \text{HOL } p$. En d'autres termes la formule est vue *holophrastiquement* (comme un tout homogène). Si $i > 0$, on applique la définition suivante proposée par E. Gillet :

$$\begin{array}{ll}
 \chi = \neg \varphi & A^i(\chi) = \neg A^{i-1}(\varphi) \\
 \chi = \phi \wedge \psi & A^i(\chi) = A^j(\phi) \wedge A^k(\psi) \\
 \chi = \phi \vee \psi & A^i(\chi) = A^j(\phi) \vee A^k(\psi) \\
 \chi = \phi \supset \psi & A^i(\chi) = A^j(\phi) \supset A^k(\psi) \\
 \chi = L \phi & A^i(\chi) = L A^{i-1}(\phi) \\
 \chi = B^i \phi & A^i(\chi) = B^i A^{i-1}(\phi)
 \end{array}$$

avec $(^1) \quad j = i \div (d(\chi) - d(\varphi))$
 $k = i \div (d(\chi) - d(\psi)).$

4.5. Exemple d'application de la fonction d'analyse à une formule

Considérons la formule suivante qui est de profondeur $D = 4$. et déterminons récursivement la structure qui devient perceptible au niveau d'analyse de degré $i = 2$

$$((p \Rightarrow (q \Rightarrow r)) \Rightarrow (r \vee (p \Rightarrow (q \Rightarrow r))))$$

¹ $a - b = \max(0, a - b)$ et $j = \max(0, i - [d(\chi) - d(\varphi)])$

Le membre de gauche de l'implication principale est de profondeur 2 et le membre de droite de profondeur 3. On dévoile le connecteur principal du constituant de droite, à savoir "v" et on diminue l'indice d'analyse d'une unité. Pour le constituant de gauche de l'implication principale, on calcule la quantité de structure que l'on doit dévoiler en appliquant la formule suivante où i a la valeur 2, comme convenu plus haut :

$$i - (1 + (Df - Dg))$$

Df signifie ici "la profondeur du constituant le plus profond" (qui était le membre de droite de l'implication principale) et Dg signifie "la profondeur du constituant le moins profond" (qui était le membre de gauche de l'implication principale). On a $Df = 3$ et $Dg = 2$. Donc $i = 0$. En d'autres termes, à ce stade, on voit le membre de gauche de l'implication principale comme un bloc. Il n'y a plus rien à faire en ce qui le concerne. On travaille alors sur le membre de droite dont le connecteur principal est une disjonction. Le membre de droite de celle-ci est de profondeur 2 et le membre de gauche de profondeur 0 (C'est un atome). On travaille donc sur le membre de droite pour lequel on soustrait une unité à $i = 1$. On a donc pour consigne d'inspecter ce constituant de la disjonction au niveau d'analyse 0, c'est-à-dire de le regarder comme un bloc. Pour le membre gauche de la disjonction, on applique à nouveau la formule en exergue ci-dessus et on obtient :

$1 - (1 + (2 - 0)) = -2 = 0$. En effet 0 est le degré le plus bas de l'analyse. On voit une formule comme un bloc. Il n'y a pas d'analyse négative. Pour nous résumer, au niveau d'analyse 2, ce qu'on voit de la formule, c'est la structure :

$$p \Rightarrow (q \vee p)$$

Ce qui est suffisant pour savoir que la formule est valide. En effet, comme on l'a vu, l'analyse est "monotone" pour ce qui est de la tautologie. Celle-ci une fois installée ne disparaît pas.

On se souvient que dans la logique générale de la conscience, si la formule p est crue explicitement et si la formule $p \Rightarrow q$ l'est aussi, dès que l'agent est conscient de q , il croit q . En d'autres termes, si les deux prémisses du *modus ponens* sont crues explicitement, il ne peut arriver que l'agent, conscient de q , croie explicitement $-q$. Au contraire, dans la logique de l'analyse qui vient d'être exposée, la croyance de $-q$ par l'agent est parfaitement possible. Si le degré d'analyse se calculait uniquement à partir du premier connecteur rencontré en surface lors de l'exploration de l'arbre d'analyse d'une formule, ce ne serait pas le cas, mais si l'on fait usage de la fonction d'analyse que nous venons d'évoquer, le calcul prend en compte non pas le connecteur de surface, mais tout l'arbre de la formule. Voici un exemple, figurant dans la thèse, qui montre bien que la fonction d'analyse n'est pas triviale. Soit la formule :

$$(p \& q) \Rightarrow (p \vee (r \& s))$$

Elle paraît avoir la forme valide $(p \& q) \Rightarrow (p \vee r)$

et cependant la procédure d'analyse présentée plus haut lui confère au niveau 2 la forme non valide $p \Rightarrow (q \vee r)$.

4.6. L'élimination de l'omniscience logique dans le système de E. Gillet

La thèse dont nous présentons ici quelques-uns des apports les plus significatifs comporte un système axiomatique pour une logique de la croyance implicite et de la croyance explicite étendue au calcul des prédicats du premier ordre. Le système, qui est une adaptation originale de KD45, est assorti d'une preuve de correction et de complétude.

Le but visé était d'éliminer les paradoxes de l'omniscience logique. Nous allons montrer succinctement comment sont traités quatre d'entre eux.

L'ensemble des tautologies cesse d'être fermé sur l'opération de nécessité doxastique. Si la formule *input* de l'exemple ci-dessous est analysé seulement au niveau de profondeur 1, l'agent ne voit pas qu'il s'agit d'une tautologie et dès lors ne s'estime pas tenu de la croire :

$$\underline{p \Rightarrow (q \Rightarrow p)}$$

$$B(p \Rightarrow (q \Rightarrow p))$$

La croyance explicite cesse d'être fermée sur l'implication valide. Considérons à nouveau une formule analysée au niveau 1, à savoir la formule valide :

$$p \Rightarrow (q \Rightarrow (p \& q))$$

On voit tout de suite qu'au niveau d'analyse 1, l'agent ne dévoile pas assez de structure dans le deuxième constituant de l'implication principale pour voir que cette dernière est une formule valide. La portion qu'il voit de cette dernière lui permet de construire un modèle où p soit vrai et la schéma $q \Rightarrow r$ qu'il voit, faux.

La croyance cesse d'être fermée sur l'implication matérielle.

Considérons le syllogisme suivant analysé, cette fois encore, au niveau 1 :

$$(p \Rightarrow q) \& (q \Rightarrow (s \Rightarrow (t \vee r)))$$

$$\text{donc } p \Rightarrow (s \Rightarrow (t \vee r))$$

L'agent qui analyse la prémisse au niveau 1 traite le premier membre de la conjonction comme un bloc, ce qui le rend incapable d'exploiter la co-occurrence de p dans la prémisse et la conclusion, co-occurrence qui joue un rôle essentiel dans le raisonnement en question.

Les contradictions qui ne sont pas reconnues comme telles sont susceptibles d'être crues. Voici un exemple d'une telle contradiction dont l'agent n'apercevra pas la nature si son niveau d'analyse est le niveau 2 :

$$((p \& \neg p) \& q) \& r$$

4.7. Recherches futures

La fonction d'analyse proposée n'est nullement *ad hoc*. L'auteur s'est livré dans son travail à une exploration systématique des fonctions d'analyse possibles pour se prémunir contre le reproche qu'on pourrait lui faire d'utiliser un filtre d'analyse arbitraire qui lie trop étroitement la notion de croyance explicite à la morphologie. Il n'en demeure pas moins que - nous devons cette observation à J. Dubucs - la capacité calculatoire d'un agent ne dépend pas uniquement de sa capacité d'analyser finement une formule. Il y a d'autre ingrédients dans la compétence logique. Il suffit pour s'en convaincre de reconsidérer les prémisses des dilemmes d'Eulathus et de Protagoras mentionnés plus haut. Pour voir leur caractère contradictoire, nous n'utilisons pas la technique duale de la substitution, technique que Quine appelle *superstitution* dans ses "Méthodes de logique". Nous ne cherchons pas comment ces prémisses pourraient être obtenues par substitution à partir du schéma contradictoire $p \& \neg p$. La démarche naturelle, c'est de transformer les prémisses à l'aide de la règle d'inférence du dilemme qui nous donne la conjonction de disjonctions $(q \vee r) \& (\neg q \vee \neg r)$, puis, par une transformation sur la base de la règle d'idempotence, d'obtenir la contradiction flagrante $q \& \neg q$, d'où, enfin, nous concluons *ab absurdo*, que les prémisses des dilemmes étaient contradictoires. C'est dans cette direction, pensons-nous qu'il faudrait continuer la recherche dont nous venons de présenter quelques résultats : il faudrait élaborer une logique épistémique qui tiendrait compte de la dynamique de la preuve.

Il ne serait pas juste de terminer sur une note négative et d'insister sur ce qui manque à la logique doxastique que nous avons présentée, - très partiellement - car, à notre connaissance, ce manque affecte aussi les autres logiques doxastiques proposées jusqu'ici. La logique de E. Gillet ne se limite pas à fournir un traitement rigoureux et approfondi de la croyance implicite, de la croyance explicite et de la prise de conscience conçue en termes de faculté d'analyse. Elle permet, par surcroît, d'axiomatiser et de modéliser dans une théorie unitaire les concepts de connaissance multi-agent "transparente" (Si A sait que B sait que p alors A sait que p) et de connaissance multi-agent "opaque" (dans laquelle l'ce même énoncé n'est pas un thèse) concepts qui ont été introduits par P. Weingartner.

Bibliographie

- Dretske F. "Epistemic Operators" *The Journal of Philosophy*, 67, n° 24, 1970.
- Fagin R. & Halpern J.Y. "Belief, Awareness, and Limited Reasoning" *Artificial Intelligence*, 34 (1988) 39-76.
- Gillet E. *Contribution à la logique de la connaissance. Omniscience logique et connaissance faible*, thèse de doctorat, Univ. de Liège, Dept. de Philosophie, 1992.
- Gillet E. & Gochet P. "Le problème de l'omniscience logique" *Dialectica*, sous presse.
- Gillet E. & Gochet P. "On Professor Weingartner's Contribution to Epistemic Logic", *Advances in Scientific Philosophy. Essays in Honour of Paul Weingartner on the Occasion of the 60th Anniversary of his Birthday*.
- Ed. Shurz G. & Dorn G.J.W., Rodopi, Amsterdam 1991, 97-115.

- Hintikka J. "Knowledge, Belief and Logical Consequence", *Ajatus* 1970, 32-47.
- van der Hoek W. & Meyer J.-J. Ch. "Possible Logics for Belief" mss. Amsterdam 1989, 1-11.
- Konolige K. "What Awareness Isn't : A Sentential View of Impollicit and Explicit Belief" *Theoretical Aspects of Reasoning About Knowledge*, Proceedings of the 1986 Conference. Ed. by Halpern J.Y. Monterey California, 241-250.
- Stalnaker R. "The Problem of Logical Oniscience I". *Synthese*, 89 (1991) 425-440.
- Thijssse E. "On total awareness logics" Research Report Tilburg University 191 1-25 publié ultérieurement dans *Colloquium on Modal Logic* 1991, ed. by M. de Rijke, Amsterdam 1991.
- Voorbraak F. "Generalized Kripke Models for Epistemic logic" *Theoretical Aspects of Reasoning About Knowledge*, Proceedings of the Fourth Conference (TARK 1992) Ed. by Moses Y., Monterey California 1992, 214-228.

Paul GOCHET

Chapitre 11

L'analyse de la forme logique des propositions en grammaire universelle

Depuis sa fondation, la logique s'efforce d'analyser la forme des *propositions* qui procèdent à une *généralisation universelle* ou *existentielle* comme les propositions que tous les hommes sont mortels et que quelque chose est beau. D'un point de vue linguistique, de telles *propositions généralisées* sont exprimées par des *énoncés* contenant des *quantificateurs*. Comme Aristote lui-même l'avait déjà remarqué, la quantification universelle et existentielle est une activité indispensable à l'abstraction et au raisonnement humain pratique et théorique. Ainsi l'analyse de la forme des propositions généralisées est un objectif important de la logique.

Malheureusement, les différentes logiques de la quantification qui ont été développées jusqu'à présent aussi bien dans l'approche objectuelle¹ et substitutionnelle² que dans l'approche algébrique³ ont seulement procédé à l'analyse des *conditions de vérité* des propositions généralisées. Ces logiques ont eu tendance à identifier les propositions généralisées avec les *sens* (ou les valeurs sémantiques) des *énoncés pourvus de quantificateurs*. Elles ont négligé le fait que les propositions généralisées sont également les *contenus des pensées conceptuelles* que nous avons la capacité d'accomplir et d'exprimer dans l'usage du langage. Ainsi de nombreux actes illocutoires et attitudes ont des propositions généralisées comme contenu. Par exemple, nous pouvons affirmer et croire la proposition que tous les hommes sont

1. L'approche objectuelle est due à Tarski. Voir son article "The concept of truth in formalized languages" présenté à Varsovie en 1931.

2. Voir J.M. Dunn & N. Belnap, "The Substitution Interpretation of the Quantifiers", *Noûs*, vol. 2, 1968.

3. Voir W. Craig, *Logic in Algebraic Form*, North Holland, 1974.

mortels. Nous pouvons également promettre et avoir l'intention d'être bons envers tous les hommes. La logique de la quantification doit donc être compatible avec les lois et observations de base de la théorie des actes de discours⁴ et de la philosophie de l'esprit et des états mentaux. En particulier, cette logique doit tenir compte du fait que toute proposition est en principe *exprimable* lors de l'accomplissement d'un acte de discours. Comme Austin⁵ et Searle l'ont montré, ce sont des actes illocutoires complets pourvus d'une force et d'un contenu propositionnel et non pas des propositions isolées qui sont les unités premières de signification dans l'usage du langage. Par conséquent, on a besoin d'une logique des propositions qui tienne compte de leur expression lors de l'accomplissement des actes illocutoires en grammaire universelle. Une telle logique propositionnelle doit être plus fine que la logique intentionnelle actuelle.

Dans *Meaning and Speech Acts*, j'ai formulé une nouvelle analyse de la forme logique des propositions qui tient compte de leur double nature de *sens* d'énoncés et de contenu de pensées. Selon cette analyse, toute proposition qui est le sens d'un énoncé dans un contexte d'emploi d'une interprétation sémantique est aussi le *contenu* de l'acte illocutoire possible que l'énonciation littérale de cet énoncé pourrait servir à accomplir dans ce contexte. Plus récemment⁶, j'ai développé sur la base de ces idées (et indépendamment de la théorie des actes de discours) une *logique minimale naturelle complète des propositions* dont les théorèmes expriment des lois valides pour les propositions élémentaires prédisant des attributs extensionnels quelconques et les propositions complexes qui sont des fonctions de vérité de telles propositions élémentaires.

Contrairement aux logiques modale⁷ et intensionnelle⁸ classiques, cette nouvelle logique des propositions décrit la *structure de constituants* et le *contenu* des propositions. En particulier, elle identifie l'ensemble des propositions atomiques dont est composée chaque proposition⁹. Elle décrit *comment les sens qui sont les constituants d'une proposition sont logiquement liés par la prédication* dans les propositions atomiques. Finalement, à la différence des autres logiques modale,

4. Voir J.R. Searle & D. Vanderveken, *Foundations of Illocutionary Logic*, Cambridge University Press, 1985 et D. Vanderveken, *Meaning and Speech Acts*, Volume I: *Principles of Language Use* et Volume II : *Formal Semantics of Success and Satisfaction*, Cambridge University Press, 1990-1.

5. Austin, J.L. *How to do Things with Words*, Oxford: Clarendon Press, 1962.

6. Voir "La forme logique des contenus de nos pensées conceptuelles" dans D. Laurier & F. Lepage (dirs.), *Essais sur le langage et l'intentionnalité*, Bellarmin, 1992 et "A new Formulation of the Logic of Propositions" dans M. Marion & R. Cohen (dirs.), *Quebec Studies in the Philosophy of Science*, Kluwer, 1995.

7. Carnap, R., *Meaning and Necessity*, University of Chicago Press, 1956.

8. Voir A. Church, "A Formulation of the Logic of Sense and Denotation", in P. Henle et al. (eds.), *Structure, Method and Meaning*, New York: Liberal Arts Press, 1951 et R. Montague, *Formal Philosophy*, Yale University Press, 1974.

9. L'ensemble des propositions atomiques d'une proposition est appelé le *contenu* de cette proposition.

intensionnelle et hyperintensionnelle¹⁰ et de la logique de l'implication analytique de Parry¹¹, ma nouvelle logique propositionnelle explique formellement les conditions de vérité d'une façon plus constructive que Carnap dans *Meaning and Necessity*. Les conditions de vérité de chaque proposition ne sont pas réduites à la fonction de l'ensemble des mondes possibles dans l'ensemble des valeurs de vérité qui associe à chaque monde possible la valeur de vérité de cette proposition dans ce monde. Au contraire, les *conditions de vérité des propositions* sont expliquées en tenant compte de la façon dont nous les déterminons en pensant à partir des possibilités de vérité de leurs propositions atomiques. Ainsi je propose également un *traitement intensionnel des conditions de vérité* qui formalise des aspects cognitifs sous-jacents à l'appréhension même des conditions de vérité.

Le présent article entend poursuivre ces investigations logico-philosophiques sur les propositions. Son but principal est d'élaborer une logique de la quantification et des propositions généralisées qui soit adéquate pour les besoins de la philosophie du langage et de la sémantique formelle. Une telle recherche logique soulève les questions suivantes :

Quelle est la structure de constituants des propositions généralisées ? Quelle est la nature de leurs propositions atomiques ? En particulier, comment la quantification et la prédication sont-elles liées logiquement dans la structure logique qui leur est propre ? Et de quelle façon comprenons-nous les conditions de vérité de telles propositions ?

La première section spécifie des critères d'adéquation que toute logique de la quantification doit satisfaire pour des raisons philosophiques. Cette section formule ensuite de nouveaux principes plus fins d'analyse logique du contenu et des conditions de vérité des propositions généralisées. La seconde section définit la langue-objet idéale de ma logique des propositions généralisées. Cette nouvelle langue artificielle est une extension de la langue-objet de la logique minimale des propositions mentionnée plus haut. Son lexique contient quelques nouvelles constantes logiques exprimant l'identité et la quantification existentielle à l'intérieur des propositions ainsi que les opérations classiques de la logique des relations. Comme nous le verrons, la structure grammaticale superficielle des formules exprimant des propositions généralisées exhibe diagrammatiquement leur forme logique. La troisième section définit la structure sémantique d'un modèle pour la nouvelle langue formelle et la quatrième section présente un système axiomatique dont les théorèmes sont des énoncés logiquement vrais. Enfin, la dernière section énumère une série de lois valides fondamentales de notre logique.

10. Cresswell, Max J. "Hyperintensional Logic", *Studia Logica*, 34: 25-38, 1975

11. Voir W.T. Parry, "Ein Axiomensystem für eine neue Art von Implikation (analytische Implikation)", *Ergebnisse eines Mathematisches Colloquiums*, Volume 4, 1933 et "Comparison of entailment theories", *The Journal of Symbolic Logic*, Volume 37, 1972 et K. Fine, "Analytic implication", in *Notre Dame Journal of formal Logic*, Volume 27, numéro 2, avril 1986.

1. Principes généraux de la logique de la quantification

Voici quelques considérations philosophiques qui imposent des contraintes formelles décisives à la logique des propositions généralisées.

1) Le contenu d'une proposition généralisée ne peut être identifié à l'ensemble des propositions atomiques de toutes les propositions plus simples qui sont ses instances de substitution.

Si chaque proposition est en principe exprimable dans l'usage du langage, alors on ne peut adopter l'analyse standard objectuelle ou substitutionnelle de la quantification pour définir adéquatement en grammaire universelle l'ensemble des propositions atomiques de chaque proposition généralisée. En effet, nous autres, les êtres humains, qui sommes par nature pourvus de compétence linguistique créative¹², avons des *capacités cognitives* intrinsèques *limitées*. D'abord, nous ne pouvons accomplir *qu'un nombre fini d'actes de référence et prédication* dans un contexte d'énonciation. Ainsi le nombre des constituants propositionnels (attributs et concepts) et celui des propositions atomiques que nous sommes capables d'exprimer dans un contexte d'énonciation sont par hypothèse *finis*. Qui plus est, nous ne pouvons nous référer à des entités sous des concepts ni prédiquer des attributs (des relations ou des propriétés) de telles entités sans avoir à l'esprit ces concepts et ces attributs. Par conséquent, la proposition universelle que tout objet est identique à lui-même ne peut être analysée comme étant la longue conjonction que Socrate est identique à lui-même, que Platon est identique à lui-même, qu'Alexandre le Grand est identique à lui-même *et* ainsi de suite pour tous les autres objets. De même, la proposition existentielle qu'un objet est identique à lui-même ne peut être analysée comme étant la disjonction correspondante de telles propositions.

Ainsi, les généralisations universelle et existentielle ne sont pas réductibles respectivement à des conjonctions et à des disjonctions de leurs instances de substitution, comme Wittgenstein l'a dit dans son *Tractatus logico-philosophicus*. En effet, lorsque nous exprimons des propositions généralisées à propos des entités d'un certain type, nous ne pensons pas spécifiquement pour autant à toutes les entités particulières en question.

Le plus souvent, nous appréhendons les propositions généralisées, qui sont les sens d'énoncés avec quantificateurs, sans faire de référence à toutes les entités particulières qui pourraient être assignées comme valeur aux variables liées de tels énoncés. De même, nous ne prédisons pas de toutes ces valeurs possibles un certain attribut dans l'expression de telles propositions. Ainsi, d'un point de vue cognitif empirique, le contenu d'une proposition généralisée n'est *pas* l'ensemble des propositions atomiques dont sont composées toutes les propositions plus simples qui sont ses instances de substitution. Qui plus est, d'un point de vue logique, on ne peut identifier ce contenu à un tel ensemble sans violer le principe même d'exprimabilité. En effet, nous sommes capables de quantifier sur des domaines

12. Nous sommes capables d'apprendre des langues naturelles contenant un nombre infini d'énoncés.

infinis, comme l'ensemble des nombres naturels et d'exprimer des propositions généralisées ayant un nombre infini d'instances de substitution différentes. Cependant, il est bien clair que notre finitude nous empêche de référer à chaque élément particulier d'un ensemble infini. Nous ne pouvons jamais avoir à l'esprit l'ensemble infini des propositions atomiques des instances de substitution de telles propositions généralisées.

Pour décrire la structure *finie* de constituants des propositions généralisées, je m'inspirerai plutôt ici de la conception fré géenne de la quantification selon laquelle généraliser, c'est toujours prédiquer d'un attribut une certaine propriété logique importante d'ordre supérieur. En particulier, généraliser universellement que tous les objets ont une certaine propriété Φ , c'est selon Frege¹³ prédiquer de cette propriété Φ la propriété logique d'ordre supérieur qu'elle est satisfaite par toute chose. De même, généraliser existentiellement qu'au moins un objet a une propriété Φ , c'est tout simplement prédiquer de Φ l'autre propriété logique d'ordre supérieur qu'elle est possédée par quelque chose.

Dans les dernières décennies, l'analyse fré géenne de la quantification a été adoptée par Church et Montague en logique intensionnelle afin de définir les conditions de vérité des propositions généralisées exprimées par les énoncés contenant des quantificateurs. Ainsi, en logique intensionnelle classique, le quantificateur universel $\forall$ est-il souvent dérivé grâce à l'opérateur λ d'abstraction et à la formule nommant le vrai selon la règle d'abréviation suivante : $\forall x_\alpha A_t =_{\text{def}} \lambda x_\alpha A_t = \lambda x_\alpha 1_t$. Dans cette optique, un énoncé de la forme $\forall x_\alpha A$ signifie simplement que la fonction qui associe le vrai à chaque entité de type α satisfaisant A est celle qui associe le vrai à toutes les entités de type α . Une telle règle d'abréviation est bien entendu une reformulation naturelle de l'idée de Frege, à l'intérieur du λ -calcul modal typé, comme Church lui-même l'a reconnu. Cependant, les logiques intensionnelles de Church et Montague n'ont pas tenu compte des rapports existant entre la généralisation et la prédication selon Frege. Elles en sont malheureusement restées à l'analyse des seules conditions de vérité des propositions généralisées.

Comme je le montrerai plus loin, l'analyse fré géenne de la quantification permet d'analyser aussi bien le contenu que les conditions de vérité des propositions généralisées. Pour cela, il vaut mieux remplacer les approches objectuelle et substitutionnelle classiques de la quantification par une nouvelle approche algébrique¹⁴ s'inspirant de la logique intensionnelle¹⁵.

13. Voir G. Frege, *The Basic Laws of Arithmetic*, University of California Press, 1967, édition originale 1893 et *Foundations of Arithmetic*, Oxford: Blackwell, 1950, édition originale 1884.

14. Henkin L., Monk, D. et Tarski A., *Cylindric Algebras*, Part I, North Holland, 1971.

15. Une approche intensionnelle algébrique des conditions de vérité des propositions généralisées est développée dans G. Bealer *Quality and Concept*, Oxford University Press, 1982.

2) Les propositions strictement équivalentes ne sont pas pour autant identiques. Et il existe une relation logique d'implication entre les propositions généralisées plus fine que l'implication stricte et plus générale que l'implication forte.

Selon la philosophie du langage et de l'esprit, la plupart des propositions strictement équivalentes¹⁶ ne sont pas substituables *salva felicitate* à l'intérieur des forces illocutoires et des modes psychologiques. Car il est souvent possible d'accomplir avec succès un acte illocutoire de la forme F(P) dans un contexte d'énonciation sans pour autant accomplir dans ce même contexte d'autres actes illocutoires de la forme F(Q) dont le contenu propositionnel Q est strictement équivalent avec la proposition P.

(Et semblablement pour les attitudes avec le même mode.) Par exemple, on peut affirmer lors d'une énonciation que Paul a gagné la course sans pour autant affirmer à ce moment que Paul a gagné cette course et que $2 = \sqrt{4}$. Et l'on peut avoir la première croyance sans pour autant avoir la seconde. Cependant, il est clair que ces deux assertions (ou croyances) ont comme contenu des propositions strictement équivalentes qui sont vraies dans les mêmes conditions. Par conséquent, il nous faut absolument un critère d'identité propositionnelle plus fin que l'équivalence stricte en logique des propositions.

Comme je l'ai expliqué ailleurs¹⁷, les propositions strictement équivalentes ne sont pas substituables *salva felicitate* à l'intérieur des pensées conceptuelles pour la raison suivante : L'existence d'une implication stricte entre deux propositions n'engendre pas systématiquement d'engagements illocutoires ou psychologiques forts entre les actes illocutoires de même force et les attitudes de même mode ayant ces propositions comme contenu. Plus précisément, il arrive souvent qu'une proposition P *implique strictement* une autre Q sans que l'accomplissement d'un acte de discours de la forme F(P) avec une force primitive F et le contenu P engage le locuteur à accomplir *ipso facto* l'acte de discours correspondant de la forme F(Q)¹⁸. (Et semblablement pour les attitudes.) Ainsi, la proposition que Jean m'aidera demain implique strictement la proposition qu'il m'aidera ou me tuera demain. Si la première proposition est vraie dans un monde possible, la seconde l'est également. Mais, il est clair que je peux affirmer (et croire) la première sans pour autant affirmer (et croire) la seconde qui contient une nouvelle prédication. Par conséquent, les engagements illocutoires et psychologiques ne sont pas fermés sous la stricte implication, comme Hintikka¹⁹ et d'autres l'ont à tort supposé en logique

16. En logique modale, deux propositions sont *strictement équivalentes* si et seulement si elles ont les mêmes valeurs de vérité dans les mêmes contextes (ou mondes) possibles.

17. Voir *Meaning and Speech Acts*.

18. En logique illocutoire, on dit qu'un acte illocutoire *engage fortement le locuteur* à un autre acte de discours quand toutes les conditions de succès du second acte sont également des conditions du premier. Dans un tel cas, il n'est pas possible d'accomplir le premier acte dans un contexte d'énonciation sans accomplir le second.

19. J. Hintikka, *Knowledge and Belief*, Cornell University Press, 1962.

épistémique. En particulier, la loi d'introduction de la disjonction n'engendre pas généralement des engagements illocutoires ou psychologiques. Des contre-exemples similaires d'échec de la compatibilité de la stricte implication avec les engagements illocutoire et psychologique forts peuvent être donnés pour des lois classiques gouvernant les propositions généralisées comme la loi de l'instanciation universelle. Il est clair que l'on peut affirmer (ou croire) que tous les hommes sont mortels sans pour autant affirmer (ou croire) que si Nabuchodonosor est un homme alors il est mortel. En effet la seconde assertion (ou croyance) exige une référence supplémentaire à Nabuchodonosor.

Quelle est la relation non classique d'implication dont on a besoin en logique des propositions pour expliquer et prédire les engagements illocutoires et psychologiques ? Dans *Meaning and Speech Acts*, j'ai argumenté qu'une telle relation d'implication doit être *cognitivement réalisée par les locuteurs et les allocutaires en vertu de leur compétence linguistique*. Appelons par souci de commodité cette relation la *relation d'implication cognitive forte*. Dans mon optique, étant donné la forme logique des propositions, on peut dire qu'une proposition en implique cognitivement une autre de façon forte si et seulement si les deux conditions suivantes sont remplies : Premièrement, il n'est pas possible d'exprimer la première proposition dans l'usage du langage ou l'exercice de la pensée sans pour autant avoir à l'esprit la seconde. Deuxièmement, on ne peut comprendre les conditions de vérité de la première proposition sans savoir pour autant qu'elle implique la seconde.

Deux hypothèses fondamentales de la théorie des actes de discours sont le principe que le langage est l'œuvre de la raison et que les êtres humains sont minimalement rationnels lorsqu'ils parlent et pensent. Sur la base de tels principes, il est facile d'expliquer pourquoi l'implication cognitive engendre systématiquement des engagements illocutoires et psychologiques forts dans la parole et la pensée. Supposons qu'un locuteur exprime lors d'une énonciation une proposition P avec le but d'établir une correspondance entre les mots et les choses selon la direction d'ajustement d'un but illocutoire. Pourquoi se fait-il qu'il lie également au monde dans le même contexte avec la même direction d'ajustement les autres propositions plus faibles qui sont cognitivement impliquées par la première et qui satisfont les conditions pertinentes sur le contenu propositionnel ? Par hypothèse, ce locuteur ainsi que l'allocutaire savent mutuellement *a priori* en vertu de leur compétence linguistique que la vérité de ces autres propositions est une condition nécessaire au succès d'ajustement de l'énonciation. C'est pourquoi il y a engagement illocutoire fort. Incidemment, la relation d'implication cognitive qu'il s'agit de formaliser doit être *décidable*. Autrement, elle ne pourrait pas être cognitivement réalisée par nous (qui sommes des êtres finitaires) en vertu de la compétence linguistique dans notre usage et compréhension du langage.

Comme je l'ai montré, c'est l'implication forte qui est à la base de l'engagement illocutoire fort dans la logique minimale des actes illocutoires élémentaires. Par définition, une proposition P en implique fortement une autre Q si et seulement si premièrement, toutes les propositions atomiques de Q sont également des propositions atomiques de P et deuxièmement, toutes les assignations de valeur de vérité à des propositions atomiques qui rendent P vraie rendent également Q vraie. En cas d'implication forte de Q par P, il est clair que l'on ne peut exprimer la

première proposition dans un acte de pensée sans également *avoir à l'esprit la seconde*. En effet, il y a inclusion de contenu. En outre, comprendre les conditions de vérité de la première proposition P consiste à identifier les différentes possibilités de vérité de ses propositions atomiques selon lesquelles elle est vraie. Comme toutes ces possibilités de vérité rendent par hypothèse la proposition Q vraie, on ne peut comprendre la proposition P sans comprendre *ipso facto* qu'elle implique Q. Ainsi toute implication forte est bien une implication cognitive forte.

En logique illocutoire minimale, les contenus propositionnels exprimables sont des propositions élémentaires avec des attributs extensionnels quelconques. (Il n'y a pas de propositions élémentaires d'identité). Ou bien ce sont des propositions complexes obtenues à partir de telles propositions élémentaires par l'application des fonctions de vérité. Ainsi il n'y a pas d'implication cognitive forte entre contenus propositionnels qui soit autre que l'implication forte dans la logique minimale des actes illocutoires élémentaires. C'est pourquoi les lois d'implication forte engendrent systématiquement des engagements illocutoires entre les actes de discours avec une force primitive et une direction non vide d'ajustement (lorsque les conditions sur le contenu propositionnel sont satisfaites). Par exemple, l'assertion d'une conjonction est aussi une assertion de ses conjoints. En effet une conjonction implique fortement ses conjoints. En revanche, les lois d'implication stricte comme l'introduction de la disjonction, qui ne sont pas des instances d'implication forte, n'engendrent *pas* d'engagement illocutoire.

Bien entendu, il nous faut en logique des propositions généralisées une relation *d'implication cognitive* qui soit plus générale que l'implication forte. Car la logique des propositions généralisées nécessite une logique formelle des attributs dont les capacités expressives sont beaucoup plus riches que celles de la logique propositionnelle minimale. Cette logique est capable d'exprimer la relation binaire d'identité ainsi que des opérations logiques fondamentales sur les attributs comme la négation et la conjonction booléennes et la généralisation existentielle et universelle. On peut donc y exprimer des propositions d'identité et y représenter différentes formes logiques d'attributs complexes des propositions élémentaires. Or, il est clair que nous sommes capables de dériver de telles propositions, en vertu de notre compétence linguistique, d'autres propositions que celles qui sont impliquées fortement.

Premièrement, lorsque nous pensons que les objets qui tombent sous deux concepts sont identiques, par exemple que Jean est le père de Marie, nous sommes capables de faire des inférences nouvelles en faisant des substitutions de sens sur la base de telles prédications d'identité. Ainsi, il n'est pas possible d'appréhender la proposition que Jean est le père de Marie et qu'il joue aux échecs sans en dériver par substitution la proposition que le père de Marie joue aux échecs. En particulier, l'on ne peut affirmer (ou croire) la première proposition sans pour autant affirmer (ou croire) la seconde. Cependant cette implication cognitive forte n'est pas de l'implication forte au sens défini plus haut. Car la seconde proposition est composée d'une proposition atomique nouvelle. En outre, il n'y a pas d'implication tautologique. D'un point de vue cognitif, on peut dire que notre capacité de faire de telles inférences s'explique par la seule compréhension que nous avons de la relation logique d'identité.

Deuxièmement, lorsque nous prédisons une relation quelconque de deux objets de référence, par exemple, lorsque nous pensons que Jean aime l'art, nous sommes également capables en vertu de notre compétence linguistique de faire d'autres inférences valides qui sont liées à la nature même de la prédication. Ainsi nous avons la capacité de dériver de la proposition que Jean aime l'art les deux propositions que Jean aime quelque chose et que l'art est aimé par Jean en faisant respectivement une généralisation existentielle et une conversion²⁰. De telles implications cognitives ne sont pas non plus des implications fortes. Elles s'expliquent par notre connaissance intériorisée des relations logiques qui existent entre les attributs de certaines formes.

Dans mon optique, deux recherches théoriques doivent être effectuées pour expliquer et prédire les implications cognitives entre propositions.

D'abord, il faut tenir compte des capacités expressives plus riches de la logique des attributs et tenter de définir une relation intensionnelle entre les propositions qui soit plus générale que l'inclusion de l'ensemble des propositions atomiques de l'une dans l'ensemble des propositions atomiques de l'autre. Dans ce but, j'associerai à chaque proposition P un nouvel ensemble de propositions atomiques que j'appellerai l'*expansion cognitive du contenu de P*. Quand un locuteur *exprime* (ou *comprend*) dans un contexte une proposition P de notre langue formelle, il est *ipso facto* capable, en vertu de sa compétence linguistique, d'exprimer dans ce contexte en plus des propositions atomiques de P un certain nombre d'autres propositions atomiques. Il est capable de prédiquer des différents objets de référence certains attributs nouveaux qu'il n'a pas exprimés mais qui peuvent être dérivés des attributs qu'il a exprimés en exprimant la proposition P. Ainsi un locuteur qui prédique dans un contexte une relation binaire de deux objets de référence en pensant, par exemple, que Jean aime l'art est capable de concevoir dans ce contexte la propriété d'aimer quelque chose qui est obtenu par généralisation existentielle à partir de cette relation. De même, il a également à l'esprit dans ce contexte la propriété d'aimer l'art qui est le résultat de l'application de la relation prédiquée au second objet de référence. Qui plus est, il prédique également du premier objet de référence (Jean) dans ce contexte ces deux propriétés nouvelles. En effet, il sait *a priori* en vertu de sa compétence que la relation binaire en question (la relation d'aimer) existe entre les deux objets (Jean et l'art) si et seulement si le premier (Jean) a la dernière propriété (celle d'aimer l'art). En outre, il sait également *a priori* que cette propriété (d'aimer l'art) contient l'autre (celle d'aimer quelque chose). Dans le même ordre d'idées, la capacité de prédiquer deux propriétés contient la capacité de concevoir et de prédiquer la conjonction booléenne de ces deux propriétés ainsi que la négation booléenne de chacune d'entre elles. Ainsi, un locuteur qui pense, par exemple, que Paul est stable et Jim imperturbable a d'une certaine façon à l'esprit la propriété complexe d'être à la fois stable et imperturbable (qui est la conjonction booléenne de ces deux propriétés) et les deux propriétés d'être instable et d'être perturbable (qui sont leur négation

20. De même, lorsque nous pensons qu'un objet de référence est dans une certaine relation par rapport à lui-même, nous sommes capables en vertu de notre compétence d'en inférer qu'il a la propriété correspondant à la réflexivisation de cette relation. Par exemple, nous avons la capacité de dériver par réflexivisation de la proposition que Jean aime Jean la proposition que Jean s'aime lui-même.

booléenne). Il fait même des prédications nouvelles avec ces autres propriétés. Par exemple, un locuteur qui pense que Jim est imperturbable pense également qu'il n'est pas vrai qu'il soit perturbable.

De toute évidence, l'accomplissement d'un acte de discours dans un contexte d'énonciation peut engager *illocutoirement* le locuteur à des actes de discours dont les contenus propositionnels ont comme constituants des sens nouveaux que ce locuteur n'a pas explicitement exprimés. Ainsi, l'assertion que Jean est stable et que Paul aime Marie contient celle que Jean n'est pas instable et que Paul aime quelque chose. Dans mon approche logique, l'expansion cognitive du contenu d'une proposition P contient en plus des propositions atomiques de P toutes les propositions atomiques qu'un locuteur a d'une certaine façon nécessairement à l'esprit lorsqu'il exprime la proposition P dans un contexte. Plus précisément, cette expansion cognitive est le plus petit ensemble qui contient en plus des propositions atomiques de P toutes les autres propositions atomiques prédisquant des objets de référence de P les attributs qui sont obtenus en appliquant les opérations classiques de la logique des relations et les opérations de généralisation aux attributs de P²¹. Par exemple, toutes les propositions atomiques prédisquant de Jean et de Paul les propriétés d'être stable, d'être instable, d'aimer Marie et d'aimer quelque chose appartiennent à l'expansion cognitive de la proposition que Jean est stable et que Paul aime Marie. En outre, certaines de ces propositions atomiques ont une valeur de vérité bien déterminée dans les possibilités de vérité selon lesquelles nous considérons que la proposition P est vraie. Par exemple, on ne peut exprimer la proposition que Jean est stable sans considérer qu'elle est vraie si et seulement si la proposition atomique prédisquant de Jean l'instabilité est fausse. C'est pourquoi une assertion (ou croyance) de la première proposition (que Jean est imperturbable) contient une négation de la seconde proposition (que Jean est perturbable).

Une deuxième tâche de notre logique de la quantification est de *découvrir les différentes règles d'inférence que nous sommes capables d'appliquer en vertu de notre compétence linguistique* lors de l'expression des propositions élémentaires, complexes et généralisées qui sont considérées. En logique propositionnelle minimale, les règles d'inférence valides consistent à dériver des conséquences tautologiques. D'un point de vue cognitif, toutes ces règles sont intériorisées dans l'apprentissage du langage. Nous disposons en effet en vertu de notre compétence d'une méthode effective de décision²² qui nous permet de déterminer rapidement pour chaque paire de propositions en forme canonique si l'une implique ou non tautologiquement l'autre. Cependant, comme nous l'avons vu, les règles qui consistent à dériver des conséquences tautologiques ne sont pas les seules règles d'inférence valide qui soient intériorisées. Reste donc à déterminer quelles autres

21. Des exemples de tels attributs sont la négation, la conversion, la conjonction booléennes des attributs de la proposition P ainsi que ceux qui résultent de la généralisation existentielle de ces attributs.

22. La méthode des tables de vérité découverte par Wittgenstein [1921] et celle des tableaux sémantiques de Beth sont des approximations formelles de la méthode de décision effective que nous appliquons en parlant.

règles d'inférence valide ont une pareille réalité psychologique dans l'usage et la compréhension du langage.

Pour découvrir de telles règles, il me semble utile de considérer les engagements illocutoires ou psychologiques faibles qui existent entre les pensées conceptuelles ayant des propositions généralisées comme contenu. Comme Searle et moi l'avons signalé²³, il faut distinguer l'*accomplissement manifeste d'un acte illocutoire* d'un *simple engagement illocutoire faible à cet acte*.

L'*accomplissement manifeste d'un acte illocutoire* peut engager le locuteur à un autre acte, même si cet engagement n'implique pas un *accomplissement manifeste* de ce nouvel acte. Par exemple, un locuteur faisant l'assertion que Jean parle français est engagé à affirmer qu'il est possible que Jean parle français, même s'il n'a pas fait explicitement la seconde assertion. L'existence d'un tel engagement illocutoire faible se montre dans le fait qu'il est impossible d'affirmer avec succès la première proposition (que Jean parle français) et de nier simultanément la seconde (qu'il est possible que Jean parle français). Ainsi l'énoncé "Jean parle français et il n'est pas possible qu'il parle français" est-il paradoxal et illocutoirement incohérent.

Lorsqu'un acte illocutoire $F(P)$ *engage faiblement le locuteur* à un autre acte $F(Q)$, il est parfois possible d'accomplir le premier acte sans accomplir *ipso facto* le second. Mais il n'est jamais possible d'accomplir simultanément avec succès dans un même contexte d'énonciation l'acte $F(P)$ et une dénégation illocutoire $\neg F(Q)$ du second²⁴. En particulier, quand la force F a une direction non vide d'ajustement, nous ne pouvons même pas avoir l'intention d'accomplir simultanément les deux actes de discours $F(P)$ et $F(\neg Q)$. Car nous savons *a priori* en vertu de notre compétence linguistique que leurs conditions de satisfaction sont incompatibles et que le succès d'ajustement entre les mots et les choses est dans ce cas impossible.

En logique illocutoire minimale, c'est la relation d'*implication tautologique* qui est à la base des engagements illocutoires faibles entre les actes de discours avec les mêmes forces illocutoires. Quand une proposition P implique tautologiquement une autre Q , les actes illocutoires de la forme $F(P)$ pourvus d'une force primitive avec une direction non vide d'ajustement *engagent faiblement* le locuteur aux actes illocutoires correspondants de la forme $F(Q)$ en cas de satisfaction par Q des conditions sur le contenu propositionnel de F . Par exemple, une assertion que Paul ne parle pas français engage faiblement le locuteur à l'assertion qu'il ne parle pas à la fois français et flamand. Sans doute est-il possible de faire la première assertion dans un contexte sans avoir à l'esprit le contenu propositionnel de la seconde, car celui-ci contient une prédication nouvelle. L'engagement illocutoire est donc faible. Cependant, l'implication tautologique entre deux contenus propositionnels a une certaine réalité psychologique minimale : nous ne pouvons avoir à l'esprit dans un

23. Voir notre livre *Foundations of Illocutionary Logic*.

24. La dénégation d'un acte de discours $\neg F(P)$ est l'acte illocutoire complexe dont le but est de manifester que le locuteur n'accomplit pas l'acte $F(P)$. Par exemple, un refus est la dénégation illocutoire d'une acceptation. Voir *Foundations of Illocutionary Logic* et mon article "On the Logic of Illocutionary Denegations and Conditional Speech Acts" pour plus d'information sur les actes illocutoires de dénégation.

même contexte d'énonciation les deux propositions en question sans savoir *a priori* en vertu de notre compétence que la première ne peut être vraie sans que la seconde le soit également.

Il est bien clair qu'il existe aussi des engagements illocutoires faibles qui ne sont pas forts entre les actes de discours dont les contenus sont des propositions généralisées. Ainsi la loi d'instanciation universelle engendre également des engagements illocutoires faibles qui ne sont pas forts. Par exemple, il n'est pas possible d'affirmer que tous les hommes sont mortels sans être engagé à affirmer à cause de la généralisation universelle que si Nabuchodonosor est un homme, alors il est mortel. D'un point de vue logique, cet engagement illocutoire est faible. Car en faisant la première assertion, l'on ne fait pas d'acte de référence à Nabuchodonosor. Donc l'on ne fait pas ouvertement la seconde assertion. Cependant, il est paradoxal de dire "Tous les hommes sont mortels, Nabuchodonosor est un homme et il n'est pas mortel". L'existence de cet engagement faible se montre dans le fait linguistique qu'un tel énoncé est illocutoirement incohérent.

Dans mon approche logique, les trois espèces suivantes d'inférences valides sont *intériorisées* dans l'apprentissage du langage lors de l'expression des propositions considérées : premièrement, *l'inférence des conséquences tautologiques*, deuxièmement, *l'inférence des instances de substitution d'une proposition universelle* et troisièmement, *l'inférence des équivalences matérielles qui sont les conséquences d'une proposition d'identité*. La réalité psychologique de ces trois espèces de règles d'inférence est bien entendu moindre que celle des règles d'inférences qui consistent à dériver les implications cognitives fortes. Car nous pouvons exprimer les prémisses de telles règles d'inférence sans pour autant avoir à l'esprit leur conclusion. Ainsi, nous pouvons exprimer une proposition sans avoir à l'esprit une conséquence tautologique de cette proposition dont l'expression exige des actes de référence ou de prédication totalement nouveaux. De même, nous pouvons exprimer une proposition généralisée universelle sans pour autant avoir à l'esprit une instance particulière de cette proposition exigeant un nouvel acte de référence. Enfin, nous pouvons identifier les objets qui tombent sous deux concepts en exprimant une proposition d'identité sans pour autant avoir à l'esprit toutes les équivalences matérielles entre propositions qui résultent de cette prédication d'identité. Néanmoins, il est bien clair que, dans tout contexte possible d'énonciation où nous exprimons à la fois les prémisses et la conclusion de ces trois espèces de règles d'inférence, nous savons toujours *a priori* en vertu de notre compétence linguistique que les prémisses en question ne peuvent être vraies sans que la conclusion le soit également. La *réalité psychologique* de ces différentes règles d'inférence valide se montre en particulier dans l'*existence d'une méthode effective de décision* correspondant à chacune d'entre elles. Ainsi, pour tout ensemble fini Γ de propositions et n'importe quelle proposition P en forme canonique, nous avons la capacité de déterminer en un nombre fini d'étapes si la proposition P peut ou non être dérivée en appliquant les règles d'implication tautologique, d'instanciation universelle ou de substitution d'identiques aux propositions de l'ensemble Γ .

Quand une proposition P peut effectivement être dérivée d'un ensemble fini Γ de propositions en appliquant les trois types de règles d'inférence en question, je dirai dorénavant que la proposition $\&\Gamma$ qui est la conjonction de toutes les propositions de l'ensemble Γ implique alors *cognitivement de façon faible* la proposition P . Comme

on l'a vu, l'on peut exprimer une proposition sans pour autant avoir à l'esprit de nombreuses autres propositions avec des constituants propositionnels indépendants qu'elle implique pourtant cognitivement. C'est pourquoi l'implication cognitive qui vient d'être définie est bien faible d'un point de vue psychologique. Cependant, aucun locuteur compétent ne peut avoir simultanément à l'esprit (et comprendre pleinement) deux propositions en relation d'implication cognitive faible sans savoir *a priori* à ce moment que l'une ne peut être vraie sans que l'autre le soit également. Il y a donc une certaine réalité psychologique minimale dans l'implication cognitive faible.

Étant donné le principe que le langage est l'œuvre de la raison, l'implication cognitive faible entre propositions est l'origine logique des engagements illocutoires faibles entre actes de discours avec une même force. Ainsi, lorsqu'une proposition P *en implique cognitivement de façon faible* une autre Q, un acte de discours de la forme F(P) pourvu d'une force primitive avec une direction non vide d'ajustement *engage faiblement* le locuteur à l'acte illocutoire correspondant de la forme F(Q), quand les conditions sur le contenu propositionnel sont satisfaites par Q. Par exemple, l'assertion que Jean est le père de Marie engage faiblement le locuteur à l'assertion que Jean joue aux échecs si et seulement si le père de Marie joue aux échecs. Comme on l'a vu, en cas d'implication cognitive faible de Q par P, la proposition impliquée Q peut contenir des sens nouveaux dont la conception est indépendante de celle des constituants de la proposition P. C'est pourquoi l'engagement illocutoire de F(P) à F(Q) qui résulte d'une implication cognitive faible entre deux contenus propositionnels est souvent également faible du point de vue de la logique des actes de discours. Dans de nombreux cas, l'engagement illocutoire faible à un acte de discours F(Q) n'est pas un accomplissement manifeste de cet acte. Ainsi l'assertion que Jean est le père de Marie ne contient pas celle que Jean joue aux échecs si et seulement si le père de Marie joue aux échecs. Car la seconde assertion exige la prédication d'une propriété nouvelle indépendante (celle de jouer aux échecs).

Cependant, quand il y a implication cognitive faible d'une proposition Q par une proposition P et que toutes les propositions atomiques de la proposition Q appartiennent à l'expansion cognitive du contenu de P, mon hypothèse est qu'il y a alors une implication cognitive forte de Q par P. Dans les conditions habituelles, cette implication cognitive engendre un engagement illocutoire fort de l'acte de discours F(P) à l'acte F(Q). Par exemple, la proposition plus complexe que Jean est le père de Marie et qu'il joue aux échecs implique cette fois cognitivement de façon forte celle que le père de Marie joue aux échecs. C'est pourquoi l'assertion de cette proposition contient-elle l'assertion de la seconde.

Autrement dit, selon mon hypothèse, une proposition P *en implique cognitivement* une autre Q *de façon forte* si et seulement si premièrement, toutes les propositions atomiques de Q appartiennent à l'expansion cognitive du contenu de P et deuxièmement, la proposition P implique cognitivement de façon faible la proposition Q. On a vu que l'implication forte exige l'inclusion de contenu en plus de l'implication tautologique. De même, l'implication cognitive exige l'inclusion dans l'expansion cognitive du contenu en plus de l'implication tautologique ou de l'implication par instanciation universelle ou par substitution d'identiques. Ces quelques remarques générales seront formalisées dans les sections suivantes.

2. La langue-objet idéale

Pour formuler la logique de la quantification, je commencerai par définir une *langue formelle* $\mathcal{L}$ appropriée capable d'exprimer les principales formes d'attributs des propositions élémentaires ainsi que les opérations de généralisation. Comme les langues logiques habituelles, la nouvelle langue formelle $\mathcal{L}$ sera *idéale* et *idéographique*. Ses formules seront dépourvues d'ambiguïté et diagrammatiques. Leur structure de surface exhibera clairement la forme logique des propositions élémentaires, complexes et généralisées qu'elles servent à exprimer. Du point de vue de la syntaxe logique, la langue $\mathcal{L}$ de la logique des propositions généralisées est *une extension de la langue-objet de la logique minimale des propositions*. Elle contient quelques constantes logiques additionnelles qui expriment d'une façon algébrique la quantification existentielle ainsi que les opérations de base de la logique des attributs.

Le Vocabulaire de $\mathcal{L}$

La langue-objet idéale $\mathcal{L}$ de LQP contient dans son *lexique* :

- (1) Une série de *constantes individuelles*
 $c_0, c_1, c_2, \dots$
- (2) Une série de *variables individuelles*
 $x_0, x_1, x_2, x_3, \dots$
- (3) Pour chaque $n \in \omega$, une série de *prédicats de degré* n
 $r_n^0, r_n^1, r_n^2, \dots$ incluant le prédicat d'identité $=_2$
- (4) les *expressions syncatégorématiques* :
 $t, =, \neg, \wedge, \exists, \square, \neg, <, >, (\text{ et })$.

Les règles de formation de $\mathcal{L}$

Les termes individuels

Chaque constante ou variable individuelle du lexique est un terme individuel de $\mathcal{L}$ exprimant un concept d'objet individuel.

Définitions

Une séquence $t_1, \dots, t_n$ de n termes individuels est en *ordre canonique* si et seulement si, premièrement, chaque terme individuel apparaît seulement une fois dans cette séquence, deuxièmement, s'il y a des constantes, cette séquence commence par une suite de ces constantes et s'il y a des variables, elle se termine par une suite de ces variables et, finalement, une constante (ou une variable) précède une autre constante (ou variable) dans cette séquence quand son index numérique est inférieur. Ainsi la séquence c_2, c_5, c_7, x_3, x_9 est en ordre canonique alors que les séquences x_8, c_9 et c_0, c_7, x_9, x_8 ne le sont pas.

Soit T un ensemble fini non vide de termes individuels dont $t_1, \dots, t_n$ est la séquence canonique des éléments. La fonction $[]$ de T dans $\{1, \dots, n\}$ telle que pour

chaque $t \in T$, $[t] = k$ si et seulement si t est identique à t_k est appelée la *fonction canonique* de l'ensemble T . Autrement dit, la fonction canonique d'un ensemble de termes individuels associe à chaque terme de cet ensemble le nombre indiquant sa place dans l'ordre canonique de cet ensemble.

Prédicats

Chaque prédicat de degré n du lexique est un prédicat de $\mathcal{L}$ exprimant un attribut de même degré. D'un point de vue logique, l'ordre dans lequel un attribut de degré $n > 1$ est prédiqué de plusieurs objets de référence est important. Car il affecte souvent les conditions de vérité. Beaucoup de relations binaires ne sont pas symétriques. Par exemple, un individu peut aimer un autre qui ne l'aime pas. C'est pourquoi l'extension d'un attribut de degré n dans un contexte en logique est toujours un ensemble de séquences ordonnées de n objets de référence. Dire qu'une séquence d'objets de la forme $\langle u_1, \dots, u_n \rangle$ satisfait un attribut dans un contexte, c'est dire que les n objets en question le satisfont dans ce contexte quand il est prédiqué à eux dans l'ordre de la séquence.

Si R_n est un prédicat de degré n , il en va de même pour $\neg R_n$. Un prédicat complexe de la forme $\neg R_n$ nomme l'attribut de degré n dont l'extension dans un contexte est l'ensemble des séquences de n objets individuels qui n'appartiennent pas à l'extension de l'attribut nommé par R_n dans ce contexte. L'attribut nommé par $\neg R_n$ est la *négation Booléenne* de celui nommé par R_n .

Si R_n est un prédicat de degré $n \geq 1$, $1 \leq k \leq n$ et t est un terme individuel, alors $[t/k]R_n$ est un nouveau prédicat de degré $n - 1$. Un prédicat complexe de la forme $[t/k]R_n$ nomme l'attribut qui est satisfait par une séquence de $n - 1$ individus $\langle u_1, \dots, u_{n-1} \rangle$ dans un contexte si et seulement si l'extension de R_n dans ce contexte contient la séquence $\langle u_1, \dots, u_{k-1}, u, u_k, \dots, u_{n-1} \rangle$, où u est la dénotation du terme t . Par exemple, le prédicat $[t/2]_2$ exprime la propriété d'être identique à la dénotation de t . D'un point de vue logique, une relation binaire comme la relation d'aimer peut être prédiquée d'abord d'un seul objet de référence par exemple de Marie en deux endroits différents. On obtient ainsi les deux propriétés différentes qui sont celles d'aimer Marie et d'être aimé par Marie. De façon plus générale, l'attribut nommé par le prédicat complexe $[t/k]R_n$ est celui de degré $n - 1$ qui est obtenu après avoir prédiqué l'attribut R_n à l'objet de référence t en la $k^{\text{ième}}$ place²⁵.

25. Un des avantages de l'expression d'attributs de la forme $[t/k]R_n$ est de pouvoir rendre compte de l'identité de certaines propositions élémentaires obtenues en prédiquant de façon différente des attributs différents dont la forme logique est liée. Considérons, par exemple, la proposition que Jean aime Marie. On peut l'exprimer en prédiquant simultanément la relation binaire d'aimer à Jean et à Marie. On peut également l'exprimer en prédiquant à Jean la propriété Marie ou en prédiquant à Marie la propriété d'être aimée par Jean. Cependant quel que soit le type de prédication, la même proposition est exprimée.

Si R_n est un prédicat de degré n et $1 \leq k \leq n$, alors $(\exists_k R_n)$ est un nouveau prédicat de degré $n - 1$. Un prédicat complexe de la forme $\exists_k R_n$ nomme l'attribut de degré $n - 1$ qui est satisfait par une séquence d'objets individuels $\langle u_1, \dots, u_{n-1} \rangle$ dans un contexte si et seulement si l'attribut nommé par R_n est satisfait dans ce contexte par au moins une séquence de la forme $\langle u_1, \dots, u_{k-1}, u, u_k, \dots, u_{n-1} \rangle$, où u est un individu quelconque²⁶. L'attribut nommé par un prédicat de la forme $\exists_k R_n$ est celui obtenu en généralisant existentiellement à la $k^{\text{ième}}$ place de l'attribut nommé par R_n . Cet attribut sera appelé une *quantification existentielle* de celui nommé par R_n ²⁷. Par exemple, l'attribut $\exists_{2=2}$ est la propriété universelle d'être identique à quelque chose.

Soit R_n et R_m deux prédicats de degré n et m respectivement. Si $\{l_1, \dots, l_{n+m}\}$ est un ensemble contenant les k premiers nombres naturels (où $1 \leq k \leq n + m$), alors $\langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n \wedge R_m)$ est un nouveau prédicat de degré k . Un tel prédicat nomme l'attribut complexe de degré k dont l'extension est la suivante : Une séquence de k individus $\langle u_1, \dots, u_k \rangle$ satisfait cet attribut complexe dans un contexte i si et seulement si les deux séquences d'individus $\langle u_1, \dots, u_n \rangle$ et $\langle u_{n+1}, \dots, u_{n+m} \rangle$ satisfont respectivement dans le même contexte les deux attributs plus simples exprimés par R_n et R_m . L'attribut complexe $\langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n \wedge R_m)$ sera appelé une *conjonction transformationnelle* des deux attributs plus simples. En particulier, l'attribut exprimé par le prédicat $\langle 1, \dots, n \rangle \langle n+1, \dots, n+m \rangle (R_n \wedge R_m)$ est la *conjonction Booléenne* classique de degré $n + m$ des deux attributs nommés par R_n et R_m . Une séquence $\langle u_1, \dots, u_{n+m} \rangle$ satisfait en effet cette conjonction dans un contexte si et seulement si les deux séquences $\langle u_1, \dots, u_n \rangle$ et $\langle u_{n+1}, \dots, u_{n+m} \rangle$ satisfont respectivement dans ce même contexte les deux attributs exprimés par les prédicats plus simples R_n et R_m .

Soit dit en passant, la formation de ce nouveau type de prédicat complexe nous permet d'exprimer la *réflexivisation*, la *conversion*, l'*expansion* et la *permutation* aussi bien que la *conjonction* Booléenne et les autres opérations classiques de l'algèbre de Peirce et de Schroeder des relations (qu'on appelle aussi l'*algèbre des transformations*). Par exemple, le prédicat $\langle 1, 1 \rangle \langle 1, 1 \rangle (=_2 \wedge =_2)$ exprime la propriété d'être identique à soi-même qui est la réflexivisation de la relation binaire d'identité. De même, le nouveau prédicat $\langle 1, 2 \rangle \wedge \langle 1, 3 \rangle (=_2, \neg =_2)$ nomme la relation ternaire qui existe entre trois individus si et seulement si le premier est identique au deuxième et différent du troisième. Et finalement le prédicat $\langle 2, 1 \rangle \langle 2, 1 \rangle (r_2 \wedge r_2)$ exprime la converse de la relation binaire exprimée par r_2 .

26. Ainsi $\exists_k$ est l'analogue intensionnel de l'opérateur spécial exprimant la quantification existentielle dans les algèbres polyadiques.

27. De même, l'attribut nommé par un prédicat de la forme $\forall_k R_n$ où $\forall_k = \neg \exists_k \neg$ sera appelé une *quantification universelle* de celui nommé par R_n .

L'ensemble L_p des termes propositionnels

Si R_n est un prédicat de degré n et $t_1, \dots, t_n$ est une suite de n termes individuels, alors $(R_n(t_1, \dots, t_n))$ est un *terme propositionnel*. Ainsi, $(R_0()) \in L_p$ quand R_0 est un prédicat de degré 0.

Si R_n est un prédicat du lexique, $(R_n(t_1, \dots, t_n))$ exprime la *proposition élémentaire* qui est vraie dans un contexte si et seulement si la séquence des n objets individuels qui tombent dans ce contexte sous les concepts exprimés par $t_1, \dots, t_n$ appartient à l'extension dans ce même contexte de l'attribut exprimé par R_n .

Quand le prédicat R_n est la constante logique d'identité $=_2$, nous écrirons souvent $(t_1 = t_2)$ au lieu de $(=_2(t_1, t_2))$ pour respecter la notation ordinaire du calcul des prédicats. Un terme de la forme $(t_1 = t_2)$ nomme la *proposition élémentaire d'identité* qui est vraie dans un contexte si et seulement si les concepts individuels exprimés par t_1 et t_2 s'appliquent au même individu dans ce contexte.

Un terme propositionnel de la forme $([t/k]R_n(t_1, \dots, t_{k-1}, t_{k+1}, t_n))$ exprime la même proposition que le terme $(R_n(t_1, \dots, t_{k-1}, t, t_{k+1}, t_n))$.

Le terme propositionnel $(\exists_k R_n(t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_n))$ exprime la *proposition existentielle* obtenue en généralisant à la k -ième place la proposition exprimée par $(R_n(t_1, \dots, t_n))$.

Le terme propositionnel $(\neg R_n(t_1, \dots, t_n))$ exprime la *négation* de la proposition exprimée par $(R_n(t_1, \dots, t_n))$. Semblablement, le terme propositionnel $(<l_1, \dots, l_n><l_{n+1}, \dots, l_{n+m}>(R_n \wedge R_m')(t_1, \dots, t_k))$ exprime la *conjonction* des propositions exprimées par les termes $(R_n(t_1, \dots, t_n))$ et $(R_m'(t_{n+1}, \dots, t_{n+m}))$.

L'ensemble L des énoncés

Les énoncés élémentaires de L

Si A_p et B_p sont des termes propositionnels, alors *Vrai*(A_p), $(A_p = B_p)$ et $(A_p \succ B_p)$ sont des *énoncés élémentaires* de L .

L'énoncé *Vrai*(A_p) signifie que la proposition exprimée par A_p est vraie. L'énoncé $(A_p = B_p)$ signifie que A_p et B_p expriment la même proposition. Finalement, l'énoncé $A_p \succ B_p$ signifie que la proposition exprimée par A_p implique cognitivement de façon faible la proposition exprimée par B_p . Dans ce cas, comme je l'ai expliqué plus haut, la proposition exprimée par B_p peut être dérivée de la proposition exprimée par A_p en appliquant les règles d'inférence suivantes qui ont une réalité psychologique :

La règle d'inférence tautologique :

D'une proposition, dériver toute conséquence tautologique de cette proposition ;

La règle de substitution des identiques :

D'une proposition d'identité $(t_1 = t_2)$, dériver toute proposition de la forme $A_p \Leftrightarrow [t_1/t_2]A_p$ où $[t_1/t_2]A_p$ est obtenu en remplaçant toute occurrence de t_1 dans A_p par t_2 .

La règle de généralisation existentielle :

D'une proposition quelconque de la forme $(R_n(t_1, \dots, t_n))$ dériver toute proposition existentielle de la forme $(\exists_k R_n(t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_n))$ où $k \leq n$ ²⁸.

Les énoncés complexes de $\mathcal{L}$

Si A et B sont des énoncés de $\mathcal{L}$ et v est une variable, alors $\neg A$, $\exists v(A)$, $\Box A$ et $(A \wedge B)$ sont de nouveaux énoncés complexes de $\mathcal{L}$ à interpréter de la façon habituelle. Ainsi, un énoncé de la forme $\neg A$ est vrai dans un contexte i selon une interprétation si et seulement si A n'est pas vrai dans ce contexte selon cette interprétation. Un énoncé de la forme $\Box A$ est vrai dans un contexte i selon une interprétation si et seulement si A est vrai dans tous les contextes selon cette interprétation. Et semblablement pour les énoncés des autres formes.

Règles d'abréviation

On peut dériver par l'application des règles d'abréviation suivantes les autres notions fondamentales de notre logique des propositions généralisées.

Transformation d'un attribut :

$\langle l_1, \dots, l_n \rangle R_n =_{df} \langle l_1, \dots, l_n \rangle \langle l_1, \dots, l_n \rangle (R_n \wedge R_n)$, où R_n est un prédicat de degré n et $\langle l_1, \dots, l_n \rangle$ est une séquence de nombres tels que $\{l_1, \dots, l_n\} = \{1, \dots, k\}$ et $1 \leq k \leq n$.

Réflexivisation :

$Ref(R_n) =_{df} \langle 1, \dots, n-1, n-1 \rangle \langle 1, \dots, n-1, n-1 \rangle (R_n \wedge R_n)$

Conversion :

$Conv(R_n) =_{df} \langle 2, \dots, n, 1 \rangle \langle 2, \dots, n, 1 \rangle (R_n \wedge R_n)$

Expansion :

$Exp(R_n) =_{df} \langle 1, \dots, n \rangle \langle n+1, n+1 \rangle (R_n \wedge =_2)$

Inversion :

$Inv(R_n) =_{df} \langle 1, \dots, n-1, n, n-2 \rangle \langle 1, \dots, n-1, n, n-2 \rangle (R_n \wedge R_n)$

Négation propositionnelle :

$\neg(R_n(t_1, \dots, t_n)) =_{def} (\neg R_n(t_1, \dots, t_n))$

28. Cette règle est équivalente à la règle d'instanciation universelle: D'une proposition universelle de la forme $(\forall_k R_n(t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_n))$ où $k \leq n$, dériver n'importe quelle instance particulière de substitution $R_n(t_1, \dots, t_{k-1}, t, t_{k+1}, \dots, t_n)$ de cette proposition.

Conjonction propositionnelle :

$(R_n(t_1, \dots, t_n)) \& (R_m'(t'_1, \dots, t'_m)) =_{\text{def}} \langle [t_1], \dots, [t_n] \rangle \langle [t'_1], \dots, [t'_m] \rangle$
 $(R_n, R_m')(t_{j_1}, \dots, t_{j_k})$, où $t_{j_1}, \dots, t_{j_k}$ est la séquence canonique de l'ensemble $\{t_1, \dots, t_n, t'_1, \dots, t'_m\}$ et $[\]$ est la fonction canonique de cet ensemble.

Quantification propositionnelle existentielle :

Soit $R_n(t_1, \dots, t_n)$ un terme propositionnel et $\langle i_1, \dots, i_j \rangle$ la séquence des nombres ($i < i_k + 1$) indiquant où la variable v a une occurrence dans la séquence des termes individuels $\langle t_1, \dots, t_n \rangle$. Autrement dit, $t_{i_1}, t_{i_2}, \dots, t_{i_j}$ sont tous les termes individuels de la séquence qui sont identiques à la variable v .

— Quand $j = 0$ (c'est-à-dire quand $v \notin \{t_1, \dots, t_n\}$),

$$\exists v(R_n(t_1, \dots, t_n)) =_{\text{def}} (R_n(t_1, \dots, t_n)).$$

— Autrement, $\exists v(R_n(t_1, \dots, t_n)) =_{\text{def}} (\exists i_1 \langle i_1, \dots, i_2 - 1, i_1, i_2, i_2 + 1, \dots, i_j - 1, i_1, i_j, i_j + 1, \dots, n - j + 1 \rangle R_n(t_1, \dots, t_{i_1 - 1}, t_{i_1 + 1}, \dots, t_{i_2 - 1}, t_{i_2 + 1}, \dots, t_{i_j - 1}, t_{i_j + 1}, \dots, t_n))$.

Disjonction :

$A \vee B =_{\text{df}} \neg(\neg A \wedge \neg B)$ quand $A, B \in L_p$ ou $A, B \in L$.

Implication matérielle :

$$A \Rightarrow B =_{\text{df}} \neg A \vee B$$

Équivalence matérielle :

$$A \Leftrightarrow B =_{\text{df}} (A \Rightarrow B) \wedge (B \Rightarrow A)$$

Quantification universelle :

$$\forall_k R_n =_{\text{df}} \neg \exists_k \neg R_n \text{ et } \forall v(A) =_{\text{df}} \neg \exists v(\neg A)$$

Une proposition élémentaire nécessaire :

$$1_p =_{\text{df}} ((\exists_1 \exists_2 =_2)(\))$$

Implication stricte :

$$A \prec B =_{\text{df}} \Box(A \Rightarrow B)$$

Équivalence stricte :

$$A \succ\!\!\prec B =_{\text{df}} (A \prec B) \wedge (B \prec A)$$

Tautologies :

$$\mathbf{T}(A_p) =_{\text{df}} A_p = (A_p \Rightarrow A_p)$$

Inclusion de contenu

Si A_p et B_p sont deux termes propositionnels, on peut former dans notre langue formelle un énoncé $(A_p) B_p$ signifiant que toutes les propositions atomiques de B_p sont aussi des propositions atomiques de A_p . Il suffit d'appliquer les règles d'abréviation suivantes :

$r_m'(t_1', \dots, t_m') \rangle r_n(t_1, \dots, t_n) =_{\text{def}} r_m'(t_1', \dots, t_m') = r_n(t_1, \dots, t_n)$, où r_n est un prédicat simple du lexique de degré n ou un prédicat complexe de la forme $\exists i R_{n+1}$.

$$\begin{aligned}
[t/k]R_m(t_1', \dots, t_{m-1}') \rangle r_n(t_1, \dots, t_n) &=_{\text{def}} \\
R_m(t_1', \dots, t_{k-1}', t, t_k', \dots, t_{m-1}') \rangle r_n(t_1, \dots, t_n). \\
\neg R_m(t_1', \dots, t_m') \rangle r_n(t_1, \dots, t_n) &=_{\text{def}} R_m(t_1', \dots, t_m') \rangle r_n(t_1, \dots, t_n). \\
\langle l_1, \dots, l_i \rangle \langle l_{i+1}, \dots, l_{i+j} \rangle (R_i \wedge R_j)(t_1', \dots, t_k') \rangle r_n(t_1, \dots, \\
t_n) &=_{\text{def}} (R_i(t_{l_1}', \dots, t_{l_i}')) \rangle r_n(t_1, \dots, t_n) \vee (R_j(t_{l_{i+1}}', \dots, t_{l_{i+j}}')) \rangle r_n(t_1, \dots, t_n). \\
R_m'(t_1', \dots, t_m') \rangle [t/k]R_n(t_1, \dots, t_{n-1}) &=_{\text{def}} \\
R_m'(t_1', \dots, t_m') \rangle R_n(t_1, \dots, t_{k-1}, t, t_k, \dots, t_{n-1}). \\
R_m'(t_1', \dots, t_m') \rangle \neg R_n(t_1, \dots, t_n) &=_{\text{def}} \\
R_m'(t_1', \dots, t_m') \rangle R_n(t_1, \dots, t_n). \\
R_i'(t_1', \dots, t_i') \rangle \langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n, R_m)(t_1, \dots, t_k) &=_{\text{def}} \\
(R_i'(t_1', \dots, t_i') \rangle R_n(t_1, \dots, t_{l_n})) \wedge R_i'(t_1', \dots, t_i') \rangle R_m(t_{l_{n+1}}, \dots, t_{l_{n+m}}))
\end{aligned}$$

Implication forte :

$$A_p \vdash B_p =_{\text{def}} (A_p \rangle B_p) \wedge T(A_p \Rightarrow B_p)$$

Constituants propositionnels

On peut former avec chaque terme propositionnel A_p et un prédicat ou terme individuel quelconque C un énoncé $A_p(C)$ signifiant que l'attribut ou concept individuel exprimé par C est un constituant propositionnel de la proposition A_p .

$$\begin{aligned}
(r_n(t_1, \dots, t_n))(C) &=_{\text{def}} C = r_n \vee C = t_1 \vee \dots \vee C = t_n \\
[t/k]R_m(t_1, \dots, t_{m-1})(C) &=_{\text{def}} (C = t_1) \vee \dots (C = t_{m-1}) \vee (C = t) \\
\neg R_m(t_1', \dots, t_m')(C) &=_{\text{def}} R_m(t_1, \dots, t_m)(C)
\end{aligned}$$

Et ainsi de suite.

Expansion cognitive du contenu

On peut former avec deux termes propositionnels A_p et B_p quelconques un énoncé $(A_p \supset B_p)$ signifiant que toutes les propositions atomiques de B_p appartiennent à l'expansion cognitive du contenu de A_p .

$$\begin{aligned}
A_p \supset (r_n(t_1, \dots, t_n)) &=_{\text{def}} A_p(r_n) \wedge A_p(t_1) \wedge \dots \wedge A(t_n) \\
A_p \supset (\neg R_n(t_1, \dots, t_n)) &=_{\text{def}} A_p \supset (R_n(t_1, \dots, t_n)) \\
A_p \supset ([t/k]R_n(t_1, \dots, t_{k-1}, t_k+1, t_n)) &=_{\text{def}} A_p \supset \\
(R_n(t_1, \dots, t_{k-1}, t, t_k+1, t_n)) \\
A_p \supset (\exists_k R_n)(t_1, \dots, t_{k-1}, t_k+1, \dots, t_n) &=_{\text{def}} \\
A_p \supset R_n(t_1, \dots, t_{k-1}, t_1, t_k+1, \dots, t_n) \\
A_p \supset (\langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n \wedge R_m')(t_1, \dots, t_k)) &=_{\text{def}} A_p \supset
\end{aligned}$$

$$(R_n(t_{i_1}, \dots, t_{i_n})) \wedge A_p \supset (R_m'(t_{i_{n+1}}', \dots, t_{i_{n+m}}'))$$

Implication cognitive forte:

$$A_p \succ_s B_p =_{\text{df}} A_p \succ B_p \wedge (A_p \supset B_p)$$

3. La sémantique formelle

Un *modèle standard* $\mathcal{M}$ pour $\mathcal{L}$ est un triplet $\langle I, U_e, \|\cdot\| \rangle$ ayant la forme suivante : I et U_e sont deux ensembles non vides dont les éléments représentent respectivement des contextes (ou mondes) possibles et des objets individuels.

$\|\cdot\|$ est une fonction qui associe à chaque formule A (prédicat ou terme) de $\mathcal{L}$ relativement à chaque assignation σ de valeurs dans U_e à des variables, une signification qui est la valeur sémantique de A sous l'assignation σ dans le modèle $\mathcal{M}$. Cette valeur (en symboles : $\|A\|^\sigma$) satisfait les clauses suivantes :

(i) Pour chaque variable v , $\|v\|^\sigma \in (U_e)^I$. C'est une fonction constante telle que, pour tout $i \in I$, $\|v\|^\sigma(i) = \sigma(v)$.

(ii) Pour chaque constante individuelle c , $\|c\|^\sigma$ est une fonction quelconque de l'ensemble $(U_e)^I$ telle que, pour chaque assignation σ' , $\|c\|^\sigma = \|c\|^{\sigma'}$.

(iii) $\|R_n\|^\sigma \in (2^{(U_e)^n})^I$ où $2 = \{0, 1\}$. En particulier, $\|_{=2}\|^\sigma(i)(u_1, u_2) = 1$ si et seulement si $u_1 = u_2$.

(iv) Pour chaque $u_1, \dots, u_n \in U_e$, $\|\neg R_n\|^\sigma(i) \langle u_1, \dots, u_n \rangle = 1$ si et seulement si $\|R_n\|^\sigma(i) \langle u_1, \dots, u_n \rangle = 0$.

(v) $\|[t/k]R_n\|^\sigma(i) \langle u_1, \dots, u_{n-1} \rangle = 1$ si et seulement si $\|R_n\|^\sigma(i) \langle u_1, \dots, u_{n-1}, \|t\|^\sigma(i), u_n \rangle = 1$.

(vi) $\|\exists_k R_n\|^\sigma(i) \langle u_1, \dots, u_{n-1} \rangle = 1$ si et seulement si il existe un objet $u \in U_e$ tel que $\|R_n\|^\sigma(i) \langle u_1, \dots, u_{n-1}, u, u_n \rangle = 1$.

(vii) $\|\langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n \wedge R_m')\|^\sigma(i) \langle u_1, \dots, u_k \rangle = 1$ si et seulement si $(\|R_n\|^\sigma(i) \langle u_1, \dots, u_{i_n} \rangle = 1 \text{ et } \|R_m'\|^\sigma(i) \langle u_{i_{n+1}}, \dots, u_{i_{n+m}} \rangle = 1)$.

Définition

Soit $U_a = \{ \|A_a\|^\sigma / A_a \in L_a \text{ et } \sigma \in U_e^{\text{Var}} \}$ l'ensemble de toutes les *propositions atomiques* qui sont exprimables dans le modèle $\mathcal{M}$, à savoir l'ensemble de toutes les entités de la forme $\|R_n\|^\sigma (\|t_1\|^\sigma, \dots, \|t_n\|^\sigma)$ qui sont des paires ordonnées de la forme

$$< \{ \|\mathbf{R}_n\|^\sigma, \|\mathbf{t}_1\|^\sigma, \dots, \|\mathbf{t}_n\|^\sigma \}, \{ i / \|\mathbf{R}_n\|^\sigma(i) < \|\mathbf{t}_1\|^\sigma(i), \dots, \|\mathbf{t}_n\|^\sigma(i) = 1 \} >.$$

Comme dans la logique minimale des propositions, une *proposition complète* est une paire ordonnée dont le premier terme est l'ensemble non vide de ses propositions atomiques et le deuxième terme l'ensemble de toutes les assignations de valeurs de vérité à des propositions atomiques qui la rendent vraie.

Sur la base des remarques précédentes, l'ensemble U_p de toutes les propositions complètes qui sont exprimées par des termes propositionnels dans un modèle $\mathcal{M}$ est défini par l'induction suivante :

(vii) $\|(\mathbf{r}_n(\mathbf{t}_1, \dots, \mathbf{t}_n))\|^\sigma = < \{A\}, \{f \in \{0, 1\}^{U_a} : f(A) = 1\} >$ où la proposition atomique $A = < \{\|\mathbf{r}_n\|^\sigma, \|\mathbf{t}_1\|^\sigma, \dots, \|\mathbf{t}_n\|^\sigma\}, \{i \in I / \|\mathbf{r}_n\|^\sigma(i) (\|\mathbf{t}_1\|^\sigma(w), \dots, \|\mathbf{t}_n\|^\sigma(i)) = 1\} >$.

$$(viii) \ \|([t/k]\mathbf{R}_n(\mathbf{t}_1, \dots, \mathbf{t}_{n-1}))\|^\sigma = \|\mathbf{R}_n(\mathbf{t}_1, \dots, \mathbf{t}_{k-1}, t, \mathbf{t}_k, \dots, \mathbf{t}_{n-1})\|^\sigma.$$

(ix) $\|(\exists_k \mathbf{R}_n(\mathbf{t}_1, \dots, \mathbf{t}_{n-1}))\|^\sigma = < \{A\}, \{f \in \{0, 1\}^{U_a} : f(A) = 1\} >$ où la proposition atomique $A = < \{\|\exists_k \mathbf{R}_n\|^\sigma, \|\mathbf{t}_1\|^\sigma, \dots, \|\mathbf{t}_{n-1}\|^\sigma\}, \{i \in I / \|\exists_k \mathbf{R}_n\|^\sigma(i) (\|\mathbf{t}_1\|^\sigma(i), \dots, \|\mathbf{t}_{n-1}\|^\sigma(i)) = 1\} >$

Notation :

Si $\langle a, b \rangle$ est une paire ordonnée, soient $id_1(\langle a, b \rangle)$ le premier terme a et $id_2(\langle a, b \rangle)$ le deuxième terme b de cette paire.

$$(x) \ \|(\neg \mathbf{R}_n(\mathbf{t}_1, \dots, \mathbf{t}_n))\|^\sigma = < id_1(\|(\mathbf{R}_n(\mathbf{t}_1, \dots, \mathbf{t}_n))\|^\sigma), \{0, 1\}^{U_a} - id_2(\|(\mathbf{R}_n(\mathbf{t}_1, \dots, \mathbf{t}_n))\|^\sigma) >.$$

$$(xi) \ \|(\langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (\mathbf{R}_n, \mathbf{R}_m')(\mathbf{t}_1, \dots, \mathbf{t}_k))\|^\sigma = < id_1(\|(\mathbf{R}_n(\mathbf{t}_1, \dots, \mathbf{t}_{1_n}))\|^\sigma) \cup id_1(\|(\mathbf{R}_m'(\mathbf{t}_{1_{n+1}}, \dots, \mathbf{t}_{1_{n+m}}))\|^\sigma), id_2(\|(\mathbf{R}_n(\mathbf{t}_1, \dots, \mathbf{t}_{1_n}))\|^\sigma) \cap id_2(\|(\mathbf{R}_m'(\mathbf{t}_{1_{n+1}}, \dots, \mathbf{t}_{1_{n+m}}))\|^\sigma) >.$$

La définition des conditions de vérité des énoncés

Si $\mathcal{M}$ est un modèle, σ une assignation de valeurs à des variables et i un contexte, la valeur de vérité de l'énoncé A dans le contexte i sous l'assignation σ selon le modèle $\mathcal{M}$ (en symboles: $\|\mathbf{A}\|_i^\sigma$) est définie inductivement comme suit :

$$\text{— } \|\mathbf{A}_p\| = \mathbf{B}_p\|_i^\sigma \text{ est le vrai si et seulement si } \|\mathbf{A}_p\|^\sigma = \|\mathbf{B}_p\|^\sigma.$$

— $\|Vrai(A_p)\|_i^\sigma$ est *le vrai* si et seulement si, selon l'interprétation possible $\mathcal{M}$, il existe au moins une fonction $f \in id_2(\|A_p\|^\sigma)$ telle que, pour chaque proposition atomique $u_a \in id_1(\|A_p\|^\sigma)$, $f(u_a) = 1$ si et seulement si $i \in id_2(u_a)$.

— $\|A \succ B\|_i^\sigma$ est *le vrai* si et seulement si $Int[id_2(\|A_p\|^\sigma)] \subseteq id_2(\|B\|^\sigma)$, où pour n'importe quel ensemble d'assignations de valeurs de vérité à des propositions atomiques $\mathcal{F} \subseteq \{0, 1\}^{U_a}$, l'*Intérieur* de $\mathcal{F}$, $Int[\mathcal{F}]$, est l'ensemble de toutes les assignations $f \in \mathcal{F}$ qui sont compatibles avec les postulats de signification gouvernant les constantes logiques d'identité et de quantification. Plus précisément, une assignation f de $\mathcal{F}$ appartient à $Int[\mathcal{F}]$ si et seulement si

1) $f(\|R_n\|^\sigma(\|t_1\|^\sigma, \dots, \|t_n\|^\sigma)) = f(\|R_n\|^\sigma(\|t_1'\|^\sigma, \dots, \|t_n'\|^\sigma))$ quand $f(\|=\|^\sigma(\|t_1\|^\sigma, \|t_2'\|^\sigma)) = 1$ et un terme t_k' est différent de t_k si et seulement si $t_k = t$ et $t_k' = t'$ et

2) $f(\|(\exists_k R_n)\|^\sigma(\|t_1\|^\sigma, \dots, \|t_n\|^\sigma)) = 1$ quand $1 \leq k \leq n$ et $f(\|R_n\|^\sigma(\|t_1\|^\sigma, \dots, \|t_n\|^\sigma)) = 1$.

— Les conditions de vérité des énoncés complexes sont définies de la façon habituelle. Ainsi $\|\forall A\|_i^\sigma$ est *le vrai* si et seulement si $\|A\|_i^{\sigma'}$ est *le vrai* pour au moins une assignation σ' qui diffère au plus de σ par le fait que $\sigma(v) \neq \sigma'(v)$.

Définition de la validité

Un énoncé A de $\mathcal{L}$ est *nécessairement vrai dans un modèle* $\mathcal{M} = \langle I, U_e, \|\cdot\| \rangle$ de $\mathcal{L}$ si et seulement si $\|A\|_i^\sigma$ est *le vrai* pour tout contexte $i \in I$ et toute assignation σ de valeurs aux variables. L'énoncé A est *valide* (ou *logiquement vrai*) quand il est nécessairement vrai dans tous les modèles standards $\mathcal{M}$ de $\mathcal{L}$.

4. Un système axiomatique

Je conjecture que le système axiomatique S suivant²⁹ est à la fois complet et valide. Ses théorèmes expriment toutes les lois logiquement vraies qui gouvernent les propositions qui sont exprimables dans la langue formelle $\mathcal{L}$.

29. Ce système formel est une extension conservative du calcul de logique propositionnelle minimale que j'ai défini dans "A new formulation of the logic of propositions" à paraître dans *Quebec Studies in the Philosophy of Science*, Volume 1 *Logic, Mathematics, Physics* in the *Boston Studies in the Philosophy of Science*, Kluwer, 1995.

Les axiomes du système S sont :

- les axiomes de la logique des connecteurs de vérité, et du calcul des prédicats du premier ordre³⁰,
- les axiomes de la logique modale S5³¹,
- les axiomes suivants d'identité propositionnelle :

Schéma d'axiomes 1 :

$$A_p = A_p$$

Schéma d'axiomes 2 :

$$(A_p \vdash B_p \wedge B_p \vdash A_p) \Rightarrow A_p = B_p$$

Schéma d'axiomes 3 :

$$\square (Vrai(r_n(t_1, \dots, t_n)) \Leftrightarrow Vrai(r_n(t_1', \dots, t_n')) \Rightarrow (r_n(t_1, \dots, t_n) = r_n(t_1', \dots, t_n')), \text{ quand } \{t_1, \dots, t_n\} = \{t_1', \dots, t_n'\}.$$

Schéma d'axiomes 4 :

$$\neg(r_n(t_1, \dots, t_n)) = r_m(t_1', \dots, t_m')), \text{ quand } n \neq m.$$

Schéma d'axiomes 5 :

$$[t/k]R_n(t_1, \dots, t_{n-1}) = R_n(t_1, \dots, t_{k-1}, t, t_k, \dots, t_{n-1})$$

Schéma d'axiomes 6 :

$$\langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n, R_m')(t_1, \dots, t_k) = \langle [t_1], \dots, [t_n] \rangle \langle [t_{n+1}], \dots, [t_{n+m}] \rangle (R_n, R_m')(t_{j1}, \dots, t_{ji}),$$

où $\langle t_{j1}, \dots, t_{ji} \rangle$ est la séquence canonique des termes $t_1, \dots, t_k$.

Le schéma d'axiomes 2 dit que deux propositions qui s'impliquent fortement sont identiques. Les schémas d'axiomes 3 et 4 décrivent des conditions d'identité pour les propositions dont l'attribut est de forme simple ou une généralisation. Enfin, les schémas d'axiomes 5 et 6 spécifient des formes normales pour les propositions qui sont des attributs complexes comme le résultat de l'application d'un attribut de degré supérieur ou une conjonction.

- les axiomes suivants pour les propositions identiques³² :

Schéma d'axiomes 7 :

$$(r_n(t_1, \dots, t_n) = r_n'(t_1', \dots, t_n')) \Rightarrow (r_n(\tau_1, \dots, \tau_n) = r_n'(\tau_1, \dots, \tau_n))$$

30. Voir Church, *Introduction to Mathematical Logic*, Volume 1, Princeton University Press, 1956.

31. Voir C.I. Lewis, *A Survey of Symbolic Logic*, Berkeley et Los Angeles, University of California Press, 1918.

32. Comme d'habitude, m est une variable métalinguistique pour prédicats de degré n du lexique et prédicats complexes de la forme $\exists k R_{n+1}$ composé d'un prédicat R_{n+1} de degré $n+1$.

Schéma d'axiomes 8 :

$$(\Box \text{Vrai}(t_1 = t_1') \wedge \dots \wedge \Box \text{Vrai}(t_n = t_n')) \Rightarrow ((r_n(\tau_1, \dots, \tau_n) = r_n'(\tau_1, \dots, \tau_n)) \Rightarrow (r_n(t_1, \dots, t_n) = r_n'(t_1', \dots, t_n')))$$

Schéma d'axiomes 9 :

$$(r_n(t_1, \dots, t_n) = r_n'(t_1', \dots, t_n')) \Rightarrow (\Box \text{Vrai}(t_i = t_1') \vee \dots \vee \Box \text{Vrai}(t_i = t_n')), \text{ où } 1 \leq i \leq n.$$

Schéma d'axiomes 10 :

$$A_p = B_p \Rightarrow (C_p = [B_p/A_p]C_p)$$

Schéma d'axiomes 11 :

$$A_p = B_p \Rightarrow \Box(A_p = B_p)$$

Schéma d'axiomes 12 :

$$A_p = B_p \Rightarrow (A \Leftrightarrow [B_p/A_p]A), \text{ où } A \text{ est un énoncé quelconque.}$$

Les schémas d'axiomes 8-12 décrivent des conditions que doivent satisfaire les propositions identiques. En particulier, le schéma d'axiomes 11 dit que toute proposition a les mêmes propositions atomiques dans tous les contextes.

— les axiomes suivants pour les conditions de vérité :

Schéma d'axiomes 13 :

$$\text{Vrai}(t = t)$$

Schéma d'axiomes 14 :

$$\text{Vrai}(t = t') \Rightarrow (\text{Vrai}(R_n(t_1, \dots, t_{k-1}, t, t_{k+1}, \dots, t_n)) \Rightarrow \text{Vrai}(R_n(t_1, \dots, t_{k-1}, t', t_{k+1}, \dots, t_n)))$$

Schéma d'axiomes 15 :

$$(R_n(t_1, \dots, t_n) = R_m'(t_1', \dots, t_m')) \Rightarrow (\text{Vrai}(R_n(t_1, \dots, t_n)) \Leftrightarrow \text{Vrai}(R_m'(t_1', \dots, t_m')))$$

Schéma d'axiomes 16 :

$$\text{Vrai}(\neg R_n(t_1, \dots, t_n)) \Leftrightarrow \neg \text{Vrai}(R_n(t_1, \dots, t_n))$$

Schéma d'axiomes 17 :

$$\text{Vrai}(\langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n, R_m')(t_1, \dots, t_k)) \Leftrightarrow (\text{Vrai}(R_n(t_{l_1}, \dots, t_{l_n})) \wedge \text{Vrai}(R_m'(t_{l_{n+1}}, \dots, t_{l_{n+m}}))),$$

Schéma d'axiomes 18 :

$$\text{Vrai}(\exists_k R_n(t_1, \dots, t_{n-1})) \Leftrightarrow \exists v (\text{Vrai}(R_n(t_1, \dots, t_{k-1}, v, t_k, \dots, t_{n-1}))),$$

où v est différent des $t_1, \dots, t_{n-1}$.

Le schéma d'axiomes 13 affirme la vérité de toute proposition élémentaire selon laquelle un objet est identique à lui-même. Les schémas d'axiomes 14-18 décrivent

comment les valeurs de vérité des propositions de certaines formes sont logiquement liées. Ainsi le schéma d'axiomes 14 dit que la substitution des termes individuels ayant même dénotation dans les termes propositionnels préserve la vérité des propositions exprimées par ces termes³³. Les schémas d'axiomes 16, 17 et 18 disent comment déterminer la valeur de vérité des propositions dont l'attribut est la négation, la quantification ou la conjonction d'attributs plus simples à partir de la valeur de vérité de certaines propositions plus élémentaires prédisquant les attributs plus simples en question aux mêmes objets de référence. Ce faisant, ces trois schémas d'axiomes établissent un pont entre les opérations de négation, de conjonction et de généralisation sur les attributs des propositions et les opérations correspondantes de négation, de conjonction et de généralisation sur les énoncés.

— les axiomes suivants pour les tautologies :

Schéma d'axiomes 19 :

TA_p , quand A_p a la forme d'une tautologie selon la méthode des tables de vérité.

Schéma d'axiomes 20 :

$$(TA_p \wedge T(A_p \Rightarrow B_p)) \Rightarrow T(B_p)$$

Dans les trois axiomes suivants, les termes propositionnels $A_p^1, \dots, A_p^n$ et $B_p^1, \dots, B_p^m$ sont tous de la forme $r_n(t_1, \dots, t_n)$.

Schéma d'axiomes 21 :

$$\neg T(A_p^1 \vee \dots \vee A_p^n)$$

Schéma d'axiomes 22 :

$$\neg T(\neg A_p^1 \vee \dots \vee \neg A_p^n)$$

Schéma d'axiomes 23 :

$$T(A_p^1 \vee \dots \vee A_p^n \vee \neg B_p^1 \vee \dots \vee \neg B_p^m) \Rightarrow (((A_p^1 = B_p^1) \vee \dots \vee (A_p^1 = B_p^m) \vee \dots \vee (A_p^n = B_p^1) \vee \dots \vee (A_p^n = B_p^m))),$$

Le schéma d'axiomes 19 dit comment reconnaître les tautologies. Le schéma 20 affirme que l'ensemble des tautologies est fermé sous le *modus ponens*. Les schémas d'axiomes 21 et 22 affirment qu'aucune proposition élémentaire n'est une tautologie ou une contradiction. Finalement, le schéma d'axiomes 23 décrit la forme canonique des tautologies.

Les axiomes suivants pour l'implication cognitive faible

33. N'oublions pas que les attributs des propositions exprimables dans notre logique sont tous *extensionnels*.

Schéma d'axiomes 24 :

$$(A_p \succ B_p) \Rightarrow ((B_p \succ C_p) \Rightarrow (A_p \succ C_p))$$

Schéma d'axiomes 25 :

$$T(A_p \Rightarrow B_p) \Rightarrow A_p \succ B_p$$

Schéma d'axiomes 26 :

$(t = t') \succ ((R_n(t_1, \dots, t_n)) \Leftrightarrow (R_n(t'_1, \dots, t'_n)))$, quand les deux suites $t_1, \dots, t_n$ et $t'_1, \dots, t'_n$ diffèrent au plus par le fait que $t_k = t$ et $t'_k = t'$.

Schéma d'axiomes 27 :

$$(R_n(t_1, \dots, t_{k-1}, t, t_{k+1}, \dots, t_{n-1})) \succ (\exists_k R_n(t_1, \dots, t_{n-1}))$$

Schéma d'axiomes 28:

$(\neg(T(r_n(t_1, \dots, t_n) \Rightarrow R_m'(t'_1, \dots, t'_m))) \wedge \neg(R_m(t'_1, \dots, t'_m) = \exists_k R_n(t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_n))) \Rightarrow \neg(r_n(t_1, \dots, t_n) \succ R_m'(t'_1, \dots, t'_m))$, où $1 \leq k \leq n$ et $r_n \neq 2$.

Les schémas d'axiomes 24-27 décrivent comment s'engendre l'engagement cognitif à des propositions. Le schéma d'axiomes 28 dit que l'engagement cognitif entre propositions s'engendre seulement de cette façon.

Les trois règles d'inférence de notre système axiomatique sont :

la règle du modus ponens : De A_p et de $A_p \Rightarrow B_p$ inférer B_p ;

la règle de nécessité : D'un théorème A_p inférer $\Box A_p$; et

la règle de généralisation : D'un théorème A_p inférer $\forall v A_p$.

Les notions de preuve et de dérivation sont définies de la façon habituelle.

5. Lois valides

Voici quelques lois fondamentales qui sont valides dans notre logique des propositions généralisées :

1. Lois concernant les conditions de vérité

$$\models \text{Vrai}(((\neg R_n)(t_1, \dots, t_n))) \Leftrightarrow \text{Vrai}(\neg(R_n(t_1, \dots, t_n)))$$

$\models \text{Vrai}(((\exists_k R_n)(t_1, \dots, t_{n-1}))) \Leftrightarrow \exists v \text{Vrai}((R_n(t_1, \dots, t_{k-1}, v, t_k, \dots, t_{n-1})))$, où v est différent des $t_1, \dots, t_{n-1}$. Et semblablement pour $\forall$.

$$\models \text{Vrai}(\langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n \wedge R_m)(t_1, \dots, t_k)) \Leftrightarrow (\text{Vrai}(R(t_1, \dots, t_n)) \wedge \text{Vrai}((R_n(t_{1_{n+1}}, \dots, t_{1_{n+m}}))))$$

2. Lois concernant l'expansion cognitive du contenu

La relation d'expansion cognitive exprimée par $\nearrow$ est réflexive et transitive.

$$\models (A_p \succ B_p) \Rightarrow (A_p \nearrow B_p)$$

$$\models (R_n(t_1, \dots, t_n)) \nearrow (R_n(t'_1, \dots, t'_n)), \text{ quand } \{t'_1, \dots, t'_n\} \subseteq \{t_1, \dots, t_n\}.$$

$$\models A_p \nearrow (R_n(t_1, \dots, t_n)) \Leftrightarrow A_p \nearrow (\neg R_n(t_1, \dots, t_n))$$

$$\models (R_n(t_1, \dots, t_n)) \nearrow ((\forall_k R_n)(t_1, \dots, t_{k-1}, t_{k+1}, t_n))$$

$$\models A_p \nearrow ((\forall_k R_n)(t_1, \dots, t_{k-1}, t_{k+1}, t_n)) \wedge A_p \nearrow B_p(t) \Rightarrow (A_p \nearrow (R_n(t_1, \dots, t_n))), \text{ quand } t_k = t.$$

$$\models (A_p \nearrow (R_n(t_1, \dots, t_n)) \wedge A_p \nearrow (R_m(t'_1, \dots, t'_m))) \Rightarrow (A_p \nearrow (R_m(\tau_1, \dots, \tau_m))) \text{ quand } \{\tau_1, \dots, \tau_m\} \subseteq \{t_1, \dots, t_n, t'_1, \dots, t'_m\}.$$

$$\models (A_p \nearrow (<l_1, \dots, l_n > <l_{n+1}, \dots, l_{n+m}> (R_n \wedge R_m)(t_1, \dots, t_k)) \Leftrightarrow (A_p \nearrow (R(t_{l_1}, \dots, t_{l_n})) \wedge (A_p \nearrow ((R_n(t_{l_{n+1}}, \dots, t_{l_{n+m}}))))$$

3. Lois d'implication cognitive forte

La relation d'implication cognitive forte est, comme il se doit, une relation réflexive et transitive. Elle est engendrée par l'implication forte.

$$\models (A_p \vdash B_p) \Rightarrow (A_p \succ_s B_p)$$

Elle est engendrée par la substitution des identiques dans une prémisses donnée,

$$\models (A_p \wedge (t_1 = t_2)) \succ_s [t_1/t_2]A_p, \text{ quand il n'y a pas de confusion de variables liées. Elle est engendrée par la généralisation existentielle.}$$

$$(R_n(t_1, \dots, t_n)) \succ_s ((\exists_k R_n)(t_1, \dots, t_{k-1}, t_{k+1}, t_n))$$

Elle est également engendrée par l'instanciation universelle à un objet sous un concept exprimé dans les prémisses.

$$\models (\forall v A_p \wedge B_p(t)) \succ_s [t/v]A_p, \text{ quand il n'y a pas de confusion de variables liées}^{34}.$$

34. Ainsi, lorsque nous exprimons une proposition universelle (comme la proposition que tous les hommes sont mortels) et que nous faisons simultanément un acte de référence à un individu (par exemple, à Napoléon), nous sommes aussi capables en vertu de notre compétence linguistique d'inférer l'instance particulière de substitution concernant cet individu (à savoir que si Napoléon est un homme alors il est mortel). C'est pourquoi l'assertion que tous les hommes sont mortels et que Napoléon est un homme contient l'assertion que Napoléon est mortel. Nous appliquons souvent la loi d'instanciation universelle dans les contextes où nous faisons référence à des objets qui peuvent être assignés comme valeurs à des variables que nous lions universellement.

En particulier,

$$\models (\forall v (A_p \Rightarrow B_p) \wedge ([t/v]A_p) \succ_s [t/v]B_p).$$

$$\models \langle \langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n \wedge R_m)(t_1, \dots, t_k) \rangle \succ_s \\ (R_n(t_1, \dots, t_{l_n})) \wedge (R_m(t_{l_{n+1}}, \dots, t_{l_{n+m}})) \text{ et réciproquement.}$$

$$\text{Cependant, } \not\models t_1 = t_2 \wedge (R_1(t_1)) = (t_1 = t_2 \wedge (R_1(t_2)))$$

Donc, à la différence de l'implication forte, l'implication cognitive forte, elle, n'est pas anti-symétrique.

$\models \exists v_1 \forall v_2 A_p \succ_s \forall v_2 \exists v_1 A_p$ et semblablement pour les autres lois valides d'implication stricte dues à des permutations de quantificateurs³⁵.

Voici quelques nouvelles lois d'identité propositionnelle valides :

$$\models [t/k]R_n(t_1, \dots, t_{k-1}, t_{k+1}, t_n) = R_n(t_1, \dots, t_{k-1}, t, t_{k+1}, t_n)$$

$$\text{En particulier, } \models ([t/2]R_2(t')) = (R_2(t', t)) \text{ et } \models ([t/1]R_2(t')) = (R_2(t, t'))$$

Comme en λ -calcul, disons qu'un terme propositionnel A_p se contracte en un autre B_p si et seulement si B_p est obtenu en remplaçant une partie de A_p de la forme $[t/k]R_n(t_1, \dots, t_{k-1}, t_{k+1}, t_n)$ par $R_n(t_1, \dots, t_{k-1}, t, t_{k+1}, t_n)$. Disons aussi que A_p se réduit en B_p si et seulement si B_p est obtenu à partir de A_p par une série finie de telles contractions. Tout terme propositionnel de la forme $[t/k]R_n(t_1, \dots, t_{k-1}, t_{k+1}, t_n)$ sera appelé un *redex* et $R_n(t_1, \dots, t_{k-1}, t, t_{k+1}, t_n)$ son *contractum*. Il est clair que l'on peut faire un grand nombre de prédications différentes avec un attribut R_n de degré n et n objets de référence. D'une part, l'on peut prédiquer cet attribut de degré n simultanément des n différents objets dans différents ordres. D'autre part, on peut également le prédiquer d'un nombre $k < n$ d'objets en k places différentes. D'un point de vue philosophique, l'attribut de degré $n-k$ qui est ainsi obtenu représente une certaine étape dans le processus de prédiquer l'attribut R_n à n objets de référence en exprimant une proposition. Ainsi on peut réitérer la prédication de l'attribut de degré $n-k$ obtenu à de nouveaux objets de référence.

Un terme propositionnel est en *forme normale* quand il ne contient aucun redex. Par exemple, $(t =_2 t')$ est la forme normale des deux termes $([t/2] =_2 (t'))$ et $([t/1] =_2 (t'))$.

De façon générale, il y a un *théorème de Church-Rosser* pour la prédication en logique propositionnelle. Si un terme A_p se réduit en deux termes différents B_p et C_p , il existe un troisième terme D_p en lesquels B_p et C_p se réduisent.

35. Ainsi la proposition existentielle qu'il existe une femme qui est aimée par tous les hommes implique-t-elle cognitivement de façon forte la proposition universelle plus faible que tous les hommes aiment une femme. C'est pourquoi l'on ne peut affirmer (ou croire) la première proposition généralisée sans également affirmer (ou croire) la seconde.

De même, il y a un *théorème d'unicité de la forme normale*. Tout terme propositionnel se réduit en une forme normale unique.

Ainsi la même proposition est-elle exprimée quand l'on prédique le même attribut de degré n des mêmes objets de référence dans le même ordre de prédication peu importe que cela soit fait simultanément ou en différentes étapes successives. Par exemple, l'on peut prédiquer la relation d'aimer simultanément de Jean et de Marie dans cet ordre. L'on peut également d'abord appliquer la relation d'aimer à Marie à la place du second argument. Ce faisant, on exprime la propriété d'aimer Marie que l'on peut ensuite prédiquer de Jean. Dans les deux cas, on exprime la même proposition en vertu du théorème d'unicité de la forme normale.

Ce théorème est une conséquence du schéma d'axiomes 5.

$$\models \langle l_1, \dots, l_n \rangle R_n(t_1, \dots, t_k) = (R_n(t_1, \dots, t_{l_n})).$$

$$\text{En particulier, } \models (=_2(t, t) = \langle 1, 1 \rangle =_2(t))^{36}.$$

$$\models \forall v_1 A_p = \forall v_2 [v_2/v_1]A_p \text{ en cas d'invariance alphabétique.}$$

$$\models (R_n(t_1, \dots, t_n)) = (\text{Conv}(\text{Conv}(R_n(t_1, \dots, t_n)))$$

$$\models \forall v (B_p \wedge C_p) = (\forall v (B_p) \wedge \forall v (C_p))$$

$$\models \langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n \wedge R_m)(t_1, \dots, t_k) = (R_n(t_{l_1}, \dots, t_{l_n}) \wedge (R_m(t_{l_{n+1}}, \dots, t_{l_{n+m}}))).$$

Ainsi, la conjonction d'attributs engendre de l'implication cognitive forte.

$$\models \langle l_1, \dots, l_n \rangle \langle l_{n+1}, \dots, l_{n+m} \rangle (R_n \wedge R_m)(t_1, \dots, t_k) \succ_s (R_n(t_{l_1}, \dots, t_{l_n}) \& (R_m(t_{l_{n+1}}, \dots, t_{l_{n+m}})))$$

4. Lois d'implication cognitive faible

$$\models T(A_p \Rightarrow B_p) \Rightarrow (A_p \succ B_p)$$

La loi de l'existence d'une équivalence matérielle entre les propositions qui diffèrent au plus par le fait qu'elles subsument certains objets identiques de référence sous des concepts différents.

$$\models (t =_2 t') \Rightarrow (A_p \succ [t'/t]A_p), \text{ quand il n'y a pas de confusion de variables liées.}$$

La loi d'instanciation universelle

$$\models \forall v A_p \succ [t/v]A_p, \text{ quand il n'y a pas de confusion de variables liées.}$$

36. Ainsi les propositions que Jean est identique à Jean et que Jean est identique à lui-même sont identifiées en logique propositionnelle même si elles prédisent des attributs de longueur différente.

Bibliographie

- Aristote, *Analytica Priora et Posteriora*, W. D. Ross (ed), Oxford : Clarendon Press, 1968.
- Austin, J.L. *How to do Things with words*, Oxford : Clarendon Press, 1962.
- Bealer, G., *Quality and Concept*, Oxford : Clarendon Press, 1982.
- Carnap, R., *Meaning and Necessity*, University of Chicago Press, 1956.
- Bigelow, J., "Semantic of Thinking, Speaking and Translation", in F. Guenther and M. Guenther-Reutter (eds), *Meaning and Translation*, New York University Press, 1978.
- Cherniak, C., *Minimal Rationality*, Cambridge, Mass. : MIT Press, Bradford Books, 1986.
- Church A., "A formulation of the logic of sense and denotation", in P. Henle et al. (eds.), *Structure, Method and Meaning*, New York : Liberal Arts Press, 1951.
- , "The Need for Abstract Entities in Semantic Analysis", *Proceedings of the American Academy of Arts and Sciences*, lxxx, 1951.
- , "Intensional Isomorphism and identity of Belief", *Philosophical Studies* 5, 1954.
- , *The Problem of Universals*, Notre Dame, Indiana : University of Notre Dame Press, 1956.
- Craig, W., *Logic in Algebraic Form*, North Holland, 1974.
- Cresswell, Max J. "Hyperintensional Logic", *Studia Logica*, 34 : 25-38. 1975.
- Descartes, R., *Méditations métaphysiques*, 1641, réédité in Descartes, R., *Oeuvres et Lettres*, A Bridoux (ed.), Bibliothèque de la Pléiade, Paris : Editions Gallimard, 1953.
- Dunn, J.M. & N. Belnap, "The Substitution Interpretation of the Quantifiers", *Noûs*, vol. 2, 1968.
- Fine, K., "Analytic implication", in *Notre Dame Journal of formal Logic*, Volume 27, number 2, April 1986.
- Frege, G., "On sense and reference", in P. Geach et M. Black (eds), *Translation from the Philosophical Writings of Gottlob Frege*, Oxford : Blackwell, 1970, édition originale 1892.
- Frege, G., *The Basic Laws of Arithmetic*, University of California Press, 1967, édition originale 1893.
- "Gedankengefüge", in *Beiträge zur Philosophie des Deutschen Idealismus*, 3, 1923-6
- Foundations of Arithmetic*, Oxford : Blackwell, 1950, original edition in 1884.
- Gentzen, G., "Untersuchungen über das Logische Schliessen", *Mathematische Zeitschrift*, 39, 1934.
- Henkin L., Monk. D. et Tarski A., *Cylindric Algebras*, Part I, North Holland, 1971.
- Kaplan, D., "On the logic of demonstratives", *Journal of Philosophical Logic*, 8, 1 (1970), 81-98.
- Lewis, C.I., *A Survey of Symbolic Logic*, Berkeley et Los Angeles University of California Press, 1918.
- Lewis, D., "General semantics", dans D. Davidson et G. Harman (eds), *Semantics of Natural Language*, Dordrecht, Netherlands : Reidel, 1972.
- Montague R., *Formal Philosophy*, Yale University Press, 1974.

- Nowak M. et D. Vanderveken, "A complete minimal logic of the propositional contents of thought", à paraître dans *Studia Logica*.
- Nuchelmans, G., *Theories of Propositions, Ancient and Medieval Conceptions of the Bearers of Truth and Falsity*, Amsterdam : North-Holland, 1973.
- Parry, W.T., "Ein Axiomsystem für eine neue Art von Implikation (analytische Implication)", *Ergebnisse eines Mathematisches Colloquiums*, Volume 4, 1933.
- "Comparison of entailment theories", *The Journal of Symbolic Logic*, Volume 37, 1972.
- Russell, B., "Mathematical Logic as Based on the Theory of Types", *American Journal of Mathematics*, xxx, 1908.
- Searle, J.R., *Speech Acts*, Cambridge University Press, 1969.
- Searle, J.R. et D. Vanderveken, *Foundations of Illocutionary Logic*, Cambridge University Press, 1985.
- Strawson P.F., *Subject and Predicate in Logic and Grammar*, London, Methuen, 1974.
- Tarski, A., *Logic, Semantics, Metamathematics*, Oxford University Press, 1956.
- Vanderveken, D., "On the unification of speech act theory and formal semantics", in P.R. Cohen, J. Morgan, et M.E. Pollack, (eds), *Intentions in Communication*, Bradford Books, MIT Press, 1990.
- Vanderveken, D., *Meaning and Speech Acts*, Volume I : Principles of Language Use et Volume II : Formal Semantics of Success and Satisfaction, Cambridge University Press, 1990-1.
- Vanderveken, D., "A new formulation of the logic of propositions" à paraître dans *Quebec Studies in the Philosophy of Science*, Volume 1 Logic, mathematics, physics in the Boston Studies in the Philosophy of Science, Kluwer, 1995.
- Vanderveken, D., "La forme logique des contenus de nos pensées conceptuelles", in D. Laurier et F. Lepage (éds), *Essais sur le langage et l'intentionnalité*, Bellarmin-Vrin, 1992.
- Whitehead, A.N. et B. Russell, *Principia Mathematica*, Cambridge University Press, 1910.
- Wittgenstein, L., *Tractatus logico-philosophicus*, London : Routledge & Kegan Paul, 1961, édition originale 1921.

Remerciements

Je remercie fortement les collègues et amis qui m'ont aidé par leurs remarques critiques lors de la rédaction de cet article. Je suis particulièrement reconnaissant à l'égard de G. Malinowski et de M. Nowak qui ont contribué à cette recherche. Je remercie également le Fonds FCAR du Québec et les deux Conseils canadiens de Recherche en Sciences Humaines et en Science Naturelle et en Génie qui m'ont octroyé des subventions pour des projets liés à la présente recherche.

Daniel VANDERVEKEN

PARTIE III

LOGIQUE ET LANGAGE

Chapitre 12

Logique partielle et inférence valide

Pourquoi une logique partielle ? Et tout d'abord qu'entend-on par logique partielle ? L'expression elle-même n'est pas très heureuse et connote l'idée d'une logique dont le domaine d'application serait restreint ou encore d'une logique qui ne serait que *partiellement* logique. Paradoxalement, les différents systèmes de logique partielle se présentent plutôt comme des systèmes plus généraux que les systèmes de logiques classiques, ces derniers n'étant que des cas limites, des idéalizations où tout est défini dans la plus grande transparence.

La meilleure manière d'introduire la logique partielle est peut-être de présenter les motivations philosophiques qui lui ont donné naissance. L'idée d'élaborer une logique partielle en théorie des types est relativement récente. C'est Pavel Tichý [Tic82] qui non seulement en formula explicitement le projet mais donna les motivations d'une telle entreprise. Il vaut la peine d'en reprendre l'essentiel ici.

La première « est que la logique sous-jacente au langage ordinaire - et donc à nos schèmes conceptuels - est celle de la théorie simple des types. » Sans doute faut-il comprendre par là que la théorie des types possède les ressources suffisantes pour exprimer les structures inférentielles que l'on retrouve dans les langues naturelles, structures dont la complexité excède souvent le pouvoir expressif des calculs de premier ordre. La seconde est qu'il faut enrichir la théorie des types habituelle pour ainsi dire de l'intérieur « en laissant tomber l'hypothèse de la totalité et en traitant les fonctions partielles sur un pied d'égalité avec les totales. » En fait, « [1] la théorie partielle des types nous fournit un support qui permet l'analyse de n'importe quelle expression linguistique et nous procure une explication universelle de l'implication logique. »

On voit que pour Tichý l'idée d'introduire des fonctions partielles en théorie des types n'est pas présentée comme l'adoption d'une hypothèse particulière mais, au contraire, comme l'abandon d'une restriction arbitraire, celle qui consiste à n'utiliser comme valeurs sémantiques qu'un sous-ensemble des fonctions, le sous-ensemble des fonctions totales, celles qui sont définies pour tous les arguments de la bonne catégorie.

Tichý n'a pas réalisé complètement son projet, pas plus qu'il n'a su déployer toute la puissance de la théorie partielle des types et ce parce qu'il lui manquait une notion technique fondamentale sur laquelle je reviendrai plus loin. Avec Tichý apparaît cependant l'idée fondamentale que les logiques partielles permettent de définir une notion plus générale de l'implication logique, une notion d'implication capable de couvrir les cas où certaines des prémisses sont sémantiquement indéfinies, c'est-à-dire dont les valeurs ne sont que *partielles*.

Cette « indéfinitude » sémantique est susceptible d'être interprétée de multiples façons¹ : présuppositions non satisfaites (l'actuel roi de France...), erreur de catégorie (les idées vertes...), indécidabilité de certains énoncés, etc.. La motivation première de la logique partielle que nous allons présenter ici - sans présumer de l'universalité de ses applications - est épistémique. Il est en effet bien connu que l'identité de valeur sémantiques classiques de deux expressions n'est pas une condition suffisante pour permettre leur inter-substitutivité dans les contextes d'attitudes propositionnelles. L'exemple le plus célèbre est le fameux *puzzle* de Kripke où deux noms propres du même objet ne sont absolument pas substituables dans le contexte de l'expressions de croyance d'un locuteur qui ignore que les deux noms propres désignent le même objet et croit fermement même que ces deux objets sont différents². C'est ici que la notion d'objet sémantique partiel montre son utilité : deux expressions du langage peuvent très bien avoir la même valeur sémantique (nous supposons que cette valeur est une fonction au sens de la théorie des ensembles) sans que les représentations de ces fonctions chez le locuteur coïncident. La seule contrainte - qui sera omniprésente en logique partielle - est que ces deux représentations soient compatibles, ce qui se traduit par une exigence générale de *monotonie*, c'est-à-dire qu'alors que fonctions et arguments prennent des valeurs de mieux en mieux définies la valeur de la fonction appliquée à l'argument converge vers une valeur bien définie, la valeur totale. De là est née l'idée de décrire la hiérarchie des fonctions partielles en théorie des types comme une hiérarchie de fonctions monotones par rapport à une relation d'ordre dont l'interprétation intuitive est celui d'être plus défini.

Cette étude comportera deux parties. Dans la première, je retracerai les grandes étapes qui ont jalonné l'élaboration de la théorie partielle des types en accordant une attention particulière à la théorie propositionnelle des types qui est moins bien

1. Voir [Bla86], section 2 p. 15 à 30.

2. Voir [Lep92] pour une solution de l'énigme dans le cadre de la logique partielle.

connue et dans la seconde, je ferai états de résultats récents principalement dû à Serge Lapierre et à moi-même concernant différents concepts de validité et je terminerai en présentant quelques résultats de complétude pour la logique partielle.

1. La théorie des types propositionnels

La théorie partielle des types est née de l'idée de généraliser cette théorie bien connue qu'est la théorie simple des types en autorisant la présence d'objet partiellement définis. Rappelons quelques faits élémentaires à propos de la théorie des types, en particulier à propos de la théorie des types propositionnels.

La théorie simple des types est une version simplifiée de la théorie des types ramifiés de Russell. Elle a été initialement formulée par Chwistek et Ramsey et la première présentation « achevée » est celle de Church [Chu40].

Dans ses versions contemporaines, la théorie des types consiste d'abord en la construction d'une hiérarchie de fonctions de la façon suivante³.

Définition 1 : L'ensemble des types T est le plus petit ensemble tel que

- (i) $e \in T$ (e est le type des individus)
- (ii) $t \in T$ (t est le type des propositions)
- (iii) si $\alpha, \beta \in T$, alors $\langle \alpha \beta \rangle \in T$.

Les domaines de chaque type sont :

Définition 2 : Pour chaque $\alpha \in T$, l'ensemble des entités de type α est l'ensemble D_α tel que :

- (i) $D_e = E$ (où E est un ensemble non vide ; E est l'ensemble des individus de l'univers du discours) ;
- (ii) $D_t = \{0, 1\}$ (l'ensemble des valeurs de vérité) ;
- (iii) $D_{\langle \alpha \beta \rangle} = D_\beta^{D_\alpha}$ (l'ensemble des fonctions de D_α dans D_β ; nous écrirons simplement $D_{\alpha\beta}$).

3. Pour des raisons évidentes nous utiliserons un formalisme le plus uniforme possible. Cela impose parfois des reformulations relativement importantes que nous signalerons en notes.

Il est possible d'élaborer un calcul logique comprenant un nombre infini dénombrable de variables de chaque type et un très petit nombre de constantes logiques, en fait trois, soit l'abstracteur fonctionnel, son inverse l'application fonctionnelle et l'identité. Henkin [Hen50] a proposé une axiomatisation d'un tel système et en a montré la complétude au sens général. Gallin [Gal75] a même proposé une version intensionnelle et fourni une axiomatisation complète toujours, bien sûr, pour des modèles généraux.

Nous allons nous intéresser ici à un calcul essentiellement plus simple, un calcul basé sur la hiérarchie construite exclusivement à partir de $D_t = \{0, 1\}$ et en laissant tomber les clauses (i) des définitions 1 et 2 et que nous appellerons théorie des types propositionnels. Un tel calcul possède un double intérêt. Le premier est d'ordre purement théorique : il fait l'économie de l'hypothèse qu'il existe des objets. Tous les énoncés valides de ce calcul expriment des propriétés qui reposent ultimement sur les formes propositionnelles en général. En ce sens ce calcul est une véritable logique (bien qu'il contienne des variables de foncteurs propositionnels de tous ordres et qu'il autorise la quantification sur ces variables) et possède vis-à-vis la théorie des types en général, un statut analogue au calcul des propositions vis-à-vis le calcul des prédicats du premier ordre : ses énoncés valides constituent un « noyau dur » de l'ensemble des énoncés valides de la théorie simple des types.

Le second intérêt est plus pragmatique : chaque domaine D_α est fini ce qui quelquefois, ainsi que nous allons le voir, simplifie grandement les choses.

Il serait délicat de donner une date et un lieu de naissance à la théorie des types propositionnels à la fois pour des raisons internes - ses différents concepts fondamentaux sont également des concepts de la théorie simple des types et ne sont pas apparus simultanément - et pour des raisons externes - certains concepts ont été introduits dans des articles qui sont passés tellement inaperçus qu'ils ont été tout simplement réinventés plus tard. Voici une ébauche de la genèse en autant que je sois capable de la reconstituer. Pour simplifier les choses, partons du texte qui présente la théorie d'une manière achevée. Il s'agit d'un texte assez peu connu : « A Theory of Propositional Types » de Leon Henkin [Hen63]. Le calcul présenté par Henkin n'utilise comme symboles primitifs que l'abstracteur λ , sa converse et l'identité, est complet et la preuve de sa complétude est constructive (en fait, pour chaque élément de chaque domaine D_α il existe une expression du langage que l'on peut construire explicitement dont cette fonction est la valeur).

Voici schématiquement comment se présente ce calcul. Tout d'abord pour chaque type α , il y a un ensemble $Var_\alpha = \{x_{\alpha i}\}_{i \in \omega}$ dénombrable de variables de ce type.

Définition 3 : L'ensemble Trm_α des termes de type α est le plus petit ensemble tel que

- (i) $Var_\alpha \in Trm_\alpha$
- (ii) si $A, B \in Trm_\alpha$, alors $[A \equiv B] \in Trm_t$

(iii) $A \in Trm_\alpha$ et $x \in Trm_\beta$, alors $\lambda x A \in Trm_{\beta\alpha}$

(iv) si $A \in Trm_{\alpha\beta}$ et $B \in Trm_\alpha$, alors $[AB] \in Trm_\beta$

On appellera *énoncés* les termes de type t . Les constantes logiques sont définies de la manière suivante⁴ :

Définition 4 :

$$T =_{\text{def}} [\lambda x_t x \equiv \lambda x x]$$

$$F =_{\text{def}} [\lambda x_t x \equiv \lambda x T]$$

$$\neg t =_{\text{def}} \lambda x_t [x \equiv F]$$

$$\wedge_{t<t'} =_{\text{def}} \lambda x_t \lambda y_{t'} [\lambda f_{t'} [\lambda f_t [\lambda y] y] \equiv \lambda f [\lambda f T] T]]$$

$$\forall x_\alpha A =_{\text{def}} [\lambda x_\alpha A \equiv \lambda x T]$$

etc. (les autres constantes étant introduites de la manière habituelle)⁵.

Nous écrirons $[A \wedge B]$ pour $[[\wedge A]B]$ et utiliserons $x_\alpha, y_\alpha, z_\alpha, f_\alpha, g_\alpha, h_\alpha$, comme variables quelconque de type α .

Nous définissons ensuite une *assignation de valeur*.

Définition 5 : Une assignation de valeur μ est une fonction

$$\mu : \bigcup_{\alpha \in T} Var_\alpha \rightarrow \bigcup_{\alpha \in T} D_\alpha$$

telle que $\mu(x_\alpha) \in D_\alpha$. On note $\mu(a/x)$ l'assignation qui diffère au plus de μ par le fait qu'elle assigne à x la valeur a .

Enfin, nous définissons une *valuation basée sur μ* .

Définition 6 : Une valuation basée sur μ est une fonction

$$V_\mu : \bigcup_{\alpha \in T} Trm_\alpha \rightarrow \bigcup_{\alpha \in T} D_\alpha$$

4. Nous présentons une variante inspirées de Andrews [And63].

5. Nous omettrons les indices de types lorsque cela ne crée pas de confusion.

telle que

- (i) $V_\mu(x_\alpha) = \mu(x_\alpha)$;
- (ii) $V_\mu([A \equiv B]) = 1$ ssi $V_\mu(A) = V_\mu(B)$ et 0 autrement ;
- (iii) $V_\mu([AB]) = V_\mu(A)(V_\mu(B))$
- (iv) $V_\mu(\lambda x_\alpha A)$ est la fonction qui à chaque $a \in D_\alpha$ associe $V_\mu(a/x)(A)$.

Lorsqu'une formule A est close $V_\mu(A)$ est indépendant de μ et $V_\mu(A)$ est appelé la *dénotation* de A et est notée A^d .

Le système axiomatique proposé par Henkin comporte sept schémas d'axiomes et une règle de remplacement.

$$H_1 : [A_\alpha \equiv A]$$

$$H_2 : [[A_t \equiv T] \equiv A]$$

$$H_3 : [[T \wedge F] \equiv F]$$

$$H_4 : [[f_{tt}T \wedge fF] \equiv \forall x_t[f x]]$$

$$H_5 : [x_\alpha \equiv y_\alpha] \supset [[f_{\alpha\beta} \equiv g_{\alpha\beta}] \supset [[f x] \equiv [g y]]]$$

$$H_6 : [\forall x_\alpha [[f_{\alpha\beta} x_\alpha] \equiv [g_{\alpha\beta} x_\alpha]] \supset [f \equiv g]]$$

$H_7 : [[\lambda x_\alpha AB] \equiv A\{B/x\}]$ où B est libre pour x et $A\{B/x\}$ est la formule obtenue à partir de A en substituant B à chaque occurrence libre de x .

Règle R De $(A \equiv B)$ et C on peut tirer D , où D provient de C par substitution d'une occurrence de A par une occurrence de B . Si A est un théorème, on écrit $\vdash A$.⁶

Pour montrer la complétude de ce système Henkin procède de la manière suivante :

1) Pour n'importe quelle $a \in D_\alpha$, il existe une formule close, appelons-la a^n telle que $(a^n)^d = a$. Cette formule est un nom canonique de a et Henkin donne un algorithme explicite pour sa construction. En particulier, $T^n = T$ et $F^n = F$ et $T^d = 1$ et $F^d = 0$.

6. Nous avons, bien sûr, adapté le système de Henkin à notre notation.

2) Soit A^μ la formule obtenue à partir de A en substituant à chaque occurrence de variable libre x_α la formule $(\mu(x_\alpha))^n$. Henkin montre que pour n'importe quelle formule A on a $\vdash [A^\mu \equiv (\bigvee_\mu(A))^\mu]$.

3) En particulier, si A est valide et close, on aura $\vdash [A \equiv T]$ et de là, on montre facilement que $\vdash A$. (Si A est valide et non close, on passe par sa clôture universelle.)

Cet article de Henkin ne semble pas avoir reçu l'attention qu'il mérite. En fait, je n'ai trouvé que deux textes qui y font référence soit celui de Gallin [Gal75] et celui de Farmer [Far90]. Il semble pourtant que ce texte soit la première présentation d'un système logique d'ordre supérieur ne contenant que trois primitifs, soit l'abstraction fonctionnelle, l'application fonctionnelle et l'identité. Henkin croyait d'ailleurs être le premier à suggérer la réduction des quantificateurs à ces opérateurs. À la page 324, dans la première version du texte, après avoir rendu hommage à Tarski [Tar23] d'avoir montré comment réduire les constantes logiques habituelles en termes des quantificateurs et de l'identité, il affirme qu'il est, semble-t-il, le premier à proposer une réduction explicite des quantificateurs en termes de l'opérateur λ et de l'identité. Dans la version publiée, il ajoute une note remerciant Peter Andrews de lui avoir signalé que cette réduction a été explicitement faite par Quine [Qui56]⁷.

Si le système de Henkin est le premier à n'utiliser que les trois primitifs en question, la première théorie des types propositionnels remonte à la protothétique de Leśniewski. Nous n'avons malheureusement pas le texte de Leśniewski qui aurait été détruit durant le soulèvement de Varsovie en 1944. La version la plus fiable dont nous disposons est le texte de Jerzy Śłupecki [Ślu53] publié dans la première livraison de *Studia Logica* intitulé « St. Leśniewski protothetics » rédigé à partir de notes prises par des étudiants de Leśniewski. Curieusement, Henkin ne connaissait pas ce texte, il remercie en effet le *referee* de lui en avoir signalé l'existence. Cela est d'autant plus curieux que Henkin renvoie lui-même à la protothétique de Leśniewski dans la version de Andrzej Grzegorzczak [Grz55] « The System of Leśniewski in relation to contemporary Logical Research » également paru dans *Studia Logica* deux ans plus tard. Il faut dire que Grzegorzczak ne signale pas l'existence du texte de Śłupecki. Terminons sur ce point en signalant que Montague [Mon74] reprend

7. Ce texte de Quine où est pour la première fois présentée la réduction des quantificateurs à l'opérateur λ n'est lui non plus pas très connu. Je ne l'ai vu cité nulle part sauf dans le texte de Henkin. Je commençais sérieusement à me demander si les logiciens lisent quelquefois les travaux des autres lorsque j'ai découvert qu'il s'agissait, en fait, de l'allocation de président sortant que Quine a prononcée en séance plénière devant les membres de l'*Association for Symbolic Logic* réunis en session conjointe avec ceux de l'*Eastern Division of the American Philosophical Association* à Boston lors de leur congrès annuel le 29 décembre 1955. Non seulement ils ne se lisent pas, mais ils ne s'écoutent pas !

l'essentiel des traductions de Henkin dans « Universal Grammar » sans le signaler alors qu'il renvoie explicitement à Tarski.

Voilà pour ce qui est de la théorie des types propositionnels. Examinons maintenant ce qui en est de la *partiellité*.

2. La logique partielle

Tout comme pour la théorie des types propositionnels, il serait difficile de donner une date de naissance de la logique partielle. J'ai signalé plus que la première formulation explicite d'un projet de système de logique partielle remonte à Tichý mais on peut trouver retracer beaucoup des concepts qu'il utilise dans des articles ou ouvrages qui dont la raison première n'était pas d'apporter une contribution à la logique partielle.

La première tentative sérieuse de faire une présentation systématique de la logique partielle est celle de Blamey [Bla86]. Ce texte de Blamey contient en fait les ingrédients essentiels pour élaborer une logique partielle générale de même que les directives pour le faire mais il s'arrête au seuil de l'entreprise et en fait se contente d'élaborer une logique du premier ordre.

Parlons tout d'abord de ces éléments essentiels. Le premier de ces éléments est l'introduction de l'indéfini comme objet sémantique, c'est-à-dire l'introduction dans la métalangue d'un nom de l'indéfini. Cela peut sembler banal mais c'est une étape cruciale : sans cet « engagement ontologique », il ne semble pas possible de décrire de façon cohérente la réitération de l'application fonctionnelle. C'est cette notion fondamentale manquait à Tichý ainsi que nous l'avons signalé plus tôt et c'est ce qui l'a amené à rejeter les domaines de fonctions au profit des domaines de relations et donc de demeurer essentiellement au premier ordre.⁸

L'argument est simple. Supposons - nous raisonnerons sur un exemple dont la généralisation va de soi - que A_α , B_β et C_γ sont trois expressions telles que $[BC]$ et $[A[BC]]$ sont des termes. Les notions métalinguistiques d'« être définie » et d'« être indéfinie » s'appliquent aux valeurs des expressions (qui sont des fonctions) et ce n'est que par extension, par un abus de langage trivial que la notion s'applique aux expressions selon que sa valeur est ou n'est pas définie. Ici, il ne faut pas confondre avec une autre utilisation de la notion d'« indéfinie », celle qui consisterait à qualifier d'indéfinie une expression qui ne serait pas bien formée, c'est-à-dire qui ne serait pas un terme. Dans ce cas il faudrait plutôt parler de non-sens.

8. Malgré cette carence, le texte de Tichý demeure la première véritable tentative de systématiser la logique partielle. Bizarrement, Blamey semble ignorer son existence. L'argument de Tichý est repris par Musken dans [Mus88].

Dans le cadre classique, les expressions A, B, C recevront comme valeur des fonctions totales telles que la valeur de $[BC]$ est une fonction totale du type des arguments de A de sorte que la valeur de $[A[BC]]$ soit aussi une fonction totale. Que se passera-t-il maintenant si le valeur de ces expressions tout en étant du bon type, sont partielles ? Plus spécifiquement, qu'arrivera-t-il si la valeur de B est une fonction partielle qui est indéfinie pour l'argument qui est la valeur de C ? La valeur de $[BC]$ n'est pas définie. Dans ce cas, $[A[BC]]$ n'a pas de valeur c'est-à-dire qu'il n'est pas possible de donner un sens à l'expression « valeur de $[A[BC]]$ ».

Une solution générale à ce problème consiste à donner un statut intra-théorique à l'indéfini. L'indéfini devient un objet comme les autres et peut ainsi être valeur et argument de fonction. Ce simple ajout enrichit suffisamment le métalangage pour que nous soyons maintenant capable de décrire toute la hiérarchie des fonctions partielles.

L'idée de donner explicitement le statut d'« objet » à l'indéfini remonte à Dana Scott [Sco73] qui, curieusement, l'introduisit pour se débarrasser de la hiérarchie en construisant ses domaines réflexifs qui serviront à interpréter le λ -calcul non typé. On pourrait également faire remonter cette idée à Kleene avec ses connecteurs « forts » où l'indéfini apparaît explicitement dans les tables de vérité. Mon idée [Lep84] a été d'utiliser cette astuce pour définir la hiérarchie des fonctions partielles. Voyons comment.

Définition 7 : Pour tout $\alpha \in T$, l'ensemble PM_α des fonctions partielles de type α est

- (i) $PM_t = \{0, 1, \varphi\}$
- (ii) $PM_{\alpha\beta} = (PM_\alpha \rightarrow PM_\beta)$

où $(PM_\alpha \rightarrow PM_\beta)$ est l'ensemble des fonctions monotones de PM_α dans PM_β , la monotonie étant relative à l'ordre suivant :

- (i) pour tout $x \in PM_t$, $x \leq x$ et $\varphi \leq x$;
- (ii) pour tout $f, g \in PM_{\alpha\beta}$, $f \leq g$ si et seulement si pour tout $x \in PM_\alpha$, $f(x) \leq g(x)$;

Proposition 8 : Pour tout α , PM_α est un inf-semi-treillis, où l'inf $\bigwedge$ et (lorsque, bien sûr, il existe) le sup $\bigvee$ sont respectivement définis par la clause récursive

- (i) pour $x, y \in PM_t$, $x \bigwedge y = x$ si $x = y$ et φ autrement ;
- (ii) pour $f, g \in PM_{\alpha\beta}$, $f \bigwedge g$ est la fonction h telle que pour tout $x \in PM_\alpha$, $h(x) = f(x) \bigwedge g(x)$;

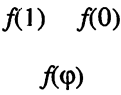
et

- (iii) pour $x, y \in PM_t, x \vee y = x$ si $y = \varnothing$ ou $x = y$
 $x \vee y = y$ si $x = \varnothing$ et autrement n'existe pas ;
- (iv) pour $f, g \in PM_{\alpha\beta}, f \vee g$ est la fonction h telle que pour tout $x \in PM_{\alpha},$
 $h(x) = f(x) \vee g(x)$ si $f(x) \vee g(x)$ existe.

On peut représenter les objets partiels en utilisant l'artifice graphique suivant. Tout d'abord PM_t formant un inf-semi-treillis plat il est naturel de le représenter par



et une fonction $f \in PM_{tt}$ par



Généralisant, PM_{tt} peut être représenté par

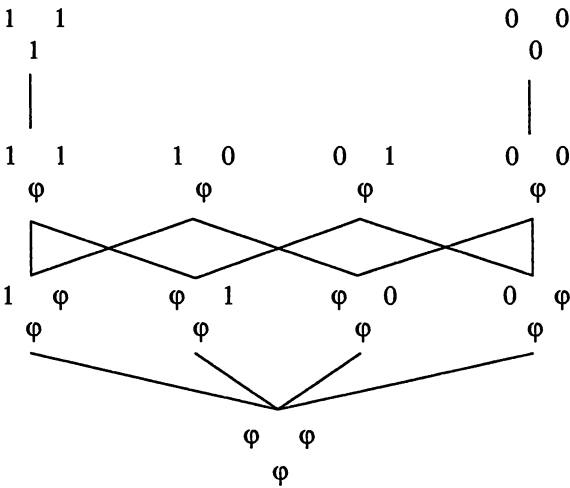


Diagramme 1.

Certains éléments du sommet de la hiérarchie (que nous appellerons *fonctions partielles totales*) se comportent comme des fonctions totales habituelles (que nous appellerons dorénavant *fonctions classiques*) du moins lorsque leurs arguments eux-mêmes se comportent comme des fonctions classiques. On peut les caractériser formellement de la manière suivante.

Définition 9 : Pour tout $\alpha \in T$, l'ensemble $PT_\alpha \subseteq PM_\alpha$ (l'ensemble des fonctions partielles totales) est le plus petit ensemble tel que

- (i) pour $\alpha = t$, et $x \in PM_t$, $x \in PT_t$ ssi $x \neq \phi$;
- (ii) pour $\alpha = \beta\gamma$, et $f \in PM_{\alpha\beta}$, $f \in PT_{\alpha\beta}$ ssi pour tout $x \in PT_\alpha$, $f(x) \in PT_\beta$.

Curieusement, l'espace des fonctions classiques n'est pas isomorphe à l'espace des fonctions partielles totales. On peut cependant montrer qu'il existe un épimorphisme qui plonge l'ensemble de fonctions partielles totales dans l'ensemble des fonctions classiques tel que l'image inverse d'une fonction classique par cet épimorphisme est un treillis complet de fonctions partielles totales [Lep92b].

Cette propriété permet de définir récursivement une relation d'équivalence entre fonctions partielles : deux fonctions sont équivalentes lorsqu'elles prennent des valeurs équivalentes pour les arguments *totaux*. Formellement,

Définition 10 : Soit « $\approx$ » la relation suivante sur PM_α

- (i) pour tout $x \in PM_t$, $x \approx x$;
- (ii) pour tout $f, g \in PM_{\alpha\beta}$, $f \approx g$ ssi pour tout $x \in PT_\alpha$, $f(x) \approx g(x)$.

On vérifie facilement que « $\approx$ » est une relation d'équivalence. Une des propriétés intéressantes de cette notion est exprimé par la proposition suivante.

Proposition 11 : Pour tout $f, g \in PT_{\alpha\beta}$, $f \approx g$ ssi $f \bigvee g$ existe.

Intuitivement, des fonctions partielles du même type peuvent différer de deux manières. La première, classique, est qu'elles peuvent prendre des valeurs incompatibles. Elle peuvent également différer en ce qu'elles sont compatibles mais plus ou moins définies. Cette intuition peut être formalisée de la manière suivante.

Définition 12 : Pour tout $\alpha \in T$, soit $\neq_\alpha^* \subseteq PM_\alpha \times PM_\alpha$ la relation ($\neq_\alpha^*$ est l'"incompatibilité" ou la "différence forte" et " $x \neq_\alpha^* y$ " se lit " x est incompatible avec y " ou " x diffère fortement de y ")⁹

9. Aucune confusion n'étant possible, nous omettrons les indices de types et écrirons simplement $\neq^*$.

- (i) pour $x, y \in PM_t$, $x \neq^* y$ ssi $x \neq \varnothing$ et $y \neq \varnothing$ et $x \neq y$;
- (ii) pour $f, g \in PM_{\alpha\beta}$, $f \neq^* g$ ssi il y a un $x \in PM_\alpha$ tel que $f(x) \neq^* g(y)$.

Nous pouvons maintenant définir une interprétation partielle pour les termes de la théorie des types propositionnels.

Définition 13 : Une assignation de valeur partielle μ est une fonction

$$\mu : \bigcup_{\alpha \in T} Var_\alpha \rightarrow \bigcup_{\alpha \in T} PM_\alpha$$

telle que $\mu(x_\alpha) \in PM_\alpha$. Comme précédemment, on note $\mu(a/x)$ l'assignation qui diffère au plus de μ par le fait qu'elle assigne à x la valeur a .

Enfin, nous définissons une *valuation partielle basée sur μ* .

Définition 14 : Une valuation partielle basée sur μ est une fonction

$$V_\mu : \bigcup_{\alpha \in T} Trm_\alpha \rightarrow \bigcup_{\alpha \in T} PM_\alpha$$

telle que

- (i) $V_\mu(x_\alpha) = \mu(x_\alpha)$;
- (ii) $V_\mu([A_\alpha \equiv B_\alpha]) = 1$ ssi $V_\mu(A) \in PT_\alpha$ et $V_\mu(A) \approx V_\mu(B)$
 0 ssi $V_\mu(A) \neq^* V_\mu(B)$
 $\varnothing$ autrement ;
- (iii) $V_\mu([AB]) = V_\mu(A)(V_\mu(B))$
- (iv) $V_\mu(\lambda x_\alpha A)$ est la fonction qui à chaque $a \in D_\alpha$ associe $V_\mu(a/x)(A)$

Les clauses (i), (iii) et (iv) sont identiques aux clauses correspondantes de la définition 6. La clause (ii) mérite quelques commentaires. L'adoption de la clause 6(ii) n'est pas possible car dans ce cas $V_\mu([A_\alpha \equiv B_\alpha])$ ne serait pas monotone par rapport à $V_\mu(A)$ et à $V_\mu(B)$ (par exemple, soit A_t et B_t tels que $V_\mu(A) = \varnothing$ et $V_\mu(B) = \varnothing$; d'après 6 (ii) on devrait avoir $V_\mu([A \equiv B]) = 1$ et pour C et D tel que $V_\mu(C) = 1$ et $V_\mu(D) = 0$ et donc tels que $V_\mu(A) \leq V_\mu(C)$ et $V_\mu(B) \leq V_\mu(D)$, et par monotonie, $V_\mu([C \equiv D]) = 1$. Donc $1 = 0$ ce qui est pour le moins indésirable).

Par ailleurs, la relation « $\approx$ » entre fonctions partielles totales est plus faible que l'identité mais demeure suffisante en vertu de la proposition 11 : si deux fonctions totales sont différentes mais équivalentes, leur supremum existe et appartient à la

même classe d'équivalence. Le fait que ces fonctions prennent des valeurs totales équivalentes pour les mêmes arguments totaux garantit la monotonie.

On peut montrer que l'ordre sur les fonctions partielles induit un ordre similaire sur les valuations partielles. Nous pouvons ainsi parler de *valuations partielles totales* et de *valuations classiques*. L'existence d'un épimorphisme entre les fonctions partielles totales et les fonctions classiques nous autorise à assimiler une valuation classique à n'importe quel élément de la classe d'équivalence de valuations partielles totales correspondante, car tous les énoncés reçoivent exactement la même valeur (bien sûr totale) par la valuation classique et par toutes les valuations partielles totales de la classe d'équivalence correspondante. Pour une preuve détaillée, voir [Læp92].

On vérifie facilement que (voir la définition 4 et la définition de la dénotation),

$$T^d = [\lambda x_t x \equiv \lambda x x]^d = 1$$

$$F^d = [\lambda x_t x \equiv \lambda x T]^d = 0$$

Revenons au diagramme 1. On voit que

$$1 \quad 0$$

$$\varphi$$

représente l'identité propositionnelle et que $\neg_{tt}^d = \lambda x_t [x \equiv F]^d$ sera la fonction représenté par le diagramme suivant

$$0 \quad 1$$

$$\varphi$$

Qu'en sera-t-il maintenant des connecteurs binaires ? Un élément f de $PM_{t \times t}$ peut être représenté selon la convention suivante

$$f(1)(1) \quad f(1)(0)$$

$$f(0)(1) \quad f(0)(0)$$

$$f(1)(\varphi)$$

$$f(0)(\varphi)$$

$$f(\varphi)(1) \quad f(\varphi)(0)$$

$$f(\varphi)(\varphi)$$

Diagramme 2.

Si nous effectuons le calcul en prenant pour f la dénotation de $\wedge_{t < tt}$, soit

$$\wedge_{t < tt}^d = \lambda x_t \lambda y_t [\lambda f_{t < tt} [[fx]y] \equiv \lambda f [[fT]T]]^d$$

nous obtenons la fonction représentée par le diagramme 3

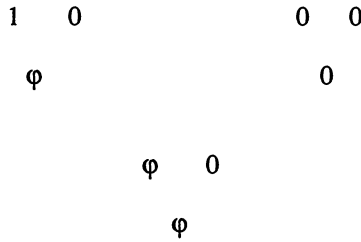


Diagramme 3.

Bien qu'il n'y ait pas de surprise en logique, il est remarquable de retrouver ici la conjonction forte de Kleene. On obtient évidemment la disjonction forte de Kleene par la traduction habituelle $A \vee B =_{\text{def}} \neg(A \wedge B)$.

Terminons ces remarques sémantiques par l'examen des quantificateurs. On vérifie directement que

$$\begin{aligned} \forall x A^d &= [\lambda x_{\alpha} A \equiv \lambda x T]^d = 1 \text{ ssi pour tout } a \in PT_{\alpha}, V_{\mu}(a/x)(A) = 1 \\ &0 \text{ ssi pour un } a \in PM_{\alpha}, V_{\mu}(a/x)(A) = 0 \\ &\varnothing \text{ autrement.} \end{aligned}$$

et

$$\begin{aligned} \exists x A^d &= [\neg \forall x [\neg A]]^d = [\neg [\lambda x_{\alpha} [\neg A] \equiv \lambda x T]]^d \\ &= 1 \text{ ssi pour un } a \in PM_{\alpha}, V_{\mu}(a/x)(A) = 1 \\ &= 0 \text{ ssi pour tout } a \in PT_{\alpha}, V_{\mu}(a/x)(A) = 0 \\ &= \varnothing \text{ autrement.} \end{aligned}$$

3. Deux notions de validité

En logique classique, au sens où nous l'entendons ici, la notion de validité et celle d'inférence valide sont monolithiques. A est une conséquence valide d'un ensemble Γ d'énoncés, en symboles $\Gamma \models A$, si et seulement si, pour toute valuation μ , si pour tout $B \in \Gamma$, $V_\mu(B) = 1$, alors $V_\mu(A) = 1$. De manière similaire, A est valide, symboliquement $\models A$, ssi pour toute valuation μ , $V_\mu(A) = 1$.

Une manière équivalente de présenter les choses dans le cadre classique est d'utiliser le faux plutôt que le vrai. On aurait alors que A est une conséquence valide de Γ si et seulement si, pour toute valuation μ , si pour tout $B \in \Gamma$, $V_\mu(B) = 1$, alors $V_\mu(A) \neq 0$. De manière similaire, A est valide ssi pour aucune valuation μ , $V_\mu(A) = 0$.

Ces deux notions, on s'en doutera, ne coïncident pas en logique partielle à cause de la présence d'un tiers, l'indéfini. Nous pouvons définir deux concepts de validité.¹⁰

Définition 15 : Soit Γ un ensemble d'énoncés et A_I un énoncé. On dira que A est une conséquence *vérifiablement valide* de Γ , en symbole $\Gamma \stackrel{\perp}{\models} A$ si et seulement si pour toute valuation partielle μ , si pour tout $B \in \Gamma$, $V_\mu(B) = 1$, alors $V_\mu(A) = 1$. De la même façon, A est vérifiablement valide, en symbole $\stackrel{\perp}{\models} A$, ssi pour toute valuation partielle μ , $V_\mu(A) = 1$.

Définition 16 : Soit Γ un ensemble d'énoncés et A_I un énoncé. On dira que A est une conséquence *falsifiablement valide* de Γ , en symbole $\Gamma \stackrel{0}{\models} A$ si et seulement si pour toute valuation partielle μ , si pour tout $B \in \Gamma$, $V_\mu(B) = 1$, alors $V_\mu(A) \neq 0$. De la même façon, A est falsifiablement valide, en symbole $\stackrel{0}{\models} A$, ssi pour aucune valuation partielle μ , $V_\mu(A) = 0$.

Ces deux notions ne coïncident pas ; on peut cependant facilement montrer que [Læp92] :

Proposition 17 : $\Gamma \models A$ ssi $\Gamma \stackrel{0}{\models} A$

10. Encore une fois, il est difficile de dater la naissance de ces notions. Comme le remarque à juste titre Elias G. C. Thijsse [Thi89] l'histoire de la logique partielle se confond souvent avec celle de la logique trivalente bien que les motivations qui ont conduit à leur élaboration soient très différentes de celles qui sont mises de l'avant ici. Par exemple chez Kleene [Kle38] l'incalculable. Pour une discussion de ces questions, voir [Bla86] et [Urq86]. Les définitions présentées ici sont celles de Thijsse dans la variante de [Lap92]. Voir également [Ben86].

Qu'en est-il alors de la classe des énoncés vérifiablement valides ? Cette classe n'est ni vide ni identique à la classe des énoncés classiquement valides. Par exemple, $[A_\alpha \equiv A_\alpha]$ n'est pas vérifiablement valide mais $[\lambda x_\alpha x \equiv \lambda x_\alpha x]$ l'est. Le problème, on l'aura deviné, est de fournir un système (complet si possible) pour la classe des énoncés vérifiablement valides. La question n'est pas encore résolue et la grande difficulté repose sur l'impossibilité absolument incontournable de construire une expression E_{tt} de la langue objet dont l'interprétation serait $[EA]$ est vrai si et seulement si A n'est pas définie : il est facile de se convaincre qu'un tel foncteur ne serait pas monotone.

Lapierre et moi-même avons cependant obtenus les résultats suivants. Tout d'abord, s'il n'est pas possible d'introduire dans la langue objet un foncteur $[EA]$ dont l'interprétation serait « A est indéfinie » il est possible d'introduire un foncteur $\mathfrak{Z}(A)$ dont l'interprétation est à peu près « A est totale ».

Définition 18 : $\mathfrak{Z}(A_\alpha) =_{\text{def}} \exists x_\alpha [x \equiv A]$

On comprend pourquoi j'utilise l'expression « à peu près » : $\mathfrak{Z}(A_\alpha)$ est vraie quand A est totale et indéfinie autrement. $\mathfrak{Z}(A_\alpha)$ n'est jamais fausse et ainsi la monotonie est préservée. Un tel foncteur est d'une très grande utilité. Il permet en effet d'introduire des axiomes de la forme « si $A, B, \dots$, sont totaux, alors... » et permet l'élaboration d'un système de logique partielle.

4. Un système déductif pour la logique partielle

Une première étape consiste à caractériser le sous-ensemble des formules du calcul propositionnel des *tautologies vérifiablement valides*. On procède de la façon suivante.

Définition 19 : Soit $FP \subseteq Trm_t$, l'ensemble des formules propositionnelles, le plus petit ensemble tel que

- (i) $Var_t \cup \{T, F\} \subseteq FP$
- (ii) si $A, B \in FP$, alors $\neg A, [A \wedge B], [A \supset B], [A \vee B], [A \equiv B] \in FP$.

Reprenant une idée de Henkin, nous définissons récursivement deux sous-ensembles de FP :

Définition 20 : Soit FP^+ et FP^- tels que

- (i) $T \in FP^+$ et $F \in FP^-$;
- (ii) si $A \in FP^+$ et $B \in FP^-$, alors $\neg B \in FP^+$ et $\neg A \in FP^-$;

(iii) si $A, B \in FP^+$, alors $[A \wedge B] \in FP^+$, et si $A \in FP^-$, alors pour tout $B \in FP$,

$$[A \wedge B], [B \wedge A] \in FP^-;$$

(iv) si $A \in FP^+$ et $B \in FP^-$, alors $[A \supset B] \in FP^-$, et pour tout $C \in FP$, $[C \supset A]$,

$$[B \supset C] \in FP^+;$$

(v) si $A \in FP^+$ et $B, C \in FP^-$, alors $[B \vee C] \in FP^-$ et pour tout $D \in FP$, $[A \vee D]$,

$$[D \vee A] \in FP^+;$$

(vi) si $A, B \in FP^+$ et $C, D \in FP^-$, alors $[A \equiv B], [C \equiv D] \in FP^+$ et $[A \equiv C]$,

$$[C \equiv A] \in FP^-;$$

(vii) rien d'autre n'appartient à FP^+ or FP^- .

On montre facilement et directement que A est une tautologie vérifiablement valide ssi $A \in FP^+$. On voit que l'existence d'une procédure de décision facilite grandement les choses. Nous pouvons maintenant présenter un système formel. Le système comprend 16 règles de bases et 3 règles impropres. On peut regrouper les règles propres en trois catégories : celles qui sont classiquement valides, celles qui ont une contrepartie classiquement valide et enfin celles qui n'ont pas de contrepartie classique.

Règles classiquement valides

$$R1. F \Rightarrow A$$

R2. $[\lambda x_1 \dots [\lambda x_n A_t B_1] \dots B_n] \Leftrightarrow A_t \{B_1 \dots B_n / x_1 \dots x_n\}$, pourvu que chaque B_i soit libre pour x_i dans A^{11} .

11. Nous utilisons la notation $A\{B/x\}$ pour signifier que *toutes* les occurrences libres de x ont été remplacées par des occurrences de B , et $A[B/x]$ pour signifier que *certaines* (peut-être aucune) occurrences libres de x ont été remplacées par des occurrences de B .

R3. $A_t[C/x_t], [C \equiv B_t] \Rightarrow A[B/x]$, pourvu que B soit libre pour x dans A .

R4. $[[A_{tt}T] \wedge [AF]] \Leftrightarrow \forall x_t[Ax]$

R5 $\Rightarrow \forall z_{\alpha\beta} \forall x_{\alpha} \forall y_{\alpha} [[x \equiv y] \supset [zx \equiv zy]]$

R6 $\Rightarrow \forall z_{\alpha\beta} \forall v_{\alpha\beta} [\forall x_{\alpha} [zx \equiv vx] \equiv [z \equiv v]]$

Règles ayant une contrepartie classique

R7. $\mathfrak{I}(B_{\alpha}), \mathfrak{I}(A_{\beta}\{B/x_{\alpha}\}) \Rightarrow [[\lambda x_{\alpha} AB] \equiv A_{\beta}\{B/x_{\alpha}\}]$, pourvu que B soit libre pour x dans A .

R8. $A_t[C/x_{\alpha}], [C_{\alpha} \equiv B_{\alpha}], \mathfrak{I}(A[B/x_{\alpha}]) \Rightarrow A[B/x_{\alpha}]$, pourvu que B soit libre pour x dans A .

R9 $\mathfrak{I}(A_t) \Rightarrow [A \equiv T] \equiv A$

Règles n'ayant pas de contrepartie classique

R10 $A \Rightarrow \mathfrak{I}(A)$

R11 $\mathfrak{I}(A_{\alpha}), \mathfrak{I}(B_{\alpha}) \Rightarrow \mathfrak{I}([A \equiv B])$

R12 $[A_{\alpha} \equiv B_{\alpha}] \Rightarrow \mathfrak{I}(A)$

R13 $[A_{\alpha} \equiv B_{\alpha}] \Rightarrow \mathfrak{I}(B)$

R14 $\mathfrak{I}(A_{\alpha\beta}), \mathfrak{I}(B_{\alpha}) \Rightarrow \mathfrak{I}([AB])$

R15 $\forall x_1 \dots \forall x_n \mathfrak{I}(A_{\beta}) \Leftrightarrow \mathfrak{I}(\lambda x_1 \dots \lambda x_n A_{\beta})$, où $x_1, \dots, x_n$ sont des variables distinctes de types quelconques.

R16 $\Rightarrow \forall x_{\alpha} \mathfrak{I}(x)$

On vérifie facilement que toutes ces règles sont vérifiablement valides.

Définition 21 : Il existe une *preuve* de A à partir d'un ensemble H d'hypothèses (en symbole $H \vdash A$) si et seulement si il existe une suite de n formules $A_1, \dots, A_n$ telle que $A = A_n$ et pour tout $i \leq n$:

- (i) $A_i \in H$, ou
- (ii) A_i est telle que $\Rightarrow A_i$ est une règle, ou
- (iii) $A_k, \dots, A_l \Rightarrow A_i$ (avec $k, l < i$) est une règle, ou
- (iv) A_i résulte de l'application d'une des règles impropres suivantes :

R1* $A_i = \forall x_1 \dots \forall x_n B_i$ et il existe une preuve de B à partir de $\mathfrak{I}(x_1), \dots, \mathfrak{I}(x_n)$.

R2* A_i est $[B \equiv C]$ et il existe une preuve de C à partir de B et une preuve de B à partir de C , et pour un certain $k < i$, $A_k = \mathfrak{I}(B)$ ou $A_k = \mathfrak{I}(C)$.

R3* $A_i = B_i \{C_\alpha / x_\alpha\}$ et il existe une preuve de B à partir de l'ensemble vide et C est libre pour x dans A . Si H est vide, on écrit $\vdash A$.

Proposition 22 : Si $A \in FP^+$, alors $\vdash A$.

On obtient assez facilement la généralisation suivante :

Proposition 23 : Si $A \in FP$ est une forme tautologique classique dont les variables libres sont $x_1, \dots, x_n$, alors $\vdash \forall x_1 \dots \forall x_n A$.

Il ne semble pas possible, à l'instar de Henkin, de généraliser ce résultat à la totalité des formules valides de la théorie partielle des types propositionnels. Le problème, incontournable, lié à cette approche est l'impossibilité d'avoir un nom canonique dans la langue objet pour chaque fonction partielle sans modifier de façon essentielle le cadre théorique. Supposons, par exemple, que ϕ ait un nom, soit ϕ^n . Une conséquence immédiate sera que, même pour les valuations partielles totales, certaines expressions demeurent indéfinies. Bien que de tels systèmes de logiques puissent avoir de l'intérêt, ils sont, selon le point de vue adopté ici, inacceptables.

Une autre approche, que nous ne ferons qu'évoquer en terminant, est d'utiliser la technique des modèles généraux. Il semble qu'il soit alors possible d'élaborer des modèles pour lesquels le système est complet mais où seulement une partie de l'ensemble des fonctions partielles apparaît, celles, bien sûr, qui ont un nom dans la langue objet. Mais ceci est une autre histoire que nous n'aborderons pas ici.

Bibliographie

- [And63] Andrews, P. B., « A Reduction of the Axioms for the Theory of Propositional Types », *Fundamenta Mathematicæ* LII, 1963, 345-350.
- [Ben86] van Benthem, J., « Partiality and Nonmonotonicity in Classical Logic », *Logique et Analyse*, 29, 1986, 225-247.
- [Bla86] Blamey, S., « Partial Logic », in *Handbook of Philosophical Logic Vol. III*, Gabbay, D., et al. (eds), Dordrecht, Reidel, 1986, 1-70.
- [Chu40] Church, A., « A Formulation of the Simple Theory of Types », *The Journal of Symbolic Logic*, 5, 1940, 56-68.
- [Far90] Farmer, W. M., « A Partial Functions Version of Church's Simple Theory of Types », *The Journal of Symbolic Logic*, 55, 3, 1990, 1269-1271.
- [Gal75] Gallin, D., *Intensional and Higher-Order Modal Logic*, Amsterdam, North-Holland, 1975.
- [Grz55] Grzegorezyk, A. "The Systems of Leśniewski in Relation to Contemporary Logical Research », *Studia Logica* 3, 1955, 77-95.
- [Hen50] Henkin, L., « Completeness in the Theory of Types », *The Journal of Symbolic Logic*, 15, 1950, 81-91.
- [Hen63] Henkin, L., « A Theory of Propositional Types », *Fundamenta Mathematicæ*, 52, 1983, 323-344.
- [Kri79] Kripke, S., « A Puzzle About Belief », Margalit, A., (ed), *Meaning and Use*, Dordrecht, Reidel, 1979, 239-283.
- [Lap92] Lapierre, S., « A Partial Semantics for Intensional Logic », *Notre Dame Journal of Formal Logic*, 33, n°4, 1992, 517-541.
- [Læp92] Lapierre, S. et Lepage F., « Toward a Calculus of Partial Propositional Types », Cahier du département de philosophie n° 92-11, Université de Montréal, 1992.
- [Lep92a] Lepage, F. « Contenu sémantique, contenu cognitif et l'énigme de Kripke », Laurier D. et al. (eds.) *Essais sur le langage et l'intentionnalité*, Paris et Montréal, Vrin et Bellarmin, 1992, 227-243.
- [Lep92b] Lepage, F., « Partial Functions in Type Theory », *Notre Dame Journal of Formal Logic*, 33, n°4, 1992, 493-516.
- [Lep84] Lepage F., « The Object of Belief », *Logique et Analyse*, 36, 106, 193-210, 1984.
- [Mon74] Montague, R., « Universal Grammar », in *Formal Philosophy*, New Haven, Yale University Press, 1974, 222-246.
- [Mus88] Musken, R., « Going Partial in Montague Grammar », *ITLI Prepublication Series*, LP-88-04, University of Amsterdam, 1988.
- [Qui56] Quine, W. V., « Unification of Universes in Set Theory », *The Journal of Symbolic Logic*, 21, 1956, 267-279.
- [Sco73] Scott, D., « Models for Various Type-free Calculi », in P. Suppes et al. (eds) *Logic, Methodology and Philosophy of Science IV*, Amsterdam, North-Holland, 1973, 157-187.

- [Slu53] Slupecki, J., « St. Leśniewski protothetics », *Studia Logica*, 1, 1953, 44-111.
- [Tar23] Tarski, A., « Sur le terme primitif de la logistique », *Fundamenta Mathematicæ* IV, 1923, 59-74.
- [Thi87] Thijsse, G. C. E., « Partial Logic and Modal Logic : A Systematic Survey », manuscrit, 1987.
- [Tic82] Tichý, P., « Foundations of Partial Type Theory », *Reports on Mathematical Logic*, 14, 59-72, 1982.

François LEPAGE

Chapitre 13

Conditionnels, quantificateurs et inférences

Introduction

La présente intervention exploite l'étroite analogie qui existe entre l'analyse des quantificateurs et les conditionnels. Cette analogie fut pour la première fois mise en évidence par Johan van Benthem [1984]. Il s'agit de considérer que tout énoncé conditionnel de la forme *Si X, (alors) Y* exprime une certaine relation entre l'ensemble des situations vérifiant *X* et l'ensemble des situations vérifiant *Y*. En d'autres termes, la particule *Si* dénote un *quantificateur généralisé binaire*, c'est-à-dire une fonction qui assigne à chaque univers non vide (en l'occurrence, à chaque univers de *situations*) une certaine relation binaire entre des sous-ensembles de cet univers. Un exemple de quantificateur qui serait plausible ici est *tout*, c'est-à-dire l'inclusion ; dans ce cas, étant donné un univers *E* de situations, l'énoncé *Si X, (alors) Y* est vrai si, et seulement si, toutes les situations de *E* vérifiant *X* sont également des situations vérifiant *Y*.

L'analogie entre quantificateur et conditionnel peut être exploitée au moins de deux façons. D'abord, on peut tenter de déterminer parmi tous les quantificateurs possibles ceux qui peuvent être considérés comme des relations conditionnelles. Cette tâche, qui relève de la *logique inverse*, procède en deux temps : d'abord formuler en termes de propriétés de quantificateurs un certain nombre d'intuitions *a priori* censées caractériser les relations conditionnelles ; ensuite déterminer

précisément les quantificateurs qui satisfont toutes ces propriétés. Au terme de cette recherche, on peut naturellement s'engager dans une *logique directe*, c'est-à-dire l'étude des différentes théories axiomatiques des quantificateurs sélectionnés. Comme on le voit, aucune de ces deux perspectives ne présuppose qu'il existe une « unique et véritable » logique conditionnelle, mais plutôt qu'il existe un ensemble d'intuitions plus ou moins spécifiques des conditionnels et que ces intuitions peuvent être exprimées et étudiées de différentes manières.

Quant à notre intervention, elle ressortit d'avantage à la logique directe qu'à la logique inverse. En effet notre propos est de présenter certains résultats nouveaux concernant les logiques conditionnelles déterminées par certains quantificateurs. Les quantificateurs qui retiennent notre attention sont *tout*, *tout ou au moins* un et *au moins la moitié*. Si la logique déterminée par *tout* nous est familière (il s'agit ni plus ni moins de la logique de l'implication stricte), les autres nous le sont beaucoup moins, d'où l'intérêt d'en connaître plus à leur sujet.

Pour introduire les notions que nous utiliserons régulièrement par la suite, nous allons d'abord présenter brièvement certains éléments fondamentaux de la théorie des quantificateurs généralisés, pour ensuite expliquer plus en détail l'analogie entre les quantificateurs et les conditionnels, ce qui nous conduira à identifier un certain nombre de quantificateurs qui se comportent comme d'authentiques relations conditionnelles, dont notamment les trois quantificateurs sus-nommés. Quel que soit le quantificateur considéré, l'étude de ses propriétés inférentielles dépend d'un choix préalable de formalisme. Dans la deuxième section nous présentons les formalismes que nous utiliserons. Deux formalismes retiennent notre attention. Le premier formalisme (FCE) comprend uniquement des formules conditionnelles de la forme *Si X, (alors) Y*, où *X, Y* sont des termes Booléens ; le second (FCB) inclut FCE et comprend de plus des composés logiques de formules conditionnelles. Les troisième et quatrième sections constituent une étude systématique des logiques conditionnelles déterminées par nos quantificateurs. La troisième section porte plus spécifiquement sur les logiques déterminées dans le cadre du formalisme FCE. Ces logiques ont déjà été assez bien étudiées jusqu'à présent (van Benthem [1984], [1986] et Lapierre [1991a]) ; nous allons donc nous contenter de résumer les derniers résultats les concernant. La quatrième section porte sur les logiques déterminées dans le cadre du formalisme FCB. Nous y présentons notamment des résultats intéressants portant sur l'axiomatisation de ces logiques.

1. Quantificateurs et conditionnels

1.1. *Qu'est-ce qu'un quantificateur généralisé ?*

La première utilisation explicite de la notion de quantificateur généralisé remonte à Lindström [1996]. L'application de cette notion était toutefois restreinte à l'étude des langages formalisés. On la retrouve cependant implicitement appliquée à l'étude des langues naturelles dans le grand article « The Proper Treatment of Quantification

in Ordinary English » de Richard Montague [1973]¹. La première analyse systématique et détaillée de cette notion ne fut faite que dix années plus tard par Barwise et Cooper [1981]. L'idée fondamentale est la suivante. Plusieurs phrases des langues naturelles (« Au moins un étudiant réussit », « La plupart des étudiants réussissent », etc.) peuvent être analysés comme étant de la forme QXY , où X est le sujet (« étudiant »), Y est le prédicat (« réussir ») et Q est une expression de quantification (« au moins un », « la plupart »). En théorie des modèles, on considère habituellement que le sujet et le prédicat dénotent des sous-ensembles d'un certain univers non vide E d'individus. Qu'en est-il de l'expression de quantification ? Étant donné la précédente analyse, il semble naturel de considérer que l'expression de quantification dénote une relation entre des sous-ensembles de cet univers. Par exemple :

au moins un dénote $\{ \langle A, B \rangle \mid A, B \subseteq E \text{ \& } A \cap B \neq \emptyset \}$;
la plupart dénote $\{ \langle A, B \rangle \mid A, B \subseteq E \text{ \& } |A \cap B| > |A - B| \}$.

En effet, étant donné un modèle M basé sur E et $[[X]]$, $[[Y]]$ les dénnotations du sujet X et du prédicat Y dans ce modèle, nous avons que :

au moins un XY est vrai dans $M \Leftrightarrow [[X]] \cap [[Y]] \neq \emptyset$;
la plupart XY est vrai dans $M \Leftrightarrow |[[X]] \cap [[Y]]| > |[[X]] - [[Y]]|$.

Avant de poursuivre, faisons une petite convention de notation. Nous écrirons QXY pour désigner une expression de quantification suivie des expressions linguistiques des types appropriés. Pour désigner une relation opérant sur des ensembles (les dénnotations de ces expressions linguistiques), nous écrirons $Q_E AB$ (où parfois seulement QAB , comme nous le verrons). Cet abus de notation est de loin préférable à une multiplication de symboles spéciaux.

Définissons maintenant la notion de quantificateur généralisé dans toute sa généralité. Par un *quantificateur généralisé n -aire* ($n \geq 1$), on entend une fonction Q qui assigne, à chaque univers non vide E , une relation n -aire Q_E entre des sous-ensembles de E ². La grande majorité des quantificateurs des langues naturelles sont binaires. Les quantificateurs *au moins un* et *la plupart* en sont des exemples. Le quantificateur *au moins un* assigne à chaque univers E l'ensemble des paires ordonnées $\langle A, B \rangle$ telles que $A, B \subseteq E$ et $A \cap B \neq \emptyset$; *la plupart* assigne à chaque univers E l'ensemble des paires ordonnées $\langle A, B \rangle$ telles que $A, B \subseteq E$ et $|A \cap B| > |A - B|$.

1. Nous disons bien *implicitement* car Richard Montague n'utilise pas le terme «generalized quantifier» ni ne mentionne les travaux de Lindström.

2. On appelle parfois *quantificateur local* la valeur Q_E d'un quantificateur Q pour un univers E .

Une classe importante de quantificateurs est celle des *quantificateurs logiques*. Un quantificateur binaire est dit logique s'il satisfait les trois propriétés que voici :

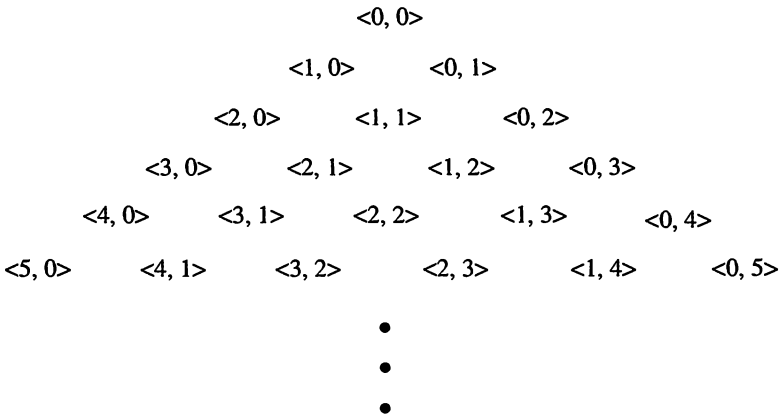
CONS	Pour tous les $A, B \subseteq E$, $Q_E AB \Leftrightarrow Q_E A(B \cap A)$	<i>Conservativité</i>
EXT	Si $A, B \subseteq E \subseteq E'$, alors $Q_E AB \Leftrightarrow Q_{E'} AB$	<i>Extension</i>
QUANT	Pour toute permutation $\pi: E \rightarrow E$ et $A, B \subseteq E$, $Q_E AB \Leftrightarrow Q_E \pi[A]\pi[B]$	<i>Quantité</i>

On vérifie facilement que selon ces trois propriétés, le fait que $Q_E AB$ dépend uniquement des cardinalités $|A \cap B|$ et $|A - B|$ ³. Par conséquent, Q peut être entièrement décrit par une relation binaire Q^c entre *nombre cardinal* telle que $Q^c xy$ si et seulement s'il existe trois ensemble A, B, E tels que $A, B \subseteq E$, $Q_E AB$, $x = |A - B|$ et $y = |A \cap B|$. Ainsi, pour reprendre nos mêmes exemples :

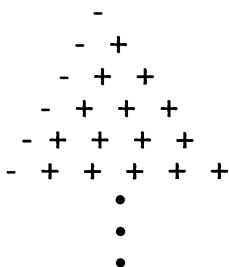
$$\begin{aligned} \text{au moins un}^c xy &\Leftrightarrow y \neq 0 ; \\ \text{la plupart}^c xy &\Leftrightarrow y > x. \end{aligned}$$

La caractérisation numérique devient encore plus intéressante si on ne considère que des univers *finis*. Les relations numériques correspondant aux quantificateurs binaires logiques deviennent en effet des sous-ensembles finis de $\omega \times \omega$. Or $\omega \times \omega$ peut être représenté par un *arbre numérique*, où chaque paire $\langle x, y \rangle$ a deux *successeurs immédiats* $\langle x + 1, y \rangle$ et $\langle x, y + 1 \rangle$, lesquels sont les *prédécesseurs immédiats* de la paire $\langle x + 1, y + 1 \rangle$.

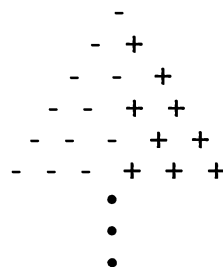
3. En effet, EXT signifie la neutralité du contexte : le fait que $Q_E AB$ dépend uniquement des ensembles A, B . Donc en conjonction avec CONS, EXT signifie que le fait que $Q_E AB$ dépend uniquement des ensembles $A \cap B$ et $A - B$. D'un autre côté, selon QUANT le fait que $Q_E AB$ dépend uniquement des cardinalités des ensembles $A \cap B$, $A - B$, $B - A$ et $E - (A \cup B)$ (la nature des éléments ne joue absolument aucun rôle). Donc la conjonction de CONS, EXT et QUANT signifie que le fait que $Q_E AB$ dépend uniquement du nombre d'éléments dans $A \cap B$ et du nombre d'éléments dans $A - B$.



On peut ainsi représenter n'importe quel quantificateur binaire logique Q par un sous-ensemble de cet arbre, en indiquant par le signe «+» que la paire $\langle x, y \rangle$ appartient à la relation numérique correspondante, et par le signe «-» que la paire $\langle x, y \rangle$ n'appartient pas à cette relation. Par exemple :



Au moins un



La plupart

1.2. Analogie entre les conditionnels et les quantificateurs

Comme nous l'avons annoncé dans notre introduction, nous analysons les énoncés conditionnels *Si X, (alors) Y* comme des expressions de relation entre des ensembles de situations. Plus précisément, étant donné un univers non vide E de situations (ou *événements*, *mondes possibles*, etc.), tout conditionnel *Si X, (alors) Y* exprime que $Si_E[[X]][[Y]]$, où Si_E est une relation binaire, appelée *relation conditionnelle*, entre des sous-ensembles de E et $[[X]]$, $[[Y]]$ sont des sous-ensembles

de E tenant lieu des dénnotations des termes X, Y respectivement⁴. Selon cette analyse, Si_E est la valeur d'un quantificateur binaire Si pour l'univers E . Par définition, Si est un *quantificateur conditionnel*.

Ce ne sont évidemment pas tous les quantificateurs binaires qui sont conditionnels. Nos intuitions concernant les relations conditionnelles interdisent à plusieurs quantificateurs ce privilège. Dans le but de circonscrire la classe des quantificateurs admissibles, J. van Benthem a suggéré une courte liste de propriétés qui traduisent un certain nombre de ces intuitions. Les trois propriétés suivantes semblent être partagées par toutes les relations conditionnelles, sinon par la plupart.

D'abord, il y a la *Confirmation*. Supposons que nous sommes disposés à donner notre assentiment à un énoncé conditionnel $Si X, alors Y$ sur la base de nos croyances à propos des extensions des termes X, Y . Nous devons donc estimer que l'importance relative des *cas positifs*, c'est-à-dire les cas tombant à la fois sous X et Y , domine l'importance relative des *cas négatifs*, c'est-à-dire les cas tombant sous X mais non sous Y . Supposons maintenant que des informations additionnelles nous obligent à réviser nos croyances à propos des extensions des termes. En principe, une révision à la hausse du nombre de cas positifs, ou une révision à la baisse du nombre de cas négatifs, ne devrait pas modifier notre jugement concernant le conditionnel. Cette intuition peut se formuler de manière plus concrète en termes de propriétés des relations conditionnelles elles-mêmes (avec $A, B, C \subseteq E$) :

C1	$Si_E AB \Rightarrow Si_E A(B \cup C)$	<i>addition de cas positifs</i>
C2	$Si_E A(B \cap C) \Rightarrow Si_E (A \cap B)C$	<i>retrait de cas négatifs</i>
C3	$Si_E AB \Rightarrow Si_E (A \cup C)(B \cup C)$	<i>à la fois addition de cas positifs et retrait de cas négatifs</i>
C4	$A \subseteq B \Rightarrow Si_E AB$	<i>l'absence de cas négatifs confirme ipso facto le conditionnel !</i>

Incidentement, une conséquence immédiate de C4 est la *réflexivité* : $Si_E AA$ pour tout $A \subseteq E$.

De toute évidence, toute proposition possible implique certaines propositions mais pas toutes, car autrement le possible impliquerait n'importe quoi, y compris l'impossible, ce qui est absurde. De cette observation on peut conclure que les relations conditionnelles satisfont cette autre propriété :

4. Intuitivement, les termes X, Y sont vus comme des *énoncés*. Leurs dénnotations correspondent à ce que certains logiciens appellent *propositions*, caractérisées comme des ensembles de mondes possibles.

VAR Pour tout $A \subseteq E$, A étant non vide, il existe $B, C \subseteq E$
 tels que $Si_E AB$ et non $Si_E AC$ Variété

Soulignons que C1 et VAR entraînent la *cohérence* : $Si_E A \emptyset$ seulement si $A = \emptyset$.

Enfin, tout énoncé conditionnel nous invite à faire une espèce de voyage mental dans les situations de l'antécédent, si bien que dans l'assertion d'un conditionnel, l'assertion du conséquent est toujours faite implicitement en conjonction avec l'antécédent. Par exemple, le conditionnel « S'il pleuvait, je sortirais », est équivalent au conditionnel « S'il pleuvait, je sortirais et il pleuvrait ». Un autre exemple est l'énoncé universel « Tous les corbeaux sont noirs », logiquement équivalent à l'énoncé « Tous les corbeaux sont des corbeaux noirs ». Ainsi les relations conditionnelles semblent posséder la propriété Conservativité (CONS), déjà formulée dans la précédente section.

Il est facile de vérifier que les quantificateurs *tout*, *tout ou au moins un* et *au moins la moitié*, définis ci-dessous, satisfont Confirmation, Variété et Conservativité. Cette dernière propriété découle de la logicalité, laquelle nous permet d'omettre de spécifier l'univers de la relation :

$$\begin{aligned} tout AB &\Leftrightarrow A \subseteq B ; \\ tout\ ou\ au\ moins\ un AB &\Leftrightarrow A \subseteq B\ ou\ A \cap B \neq \emptyset ; \\ au\ moins\ la\ moitié AB &\Leftrightarrow |A \cap B| \geq |A - B|. \end{aligned}$$

Toutefois, d'autres quantificateurs passent aussi le test. Parmi les quantificateurs logiques, nous avons notamment *tout ou la plupart* ; parmi les quantificateurs *non logiques*, nous avons entre autres *tout préféré*, défini pour des univers structurés $\langle E, R \rangle$ où R est une relation binaire sur E :

$$tout\ préféré_{\langle E, R \rangle} AB \Leftrightarrow \text{tous les éléments } R\text{-maximaux dans } A \text{ sont dans } B.$$

Ce dernier quantificateur est particulièrement intéressant. Nous savons en effet (voir van Benthem [1984]) que restreint à des univers $\langle E, R \rangle$ avec E fini et R *irréflexive* et *transitive*, il valide précisément la *Logique subjunctive de base S* de Burgess [1981]. Nous reparlerons de cette logique dans la section 3.

Évidemment, le nombre de quantificateurs conditionnels admissibles peut être diminué en imposant des contraintes additionnelles. Parmi celles que l'on peut envisager, il y a *Uniformité*. Soit $A, B \subseteq E$ tels que $Si_E AB$ (ou non $Si_E AB$, peu importe) et soit deux situations distinctes $a, b \in E - (A \cup B)$. Effectuons les opérations suivantes : d'abord vérifions si $Si_E (A \cup \{a\})B$; ensuite vérifions si $Si_E (A \cup \{a\})(B \cup \{a\})$; vérifions enfin si $Si_E (A \cup \{a, b\})(B \cup \{b\})$. Appelons respectivement O1, O2 et O3 la première, deuxième et troisième de ces opérations. Ces opérations et leurs résultats peuvent être représentés dans cette figure losangique («+» = oui, «-» = non) :

$SiEAB (+/-)$

O1 (+/-)

O2 (+/-)

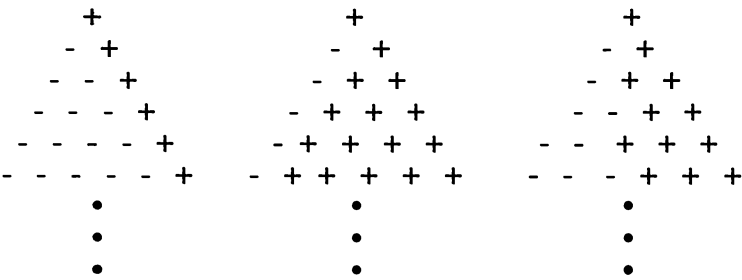
O3 (+/-)

Par rapport à l'état initial, indiqué au sommet, la relation SiE a été testée avec un cas négatif de plus dans O1, avec un cas positif de plus dans O2, et enfin, avec à la fois un cas positif et un cas négatif de plus dans O3. Maintenant, il semble raisonnable de s'attendre à ce que le résultat de O3 soit *indépendant* des ensembles A, B initiaux et donc qu'il soit *uniquement déterminé* par les *résultats* de O1 et O2. Ainsi, le nombre de répartitions possibles de + et de - dans la figure ci-dessus est 2^4 . De ce nombre, dix sont exclues par Confirmation. Les six possibilités qui restent sont :

+		+		+		-		-		-	
+	+	-	+	-	+	-	+	-	+	-	-
+		+		-		-		+		-	

Or, si la relation est uniforme, les deuxième et troisième (et les quatrième et cinquième) possibilités ne se produisent pas ensemble dans la relation. Dans le cas où Si est logique, cela revient à dire que pour deux nombres cardinaux quelconques x et y , l'appartenance ou la non appartenance de la paire $\langle x + 1, y + 1 \rangle$ à la relation numérique correspondante dépend uniquement des résultats portant sur les paires $\langle x, y \rangle$, $\langle x + 1, y \rangle$ et $\langle x, y + 1 \rangle$.

L'examen des arbres numériques de *tout*, *tout ou au moins un* et *au moins la moitié* nous permet de constater que ces trois quantificateurs sont uniformes :



tout *tout ou au moins un* *au moins la moitié*

Ce qui est plus intéressant cependant, c'est qu'en dessinant les arbres des quantificateurs logiques qui satisfont à la fois Confirmation, Variété et Uniformité, on obtient finalement uniquement les arbres de ces trois quantificateurs ⁵.

Nous pourrions aussi avoir la tentation d'imposer une autre contrainte, intuitivement désirable dans bien des cas, à savoir la Conjonction :

$$\text{Si}_E AB, \text{Si}_E AC \text{ seulement si } \text{Si}_E A(B \cap C).$$

Mais réflexion faite, il existe des relations conditionnelles plausibles qui n'ont pas cette propriété. Par exemple, il y a les conditionnels contrefactuels possibilistes de la forme *S'il avait été le cas que X, il aurait pu être le cas que Y* ⁶. Certaines analyses probabilistes des conditionnels admettent aussi des modèles réfutant Conjonction. L'univers E des situations peut être associé à une fonction de probabilité P de $\wp(E)$ dans $[0, 1]$ satisfaisant certaines contraintes de base telles que la *normalité* : $P(\emptyset) = 0$ et $P(E) = 1$; ainsi que l'*additivité* : pour $A, B \subseteq E$, $P(A \cup B) = (P(A) + P(B)) - P(A \cap B)$. Soit maintenant la relation conditionnelle $\text{Si}_{E,P}$ telle que $\text{Si}_{E,P} AB$ si et seulement si $P(B|A) \geq 0.5$. Cette relation conditionnelle ne satisfait pas nécessairement Conjonction : on peut en effet trouver des contre-exemples, notamment si P monotone en fonction de la cardinalité des ensembles, c'est-à-dire lorsque $P(A) \geq P(B)$ si $|A| \geq |B|$.

Nous devons toutefois reconnaître que Conjonction n'est pas la seule propriété désirable mais non universelle. Par exemple, Variété n'a pas besoin d'être satisfaite dans la perspective probabiliste, où les ensembles non vides peuvent recevoir la probabilité zéro (car ce qui est logiquement possible n'est pas forcément probable). Cependant, à la différence de Variété, Conjonction réduit dramatiquement la classe des quantificateurs admissibles. Ainsi, dans les univers *finis*, il s'avère que *tout* est le seul quantificateur logique et cohérent satisfaisant Conjonction ⁷. Telle est la principale raison qui nous incite à ignorer cette contrainte.

5. Nous ne reformulons pas ici la preuve, incidemment simple, que l'on peut trouver dans van Benthem [1984] et aussi dans van Benthem [1986], chap. 4, sec. 3.

6. Cela est clair dans l'exemple suivant : « Je n'ai pas gagné simplement parce que je n'ai pas joué ; bien sûr, si j'avais joué, j'aurais pu gagner, mais j'aurais pu perdre aussi ».

7. La preuve est simple. Supposons que Q satisfait Conjonction, que QAB mais que $A \not\subseteq B$. Donc $(A \cap B) \subset A$ et par Conservativité, $QA(A \cap B)$. Donc par Quantité, pour tout $C \subseteq E$ tel que $|A \cap C| = |A \cap B|$, $QA(A \cap C)$ et donc $QA((A \cap B) \cap C)$. On peut alors continuer ce raisonnement jusqu'à ce qu'on obtienne $QA\emptyset$. Donc par la cohérence, $A = \emptyset$. Donc $A \subseteq B$: une contradiction.

En combinant Confirmation, Variété et Uniformité, nous obtenons donc seulement trois quantificateurs conditionnels logiques : *tout*, *tout ou au moins un* et *au moins la moitié*. L'étude des propriétés inférentielles de ces trois quantificateurs suppose évidemment un choix préalable de formalisme. Bien qu'il existe de nombreuses possibilités, allant du formalisme le faible jusqu'au plus fort, nous allons nous limiter ici à deux formalismes.

2. Quantificateurs et inférences

Pour pouvoir exprimer des propriétés fondamentales telle la Confirmation ou la Conservativité, nous avons besoin d'un formalisme qui comprend au moins des formules conditionnelles de la forme élémentaire *Si X, (alors) Y* (de manière plus concise, *SiXY*), où *X, Y* sont des termes Booléens atomiques ou complexes. Soit FCE le formalisme qui ne comprend que des formules conditionnelles de cette forme élémentaire. Les inférences que l'on peut étudier dans le cadre de FCE sont donc du type :

$$2.1. \quad \frac{SiX_1Y_1 \dots SiX_nY_n}{SiXY}$$

où les équivalents Booléens sont intersubstituables sans restriction. Appartiennent à ce type d'inférence les « figures syllogistiques pures », comme la loi de la *transitivité* :

$$\frac{SiXY \quad SiYZ}{SiXZ}$$

ainsi que des principes où interviennent des termes Booléens complexes, tel le principe de la *conjonction* :

$$\frac{SiXY \quad SiXZ}{SiX(Y \wedge Z)}$$

Une inférence du type 2.1 peut être valide selon un quantificateur mais non selon un autre. La notion de validité selon un quantificateur peut être formellement définie de la manière habituelle sur la base de la notion de *formule vraie dans un modèle et selon un quantificateur*. Pour ce faire, nous allons interpréter FCE par des modèles de la forme $\langle E, [[]] \rangle$, où *E* est un ensemble non vide et *fini* de situations et $[[]]$ est une fonction qui assigne à chaque terme Booléen *X* un sous ensemble $[[X]]$ de *E* selon les règles que voici :

- i) si *X* est atomique, alors $[[X]] \subseteq E$;
- ii) $[[\neg X]] = E - [[X]]$;
- iii) $[[X \wedge Y]] = [[X]] \cap [[Y]]$;

$$\text{iv) } [[X \vee Y]] = [[X]] \cup [[Y]].$$

Ici, la restriction à des modèles finis est motivée par deux considérations. Premièrement, le cadre finitaire est en soi commode car il simplifie les résultats et les preuves. Deuxièmement, contrairement à certains quantificateurs comme *infinitement plusieurs*, aucun des quantificateurs *tout*, *tout ou au moins un* et *au moins la moitié* ne requiert des univers finis. Dans ce contexte, le principe d'économie nous incite à ignorer le domaine de l'infini, d'autant plus que la plupart des résultats que l'on obtient dans le fini se généralisent à des univers infinis.

Nous pouvons maintenant définir la notion de *vérité* (relativement à un modèle $M = \langle E, [[]] \rangle$ et selon un quantificateur conditionnel Si) et de *validité* selon un quantificateur conditionnel Si .

$$\bullet \text{ Vérité : } M \models SiXY \Leftrightarrow Si_E[[X]][[Y]].$$

$$\bullet \text{ Inférence valide : } SiX_1Y_1 \dots SiX_nY_n \models SiXY \Leftrightarrow M \models SiXY \text{ pour tout modèle } M \text{ tel que pour chaque } i \ (1 \leq i \leq n), M \models SiX_iY_i.$$

$$\bullet \text{ Validité : } \models SiXY \Leftrightarrow M \models SiXY \text{ pour tout modèle } M.$$

Chacun des quantificateurs *tout*, *tout ou au moins un* et *au moins la moitié* détermine, dans le formalisme FCE, une logique conditionnelle. Nous appelons *Logique conditionnelle classique C* la logique déterminée par *tout*, *Logique conditionnelle exemplaire E* la logique déterminée par *tout ou au moins un*, et *Logique conditionnelle quasi démocratique QD* la logique déterminée par *au moins la moitié*.

L'expression de certains principes bien connus en logique des conditionnels nécessite un formalisme autorisant l'emploi d'opérations logiques à l'extérieur des formules conditionnelles elles-mêmes. Un exemple est le *Tiers exclu conditionnel* :

$$\text{TEC} \quad SiXY \vee SiX \neg Y$$

Afin de pouvoir exprimer des formules de cette sorte, nous utiliserons le formalisme FCB des *formules conditionnelles Booléennes*. Ce formalisme n'est ni plus ni moins que la plus petite classe FCB telle que : i) $FCE \subseteq FCB$; ii) si ϕ, ψ appartiennent à FCB, alors $\neg\phi, (\phi \vee \psi)$ appartiennent à FCB. Dans le cadre de ce formalisme, les inférences que l'on peut étudier sont du type :

$$\begin{array}{c} 2.1. \quad \phi_1 \dots \phi_n \\ \hline \psi \end{array}$$

où $\phi_1, \dots, \phi_n, \psi \in FCB$. Évidemment, dans ce type d'inférences les termes Booléens équivalents ainsi que les formules logiquement équivalentes sont

intersubstituables sans restriction. En outre, l'application de toutes les règles classiquement valides de la déduction naturelle est autorisée. En particulier, toute inférence du type 2.1 implique la thèse :

$$\vdash (\neg\phi_1 \vee \dots \vee \neg\phi_n) \vee \psi$$

laquelle justifie, en vertu du *modus tollens*, l'inférence contraposée :

$$\frac{\neg\psi}{\neg\phi_1 \vee \dots \vee \neg\phi_n}$$

Par exemple, appartient au type 2.1 l'inférence suivante :

$$\frac{\neg SiXY}{SiX \rightarrow Y}$$

laquelle implique la thèse TEC, autorisant l'inférence contraposée :

$$\frac{\neg SiX \neg Y}{SiXY}$$

La notion de vérité dans un modèle et selon un quantificateur, définie pour les formules conditionnelles de FCE, peut être facilement étendue pour les formules conditionnelles Booléennes. À partir de cette notion, la notion de formule conditionnelle Booléenne valide selon un quantificateur est définie de la manière habituelle.

Encore une fois, chacun des quantificateurs *tout*, *tout ou au moins un* et *au moins la moitié* détermine, dans le formalisme FCB, une logique conditionnelle. Ainsi nous appelons *Logique conditionnelle classique Booléenne CB* la logique de *tout*, *Logique conditionnelle exemplaire Booléenne EB* la logique de *tout ou au moins un*, et *Logique conditionnelle quasi démocratique Booléenne QDB* la logique de *au moins la moitié*. Avant de nous pencher sur ces logiques, résumons d'abord les derniers faits connus concernant *C*, *E* et *QD*.

3. Les logiques *C*, *E* et *QD*

Par définition, chacun de nos trois quantificateurs conditionnels satisfait les propriétés Confirmation, Variété et Conservativité. Or, ces trois propriétés déterminent à elles seules une logique conditionnelle, appelée *Logique conditionnelle minimale M*, dont les principes appartiennent au type d'inférences 2.1. Les principes élémentaires de *M* sont les suivants (on notera une certaine redondance dans cette présentation) :

R	$SiXX$	<i>Réflexivité</i>
CONS1	$\frac{SiXY}{Si\ X(Y \wedge X)}$	<i>Conservativité</i>
CONS2	$\frac{SiX(Y \wedge X)}{SiXY}$	
C1	$\frac{Si\ XY}{Si\ X(Y \vee Z)}$	<i>Confirmation</i>
C2	$\frac{SiX(Y \wedge Z)}{Si\ (X \wedge Y)Z}$	
C3	$\frac{SiXY}{Si\ (X \vee Z)(Y \vee Z)}$	

On constate que tous les principes de M sont à zéro prémisses ou à une seule prémisses. Nous pouvons déjà dériver dans cette logique certains principes utiles⁸ (les preuves sont toutes simples).

C4	$\frac{Si(X \wedge \neg Y)\perp}{SiXY}$	<i>Confirmation no.4</i>
RTA	$\frac{SiX\perp}{Si(X \wedge Y)\perp}$	<i>Renforcement trivial de l'antécédent</i>
IMP	$\frac{Si\ T(X \wedge Y)}{SiXY}$	<i>Importation</i>
SIMP	$\frac{SiX(Y \wedge Z)}{SiXY}$	<i>Simplification</i>

Toutes les autres logiques sont bien sûr des extensions conservatives de M . Pour commencer, la logique C contient évidemment, outre tous les principes de M , les principes additionnels suivants.

8. Nous utilisons le symbole $\perp$ pour abrégier $X \wedge \neg X$, où X est n'importe quel terme. D'autre part, nous utilisons le symbole T pour abrégier $X \vee \neg X$. La connexion entre C4 ci-dessous et la propriété Confirmation 4 ($A \subseteq B$ seulement si $SiEAB$) est la suivante. Selon n'importe quel quantificateur conditionnel cohérent, $SiX\perp$ signifie exactement que X dénote l'ensemble vide. Donc $Si(X \wedge \neg Y)\perp$ signifie exactement que la dénotation de X est incluse dans celle de Y . Donc $Si(X \wedge \neg Y)\perp$ seulement si $SiXY$, selon Confirmation 4.

CNT	$\frac{SiXY}{Si \neg Y \neg X}$	Contraposition
TRN	$\frac{SiXY \quad SiYZ}{SiXZ}$	Transitivité
RA	$\frac{SiXY}{Si (X \wedge Z)Y}$	Renforcement de l'antécédent
CNJ	$\frac{SiXY \quad SiXZ}{SiX(Y \wedge Z)}$	Conjonction
DSJ	$\frac{SiXZ \quad SiYZ}{Si (X \vee Y)Z}$	Disjonction

De toute évidence C correspond à la logique de l'implication stricte dans le système S_5 . Nous avons également une axiomatisation complète de C .

3.1. Théorème

Tous les principes élémentaires de M plus TRN constituent une axiomatisation complète pour l'inclusion (*tout*).

Preuve. D'abord, on peut facilement vérifier que DSJ ainsi que :

CC4	$\frac{SiXY}{Si(X \wedge \neg Y)\perp}$	Converse de C4

sont dérivables dans $M+TRN$.

Supposons maintenant que $SiX_1Y_1 \dots SiX_nY_n \not\vdash SiXY$ dans $M+TRN$. Puisqu'il y a seulement un nombre fini de termes impliqués ici, il doit exister une algèbre Booléenne finie - donc une algèbre atomique et complète - $(U, \wedge, \vee, 0, 1, ')$, munie d'une relation binaire additionnelle Q ayant ni plus ni moins toutes les propriétés formelles de Si dans $M+TRN$ et selon laquelle, par l'entremise d'une assignation V allant des termes dans U , chaque prémisses est vérifiée et la conclusion est réfutée. Or nous savons que cette algèbre Booléenne est représentée isomorphiquement dans l'algèbre $\wp(\text{At}(U))$ par la fonction $\theta : B \rightarrow \wp(\text{At}(U))$ telle que $\theta(x) = \{y \in \text{At}(U) \mid y \wedge x' = 0\}$. Donc par l'entremise de θ , Q devient la relation binaire $Q = \{\langle \theta(x), \theta(y) \rangle \mid Qxy\}$ ayant les propriétés suivantes :

P1. QAA	(R)
P2. $QAC, QBC \Rightarrow Q(A \cup B)C$	(DSJ)
P3. $QA\emptyset \Rightarrow Q(A \cap B)\emptyset$	(RTA)
P4. $Q(A \cap B')\emptyset \Rightarrow QAB$	(C4)
P5. $QAB \Leftrightarrow Q(A \cap B')\emptyset$	(C4+CC4)

Or Q peut être complètement remplacée par le prédicat unaire Q^* sur $\wp(\text{At}(U))$ tel que Q^*A si et seulement si $QA\emptyset$. Par P5 nous avons en effet que $Q^*A \cap B'$ si et seulement si QAB , d'où les propriétés suivantes :

- P1*. $Q^*\emptyset$ (de P1 - $Q\emptyset\emptyset$)
 P2*. $Q^*A, Q^*B \Rightarrow Q^*A \cup B$ (de P2)
 P3*. $Q^*A \Rightarrow Q^*A \cap B$ (de P3)

À ce stade, il est clair que nous obtiendrons le résultat désiré si nous pouvions montrer que pour tous les $A, B \subseteq \text{At}(U)$, $Q^*A \cap B'$ si et seulement si $A \subseteq B$, car cela signifierait que le modèle $\langle \text{At}(U), [[]] \rangle$ tel que $[[Z]] = \theta(V(Z))$ pour tout terme atomique Z , réfute l'inférence selon l'inclusion. Malheureusement, si par P1* nous avons que $A \subseteq B$ seulement si $Q^*A \cap B'$, la converse n'est pas nécessairement vraie. Nous devons restreindre Q^* d'une certaine façon.

Soit $K = \bigcup \{A \mid Q^*A\}$. Donc Q^*K par P2*, le fait que $\text{At}(U)$ est fini et la définition de K . Donc Q^* restreint à $\wp(\text{At}(U) - K)$ contient seulement l'ensemble vide. Par conséquent pour tous les $A, B \subseteq \text{At}(U) - K$, $Q^*A \cap B'$ si et seulement si $A \subseteq B$. Soit $F = \lambda D. D - K$. On vérifie facilement que F est un homomorphisme Booléen de $\wp(\text{At}(U))$ dans $\wp(F(\text{At}(U)))$. En outre $F(\text{At}(U)) \neq \emptyset$, car autrement, en vertu de P3*, Q serait universelle, c'est-à-dire que nous aurions que Q^*A pour tout $A \subseteq \text{At}(U)$, et donc que QAB pour tous les $A, B \subseteq \text{At}(U)$, ce qui implique qu'il n'y aurait aucune formule réfutée selon Q^* , ce qui est contraire à l'hypothèse. Nous obtenons donc le résultat souhaité en montrant que pour tous les $A, B \subseteq \text{At}(U)$:

$$A \cap B' \Leftrightarrow Q^*F(A \cap B'),$$

car alors $QAB \Leftrightarrow Q^*A \cap B' \Leftrightarrow F(A) \subseteq F(B)$, ce qui implique qu'il existe un modèle *non vide*, à savoir $\langle F(\text{At}(U)), [[]] \rangle$ tel que $[[Z]] = F(\theta(V(Z)))$ pour tout terme atomique Z , qui réfute l'inférence selon l'inclusion. Voici la preuve.

D'abord, si $Q^*A \cap B'$, alors par P3*, $Q^*(A \cap B') - K$, i.e., $Q^*F(A \cap B')$. Conversement, supposons que $Q^*F(A \cap B')$. Puisque Q^*K , nous avons que $Q^*K \cup F(A \cap B')$, par P2*. Or, par la définition de F et identité Booléenne, $K \cup F(A \cap B') = K \cup (A \cap B')$. Donc $Q^*K \cup (A \cap B')$. Donc par P3*, $Q^*(K \cup (A \cap B')) \cap (A \cap B')$, et donc $Q^*A \cap B'$, par identité Booléenne. ■

Incidentement, la démonstration précédente nous indique que les principes R, RTA, DSJ, C4 et CC4 forment également une base axiomatique complète pour C.

Concernant E , la logique de *tout ou au moins un*, une axiomatisation complète reste à trouver. Toutefois, nous connaissons un certain nombre de choses intéressantes sur cette logique. D'abord, nous savons qu'elle est incluse dans C, comme l'indique le théorème suivant.

3.2. Théorème

Toute inférence valide selon *tout ou au moins un* est aussi une inférence valide selon l'inclusion (*tout*).

Preuve. Soit $SiX_1Y_1 \dots SiX_nY_n / SiXY$ une inférence réfutée selon l'inclusion dans un certain modèle. Nous avons d'emblée que toute ou au moins une X_i -situation est une Y_i -situation ($1 \leq i \leq n$). D'autre part, il existe au moins une X -situation qui n'est pas une Y -situation et donc, s'il n'y a aucune $X \wedge Y$ -situation, alors nous avons déjà un modèle qui réfute l'inférence à la fois selon *tout* et *tout ou au moins un*. Autrement, considérez le sous-modèle homomorphique comprenant seulement les non- $X \wedge Y$ -situations et où chacune d'elles se comporte exactement de la même façon relativement à leur appartenance (et non appartenance) aux extensions des termes atomiques. Dans ce nouveau modèle, chaque prémisses est encore vraie selon l'inclusion ; d'autre part, la conclusion $SiXY$ est toujours fausse selon l'inclusion, bien qu'il n'existe plus de $X \wedge Y$ -situation. Nous avons donc un modèle qui réfute l'inférence à la fois selon *tout* et *tout ou au moins un*. ■

La converse de ce théorème est évidemment fausse. Aucun des principes CNT, TRN, RA et CNJ n'est valide selon *tout ou au moins un*. Cependant DSJ est toujours valide, comme on peut facilement le vérifier. Voici d'autres principes valides selon *tout ou au moins un* (et donc également valides selon l'inclusion).

CNJP	$\frac{SiXY \quad Si(X \wedge Y)Z}{SiX(Y \wedge Z)}$	Conjonction prudente
SYMC	$\frac{SiTX \quad SiXY}{SiYX}$	Symétrie conditionnelle
TRSN	$\frac{SiTX \quad SiXY}{SiTY}$	Transmissibilité de la nécessité
ACA	$\frac{SiTX \quad SiXY}{Si(X \vee Z)Y}$	Affaiblissement conditionnel de l'antécédent

Parmi tous ces principes, certains sont plus fondamentaux que d'autres. En fait, on vérifie aisément que tous les principes élémentaires de M plus CNJP suffisent pour dériver tous les principes précédents, ainsi que DSJ. Cela nous incite à émettre la conjecture suivante.

3.3. Conjecture

Tous les principes élémentaires de M plus CNJP constituent une axiomatisation complète pour *tout ou au moins un*. En outre, des précédents faits et du théorème 3.2, nous inférons le théorème suivant.

3.4. Théorème. $M \subset E \subset C$.

On constate sans peine qu'à la différence de C , qui correspond à la logique de l'implication modale, E ne semble pas avoir d'analogue parmi les logiques conditionnelles déjà connues. Mais elle n'est toutefois pas complètement dénuée d'intérêt. D'abord, elle ne contient aucun des principes CNT, TRN et RA, non valides dans les logiques des conditionnels subjonctifs ou contraires aux faits. Elle n'est cependant pas une candidate sérieuse pour la logique subjonctive, car elle ne contient pas CNJ. Or ce fait indique-t-il qu'elle constitue une logique affaiblie par rapport à la logique subjonctive ? Considérons en particulier la *Logique subjonctive de base S* de Burgess - celle que nous avons évoquée dans la Section 1. Cette logique a l'heureuse particularité d'être présentée dans un formalisme équivalent à FCE ; elle est en effet axiomatisée par les cinq principes R, CNJ, DSJ, SIMP et

RPA	$\frac{SiXZ \quad SiXY}{Si(X \wedge Z)Y}$	Renforcement prudent de l'antécédent

On peut vérifier que cette base axiomatique est déductivement équivalente à M plus CNJ.

En outre, CNJP est un principe de S , dérivable de $R+SIMP+CI+DSJ+CNJ$:

$$\begin{array}{l}
 \frac{Si(X \wedge \neg Y)(X \wedge \neg Y)}{Si(X \wedge \neg Y)\neg Y} \quad \frac{Si(X \wedge Y)Z}{Si(X \wedge Y)(Z \vee \neg Y)} \\
 \frac{Si(X \wedge \neg Y)(\neg Y \vee Z)}{Si((X \wedge \neg Y) \vee (X \wedge Y))(\neg Y \vee Z)} \\
 \frac{SiXY \quad SiX(\neg Y \vee Z)}{SiX((\neg Y \vee Z) \wedge Y)} \\
 SiX(Z \wedge Y) \\
 SiX(Y \wedge Z)
 \end{array}$$

Nous pouvons donc nous poser cette question : $E \subset S$? Jusqu'à présent aucun contre-exemple de cette inclusion n'a été trouvé. Mais à défaut d'une axiomatisation complète de E , le seul espoir qui reste pour établir cette inclusion est l'argument sémantique, qui pose malheureusement certaines difficultés⁹.

9. Depuis la rédaction de cet article, nous avons pu établir l'inclusion de E dans S . Les détails ont été publiés dans S. Lapierre [1991b]. Ce dernier article a en fait été rédigé en 1993, à titre de compte rendu d'une communication donnée dans le cadre l'*International Symposium on Epistemic Logic*, tenu à l'Université de Liège en mars 1993. L'année 1991 comme date de publication est une contrainte de la revue.

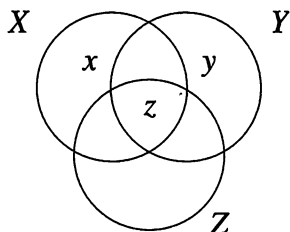
La logique *QD* présente aussi un intérêt non négligeable. Au début, ses principes fondamentaux semblaient difficiles à localiser. Néanmoins, nous savions que *QD* est aussi incluse dans *C*.

3.5. Théorème (van Benthem [1984])

Toute inférence valide selon *au moins la moitié* est aussi une inférence valide selon l'inclusion (*tout*).

Preuve. Voici une démonstration plus simple que l'originale. Soit $SiX_1Y_1 \dots SiX_nY_n / SiXY$ une inférence réfutée selon l'inclusion dans un certain modèle. Transformons ce modèle en un sous-modèle qui réfute la même inférence à la fois selon *tout* et *tout ou au moins un* (théorème 3.2). Dans ce modèle chaque prémisses est vérifiée selon *au moins la moitié*, puisqu'elle est vérifiée selon l'inclusion ; en outre, la conclusion est réfutée selon *au moins la moitié*, puisqu'elle est réfutée selon *tout ou au moins un*. ■

La converse de ce théorème est également fausse. On constate facilement qu'aucun des principes CNT, TRN, RA et CNJ n'est valide selon *au moins la moitié*. Peut être plus surprenant est le fait que DSJ n'est pas valide non plus, comme en témoigne ce contre-exemple (les extensions de *X*, *Y* et *Z* sont telles qu'indiquées) :



Toutefois, *QD* ne coïncide pas avec *M*. D'abord quelques petits calculs nous montrent que :

$$|(A \Delta B) \cap B| \geq |(A \Delta B) - B| \Leftrightarrow |B - A| \geq |A - B| \Leftrightarrow |B| \geq |A|$$

où $A \Delta B$ abrège la différence symétrique $(A - B) \cup (B - A)$. Soit maintenant ces lois bien connues de l'arithmétique additive avec inégalité restreinte aux clauses universelles :

A1. $x \leq x$

A2. $x \leq y \leq z \Rightarrow x \leq z$

A3. $x \leq x + y$

A4. $x \leq y \Rightarrow x + z \leq y + z$

A5. $x + z \leq y + z \Rightarrow x \leq y$

À la lumière de l'équivalence ci-dessus, on vérifie facilement que les cinq principes que voici, qui ne sont que les traductions respectives de A1-A5, sont valides selon *au moins la moitié* (où $X \Delta Y$ abrège $(X \wedge \neg Y) \vee (Y \wedge \neg X)$) :

$$A1' \quad Si(X \Delta X)X$$

$$A2' \quad \frac{Si(X \Delta Y)Y \quad Si(Y \Delta Z)Z}{Si(X \Delta Z)Z}$$

$$A3' \quad \frac{Si(X \wedge Y)\perp}{Si(X \Delta (X \vee Y))(X \vee Y)}$$

$$A4' \quad \frac{Si((X \vee Y) \wedge Z)\perp \quad Si(X \Delta Y)Y}{Si((X \vee Z) \Delta (Y \vee Z))(Y \vee Z)}$$

$$A5' \quad \frac{Si((X \vee Y) \wedge Z)\perp \quad Si((X \vee Z) \Delta (Y \vee Z))(Y \vee Z)}{Si(X \Delta Y)Y}$$

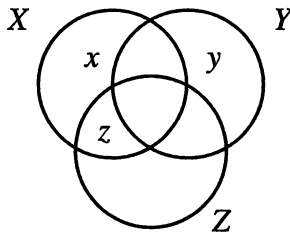
A1' et A3' sont des principes de M , comme on peut facilement le vérifier, mais non A2', A4' et A5', puisque ceux-ci sont à plus d'une prémisse. Or nous avons de bonnes raisons de croire que ces derniers principes sont vraiment des lois fondamentales de QD , étant donné la nature de la relation sur laquelle cette logique est basée. Jusqu'à maintenant cette conjecture tient le coup.

3.6. Conjecture

Tous les principes élémentaires de M plus les principes additionnels A2', A4' et A5' constituent une axiomatisation complète pour *au moins la moitié*. En outre, des précédents faits et du théorème 3.5, nous inférons le théorème suivant.

3.7. Théorème. $M \subset QD \subset C$

Pour terminer cette section, nous allons dresser un portrait complet des relations qui existent entre les différentes logiques vues jusqu'ici. Tout d'abord, $QD \not\subseteq E$, car A2' n'est pas valide selon *tout ou au moins un*, comme l'indique ce contre-exemple :

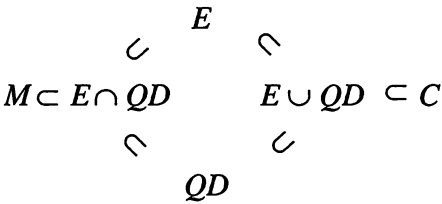


D'autre part, puisque DSJ est un principe de E mais non de QD , $E \not\subseteq QD$. En outre, $E \cap QD \neq M$, car un examen attentif indique qu'il y a des principes qui

n'appartiennent pas à M mais qui sont valides à la fois selon *au moins la moitié* et *tout ou au moins un*. En voici une courte liste :

DSJC	$\frac{Si (X \wedge Y) \perp \quad Si XZ \quad Si YZ}{Si (X \vee Y)Z}$	Disjonction conditionnelle
RCC	$\frac{Si (Y \wedge \neg Z) \perp \quad Si XY}{Si X(Y \wedge Z)}$	Renforcement conditionnel du conséquent
RCA	$\frac{Si (X \wedge \neg Z) \perp \quad Si XY}{Si (X \wedge Z)Y}$	Renforcement conditionnel de l'antécédent

Maintenant, puisque $QD \not\subseteq E$ et $E \not\subseteq QD$, $QD \neq E \cup QD \neq E$. En outre, $E \cup QD \neq C$, puisque CNT, TRN et RA, par exemple, n'appartiennent pas à E ni à QD . Cela, en conjonction avec les théorèmes 3.4 et 3.7, implique que $E \cup QD \subset C$. Et c'est ainsi qu'en tout, nous obtenons le diagramme suivant.



4. Les logiques CB , EB et QDB

On constate d'emblée que les relations entre les logiques conditionnelles déterminées par nos quantificateurs se modifient dans le cadre FCB. Considérons par exemple les trois principes suivants :

TEC	$Si XY \vee Si X \neg Y$
CV	$\frac{Si XY \quad \neg Si X \neg Z}{Si (X \wedge Z)Y}$
ALT	$\frac{Si X(Y \vee Z)}{Si XY \vee Si XZ}$

Il est facile de vérifier que ces trois principes sont valides selon *tout ou au moins un*, que seulement TEC et CV le sont selon *au moins la moitié* et que seulement CV, qui découle trivialement de RA, est valide selon *tout*. Une étude systématique des logiques CB , EB et QDB devrait donc nous fournir des indications plus précises concernant ces modifications.

Commençons d'abord par *CB*. De prime abord, tous les principes élémentaires de *M* plus TRN suffisent pour axiomatiser complètement cette logique. En effet, selon la preuve du théorème 3.1, les propriétés nécessaires à la démonstration de l'existence d'un modèle réfuteur selon l'inclusion s'obtiennent au moyen de principes R, RTA, DSJ, C4 et CC4, dérivables dans *M*+TRN. Pourtant, *CB* ne peut pas coïncider avec *C*. Le principe suivant par exemple :

$$\text{NT} \quad \frac{SiX \perp}{\neg SiTX} \quad \text{Non trivialité}$$

est manifestement valide selon *tout*, mais il n'est aucunement dérivable dans *C* : un principe introduisant la négation est en effet nécessaire.

Cette curiosité s'explique cependant très bien après réflexion. Dans la preuve du théorème 3.1, la démonstration de l'existence d'un modèle réfuteur selon l'inclusion ne repose pas seulement sur des principes dérivables dans *M*+TRN. Elle repose aussi sur le fait que le formalisme FCE ne contient aucune formule *réfutable* selon la relation *universelle*. Or, contrairement à FCE, le formalisme FCB contient des formules réfutables selon *toute* relation et donc selon la relation universelle. La relation universelle réfute bien entendu toutes les contradictions formelles (ex. : $SiXY \wedge \neg SiXY$), mais elle réfute aussi toute formule de la forme $\neg SiXY$ et *a fortiori* toute inférence de la sorte $SiX_1Y_1 \dots SiX_nY_n / \neg SiXY$, dont en particulier NT. L'existence d'un principe interdisant la relation universelle est donc indispensable dans l'axiomatisation de *CB*.

4.1. Théorème

Tous les principes élémentaires de *M* plus les principes additionnels TRN et NT constituent, dans le cadre du formalisme *FCB*, une axiomatisation complète pour l'inclusion (*tout*).

Preuve. Supposons que $\varphi_1, \dots, \varphi_n \not\vdash \psi$ dans *M*+TRN+NT. Encore une fois, il doit exister une algèbre Booléenne finie $(U, \wedge, \vee, 0, 1, ')$, munie d'une relation binaire additionnelle *Q* ayant ni plus ni moins toutes les propriétés formelles de *Si* dans *M*+TRN+NT et selon laquelle, par l'entremise d'une assignation *V* allant des termes Booléens dans *U*, chaque prémisse est vérifiée et la conclusion est réfutée. Or cette relation n'est pas universelle, car en vérifiant NT et R, elle vérifie $\neg SiTX \perp$. Nous pouvons donc poursuivre comme dans la preuve du théorème 3.1 pour finalement conclure qu'il existe un modèle, à savoir $\langle F(At(U)), [[]] \rangle$ tel que $[[Z]] = F(\theta(V(Z)))$ pour tout terme atomique *Z*, qui réfute l'inférence selon l'inclusion. ■

Considérons maintenant la logique *EB*, celle de *tout ou au moins un*. D'abord, on peut facilement vérifier que non seulement *NT*, mais également :

$$\text{CVNT} \quad \frac{\neg Si \, TX}{Si X \perp} \quad \text{Converse de NT}$$

est valide ici. Or ce dernier principe n'est pas valide selon *tout* et n'appartient donc pas à *CB*. Un autre principe valide selon *tout ou au moins un* mais non selon *tout* est le suivant :

$$\text{AEB} \quad \frac{SiXY}{Si (X \wedge Y) \perp \vee Si \, \bar{T}(X \wedge Y)}$$

qui exprime simplement la condition de l'inclusion ou de l'intersection non vide. Maintenant, les principes *NT*, *CNT* et *AEB*, combinés avec ceux de *M* plus *DSJ*, se révèlent suffisants pour axiomatiser complètement *EB*.

4.2. Théorème

Tous les principes élémentaires de *M* plus les principes additionnels *NT*, *CVNT*, *AEB* et *DSJ* constituent une axiomatisation complète pour l'inclusion ou l'intersection non vide (*tout ou au moins un*).

Preuve. Nous procédons toujours selon la stratégie appliquée dans la preuve du Théorème 3.1. Supposons que $\varphi_1, \dots, \varphi_n \not\models \psi$ dans $M + \text{AEB} + \text{NT} + \text{CVNT} + \text{DSJ}$. Encore une fois, il doit exister une algèbre Booléenne finie $(U, \wedge, \vee, 0, 1, ')$, munie d'une relation binaire additionnelle Q ayant ni plus ni moins toutes les propriétés formelles de ce système et selon laquelle, par l'entremise d'une assignation V allant des termes Booléens dans U , chaque prémisses est vérifiée et la conclusion est réfutée. Comme d'habitude, par l'entremise de la représentation par atomes θ , Q devient la relation $Q = \{ \langle \theta(x), \theta(y) \rangle \mid Qxy \}$. On obtient ainsi le résultat souhaité en montrant que Q ou une restriction appropriée de Q devient l'inclusion ou l'intersection non vide entre ensembles.

D'abord, on note que Q a les propriétés suivantes (avec $E = \text{At}(U)$).

P1. QAA	(R)
P2. $QAC, QBC \Rightarrow Q(A \cup B)C$	(DSJ)
P3. $QA\emptyset \Rightarrow Q(A \cap B)\emptyset$	(RTA)
P4. $Q(A \cap B')\emptyset \Rightarrow QAB$	(C4)
P5. $QE(A \cap B) \Rightarrow QAB$	(IMP)
P6. $QA\emptyset \Leftrightarrow \neg QEA$	(NT + CVNT)
P7. $QAB \Rightarrow Q(A \cap B')\emptyset$ ou $QE(A \cap B)$	(AEB)
P8. $QAB \Leftrightarrow Q(A \cap B')\emptyset$ ou $QE(A \cap B)$	(de P4, P5 et P7)
P9. $\neg QE\emptyset$	(de P1 - $Q\emptyset\emptyset$ - et P6)

Maintenant, Q peut être entièrement définie par deux prédicats unaires sur E , disons Q_a et Q_b , tels que $Q_a A \Leftrightarrow Q A \emptyset$ et $Q_b A \Leftrightarrow Q E A$. Par P8 nous avons en effet que $Q A B \Leftrightarrow Q_a A \cap B'$ ou $Q_b A \cap B$, d'où les propriétés suivantes :

$$P1a. Q_a \emptyset \quad (\text{de P1})$$

$$P2a. Q_a A, Q_a B \Rightarrow Q_a A \cup B \quad (\text{de P2})$$

$$P3a. Q_a A \Rightarrow Q_a A \cap B \quad (\text{de P3})$$

$$P6ab. Q_a A \Leftrightarrow \neg Q_b A \quad (\text{de P6})$$

$$P9a. \neg Q_a E \quad (\text{de P9})$$

Soit $K = \bigcup \{A \mid Q_a A\}$. Donc $Q_a K$, par P2a, le fait que E est fini et la définition de K , si bien que pour tous les $A, B \subseteq E - K$, $Q_a A \cap B' \Leftrightarrow A \subseteq B$. En outre, pour tous les $A, B \subseteq E - K$, $Q_b A \cap B \Leftrightarrow A \cap B \neq \emptyset$. En effet, si $A \cap B = \emptyset$, alors $Q_a A \cap B$ (P1a) et donc $\neg Q_b A \cap B$ (P6ab) ; conversément, si $\neg Q_b A \cap B$, alors $Q_a A \cap B$ (P6ab), donc $A \subseteq B'$, i.e., $A \cap B = \emptyset$.

Soit encore une fois l'homomorphisme Booléen $F = \lambda D.D - K$. Nous avons que $F(E) \neq \emptyset$, par P9a. Il suffit donc de montrer que pour tous les $A, B \subseteq E$:

$$Q_a A \cap B' \Leftrightarrow Q_a F(A \cap B')$$

$$Q_b A \cap B \Leftrightarrow Q_b F(A \cap B)$$

car cela implique que $Q A B$ si et seulement si $F(A) \subseteq F(B)$ ou $F(A) \cap F(B) \neq \emptyset$, ce qui signifie qu'il existe un modèle, à savoir $\langle F(E), [[]] \rangle$ tel que $[[Z]] = F(\theta(V(Z)))$ pour tout terme atomique Z , qui réfute l'inférence selon *tout ou au moins un*.

D'abord, la démonstration que $Q_a A \cap B' \Leftrightarrow Q_a F(A \cap B')$ procède comme dans la preuve du théorème 3.1 à l'aide des propriétés P2a et P3a. De ce résultat, on établit facilement que $Q_b A \cap B \Leftrightarrow Q_b F(A \cap B)$. En effet, supposons que $Q_b A \cap B$. Donc $\neg Q_a A \cap B$ (P6ab), donc $\neg Q_a F(A \cap B)$ et donc $Q_b F(A \cap B)$ (P6ab). Conversément, supposons que $Q_b F(A \cap B)$. Donc $\neg Q_a F(A \cap B)$ (P6ab), donc $\neg Q_a A \cap B$ et donc $Q_b A \cap B$ (P6ab). ■

Incidentement, DSJ est dérivable dans $M + \text{CNJP}$, si bien que ce principe peut être remplacé par CNJP dans l'axiomatisation de EB .

Enfin, concernant QDB , une axiomatisation complète reste à trouver. La nature non Booléenne de la relation présente quelques difficultés. Toutefois, on constate que NT est valide selon *au moins la moitié*, mais que CVNT ne l'est pas. Or, TEC et CV sont toujours valides ici et il est difficile de repérer des principes plus fondamentaux que ces derniers. Notre conjecture (prudente) est donc que les principes élémentaires de M plus $A2', A4', A5', TEC$ et CV forment une base axiomatique complète pour *au moins la moitié*.

5. Conclusion

Les résultats que nous venons de présenter sont modestes et encore incomplets. Beaucoup de recherches restent à faire dans la direction que nous avons choisie. En particulier, nous devons mettre au point une technique de preuve de complétude pour la logique *QDB*. Une représentation numérique pourrait être une bonne stratégie ici. Nous pouvons également étudier les logiques que nos quantificateurs déterminent dans le cadre d'un formalisme plus puissant, par exemple une extension de FCB permettant d'exprimer en plus la quantification sur des variables de termes Booléens. À l'opposé, il serait intéressant de restreindre notre formalisme à celui de la théorie syllogistique, qui ne comprend que des formules conditionnelles de la forme *SiXY*, avec *X*, *Y* atomiques.

L'étude des conditionnels dans la perspective de la théorie des quantificateurs généralisés offre aussi de nombreuses autres avenues de recherche. Soulignons à titre d'exemples quelques possibilités.

- Admettre des univers infinis, lesquels offrent la possibilité de définir davantage de quantificateurs conditionnels ¹⁰.
- Cerner plus précisément la classe des quantificateurs conditionnels logiques non uniformes et étudier les logiques conditionnelles que ceux-ci déterminent.
- Étendre notre étude aux quantificateurs conditionnels *non logiques*.
- Dans une perspective plus fondamentale, réviser ou compléter la liste des contraintes intuitives délimitant la classe des quantificateurs conditionnels admissibles.

Comme on le voit, le programme est vaste et mérite d'être poursuivi.

Bibliographie

- BARWISE J. and R. Cooper, "Generalized Quantifiers and Natural Language", *Linguistics and Philosophy* **4**, p. 159-219, 1981.
- van BENTHEM Johan, « Foundations of Conditional Logic », *Journal of Philosophical Logic* **13**, p. 303-349, 1984.
- Essays on Logical Semantics*, D.Reidel Publishing Company, 1986.
- BURGESS John P., « Quick Completeness Proofs for some Logics of Conditionals », *Notre Dame Journal of Formal Logic* **22**, p. 71-84, 1981.
- LAPIERRE Serge, « Conditionals and Quantifiers », in Jaap van der Does et Jan van Eijck (eds.), *Generalized Quantifiers and Applications*, Dutch Network for Language, Logic and Information, Amsterdam, p. 155-174, 1991.

10. À ce sujet, voir S. Lapierre [1991b].

- « New results in the analysis of some conditional quantifiers and their logics », *Logique & Analyse*, n°133-134, p. 105-120, 1991.
- LINDSTRÖM P., "First-order Predicate Logic with Generalized Quantifiers", *Theoria* 32, p. 1-11, 1966.
- MONTAGUE R., "The Proper Treatment of Quantification in Ordinary English", in J. Hintikka, J. Moravcsik and P. Suppes (eds.), *Approaches to Natural Language : Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, D. Reidel Publishing Company, Dordrecht, p. 221-242, 1973.
- WESTERSTÅHL Dag, « Quantifiers in Formal and Natural Languages », in D. Gabbay & F. Guenther (eds.), *Handbook of Philosophical Logic*, volume IV, D. Reidel Publishing Company, p. 1-131, 1989.

Serge LAPIERRE

Chapitre 14

Logique constructive et donkey-anaphore

1. Introduction

1.1. *L'exigence d'une théorie de la compréhension*

A partir du début des années 70 ont été construits de nombreux systèmes d'élaboration automatique des langues naturelles : systèmes qui développent l'analyse syntaxique de phrases d'une langue naturelle, qui répondent à des questions sur le contenu de textes en langue naturelle, qui permettent d'interroger en langue naturelle une base de données, qui résument des textes, etc. Parmi les nombreux problèmes qu'on doit résoudre pour construire de tels systèmes, il y a celui de les doter d'une *compétence sémantique* comparable à celle d'un être humain. Plus précisément, pour traiter intelligemment un texte en langue naturelle il ne suffit pas qu'un système dispose d'informations syntaxiques ; il doit disposer aussi de connaissances et de capacités usuellement classées comme sémantiques : la capacité de reconstruire des chaînes anaphoriques, la connaissance de la logique des connecteurs et des quantificateurs, la connaissance des relations de sens entre éléments lexicaux ; le système doit être capable de déduire "Il y a des fleurs" de "Certaines fleurs sont fanées", et "Certains animaux sont des mammifères" de "Les chats sont des mammifères". L'exigence de construire des compétences sémantiques 'en chair et en os', c'est-à-dire de simuler la compétence effective d'un locuteur, a amené de façon naturelle à s'interroger sur le type d'informations et de capacités mises en jeu dans la compréhension d'un texte (par un locuteur humain) ; autrement dit, elle a transformé le problème sémantique (qu'est-ce que le signifié ?) en problème de compréhension (quelles informations et capacités sont pertinentes dans la compréhension d'un texte ?)" ([Ma92], p.435-6).

Mais dans ce cadre se pose aussi un problème plus fondamental, un problème de nature philosophique. Comprendre une expression veut dire connaître son signifié, donc le but d'une théorie de la compréhension sera de rendre compte de la *connaissance* du signifié ; par conséquent le signifié caractérisé par une telle théorie devra être essentiellement *connaissable*, c'est-à-dire accessible aux capacités cognitives d'un sujet épistémiquement limité. M. Dummett a soutenu que la compréhension du signifié d'une expression doit être identifiée avec la capacité de l'employer correctement et de répondre correctement à son usage par autrui. Sur cette base il a critiqué, d'un côté, les théories du signifié qui utilisent, comme notion centrale, la notion réaliste ou transcendante de vérité, selon laquelle chaque phrase vraie ou fausse est vraie ou fausse indépendamment de notre capacité à la reconnaître comme telle ; et il a développé, de l'autre côté, une théorie du signifié qui identifie le signifié d'une constante logique avec les conditions d'assertibilité (ou de reconnaissance de la vérité) des phrases qui la contiennent comme opérateur principal. Par cette route il est arrivé à proposer comme paradigme de validité logique, alternatif au paradigme classique, la logique intuitionniste, conçue - selon l'indication initiale de Heyting - comme *logique du savoir*.

1.2. Le problème des donkey sentences

Il y a donc des raisons *générales* - on pourrait même dire *a priori* - pour lesquelles un intérêt réciproque entre constructivisme logique et intelligence artificielle est souhaitable. Mais il y a aussi des raisons '*a posteriori*', liées à la structure de problèmes spécifiques. Dans les pages suivantes je me propose d'en illustrer justement une ; plus spécifiquement, je me propose de montrer comment l'adoption d'une logique constructive (notamment d'une sémantique constructive pour un langage logique) permet de résoudre un problème qui présente (ou devrait présenter) un intérêt remarquable pour l'intelligence artificielle.

Les *donkey sentences* sont des propositions qui contiennent un syntagme nominal indéfini dans une subordonnée relative ou de la forme "Si...", et un pronom qui n'appartient pas à la subordonnée mais qui est lié anaphoriquement au pronom. Des exemples typiques sont les suivants :

- (1) If a farmer owns a donkey, he beats it.
- (2) Every farmer who owns a donkey, beats it.

Le problème qu'elles posent peut être décrit dans les termes suivants. Considérons le processus de formalisation, c'est-à-dire de traduction dans une langue formelle, d'une phrase d'une langue naturelle, et supposons que nous cherchions, pour quelque raison, à rendre ce processus *algorithmiquement déterminé*, et donc, en principe, automatisable. La façon la plus naturelle de poursuivre ce but est de faire que la traduction respecte rigoureusement le *principe de compositionnalité*, selon lequel la traduction d'une expression complexe est fonction des traductions des expressions qui la composent. C'est, par exemple, la stratégie suivie par Montague dans [Mo73], et, à sa suite, par toute la *Montague Grammar* ; mais c'est, plus généralement, la stratégie la plus naturelle pour tout projet (par exemple, de traduction automatique) qui prévoit une phase de 'représentation du signifié' des

phrases d'une langue naturelle, c'est-à-dire de traduction dans un langage formel. Les *donkey sentences* constituent un défi justement pour cette stratégie de traduction 'compositionnelle'. Considérons par exemple la phrase anglaise (1) et le langage prédicatif du premier ordre L . La traduction standard de (1) dans L est

$$(3) \forall v_1 \forall v_2 [(F(v_1) \wedge D(v_2) \wedge O(v_1, v_2)) \rightarrow B(v_1, v_2)] ;$$

mais cette traduction ne respecte pas le principe de compositionnalité. En fait, (1) a la forme "If A, B", donc sa traduction devrait être fonction des traductions de A, B et "if" ; mais la traduction plus naturelle de l'antécédent est

$$(4) \exists v_1 \exists v_2 [F(v_1) \wedge D(v_2) \wedge O(v_1, v_2)],$$

et une traduction très naturelle du conséquent est la formule *ouverte*

$$(5) B(v_1, v_2),$$

ou, de façon équivalente, une des formules $B(v_2, v_3)$, $B(v_3, v_4)$, ... Si on met ensemble (4) et (5) au moyen de la traduction plus naturelle de "if...then", c'est-à-dire $\rightarrow$, on obtient

$$(6) \exists v_1 \exists v_2 [F(v_1) \wedge D(v_2) \wedge O(v_1, v_2)] \rightarrow B(v_1, v_2),$$

qui, d'un côté, est la meilleure approximation concevable d'une traduction 'compositionnelle', mais, de l'autre côté, ne représente pas du tout le signifié de (1), parce que les variables v_1 et v_2 du conséquent ne sont pas liées par les quantificateurs de l'antécédent. De toute façon, (6) n'est pas équivalent à (3) qui, même si elle n'est pas 'compositionnelle', représente exactement le signifié de (1).

Un problème analogue est posé par le phénomène de l'anaphore interpropositionnelle. Considérons par exemple le texte suivant :

$$(7) A \text{ man walks. He talks.}$$

Le pronom "he" de la deuxième proposition peut être interprété soit comme une expression démonstrative dont la référence est déterminée par le contexte pragmatique dans lequel (7) est utilisé, soit comme une expression anaphoriquement liée au syntagme indéfini "A man" de la première proposition. Dans cette deuxième interprétation nous avons un cas d'anaphore interpropositionnelle. L'ambiguïté en question peut être éliminée, et *est* éliminée dans la littérature linguistique, en marquant les expressions anaphoriquement liées par la même *marque référentielle*. Dans le cas de (7), les deux interprétations possibles seront représentées syntaxiquement de la façon suivante :

$$(8) \quad (a) A_1 \text{ man walks. He}_2 \text{ talks.}$$

$$(b) A_1 \text{ man walks. He}_1 \text{ talks.}$$

Dans le même style, les interprétations préférées de (1) et (2) seront représentées syntaxiquement de la façon suivante :

(9) If a_1 farmer owns a_2 donkey, he₁ beats it₂.

(10) Every₁ farmer who owns a_2 donkey, beats it₂.

Or, les traductions plus naturelles des deux propositions dans (8)(b) sont, respectivement,

(11) $\exists v_1 [M(v_1) \wedge W(v_1)]$

(12) $T(v_1)$,

et la traduction plus naturelle de la mise en séquence des deux propositions pour former le texte est la conjonction $\wedge$. Mais si on met ensemble (11) et (12) au moyen de $\wedge$, ce qu'on obtient,

(13) $\exists v_1 [M(v_1) \wedge W(v_1)] \wedge T(v_1)$,

n'est pas du tout la représentation correcte du signifié de (8)(b), pour la même raison que dans le cas des *donkey sentences* ; la traduction correcte serait plutôt

(14) $\exists v_1 [(M(v_1) \wedge W(v_1)) \wedge T(v_1)]$,

mais celle-ci n'est pas compositionnelle.

Ces exemples montrent que si on choisit **L** comme langage logique, la traduction des *donkey sentences* (et d'autres phrases affines) ne pourra pas être compositionnelle, et par conséquent le processus de traduction ne pourra pas être algorithmique, mais devra être essentiellement 'intelligent'.

1.3. Une stratégie de solution

Une stratégie de solution très intéressante au problème décrit a été proposée récemment par Groenendijk et Stokhof dans les articles [GS90] et [GS91]. Je parle de *stratégie* parce qu'elle peut être commune à des *tactiques* de solution très différentes. La stratégie, appliquée au langage de la logique prédicative du premier ordre, peut être décrite très synthétiquement dans les termes suivants.

Comme (6) serait la traduction 'compositionnelle' recherchée de (1) *si* les deux quantificateurs existentiels de l'antécédent pouvaient lier les deux variables du conséquent, modifions la syntaxe du langage logique afin d'obtenir exactement cet effet. Et, comme (3) est la représentation correcte du signifié de (1), modifions la sémantique du langage logique de façon que (6) soit équivalent à (3).

D'une façon analogue, comme (13) serait la traduction 'compositionnelle' recherchée de (8)(b) *si* le quantificateur existentiel pouvait lier la deuxième occurrence de la variable, modifions la syntaxe du langage logique afin d'obtenir exactement cet effet. Et, comme (14) est la représentation correcte du signifié de (8)(b), modifions la sémantique du langage logique de façon que (13) soit équivalent à (14).

Le but de cette stratégie de solution est donc d'obtenir que les équivalences

$$(15) [\exists v \varphi \rightarrow \psi] = \forall v [\varphi \rightarrow \psi]$$

et

$$(16) [\exists v \varphi \wedge \psi] = \exists v [\varphi \wedge \psi]$$

soient sémantiquement valides.

La stratégie est essentiellement la même dans le cas du langage typé, avec λ -abstraction, employé dans [GS90]. La différence est qu'au niveau syntaxique on pourra distinguer entre les cas où l'interprétation pertinente des constantes est l'interprétation modifiée et les cas où est pertinente l'interprétation usuelle

Dans les pages suivantes je vais présenter une approche au problème des donkey sentences qui est une application de la même stratégie inventée par Groenendijk et Stokhof (d'ici en avant G&S), mais qui est très différente du point de vue tactique : G&S utilisent, afin de modifier la sémantique classique, l'interprétation *dynamique* des constantes logiques, tandis que je vais utiliser (une modification opportune de) l'interprétation *constructive* de Heyting.

2. La logique constructive LC

Comme l'approche de G&S, la solution que je vais présenter peut être appliquée soit à un langage du premier ordre soit à un langage d'ordre supérieur, pareil au langage employé par Montague dans [Mo73]. Ici je vais illustrer la deuxième alternative.

2.1. Syntaxe

Les idées qui déterminent la structure du langage sont essentiellement deux. La première est que les formules (expressions de type t) sont obtenues par une opération d'"extensionnalisation" $\sim$ à partir de *termes* qui dénotent des propositions conçues, constructivement, comme ensembles de justifications.¹ La deuxième est liée à la nécessité de distinguer entre les cas où quantificateurs et variables sont employés comme d'habitude, et les cas où il y a donkey-anaphore (c'est-à-dire, ils sont employés comme dans les donkey sentences et les autres cas illustrés ci-dessus). J'ai choisi d'introduire, à côté des constantes logiques usuelles ($\wedge, \vee, \rightarrow, \forall, \exists, =$ plus la

1. La notion de justification joue ici un rôle analogue à celui que la notion de preuve joue dans l'explication de Heyting. La différence essentielle entre les deux est que s'il y a une preuve de A alors A est vraie, tandis que du fait qu'il existe une justification de A il ne suit pas que A soit vraie. Cette différence justifie les différences entre la sémantique présentée ici et d'autres 'interprétations fonctionnelles' basées sur la notion de preuve.

constante $\perp$ pour le Faux et le λ -opérateur), censées recevoir l'interprétation (constructive) standard, deux nouvelles constantes $\&$ et $\Rightarrow$, censées recevoir les interprétations, respectivement, de la conjonction et de l'implication, modifiées afin de représenter la donkey-anaphore. On verra plus avant que dans les cas où il n'y a pas de donkey-anaphore les deux constantes nouvelles fonctionnent exactement comme $\wedge$ et $\rightarrow$, respectivement (voir faits 1-3).

Voici les définitions formelles.

Types

L'ensemble des *types* est le plus petit ensemble T tel que

1. $e, t, j \in T$
2. Si $a, b \in T$, alors $(a, b) \in T$

Expressions significantes

Etant donné deux ensembles CON_a et VAR_a de constantes et de variables de type a , pour chaque type a , l'ensemble ES_a des *expressions significantes de type a* est défini par induction :

1. Si $\varphi \in CON_a \cup VAR_a$, alors $\varphi \in ME_a$
2. $\perp \in CON_j$
3. Si $\varphi \in ME_{(a,b)}$ et $\psi \in ME_a$, alors $\varphi(\psi) \in ME_b$
4. Si $\varphi, \psi \in ME_j$, alors $[\varphi \wedge \psi], [\varphi \vee \psi], [\varphi \rightarrow \psi], [\varphi \& \psi], [\varphi \Rightarrow \psi] \in ME_j$
5. Si $\varphi \in ME_j$ et $v \in VAR_a$, alors $\forall v \varphi, \exists v \varphi \in ME_j$
6. Si $\varphi, \psi \in ME_a$, alors $[\varphi = \psi] \in ME_j$
7. Si $\varphi \in ME_a$ et $v \in VAR_b$, alors $\lambda v \varphi \in ME_{(b,a)}$
8. Si $\varphi \in ME_j$, alors $\sim \varphi \in ME_t$

On emploiera $\neg \varphi$ comme abréviation de $\varphi \rightarrow \perp$.

Scope et binding

Je vais adapter au langage de CL les définitions de $aq(\varphi)$ (l'ensemble des *active quantifier occurrences* de φ), $fv(\varphi)$ (l'ensemble des *free occurrences of variables* de φ), $bp(\varphi)$ (l'ensemble des *binding pairs* de φ) et $sp(\varphi)$ (l'ensemble des *scope pairs* de φ) données dans [GS91].

Si $\varphi \in CON_a \cup VAR_a$, alors $aq(\varphi) = \emptyset$

$aq(\perp) = aq(\forall v \varphi) = \emptyset$

$aq(\varphi(\psi)) = aq(\varphi \wedge \psi) = aq(\varphi \vee \psi) = aq(\varphi \rightarrow \psi) = aq(\varphi \Rightarrow \psi) = aq(\varphi) \cup aq(\psi)$

$aq(\sim \varphi) = aq(\varphi)$

$aq(\varphi \& \psi) = aq(\psi) \cup \{\exists v \in aq(\varphi) \mid \exists v \notin aq(\psi)\}$

$\mathbf{aq}(\exists v\varphi) = \mathbf{aq}(\varphi) \cup \{\exists v\}$ si $\exists v \notin \mathbf{aq}(\varphi)$

$= \mathbf{aq}(\varphi)$ autrement.

Si $\varphi \in \text{CON}_a$, alors $\mathbf{fv}(\varphi) = \emptyset$

Si $\varphi \in \text{VAR}_a$, alors $\mathbf{fv}(\varphi) = \{\varphi\}$

$\mathbf{fv}(\perp) = \emptyset$

$\mathbf{fv}(\varphi(\psi)) = \mathbf{fv}(\varphi \wedge \psi) = \mathbf{fv}(\varphi \vee \psi) = \mathbf{fv}(\varphi \rightarrow \psi) = \mathbf{fv}(\varphi) \cup \mathbf{fv}(\psi)$

$\mathbf{fv}(\varphi \& \psi) = \mathbf{fv}(\varphi \Rightarrow \psi) = \mathbf{fv}(\varphi) \cup \{v \in \mathbf{fv}(\psi) \mid \exists v \notin \mathbf{aq}(\varphi)\}$

$\mathbf{fv}(\exists v\varphi) = \mathbf{fv}(\forall v\varphi) = \mathbf{fv}(\varphi)$ moins les occurrences de v

$\mathbf{fv}(\sim \varphi) = \mathbf{fv}(\varphi)$.

Si $\varphi \in \text{CON}_a \cup \text{VAR}_a$, alors $\mathbf{bp}(\varphi) = \emptyset$

$\mathbf{bp}(\perp) = \emptyset$

$\mathbf{bp}(\varphi(\psi)) = \mathbf{bp}(\varphi \wedge \psi) = \mathbf{bp}(\varphi \vee \psi) = \mathbf{bp}(\varphi \rightarrow \psi) = \mathbf{bp}(\varphi) \cup \mathbf{bp}(\psi)$

$\mathbf{bp}(\varphi \Rightarrow \psi) = \mathbf{bp}(\varphi \& \psi) = \mathbf{bp}(\varphi) \cup \mathbf{bp}(\psi) \cup \{\langle \exists v, v \rangle \mid \exists v \in \mathbf{aq}(\varphi) \text{ et } v \in \mathbf{fv}(\psi)\}$

$\mathbf{bp}(\exists v\varphi) = \mathbf{bp}(\varphi) \cup \{\langle \exists v, v \rangle \mid v \in \mathbf{fv}(\varphi)\}$

$\mathbf{bp}(\forall v\varphi) = \mathbf{bp}(\varphi) \cup \{\langle \forall v, v \rangle \mid v \in \mathbf{fv}(\varphi)\}$.

$\mathbf{bp}(\sim \varphi) = \mathbf{bp}(\varphi)$

Si $\varphi \in \text{CON}_a \cup \text{VAR}_a$, alors $\mathbf{sp}(\varphi) = \emptyset$

$\mathbf{sp}(\perp) = \emptyset$

$\mathbf{sp}(\varphi(\psi)) = \mathbf{sp}(\varphi \wedge \psi) = \mathbf{sp}(\varphi \vee \psi) = \mathbf{sp}(\varphi \rightarrow \psi) = \mathbf{sp}(\varphi \Rightarrow \psi) = \mathbf{sp}(\varphi \& \psi) = \mathbf{sp}(\varphi) \cup \mathbf{sp}(\psi)$

$\mathbf{sp}(\exists v\varphi) = \mathbf{sp}(\varphi) \cup \{\langle \exists v, v \rangle \mid v \in \mathbf{fv}(\varphi)\}$

$\mathbf{sp}(\forall v\varphi) = \mathbf{sp}(\varphi) \cup \{\langle \forall v, v \rangle \mid v \in \mathbf{fv}(\varphi)\}$.

2.2. Sémantique

Je vais maintenant associer au langage logique une sémantique constructive basée sur une interprétation fonctionnelle des constantes logiques. La raison essentielle de ce choix est la suivante. On a vu que dans les donkey sentences un syntagme nominal indéfini (dont la traduction emploiera évidemment un quantificateur existentiel) lie anaphoriquement un pronom, c'est-à-dire le force à dénoter le même individu dénoté par le syntagme nominal. L'avantage décisif d'une interprétation fonctionnelle est que les objets qu'elle associe à une formule existentielle $\exists v\varphi$ (qu'ils soient des preuves, des justifications, des fonctions réalisantes, ou autre chose) exhibent explicitement un *témoin* de φ , c'est-à-dire un individu qui a la propriété φ . Considérons par exemple la formule $[\exists v\varphi \rightarrow \psi]$. Une preuve (justification) de l'antécédent est un couple $j = \langle d, j' \rangle$, où d est un individu et j' une preuve (justification) de $\varphi[d/v]$; on pourra donc se référer au témoin en l'appelant simplement la *projection gauche* de j . A ce point, tout ce qui nous reste à faire pour rendre compte de la donkey-anaphore est de modifier l'interprétation fonctionnelle de

l'implication afin de forcer la variable du conséquent à dénoter le témoin ; c'est précisément l'interprétation définie ci-dessous pour le connecteur $\Rightarrow$.²

Passons aux définitions formelles.

Domaines

Etant donné un ensemble non vide D d'objets (individus) et une collection $\mathbf{J}$ d'ensembles finis d'objets (propositions, vues comme ensembles de justifications), on définit de la façon suivante le domaine D_a des entités de type a , pour chaque type a :

1. $D_e = D$
 $D_t = \{0,1\}$
 $D_j = \mathbf{J}$
2. $D_{(a,b)} = D_b^{D_a}$

Interprétation

Une interprétation est une triple $\mathbf{I} = \langle D, \mathbf{J}, F \rangle$, où D et $\mathbf{J}$ sont comme ci-dessus et F est une fonction telle que, si $\alpha \in \text{CON}_a$, $F(\alpha) \in D_a$.

Assignment

Une assignation est une fonction g telle que, si $v \in \text{VAR}_a$, $g(v) \in D_a$.

Dénotations

Etant donné une expression α de type $a \neq t$, une interprétation $\mathbf{I}$ et une assignation g , on définit la dénotation $[\alpha]_{\mathbf{I},g}$ de α dans $\mathbf{I}$ et g par l'induction suivante :

1. Si $\alpha \in \text{CON}_a$, $[\alpha]_{\mathbf{I},g} = F(\alpha)$
 Si $\alpha \in \text{VAR}_a$, $[\alpha]_{\mathbf{I},g} = g(\alpha)$
2. Si α est $[\varphi(\psi)]$, $[\alpha]_{\mathbf{I},g} = [\varphi]_{\mathbf{I},g}([\psi]_{\mathbf{I},g})$
3. Si α est $[\varphi \wedge \psi]$, $[\alpha]_{\mathbf{I},g} = [\varphi]_{\mathbf{I},g} \times [\psi]_{\mathbf{I},g}$
 Si α est $[\varphi \vee \psi]$, $[\alpha]_{\mathbf{I},g} = [\varphi]_{\mathbf{I},g} \cup [\psi]_{\mathbf{I},g}$
 Si α est $[\varphi \rightarrow \psi]$, $[\alpha]_{\mathbf{I},g} = [\psi]_{\mathbf{I},g}$
 Si α est $[\varphi \& \psi]$, $[\alpha]_{\mathbf{I},g} = \{ \langle j_1, j_2 \rangle \mid j_1 \in [\varphi]_{\mathbf{I},g} \text{ et } j_2 \in [\psi \uparrow (\varphi \& \psi)]_{j_1} \}$

2. La considération de la formule $\exists v \varphi \wedge \psi$ nous amènera d'une façon analogue à l'interprétation du connecteur $\&$.

Si α est $[\varphi \Rightarrow \psi]$,

$$[\alpha]_{\mathbf{I},g} = \{j \mid \text{pour chaque } j' \in [\varphi]_{\mathbf{I},g},$$

$$j(j') \in [\psi[\uparrow(\underline{\varphi} \Rightarrow \underline{\psi})]_{\mathbf{I},g} / \uparrow(\varphi \Rightarrow \psi)]_{\mathbf{I},g}\}$$

4. Si α est $[\forall v \varphi]$, $[\alpha]_{\mathbf{I},g} = \{j \mid \text{pour chaque } f \text{ telle que}$

$$f(v) \in D_a, j(f) \in [\varphi[\underline{f}(\underline{v})/v]]_{\mathbf{I},g}\}$$

Si α est $[\exists v \varphi]$, $[\alpha]_{\mathbf{I},g} = \{ \langle d, j \rangle \mid d = g(x) \text{ et } j \in [\varphi[\underline{d}/v]]_{\mathbf{I},g} \}$

5. Si α est $[\alpha = \beta]$, $[\alpha]_{\mathbf{I},g} = \{ \alpha = \beta \}$

6. Si α est $[\lambda v \varphi]$, $[\alpha]_{\mathbf{I},g}$ = la fonction f de D_b à D_a telle que,

$$\text{pour chaque } d \in D_a, f(d) = [\varphi[\underline{d}/v]]_{\mathbf{I},g}$$

Voici les explications des notations employées.

Si E et F sont deux ensembles,

$$E \times F = \{ \langle e, f \rangle \mid e \in E \text{ et } f \in F \};$$

$$E \cup F = \{ \langle e, 0 \rangle \mid e \in E \} \cup \{ \langle f, 1 \rangle \mid f \in F \}.$$

$p(\langle d, e \rangle) = d$; $q(\langle d, e \rangle) = e$. $p(x)$ est appelée la projection gauche de x , $q(x)$ la projection droite de x .

$\uparrow \varphi = \langle v_1, \dots, v_n \rangle$, où v_i appartient à la séquence ssi $\langle \exists v_i, v_i \rangle \in \mathbf{bp}(\varphi) - \mathbf{sp}(\varphi)$.

J'appellerai $\uparrow \varphi$ la séquence des *occurrences capturées de variables* dans φ .

Si $\uparrow \varphi = \langle v_1, \dots, v_n \rangle$ et $\psi = \langle \mu_1, \dots, \mu_n \rangle$, $\varphi[\psi/\uparrow \varphi]$ est le résultat de la substitution, dans φ , de v_1 par $\mu_1, \dots$, et de v_n par μ_n .

Si $\mathbf{I} = \langle D, J, F \rangle$ et $d \in D$, $\underline{d}$ est une constante c telle que $F(c) = d$.

Si $j \in [\exists v \varphi]_{\mathbf{I},g}$, $lv_j = p(j)(v)$

Si $j \in [\forall v \varphi]_{\mathbf{I},g}$, $lv_j = lv_j(f(v))$, pour chaque f telle que $f(v) \in D_a$ et $j(f) \in [\varphi[\underline{f}(\underline{v})/v]]_{\mathbf{I},g}$.

Modèle

Etant donné une interprétation $\mathbf{I}=\langle D, J, F \rangle$, un modèle est une fonction $\mathbf{M}$ de $\cup J$ à D_I telle que :

1. Si $j \in [\perp]_{\mathbf{I},g}$, $\mathbf{M}(j)=0$
2. Si $j \in [\varphi \wedge \psi]_{\mathbf{I},g}$, $\mathbf{M}(j)=1$ ssi $\mathbf{M}(p(j))=\mathbf{M}(q(j))=1$
3. Si $j \in [\varphi \vee \psi]_{\mathbf{I},g}$, $\mathbf{M}(j)=1$ ssi $\mathbf{M}(p(j))=1$
4. Si $j \in [\varphi \rightarrow \psi]_{\mathbf{I},g}$, $\mathbf{M}(j)=1$ ssi, pour chaque $j' \in [\varphi]_{\mathbf{I},g}$ telle que $\mathbf{M}(j')=1$, $\mathbf{M}(j(j'))=1$
5. Si $j \in [\varphi \& \psi]_{\mathbf{I},g}$, $\mathbf{M}(j)=1$ ssi $\mathbf{M}(p(j))=\mathbf{M}(q(j))=1$
6. Si $j \in [\varphi \Rightarrow \psi]_{\mathbf{I},g}$, $\mathbf{M}(j)=1$ ssi, pour chaque $j' \in [\varphi]_{\mathbf{I},g}$ telle que $\mathbf{M}(j')=1$, $\mathbf{M}(j(j'))=1$
7. Si $j \in [\forall v \varphi]_{\mathbf{I},g}$, $\mathbf{M}(j)=1$ ssi, pour chaque f telle que $f(v) \in D_a$, $\mathbf{M}(j(f))=1$
8. Si $j \in [\exists v \varphi]_{\mathbf{I},g}$, $\mathbf{M}(j)=1$ ssi $\mathbf{M}(q(j))=1$
9. Si $j \in [\alpha = \beta]_{\mathbf{I},g}$, $\mathbf{M}(j)=1$ ssi $[\alpha]_{\mathbf{I},g} = [\beta]_{\mathbf{I},g}$

Vérité et validité constructive

Etant donné une interprétation $\mathbf{I}=\langle D, J, F \rangle$, une assignation g et un modèle $\mathbf{M}$, on dira qu'une formule $\sim \varphi$ est justifiée relativement à $\mathbf{I}$, g et $\mathbf{M}$ ssi il y a une $j \in [\varphi]_{\mathbf{I},g}$ telle que $\mathbf{M}(j)=1$; $\sim \varphi$ est vraie (ou conclusivement justifiée) relativement à $\mathbf{I}$ et g ssi il y a une $j \in [\varphi]_{\mathbf{I},g}$ telle que, pour chaque modèle $\mathbf{M}$, $\mathbf{M}(j)=1$. On dira que $\sim \varphi$ est valide ssi, pour chaque interprétation $\mathbf{I}$, $\sim \varphi$ est vraie relativement à $\mathbf{I}$.

2.3. Quelques remarques**Fait 1**

Si v n'a pas d'occurrences capturées dans ψ , alors

$$a) [\exists v \varphi \& \psi] = [\exists v \varphi \wedge \psi]$$

$$b) [\exists v \varphi \Rightarrow \psi] = [\exists v \varphi \rightarrow \psi]$$

Démonstration

$\psi[\uparrow(\varphi \& \psi)]_{j_1} / \uparrow(\varphi \& \psi) = \psi$ si v n'a pas d'occurrences capturées dans ψ .
 Dans ce cas-là, $[\varphi \& \psi]_{I,g} = [\varphi \wedge \psi]_{I,g}$ et $[\varphi \Rightarrow \psi]_{I,g} = [\varphi \rightarrow \psi]_{I,g}$.

Fait 2

Si ψ ne contient pas d'occurrences capturées de variables sauf v , alors

- a) $\exists v[\varphi \& \psi] = \exists v[\varphi \wedge \psi]$
- b) $\exists v[\varphi \Rightarrow \psi] = \exists v[\varphi \rightarrow \psi]$

Démonstration

a)
 Comme ψ ne contient pas d'occurrences capturées de variables sauf v ,
 1) $\psi[\uparrow(\varphi \& \psi)]_{j_1} / \uparrow(\varphi \& \psi) = \psi$ si $\uparrow \varphi \neq v$, et 2) $\psi[\uparrow(\varphi \& \psi)]_{j_1} / \uparrow(\varphi \& \psi) = \psi[\uparrow(\varphi \& \psi)]_{j_1} / v$ si $\uparrow \varphi = v$. Dans le premier cas $[\varphi \& \psi]_{I,g} = [\varphi \wedge \psi]_{I,g}$ et le fait est démontré. Dans le deuxième cas, considérons $[\exists v[\varphi \& \psi]]_{I,g}$, c'est-à-dire $\{ \langle f, \langle j_1, j_2 \rangle \rangle \mid f(v) \in D_a \text{ et } j_1 \in [\varphi[f(v)/v]]_{I,g} \text{ et } j_2 \in [\psi[f(v)/\uparrow(\varphi \& \psi)]_{j_1}]_{I,g} \}$; en conséquence de 2), $[\psi[f(v)/\uparrow(\varphi \& \psi)]_{j_1}]_{I,g} = [\psi[f(v)/v]]_{I,g}$, donc $[\exists v[\varphi \& \psi]]_{I,g} = [\exists v[\varphi \wedge \psi]]_{I,g}$.

b) est pareil.

Fait 3

Si φ ne contient pas de sous-formule de la forme $\exists v\chi$ ou de la forme $\forall v\chi$, alors

- a) $\varphi \& \psi = \varphi \wedge \psi$
- b) $\varphi \Rightarrow \psi = \varphi \rightarrow \psi$

Démonstration

Comme φ ne contient pas de sous-formule de la forme $\exists v\chi$ ou de la forme $\forall v\chi$, $\uparrow(\varphi \& \psi) = \uparrow(\varphi \Rightarrow \psi) = \emptyset$, donc $\psi[\underline{d}/\uparrow(\varphi \& \psi)] = \psi[\underline{d}/\uparrow(\varphi \Rightarrow \psi)] = \psi$.

Fait 4

- a) $[\exists v\varphi \& \psi] = \exists v[\varphi \wedge \psi]$
- b) $[\exists v\varphi \Rightarrow \psi] = \forall v[\varphi \rightarrow \psi]$

Démonstration

a)

$$[\exists v \varphi \& \psi]_{\mathbf{I},g} =$$

$$= \{ \langle j_1, j_2 \rangle \mid j_1 = \langle f, j_3 \rangle \text{ et } f(v) \in D_a \text{ et } j_3 \in [\varphi[f(v)/v]]_{\mathbf{I},g} \text{ et } j_2 \in [\psi[\uparrow(\varphi \& \psi)]_{j_1} / \uparrow(\varphi \& \psi)]_{\mathbf{I},g} \}$$

(en conséquence des définitions)

$$= \{ \langle j_1, j_2 \rangle \mid j_1 = \langle f, j_3 \rangle \text{ et } f(v) \in D_a \text{ et } j_3 \in [\varphi[f(v)/v]]_{\mathbf{I},g} \text{ et } j_2 \in [\psi[f(v)/v]]_{\mathbf{I},g} \}$$

(comme $\uparrow(\varphi \& \psi) = v$ et $\text{Iv}|_{j_1} = \mathbf{p}(j_1)(v) = f(v)$)

$$= \{ \langle f, j \rangle \mid f(v) \in D_a \text{ et } j = \langle j_3, j_2 \rangle \text{ et } j_3 \in [\varphi[f(v)/v]]_{\mathbf{I},g} \text{ et } j_2 \in [\psi[f(v)/v]]_{\mathbf{I},g} \}$$

(comme on peut identifier le couple $\langle \langle f, j_3 \rangle, j_2 \rangle$ avec le couple $\langle f, \langle j_3, j_2 \rangle \rangle$)

$$= [\exists v [\varphi \wedge \psi]]_{\mathbf{I},g}$$

b)

$$[[\exists v \varphi \Rightarrow \psi]]_{\mathbf{I},g} =$$

$$= \{ j \mid \text{pour chaque } j' = \langle f, j'' \rangle \text{ telle que } f(v) \in D_a \text{ et } j'' \in [\varphi[f(v)/v]]_{\mathbf{I},g}, j(j') \in [\psi[\uparrow(\varphi \Rightarrow \psi)]_{j'} / \uparrow(\varphi \Rightarrow \psi)]_{\mathbf{I},g} \}$$

(en conséquence des définitions)

$$= \{ j \mid \text{pour chaque } j' = \langle f, j'' \rangle \text{ telle que } f(v) \in D_a \text{ et } j'' \in [\varphi[f(v)/v]]_{\mathbf{I},g}, j(j') \in [\psi[f(v)/v]]_{\mathbf{I},g} \}$$

(comme $\uparrow(\varphi \Rightarrow \psi) = v$ et $\text{Iv}|_{j'} = \mathbf{p}(j')(v) = f(v)$)

$$= \{ j \mid \text{pour chaque } f \text{ telle que } f(v) \in D_a \text{ et pour chaque } j'' \in [\varphi[f(v)/v]]_{\mathbf{I},g}, j(f)(j'') \in [\psi[f(v)/v]]_{\mathbf{I},g} \}$$

(comme on peut identifier $j(\langle f, j'' \rangle)$ avec $j(f)(j'')$)

$$= [\forall v [\varphi \rightarrow \psi]]_{\mathbf{I},g}.$$
3. Un fragment d'anglais

Le fragment d'anglais auquel je ferai référence est le même que celui de [GS90]. Voici ses *catégories* et les expressions de base EB de chaque catégorie.

Catégories de base :

VI (syntagmes verbales intransitifs).EB: {walk, talk}

NC (syntagmes nominaux communs).EB: {man, donkey, farmer}

S ((séquences de) propositions).EB: $\emptyset$

Catégories dérivées :

Si A,B sont des catégories, A/B est une catégorie.

Voici quelques noms de catégories dérivées :

SN (syntagmes nominaux) = S/VI. EB: {Pedro, he₁, he₂, ...}

DET (déterminateur) = SN/NC. EB: {a₁, ..., every₁, ...}

VT (syntagmes verbaux transitifs) = VI/SN. EB: {own, beat}

Constructions syncatégorématiques

Formation de séquences de propositions :

Si ϕ, ψ sont des S, " $\phi. \psi$ " est une S.

Formation de propositions conditionnelles :

Si ϕ, ψ sont des S, "*if* ϕ *then* ψ " est une S.

Formation de propositions relatives restrictives:

Si ϕ est un NC et ψ est un VI, " ϕ *who* ψ " est un NC.

4. Traduction

Décrivons d'abord l'application f des catégories du fragment dans les types de LC :

$$1) f(S) = j$$

$$f(NC) = f(VI) = (c, j)$$

$$2) f(A/B) = (f(B), f(A))$$

Définissons maintenant la fonction g telle que

1) son domaine est l'ensemble des expressions de base du fragment exceptés les éléments de SN et DET

2) si ϕ est une expression de base de catégorie A et ϕ est dans le domaine de g , $g(u) \in \text{CON}_{f(A)}$.

Pour compléter la description de l'algorithme de traduction il nous reste à donner les règles de traduction des éléments de SN et DET et pour les constructions syncatégorématiques.

<i>Pedro</i>	est traduit en	$\lambda P[P(p)]$
<i>he_i</i>	est traduit en	$\lambda P[P(v_i)]$
<i>a_i</i>	est traduit en	$\lambda P\lambda Q\exists v_i[P(v_i)\&Q(v_i)]$
<i>every_i</i>	est traduit en	$\lambda P\lambda Q\forall v_i[P(v_i)\Rightarrow Q(v_i)]$

Si ϕ, ψ sont les traductions de ϕ, ψ , respectivement,

$\phi. \psi$	est traduit en	$\phi' \& \psi'$
<i>if</i> ϕ <i>then</i> ψ	est traduit en	$\phi' \Rightarrow \psi'$
ϕ <i>who</i> ψ	est traduit en	$\lambda v[\phi'(v) \& \psi'(v)]$

Il y aura aussi des *meaning postulates* pour VI et TV extensionnels (comme c'est le cas pour tous les verbes de notre fragment), qui garantissent que, si ϕ est la traduction d'un VI ou d'un VT, ϕ est définissable en termes de ϕ_* , qui est la traduction 'naturelle', c'est-à-dire une expression dénotant l'ensemble ou relation opportune, respectivement. Pour simplifier, dans les exemples suivants j'éviterai de distinguer entre ϕ et ϕ_* .

5. Quelques exemples

a) A_1 man walks. He_1 talks.

1) $\lambda P\lambda Q\exists v_1[P(v_1)\&Q(v_1)](\text{man})(\text{walk})\&\text{talk}(v_1)$
(application fonctionnelle et règle pour séquences)

2) $\exists v_1[\text{man}(v_1)\&\text{walk}(v_1)]\&\text{talk}(v_1)$
(λ -réductions)

3) $\exists v_1[\text{man}(v_1)\wedge\text{walk}(v_1)\wedge\text{talk}(v_1)]$
(faits 2 et 4)

4) $\sim(\exists v_1[\text{man}(v_1)\wedge\text{walk}(v_1)]\wedge\text{talk}(v_1))$
('extensionalization')

b) A_1 man walks. He_2 talks.

1) $\lambda P\lambda Q\exists v_1[P(v_1)\&Q(v_1)](\text{man})(\text{walk})\&\text{talk}(v_2)$
(application fonctionnelle et règle pour séquences)

2) $\exists v_1[\text{man}(v_1)\&\text{walk}(v_1)]\&\text{talk}(v_2)$
(λ -réductions)

$$3) \exists v_1 [\mathbf{man}(v_1) \wedge \mathbf{walk}(v_1)] \wedge \mathbf{talk}(v_2)$$

(fait 1)

$$4) \sim (\exists v_1 [\mathbf{man}(v_1) \wedge \mathbf{walk}(v_1)] \wedge \mathbf{talk}(v_2))$$

('extensionalization')

c) *If a₁ man walks, he₁ talks.*

$$1) [\lambda P \lambda Q \exists v_1 [P(v_1) \& Q(v_1)](\mathbf{man})(\mathbf{walk})] \Rightarrow \mathbf{talk}(v_1)$$

(application fonctionnelle et règle pour conditionnels)

$$2) \exists v_1 [\mathbf{man}(v_1) \& \mathbf{walk}(v_1)] \Rightarrow \mathbf{talk}(v_1)$$

(λ-réductions)

$$3) \exists v_1 [\mathbf{man}(v_1) \wedge \mathbf{walk}(v_1)] \Rightarrow \mathbf{talk}(v_1)$$

(fait 2)

$$4) \forall v_1 [\mathbf{man}(v_1) \wedge \mathbf{walk}(v_1) \rightarrow \mathbf{talk}(v_1)]$$

(fait 4)

$$5) \sim (\forall v_1 [\mathbf{man}(v_1) \wedge \mathbf{walk}(v_1) \rightarrow \mathbf{talk}(v_1)])$$

('extensionalization')

d) *If a₁ farmer owns a₂ donkey, he₁ beats it₂.*

$$1) \exists v_1 [\mathbf{farmer}(v_1) \& \exists v_2 [\mathbf{donkey}(v_2) \& \mathbf{own}(v_2)(v_1)]] \Rightarrow \mathbf{beat}(v_2)(v_1)$$

(application fonctionnelle, règle pour conditionnels et λ-réductions)

$$2) \exists v_1 [\mathbf{farmer}(v_1) \wedge \exists v_2 [\mathbf{donkey}(v_2) \wedge \mathbf{own}(v_2)(v_1)]] \Rightarrow \mathbf{beat}(v_2)(v_1)$$

(fait 2)

$$3) \exists v_1 \exists v_2 [\mathbf{farmer}(v_1) \wedge \mathbf{donkey}(v_2) \wedge \mathbf{own}(v_2)(v_1)] \Rightarrow \mathbf{beat}(v_2)(v_1)$$

(logique)

$$4) \forall v_1 \forall v_2 [\mathbf{farmer}(v_1) \wedge \mathbf{donkey}(v_2) \wedge \mathbf{own}(v_2)(v_1)] \rightarrow \mathbf{beat}(v_2)(v_1)$$

(fait 4)

$$5) \sim (\forall v_1 \forall v_2 [\mathbf{farmer}(v_1) \wedge \mathbf{donkey}(v_2) \wedge \mathbf{own}(v_2)(v_1)] \rightarrow \mathbf{beat}(v_2)(v_1))$$

('extensionalization')

e) *Every₁ farmer who owns a₂ donkey, beats it₂.*

L'application fonctionnelle, la règle pour la formation des relatives, les *meaning postulates* et les λ-réductions nous donnent la traduction suivante de *farmer who owns a₂ donkey*:

$$1) \lambda v [\mathbf{farmer}(v) \& \exists v_2 [\mathbf{donkey}(v_2) \& \mathbf{own}(v_2)(v)]]$$

Par conséquent $Every_1$ *farmer who owns a* a_2 *donkey* reçoit les traductions suivantes :

2) $\lambda P \lambda Q \forall v_1 [P v_1 \Rightarrow Q v_1] (\lambda v [\text{farmer}(v) \& \exists v_2 [\text{donkey}(v_2) \& \text{own}(v_2)(v)]])$
(application fonctionnelle)

3) $\lambda Q \forall v_1 [[\text{farmer}(v_1) \& \exists v_2 [\text{donkey}(v_2) \& \text{own}(v_2)(v_1)]] \Rightarrow Q v_1]$
(λ -réduction)

Les *meaning postulates* et les λ -réductions nous donnent pour le verbe intransitif *beats it* $_2$ la traduction suivante :

4) $\lambda v (\text{beat}(v_2)(v))$

Par conséquent e) reçoit les traductions suivantes :

5) $\lambda Q \forall v_1 [[\text{farmer}(v_1) \& \exists v_2 [\text{donkey}(v_2) \& \text{own}(v_2)(v_1)]] \Rightarrow Q(v_1)] (\lambda v (\text{beat}(v_2)(v)))$
(application fonctionnelle)

6) $\forall v_1 [[\text{farmer}(v_1) \& \exists v_2 [\text{donkey}(v_2) \& \text{own}(v_2)(v_1)]] \Rightarrow \text{beat}(v_2)(v_1)]$
(λ -réduction)

7) $\forall v_1 [[\text{farmer}(v_1) \wedge \exists v_2 [\text{donkey}(v_2) \wedge \text{own}(v_2)(v_1)]] \Rightarrow \text{beat}(v_2)(v_1)]$
(faits 2 et 3)

8) $\forall v_1 [\exists v_2 [\text{farmer}(v_1) \wedge \text{donkey}(v_2) \wedge \text{own}(v_2)(v_1)] \Rightarrow \text{beat}(v_2)(v_1)]$
(logique)

9) $\forall v_1 \forall v_2 [\text{farmer}(v_1) \wedge \text{donkey}(v_2) \wedge \text{own}(v_2)(v_1)] \rightarrow \text{beat}(v_2)(v_1)]$
(fait 4)

10) $\sim (\forall v_1 \forall v_2 [\text{farmer}(v_1) \wedge \text{donkey}(v_2) \wedge \text{own}(v_2)(v_1)] \rightarrow \text{beat}(v_2)(v_1))$
(*'extensionalization'*)

6. Conclusion

Je ne me propose pas de faire ici une comparaison avec d'autres approches possibles. Je voudrai seulement noter les différences plus importantes entre mon approche et les approches de G&S dans [GS90] et de A.Ranta dans [R92].

Comme je l'ai dit plus haut, mon approche applique la même stratégie de solution adoptée par G&S, mais la tactique est différente. Les différences principales concernent (i) l'interprétation des constantes logiques qui permet d'obtenir les équivalences (15) et (16), et (ii) la technique pour distinguer les cas où

l'interprétation pertinente des constantes est celle usuelle des cas où elle est celle dynamique.

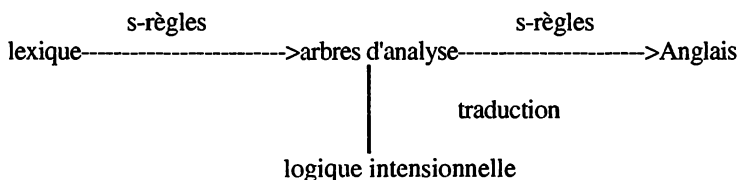
Pour ce qui concerne (i), G&S emploient l'interprétation dynamique des constantes logiques. Intuitivement, l'idée à la base de cette interprétation est que "the meaning of a sentence resides in its information change potential. Sentences carry information, and the best way that this information can be characterized is by showing how it changes information, for example the information of someone who processes the sentence. This idea is quite general, and one way to capture part of it is by looking upon the meaning of a sentence as something which tells us which propositions are true after its contents have been processed. We may assume that this processing takes place in some situation, so that we can regard the extension of a sentence in a situation as those propositions which are true after the sentence has been processed in that situation." ([GS91], p. 48) Par conséquent les phrases sont traduites dans leur langage logique par des expressions de type $((s,t),t)$, c'est-à-dire des expressions dénotant des ensembles de propositions.

Pour ce qui concerne (ii), tandis que mon approche prévoit l'introduction de deux constantes logiques nouvelles dont le signifié est défini dans le métalangage sémantique, G&S introduisent dans le langage logique deux catégories syntaxiques nouvelles : celle des *discourse markers* et celle des *state switchers*. L'idée générale est la suivante. "Discourse markers are used in the translations of indefinite terms and pronouns. Part of the meaning of the indefinite term is to assign a value, one that is a witness of the term, to the discourse marker. Also, provisions are taken to pass on this value to further occurrences of the same discourse marker. This is taken care of, among other things, by the state switchers. The denotations of discourse markers are determined with reference to a state, and the state switcher will allow us to pick out certain states. More specifically, it will allow us to identify states in which discourse markers have certain values, by switching from the current state to such a state. Finally, the meanings of sentences will be constructed in such a way that they determine that subsequent sentences are interpreted with respect to the most recently identified state. This they bring about by being themselves a kind of descriptions of states [...]" (ibid., p. 42)

L'idée d'employer l'interprétation constructive pour affronter le problème des donkey sentences n'est pas nouvelle. Elle a été suggérée, par exemple, par G. Sundholm dans [S86] et [S89], et systématiquement développée par A. Ranta dans [R92]. Ces approches utilisent comme langage logique le langage de la théorie des types de Martin-Löf, qui permet de faire une référence directe à une preuve d'une proposition. Le langage que j'emploie est différent dans le sens où il dispose de termes pour dénoter des ensembles (pas de preuves mais) de justifications pour une proposition, mais pas de termes pour dénoter des éléments de ces ensembles. La référence à une justification spécifique, à sa projection gauche ou droite, etc., est possible seulement au niveau du métalangage sémantique.

Une autre différence entre l'approche de [R92] et non seulement la mienne mais toutes les approches qui s'inspirent de la Montague Grammar concerne la structure de la grammaire. Voici comment Ranta lui-même schématise la différence. Les deux types de grammaire ont les structures suivantes :

MONTAGUE GRAMMAR



RANTA



D'un côté les s-règles (règles syntaxiques) de Montague jouent le double rôle de combiner expressions de base en arbres d'analyse et de transformer [*sugaring*] les arbres d'analyse en anglais. De l'autre côté Ranta ne distingue pas entre le langage des arbres d'analyse et le langage logique, en employant un seul langage formel qui sert comme base aussi bien pour la représentation du signifié des phrases que pour l'opération de *sugaring*. Donc il n'y a pas dans son approche un processus de traduction, et ce fait rend plus difficile une comparaison avec la Montague Grammar. En particulier, ce n'est pas clair si une traduction de l'anglais (ou plus précisément du langage des arbres d'analyse) dans le langage de la théorie des types serait compositionnelle dans le sens défini au début.

Bibliographie

- [GS90] GROENENDIJK J. and M. STOKHOF, "Dynamic Montague Grammar", in L. KALMAN and L. POLOS (eds.), *Papers from the Second Symposium on Logic and Language*, Akademiai Kiado, Budapest.
- [GS91] GROENENDIJK J. and M. STOKHOF, "Dynamic Predicate Logic", *Linguistics and Philosophy* 14, 39-100.
- [Ma92] MARCONI D., "Semantica cognitiva", in M. SANTAMBROGIO (ed.), *Introduzione alla filosofia analitica del linguaggio*, Laterza, Roma-Bari.
- [Mo73] MONTAGUE R., "The proper treatment of quantification in ordinary English", in J. HINTIKKA et al. (eds.), *Approaches to Natural Language*, Reidel, Dordrecht.
- [R92] RANTA A., "Intuitionistic Categorical Grammar", *Linguistic and Philosophy* 15.
- [S86] SUNDHOLM G., "Proof Theory and Meaning", in D. GABBAY and F. GUENTHNER (eds.), *Handbook of Philosophical Logic*, Reidel, Dordrecht.
- [S89] SUNDHOLM G., "Constructive Generalized Quantifiers", *Synthese* 79, 1-12.

Gabriele USBERTI

Chapitre 15

Approximations et négation

1. Introduction

On ne saurait nier l'importance des approximations et de la négation dans l'étude de la connaissance. Ces notions impliquent en elles-mêmes des difficultés évidentes. Plus encore, leur combinaison amène des complications redoutables et rend nécessaire un retour sur des présupposés de l'analyse logique moderne. Ainsi dans un contexte d'approximations, le concept de négation se divise en deux sous-concepts dont les propriétés sont logiquement différentes. De façon surprenante, puisqu'il s'agit d'un corollaire inattendu, certaines subtilités de la langue naturelle deviennent alors analysables dans un contexte purement logique¹.

Nous allons tout d'abord présenter la problématique rattachant la négation aux approximations. Des considérations élémentaires nous permettront rapidement de déterminer comment la négation doit être formalisée. Nous serons ensuite conduits à une variante mineure de la construction de Higgs de la catégorie des ensembles sur des algèbres de Heyting qui fournit le contexte approprié pour rendre compte des principaux aspects logiques des approximations numériques.

2. La problématique fondamentale des approximations et de la négation

Précisons d'abord le genre d'approximations dont il est question ici. La connaissance scientifique repose sur des approximations numériques, tant sur le plan expérimental que sur le plan théorique. Dans les deux cas, les énoncés de base sont de la forme ' $F(x_1, \dots, x_n) = r$ ', où F est une fonction n -aire qui prend généralement

1. Ce lien avec la négation dans les langues naturelles est exploré plus à fond par Lapalme-Reyes et. al. 1994.

une valeur dans un espace métrique. Un énoncé p peut alors être dit partiellement vrai de la manière suivante. Soit $p = 'F(x_1, \dots, x_n) = r'$, une prédiction théorique par exemple. Si cette prédiction est vérifiable, alors il existe un énoncé $q = 'F(x_1, \dots, x_n) = s'$ auquel sera confronté p et qui est considéré comme étant totalement vrai. La distance entre les valeurs r et s permet alors de construire la distance entre l'énoncé p et la vérité, donc la valeur de vérité (partielle) de p . La construction consiste à passer de l'espace d'évaluation de la fonction F à un certain espace quotient, appelé l'espace des valeurs de vérité (Voir Marquis 1991 pour les détails.) Le plus souvent, l'espace des valeurs de vérité est un ensemble linéairement ordonné, isomorphe à l'intervalle des réels $[0, 1]$. Nous dénoterons par $\|F(x) = r\|$ la valeur de vérité de l'énoncé $'F(x) = r'$, valeur de vérité qui peut être entre le vrai et le faux.

La définition des opérateurs logiques sur un espace de valeurs de vérité $[0, 1]$ est particulièrement délicate. En effet, il est possible de définir plusieurs opérateurs distincts sur l'intervalle des réels $[0, 1]$ et d'obtenir ainsi des structures différentes. La question est de savoir lequel ou lesquels de ces opérateurs sont le plus adéquats et selon quels critères.

Une remarque informelle indique la direction à prendre pour la formalisation.

"Jean vient d'avoir 10 ans. Son ami Pierre croit qu'il a 11 ans, tandis que son amie Jeanne soupçonne qu'il en ait 9. Les deux se trompent, mais de peu... Charles, ..., n'est pas sûr de l'âge de Jean et, étant une personne très prudente, ne prend pas de risque et affirme "Jean n'a pas 9 ans". Il a évidemment raison..." [Bunge, 1981]².

Il s'agit bel et bien d'un exemple simple d'énoncés qui ont la forme appropriée. Ce que cet exemple illustre, c'est le fait que les énoncés négatifs ont dans un tel contexte un comportement déterminé. En effet, dans ce cadre, une proposition négative a "plus de chances" d'être vraie qu'une proposition affirmative. Il n'y a qu'une seule valeur pour laquelle une proposition exprimant une approximation quantitative est totalement vraie, pour toutes les autres valeurs, la négation de cette proposition sera vraie. Cette opération, appelée 'supplément', peut donc être formellement représentée sur l'intervalle $[0, 1]$ de la manière suivante :

$$\text{pour tout } a \in [0, 1], \quad \sim a = \begin{cases} 0 & \text{si } a = 1 \\ 1 & \text{si } a \neq 1 \end{cases}$$

Cet opérateur est bien connu ; il a d'abord été défini et étudié systématiquement par Ribenboim dans les années quarante (voir Ribenboim 1949). Il est parfois désigné le 'sup-complément'; c'est le dual de l'inf-complément, utilisé dans les modélisations algébriques de la logique intuitionniste. Il ne figure cependant pas dans la littérature des années cinquante et soixante des logiques polyvalentes et ne réapparaît que dans les années soixante-dix et quatre-vingt dans la littérature sur la théorie des ensembles flous (voir, par exemple, Lowen 1978, Ovchinilov 1983).

2. Voir également Bunge 1974, 1983, 1990 et Marquis 1992.

Il semble que Bunge ait été le premier à indiquer comment cet opérateur est lié au contexte d'approximations propre aux sciences.

On peut aisément penser que la négation discutée dans l'exemple de Bunge affecte l'égalité même d'une équation ' $F(x) = r$ ' : ainsi, affirmer "Jean n'a pas dix ans" correspondrait à déclarer "l'âge de Jean n'est pas égale à dix" ou, pour employer une forme plus proche du symbolisme logique, " $\text{âge}(\text{Jean}) \neq 10$ ". Évidemment, il y a une autre façon de 'nier' cette affirmation : en attribuant un âge différent à Jean, par exemple 12 ans. Cette forme de négation ne porte pas sur l'égalité, mais sur la valeur de la propriété elle-même ; son comportement logique est considérablement différent de celui du supplément. En effet, supposons que Jean a 11 ans. Quiconque affirme que Jean n'a pas 10 ans aura raison et par conséquent fera une affirmation vraie. Toutefois, quiconque affirme que Jean a 12 ans, ce que nous considérons également comme une négation de la vérité, a tort et fait par conséquent une affirmation fausse, bien que très proche de la vérité. Ce qu'il importe de remarquer ici, c'est qu'il semble y avoir deux manières distinctes de nier une affirmation dans le contexte des approximations : la première s'applique à la relation d'égalité, alors que la seconde ne porte que sur la valeur elle-même, qui se retrouve ainsi d'une certaine manière, à l'intérieur de la relation d'égalité. Dans la sémantique, cette distinction peut être introduite de la manière suivante :

$$\| \sim (F(x) = r) \| = 1 \text{ si et seulement si } \| F(x) = r \| \neq 1 ;$$

$$\| \neg (F(x) = r) \| = 1 \text{ si et seulement si } \| F(x) = r \| = 0.$$

Ce qui se lit informellement comme suit : dans le premier cas, la négation 'prudente' est vraie si et seulement si l'affirmation n'est pas vraie ; dans le second cas, la négation est vraie si et seulement si l'affirmation est fausse. Ne pas être vrai n'équivaut donc pas à être faux.

Étant donné cette définition, il importe alors de déterminer comment les autres opérateurs logiques doivent être définis et ensuite la classe des énoncés valides qui en découle. Certains résultats élémentaires sur cette question ont déjà été présentés dans un autre contexte (voir Marquis 1992). Une des questions fondamentales qui s'impose est celle de la généralisation de ces résultats. La voie suivante semble être la plus adéquate.

Soit $\mathbf{B} = \langle B, 0, 1, \wedge, \vee \rangle$ un treillis possédant un élément maximal, 1, et un élément minimal, 0.

Définition : une algèbre de Heyting $\mathbf{B}$ est un treillis tel que pour toute paire (a, b) d'éléments il existe un élément $(a \rightarrow b)$ tel que $c \leq (a \rightarrow b)$ si et seulement si $c \wedge a \leq b$.

Les structures recherchées sont les structures duales, dans le sens algébrique de cette expression, des algèbres de Heyting. Pour cette raison, elles sont appelées algèbres de co-Heyting. Nous les appelons également algèbres de Bunge, pour les raisons mentionnées précédemment. Elles peuvent être définies directement de la manière suivante :

Définition : une algèbre de co-Heyting est un treillis tel que pour toute paire (a, b) d'éléments il existe un élément $(a \setminus b)$ tel que $(a \setminus b) \leq c$ si et seulement si $a \leq b \vee c$.

Il importe de rappeler ici quelques faits élémentaires. Il est bien connu que le complément se définit, dans le cas des algèbres de Heyting, de la manière suivante : $\neg a =_{df} a \rightarrow 0$. Ainsi $a \wedge \neg a = 0$ pour tout a , mais l'égalité $a \vee \neg a = 1$ n'est pas toujours vérifiée. Par ailleurs, le supplément ' $\sim$ ' se définit par $\sim a =_{df} 1 \setminus a$. On peut alors facilement vérifier qu'il satisfait les lois duales du complément : $a \vee \sim a = 1$ est toujours vraie et $a \wedge \sim a = 0$ ne l'est pas toujours ; on obtient la définition du supplément lorsque le treillis est l'intervalle des réels $[0, 1]$ avec comme infimum l'opérateur min et comme supremum l'opérateur max. Le supplément semble ainsi capturer les propriétés essentielles des contradictoires et le complément les propriétés essentielles des contraires. En effet, des termes contraires peuvent très bien être simultanément faux. La loi du tiers exclu ne s'applique donc pas à des termes contraires, bien que la loi de la (non)contradiction demeure valide. Comme des termes contradictoires ne peuvent être simultanément faux, dans leur cas la loi du tiers exclu et la loi de la (non)contradiction sont valides.

La sémantique générale s'élabore de la manière suivante. Soit les fonctions $F : X_1 \times \dots \times X_n \rightarrow Y$ et les énoncés de la forme $F(x_1, \dots, x_n) = y$. Cette dernière égalité est une relation dans l'ensemble Y , donc un sous-ensemble de $Y \times Y$. Par conséquent, il convient de prendre comme point de départ une fonction $e : Y \times Y \rightarrow B$, où B est une algèbre de bi-Heyting fixe, qui attribue aux égalités dans Y une certaine valeur dans B , valeur qui représente la valeur de vérité de l'égalité en question.

Il est intuitif d'admettre que la fonction e doit satisfaire au moins les conditions suivantes :

1. $e(x, y) = e(y, x)$ (symétrie);
2. $e(x, y) \wedge e(y, z) = e(x, z)$ (transitivité).

Nous pouvons ensuite définir une catégorie $\mathbf{B}$ comme suit :

Les objets de $\mathbf{B}$ sont les fonctions $e : Y \times Y \rightarrow B$ — où B dénote une algèbre de bi-Heyting fixe et Y est un ensemble arbitraire — qui satisfont les conditions 1.-2. précédentes.

Un morphisme de $\mathbf{B}$ entre $e : Y \times Y \rightarrow B$ et $f : X \times X \rightarrow B$ est une application $\alpha : Y \times X \rightarrow B$ telle que pour tout y, y', x, x'

- i. $\alpha(y, x) \wedge e(y, y') \leq \alpha(y', x)$, respect de l'égalité dans Y ,
- ii. $\alpha(y, x) \wedge f(x, x') \leq \alpha(y, x')$, respect de l'égalité dans X ,
- iii. $\alpha(y, x) \wedge \alpha(y, x') \leq f(x, x')$,
- iv. $\bigcup \alpha(y, x) = 1$.

La composition de deux morphismes $\alpha : Y \times X \rightarrow B$ et $\beta : X \times Z \rightarrow B$ est définie par

$$\beta \alpha(y, z) = \bigcup (\alpha(y, x) \wedge \beta(x, z)).$$

Quelques commentaires s'imposent. La définition précédente est la définition habituelle des ensembles évalués sur une algèbre de Heyting complète, notion

introduite par Higgs dans les années soixante-dix et généralisée ultérieurement par plusieurs (voir par exemple Makkai & Reyes 1977, Goldblatt 1979 et Higgs 1984). L'unique différence, ici, tient dans le fait que les ensembles sont évalués sur une algèbre de *bi*-Heyting. Il est bien connu que, dans le cas d'une évaluation sur une algèbre de Heyting complète, la catégorie obtenue est alors équivalente à un topos de faisceaux sur l'algèbre de Heyting en question. On peut aisément montrer que l'évaluation sur une algèbre de *bi*-Heyting donne une catégorie qui est équivalente à un topos de faisceaux sur l'algèbre de *bi*-Heyting, qu'un tel topos est un topos de *bi*-Heyting, c'est-à-dire que le treillis des sous-objets d'un objet *X* quelconque du topos est une algèbre de *bi*-Heyting.

Certains pourraient s'étonner de l'absence de la condition de réflexivité sur les fonctions *e*, c'est-à-dire de l'équation $e(x, x) = 1$. Comme c'est le cas pour la construction de Higgs, son ajout apparemment banal a des conséquences dévastatrices sur la structure globale de la catégorie résultante et en particulier sur ce qu'il convient d'appeler sa structure logique. En effet, la catégorie n'ayant plus d'égalisateurs elle n'a plus la totalité des limites finies. Mais peut-être existe-t-il malgré tout suffisamment de structure pour développer la logique sur une autre base. Cerruti & Höhle ont effectivement montré que, pour des ensembles évalués sur des structures un peu plus générales auxquelles on imposait la réflexivité, la catégorie résultante est "topos faible" (*weak-topos*), pouvant servir de fondements à la théorie des ensembles flous (voir Cerruti & Höhle 1986, Höhle 1991, Stout 1991, Höhle & Stout 1991). Cette question exigerait cependant d'être examinée de plus près.

Par ailleurs, d'autres formalisations demeurent possibles. La Palme Reyes et al. 1994 en ont proposé une autre, bien que présentée dans un tout autre contexte. Nous l'aborderons ultérieurement, en lieu avec la question des langues naturelles, puisque c'est dans ce contexte que La Palme Reyes et al. ont élaboré leur modèle.

Nous nous attarderons d'abord à des considérations syntaxiques, où quelques surprises nous attendent. En effet, de prime abord, il est raisonnable de croire que le supplément doit se formaliser sans difficulté dans un système déductif et qu'un tel système peut se rattacher à la sémantique précédente ; il suffit de modifier le langage interne d'un topos de *bi*-Heyting pour obtenir le système déductif approprié. C'est ici que la surprise surgit. Car, comme nous allons le montrer, si l'opérateur de supplément est définissable dans un tel contexte, alors le topos est nécessairement booléen ! En d'autres termes, il est impossible de définir ou de distinguer de manière interne l'opérateur de supplément, ce qui toutefois ne veut pas dire qu'il ne soit pas présent.

Nous allons tout d'abord présenter un système de séquents (propositionnels) pour ce nouvel opérateur de négation. Nous montrerons par la suite dans quelles conditions le système s'effondre pour devenir classique, i.e. booléen. Nous adopterons une présentation sous la forme d'un calcul de séquents en terme de règles d'adjonction, que l'on doit à Lawvere. La raison de ce choix est très simple : nous ne connaissons aucune autre présentation des nouveaux opérateurs logiques. En particulier, il n'existe à notre connaissance aucune présentation axiomatique (à la Hilbert) d'un tel système.

Certaines des règles sont familières, d'autres sont étranges à première vue. Il importe de noter qu'il s'agit ici d'une transcription pratiquement littérale des règles

d'adjonctions pour les différents opérateurs ce qui donne une apparence mystérieuse à certaines règles, comme dans le cas de la règle (T). L'avantage de cette présentation est de rendre la démonstration de la complétude du système pratiquement transparente, puisque les règles sont en fait très proches de la sémantique.

Note : une ligne horizontale double signifie que l'on peut passer des séquents du haut à ceux du bas et vice-versa.

$$(Id) \frac{\quad}{\Gamma, \phi \vdash \phi}$$

$$(Cut) \frac{\Gamma \vdash \phi \quad \Delta, \phi \vdash \varphi}{\Delta, \Gamma \vdash \varphi}$$

$$(T) \frac{\quad}{\Gamma \vdash T}$$

$$(\wedge) \frac{\Gamma \vdash \phi \quad \Gamma \vdash \varphi}{\Gamma \vdash \phi \wedge \varphi}$$

$$(\perp) \frac{\quad}{\Gamma, \perp \vdash \phi}$$

$$(\vee) \frac{\Gamma, \phi \vdash \theta \quad \Gamma, \varphi \vdash \theta}{\Gamma, \phi \vee \varphi \vdash \theta}$$

$$(\rightarrow) \frac{\Gamma, \phi \vdash \varphi}{\Gamma \vdash \phi \rightarrow \varphi}$$

$$(\neg) \frac{\Gamma, \phi \vdash \perp}{\Gamma \vdash \neg \phi}$$

$$(\wedge) \frac{\phi \wedge \varphi \vdash \theta}{\phi \vdash \varphi \vee \theta}$$

$$(\sim) \frac{\sim \phi \vdash \theta}{T \vdash \varphi \vee \theta}$$

Remarquons d'emblée l'absence de Γ dans les deux dernières règles. Cette absence est volontaire et, en fait, capitale. En effet, l'introduction de Γ provoquerait un effondrement du système, qui deviendrait simplement booléen. On obtient alors une présentation un peu tordue de la logique classique, dans la mesure où les deux opérateurs de négation seraient identiques.

Proposition : L'introduction de Γ dans les deux dernières règles nous donne une logique booléenne, c'est-à-dire que, dans ce cas, $\sim \alpha = \neg \alpha$; cette négation satisfait alors tous les théorèmes de la négation de la logique classique³.

3. Ce fait a été remarqué par plusieurs personnes plus ou moins simultanément.

Preuve : Soit $\Gamma := \{\alpha\}$. Il est suffisant de démontrer que $\alpha \vee \neg\alpha$ est un théorème. En effet, si la dernière formule est démontrable, alors il s'ensuit automatiquement que les deux négations sont équivalentes et que tous les théorèmes de la négation de la logique classique sont démontrables.

Nous avons alors la déduction suivante :

$$\underline{\underline{\alpha, \phi \backslash \psi \vdash \theta}}$$

$$\underline{\underline{\alpha, \phi \vdash \psi \vee \theta}} \quad \underline{\underline{\alpha, \phi \vdash \alpha}}$$

$$\underline{\underline{\alpha, \phi \vdash \alpha \wedge (\psi \vee \theta)}} \quad \underline{\underline{\alpha \wedge (\psi \vee \theta) \vdash (\alpha \wedge \psi) \vee (\alpha \wedge \theta)}}$$

$$\underline{\underline{\alpha, \phi \vdash (\alpha \wedge \psi) \vee (\alpha \wedge \theta)}} \quad \underline{\underline{(\alpha \wedge \psi) \vee (\alpha \wedge \theta) \vdash (\alpha \wedge \psi) \vee \theta}}$$

$$\underline{\underline{\alpha, \phi \vdash (\alpha \wedge \psi) \vee \theta}}$$

$$\underline{\underline{\alpha, \phi \backslash (\alpha \wedge \psi) \vdash \theta}}$$

$$\underline{\underline{\alpha, \top \backslash (\alpha \wedge \alpha) \vdash \theta}}$$

$$\underline{\underline{\alpha, \top \backslash \alpha \vdash \theta}}$$

$$\underline{\underline{\alpha, \sim\alpha \vdash \theta}}$$

Il est ainsi possible de déduire n'importe quelle formule θ de $\alpha, \sim\alpha$ sous l'hypothèse que Γ n'est pas vide. Donc $\alpha \wedge \sim\alpha = \perp$. Par ailleurs, nous avons que $\sim\alpha \vdash \neg\alpha$. Mais, puisque nous avons également $\alpha \vee \sim\alpha = \top$, α est complémenté pour tout α , ce qui signifie que $\sim\alpha = \neg\alpha$.

Ce système déductif, sans être booléen, demeure presque booléen. Nous frappons aux portes de la logique classique sans y entrer. De plus, comme l'ont montré Makkai & Reyes 1992, ce système déductif constitue une extension conservatrice de la logique intuitionniste propositionnelle : si $\phi \vdash \psi$, où ϕ et ψ sont des formules de la logique intuitionniste, est démontrable dans le système précédent, alors $\phi \vdash \psi$ est démontrable dans la logique intuitionniste. Notre opérateur de négation a par ailleurs d'autres propriétés fondamentales et étonnantes, dont la plus importante tient certainement dans le fait qu'il n'est pas fonctoriel (voir Reyes & Zolfaghari 1994).

Il vaut la peine, à ce point-ci, de nous attarder à quelques remarques historiques ainsi qu'à quelques comparaisons.

3. Logique classique et négation

La logique classique contemporaine ne reconnaît qu'un seul opérateur de négation ; cet opérateur est propositionnel⁴. Ce postulat semble être apparu explicitement avec Frege qui, dans son article sur ce sujet, déclare sans sourciller qu'"il semble souvent que le mot ou la syllabe qui portent la négation sont associés plus étroitement à une certaine partie de la proposition, par exemple le prédicat. On pourrait en garder le sentiment que seul le contenu de cette partie est nié, plutôt que le contenu de la proposition entière. Mais en disant d'un homme qu'il est inconnu (*unberühmt* — non célèbre), on pose comme fausse la pensée qu'il est un homme célèbre" (Frege, 1919, 1971, 209-210).

On peut aisément comprendre de quelle façon Frege en arrive à cette conviction. En effet, l'analyse de Frege est une conséquence directe de son holisme sémantique, holisme selon lequel le sens d'une expression ou d'un terme doit toujours être recherché dans des jugements complets et non pas isolément. D'où la fameuse révolution frégeenne selon laquelle, dans nos analyses, les jugements précèdent les concepts et selon laquelle les concepts sont identifiés par une analyse des jugements ou des contenus-de-jugements. Il s'agit là d'un renversement radical de l'analyse logique telle qu'elle était conçue avant Frege : dans une telle analyse, les concepts étaient d'abord donnés et les jugements étaient ensuite obtenus par leur liaison. On conçoit pourquoi, chez Frege, la négation *doit* être propositionnelle : le résultat de l'analyse logique d'un jugement (atomique) doit fournir une fonction et des arguments. Ce cadre conceptuel interdit d'avoir une fonction qui comporterait 'en elle-même' une négation ; la négation ne saurait donc s'appliquer qu'à la fonction, elle-même évaluée pour tel et tel argument⁵.

4. Pour une discussion détaillée de certains aspects historiques entourant l'analyse logique du concept de négation, voir Horn 1989 et pour un résumé lié particulièrement aux négations traitées dans ce texte, voir La Palme Reyes et al. 1994.

5. La Palme Reyes et al. affirment, suite à une suggestion faite par Lawvere, que c'est probablement à la suite de l'influence de Kant qui, selon Lawvere, déduit ses catégories à partir des jugements, que la négation des termes a été négligée. Bien qu'intéressante et ingénieuse, cette remarque nous semble néanmoins inadéquate. D'abord, Kant lui-même, dans son traité de logique, suit la tradition et procède des concepts aux jugements. De plus, il reconnaît dans ces notes la possibilité de deux types distincts de négation. Il affirme en effet, après avoir distingué les jugements affirmatifs, négatifs et indéfinis, que "dans les jugements négatifs la négation affecte toujours la copule ; dans les jugements indéfinis ce n'est pas la copule, mais le prédicat qui est affecté par la négation, ce qui peut être exprimé au mieux en latin" (Kant, *Logique*, 114). Auparavant, Kant constatait toutefois que "la logique ayant simplement affaire à la forme du jugement et non aux concepts considérés dans leur contenu, la distinction entre les jugements indéfinis et les jugements négatifs n'est pas du ressort de cette science" (Kant, *Logique*, 114). En d'autres mots, pour Kant, la distinction entre les deux types de négations n'est pas une distinction logique. Par ailleurs, il faut voir comment Kant déduit les catégories des jugements et en quoi il serait un précurseur de Frege sur le lien entre les jugements et les concepts. Le passage clé est le paragraphe B105 de l'*Analytique Transcendantale*. À cet endroit, Kant affirme :

"La même fonction qui donne l'unité aux représentations diverses *dans un jugement*, donne aussi à la simple synthèse de représentations diverses *dans un intuition* l'unité, qui,

Nul besoin d'insister sur le fait que, dans ces conditions, la forme logique s'appauvrit considérablement ; elle semble incapable de rendre justice à des phénomènes pourtant fondamentaux, dont l'exclusion du champ de la logique apparaît pour le moins douteuse.

Si l'on considère les énoncés suivants : (i) cet homme n'est pas heureux ; (ii) cet homme est malheureux ; peut-on déclarer avec Frege que, dans les deux cas, nous indiquons la fausseté de la pensée que l'homme est heureux ? Entre le bonheur et le malheur, il y a toute une gamme d'états possibles. Ne peut-on pas dire qu'il est possible d'être heureux de différentes manières ou à différents niveaux, par exemple au travail, à la maison, etc. Et, par conséquent, qu'il est possible de s'approcher du bonheur, d'être partiellement heureux, rejoignant ainsi une forme d'approximation ?

Il ne s'agit que d'un exemple d'un phénomène très répandu dans les langues naturelles. En français, ce phénomène apparaît sous la forme d'un préfixe comme "i-", "a-", "mal-", "dés-", "mé-" ou "non-", comme dans les expressions "imbattable", "impossible", "indécidable", "irrationnel", "immobile", "amoral", "aphone", "malheureux", etc.

Selon Aristote, cette négation des termes est en fait une affirmation⁶. Il n'est pas clair que ce soit le cas en général. Il existe au moins une catégorie de cas pour lesquels on peut le montrer formellement, et ce, même dans le cadre de la logique classique où il n'y a qu'une seule négation.

Proposition (triviale mais néanmoins intéressante) : dans le cadre de la logique classique, la négation 'interne' d'une relation asymétrique est toujours une affirmation, i.e. elle est équivalente à une formule positive.

Preuve : soit $R(x, y)$ une relation asymétrique et supposons que nous ayons $\neg R(x, y)$. Puisque $R(x, y) \rightarrow \neg R(y, x)$, il s'ensuit par substitution propositionnelle de $p/\neg p$ que $\neg R(x, y) \rightarrow \neg\neg R(y, x)$. Mais en logique classique, $\neg\neg R(y, x) \leftrightarrow R(y, x)$.

exprimée généralement, s'appelle le concept pur de l'entendement. Le même entendement donc, et bien par les mêmes actes par lesquels il établissait dans les concepts, au moyen de l'unité analytique, la forme logique d'un jugement, apporte aussi dans ses représentations, au moyen de l'unité synthétique du divers dans l'intuition en général, un contenu transcendantal, en raison de quoi ces représentations s'appellent purs concepts de l'entendement qui se rapportent *a priori* aux objets, ce que ne peut faire la logique générale".

Il est clair ici que c'est la *fonction logique* appliquée aux jugements, c'est-à-dire l'unité qu'apporte la copule entre le sujet et le prédicat, qui est transférée aux intuitions et qui donne les concepts purs recherchés. En d'autres mots, les jugements, qui sont toujours de la forme sujet-copule-prédicat pour Kant, ont une unité. Par ailleurs, la représentation dans l'intuition est au départ chaotique et il faut chercher du côté de la fonction logique de cette unité ce qui pourra fournir l'unité de l'objet dans la représentation. Il faut donc reconnaître qu'il n'y a pas de passage direct des jugements aux concepts comme chez Frege, mais plutôt l'identification d'une fonction logique dans les jugements qui est ensuite transférée à l'intuition, ce qui est totalement différent.

6. Bossuet, suivant certainement Aristote sur ce point, dans la *Logique du Dauphin* fait remarquer que plusieurs des ces termes négatifs sont en réalité des affirmations au même titre que leur contraire. Considérez indécis et immobile.

Par hypothèse, nous avons $\neg R(x, y)$; nous pouvons donc, par modus ponens, détacher le conséquent pour obtenir $R(y, x)$ et la preuve est complète.

Ce résultat est informellement transparent. Considérez l'adjectif "imbattable" et formalisons l'énoncé "Jocelyne est imbattable". "Imbattable" contient une négation implicite, il s'agit de l'adjectif "battable" ou, plus précisément, du verbe d'action "x bat y", nié directement. Une première formalisation conforme à la logique classique pourrait être " $\forall x \neg B(x, \text{Joce})$ " ou encore, ce qui est classiquement équivalent, " $\neg \exists x B(x, \text{Joce})$ ". Mais, puisque la relation est asymétrique, le dernier énoncé est équivalent à " $\neg \exists x \neg B(\text{Joce}, x)$ ", qui à son tour est équivalent à " $\forall x B(\text{Joce}, x)$ " ; or, il n'y a plus de trace de négation ici. Évidemment, la logique classique ne permet pas de distinguer une négation interne d'une négation externe, puisqu'il est toujours possible de déplacer une négation.

La présence de deux négations, le supplément ' $\sim$ ' et le complément ' $\neg$ ', soulève plusieurs questions importantes : ces négations sont-elles des négations propositionnelles ou des négations de termes ? Comment sont-elles reliées entre elles ? Permettent-elles de définir de nouveaux opérateurs ?⁷ Les considérations sur les approximations suggèrent que le supplément affecte en quelque sorte la copule, soit l'égalité, alors que le complément semble agir sur la 'propriété' attribuée. Il reste à savoir s'il s'agit là d'un phénomène redevable à quelque idiosyncrasie ou s'il est possible de lui donner un traitement général approprié.

Il nous semble que, sur ce point, La Palme Reyes et al. 1994 ont ouvert la voie dans la bonne direction. Nous nous proposons d'abord d'examiner rapidement le contexte informel dans lequel leur analyse se déploie. Le point de départ est évidemment un concept qui peut être nié directement, comme le concept de bonheur ou, comme le concept d'honnêteté, choisi par La Palme Reyes et al. et remontant à Aristote. Dans le cas du bonheur comme dans celui de l'honnêteté, ce qui importe, c'est que les concepts puissent être 'partiellement satisfaits' : il est possible d'être partiellement heureux ou partiellement honnête, et ce, tout simplement parce que quelqu'un peut être heureux au travail, par exemple, mais malheureux dans sa vie familiale, etc. Un individu sera alors totalement heureux s'il est heureux dans tous les secteurs de sa vie. Les négations sont introduites informellement de la manière suivante : on dira d'un individu qu'il n'est pas heureux s'il n'est pas heureux sous un certain aspect ou, dans notre terminologie, s'il ne l'est pas totalement ; et l'on dira d'un individu qu'il est malheureux s'il n'est heureux dans aucun des aspects de sa vie.

Déjà à ce niveau informel, il est manifeste qu'un individu ne peut être heureux et malheureux à la fois, bien que l'affirmation selon laquelle un individu est heureux ou n'est pas heureux est, elle, toujours vraie. Cependant, l'affirmation selon laquelle un individu est heureux ou malheureux peut ne pas être totalement vraie et la proposition selon laquelle un individu est heureux et n'est pas heureux peut ne pas être totalement fausse. Voyons maintenant comment La Palme Reyes et al. ont proposé de formaliser ces considérations.

7. Reyes et Zolfaghari ont d'abord utilisé le supplément afin de définir les opérateurs modaux de nécessité et de possibilité. Ce n'est qu'un peu plus tard qu'ils ont considéré le supplément comme un opérateur de négation.

Le point de départ est une catégorie $\mathbf{C}$, catégorie qui représente la façon dont on peut s'approcher d'un état donné, soit par exemple être heureux. Dans ce contexte, il est suffisant que $\mathbf{C}$ soit simplement un ensemble partiellement ordonné avec un élément maximal, que nous dénoterons par G . Cet ensemble représente l'état global qui nous intéresse et les points sous G représentent le 'découpage' de cet état global en états partiels, comme par exemple le milieu de travail, le milieu familial, etc. La catégorie $\mathbf{C}$ représente également comment ces états sont reliés et comment certains états dépendent d'autres. Ainsi, pour être heureux au travail, il faut être satisfait de sa charge de travail, des relations avec ses pairs et avec ses supérieurs et ainsi de suite. Un morphisme $S \rightarrow T$ indique donc qu'un aspect S est un sous-aspect de T , c'est-à-dire que pour que T soit satisfait, il faut qu'au préalable S le soit. Par conséquent, pour être heureux au travail, il faut d'abord être heureux dans chacun de ces domaines et, pour être heureux tout court, il faut être heureux au travail. La catégorie $\mathbf{C}$ n'est donc ni plus ni moins que l'analyse des différentes conditions et manières d'être heureux, ou, plus généralement, d'être dans un état. Abstraitement, la catégorie $\mathbf{C}$ est une 'décomposition' d'un prédicat, de ses conditions de satisfaction. La logique d'un terme ou la forme logique d'un terme se trouve donc à ce niveau; cependant, et il est important de le noter, ce n'est pas ici que sont introduits les opérateurs logiques.

L'étape suivante consiste à lier cette analyse d'un terme à un contexte extensionnel. On fixe d'abord un ensemble X d'individus. Un prédicat H sur cet ensemble devient une famille $(H_A)_{A \in \mathbf{C}}$ de sous-ensembles de X satisfaisant la propriété fonctorielle suivante : si $x \in H_A$ et $A' \rightarrow A$, alors $x \in H_{A'}$. Formellement, le foncteur $H_{(-)}$ associe à chaque objet A un ensemble H_A et à chaque morphisme $f : A' \rightarrow A$ une fonction $H_f : H_A \rightarrow H_{A'}$. Cette condition signifie que si l'individu x a la propriété H lorsque cette dernière est considérée sous l'aspect A , alors l'individu x a également cette propriété sous tous les sous-aspects A' de A . Ainsi, pour le prédicat 'être heureux', il y aura un ensemble d'individus heureux au travail, que l'on peut dénoter par H_T : tous ces individus sont nécessairement heureux sous tous les sous-aspects S du travail T , et ainsi de suite pour tout les aspects, y compris être heureux tout court, c'est-à-dire la propriété globale d'être heureux. La condition de fonctorialité implique qu'un individu est globalement heureux si et seulement si il est heureux sous tous les aspects donnés.

Ce que nous venons de construire est en réalité un foncteur $H_{(-)} : \mathbf{C}^{\text{op}} \rightarrow \mathbf{Ens}$ de la catégorie $\mathbf{C}$ à la catégorie des ensembles. La logique classique interprète le langage donné dans les ensembles. Pour généraliser cette interprétation au contexte des foncteurs, il faut d'abord considérer une théorie des types. Ainsi, la logique est une logique d'ordre supérieur : chaque sorte du langage devient un ensemble d'individu X et chaque prédicat H sur cet ensemble est interprété de la manière indiquée précédemment. Soit H et J deux prédicats de X , c'est-à-dire deux familles $(H_A)_{A \in \mathbf{C}}$ et $(J_A)_{A \in \mathbf{C}}$, alors :

$$H \leq J \text{ si et seulement si } \forall A \in \mathbf{C}, \forall x \in X (x \in H_A \Rightarrow x \in J_A),$$

$$x \in (H \vee J)_A \text{ si et seulement si } (x \in H_A \vee x \in J_A),$$

$$x \in (H \wedge J)_A \text{ si et seulement si } (x \in H_A \wedge x \in J_A),$$

$$x \in \neg H_A \text{ si et seulement si } \forall A' \rightarrow A \in \mathbf{C}, x \notin H_{A'},$$

$x \in \sim H_A$ si et seulement si $\exists A \rightarrow A' \in C, x \notin H_{A'}$.

Ces définitions satisfont la condition de fonctorialité, en particulier dans la définition du supplément l'inversion de A et de A' nous garantit cette fonctorialité. Selon ces définitions, d'une part, un individu est malheureux si et seulement si il n'est heureux sous aucun aspect ; et, d'autre part, un individu n'est pas heureux si et seulement si il n'est pas heureux sous un aspect.

Il faut toutefois ajouter une autre condition pour assurer la validité de certains résultats. Cette condition exige que deux individus distincts possédant une certaine propriété H_A ne sont pas identifiés à un aspect ultérieur $H_{A'}$. Plus précisément,

$$(\forall f : A' \rightarrow A) (y \in H_A \wedge H_f(y) = x_{A'} \Rightarrow y = x_A).$$

Cette condition assure principalement que le supplément est strictement global, en ce sens que pour tout aspect A , $x \in \sim H_A$ si et seulement si $x \in \sim H_G$, où G est l'aspect global. Cette propriété du supplément entraîne automatiquement le corollaire suivant : lorsqu'il est évalué globalement, le supplément est booléen, c'est-à-dire que le prédicat $(H_G \wedge \sim H_G)$ est l'ensemble vide. Selon La Palme Reyes et al., c'est cette dernière propriété qui donne lieu à la croyance que la logique est naturellement booléenne.

Ces remarques préliminaires et semi-formelles devraient nous permettre de comprendre comment la sémantique s'élabore dans toute sa généralité. Le cadre général est celui de la sémantique des préfaisceaux de la catégorie $\mathbf{Ens}^{\mathbf{C}^{\text{op}}}$, puisqu'il a été démontré que ces catégories sont en fait des topos de bi-Heyting⁸ (voir Lawvere 1990 ou Reyes & Zolfaghari 1994). Rappelons qu'un préfaisceaux est un foncteur $X : \mathbf{C}^{\text{op}} \rightarrow \mathbf{Ens}$, de la duale de la catégorie $\mathbf{C}$ à la catégorie des ensembles. Ces foncteurs X sont les objets de la catégorie. Un morphisme entre deux préfaisceaux X et Y est une famille de fonctions $\langle \eta_A \rangle_{A \in \mathbf{C}}$, telle que pour tout $A' \rightarrow A$, le diagramme suivant

$$\begin{array}{ccc} X(A) & \xrightarrow{\eta_A} & Y(A) \\ \downarrow & & \downarrow \\ X(A') & \xrightarrow{\eta_{A'}} & Y(A') \end{array}$$

est commutatif. Dans le cas présent, on exige donc que la catégorie $\mathbf{C}$ soit un ensemble partiellement ordonné avec un élément maximal et que le topos des préfaisceaux satisfasse la condition supplémentaire suivante :

$$(\forall A' \rightarrow A) (y \in X(A) \wedge X(f)(y) = x_{A'} \Rightarrow y = x_A).$$

8. Remarquons que la sémantique des préfaisceaux est un cas particulier de la sémantique de Kripke-Joyal pour les faisceaux puisqu'un topos de préfaisceaux peut être vu comme un topos de faisceaux muni de la topologie discrète, parfois aussi appelée 'minimale' ou 'triviale'.

La sémantique des préfaisceaux est en réalité une généralisation de la sémantique de Kripke et elle se construit sur la base d'une relation de forcing. Étant donné les conditions additionnelles imposées à la catégorie $\mathbf{C}$ et au topos, la sémantique se simplifie légèrement. La relation de forcing est définie de la manière suivante :

$A \Vdash \varphi(\alpha)$ si et seulement si $\alpha \in \{x : \varphi(x)\}(A)$, où le dernier ensemble est l'ensemble des éléments qui satisfont la propriété au stade A , $\varphi(x)$ est une formule et x une variable de type X et $\alpha \in X(A)$.

La monotocité et le caractère local de la relation de forcing deviennent automatiquement, dans ce cas particulier (voir, par exemple, Mac Lane & Moerdijk 1992, 316 pour ces propriétés de la sémantique en général) :

Monotonocité : si $A \Vdash \varphi(\alpha)$ et $f : A' \rightarrow A$, alors $A' \Vdash \varphi(\alpha f)$.

Localité : pour tout A dans $\mathbf{C}$, $a : A \rightarrow G$, $A \Vdash \varphi(\alpha a)$ si et seulement si $G \Vdash \varphi(\alpha)$.

Nous nous contenterons maintenant de donner les conditions pour les deux notions de négation qui nous intéressent.

$A \Vdash \sim\varphi(\alpha)$ si et seulement si $\exists f : A \rightarrow A'$ tel que $A' \nVdash \varphi(\alpha)$;

$A \Vdash \neg\varphi(\alpha)$ si et seulement si $\forall f : A' \rightarrow A$ tel que $A' \nVdash \varphi(\alpha)$.

Quelques remarques s'imposent. D'abord, le lecteur sera probablement étonné de voir la définition du complément ' $\neg$ ' sous cette forme, puisque ce n'est pas ainsi qu'il est défini en général dans un topos de faisceaux. Toutefois, la définition donnée est une conséquence immédiate du fait que nous nous trouvons dans un topos de faisceaux possédant une topologie triviale. Il importe de noter également l'inversion des morphismes dans les définitions : dans le premier cas nous examinons un morphisme partant de A et allant vers A' , alors que, dans le second cas, nous examinons les morphismes qui vont vers A . Cette différence, bien qu'importante, cache en réalité une différence plus fondamentale : dans de telles conditions, le supplément ne dépend en effet que du niveau global de l'évaluation et non pas de l'évaluation à chaque stade. De plus, lorsqu'on les exprime de manière adéquate au niveau global, la loi du tiers exclu et de la non-contradiction sont valides pour le supplément.

Cette sémantique ouvre la porte à des questions intéressantes et qui demeurent en suspens. Ainsi, il semble important de clarifier davantage le fait que le complément ' $\neg$ ' capture par moments une négation 'interne' des termes alors que le supplément ' $\sim$ ' semble plutôt correspondre à une négation 'externe' des propositions qui ne peuvent être identifiées que lorsque la logique sous-jacente est booléenne. La Palme Reyes et al. ont également apporté quelques précisions à ce sujet. Cependant une logique des termes et ses liens avec les niveaux propositionnels méritent un examen plus poussé.

4. Conclusion

Il existe plusieurs types d'approximation, mais aucune étude exhaustive de cette question ne nous permet présentement de déterminer une taxinomie appropriée. Nous espérons avoir convaincu le lecteur qu'en ce qui a trait aux approximations numériques inhérentes à la démarche scientifique contemporaine, il est nécessaire de raffiner l'analyse logique traditionnelle et de distinguer au moins deux types distincts de négation. Nous espérons l'avoir par ailleurs convaincu qu'un cadre conceptuel et formel peut être présenté pour mieux comprendre le comportement de ces négations. Remarquons pour finir que, de façon surprenante, quelques uns des résultats ainsi obtenus permettent également de mieux comprendre certains phénomènes logiques des langues naturelles éliminés de l'analyse logique traditionnelle.

Bibliographie

- Bunge M., 1974, Interpretation and Truth, Vol. 2 of Treatise on Basic Philosophy, Dorrecht : Reidel.
- Bunge M., 1981, Half Truths, Essays in Scientific Philosophy, E. Morscher & O. Neumaier & Zecha, eds, Verlag, 87-91.
- Bunge M., 1983, Understanding the World, Vol. 5 of Treatise on Basic Philosophy, Dordrecht : Reidel.
- Bunge M., 1990, Marquis on Partial Truth, Studies on Mario Bunge's Treatise, P. Weingartner & G. Dorn, eds., Amsterdam : Rodopi.
- Cerruti U. & Höhle U., 1986, An approach to uncertainty using algebras over a monoidal closed category, Supp. Rend. Circ. Mat. Palermo, Ser. II, n° 12.
- Frege G., 1919, 1971, Recherches Logiques, Ecrits logiques et philosophiques, C. Imbert, trad., Paris : Éditions du Seuil.
- Goldblatt R., 1979, Topoi, North Holland.
- Higgs D., 1984, Injectivity in the topos of complete Heyting algebra valued sets, Canadian Journal of Mathematics, Vol. XXXVI, N° 3, 550-568.
- Höhle U., 1991, Monoidal Closed Categories, Weak Topoi and Generalized Logics, Fuzzy Sets and Systems, 42, 15-35.
- Höhle U. & Stout L., 1991, Foundations of Fuzzy Sets, Fuzzy Sets and Systems, 40, 257-296.
- Horn, L. R., 1989, A Natural History of Negation, Chicago : The University of Chicago Press.
- Kant I., Critique de la Raison Pure, A. J-L. Delamarre & F. Marty, trad., Bibliothèque de la Pléiade.
- Kant I., Logique, Vrin.
- La Palme Reyes M. & Macnamara J & Reyes G.E. & Zolfaghari H., 1994, The Non-Boolean Logic of Natural Language Negation, Philosophia Mathematica, (3), vol. 2, 45-68.
- Lawvere W., 1990, Intrinsic co-Heyting boundaries and the Leibniz rule in certain toposes, Category Theory Proceedings Como 1990, A. Carboni, M.C. Pedicchio, G. Rosolini, eds., LNM 1488, Springer Verlag, 279-281.
- Lowen R., 1978, On Fuzzy Complements, Information Sciences, 14, 107-113.
- Makkai M. & Reyes G., 1977, First Order Categorical Logic, LNM 611, Springer Verlag.

- Makkai M. & Reyes G., 1992, Completeness Results for Intuitionistic and Modal Logic in a Categorical Setting, Preprint.
- Mac Lane S. & Moerdijk I., 1992, Sheaves in Geometry and Logic : a first introduction to topos theory, Springer Verlag.
- Marquis J.-P., 1990, Partial Truths about Partial Truth, Studies on Mario Bunge's Treatise, P. Weingartner & G. Dorn, eds., Amsterdam : Rodopi, 61-78.
- Marquis J.-P., 1991, Approximations and Truth Spaces, Journal of Philosophical Logic, 20, 375-401.
- Marquis J.-P., 1992, Approximations and Logic, Notre Dame Journal of Formal Logic, 32, 2, 184-196.
- Mawanda M. M., 1989, On a categorical analysis of Zadeh Generalized subsets of sets I, Categorical Algebra and its Applications, F. Borceux, ed., LNM 1348, New York : Springer-Verlag.
- Ovchinnikov S.V., 1983, General Negations in Fuzzy Set theory, Journal of Mathematical Analysis and Applications, 92, 234-239.
- Ponasse D., 1988, Categorical Studies of Fuzzy Sets, Fuzzy Sets and Systems, 28, 235-244.
- Reyes G. & Zolfaghari H., 1994, Bi-Heyting Algebras, Topos and Modalities, à paraître dans Journal of Philosophical Logic.
- Ribenboim P., 1949, Characterization of the sup-complement in a distributive lattice with last element, Sum. Bras. Math. II, Fasc. 4, 43-49.
- Rodabaugh S. E. et al., 1992, Applications of Category Theory to Fuzzy Subsets, Dordrecht : Kluwer Academic.
- Stout L., 1991, A Survey of Fuzzy Sets and Topos Theory, Fuzzy Sets and Systems, 42, 2-14.
- Zolfaghari H., 1992, Toposes and Modalities, Thèse de Doctorat, Département de Mathématiques, Université de Montréal.

Jean-Pierre MARQUIS

Chapitre 16

Paradoxes sémantiques et programmation logique

Il est remarquable que la résolution des paradoxes sémantiques et l'étude des langages de programmation à base de règles (programmation logique), aient conduit les théoriciens à se poser des problèmes analogues et à proposer des solutions semblables, sans que la culture et les techniques acquises dans l'un des domaines servent pour l'autre. Dans cet article nous expliquons ce qui rapproche les deux domaines, nous détaillons l'analogie entre les idées proposées de part et d'autre et nous indiquons comment certaines méthodes et techniques pourraient migrer d'une communauté à l'autre. Notre conclusion est qu'en fait l'échec persistant de la recherche d'une solution définitive dans les deux domaines doit être interprété comme montrant qu'ils sont sujets à ce que nous appelons un schéma de sens wittgensteinien alors qu'on avait cru se trouver face à un schéma de sens gödelien.

1. Introduction

L'intelligence artificielle a été amenée à créer ses propres langages de programmation dont Prolog est le plus original et le plus mystérieux à la fois. Plus de 20 ans après son invention — doit-on dire sa découverte ? — on ne dispose pour lui d'aucune sémantique totalement satisfaisante, c'est-à-dire d'aucune théorie qui tout en respectant sa spécificité déclarative et l'usage qui en est fait par les programmeurs, décrive sans aucune ambiguïté son fonctionnement. Les très nombreux travaux qui lui ont été consacrés n'ont pas élucidé totalement l'association presque miraculeuse que Prolog réussit entre un noyau déclaratif (pratiquement unique dans les langages de programmation) et une puissance étonnante très appréciée pour le développement d'applications en intelligence artificielle. Cependant les nombreuses voies explorées

par les théoriciens de la programmation logique, laissent apparaître une analogie frappante avec celles explorées pour résoudre les paradoxes sémantiques en logique. Pourquoi ? Jusqu'où cette similitude se prolonge-t-elle ? La résolution des paradoxes qui a, elle aussi, une certaine importance dans les projets de l'intelligence artificielle peut-elle s'en trouver facilitée ? Que signifie cette similitude et n'est-elle pas l'indice d'un problème plus profond ?

1.1. *Quel est le problème posé par les paradoxes sémantiques ?*

Des phrases comme "*cette phrase n'est pas vraie*" ne peuvent se voir attribuer de valeur de vérité sans que cela entraîne une contradiction : c'est le fameux paradoxe du **menteur**. On en tire l'impossibilité de définir un prédicat de vérité dans tout système formel suffisamment riche pour coder sa propre syntaxe comme par exemple l'arithmétique de Péano ou la théorie des ensembles ZF. Le paradoxe du menteur a été l'objet de bien des réflexions depuis plus de deux mille ans de la part des philosophes et des logiciens, mais c'est seulement depuis une soixantaine d'années (avec Tarski) qu'on l'utilise pour *démontrer* effectivement que l'idée naturelle d'introduire un prédicat $Vr(.)$ dans un langage logique ne peut se faire simplement, en respectant les principes logiques identifiés et adoptés par ailleurs et formant ce qu'on appelle la logique classique. Notons aussi que les difficultés créées par les paradoxes sémantiques concernent aussi les logiques constructivistes.

Le problème posé est alors : comment modifier la logique classique ou comment formuler en son sein le problème des paradoxes sémantiques, pour réussir l'introduction d'un prédicat $Vr(.)$ conforme à l'essentiel de nos intuitions préthéoriques sur la vérité.

1.2. *Quel est le problème posé par les langages logiques à base de règles ?*

Les systèmes experts, les bases de données déductives, et le langage de programmation Prolog sont fondés sur la notion de règles logiques comme :

alain est le père de léon.

alain est le père de lucie.

léon est le père de louis.

lucie est la mère de paul.

si X est le père de Y et Y est le père de Z alors
X est le grand-père de Z.

si X est le père de Y et Y est la mère de Z alors
X est le grand-père de Z.

(L'usage, que nous adopterons à partir de maintenant, sera de noter $A \leftarrow B$ l'implication si B alors A.)

Ces règles s'écrivent et se comprennent *naturellement*, et pourtant on découvre vite que ce sous-langage de la logique qu'elles constituent fait naître de nombreuses difficultés. En particulier leur *sens naturel*, n'est pas seulement celui que la logique classique donne à de telles règles. Le *sens naturel* suppose dans un exemple comme celui donné au-dessus qu'ayant énoncé ces règles là, **ce qui n'en résultera pas sera faux**. On déduit donc (correctement) que *alain est le grand-père de louis et de paul*, mais on déduit (abusivement en regard de la logique classique) que *paul n'est pas le grand-père de alain* ou que *lucie n'est pas le père de paul*. En fait, l'attitude *naturelle* face à de telles données est de les *compléter* implicitement en fonction de principes — qu'il s'agit justement d'éclaircir — de manière à en tirer plus, que ce que la simple logique autorise.

Le problème posé par la programmation logique est donc celui de l'identification du sens qu'on doit attribuer à des énoncés formulés dans un langage à base de règles de façon à ce que ce sens soit à la fois conforme aux calculs de Prolog et à nos attentes *naturelles* (les deux choses ne sont pas forcément compatibles d'ailleurs, ce qui est une difficulté supplémentaire).

1.3. *Quel rapport y a-t-il entre ces deux problèmes ?*

Formulés ainsi, les rapports entre les deux problèmes semblent minces : certes on veut plus ou moins à chaque fois *adapter ou modifier la logique classique* ; certes à chaque fois on découvre vite que c'est la *négation* qui introduit les plus graves difficultés, mais si l'analogie en restait là on pourrait dire qu'entre les deux domaines considérés il n'y a pas plus de liens qu'entre eux-mêmes et la logique modale, ou l'intuitionnisme ! C'est en y regardant de plus près que l'analogie se renforce.

2. Le parallèle

2.1. *Dire le plus possible sans tomber sur l'inconsistance*

L'idée dans les deux domaines est à chaque fois, qu'il faut *définir le plus possible le sens des prédicats en cause* (le prédicat $\text{Vr}(\cdot)$ quand il s'agit des paradoxes sémantiques, et les prédicats utilisés dans les programmes dans le cas des langages logiques). L'idée est aussi que *l'inconsistance doit limiter ce « remplissage » du sens*.

Dans le cas des paradoxes sémantiques on sait l'inconsistance proche à cause du "menteur". Dans le cas des langages logiques on découvre aussi qu'elle menace. Par exemple : considérer comme conséquences implicites d'une base de règles, la négation de tous les littéraux (formules atomiques, ou négations de formules atomiques) qui ne sont pas conséquences du programme au sens de la logique usuelle (principe parfois utilisé sous le nom d'hypothèse du "monde clos" [Reiter 1978] et qui permet de traiter d'une manière satisfaisante le programme logique donné plus haut) n'est pas acceptable en général, car cela conduit à l'inconsistance dans certaines situations (à partir de "si A alors A" on est conduit à ajouter A et non A).

2.2. Sur-détermination et sous-détermination

Dans les deux domaines le but est donc : « dire le plus possible sans tomber sur l'inconsistance ». Formuler autrement : il faut régler simultanément un problème de *sur-détermination* et un problème de *sous-détermination*.

Du côté des paradoxes la contradiction que le raisonnement usuel permet d'obtenir à partir de la phrase du menteur c'est la sur-détermination, alors que l'absence de conclusion justifiée concernant l'autoaffirmation ("*cette phrase est vraie*") c'est la sous-détermination.

Dans le cas de la programmation logique on a une sous-détermination de la logique classique vis-à-vis de Prolog, pour le programme :

$$a \leftarrow \neg b$$

car à la question « est-ce que a ? » Prolog répond OUI alors que a n'est pas conséquence de l'implication représentée. La réponse de Prolog provient de ce que le but a appelle le but $\neg b$, qui appelle le but b . Le but b échoue et donc le principe de négation par l'échec (« ce qui ne peut pas être établi par le mécanisme de déduction Prolog est considéré comme faux » voir [Shepherdson 1984 1985]) entraîne Prolog à donner un succès pour $\neg b$, et donc un succès pour a .

Le cas de sur-détermination de la logique classique vis-à-vis de la programmation logique se produit pour le programme :

$$a \leftarrow b$$

$$a \leftarrow \neg b$$

$$b \leftarrow b$$

En effet pour la question « est-ce que a ? » Prolog tombe sur une boucle infinie et ne répond donc rien alors que a est pourtant une conséquence logique des trois implications constituant le programme.

2.3. Rechercher des principes généraux

Une idée évidente et présente dans les deux domaines est que la solution recherchée doit se fonder sur des principes élémentaires qu'on doit pouvoir énoncer simplement, et qui doivent, autant que possible, être opératoires, c'est-à-dire conduire à des systèmes formels ou à des algorithmes praticables. Notons tout de suite que l'aspect opératoire a été mis en avant avec plus d'insistance à propos des langages logiques.

Parmi les idées avancées pour formuler des principes conduisant aux extensions recherchées on peut mentionner :

- l'idée qu'il faut formuler des *équivalences* exprimant ce qu'on attend naturellement (nous avons une idée pré-théorique assez précise par exemple de ce que doit satisfaire le prédicat de vérité $Vr(.)$) ;

- l'idée qu'il faut accepter d'être *partiel* (logique partielle ou trivaluée) ne serait-ce que momentanément, ou pour les cas les plus généraux ;

- l'idée qu'on peut *désigner un modèle* particulier (partiel ou pas) d'un système d'axiomes qui en a plusieurs, et que pour cela, on peut s'intéresser soit à la minimalité des modèles, soit au contraire à leur maximalité, soit à leur compatibilité avec les autres modèles (modèle optimal ou intrinsèque) ;

- l'idée que tout ce qu'il n'est pas incohérent de *considérer vrai* (resp. *faux*) doit l'être ;

- l'idée qu'on peut *accepter tout ce qui n'est pas arbitraire*, ou tout ce qu'on retrouve toujours (supervaluation de Van Fraassen) ;

- l'idée qu'au fond ce qui est le plus simple est *itératif* et que tout peut donc se fonder sur des schémas inductifs, éventuellement transfinis.

Détaillons quelques-unes de ces idées (qui souvent ont des relations intéressantes les unes avec les autres).

2.4. Les principes généraux s'expriment naturellement avec des équivalences

Il se trouve que les principes qu'on énonce naturellement se présentent souvent sous forme d'équivalences. C'est une grande source d'analogies entre les méthodes utilisées dans les deux domaines.

A propos du prédicat de vérité on considère comme allant de soi que le prédicat de vérité $Vr(.)$ doit vérifier : $Vr(f) \leftrightarrow f$

Cette équivalence appelée 'convention T' ([Tarski 1933]) est souvent jugée comme un point central. Elle entraîne parfois (par exemple quand on l'utilise en logique partielle) une *non unicité*, ou une *non effectivité* de la solution, ce qui conduit à lui préférer des principes moins simples mais plus constructifs. On consultera [Kripke 1975], [Barwise Etchemendy 1987], [Feferman 1982], [Fitting 1985 1986], [Martin 1984] (recueil d'articles), [Tarski 1933], [Visser 1989] (survey).

A propos des prédicats d'un programme logique on considère naturel de supposer que ce qui est exprimé implicitement dans les règles, ce sont non pas des implications mais des équivalences. Dans le cas de notre exemple, ce qu'on veut dire pour le prédicat "est le grand-père de" c'est :

$$[X \text{ est le grand-père de } Z]$$

$$\leftrightarrow$$

$$\exists Y[X \text{ est le père de } Y \text{ et } Y \text{ est le père de } Z] \text{ ou}$$

$$\exists Y[X \text{ est le père de } Y \text{ et } Y \text{ est la mère de } Z]$$

C'est l'idée du programme complété de Clark [Clark 1978] qui a une grande importance en sémantique des langages logiques. On consultera [Fitting 1985], [Lloyd 1988] (livre de synthèse), [Reiter 1978], [van Emden Kowalski 1976].

Les théoriciens des paradoxes ont recherché à axiomatiser une équivalence faible (parfois à l'aide de la logique modale) leur donnant *ce qu'il faut mais pas plus* (pas de paradoxes). Leurs systèmes d'axiomes pourraient certainement être utiles en programmation logique. Voir [Aczel Feferman 1980] [Feferman 1982]. Inversement la logique linéaire a été utilisée pour Prolog et devrait donc pouvoir l'être pour les paradoxes sémantiques. Voir [Girard 1987] [Cerrito 1990].

2.5. Les logiques partielles (ou trivaluées) s'imposent

Dans les deux domaines l'idée qu'on doit produire une extension mais qu'on n'est pas sûr d'y arriver a conduit à l'utilisation de la logique partielle ([Blamey 1986] survey), c'est-à-dire d'une logique où une troisième valeur de vérité correspondant à *inconnu* ou *indéterminé* est acceptée. A propos des langages logiques à base de règles, la troisième valeur de vérité a parfois été utilisée pour représenter la non-terminaison des algorithmes d'évaluation des questions (en particulier SLDNF). Voir [Mycroft 1984] [Fitting 1985] [Lassez Maher 1985] [Delahaye 1987a b] [Kunen 1989] [Cavendon Lloyd 1989] [Mathieu Delahaye 1990] [Thibau Delahaye 1991a] [Dix 1991] [Delahaye 1992].

Sur ce point les théoriciens des paradoxes (en particulier [Kripke 1975]) ont étudié de plus près la logique trivaluée. Les schémas de supervaluation de van Fraassen ([van Fraassen 1966] [Burgess 1986]) n'ont pas été exploités, à ma connaissance, à propos des langages logiques alors qu'ils font partie de la *culture* des théoriciens des paradoxes (voir le survey [Visser 1989]). Sur les logiques partielles en général on pourra aussi consulter [Kleene 1952] [Rosser Turquette 1958] [Haack 1974 1978] [Turner 1984] [Urquhart 1986] [Blamey 1986].

La théorie de Barwise et Etchemendy [Barwise Etchemendy 1987] est peut-être l'ultime développement de l'idée de la logique partielle à propos des paradoxes sémantiques. La sémantique des langages logiques a peut-être quelque chose à tirer des idées proposées en sémantique des situations [Barwise Perry 1983].

Le dédoublement du nom des prédicats pour obtenir moins d'interaction entre A et non A (ce qui est souvent l'idée de fond motivant l'introduction de la logique trivaluée) a été utilisé dans les deux domaines ; dans [Gilmore 1974] [Feferman 1982] pour les paradoxes ; et dans [Delahaye 1988] [Bleuzen-Guernalec Gil 1989] [Denis 1990] [Denis Delahaye 1991] pour les langages logiques, mais n'a sans doute pas donné tout ce qu'on peut en attendre.

2.6. Les principes opératoires sont jugés importants

Parmi les principes opératoires qui ont donné lieu à des développements parfois assez proches dans les deux domaines, on peut distinguer :

- (a) L'idée de la hiérarchisation qui permet de traiter progressivement un problème.
- (b) L'idée de l'itération d'un opérateur d'extension qui permet de calculer ou d'approcher des points fixes (et donc des équivalences).

2.6.1. L'idée de la hiérarchisation

C'est une des premières idées pour les paradoxes, et c'est en gros l'idée de Tarski *langage/méta-langage/méta-méta-langage/ etc.* Mais cette idée est assez faible (elle équivaut simplement à dire que le *menteur* n'est pas une phrase syntaxiquement correcte) et c'est justement lorsqu'elle a été jugée insuffisante par Kripke [Kripke 1975] qu'on a pu aller plus loin.

C'est aussi une des premières idées en sémantique des langages logiques : les programmes hiérarchisés (sans appels récursifs) « se laissent bien faire » comme l'a très tôt établi Clark [Clark 1978] et cela même en présence de négations.

Mais dans le domaine de la sémantique des programmes logiques cette idée a été poussée plus loin avec la notion de *programme stratifié* ([Apt Blair Walker 1987]). Peut-être est-ce là une idée à explorer dans le domaine des paradoxes : certaines phrases autoréférentielles auxquelles sont associés des programmes stratifiés se voyant alors attribuer un sens naturel par la sémantique du modèle minimal supporté de Apt Blair Walker.

De même, en sémantique des langages logiques on a accepté l'idée d'une dissymétrie entre le *positif* et le *négatif* (en partant du principe que ce qui n'était pas mis explicitement devait plutôt être considéré *faux* que *vrai*). Les logiques des défauts, modèles parfaits etc. doivent avoir leurs contreparties du côté des paradoxes sémantiques et il y a certainement des idées à adapter. Il est possible qu'on trouve là une justification à certaines méthodes utilisées pour les paradoxes ([Herzberger 1982] [Gupta 1982] [Belnap 1982] [Visser 1989]) ce qui les *fonderaient* d'une manière plus satisfaisante.

2.6.2. L'idée de l'itération d'un opérateur d'extension permet de calculer ou d'approcher des points fixes et donc des équivalences

Cette idée est elle aussi très naturelle, c'est celle de Kripke [Kripke 1975], c'est celle de l'opérateur conséquence immédiate de van Emden, Kowalski et Apt [van Emden Kowalski 1976] [Apt van Emden 1982]. Son succès est d'autant plus grand que l'itération sur un opérateur bien choisi conduit à un point fixe et donc à une situation 'stable' c'est-à-dire satisfaisant une certaine équivalence. L'analogie entre les deux domaines est particulièrement manifeste :

- dans les deux domaines le même théorème de point fixe ([Knaster 1928] [Tarski 1955]) est utilisé, et dans les deux domaines les points fixes minimaux (qu'on peut approcher par itération transfinie) jouent un rôle privilégié,
- dans les deux domaines l'équivalence obtenue est *faible* (donne trop peu de conséquences) d'où le souci de la renforcer.

Concernant les points fixes, la notion de *point fixe optimal* [Kripke 1975] (qu'il appelle : *point fixe intrinsèque*) a été exploitée, mais ne l'a pas été vraiment à propos des langages logiques à l'exception de [Lassez Maher 1985] [Thibau Delahaye 1991b]. Dans le domaine de la sémantique des langages de programmation classique,

elle avait été découverte indépendamment de Kripke par Manna et Shamir [Manna Shamir 1976].

En sémantique des langages logiques certaines itérations complexes (à plusieurs niveaux, avec phases de complémentation etc. [Przymusiński 1989 1990 1991] [Van Gelder Ross Schlipf 1991]) ont été étudiées sans que leur équivalent le soit en théorie des paradoxes sémantiques (voir l'utilisation de cette idée en annexe).

A l'opposé les itérations de Herzberger [Herzberger 1982], Gupta [Gupta 1982] et Belnap [Belnap 1982] (voir [Visser 1989]) ont donné lieu à des résultats susceptibles d'intéresser la sémantique des programmes logiques [Węgrzynowski Delahaye 1992].

3. Conclusion

3.1. *Non-monotonie, remplissage maximum cohérent*

La non-monotonie des logiques introduites dans les deux domaines ([Brewka Makinson Schlechta 1991] [Dix 1991] [Makinson 1992]) est bien souvent une conséquence de la volonté d'ajuster la sémantique classique pour résoudre la sous-détermination et la sur-détermination dont nous parlions au début. Nous pensons que même si on peut voir là une raison supplémentaire de prendre au sérieux les recherches qui sont faites aujourd'hui sur les logiques non-monotones, ce n'est pas en elles qu'il faut rechercher les raisons profondes de l'analogie des deux domaines. La non-monotonie ici n'est que l'absence d'une propriété et n'éclaire pas vraiment l'analogie remarquée.

Par contre dans les deux situations il y a une recherche de *remplissage maximal cohérent* ("dire le plus possible sans tomber sur l'inconsistance") qu'on peut voir comme la recherche d'un sens simple et ignoré, que le but des travaux serait de reconnaître. Dans les deux cas qui nous occupent ici, il semble bien le sens attribué aux éléments du langage concerné (celui résultant de l'introduction de $Vr(.)$, celui des programmes logiques) est toujours ajusté "en cours de route", "par l'usage", "au fur et à mesure des découvertes formelles".

Pour les paradoxes sémantiques c'est très clair, il y a d'abord eu le menteur, puis le menteur renforcé, puis le menteur contingent, puis le paradoxe de Gupta, puis celui de Löb (que Visser appelle le paradoxe de Curry). A chaque fois les solutions des paradoxes précédents sont montrées insuffisantes pour les nouveaux paradoxes (créés "exprès"). Les solutions ne sont jamais pleines. Chaque nouvelle tentative approfondit le problème, ainsi d'ailleurs que le scepticisme de ceux qui ont du mal à prendre au sérieux tous ces jeux avec la prédicat $Vr(.)$! Les problèmes en réalité n'apparaissent pas très urgents et les solutions partielles dont on dispose depuis longtemps suffisent le plus souvent (par exemple en théorie des modèles). Les classifications des paradoxes n'apportent pas d'illumination particulière et les outils proposés soulèvent toujours autant de nouveaux problèmes qu'ils n'en résolvent d'anciens. Il semblerait bien que le sens de $Vr(.)$ soit indéterminé dès qu'on aborde les situations avec autoréférences et que ce soit l'usage fixé au coup par coup — et

parfois contradictoirement d'un auteur à l'autre — qui permettent d'aller plus loin (quoique jamais assez).

Dans le cas de la sémantique de Prolog la théorie du "sens déterminé par l'usage" peut être adaptée et rend assez bien compte de ce à quoi nous assistons. Le jeu est juste un plus compliqué. Au début il y a une sémantique provenant de la théorie de la résolution de Robinson («Prolog c'est la résolution utilisée avec des clauses de Horn»), puis pour rendre compte des réponses négatives Kowalski, Apt, van Emden et Clark introduisent le complété standard et les sémantiques par itérations de l'opérateur conséquence immédiate. Ceci se révèle insuffisant, car le Prolog traité est 'idéalisé' et n'est expliqué que partiellement. Les voies de recherche divergent alors, car vouloir tout prendre en compte apparaît hors de portée. Il y a ceux qui veulent rester fidèle à Prolog, ceux qui veulent traiter la négation mais acceptent d'idéaliser le langage (par exemple en ne retenant que les stratégies équitables, alors que Prolog n'en n'utilise pas), ceux qui veulent obtenir plus de sens encore que Prolog n'en donne (c'est l'approche généralement retenue en base de données où Prolog est moins sacralisé). Jamais l'adéquation n'arrive à se faire et on modifie en fait aussi bien la pureté des objectifs (rendre compte logiquement de tout Prolog), que l'objectif lui-même (on accepte de modifier Prolog). Tout cela en définitive ne converge pas, et l'espoir aujourd'hui d'une théorie raisonnablement simple et naturelle qui satisfasse les objectifs de départ s'éloigne de plus en plus.

Même si la théorie du sens par ajustement conventionnel et approximations successives rend bien compte de l'évolution des recherches, il est clair qu'elle manque l'essentiel, car ce qu'on veut c'est y échapper. Un succès véritable serait une théorie qui traite une fois pour toutes le problème sans que plus aucune modification ou nouvelle théorie ne soit nécessaire. Ce qu'on veut c'est une illumination qui fasse dire : «je ne le savais pas, mais c'est évident maintenant que vous me le montrez, vous avez trouvé la théorie de Prolog qu'on attendait». La recherche poursuivie et jamais aboutie a pour conséquence une profusion de demi-solutions. Cela aussi rend les deux domaines analogues et mérite une réflexion.

3.2. Les schémas de sens wittgensteinien et gödélien

C'est là sans doute le lieu d'introduire une distinction qui dans le cas qui nous concerne pourrait bien éclairer les situations et montrer en quoi, au-delà de l'aspect syntaxique et formel, il y a bien analogie profonde entre les paradoxes sémantiques et la programmation logique.

Lorsqu'on cherche à rendre compte d'une pratique de langage, deux schémas au moins sont possibles et qu'à défaut d'une meilleure idée nous nommerons *wittgensteinien* et *gödélien*. Dans le premier, la théorie qu'on cherche à élaborer pour rendre compte d'un sens, le normaliser et le généraliser rencontre une barrière qu'on reconnaît faite de conventions : dire le sens c'est fixer arbitrairement par des règles la relation avec le reste du langage. Au bout du compte c'est l'usage qui détermine le sens, et celui-ci n'est pas fixé par avance par le langage dont on dispose déjà. Ce premier schéma s'impose lorsqu'il s'agit d'expressions idiomatiques comme "se lever du mauvais pied" ou "faire long feu". Que l'usage puisse avoir une histoire ne change en rien, et d'ailleurs un retournement de sens (par rapport à celui que donne

une analyse purement historique et philologique) n'est jamais exclu. Si tout le monde se mettait à considérer que "faire long feu" signifie "durer longtemps" (comme on le trouve parfois, et on comprend bien pourquoi) ceux qui majoritaires aujourd'hui considèrent en se référant à l'image de la mèche que cela signifie "ne pas durer longtemps" auraient tort (puisque'ils ne seraient pas compris !).

En science et en particulier en mathématiques, les choses sont plus complexes et la *logique* d'une théorie peut déterminer le sens d'une expression qu'on a pu croire conventionnelle à un moment. C'est même sans doute le propre des mathématiques que cette homogénéisation soudaine du sens qui échappe alors brusquement à toute explication conventionnaliste. L'exemple caractéristique de cette imposition du sens par un extérieur (et qui explique le choix de l'adjectif *gödelien* retenu pour désigner ce second schéma) est celui des *énoncés vrais d'arithmétique*. L'idée qu'on les définit de manière conventionnelle par un système de preuves choisi arbitrairement a pu être soutenue un moment, mais ne peut absolument plus l'être aujourd'hui. Elle ne résiste pas à deux attaques : celle des théorèmes d'incomplétude de Gödel et celle de la définition au second ordre de la notion d'*énoncé vrai d'arithmétique*. Pour un mathématicien la notion d'énoncé vrai d'arithmétique est parfaitement définie, même si c'est d'une manière non effective (c'est ce qu'établissent entre autres choses les théorèmes d'incomplétude). Les axiomes de Péano n'ont plus rien à voir là-dedans, on peut les ignorer, le sens n'est pas réductible à une convention, aussi complexe et déguisée soit-elle. Lorsque c'est le schéma gödelien qui s'applique pour déterminer une situation linguistique (formelle ou non) tout peut être pensé en une fois, et ce sens mis à jour est définitivement le bon. La notion de *fonction calculable* a subi elle aussi un renversement de situation : c'était une notion qu'on croyait *wittgensteinienne*, et on a soudain découvert qu'elle possédait une définition absolue que fixe la *thèse de Church*. Les notions de *figures dessinables à la règle et au compas*, d'*atomes*, de *formules vraies du calcul des prédicats du premier ordre*, constituent autant d'autres exemples de situations où on croyait avoir affaire à un schéma *wittgensteinien* et où un "progrès" de compréhension — une homogénéisation du langage — a fait voir qu'en réalité il y avait un schéma *gödelien* : la notion ne pouvait recevoir qu'un seul sens qui n'est en rien arbitraire. Tous les théorèmes de représentation en mathématiques et de complétude en logique ont pour effet de faire passer une partie de langage, d'un sens *wittgensteinien* à un sens *gödelien*. Tout n'est pas possible, c'est ce que montrent sans cesse les mathématiques et plus généralement les sciences, où le progrès est sans doute bien souvent de fixer les significations dans une nécessité insoupçonnée l'instant précédent.

A propos de l'analogie qui nous occupe je voudrais défendre l'idée que dans les deux cas mis en parallèle ici, on a cru être face à un schéma *gödelien* de sens alors qu'en fait c'était un schéma *wittgensteinien*. Une analyse non formulée donnait à croire que le sens des éléments de langage en question était *gödelien*, alors qu'en réalité il était *wittgensteinien*. La leçon à tirer de tout cela serait que non seulement le *progrès rationnel* fait passer certaines situations linguistiques d'un schéma *wittgensteinien* à un schéma *gödelien*, mais qu'inversement le progrès rationnel nous conduit à renoncer à attribuer un sens *gödelien* à des concepts qui pourtant nous apparaissaient de prime abord devoir s'y conformer. Il faut accepter parfois de revenir sur ses intuitions préthéoriques. Dans un tel cas se bloquer sur l'idée préconçue serait

alors une erreur inverse — mais aussi grossière — que celle consistant aujourd'hui à nier la thèse de Church.

Que la vérité même quand il s'agit de la vérité du langage de l'arithmétique avec $Vr(.)$ ne se laisse jamais définir que par conventions est-ce finalement si important ? Les échecs répétés pendant des siècles pour faire mieux doivent nous apparaître comme la preuve indirecte que contrairement à nos préjugés préthéoriques il n'y a pas de définition (même ineffective) qui englobe une fois pour toutes ce qu'il en est de ce prédicat. Donner un sens au *menteur* est un jeu, ce jeu se complique dès qu'on croit avoir une solution et finalement se perd en une recherche désespérée. Remarquons aussi qu'il est faux que nos intuitions et notre usage courant indiquent quoi que ce soit de définitif du *menteur*, encore moins du *menteur contingent* ou du *paradoxe de Löb*. Il était légitime de rechercher (et anecdotiquement il le sera encore) une sémantique conforme au schéma *gödélien* pour $Vr(.)$. Le fait qu'on ait échoué aujourd'hui désigne avec une presque certitude que nous sommes face à un schéma de sens *wittgensteinien*.

Dans le cas de Prolog comme nous le disions tout est un peu plus compliqué mais revient finalement au même. L'existence d'un noyau propre a donné la conviction que la situation était *gödélienne*, mais l'usage aujourd'hui, et plus encore les implémentations de Prolog, montrent qu'il est un mélange un peu confus de logique, de calcul relationnel et d'astuces procédurales. Ce n'est pas la première fois qu'un langage informatique se révèle "impossible à nettoyer". Ne faisons pas une fixation sur Prolog. Prolog est *wittgensteinien*, même si auprès de lui il y a un peu de sens *gödélien* (par exemple lorsqu'on se limite à des programmes sans négation).

Ma conclusion est donc qu'on a une analogie évidente entre les deux domaines, due principalement à une formalisation homomorphe, qu'il y a eu un approfondissement exacerbé dans les deux cas menant à une exploration tout azimut, débordante et quelque peu ridicule. L'inaboutissement de ces recherches montrent qu'on est face à un schéma de sens *wittgensteinien*, et rend vaine la poursuite effrénée d'une solution *gödélienne* qui n'existe pas; ce qui réunit le mieux les deux situations est l'illusion qu'elles ont fait naître et la quantité de travail qu'elles ont suscitée. Puisque le succès de l'une serait aussi le succès de l'autre, leur échec commun doit renforcer l'idée qu'il s'agit d'un échec définitif. Il serait sage d'admettre que dans les deux cas, l'idéal d'une solution *gödélienne* est aujourd'hui devenu improbable.

Annexe

Détails sur les possibilités de traduire complètement le problème des paradoxes sémantiques dans le formalisme des programmes logiques

Pour fixer les idées supposons que nous cherchions à ajouter à l'arithmétique de Péano un prédicat $Vr(.)$ pour représenter le prédicat de vérité.

Le vocabulaire du langage est composé de $\forall, \vee, \neg$, zéro, $s, +, \times, =, Vr(.)$, Var (ensemble des noms de variables). Nous appelons **For** ce langage.

On se donne le modèle standard de l'arithmétique ($\omega, 0, \text{succ}(\cdot), +(\cdot, \cdot), \times(\cdot, \cdot)$) et on tente de le prolonger en définissant l'interprétation de $\text{Vr}(\cdot)$. Pour simplifier la présentation nous identifions $0=\text{zéro}$, $1=\text{s(zéro)}$, $2=\text{s(s(zéro))}$, etc.

Nous avons besoin d'un codage de la syntaxe dans la théorie même. Pour l'obtenir nous fixons une bijection de l'ensemble des formules closes sur ω , ainsi qu'une bijection de l'ensemble des formules à une variable libre sur ω . Ces bijections sont notées : $P \rightarrow 'P'$; $P(x) \rightarrow 'P(x)'$. On dit que ' P ' est le numéro de P .

Lemme. Si : $\text{sub} : \omega \times \omega \rightarrow \omega$; $'P(x)'$, $n \rightarrow 'P(n)'$ est récursive, alors pour toute formule à une variable libre $P(x)$ il existe un entier n tel que :

$P(n)$ est équivalente (i.e. prend la même valeur de vérité) dans toute interprétation partielle de **For** prolongeant $(\omega, 0, \text{succ}(\cdot), +(\cdot, \cdot), \times(\cdot, \cdot))$ à une formule P° dont le numéro est n : $P('P^\circ') \equiv P^\circ$

Pour $P(x) = \text{Vr}(x)$ on peut ainsi obtenir une formule P° de numéro n équivalente à $\text{Vr}(n)$. $\text{Vr}(n)$ affirme donc sa propre vérité. Par exemple on peut imaginer que l'entier n est 2005 et qu'à $\text{Vr}(2005)$ est associée la formule équivalente $P^\circ = (\exists m \ m = ((2 \times 10) \times (10 \times 10) + 5) \wedge \text{Vr}(m))$ dont le numéro est 2005. La formule numéro 2005 affirme donc sa propre vérité.

Démonstration.

Soit $P(x)$ une formule à une variable libre construite avec le langage :

$$\forall, \vee, \neg, 0, s, +, \times, =, \text{Vr}(\cdot), \text{Var}.$$

L'hypothèse faite sur la fonction *sub* permet de montrer que la formule $\text{sub}(n, m) = p$ signifiant :

« la substitution de la variable par l'entier m dans la formule
à une variable libre numéro n ,
donne la formule (close) numéro p : $\text{sub}'(P(x)', m) = 'P(m)'$ »

est exprimable dans l'arithmétique et ne fait pas intervenir le prédicat $\text{Vr}(\cdot)$.

En choisissant bien les numérotations on pourrait par exemple s'arranger pour que $\text{sub}(n, m) = p$ soit la formule exprimant $2^n \times 3^m = p$.

On pose alors :

$$Q(x) = \{\exists y \ P(y) \wedge (y = \text{sub}(x, x))\}.$$

La formule $Q(x)$ possède une variable libre, soit m son numéro ($m = 'Q(x)'$). On pose : $P^\circ = Q(m)$: P° est donc la formule $Q(x)$ dans laquelle x a été remplacé par l'entier m , P° est une formule close.

Montrons que pour toute interprétation partielle i de $\text{Vr}(\cdot)$ on a $i(P^\circ) = i(P('P^\circ'))$. L'interprétation i est la donnée de deux ensembles disjoints d'entiers correspondant aux numéros des formules vraies et des formules fausses.

Une fois donné un tel i chaque formule du langage construit avec :

$$\forall, \vee, \neg, 0, s, +, \times, =, \text{Vr}(\cdot), \text{Var}$$

s'évalue en considérant les connecteurs forts de Kleene ou le schéma de supervaluation de Van Fraassen, ou n'importe quel autre schéma d'évaluation satisfaisant

- (a) $i(\exists y P(y) \wedge (y = n)) = i(P(n))$ et

- (b) prolongeant le schéma standard d'évaluation pour les formules de l'arithmétique n'utilisant pas $Vr(.)$.

Bien sûr toute formule close ne faisant pas intervenir le prédicat $Vr(.)$ s'évalue *Vrai* ou *Faux* indépendamment de i .

On a :

$$\begin{aligned} i(P^\circ) &= i(Q(m)) = i(\exists y P(y) \wedge (y = \text{sub}(m, m))) \\ &= i(\exists y P(y) \wedge (y = \text{sub}('Q(x)', m))) = i((\exists y P(y) \wedge (y = 'Q(m)))) \\ &= i(P('Q(m)')) = i(P('P^\circ')). \end{aligned}$$

Ceci parce que lorsqu'on évalue $y = \text{sub}('Q(x)', m)$ le résultat ne dépend pas de l'interprétation i de $Vr(.)$. Fin de la démonstration.

Il n'est pas exact en général qu'on puisse trouver un entier n tel que le numéro de la formule $Vr(n)$ est n . Cependant d'après le lemme moyennant une modification finie de la numérotation des formules, on peut supposer qu'un tel entier existe. Nous le supposons et donc : il existe une formule **aa** (pour 'autoaffirmation') ayant la propriété que :

$$\mathbf{aa} = Vr('aa').$$

De même on peut supposer sans perte de généralité (d'après le lemme) qu'il existe un entier n tel que n est le numéro de la formule $\neg Vr(n)$: la formule de numéro n affirme donc sa propre fausseté. Nous désignerons cette formule par **me** (pour 'menteur') :

$$\mathbf{me} = \neg Vr('me').$$

Pour que $Vr(.)$ représente ce qu'on en attend on souhaiterait pour toute formule avoir :

$$Vr('f') \leftrightarrow f$$

Ceci est bien sûr impossible en logique classique (c'est-à-dire avec une interprétation i définie par deux parties complémentaires de ω) car alors $\mathbf{me} \leftrightarrow Vr('me')$ et donc par définition de $\mathbf{me}$:

$$\neg Vr('me') \leftrightarrow Vr('me')$$

Nous venons simplement de redécouvrir que le paradoxe du menteur interdit toute extension simple de l'arithmétique : le prédicat $Vr(.)$ sera partiel.

Si pour toute formule f on écrit :

$$Vr('f') \leftarrow f$$

on obtient un ensemble infini de règles logiques qui se rapprochent déjà du formalisme des programmes logiques. Le seul problème étant que f peut être

n'importe quelle formule. Moyennant la mise sous forme normale disjonctive infinie (le domaine d'interprétation est fixé une fois pour toutes et que donc $\forall x P(x)$ signifie simplement $P(0)$ et $P(1)$ et) nous pouvons transformer cette règle en un nombre fini ou infini de règles de la forme :

$$Vr('f') \leftarrow \bigwedge \{ Vr(n_i) \mid n_i \in I \} \wedge \bigwedge \{ \neg Vr(n_j) \mid n_j \in J \} \quad I, J \text{ ensembles d'entiers}$$

ce qui cette fois est une règle (propositionnelle) comme on les considère dans les langages à base de règles, mis à part qu'elle a un nombre éventuellement infini de prémisses.

[A la place de parler de "formes normales disjonctives infinies" on peut dire que les ensembles I et J sont tous les ensembles minimaux I et J disjoints tels que l'interprétation partielle i qu'ils définissent donne $i(f) = \text{Vrai}$]

Nous sommes alors en mesure d'appliquer toutes les méthodes utilisées dans le domaine de la programmation logique et des bases de données déductives.

Le programme obtenu est-il hiérarchisé ? Non car il contient la règle :

$$Vr('aa') \leftarrow Vr('aa')$$

Le programme obtenu est-il stratifié ? Non car il contient la règle :

$$Vr('me') \leftarrow \neg Vr('me')$$

Il est amusant de constater que ce sont justement les phrases autoréférentielles qui rendent ces programmes difficiles à traiter par les méthodes usuelles de la sémantique des langages logiques. On sait que les programmes hiérarchisés ont toutes sortes de très bonnes sémantiques et que beaucoup d'idées en apparence différentes coïncident quand on les leur applique. On sait aussi qu'à un degré moindre il en va de même pour les programmes stratifiés.

Considérer le programme complété Clark revient à ajouter

$$\neg Vr('f') \leftarrow \neg f$$

et à mettre sous forme normale comme précédemment.

Parmi les possibilités offertes il y a :

La sémantique du *modèle faible bien fondé* (de Fitting) ou plus petit point fixe de l'opérateur conséquence immédiate trivalué de Fitting [Fitting 1985] qui revient en fait à rechercher le plus petit modèle trivalué du programme complété. Il s'agit là d'une idée qui a été proposée dans le domaine des paradoxes sémantiques par Kripke [Kripke 1975].

La sémantique du modèle optimal [Thibau Delahaye 1991b]. C'est une idée qui a déjà été considérée par Kripke à propos des paradoxes.

Par contre les sémantiques des défauts, et surtout la sémantique du modèle bien fondé [van Gelder Ross Schlipf 1988] ne semblent jamais avoir été appliquées.

Détaillons la sémantique du modèle bien fondé. On pose :

$$\mathcal{T}_I(T) = \{A \mid \exists r : A \leftarrow L_1 \wedge \dots \wedge L_n \quad \forall i \quad L_i \text{ vrai dans } I \text{ ou } L_i \in T\}$$

$$\mathcal{F}_I(F) = \{A \mid \forall r : A \leftarrow L_1 \wedge \dots \wedge L_n \quad \exists i \quad L_i \text{ faux dans } I \text{ ou } L_i \in F\}$$

$$\mathcal{T}_I^{\uparrow 0} = \emptyset \quad \mathcal{F}_I^{\downarrow 0} = H_P \quad \mathcal{T}_I^{\uparrow n+1} = \mathcal{T}_I(\mathcal{T}_I^{\uparrow n}) \quad \mathcal{F}_I^{\downarrow n+1} = \mathcal{F}_I(\mathcal{F}_I^{\downarrow n})$$

$$\mathcal{T}_I^\infty = \bigcup \{\mathcal{T}_I^{\uparrow n} \mid n \in \omega\} \quad \mathcal{F}_I^\infty = \bigcap \{\mathcal{F}_I^{\downarrow n} \mid n \in \omega\}$$

$$\mathcal{I}(I) = I \cup \langle \mathcal{T}_I^\infty; \mathcal{F}_I^\infty \rangle$$

$$M_0 = \langle \emptyset; \emptyset \rangle \quad M_{\alpha+1} = \mathcal{I}(M_\alpha) \quad M_\alpha = \bigcup \{M_\beta \mid \beta < \alpha\}$$

Le menteur $m : m \leftarrow \neg m$. On trouve que m est indéterminé pour la sémantique du modèle bien fondé (on ne peut s'appuyer ni sur le vrai ni sur le faux)

L'autoaffirmation $a : a \leftarrow a$. On trouve que a est faux pour la sémantique du modèle bien fondé.

Les cycles de menteurs $a \leftarrow b, b \leftarrow c, c \leftarrow \neg a$. Tout est indéterminé.

Les cycles d'affirmations $a \leftarrow b, b \leftarrow c, c \leftarrow a$. Tout est faux.

Les cycles ajustés $a \leftarrow \neg b, b \leftarrow \neg a$. Tout est indéterminé.

Programmes stratifiés : $a \leftarrow a, b \leftarrow \neg a, b \leftarrow b, c \leftarrow \neg b$. a faux, b vrai, c faux.

On peut voir cette sémantique comme exprimant l'idée qu'on veut déterminer le plus possible les valeurs de vérité, mais en s'arrangeant pour que le résultat puisse au moins s'appuyer sur lui-même pour se justifier et en donnant une préférence au faux sur le vrai. On note que face aux problèmes d'autoréférences et de paradoxes cette sémantique est assez peu déterminante.

Bibliographie

- [Aczel Feferman 1980] P. Aczel, S. Feferman. *Consistency of the unrestricted abstraction principle using an intensional equivalence operator*. In "To H. B. Curry : Essays on Combinatory Logic, Lambda Calculus and Formalism." J. P. Seldin and J. R. Hindley (editor). Academic Press. 1980. p. 67-98.
- [Apt Blair Walker 1987] K. Apt, H. A. Blair, A. Walker. *Toward a theory of declarative knowledge*. in "Foundation of Deductive Databases". (Minker Ed.). Morgan Kaufmann, Los Altos, 1987.
- [Apt van Emden 1982] K. Apt, M. van Emden. *Contribution to the theory of logic programming*. Journal of the Association for Computing Machinery, 29.3. 1982. p. 841-862.
- [Barwise Etchemendy 1987] J. Barwise and J. Etchemendy. *The Liar. An Essay on Truth and Circularity*. Oxford University Press, New-York Oxford. 1987.

- [Barwise Perry 1983] J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press. Bradford Books, Cambridge Mass. 1983.
- [Belnap 1982] N. D. Belnap Jr. *Gupta's rule of revision theory of truth*. Journal of Philosophical Logic. vol. 11. 1982. p. 103-116.
- [Blamey 1986] S. Blamey. *Partial Logic*. Handbook of Philosophical Logic. Volume III : Alternatives in Classical Logic. Edited by D. Gabbay and F. Guenther. D. Reidel Publ. Company. 1986. p. 1-71.
- [Bleuzen-Guernalec Gil 1989] N. Bleuzen-Guernalec, B. Gil. *A Semantical Approach of an Internal Rendering of the Negation in General Logic Programs*. Université d'Aix-Marseille II, 1989.
- [Brewka Makinson Schlechta 1991] G. Brewka, D. Makinson, K. Schlechta. *Cumulative Inference Relations for JTMS and Logic Programming*. First International Workshop on Nonmonotonic and Inductive Logic. Lecture Notes in Artificial Intelligence. Springer-Verlag LNCS 543. 1991. p. 1-12
- [Burgess 1986] J. P. Burgess. *The Truth is Never Simple*. The Journal of Symbolic Logic. V. 51, n° 3. 1986. pp. 663-681. *Addendum* V. 53, n° 2. 1988. p. 390-392.
- [Cavendon Lloyd 1989] L. Cavendon, J. W. Lloyd. *A completeness theorem for SLDNF resolution*. Journal of Logic Programming. 7. 1989. p. 177-191.
- [Cerrito 1990] S. Cerrito. *A linear semantics for allowed logic programs*. IEEE Symposium on logic in Computer Science, 1990.
- [Clark 1978] K. L. Clark. *Negation as failure*. In "Logic and Database". H. Gallaire J. Minker Ed. Plenum Press, New York, 1978. p. 293-324.
- [Delahaye 1987a] J.-P. Delahaye. *Programmation en Logique Trivaluée*. Rapport de Recherche du Laboratoire d'Informatique Fondamentale de Lille. Bat M3. 59655. Villeneuve d'Ascq. 1987.
- [Delahaye 1987b] J.-P. Delahaye. *Chaînage avant et calcul de modèles booléens et trivalués*. Les Systèmes Experts et leurs Applications, Septièmes Journées Internationales sur les Systèmes Experts et leurs Applications, Avignon, 1987.
- [Delahaye 1988] J.-P. Delahaye. *Sémantiques logiques et dénotationnelles des interpréteurs Prolog*. Theoretical Informatics and Applications. 22. 1988. p. 3-42.
- [Delahaye 1992] J.-P. Delahaye. *Chaînage avant et déduction logique*. Pour La Science. (Edition Française du Scientific American.). Février 1992. p. 104-109, 114
- [Denis Delahaye 1991] J.-P. Delahaye, F. Denis. *Is there an Axiomatic Semantics for Standard Pure Prolog ?* Theoretical Computer Science, 1991. p.373-388
- [Dix 1991] J. Dix. *Cumulativity and rationality in semantics of normal Logic Programs*. First International Workshop on Nonmonotonic and Inductive Logic. Lecture Notes in Artificial Intelligence. Springer-Verlag LNCS 543. 1991. p. 13-37.
- [Feferman 1982] S. Feferman. *Toward Useful Type-Free Theories*. Journal of Symbolic Logic. 1982. Aussi dans [Martin 1984] p. 237-287.
- [Fitting 1985] M. Fitting. *A Kripke-Kleene Semantics for Logic Programs*. Journal of Logic Programming. 2 (4). 1985. p. 295-312.
- [Fitting 1986] M. Fitting. *Notes on the Mathematical Aspects of Kripke's Theory of Truth*. Notre Dame Journal of Formal Logic. Vol. 27. January. 1986. p. 75-88.
- [Gilmore 1974] P. C. Gilmore. *The consistency of partial set theory without extensionality*. Axiomatic Set Theory, Proc. Symposia Pure Maths. XIII, Part II, Amer. Math. Soc. 147-153.

- [Girard 1987] J.-Y. Girard. *Linear Logic*. Theoretical Computer Science. 50. 1987. p.1-102
- [Gupta 1982] A. Gupta. *Truth and Paradox*. Journal of Philosophical Logic. 11. 1982. p. 1-60.
- [Haack 1974] S. Haack. *Deviant Logic : Some Philosophical Issues*. Cambridge University Press. 1974.
- [Haack 1978] S. Haack. *Philosophy of Logic*. Cambridge University Press 1978.
- [Herzberger 1982] H. G. Herzberger. *Notes on Naive Semantics*. Journal of Philosophy of Logic. 11. 1982. p. 61-102. Repris dans [Martin 1984].
- [Kleene 1952] S. C. Kleene. *Introduction to Metamathematics*. North-Holland, Amsterdam. 1952.
- [Knaster 1928] B. Knaster. *Un théorème sur les fonctions d'ensembles*. Annales de la Société Polonaise de Mathématiques. Vol. 6. 1928. p. 133-134.
- [Kripke 1975] S. Kripke. *Outline of a Theory of Truth*. The Journal of Philosophy. Vol 72. 1975. p. 690-716. Repris dans [Martin 1984].
- [Kunen 1987] K. Kunen. *Negation in Logic Programming*. Journal of Logic Programming. 1987. p. 289-308.
- [Kunen 1989] K. Kunen. *Signed data dependencies in Logic Programs*. Journal of Logic Programming. 7. 1989. p. 231-245.
- [Lassez Maher 1985] J.L. Lassez, M. Maher. *Optimal fixedpoint of logic programs*. TCS. 39, 1985, p.15-25.
- [Lloyd 1988] J. W. Lloyd. *Fondements de la programmation logique*. Editions Eyrolles. 1988. Traduction (V. Thibau) de : Foundations of Logic Programming. Springer-Verlag. 1987.
- [Makinson 1992] D. Makinson. *General patterns in nonmonotonic reasoning*. In D. Gabbay and others eds. Handbook of Logic in Artificial Intelligence and Logic Programming. Vol II Nonmonotonic and Uncertain Reasoning. Oxford University Press. 1992
- [Manna Shamir 1976] Z. Manna, A. Shamir. *The theoretical aspects to the optimal fixed point*. S.I.A.M. J. Comput. 1976. p. 414-426.
- [Martin 1984] R. L. Martin. Ed. *Recent Essays on Truth and the Liar Paradox*. Oxford University Press. New York. 1984.
- [Mathieu Delahaye 1990] J.-P. Delahaye, P. Mathieu. For which Bases Forward Chaining is Sufficient ? COGNITIVA-1990. Madrid. 20-23 novembre 1990. p. 699-702.
- [Mycroft 1984] A. Mycroft. *Logic Program and Many-valued Logic*. STACS 1984. (M. Fontet, K. Melhorn ed.) Lecture Notes in Computer Science 166. 1984. p. 74-286.
- [Przymusinski 1989] T. Przymusinski. *Non-monotonic Formalisms and Logic Programming*. Inter. Conference on Logic Programming, Lisbonne. 1989. p. 655-674.
- [Przymusinski 1990] T. Przymusinski. *Well-founded semantics coincides with three-valued stable semantics*. Fundamenta Informatica XIII. 1990. p. 445-563.
- [Przymusinski 1991] T. Przymusinski. *Well-founded completions of logic programs*. Intern. Conference in Logic Programming. 1991. p. 726-741.

- [Reiter 1978] R. Reiter. *On closed world databases*. Logic and Databases. H. Gallaire Ed. Plenum Press, New-York. 1978. p. 55-76.
- [Reiter 1980] R. Reiter. *A logic for default reasoning*. Artificial Intelligence, 13, 1980. p.81-132.
- [Rosser Turquette 1958] J. B. Rosser, A. R. Turquette. *Many-Valued Logic*. North-Holland. 1958.
- [Shepherdson 1984 1985] J. C. Shepherdson. *Negation as Failure I et II*. Journal of Logic Programming. 1984. n°1, pp. 51-79 et 1985. n°3 p. 185-202.
- [Tarski 1933] A. Tarski. *The Concept of Truth in Formalized Languages*. in "Logic, Semantics, Metamathematics". Oxford : Clarendon Press. 1956. p. 155-277. Traduction de "Pojecieprawdy w językach nauk dedukcyjnych". Prace Towarzystwa Naukowego, Wydział III matematyczno-fizycznych. N° 34. Warsaw 1933.
- [Tarski 1955] A. Tarski. *A lattice-theoretical fixpoint theorem and its applications*. Pacific Journal of Mathematics. Vol. 5. 1955. p. 285-309.
- [Thibau Delahaye 1991a] J.-P. Delahaye, V. Thibau. Programming in three-valued logic. Theoretical Computer Science, V. 78. 1991, p.189-216.
- [Thibau Delahaye 1991b] J.-P. Delahaye, V. Thibau. *The optimal model of a logic program*. European Workshop on Logic in Artificial Intelligence. Springer-Verlag, LNCS 478, 1991. p. 180-195.
- [Turner 1984] R. Turner. *Logic for Artificial Intelligence*. Ellis Horwood limited. Londres. 1984.
- [Urquhart 1986] A. Urquhart. *Many-valued logic*. In the "Handbook of Philosophical Logic" (D. Gabbay and F. Guenther eds.). Vol. 3. Reidel Publishing Company. 1986. p. 71-116.
- [Van Gelder Ross Schlipf 1991] A. van Gelder, K. Ross, J. Schlipf. *The well-founded Semantics for General Logic Programs*. Journal of the As. Computing Machinery. Vol. 38. 3, 1991. p. 620-650.
- [van Emden Kowalski 1976] M. H. van Emden, R. A. Kowalski. *The Semantics of predicates logic as a programming language*. J.A.C.M. 1976. 23-4. p. 733-742.
- [van Fraassen 1966] B. C. van Fraassen. *Singular Terms, Truth-value Gaps, and Free Logic*. The Journal of Philosophy. LXIII, 17, 1966. p. 481-495.
- [Visser 1989] A. Visser. *Semantics of the Liar Paradox*. in "D. Gabbay and F. Guenther (eds.) Handbook of Philosophical Logic. Vol. IV". D. Reidel Publishing Company. 1989. p. 617-706.
- [Wegrzynowski Delahaye 1992] E. Wegrzynowski, J.-P. Delahaye. *The grand loop semantics for logic programs* First Compulog Net Meeting on Knowledge Base Systems. Munich, 1992.

Ce travail a été représenté au Colloque Intelligence Artificielle et Logiques non Classiques, à Montréal les 17-20 octobre 1992.

Jean-Paul DELAHAYE

Conclusion

En guise de conclusion, les auteurs de ce livre aimeraient attirer l'attention du lecteur sur quelques préceptes qui leur paraissent de mise dans le domaine couvert par l'ouvrage :

- recherche des isomorphismes : la disparité des objets manifestes ne compte pas
- attention aux détails : l'essentiel est de grain fin
- sens de la variété : les activités cognitives sont diverses
- rigueur informelle : l'analyse conceptuelle importe
- mémoire : la « grande tradition » l'emporte souvent sur l'actualité du jour
- audace : rien n'est intangible
- bon sens : peu d'agents sont des archanges
- modestie : peu de choses ont encore été faites
- ouverture : rien de ce qui a été fait ne l'a été par une seule Ecole

CET OUVRAGE A ÉTÉ COMPOSÉ
PAR LES ÉDITIONS HERMÈS
REPRODUIT ET ACHEVÉ D'IMPRIMER
PAR L'IMPRIMERIE FLOCH À MAYENNE
EN MARS 1995.

DÉPÔT LÉGAL : AVRIL 1995.
N° D'IMPRIMEUR : 37482.

Langue - Raisonnement - Calcul

Cet ouvrage présente l'état de l'art dans le domaine des méthodes logiques pour les sciences cognitives. Les auteurs se sont fixé un triple objectif :

- être complets : loin de se cantonner aux problèmes traditionnels de l'intelligence artificielle, ils abordent l'étude du raisonnement taxinomique, la description formelle des systèmes de croyances dans une société d'agents et la théorie des actes de langage.
 - être actuels : plusieurs chapitres constituent la première exposition systématique en langue française de méthodes qui n'ont pas dix ans d'âge, comme la théorie des structures de connaissance ou la planification déductive.
 - « mettre les idées aux commandes » : une très large place est donnée non seulement à la justification informelle des méthodes présentées, mais aussi à une réflexion critique et philosophique sur les problèmes posés et la manière traditionnelle de les aborder.
- Ces trois caractéristiques en font un ouvrage destiné au public des sciences cognitives au sens le plus large : linguistes, informaticiens et spécialistes d'intelligence artificielle, mais aussi philosophes, psychologues et chercheurs en sciences sociales.

Jacques Dubucs est directeur de recherches au CNRS, Institut d'histoire et philosophie des sciences et des techniques, université de Paris I ; François Lepage est professeur de philosophie à l'université de Montréal.

Editions HERMES
14, rue Lantiez
75017 Paris



9 782866 014612

