UNIVERSITE AIX-MARSEILLE I - Université de Provence
U.F.R. Philosophie.

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE AIX-MARSEILLE I

Formation doctorale : Philosophie

Présentée et soutenue publiquement
par

Xavier PARENT

le 2 Decembre 2002

LOGIQUES NON-MONOTONES ET MODES D'ARGUMENTATION

Directeur de thèse :

Pierre LIVET

JURY

Gabriella CROCCO	Maître de Conférences, Université d'Aix-Marseille I
Andreas HERZIG ‡	Chargé de Recherches, CNRS/IRIT, Toulouse
Paul GOCHET * ‡	Professeur, Université de Liège
Pierre LIVET	Professeur, Université d'Aix-Marseille I
Alain MICHEL	Professeur, Université d'Aix-Marseille I

* Président du Jury ‡ Rapporteur

Introduction

Ce doctorat a pour titre très général « logiques non-monotones et modes d'argumentation ». Il constitue une tentative d'élargissement de recherches formelles entamées en cours de DEA. Ce dernier portait sur les logiques modales et, plus précisément, sur les logiques déontiques, dans lesquelles les modalités du nécessaire et du possible désignent respectivement l'obligatoire et le permis[1]. Nous y présentions un système déontique trivalent, permettant d'échapper à certains des paradoxes auxquels achoppent les systèmes traditionnels de logique déontique. Ces paradoxes soulèvent la question du rapport entre le sens des connecteurs en logique classique et leur sens dans la langue naturelle. Aussi nous sembla-t-il opportun d'essayer, à l'occasion de ces trois années de doctorat, d'élargir notre optique initiale de travail, en prenant pour fil directeur la question du rapport entre logique et argumentation. A partir des années 1950, la logique a dû faire place à un corps de théories groupées sous la désignation générale (et un peu ambiguë puisqu'elle réunit des entreprises visant un but parfois assez différent) de logique non-classique. Nous avons choisi d'axer la réflexion sur le cas des logiques non-monotones. Car, modélisant une notion de nécessité contextuelle et révisable, ces formalismes paraissent offrir un outil bien adapté à l'analyse de nos argumentations quotidiennes. Celles-ci ne s'inscrivent-elles pas dans un environnement en constante évolution ? Naturellement, il nous fallut chercher à mieux nous familiariser avec les théories existantes de l'argumentation. Au fil de nos lectures, nous nous sommes rendu compte que, lorsqu'un domaine de recherche comme celui de l'argumentation n'est pas encore véritablement constitué, il n'est guère possible d'en traiter systématiquement dans un exposé d'ensemble. Les recherches en ce domaine donnent une impression d'éparpillement, qui paraît difficilement surmontable.

Présentons brièvement le contenu de la dissertation, qui sera développé dans les chapitres suivants. Depuis quelques décennies, l'intérêt des logiciens de la non-monotonie pour le domaine de l'argumentation s'est considérablement accru. Nous verrons que les programmes de recherches partent généralement dans trois directions distinctes. Les uns étudient des schémas d'argumentation spécifiques, tels que le raisonnement causal ou le raisonnement mettant en jeu des obligations. D'autres s'intéressent à l'interface entre la sémantique (dont on dit traditionnellement qu'elle s'attache aux relations des expressions du langage à leurs références) et la pragmatique (étude empirique des langues naturelles et de leurs conditions d'usage). D'autres enfin cherchent à élaborer une théorie générale de l'interaction entre arguments. Néanmoins, les travaux qui jusque-là ont été menés restent difficilement accessibles. D'où, tout d'abord, notre souhait de fournir au lecteur une vue d'ensemble de ce domaine de recherches. Sur les trois chapitres que comporte ce doctorat, les deux premiers sont introductifs. Ils constituent un essai de

[1] Les ouvrages de Gardies [85, 86] constituent une excellente introduction à ces logiques.

panorama de l'ensemble des deux disciplines concernées, et tentent de donner le bagage nécessaire à la compréhension des articles spécialisés. Décrivons à grands traits le contenu de ces deux premiers chapitres, purement introductifs.

Le chapitre 1 porte sur les théories contemporaines de l'argumentation. Eu égard au problème du rapport entre logique et argumentation, quatre voire cinq attitudes sont possibles. Les différentes théories que, dans ce premier chapitre, nous abordons adoptent toujours l'une d'entre elles. Tout d'abord, on peut soutenir la position extrême selon laquelle nos argumentations quotidiennes ne relèvent pas d'une démarche logique. Cela revient à nier que la logique ait un quelconque droit de regard sur l'argumentation (position 1a). Ensuite, on peut défendre l'idée selon laquelle nos argumentations relèvent bien d'une telle démarche. A dire vrai, cette position en recouvre plusieurs autres. En effet, la théorie que Frege et Russell nous ont léguée et qui porte le nom de « logique classique » fut initialement conçue pour les besoins de (et dans l'espoir de fonder) la pensée mathématique. Tels Grice, certains diront que la logique classique suffit amplement pour décrire le fonctionnement de nos argumentations quotidiennes, à condition que la pragmatique vienne néanmoins l'« épauler » dans cette tâche (position 2). D'autres vont remettre en cause les prétentions de la logique classique, mais avec un degré de radicalité qui varie. L'attitude la plus radicale consiste à dire que les concepts fondamentaux de la logique traditionnelle doivent être abandonnés. Par exemple, au début de son ouvrage consacré aux usages de l'argumentation, Toulmin nous invite à « nous désemcombrer l'esprit des idées dérivées de la théorie logique existante » [253, p. 8]. Cette position (nommons-là 1b) rejoint finalement la position 1a : on refuse à la logique tout droit de regard sur l'argumentation. Cela donne naissance à une théorie de l'argumentation qui se développe en dehors de la logique, qu'elle soit classique ou non classique. C'est l'attitude d'Anscombre/Ducrot. Une autre approche possible consiste à penser que les concepts et les méthodes issus de la logique classique restent un outil d'analyse privilégié, qu'il convient tout au plus d'affiner, lorsqu'on passe de l'étude du raisonnement mathématique à l'étude du raisonnement naturel. Telle est, par exemple, la perspective que Vanderveken adopte. Ceci lui permet de donner à sa théorie des actes de langage une présentation toute formelle, dans la tradition des logiques modales. Il faut ici à nouveau distinguer deux attitudes. Les uns diront que, une fois affiné, ledit outil est le seul à pouvoir donner un fondement solide à nos argumentations quotidiennes (position 3). C'est l'attitude de Vanderveken. Ce qu'était la logique classique à la théorie des ensembles, la logique non classique le devient à l'égard de l'argumentation. D'autres vont chercher à utiliser des logiques non classiques, sans prétendre qu'elles puissent constituer autre chose que des instruments partiels, des repères pour analyser l'argumentation (position 4). Cela va de pair avec une conception de l'argumentation comme se développant partiellement en dehors de la logique, en l'occurrence non classique. Ce sera notre approche au chapitre 3. L'attitude 1a ne nous paraît pas très fructueuse. L'attitude 1b a un danger manifeste : nous risquons de tomber dans l'obscurantisme. La position 2 nous semble une position de facilité : dès qu'une difficulté apparaît, on invoque des considérations d'ordre pragmatique qui l'expliquent. La position 3 comporte elle aussi un danger : à vouloir produire un formalisme qui rende compte de tous les aspects d'une argumentation, nous risquons de ne plus proposer qu'un ersatz d'analyse pour chacun d'eux.

Le chapitre 2 porte sur les logiques pour le raisonnement en situation d'incertitude qui ont été développées dans le cadre de l'Intelligence Artificielle à partir des années 1980. On distingue généralement deux grandes familles de logiques : la classe des logiques non-monotones, la classe des théories de la révision. Les secondes mettent au premier plan la notion de changement, celui-là même qui nous est nécessaire pour mettre fin à une situation d'incertitude conçue comme indétermination ou indécision. Les premières ont pour objet l'étude formelle des énoncés comportant des règles générales sujettes à exceptions, du type : si ϕ alors normalement ψ. Généralement, on distingue dans celles-ci deux grandes variétés. La logique des défauts de Reiter et la logique autoépistémique de Moore constituent la première variété ; ce sont des approches de type point-fixe. La théorie de la circonscription de Mc Carthy et la théorie des modèles préférentiels de Gabbay, Makinson, Lehmann et Magidor, constituent la seconde variété ; ce sont des approches dites à base de minimisation. Nous étudierons plus en détail ces différents formalismes. Au terme de notre étude, l'idée se confirmera que, malgré leur apparente technicité, les théories de la non-monotonie et de la révision sont bien porteuses de projets se rapportant à l'argumentation. Par exemple, nous avons dit que certains chercheurs en IA se sont récemment intéressés à l'interface entre deux disciplines généralement considérées comme indépendantes l'une de l'autre : la sémantique et la pragmatique. Cet intérêt est né de l'observation selon laquelle certains éléments essentiels de la communication sont « défaisables », c'est-à-dire révisables. Cette particularité avait déjà été soulignée par Grice, l'un des fondateurs de la pragmatique. Celui-ci faisait de la défaisabilité une propriété définitoire des implicatures – règles permettant d'inférer A de B moyennant certaines maximes de rationalité et les intentions des locuteurs dans le contexte de la conversation. Cette observation de Grice donna naissance à un groupe de programmes, qui visent à faire de l'inférence pragmatique un cas particulier d'inférence non-monotone. Nous pouvons citer ici l'idée de Mercer de développer une analyse non-monotone des phénomènes présuppositionnels, ou encore la théorie des implicatures de Gazdar.

Ces deux premiers chapitres ne prétendent à aucune originalité. Ils n'ont d'autre ambition que de faire un rapide tour d'horizon de la littérature existante, dans le souci de s'informer. Le chapitre 3 a une facture bien à lui. Il porte sur le concept d'obligation, autour duquel nos recherches se sont toujours organisées.

Plus précisément, ce dernier chapitre porte sur l'analyse de la notion d'obligation conditionnelle en termes de modèles préférentiels. A l'origine, ce type de compte rendu fut proposé par B. Hansson et D. Lewis, en réaction à celui de Kripke. Leur idée était, en substance, de remplacer la relation d'accessibilité entre mondes, que Kripke utilisait, par une relation de préférence. Notre étude comprend deux mouvements. Dans un premier temps, reprenant des travaux dont nous avons eu l'occasion de faire état dans [189, 190], nous présentons les principes d'une sémantique de l'obligation conditionnelle, dont la particularité est d'introduire du séquentiel. Ce genre de compte rendu tire sa motivation essentielle dans la constatation, tout à fait surprenante, selon laquelle une sémantique standard de l'obligation autorise l'inférence d'un *doit* à partir d'un *est*, en laquelle chacun dénoncera une violation de la loi de Hume. La logique formelle perd ici de son crédit. Dans un deuxième temps, nous axons notre enquête sur le thème des interactions conversationnelles, et nous tentons de voir en quel sens lesdites sémantiques de l'obligation (et la logique déontique en général) peuvent contribuer à l'étude de ce thème. Nous essayons, pour l'essentiel, de donner un éclairage nouveau à certaines des difficultés logiques auxquelles nous nous heurtons, lorsque nous cherchons à rendre compte d'un échange dit réparateur. Un examen attentif révèle que la notion de révision (ou, si l'on veut, celle de changement de normalité) joue ici un rôle, de même que la dimension du temps. Qui souhaiterait acquérir immédiatement une vue d'ensemble de ce doctorat peut directement se reporter à la section 3.5. Nous y présentons nos conclusions, ainsi que quelques pistes de recherche.

Chapitre 1

Théories de l’argumentation

1.1 Introduction

Les théories contemporaines de l'argumentation se répartissent en deux catégories, selon la nature de la réponse qu'elles apportent à la question :

> La logique que Russell et Frege nous ont léguée peut-elle prétendre pouvoir rendre compte de l'argumentation ou, tout au moins, de certains de ses modes ?

Résolument non-formelles, les théories de la première catégorie répondent par la négative. Les théories de la seconde catégorie répondent par l'affirmative mais soutiennent que, pour y parvenir, la logique moderne doit remanier ses concepts fondamentaux.

Le but du présent chapitre est de présenter, au moins dans ses grandes lignes, certaines des figures centrales de ces deux catégories d'approches. Nous nous intéresserons tout d'abord aux approches non-formelles de l'argumentation. Il en est au moins trois qui peuvent difficilement être ignorées ; nous les envisageons tour à tour. Il s'agit de :

1. Le modèle de Toulmin ;
2. La théorie de la conversation de Grice ;
3. La théorie d'Anscombre et Ducrot.

Nous aborderons ensuite les approches formelles de l'argumentation. Trois d'entre elles retiendront notre attention :

1. Le programme de Hamblin ;
2. La logique naturelle de Grize ;
3. La logique de l'illocutoire de Vanderveken.

Nous devons insister sur le fait que, ne dénonçant pas les mêmes insuffisances de la logique mathématique, les auteurs sus-mentionnés élaborent toujours leur théorie indépendamment les uns des autres. Ces théories paraissent suffisamment complexes et lourdes, pour que nous nous permettions de les considérer une à une.

Il faut aussi préciser que, dans ce chapitre comme dans le suivant, notre propos est seulement de fixer le cadre général à l'intérieur duquel, au chapitre 3, notre réflexion s'inscrira. Dans ces deux premiers chapitres, nous décrivons seulement les grandes lignes de différents systèmes, plus précisément ce que nous en avons compris.

Le lecteur s'étonnera sans doute de ne pas voir figurer la nouvelle rhétorique de Perelman et Olbrechts-Tyteca dans notre typologie. C'est à contrecoeur que nous ne l'avons pas intégrée à cet exposé d'ensemble. La longue classification des types de schémas argumentatifs, qui est présentée dans la troisième partie du *Traité de l'argumentation*, constitue la charpente de la théorie de Perelman et Olbrechts-Tyteca. Les catégories fondamentales d'arguments qu'ils distinguent nous semblent assez floues. Ils distinguent tout d'abord les arguments qui procèdent par association de notions. Ceux-ci se répartissent eux-mêmes en trois classes : la classe des arguments quasi-logiques qui sont, comme leur nom l'indique, construits sur le modèle du raisonnement logique ou mathématique ; la classe des arguments qui sont basés sur la structure du réel et qui s'appuient soit sur une liaison de succession (comme dans le rapport de cause à effet) soit sur une liaison de coexistence (comme dans l'argument d'autorité) ; la classe des arguments qui fondent la structure du réel (appel au cas particulier et raisonnement par analogie). Aux formes d'argumentation qui procèdent par association de notions, Perelman et Olbrechts-Tyteca opposent celles qui procèdent au contraire par dissociation de notions. Elles consistent la plupart du temps à dissocier des notions en couples hiérarchisés comme apparence/réalité, lettre/esprit ou contingence/nécessité – ceci pour résoudre un problème d'incompatibilité. Hormis pour cette dernière forme d'argumentation, les schémas auxquels Perelman et Olbrechts-Tyteca s'intéressent restent de nature très classiques.

1.2 Les théories non-formelles

1.2.1 Le modèle de Toulmin

L'ouvrage de S. Toulmin sur les usages de l'argumentation est paru en 1958, c'est-à-dire la même année que le *Traité de l'argumentation* de Perelman et Olbrechts-Tyteca. Dans cet ouvrage, Toulmin refuse l'idée que la logique formelle puisse être un outil adéquat pour analyser nos argumentations quotidiennes. Essayons de comprendre les raisons sur lesquelles il se fonde.

Toulmin [253, p. 18] commence par introduire la notion de champ argumentatif (*argumentative field*). Deux arguments sont dits appartenir au même champ lorsque les données et la conclusion constituant chacun des deux arguments relèvent du même domaine – la géométrie, l'éloge esthétique, etc. Dans quelle mesure existe-t-il des critères communs pouvant servir à juger des arguments relevant de champs différents ? Toulmin répond à cette première question par analogie avec la théorie du droit. La procédure suivant laquelle sont réglées les questions soulevées à un tribunal peut, en général, être décomposée en trois phases : exposé de ladite requête ou du chef d'accusation ; présentation des preuves ou des témoignages à son appui ; prononciation du verdict. Considérons telle ou telle argumentation apportée à l'appui d'une conclusion donnée. Ses étapes sont identiques. Tout d'abord, nous posons le problème. Ensuite, nous examinons les solutions possibles, puis nous nous prononçons. De cette constatation, Toulmin [253,

p. 52] en tire une première objection contre l'utilisation de la logique formelle dans l'analyse de nos raisonnements quotidiens : telle que nous la connaissons, la logique formelle n'envisage pas l'argumentation dans son aspect procédural[1].

Telle n'est pas la seule objection que Toulmin adresse à la logique formelle. Non seulement la vision globale (ou, si l'on veut, la macro-analyse) qu'elle offre de l'argumentation est réductrice, mais son analyse des micro-arguments l'est aussi. La critique de Toulmin porte essentiellement sur l'agencement traditionnel des propositions dans le raisonnement en « prémisse majeure, prémisse mineure, donc conclusion ». Curieusement, l'analyse que la théorie logique donne de la structure interne des propositions est mise à l'arrière-plan. Toulmin propose le modèle suivant, qui est maintenant bien connu. Il peut être schématisé comme suit :

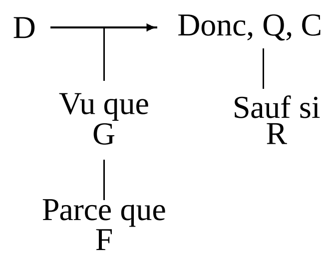


FIG. 1.1 – Modèle de Toulmin

C désigne la conclusion de l'argument et D les données (*data*) que nous invoquons à l'appui de C. Par exemple, pour appuyer l'affirmation (C) que Harry est sujet britannique, nous pouvons faire appel à la donnée (D) selon laquelle il est né aux Bermudes. G désigne ce que Toulmin nomme la garantie (*warrant*) . Dans notre exemple, elle peut s'exprimer sous la forme : « celui qui naît aux Bermudes doit généralement être sujet britannique ». Laissée la plupart du temps implicite par l'argumentateur, la garantie est donc l'énoncé général, de forme hypothétique, « Si D alors C », qui légitime le passage des données à la conclusion. Toulmin introduit un quatrième élément, le fondement F (*backing*) de la garantie − énoncé factuel, catégorique, sans lequel la garantie ne possèderait ni autorité ni créance. Dans l'exemple donné par Toulmin, F désigne l'ensemble des lois régissant la nationalité des personnes nées dans les colonies britanniques. Il reste maintenant deux éléments à expliquer brièvement. Tout d'abord, il reste à expliquer la lettre Q.

[1] Certains chercheurs en Intelligence Artifi cielle ont depuis lors tenté de dépasser cette limitation. Ainsi de Gordon et Karacapilidis [94] ou de Bench-Capon [28]. Ils construisent un système rendant justice à la conception procédurale de l'argumentation ici défendue par Toulmin. Les premiers utilisent un système de maintien de la cohérence de type Truth-Maintenance System (Doyle [58]).

Celle-ci représente la modalité à affecter à la conclusion, par exemple la modalité du « vraisemblable ». Une telle modalité est utilisée, s'il existe des circonstances dans lesquelles le passage de D à C n'est plus valide : les parents de Harry sont étrangers, Harry s'est fait naturaliser américain, ... Toulmin appelle ces circonstances les « conditions de réfutation » (*rebuttal*) de l'argument. Dans le schéma, elles sont représentées par la lettre R.Cela évoque le problème que Hempel [108] soulève dans le domaine de la logique inductive et que Davidson [55] applique au raisonnement pratique. La modalité du « vraisemblable » porte-t-elle sur la conclusion C, ou bien doit-elle plutôt être rattachée au « donc », duquel C ne serait pas détachable ? Toulmin n'aborde pas explicitement cette question. Toutefois, l'une des questions qu'il adresse à la théorie des probabilités de Kneale semble indiquer qu'il opterait plutôt pour la première solution. Kneale définit la probabilité comme une relation entre la proposition avancée avec prudence et les raisons qui nous la font émettre. Ainsi, la probabilité devient relative aux données qui la soutiennent. N'est-ce pas, demande Toulmin [253, p. 89], la voie ouverte au relativisme ?

1.2.2 Garantie, fondement

La terminologie employée par Toulmin n'est pas sans évoquer les notions utilisées dans la logique des défauts de Reiter. Il y manque toutefois la distinction entre garantie et fondement, que Toulmin juge essentielle. Nous avons besoin de cette distinction, pour répondre aux deux questions suivantes :

1. Les critères de justifications nécessaires pour défendre une conclusion varient d'un champ à l'autre. Cette absence d'uniformité est-elle effectivement irréductible ?
2. En quel sens et dans quelle mesure peut-on appliquer l'idée de validité formelle à nos (micro-)arguments ?

Considérons la première question. Il convient, selon Toulmin, de distinguer le cas où la justification est une garantie et celui où elle sert de fondement à une garantie. Dans le premier cas, la réponse est positive : un énoncé servant de garantie peut, quel que soit le champ de l'argumentation, être réduit à un énoncé de la forme :

> « On peut supposer à coup sûr (ou presque certainement)
> qu'un *A* est (ou n'est pas) un *B* ».

Mais, dans le second cas, la réponse est négative. Car la nature du fondement d'une garantie dépend évidemment du champ considéré. *F* peut correspondre à un relevé statistique (« la proportion des *A* qui sont des *B* est égale à ... »), renvoyer à un système taxinomique (« un *A* peut être classé comme un *B* »), exprimer une disposition légale (« un *A* est, aux yeux de la loi, un *B* »), etc. La forme des énoncés servant de fondement aux garanties d'inférence est ici inévitablement plurielle, parce que leur contenu renvoie à des procédures de vérification irréductibles les unes aux autres. Passons à la seconde question : peut-on appliquer l'idée de validité formelle à nos (micro-)arguments ? La réponse est positive, lorsque l'argument est de la forme « Donnée ; garantie ; donc conclusion », comme dans l'exemple suivant :

Donnée	Peterson est suédois
Garantie	On peut supposer presque certainement qu'un Suédois n'est pas catholique
Conclusion	Donc il est pratiquement certain que Peterson n'est pas catholique.

Mais, selon Toulmin, la réponse est négative, lorsque l'argument est de la forme « Donnée ; fondement ; donc conclusion » et que l'on remplace la seconde prémisse par :

Fondement	La proportion des Suédois catholiques est inférieure à 2 %.

Comme le remarque Castañeda [44, p. 285], Toulmin prend la notion de validité formelle en un sens relativement approximatif, dont on peut douter qu'il corresponde au sens que lui donnent ses contemporains logiciens. Toulmin dit d'un argument qu'il est formellement valide, si « la conclusion peut s'obtenir par un simple réarrangement des constituants des prémisses » [252, p. 146]. Pour un logicien, « être formellement valide » est généralement synonyme de « rester tel pour toute substitution faite sur les termes non-logiques que l'énoncé contient ». Il n'est pas sûr, fait remarquer Castañeda [44, p. 285], que, recourant au critère informel de Toulmin, on puisse toujours rendre compte de la validité des lois que, dans la pratique, la théorie logique retient. L'exemple du principe « p donc $p \vee q$ » semble ici significatif. Notons au passage qu'il paraîtrait étrange de mettre en parallèle le critère de Toulmin et la propriété gentzénienne de la sous-formule. Cette propriété stipule que, si le séquent $\Gamma \vdash \Delta$ est démontrable, il a une démonstration n'utilisant que des séquents formés de sous-formules des formules de $\Gamma \vdash \Delta$. Par exemple, pour démontrer le séquent $p \vdash p \vee q$, on part de $p \vdash p$, et on applique la règle d'introduction de la disjonction. Cette démonstration vérifie la propriété de la sous-formule. Dira-t-on de la disjonction $p \vee q$ qu'elle « réarrange » tout au plus la donnée p ?

De plus, nous pouvons nous demander si le point de vue de Toulmin n'est pas discutable, au regard des récents développements de la logique formelle. Si la nature du fondement apporté (c'est-à-dire le contenu de F) dépend du champ considéré, en règle générale le type de fondement à apporter ne change pas à l'intérieur d'un même champ et, dans bien des cas, ce dernier s'est révélé se prêter à un traitement formel. Reportons-nous aux quelques exemples de fondement que Toulmin [252, p.128] donne, lorsqu'il discute de la notion de validité formelle :

1. La baleine est (c'est-à-dire peut être classifiée comme) un mammifère ;
2. Le bermudien est (aux yeux de la loi) sujet britannique ;
3. Le Saoudien est (on le constate) un musulman.

Le premier énoncé exprime une loi taxinomique, le second une loi juridique et le troisième un relevé statistique relatif à la distribution des croyances religieuses. Tout d'abord, un énoncé taxinomique peut être exprimé dans les termes d'un réseau d'héritage de propriétés (tel que celui développé par Touretzky, Horty, Thomasson [249]) ou à l'aide d'une logique linéaire (comme le proposent Fouqueré et Vauzeilles [74]). Ensuite, il existe des logiques de la promulgation des normes (voir e.g. Alchourrón et Bulygin [4]). Enfin, un relevé statistique peut être exprimé à l'aide d'un quantificateur généralisé de la forme « la plupart des ... », pour lequel des axiomatiques et des sémantiques ont été proposées (Schelchta [224]). Considérons − pour en rester au seul exemple du relevé statistique − un argument de la forme :

Donnée	L'objet a vérifie la propriété P
Fondement	La plupart des P sont des Q
Conclusion	Nous sommes pratiquement certains que a vérifie Q.

Toulmin parle ici de quasi-syllogisme. En voici la traduction dans le formalisme que Schlechta propose :

$$T = \{P(a), \nabla x P(x) : Q(x)\} \models_{\prec} Q(a)$$

Liée par le quantificateur ∇ (*nabla*), la variable x ne parcourt plus qu'un sous-ensemble « très grand » des individus qui sont de type P. Du point de vue sémantique, on exige seulement des « très grands » sous-ensembles d'un ensemble donné qu'ils aient quasiment la structure d'un filtre. Le symbole « $\models_{\prec}$ » est lu « a pour conséquence plausible ». Cette notion ne s'apparente pas à une inférence classique. Car elle est définie par quantification universelle sur un sous-ensemble des modèles de T, à savoir l'ensemble de ceux qui sont minimaux pour la relation de préférence $\prec$. Intuitivement, ce sont les modèles qui réduisent au minimum les exceptions aux défauts que T contient. En gros, un modèle m est préféré à un modèle m', $m \prec m'$, s'il y a au moins un défaut $d \in T$ pour lequel m est « meilleur » que m', au sens où d a plus d'instances vraies dans m que dans m'. Dans notre exemple, les modèles de T qui sont minimaux pour $\prec$ sont donc les modèles dans lesquels non seulement $P(a)$ est vrai mais aussi $Q(a)$. Dans ce formalisme, le qualificateur « vraisemblable » est apparemment plutôt rattaché au « donc ». Pourtant, la conclusion $Q(a)$ n'en reste pas moins détachable − en tant que conclusion valable par défaut.

Notre propos était simplement de montrer en quoi les récents développements de la logique formelle peuvent nous amener à nous demander si Toulmin n'indique pas une difficulté, plus qu'il ne la résout − mais n'est-ce pas là la seule ambition qu'il se fixe au début de son livre ? Si la forme des énoncés servant de fondement aux garanties autres que mathématiques est (on le constate) plurielle, devons-nous en conclure que chacune d'elles ne puisse recevoir un traitement formel ? Notre exemple de formalisation, il est vrai, ne rend pas compte de la distinction entre garantie et fondement. Mais que penser d'une telle distinction ?Nous pouvons voir en F le simple « condensé » du débat critique au terme duquel le bien-fondé de la garantie s'est imposé. Dans ce cas, reprocher à la logique formelle de ne pas distinguer entre garantie et fondement revient essentiellement à lui reprocher de ne laisser aucune place à la révision dans l'argumentation. Et ceci nous ramène finalement à la première objection de Toulmin. Nous avons vu que celui-ci reprochait à la logique de ses contemporains de ne pas envisager l'argumentation dans son aspect procédural. Depuis lors, certains chercheurs en Intelligence Artificielle ont tenté de surmonter cette limitation. Nous étudierons plus loin la théorie qu'ils nous proposent[2].

1.2.3 La conversation selon Grice

Il est coutume de présenter la théorie de Grice [95, 96], vers laquelle nous nous tournons à présent, comme relevant de la pragmatique. Définir la pragmatique n'est pas simple pour autant. Nous dirons simplement qu'elle concerne essentiellement l'usage du langage. En fait, Grice se propose d'expliquer comment il se fait que les locuteurs puissent signifier *de facto* plus que ce qui est véhiculé par le simple contenu logique de l'énoncé. Cet excédent de sens, Grice le nomme l'« implicature » de l'énoncé.Il est des cas où les implicatures peuvent être déterminées uniquement à partir des mots et de la forme de la phrase énoncée. Dans ce cas, Grice parle d'implicatures conventionnelles. Dans la mesure où le sens des mots les expliquent, elles n'intéressent pas la théorie de la conversation que Grice nous propose. De même, il est des cas où les implicatures peuvent être calculées sur la base de règles de nature esthétique, sociale ou morale. Elles non plus n'intéressent pas la théorie de la conversation. Mais il est des cas où ces deux types d'explications échouent. Dans ces cas-là, Crice parle d'implicatures conversationnelles. Ce sont elles que la théorie de la conversation doit expliquer.

1.2.4 Les maximes conversationnelles

Comment elle y parvient est maintenant bien connu. Elle fait l'hypothèse que les participants à une conversation adoptent un comportement coopératif, qui les

[2]Cf. section 2.4, pp. 97 et *sqq*.

conduit à respecter (ou, nous verrons en quel sens, exploiter) les quatre maximes conversationnelles que voici :

1. *Maxime de quantité* : soyez le plus informatif possible, et évitez d'être moins informatif qu'il n'est nécessaire.
2. *Maxime de qualité* : n'affirmez pas ce que vous croyez être faux, ni ce pour quoi vous manquez de preuves.
3. *Maxime de relation* : parlez à propos (soyez pertinent).
4. *Maximes de manière* : soyez clair, c'est-à-dire
   – évitez de vous exprimer avec obscurité
   – évitez d'être ambigu
   – soyez bref (ne soyez pas plus prolixe qu'il n'est nécessaire)
   – soyez ordonné.

Ces maximes sont intéressantes pour le logicien. A première vue, elles lui permettent de désamorcer l'une des objections qui lui est traditionnellement adressée, selon laquelle le sens des connecteurs dans la logique ne correspond pas à leur sens dans la langue naturelle. Une réponse désormais possible consiste à soutenir que le sens des connecteurs est bien déterminé par les conditions de vérité fournies par la théorie logique, quoique leurs usages pragmatiques ne le soient pas ; ces derniers relèvent des maximes conversationnelles. Voyons-le sur un exemple. Il est banal de faire remarquer au logicien que l'opérateur de conjonction avec lequel il travaille est commutatif, de sorte que l'ordre dans lequel les énoncés conjoints apparaissent est indifférent. En général, tel n'est pas le cas dans le langage naturel. Toutefois, nous pouvons considérer cet aspect séquentiel comme une implicature conversationnelle produite par la maxime de manière « soyez ordonné ». Il arrive, notons-le, que nous ayons non seulement une succession temporelle, mais en plus une relation causale. Il arrive aussi que nous ayons une simple concomitance. La réponse de Grice consiste ici à diviser les implicatures conversationnelles en deux classes, selon qu'elles dépendent ou non du contexte particulier de l'énonciation. Il qualifie les secondes d'implicatures généralisées (*generalized*), et les premières d'implicatures particularisées (*particularized*). Grice considérerait vraisemblablement les implicatures déclenchées par une conjonction comme appartenant à la classe des implicatures particularisées ; elles se caractérisent par leur sensibilité au contexte particulier de l'énonciation. Quoique indépendantes de ce dernier, les implicatures généralisées n'en restent pas moins annulables – ce en quoi elles diffèrent des implicatures conventionnelles tout à l'heure évoquées. La raison en est qu'il peut y avoir ce que Grice appelle « exploitation » d'une maxime par sa violation. Il envisage deux cas. Le premier est celui où le locuteur viole une maxime pour ne pas en violer une autre, qu'il juge plus importante :

> *Exemple.* *A* et *B* préparent un voyage en France et aimeraient bien rendre visite à *C*. *A* demande : « où habite *C* ? ». *B* répond : « Quelque part en France ». La réponse de *B* viole la maxime de quantité, mais cette violation est conditionnée par le souhait de ne pas violer la maxime de qualité.

Le second cas est celui où le locuteur viole une maxime pour transmettre une implicature par le biais d'une figure de rhétorique :

> *Exemple de la litote.* A propos d'un homme dont on sait qu'il a tout cassé : « Il avait un peu bu. »

Comme beaucoup l'ont remarqué, de ce qu'une implicature (conversationnelle) peut être annulée sans produire une contradiction, il s'ensuit qu'une implicature n'est pas assimilable à une implication. Il faut bien sûr comprendre : à une implication matérielle. Car, au moins de prime abord, il est tentant d'identifier l'implicature à une inférence non-monotone, puisque toutes deux révisables. Qu'une telle identification soit ou non fondée – et utile, telle est la question que nous pourrons nous poser [3].

1.2.5 Développements ultérieurs

Nous indiquons, pour finir, certains des développements auxquels a donné lieu la théorie gricienne des implicatures.

La notion d'implicature conventionnelle est mise à l'arrière-plan par Grice. Karttunen et Peters [123] tentent de lui donner droit de cité dans la littérature pragmatique, en montrant que cette notion entretient un lien étroit avec la notion de présupposition. Depuis Strawson, on prend généralement le terme « présupposition » comme signifiant ce qui doit être vrai pour qu'une phrase soit vraie ou fausse. Par exemple,

(a) Même Bill aime Marie

présuppose

(b) D'autres gens que Bill aiment Marie.

Tout d'abord, nos auteurs suggèrent qu'un grand nombre de phénomènes habituellement regroupés dans la catégorie générale des présuppositions peuvent être traités comme des implicatures conventionnelles. Ils prennent ici l'exemple des locutions *seulement*, *même*, *aussi* et *non*. Ils prennent également l'exemple des verbes dits factifs (comme *réaliser*), qui présupposent la vérité de leur complément d'objet. Ainsi, ils soutiennent que (b) est une implicature de (a) et qu'elle est déclenchée par la présence de *même*. Ensuite, Karttunen et Peters indiquent comment intégrer l'analyse des implicatures conventionnelles à une sémantique

[3] Nous aborderons ce point ci-dessous p. 84 et *sqq*.

formelle des langues naturelles, comme celle que proposa Montague [173, 174]. Le cas de la négation pose apparemment un problème. Un énoncé négatif comme *Le roi de France n'est pas chauve* prête à confusion. Il peut vouloir dire que la propriété d'être chauve n'est pas satisfaite par l'individu dont on dit qu'il est roi de France. Il peut aussi être interprété comme niant l'existence d'un individu qui serait roi de France. Pour rendre compte de cela, Karttunen et Peters introduisent deux types de négations, qu'ils appellent la négation ordinaire et la négation contradictoire. En gros, alors que la première touche seulement ce qui est dit, la seconde attaque ce qui est conventionnellement implicité. Ce style d'analyse a déjà été proposé, notamment par Wilson [269]. Recourant à une logique trivalente, il distingue entre une négation interne et une négation externe. L'objection traditionnellement avancée consiste à dire que le langage naturel ne marque nullement cette différence et que celle-ci n'a donc pas de justification empirique [22, 138]. Le problème se pose avec d'autant plus d'acuité, croyons-nous, que Karttunen et Peters se placent dans le cadre de la sémantique formelle de Montague. Celle-ci n'abandonne-t-elle pas le calcul des prédicats du premier ordre au profit d'une grammaire supposée utiliser des catégories du lexique ?

Une seconde direction de recherches a été empruntée par Gazdar [87, 88, 89]. Celui-ci propose un traitement d'ensemble des processusinférentiels, qu'ils soient sémantiques (vérifonctionnels) ou pragmatiques (non-vérifonctionnels). Introduisant un ordre d'application des inférences, Gazdar suppose que les premières inférences tirées d'une phrase *P* sont ses implications (logiques), ensuite ses implicatures et enfin ses présuppositions. Pour que les secondes et les troisièmes ne soient pas annulées, il suffit, nous dit-il, qu'elles soient consistantes avec les énoncés qui constituent le contexte, et donc avec les implications. Les préoccupations de Gazdar sont assez proches de celles des théoriciens de la révision. Comme nous le verrons au chapitre suivant, ceux-ci mettent au premier plan l'opération de changement de croyances, et tentent de la formaliser[4]. Illustrons le processus d'annulation auquel Gazdar songe au moyen de quelques exemples ; cela nous donnera l'occasion d'évoquer deux des notions-clés de son système, à savoir les notions d'implicature scalaire et d'implicature clausale. La caractéristique d'une implicature scalaire est de faire appel à ce que Gazdar appelle une « échelle quantitative », comme

< tous, la plupart, beaucoup, quelque, peu, ... >,
< certain, probable, possible, >,
< bouillant, chaud >,
< et, ou>.

Une échelle quantitative peut être, on le voit, définie comme un ensemble ordonné de termes $<e_1, \ldots, e_n>$ tel que, si A est un cadre syntaxique et si $A(e_i)$ est une phrase

[4]Cf. ci-dessous, p. 72 et *sqq*.

bien formée, $A(e_1)$ implique $A(e_2)$, $A(e_2)$ implique $A(e_3)$, etc., mais non l'inverse. Aux yeux de Gazdar, la notion d'échelle quantitative permet de donner un sens plus précis à la maxime gricéenne de quantité. Soit une échelle $<e_1,\ldots,e_n>$. Imaginons que le locuteur asserte $A(e_i)$. Il aura donné l'information la plus forte si elle implicite (scalairement) $\neg A(e_{i-1})$, $\neg A(e_{i-2})$ et ainsi jusqu'à $\neg A(e_1)$. Cette définition est, de prime abord, séduisante. Par exemple, elle rend compte du fait relativement trivial que

(a1) La plupart des garçons étaient à la réception

a pour implicature

(b1) Tous les garçons n'étaient pas à la réception.

De même, elle rend compte du fait que, dans la vie de tous les jours, nous prenons généralement la disjonction en son sens exclusif plutôt qu'en son sens inclusif. Voici à présent un exemple très simple qui illustrera le processus d'annulation dont Gazdar veut rendre compte[5]. Considérons la série de phrases :

(c1) Quelques garçons étaient à la réception
(c2) Tous les garçons n'étaient pas à la réception
(c3) Quelques garçons, et en fait tous, étaient à la réception
(c4) Tous les garçons étaient à la réception.

Il nous arrive d'énoncer des phrases du type (c3). Globalement, on dira que, dans ce type d'énonciation, le 'tous' vient réviser l'implicature scalaire (c2) normalement déclenchée par le 'quelques' – révision marquée linguistiquement par la locution 'en fait'. Gazdar formule les choses ainsi. Il nous invite à distinguer deux types d'implicatures : les implicatures potentielles (ou *im-plicatures*), qui sont calculables indépendamment du contexte, et les implicatures actuelles, qui sont le résultat de l'interaction du contexte et de la phrase énoncée. Comparons (c1) et (c3). Tous deux ont (c2) pour implicature scalaire potentielle. Mais seul (c1) a (c2) pour implicature scalaire actuelle : à la différence de (c1), (c3) implique logiquement (c4) ; comme l'ordre d'application des inférences prévoit que les implications logiques sont tirées en premier, une fois (c4) ajouté au contexte, nous ne pouvons plus y ajouter (c2). Il faut ici préciser que le résultat des inférences est toujours épistémiquement modifié (*epistemically modified*), c'est-à-dire enchâssé dans un opérateur épistémique 'K_S' lu « le locuteur sait que... ».

La deuxième notion-clé du système est celle d'implicature clausale. Interprétons 'P_S' (le dual de K_S) comme signifiant « le fait que ... est compatible avec tout ce que le locuteur sait » (possibilité épistémique). Considérons une phrase complexe ϕ qui contient une phrase enchâssée ψ, et telle que ϕ n'implique ni ne

[5]Exemple que l'on trouvera dans e.g. Gazdar [89, p. 64].

présuppose la vérité de ψ. Par exemple :

$$\phi := \text{ Si Jean a commis un meurtre, c'est un délinquant}$$
$$\psi := \text{ Jean a commis un meurtre.}$$

Soit ϕ' une autre expression, contenant elle aussi ψ, mais cette fois impliquant ou présupposant ψ. Par exemple :

$$\phi' := \text{ Jean est un délinquant puisqu'il a commis un meurtre.}$$

Gazdar fait ici l'hypothèse (à première vue vraisemblable) que, si le locuteur choisit d'asserter ϕ plutôt que ϕ', il implicite que, de son point de vue, $\{P_S\psi, P_S\neg\psi\}$. L'implicature est appelée « clausale », parce qu'elle est liée au type de phrase ou de construction syntaxique qui la déclenche. Désignons par $f_c(\phi)$ l'ensemble des implicatures clausales de ϕ. Nous avons

$$f_c(\text{si } p \text{ alors } q) = \{P_S p, P_S\neg p, P_S q, P_S\neg q\},$$

car il existe une expression plus forte qui nous engagerait sur la vérité de p et celle de q. Il s'agit de « q puisque p ». Pour une raison semblable, nous avons :

$$f_c(p \text{ ou } q) = \{P_S p, P_S\neg p, P_S q, P_S\neg q\}$$
$$f_c(\text{obligatoirement} p) = \{P_S p, P_S\neg p\}.$$

Dans le premier cas, l'expression plus forte qui nous engagerait sur la vérité des termes enchâssés est « p et q ». Dans le second cas, il s'agit de l'expression « nécessairement p ». Voici maintenant un cas d'annulation d'une implicature scalaire par une implicature clausale. Considérons, nous dit Gazdar, les trois phrases suivantes :

(d1) John ou Marie l'a fait
(d2) John ou Marie l'a fait, sinon les deux
(d3) John et Marie l'ont fait.

La phrase intéressante est ici (d2). Nous sommes enclins à dire que cette phrase est de la forme

si $\neg$ (d3) alors (d1).

Si cette paraphrase est correcte, alors nous sommes aussi tentés de dire que le 'si' vient suspendre l'implicature scalaire normalement déclenchée par le 'ou'. Plus généralement, (d2) et (d1) ont toutes les deux '$\neg P_S$d3' (c'est-à-dire '$K_S\neg$d3') pour implicature scalaire potentielle ; une disjonction est d'habitude prise au sens exclusif. Mais '$P_S\neg\neg$ d3' (c'est-à-dire 'P_S d3') compte au nombre des implicatures clausales déclenchées par le 'si' de (d2). Comme l'ordre d'application des inférences prévoit que les implicatures clausales sont tirées en premier, une fois 'P_Sd3' ajouté au contexte, nous ne pouvons plus y ajouter sa négation.

Une première objection souvent adressée à la théorie de Gazdar consiste à dire que l'ordre d'application des règles d'inférence que la théorie prévoit est arbitraire. Une deuxième objection consiste à faire remarquer qu'elle ne tient pas compte de toutes les implications pragmatiques que déclenche un énoncé. Par exemple, on admettra assez volontiers qu'une phrase du type *q si p* tend souvent à faire entendre, non seulement « *p* implique *q* », mais aussi la réciproque « *q* implique *p* », comme dans :

(e1) Si tu me donnes un peu de ta glace, je te donne un peu de la mienne.

Or, le couple <si et seulement si, si ... alors > forme une échelle quantitative, de sorte que (e1) a pour implicature scalaire :

(e2) $K_S \neg$ (Si je te donne un peu de ma glace, tu me donnes un peu de la tienne).

Nous rencontrons ici un type d'inférence pragmatique que la théorie de Gazdar ne prévoit pas et qui prend le pas sur une implicature scalaire. Ceci indiquerait que le mécanisme que Gazdar imagine n'est pas suffisamment général.

Quoi qu'il en soit de cette question, nous voyons que le thème des implicatures constitue un point évident d'articulation entre le domaine de l'argumentation et celui de la révision. Nous laissons provisoirement cette question de côté, car nous ne voulons pas anticiper sur certains de nos développements ultérieurs[6]. Indiquons simplement que, depuis les travaux de Gazdar, les pragmaticiens ont accordé une attention de plus en plus grande à la défaisabilité des implicatures. En témoigne l'oeuvre de Levinson [138, 139]. Son idée maîtresse est de regrouper les maximes conversationnelles de Grice en deux principes, le *principe-Q* et le *principe-I*, dont voici l'énoncé[7] :

Principe-Q

Maxime du locuteur :

Faîtes que votre contribution soit aussi informative que le demandent les buts de l'échange conversationnel

Corollaire pour l'interlocuteur :

Comprenez que le locuteur a émis l'assertion la plus forte

[6]Nous aborderons le thème des implicatures à la section 2.3.

[7]Sperber et Wilson [238] poussent plus loin la réduction, et subsument l'ensemble sous un seul et même principe, celui de la pertinence.

Principe-I

Maxime du locuteur : la maxime de minimisation

Dîtes-en le moins possible, i.e. produisez la séquence linguistique minimale suffisante pour réaliser vos buts communicationnels

Corollaire pour l'interlocuteur : règle d'enrichissement

Amplifiez le contenu informationnel de l'énoncé du locuteur, en trouvant une interprétation plus spécifique, jusqu'à ce que vous jugiez avoir atteint l'intention informative du locuteur

Comme Gazdar, Levinson consacre une bonne partie de ses efforts à la description de données linguistiques. Or, notant que les implicatures-Q et les implicatures-I entrent souvent en conflit, il suppose que les premières l'emportent nécessairement sur les secondes. Le cas de la phrase *q si p*, que nous venons d'évoquer, rend cette hypothèse suspecte. Apparemment, c'est en vertu du principe-I que celui qui affirme *q si p* tend à faire aussi entendre l'implication réciproque (il en dit le moins possible). Et c'est en vertu du principe-Q que celui qui affirme *q si p* tend à faire entendre la négation de l'implication réciproque (le locuteur donne l'affirmation la plus forte). Nous rencontrons ainsi un cas où une implicature-I prend le pas sur une implicature-Q.

1.2.6 L'argumentation dans la langue

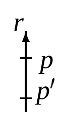
A la différence de Perelman et Olbrechts-Tyteca, Anscombre et Ducrot ne prennent pas pour objet d'étude l'argumentation au sens ordinaire du terme, conçue comme la mise en oeuvre d'une stratégie de discours dans le but de convaincre un auditoire. En témoigne le titre qu'ils donnent à l'un de leurs ouvrages les plus importants : *l'argumentation dans la langue*. Leur champ d'investigation se limite, nous le voyons à ce seul titre, aux potentialités argumentatives des énoncés, en tant qu'induites par des propriétés immanentes à la langue. Nous verrons un peu plus loin ce que cela signifie. Du reste, cela n'exclut évidemment pas qu'un locuteur ne puisse utiliser ces potentialités pour mettre en oeuvre telle ou telle stratégie de persuasion. Reste que, comme Grice et ses disciples, Anscombre et Ducrot jugent plus utiles de mettre entre parenthèse ce problème, qui leur paraît secondaire. A la différence de Grice, ils ne mettent plus au premier plan la situation de communication ou des facteurs pragmatiques.

Anscombre et Ducrot partent de l'idée selon laquelle l'acte d'argumenter ne se ramène pas à une inférence logique. Ils tentent d'étayer cette hypothèse par un grand nombre de données linguistiques, qui mettent généralement en jeu un connecteur dit « non-logique », tel que *presque*, *peut-être* ou *à peine*. Tenter de faire le point sur ces dernières est une entreprise ambitieuse ; nous ne nous y risquerons pas. Ces données montreraient qu'il existe des enchaînements dont on ne peut pas rendre compte en termes de valeurs de vérité. Voici l'un des exemples qu'ils donnent :l'enchaînement de *presque p* à *p* n'est évidemment pas acceptable du point de vue de la théorie logique, puisque *presque p* implique *non-p*. Pourtant, à la remarque « Je vais faire un tour en attendant que tu sois prêt », nous pouvons répondre « inutile, j'ai presque fini ». Ici, « j'ai presque fini » oriente l'interlocuteur vers la conclusion « j'ai fini »[8].

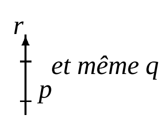
1.2.7 Echelle argumentative, topos

Anscombre et Ducrot proposent une théorie de l'argumentation, qu'ils élaborent en deux temps. Dans un premier temps, ils développent une théorie dite « des échelles argumentatives » – elle est essentiellement due à Ducrot [60, 61]. Dans un second temps, ils y introduisent un nouvel ingrédient, celui de *topos* ou de règle argumentative. Pour plus de simplicité, nous n'évoquerons pas ici la théorie de la polyphonie présentée dans Anscombre et Ducrot [10], Ducrot et al. [62] et Ducrot [64]. Indiquons seulement que, directement empruntée à Bakhtine, et remettant en cause la thèse de l'unicité du sujet parlant, la notion de polyphonie est introduite pour, principalement, rendre compte de la négation, l'ironie et le discours rapporté.

Considérant comme primitive la valeur argumentative des énoncés, la théorie des échelles argumentatives paraît relativement simple, au moins dans son principe. Ducrot dit de deux énoncés p et p' qu'ils appartiennent à une même échelle argumentative, lorsqu'ils peuvent tous deux être considérés comme des arguments vers une même conclusion r (ou une même classe de conclusions), mais que l'un des arguments, disons p, est plus fort que l'autre, ici p'. Ducrot représente cela de la manière suivante :

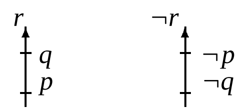


L'usage que Ducrot fait de cet outil est double. Dans un premier temps, il l'utilise pour décrire certains enchaînements auxquels, dans le discours, nous nous livrons. Par exemple, il paraphrase une séquence de la forme *p et même q* ainsi : p et q appartiennent à la même échelle argumentative, mais q est argumentativement plus fort que p. Cela donne :

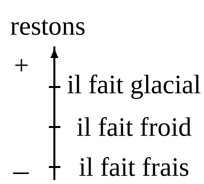


[8] Pour un contre-exemple, voir ci-dessous p. 24.

De même, il paraphrase une séquence de la forme *p mais q* par : *p* et *q* appartiennent à des classes argumentatives opposées ; *q* est argumentativement plus fort que *p*. Dans un second temps, il étudie la notion d’échelle argumentative en elle-même, et tente de dégager certaines des lois qui la régissent. Il en découvre au moins trois. Ducrot nomme la première « loi d’inversion argumentative ». Elle stipule que l’échelle des énoncés négatifs est l’inverse des énoncés positifs au sens où, si *q* est plus fort que *p* dans l’échelle argumentative déterminée par *r*, alors *non-p* est un argument plus fort que *non-q* pour la conclusion *non-r* :

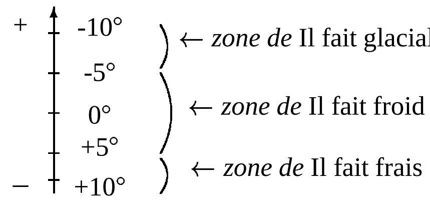


Cette loi explique par exemple le comportement de *et même* par rapport à la négation. Si nous acceptons la hiérarchie sous-jacente à « Il a le troisième cycle, et même le doctorat d’Etat », alors nous devons aussi accepter de dire, au cas où nous avouerions nous être trompés sur les faits, « Il n’a pas le doctorat d’Etat, ni même le troisième cycle ». Ducrot nomme la seconde loi « loi de faiblesse ». Elle stipule que, si *p* appartient à l’échelle argumentative déterminée par *r* mais est un argument faible pour *r*, alors dans certaines circonstances *p* peut être un argument pour *non-r*. Enfin, il appelle la troisième loi « loi d’abaissement ». Elle concerne les rapports entre une échelle argumentative donnée et l’échelle des graduations physiques à laquelle celle-ci renvoie. Soit un énoncé *p* (e.g. « il fait froid ») appartenant à l’échelle argumentative déterminée par la conclusion *r* (e.g. « restons à la maison ») :



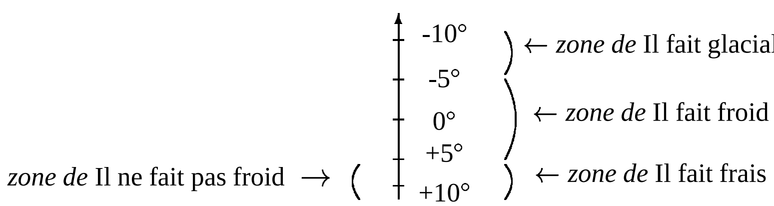
Echelle argumentative

Considérons l’échelle des graduations physiques à laquelle cette échelle argumentative renvoie. On suppose que la première tire son orientation de celle que possède déjà l’échelle argumentative. Cela donne :



Graduation physique homologue

Soit I l'ensemble des valeurs qui rendent vrai l'énoncé *p*. Dans notre exemple, il s'agit de l'ensemble des températures entre $-5°$ et $+5°$. La loi d'abaissement stipule que si *p* est vérifié dans I, alors la négation de *p* est vérifiée dans, et seulement dans, la zone de la graduation qui est inférieure à I. Dans notre exemple, il s'agit de l'ensemble des températures supérieures à $+5°$, qui devient la zone du non-froid :



Loi d'abaissement

La loi d'abaissement est destinée à rendre compte du fait que, dans de nombreux cas, la négation est équivalente à *moins que*. Ainsi, dans son usage ordinaire, un énoncé comme « il ne fait pas froid » est généralement synonyme de « il fait frais » plutôt que de « il fait glacial ».

Dans sa seconde formulation, la théorie de Anscombre-Ducrot fait intervenir un concept supplémentaire, celui de *topos* (pl. *topoi*). Afin d'illustrer la difficulté que cette notion vient résoudre, revenons à la locution *presque p* tout à l'heure évoquée. La théorie des échelles argumentatives suppose que, bien qu'impliquant logiquement *non-p*, *presque p* a la même valeur argumentative que *p*. Plus précisément, elle suppose que *presque p* appartient toujours à la même échelle argumentative que *p*, quoiqu'introduisant un argument plus faible. Il est des contre-exemples possibles à cette hypothèse. L'enchaînement « dépêche-toi : il est presque mort » en est un. Afin de lever la difficulté, force est donc de ne plus rendre automatique la relation entre un argument et l'énoncé qu'il vise, et de supposer l'existence de règles cautionnant cette dernière. Anscombre et Ducrot nomment de telles règles *topoi*, parce qu'elles s'avèrent graduelles. Elles sont soit de la forme « plus *P*, plus *Q* » (notation : $< +P, +Q >$), soit de la forme « plus *P*, moins *Q* » ($< +P, -Q >$), soit de la forme « moins *P*, plus *Q* » ($< -P, +Q >$), soit enfin de la forme « moins P, moins Q » ($< -P, -Q >$). Comme la langue admet des formes topiques contraires, la relation entre un argument et l'énoncé qu'il vise est désormais plus souple. Elle n'est cependant pas arbitraire. Anscombre et Ducrot font ici l'hypothèse que les connecteurs « non-logiques », comme *presque* ou *à peine*, ont pour fonction essentielle de déterminer l'orientation argumentative de la phrase, en sélectionnant la forme topique appropriée à l'enchaînement.Dans le cas de la phrase « dépêche-toi : il est presque mort », le chemin interprétatif qui relie l'argument à sa conclusion serait du type « plus la mort est proche, plus il faut se dépêcher ». Si tel est ce que les auteurs veulent dire, lorsqu'ils parlent de potentialités argumentatives qui seraient intrinsèques à la langue, force nous est de reconnaître que nous sommes un peu embarrassés. Tout d'abord, nous ne comprenons pas très bien comment une forme topique comme « plus la mort est proche, plus il faut se dépêcher » permet de passer de « il est presque mort » à « il n'est pas mort ». Ensuite, dans l'hypothèse où un tel passage serait possible, nous ne voyons pas très bien ce qui (dans la langue) expliquerait pourquoi, dans certains cas, *presque p* orienterait vers p (modulo un certain topos) alors que, dans d'autres cas, cette locution orienterait vers non-p (modulo un autre topos).

Quoi qu'il en soit de cette question, il est ici intéressant de remarquer que, aux yeux de Ducrot [63, p.14], un topos ne peut pas être assimilé à une règle d'inférence, du fait précisément de son caractère graduel. A l'instar de Jayez [121], nous pouvons trouver ce point de vue discutable. La forme topique « plus P, plus Q » s'exprime mal en logique classique, mais se modélise relativement bien dans la logique floue (*fuzzy logic*) de Bellman & Zadeh [23]. Prade [201] donne à un topos le nom de règle d'inférence graduelle (en abrégé, r.i.g.). Voici le principe de l'analyse qu'il en donne. Pour représenter une règle d'inférence graduelle, on part de la notion de sous-ensemble flou. Un sous-ensemble flou P de X est défini par une fonction d'appartenance qui associe à chaque élément x de X le degré $\mu_{\mathrm{P}}(x)$, compris entre 0 et 1, avec lequel x appartient à P. Etant donnés deux sous-ensembles flous P et Q, la r.i.g. « plus $x \in P$, plus $y \in Q$ » se laisse traduire par la contrainte

$$\mu_Q(y) \geq \mu_P(x) \tag{1.1}$$

Littéralement, cette inégalité nous dit que le degré d'appartenance de y à Q ne saurait être inférieur au degré d'appartenance de x à P. Ceci exprime relativement bien l'idée que, plus le degré d'appartenance de x à P est grand, plus le degré d'appartenance de y à Q est grand. Deux remarques s'imposent ici. Tout d'abord, nous voyons que (1.1) peut être généralisée sous la forme

$$\mathrm{m}(\mu_Q(y)) \geq \mu_P(x), \tag{1.2}$$

où m est une fonction monotone croissante de $[0,1]$ dans $]$-$\infty, 1]$. L'intérêt de cette généralisation est de permettre une plus grande souplesse dans l'expression d'une r.i.g.. Dans certains cas, il paraît raisonnable d'exiger du degré d'appartenance de x à P qu'il dépasse un certain seuil, pour avoir quelque incidence sur le degré d'appartenance de y à Q [9] ; il suffit de poser m(t) = t−ε avec $0 < \varepsilon < 1$. Ensuite, nous voyons également comment étendre notre inégalité initiale (1.1) à l'expression des autres formes topiques élémentaires que sont « moins ... moins », « moins ... plus », « plus ... moins ». Dire que le degré d'appartenance de x à P diminue revient à dire que le degré d'appartenance de x au complément de P augmente. Nous savons comment définir le complément P^C d'un sous-ensemble flou P ; nous avons

$$\forall x \in X, \mu_{P^C}(x) = 1 - \mu_P(x). \tag{1.3}$$

Nous obtenons le système de paraphrases :

$$\begin{aligned}
\text{plus } P, \text{ plus } Q &\quad \overset{\text{def}}{\longleftrightarrow} \quad \mu_Q(y) \geq \mu_P(x) \\
\text{plus } P, \text{ moins } Q &\quad \overset{\text{def}}{\longleftrightarrow} \quad 1 - \mu_Q(y) \geq \mu_P(x) \\
\text{moins } P, \text{ plus } Q &\quad \overset{\text{def}}{\longleftrightarrow} \quad \mu_Q(y) \geq 1 - \mu_P(x) \\
\text{moins } P, \text{ moins } Q &\quad \overset{\text{def}}{\longleftrightarrow} \quad 1 - \mu_Q(y) \geq 1 - \mu_P(x).
\end{aligned}$$

Le type d'enchaînement argumentatif que Anscombre et Ducrot nous décrivent évoque assez naturellement cette forme de raisonnement appelée « raisonnement par interpolation » :

$$\frac{\begin{array}{l}\text{plus } x \text{ est } P\text{, plus } y \text{ est } Q \\ x \text{ est très } P\end{array}}{y \text{ est très } Q.}$$

L' opérateur dit de compression (notation : COM_2) qui correspond à l'adjectif *très* est défini par : $\mu_{\mathrm{COM}_2}(P)(x) = [\mu_P(x)]^2$.

1.3 Les théories formelles

1.3.1 Le programme de Hamblin

L'optique que Hamblin adopte dans *Fallacies* (1970) est normative, au sens où il s'assigne pour but de dégager les critères nous permettant d'évaluer une argumentation donnée. Comme le titre de son ouvrage l'indique, Hamblin aborde l'argumentation correcte, si l'on ose dire, par son *négatif* : l'argumentation incorrecte. Il tente d'attirer l'attention sur un champ d'investigation nouveau, plutôt que d'en faire un examen exhaustif. Cela explique que son ouvrage ait donné lieu

[9] comme dans « plus une chaussée est âgée, plus la qualité de la fondation risque d'être mauvaise ».

à une importante littérature [10]. Dans les cinq premiers chapitres de son ouvrage, Hamblin étudie en détail l'histoire des paralogismes. Il en fait remonter l'origine aux *Réfutations sophistiques* d'Aristote et il la fait se développer sans coupure jusqu'à nos jours, aboutissant à ce qu'il nomme le « traitement standard ». Dans les trois derniers chapitres, il présente son propre traitement de certains d'entre eux. Selon la définition standard, un paralogisme est une argumentation qui « a les apparences de la validité sans être valide ». Il existe un certain nombre de paralogismes dont la logique classique ne peut apparemment pas rendre compte. Hamblin donne deux exemples : celui du raisonnement circulaire, et celui du paralogisme de la question dite « biaisée », qu'il illustre par la question

Avez-vous cessé de battre votre mari ?

Hamblin en propose un traitement dialogique, que Woods et Walton qualifieront de « pionnier » [273, p. 92]. Voyons en quoi il consiste.

1.3.2 Le système des questions-réponses

Hamblin [101, chap. 8] esquisse deux systèmes dialogiques, qu'il nomme respectivement « Jeu des obligations » (*Obligation Game*) et « système (de type) Pourquoi-Parce que » (*Why-because system*). Hamblin juge le second plus satisfaisant parce que, à la différence du premier, il autorise le locuteur à retirer en cours de partie certaines propositions de son stock d'opinions. Aussi, pour plus de simplicité, nous limiterons notre examen au second des deux systèmes. Tentant d'attirer l'attention sur un chemin d'investigation nouveau, Hamblin tâtonne dans l'élaboration de ses règles. Ceci en rend l'exposition d'autant plus délicate pour nous. A dire vrai, il nous sera parfois difficile de justifier ses choix. La question essentielle de Hamblin est celle de déterminer à quoi les locuteurs s'engagent dans un dialogue. Ce problème est également au coeur des préoccupations de Brandom [38]. Il eut été instructif de tenter d'étudier le travail de Hamblin à la lumière de celui de Brandom. Le temps nous a manqué pour le faire.

Chez Hamblin, le duel se présente comme une partie entre deux joueurs, WHITE et BLACK. Un stock d'engagements (*commitment-store*) C est associé à chaque participant. Contenant les énoncés à la vérité desquels le participant souscrit, ce stock va évoluer au cours du dialogue[11]. WHITE ouvre la partie, puis chacun des deux partenaires prennent tour à tour la parole. A chaque intervention, ils peuvent opter pour l'une des cinq stratégies suivantes :

1. défendre une série d'énoncés – ce qui est noté :

$$\lceil \text{Affirmations } \phi, \psi, \ldots, \tau \rceil.$$

[10] Woods et Walton sont les principaux représentants de ce courant de recherches.

[11] Hamblin [101, p. 264] n'exige pas du stock d'engagements qu'il soit consistant sur l'ensemble de la partie.

2. se rétracter :

$$\ulcorner\text{Pas d'engagement sur } \phi, \psi, ..., \tau\urcorner.$$

3. demander à l'autre de lever une ambiguïté :

$$\ulcorner\text{Est-ce que } \phi, \psi, ..., \tau\,?\urcorner.$$

4. inviter l'autre à justifier une affirmation antérieure :

$$\ulcorner\text{Pourquoi } \phi\,?\urcorner.$$

5. demander à l'autre de lever une contradiction

$$\ulcorner\text{Décider quant à } \phi\urcorner.$$

Le jeu des interventions est soumis aux trois règles suivantes :

S_1. Une question de la forme $\ulcorner$Est-ce que $\phi, \psi, ..., \tau\,?\urcorner$ est suivie de l'une des quatre interventions suivantes :
a. $\ulcorner$Affirmation$\neg(\phi \vee \psi \vee \ldots \vee \tau)\urcorner$;
b. $\ulcorner$Pas d'engagement sur $\phi \vee \psi \vee \ldots \vee \tau\urcorner$;
c. $\ulcorner$Affirmation $\phi\urcorner$, $\ulcorner$Affirmation $\psi\urcorner$, $\ldots$ ou $\ulcorner$Affirmation $\tau\urcorner$;
d. $\ulcorner$Pas d'engagement sur $\phi, \psi, \ldots, \tau\urcorner$.
Prenons juste un exemple. Posons p = *vous aviez un mari que vous battiez* et q = *vous avez cessé de le battre*. La question « avez-vous cessé de battre votre mari ? » se symbolise ainsi : $\ulcorner$Est-ce que (lequel des deux) $p \wedge q, p \wedge \neg q\,?\urcorner$. Une réponse affirmative à cette question est apparemment de type (c), i.e. $\ulcorner$Affirmation $p \wedge q\urcorner$. En gros, dans le cas de (a) et (b), le répondant refuse la question elle-même, qui lui semble « biaisée »[12].

S_2. Une question de la forme $\ulcorner$Pourquoi $\phi\,?\urcorner$ reçoit l'une des quatre réponses suivantes :
a. $\ulcorner$Affirmation $\neg\phi\urcorner$;
b. $\ulcorner$Pas d'engagement sur $\phi\urcorner$;
c. $\ulcorner$Affirmations $\psi, \psi \leftrightarrow \phi\urcorner$;
d. $\ulcorner$Affirmations $\psi, \psi \rightarrow \phi\urcorner$.

S_3. A une intervention de la forme $\ulcorner$Décider quant à $\phi\urcorner$, doit succéder une intervention de la forme $\ulcorner$Pas d'engagement sur $\phi\urcorner$ ou de la forme $\ulcorner$Pas d'engagement sur $\neg\phi\urcorner$.
Une intervention de la forme $\ulcorner$Décider quant à $\phi\urcorner$ doit (intuitivement) être comprise comme une invitation à lever une contradiction. Dans ce cas, le répondant annule son engagement sur ϕ ou celui sur $\neg\phi$.

Puis vient un second groupe de règles, qui coordonnent les engagements pris aux actes de langages accomplis :

[12] Voir le paragraphe 1.3.3.2 , page 32.

C_{intr} L'intervention ⌈Affirmation ϕ⌉ introduit ϕ dans le stock d'engagement du locuteur (sauf s'il y est déjà), ainsi que dans le stock d'engagement de l'interlocuteur, sauf si au coup suivant ce dernier répond par ⌈Affirmation $\neg\phi$⌉, ⌈Pas d'engagement sur ϕ⌉ ou ⌈Pourquoi ϕ ? ⌉ ;

La partie de cette règle qui concerne l'interlocuteur est peut-être discutable. Elle signifie que, si celui-ci ne s'oppose pas immédiatement au point de vue du locuteur, alors il lui concède par là même le point en discussion.

C'_{intr} L'intervention ⌈Affirmations ϕ, ψ⌉ introduit ϕ et ψ dans les stocks d'engagement sous les mêmes conditions que dans C_{intr} ;

C_{elim} Une affirmation de la forme ⌈Pas d'engagement sur $\phi, \psi, \ldots\tau$⌉ élimine chacun de ces énoncés du stock d'engagement du locuteur.

Voici un exemple de partie ; nous la représentons sous forme d'un tableau à deux colonnes, en simplifiant les notations :

	WHITE		BLACK
(1)	Est-ce que $p, \neg p$?	(2)	p
(3)	Est-ce que $q, \neg q$?	(4)	$\neg q$
(5)	q	(6)	$\neg q$
(7)	Pourquoi $\neg q$?	(8)	$p, p \to \neg q$
(9)	Pas d'engagement sur $p \to \neg q$;		
	Pourquoi $p \to \neg q$?	(10)	Pas d'engagement sur $p \to \neg q$
(11)	q	(12)	Pourquoi q ?
(13)	$p, p \to q$	(14)	Pourquoi $p \to q$?
(15)	$\neg p, \neg p \to (p \to q)$	(16)	Pas d'engagement sur $\neg p \to (p \to q)$
(17)	Décider quant à p	(18)	Pas d'engagement sur $\neg p$
(19)	$\neg p$	(20)	Décider quant à p
(21)	Pas d'engagement sur p	(22)	Pourquoi $\neg p \to (p \to q)$?
(23)	$\neg p \to (\neg p \vee q)$	(24)	Pourquoi $\neg p \to (\neg p \vee q)$?
(25)	$\neg p \to (q \vee \neg p)$,		
	$(\neg p \to (q \vee \neg p)) \to (\neg p \to (\neg p \vee q))$		
⋮	⋮	⋮	⋮

Les interventions (1) à (6) forment un premier bloc : en gros, WHITE concède p à BLACK, mais il lui refuse $\neg q$. Les interventions (7) à (10) forment un deuxième bloc : WHITE demande à BLACK de justifier sa position ; celui-ci, de son propre aveu, y échoue. Les interventions (11) à (16) peuvent elles aussi être regroupées ensemble : BLACK demande à WHITE de justifier q ; WHITE s'exécute en invoquant des raisons contradictoires, ce dont BLACK ne se rend pas immédiatement compte. Et ainsi de suite. Cet exemple est emprunté à Hamblin. Nous voyons que (4)-(6) mettent en jeu un conflit d'opinions : l'affirmation $\neg q$ est suivie de l'affirmation q puis de l'affirmation $\neg q$.

1.3.3 Application aux paralogismes

Hamblin utilise son cadre dialogique à l'analyse de deux paralogismes. Il se penche tout d'abord sur le raisonnement circulaire, puis sur le paralogisme de la question biaisée.

1.3.3.1 Pétition de principe

Voici une première argumentation au cours de laquelle, selon Hamblin, une pétition de principe est commise :

Scénario 1

	WHITE		BLACK
(1)	Pourquoi p ?	(2)	$p, p \to p$

Pour empêcher que ce scénario ne se produise, Hamblin [101, p. 271] propose d'introduire le couple suivant de règles :

R_1. Une question de la forme ⌈Pourquoi ϕ ?⌉ peut être posée seulement si ϕ figure dans le stock d'engagements de l'auditeur, mais ne figure pas dans celui du locuteur ;

R_2. Une réponse à la question ⌈Pourquoi ϕ ?⌉ qui n'est pas de la forme ⌈ Affirmation $\neg\phi$⌉ ou de la forme ⌈Pas d'engagement sur ϕ⌉ doit uniquement faire appel à des énoncés figurant déjà dans le stock des engagements de chaque participant.

En (1), WHITE invite BLACK à justifier p. Par R_1, nous en concluons que p ne figure pas dans le stock des engagements du premier. R_2 interdit alors au second d'arguer comme il le fait en (2), c-a-d en supposant ce qui, pour WHITE, est en question. L'analyse de Hamblin paraît plausible. Elle fait de la pétition de principe, pour reprendre Perelman, « non pas une faute de logique, mais de rhétorique » [192, p. 151] : le locuteur utilise une thèse que son interlocuteur est censé partager (règle R_2) alors qu'en fait il la conteste (règle R_1). Une remarque similaire s'applique aux deux dialogues suivants :

	WHITE		BLACK
(1)	Pourquoi p ?	(2)	$q, q \to p$
(3)	Pourquoi q ?	(4)	$p, p \to q$

Scénario 2

	WHITE		BLACK
(1)	Pourquoi p ?	(2)	$q, q \to p$
(3)	Pourquoi q ?	(4)	$r, r \to q$
(5)	Pourquoi r ?	(6)	$p, p \to r$

Scénario 3

Ainsi, dans le scénario 2, l'intervention (1) nous conduit à supprimer *p* du stock des engagements de WHITE – règle R_1. Et nous n'avons à l'y réintroduire ni en (2), ni en (3). Aussi, en (4), BLACK ne peut pas justifier *q* à l'aide de *p* – règle R_2. Ce scénario 2 fait songer à l'observation d'Aristote : « Il y a (...) pétition de principe (...), si on postule l'une de deux propositions qui s'impliquent nécessairement l'une l'autre : si, par exemple, ayant à démontrer que la diagonale est incommensurable avec le côté du carré, on posait que le côté est incommensurable avec la diagonale » (*Topiques*, VIII, 13, 163 a 11-14). Il y a ici pétition de principe, parce que la relation « être incommensurable avec » est symétrique.

De l'aveu même de Hamblin, cette modélisation à l'aide des règles R_1 et R_2 achoppe à une difficulté manifeste. Reprenons le scénario 2. Au coup (2), nous en concluons que *q* figure dans le stock d'engagement des deux joueurs (règle R_2). Mais alors, au coup suivant, WHITE n'a plus le droit d'exiger de BLACK qu'il justifie *q* (règle R_1). Notre couple de règles exclut donc tout questionnement en étapes – ce qui paraît absurde. Notons au passage que dans les scénarios 2 et 3 il est possible d'interpréter la stratégie de BLACK tout autrement que ne le fait Hamblin. Le but poursuivi par BLACK ne pourrait-il pas être de justifier *p*, en montrant en plusieurs temps que *p* est équivalent à *q* ou à *r* ?

Conscient de la difficulté, Hamblin propose de remplacer R_2 par :

R_3. Admettons que, à un moment donné de la partie, l'un des participants soutienne que $\lceil\phi\rceil$ (ou que $\lceil\phi, \phi \rightarrow \psi\rceil$) et que, par la suite, il ne se rétracte pas sur ϕ. Supposons que, à aucun moment, l'autre participant ne se soit engagé sur ϕ. Dans ce cas, dès qu'il en aura l'occasion, le second demandera au premier de justifier ϕ.

En présence du couple R_1-R_3, nos trois dialogues prendront fin seulement lorsque BLACK aura répondu de façon non-circulaire, c-a-d en cessant d'invoquer *p*. Hamblin juge, à juste titre, R_3 peu réaliste :

> « Elle implique, écrit-il, qu'un questionnement en`Pourquoi ?´*ne peut* prendre fin que lorsque sont atteintes des prémisses sur lesquelles les participants s'accordent *a priori* ; mais elle ne garantit aucunement que de telles prémisses seront un jour atteintes » [101, p. 272].

Malheureusement, Hamblin en reste là. Toutefois, revenant au couple R_1-R_2, nous pouvons tout d'abord nous demander si le problème que rencontre Hamblin n'est pas lié au fait que l'ordre dans lequel les deux règles sont appliquées permute : pour bloquer la pétition de principe, on applique R_1 avant R_2 ; pour rendre le questionnement par étapes impossible, on applique R_2 avant R_1. Or, il est des systèmes formels qui tentent de surmonter des difficultés similaires et qui, pour ce faire, introduisent un ordre de priorité entre les règles de déduction. Par exemple, van der Torre [256] propose un système déductif dit « à étages » (*phased*) dans lequel, pour des raisons qu'il est ici inutile d'aborder, nous pouvons, au cours d'une dérivation, appliquer la traditionnelle règle d'affaiblissement du conséquent avant la règle de renforcement de l'antécédent, mais jamais l'inverse. Pourquoi ne pas reprendre le procédé ?

Il est au moins une raison qui pourrait nous inciter à ne pas pousser davantage nos investigations dans cette direction. Elle nous est suggérée par l'objection que Woods et Walton [273] adressent au système de Hamblin. Imaginons que, initialement engagés sur (respectivement) $\{p \to q, q \to p, r \to p\}$ et $\{p, q, r, p \to q, q \to p, r \to p, s\}$ WHITE et BLACK échangent les propos suivants :

	WHITE		BLACK
(1)	Pourquoi p ?	(2)	$q, r, q \to p, r \to p$
(3)	q, r	(4)	s
(5)	Pas d'engagement sur q ;		
	Pourquoi q ?	(6)	$p, p \to q$

Il s'agit d'une variation sur le deuxième de nos dialogues. En (2), BLACK justifie p à l'aide de deux raisons indépendantes l'une de l'autre, q et r. En (3), WHITE accepte les deux raisons avancées par son adversaire. En (4), BLACK passe à un autre sujet. En (5), WHITE revient (pour une raison ou une autre) sur sa position et demande à son partenaire de justifier q. En (6), celui-ci s'exécute et justifie q à l'aide de p. Nul doute que BLACK commette ici une pétition de principe. Celle-ci n'est plus « bloquée » par le couple R_1-R_2 − et cela, indépendamment de la question de savoir s'il y a ou non lieu d'introduire un ordre de priorité entre les deux règles. En (1), comme précédemment, nous excluons p du stock des engagements de WHITE. En (3), nous y incluons p : WHITE accepte q et r. En (4), p figure toujours dans le stock des engagements de WHITE : BLACK change de sujet. Il en va de même en (5) : que WHITE se rétracte sur q ne signifie pas qu'il se rétracte aussi sur sa conséquence p ; WHITE ne se rétracte pas sur r. De la sorte, l'intervention (6) de BLACK n'est plus empêchée par R_2, quand bien même on autoriserait seulement un enchaînement du type « R_1 puis R_2 ».

Une question vient à l'esprit. Intuitivement, la règle R_2 interdit au locuteur de justifier sa position en recourant à des informations nouvelles pour l'interlocuteur. Les difficultés rencontrées ne montrent-elles tout simplement pas que cette règle est trop restrictive ?

1.3.3.2 Question biaisée

Il est tentant d'assimiler le paralogisme de la question « biaisée » (*biased*) au paralogisme par pétition de principe. En effet, qui biaise une question fait comme si son interlocuteur partageait le présupposé de sa question, alors qu'en fait il refuse celui-ci. Mais il est au moins deux raisons de ne pas confondre ces deux formes d'argumentations. Tout d'abord, nous n'avons plus nécessairement affaire à une question en « pourquoi ? ». Ensuite, le paralogisme est ici le fait de l'énonciateur, et non plus le fait du destinataire de la question. Hamblin [101, p. 216] se limite délibérément aux questions de la forme « Est-ce que $\phi, \psi, \ldots, \tau$? ». Leur présupposé correspond à la disjonction des possibilités parmi lesquelles le répondant est supposé choisir. Posons p = *vous aviez un mari que vous battiez* et q = *vous avez cessé de le battre*. La question « avez-vous cessé de battre votre mari ? » se symbolise ainsi :

$$\ulcorner\text{Est-ce que (lequel des deux) } p \wedge q, p \wedge \neg q\,?\urcorner$$

Le présupposé de cette question est $(p \wedge q) \vee (p \wedge \neg q)$ – ce qui, en logique classique, est équivalent à p. Afin de rendre compte du paralogisme de la question biaisée, Hamblin [101, p. 268] utilise, outre le principe S_1 tout à l'heure mentionné, la règle suivante :

R_3. L'intervention $\ulcorner$Est-ce que $\phi, \psi, \ldots, \tau$? $\urcorner$ introduit $\phi \vee \psi \vee \ldots \vee \tau$ dans le stock d'engagement du locuteur (à moins qu'il n'y soit déjà) ; elle l'introduit également dans le stock d'engagement du répondant, à moins que celui-ci ne réponde au coup suivant par $\ulcorner$Affirmation $\neg(\phi \vee \psi \vee \ldots \vee \tau)\urcorner$ ou par $\ulcorner$Pas d'engagement sur $(\phi \vee \psi \vee \ldots \vee \tau)\urcorner$.

Reprenons l'exemple de la question « avez-vous cessé de battre votre mari ? ». Que la question soit biaisée se traduit par le fait suivant : notre interlocutrice a le sentiment de pouvoir uniquement répondre par

$$\ulcorner\neg((p \wedge q) \vee (p \wedge \neg q))\urcorner(\text{i.e. ni } p \wedge q \text{ ni } p \wedge \neg q)$$

ou par

$$\ulcorner \text{ Pas d'engagement sur } (p \wedge q) \vee (p \wedge \neg q)\urcorner,$$

comme le lui autorise la règle S_1. Dans les deux cas, par R_3, nous en concluons que

$$(p \wedge q) \vee (p \wedge \neg q)$$

ne figure pas dans le stock des engagements de l'interlocutrice. Imaginons maintenant que la question ne soit pas biaisée, de sorte que l'auditrice puisse répondre soit par

$$\ulcorner p \wedge q\urcorner$$

soit par

$$\ulcorner p \wedge \neg q\urcorner,$$

comme le lui autorise la règle S_1. Dans les deux cas, le présupposé $(p \wedge q) \vee (p \wedge \neg q)$ est, en vertu de R_3, introduit dans le stock d'engagement de l'interlocutrice, comme escompté.

D'autres approches sont sans doute possibles. Par exemple, Belnap [24] et Åqvist [12] proposent de recourir plutôt à une logique érotétique (ou logique des questions). Voici comment, pour l'essentiel, ils nous proposent de traiter le problème des questions biaisées. On reformule tout d'abord la question originelle sous forme d'une question dépendant d'une condition. Ensuite, si cette condition n'est pas remplie, on considére que cette question n'appelle pas de réponse. Cela correspond plus ou moins au traitement de Hamblin. De même Hintikka [112] suggère-t-ild'utiliser la théorie des jeux sémantiques dont il est l'initiateur. Il propose également de l'appliquer à l'analyse de la pétition de principe.Il y aurait sans doute beaucoup à dire sur ces différentes approches. Nous préférons cependant laisser cette question pour une autre occasion. Nous souhaitions simplement présenter, dans ses grandes lignes, la cadre dialogique de Hamblin, ses motivations et certains des problèmes que ce formalisme soulève. Il est fondamental pour les interlocuteurs de pouvoir, en cours d'argumentation, revenir sur leurs engagements. Toutefois – et telle est l'objection que Walton et Krabbe [267, p.26-27] adressent à Hamblin, le modèle dialogique qu'il nous propose ne fournit aucun renseignement sur les façons par lesquelles le locuteur peut et doit modifier ses engagements. Walton et Krabbe apportent des éléments de réponse à cette question, qu'il sera intéressant de comparer avec celles qu'apportent les théories de la révision des croyances. Nous renvoyons à plus tard cette étude comparative[13].

L'approche de Hamblin soulève une seconde question. Pouvons-nous espérer rendre compte de toutes les espèces de paralogisme à l'aide d'un seul et unique modèle, dialogique ou autre ? Woods et Walton en doutent sérieusement. Explorant systématiquement diverses logiques non-classiques, ils défendent l'idée (à première vue, plausible) que, mettant en jeu une notion (causale, épistémique, ...) qui lui est propre, chaque type de paralogisme doit être modélisé à l'aide d'un système formel particulier. Ainsi, afin de comprendre la nature du paralogisme *post hoc*, ils s'aident de la logique de la connectibilité (*connectibility*) de Berder. Afin d'éclairer la nature éventuellement fallacieuse d'une argumentation par l'ignorance, ils utilisent une sémantique intuitionniste de type « Kripke ». Enfin, pour rendre compte de la circularité d'une argumentation, ils recourent aux outils de la théorie des graphes[14].Van Eemeren et Groodendorst [69, 71] s'inscrivent en faux contre cette approche, à laquelle ils reprochent de donner « uniquement des descriptions fragmentaires, et aucun aperçu général du domaine des paralogismes » [71, p. 119]. Ils élaborent une nouvelle méthode d'analyse, qu'ils nomment « pragma-dialectique », parce que combinant la théorie hamblienne du jeu dialectique et la théorie searlienne des actes de langage[15]. Leur idée est, en substance, d'interpréter chaque paralogisme comme la violation d'une des règles auxquelles la discussion critique est, selon eux, soumise. Nous pouvons nous demander si ces dernières n'ont plus avec les règles dialogiques de Ham-

[13]Cf. p. 76 et *sqq.*

[14]Nous renvoyons le lecteur intéressé à Woods et Walton [271, 272, 266].

[15]Sur la théorie des actes de langage, voir notre paragraphe 1.3.7, p. 41 et *sqq*.

blin que de lointaines ressemblances ; elles ne sont plus, à proprement parler, des règles *formelles*. Entreprendre un compte rendu plus détaillé de la théorie pragma-dialectique nous ferait sortir du cadre que nous nous sommes initialement fixé dans cette section (1.3).

1.3.4 La logique naturelle de Grize

Toute différente semble la direction en laquelle Grize s'est engagé. La théorie del'argumentation que Hamblin et ses successeurs proposent est normative, au sens où elle tente de dégager les critères nous permettant d'évaluer une argumentation donnée. La théorie que Grize propose est, non pas normative, mais descriptive. Celui-ci part du principe selon lequel, lorsqu'un locuteur A fait un discours, il propose à son interlocuteur B ce qu'il nomme une *schématisation*. Par là, Grize veut simplement dire que A donne à voir à B les représentations qu'il se fait du thème dont il traite, de A, de B et de la relation entre A et B. Grize assigne pour fonction à la logique de décrire le système des opérations (dites « logico-discursives ») qui permettent à un locuteur de construire une schématisation et à son interlocuteur de la reconstruire. Ainsi conçue, la logique se déploie à deux niveaux. Tout d'abord, il lui faut mettre en évidence les opérations qui permettent de construire des objets, de les déterminer progressivement et d'engendrer des énoncés. Ensuite, il lui faut rendre compte des activités qui conduisent à l'organisation des éléments ainsi construits.

1.3.5 Les opérations logico-discursives

Voici les principales :

1. les opérations d'ancrage α et η (ainsi que les opérations d'objets γ et θ),
2. l'opération de détermination δ,
3. l'opération de prise en charge σ
4. l'opération de configuration τ.

Présentons-les une à une. Un mot, tout d'abord, sur les opérations α et η, par l'application desquelles commence toute schématisation. La première opération ancre un objet (lui-même conçu comme une classe) et la seconde un prédicat dans ce que, reprenant la terminologie de Culioli, Grize nomme une « notion primitive » ou un « préconstruit culturel ». Soient X une notion primitive et x_1 un nom. Grize utilise en général l'expression

$$\alpha(X) = \{x_1\},$$

pour indiquer que le locuteur introduit une classe-objet particulière, qu'il nomme x_1 et qu'il suppose se rapporter au préconstruit culturel X. Par exemple,

$$\alpha(\text{MALADIE}) = \{\texttt{le malade}\}.$$

Soient $P()$ un prédicat à n places et $\overline{P}()$ le prédicat opposé. Il utilise de même l'expression

$$\eta(X) = (P(), \overline{P}())$$

ou, plus simplement, l'expression

$$\eta(X) = [P()],$$

pour indiquer que le locuteur introduit le couple de propriétés (P, non-P), qu'il suppose se rapporter au préconstruit culturel X. Par exemple,

$$\eta(\text{MALADIE}) = [\texttt{être malade}()].$$

Il est rare que le locuteur dise deux fois exactement la même chose d'un objet. Grize ici utilise la notation

$$[S()] = \{[P()], [E()][I()]\}$$

pour indiquer que l'argumentation introduit les couples de propriétés (P, non-P), (E, non-E) et (I, non-I) comme correspondant à trois modes différents du couple prédicatif (S, non-S). Grize dit de $[S()]$ qu'il est un *surprédicat*.

Grize reproche tout d'abord à la logique mathématique de ne pouvoir rendre compte du fait qu'une schématisation est sans cesse retravaillée en cours d'argumentation. Ainsi écrit-il : « Ici [en logique naturelle], objets et prédicats [] ne cessent d'être modifiés par les activités discursives qui portent sur eux » [35, p. 101]. Deux opérations en fait le permettent. La première est l'opération γ qui, appliquée à une classe-objet, l'enrichit d'ingrédients nouveaux. Par exemple :

$$\begin{aligned} \alpha(X) &= \{\texttt{le corps}\}, \\ \gamma(\{\texttt{le corps}\}) &= \{\texttt{le corps, les muscles}\}. \end{aligned}$$

La seconde est l'opération θ (dite « de spécification ») qui, appliquée à une classe-objet, l'enrichie de qualificatifs nouveaux. Par exemple :

$$\theta(\{\texttt{le corps}\}) = \{\texttt{le corps, la Machine}\}.$$

Il s'agit maintenant de relier entre eux les classes-objets et les couples prédicatifs. Pour ce faire, Grize introduit une polyopération δ dite « de détermination » , qu'il définit comme suit. Soient c une classe-objet et $[P()]$ un couple prédicatif. Nous avons :

$$\delta(c, [P()]) \quad = \quad \mu \pm P(c).$$

Ce résultat peut s'analyser en trois composantes. Tout d'abord, δ sélectionne l'un des deux termes du couple prédicatif, ce qu'indique l'expression $\pm P(.)$. Ensuite, δ instancie une classe-objet à la place libre du prédicat : $P(c)$. Enfin, la relation entre le prédicat et la classe-objet est modulée, ceci au moyen d'une modalité représentée par la lettre μ. Par exemple, on aura :

$$\delta \quad = \quad \text{Probable} + \texttt{Cancérigène}(\{\texttt{la fumée}\}).$$

Ici δ correspond au contenu de jugement « que la fumée être probablement cancérigène ». Naturellement, il peut arriver que μ soit vide, ou que δ ne choisisse pas entre les deux termes du couple prédicatif.

Reste à passer des déterminations aux énoncés. Pour ce faire, Grize introduit une opération σ, qu'il appelle « polyopération de prise en charge », et qui a deux effets :

1. préciser quel est le sujet énonciateur (Grize l'appelle aussi le témoin de l'énonciation) et quelle est sa source d'information ;
2. indiquer le type de prise en charge de la prédication par le sujet.

Supposons que, dans un texte argumentatif, nous lisions la phrase :

> « Le parti communiste français n'a pas changé ». [97, p. 234]

Cela correspond au cas de figure le plus simple – le seul que la logique mathématique envisage : n'indiquant pas quel est sa source d'information, le locuteur S_0 prend en charge une détermination, sans s'en distancier en rien. Grize reprend ici la notation de Frege ; il symbolise le contenu par une barre horizontale à gauche de $\mu \pm P(c)$, puis l'assertion par une barre verticale à gauche de la barre horizontale :

$$S_0 \vdash\!\!\!\!\text{———} \quad \mu \pm P(c)$$

Imaginons que nous rencontrions la phrase :

> « Le parti communiste français n'aurait pas changé ».

Nous sommes toujours dans le cas où le sujet énonciateur et la source d'information sont identiques. Supposons que le locuteur utilise le conditionnel pour émettre un certain doute quant à la vérité de ce qu'il rapporte. Nous avons ici affaire à une modalité, non plus *de re*, mais *de dicto*, dont la nature en fait peut être multiple. Du point de vue symbolique, Grize se contente d'inscrire celle-ci en toute lettre à l'intérieur de la barre de contenu, dans un espace blanc :

$$S_0 \vdash \texttt{Doute} \;—\; \mu \pm P(c)$$

Reste à présenter brièvement le cas où le locuteur fait explicitement appel à une source d'information. Ce cas de figure se reconnaît à la présence d'un verbe avec complétive :

« Darwin dit que l'homme descend du singe ». [98, p.179]

Grize traite « dit que » comme un prédicat binaire ayant pour premier argument « Darwin » et pour deuxième argument l'énoncé « l'homme descend du singe ». Dans le langage qui est le sien, il dit de l'opération σ de prise en charge qu'elle s'applique à deux déterminations : la première est $\texttt{Descendre}(h,s)$; la seconde est $\texttt{Dire}(d,\Delta)$, où Δ marque le renvoi à $\texttt{Descendre}(h,s)$. Le résultat de cette opération est un énoncé schématisé par :

$$S_0 \vdash \begin{array}{l} \texttt{Dire}(d,\Delta) \\ \Delta \\ \texttt{Descendre}(h,s) \end{array}$$

Les cinq opérations logico-discursives que nous venons de présenter relèvent du premier niveau d'analyse tout à l'heure évoqué : leur rôle est de rendre compte de la structure interne de nos énonciations. Pour rendre compte de leurs agencements, Grize [35, pp.130-136] introduit une sixième et dernière opération logico-discursive τ qui, admet-il, peut prendre de multiples valeurs : et, ou, si, mais, etc. Grize admet ainsi qu'il existe un grand nombre de connecteurs autres que les traditionnels opérateurs de la logique propositionnelle classique. Mais il ne nous dit pas qu'elles sont au juste leur limites, ni comment la logique naturelle surmonte ces dernières. Il nous indique seulement comment les retranscrire dans la notation de Frege.

1.3.6 Un exemple d'application

A plusieurs reprises, sans dissimuler les difficultés de son entreprise, Grize tente d'illustrer la façon d'appliquer les opérations logico-discursives sur des textes (voir e.g. [98, p. 181] et [35, p. 137]). Dans cette section, nous présentons un exemple d'application[16]. Personnellement, nous ne nous sentons pas suffisamment au clair sur les tenants et les aboutissants de la logique naturelle, pour pouvoir dire si elle éclaire véritablement la structure logique d'une argumentation.

[16] Le texte est donné dans Grize [98, p. 152].

> M. Poher nous avait dit que s'il était candidat, il abandonnerait l'intérim. Il ne l'a pas fait, c'est une première contradiction, mais enfi n passons ... Il nous dit maintenant que s'il ne l'a pas fait, c'est parce qu'il ne voulait pas laisser l'intérim à ce gouvernement que, a-t-il dit, il ne connaît que trop. Or, j'ai lu dans la presse, il y a trois jours, que M. Poher, présidant le dernier conseil des ministres, avait adressé ses félicitations et ses remerciements à ce même gouvernement. » (Discours de Georges Pompidou, *Le Monde*, 12 juin 1969)

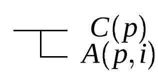
Reprenons une à une les cinq séquences de ce texte :

Séquence 1 : M. Poher nous avait dit que s'il était candidat, il abandonnerait l'interim.

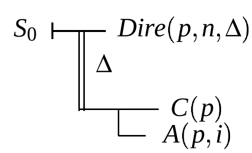
Cette séquence introduit trois classes-objets, qui sont `{M. Poher}` $=_{df}$ p, `{l'intérim}` $=_{df}$ i et `{nous}` $=_{df}$ n. Elle introduit également trois couples prédicatifs, qui sont `[Etre-candidat(.)]`$=_{df}$ $[C(.)]$ `[Abandonner(.)]` $=_{df}$ $[A(..)]$ et `[Dire()]` $=_{df}$ $[Dire(...)]$. Par δ, nous obtenons trois déterminations indépendantes les unes des autres :

1.	$C(p)$	que p être candidat
2.	$A(p,i)$	que p abandonner l'intérim
3.	$Dire(p,n,\Delta)$	que p dire à n que Δ

Par τ, nous obtenons ensuite le conditionnel :



L'application de σ au résultat ainsi obtenu ainsi qu'à la détermination $Dire(p,n,\Delta)$ conduit finalement à :



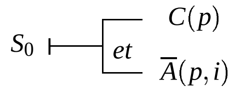
Nous voyons que, pour ce qui est du pronom anaphorique « il » qui figure dans « s'il était candidat, il abandonnerait l'intérim », la traduction utilise la lettre p, qu'on assimilera volontiers à une constante d'individu[17].

Séquence 2 : Il ne l'a pas fait, c'est une première contradiction mais enfi n passons.

Voici la traduction que donne Grize :

---

[17]Intuitivement, le pronom s'apparente plus à une variable individuelle, puisqu'il renvoie à ce qui a été antérieurement nommé. La théorie des représentations discursives de Kamp prend pour point de départ l'idée selon laquelle l'assignation d'une valeur à cette variable n'est pas toujours-déjà-faite (ce qu'un calcul ordinaire des prédicats présuppose), mais se construit pas à pas dans le discours. Nous renvoyons le lecteur intéressé par ce thème à Gamut [81].



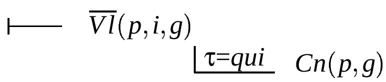
La reprise « c'est une première contradiction mais enfin passons » n'est pas traduite.

Séquence 3 : Il nous dit maintenant que, s'il ne l'a pas fait, c'est qu'il ne voulait pas laisser l'intérim à ce gouvernement, qu'il ne connaît que trop.

Procédons par étape :

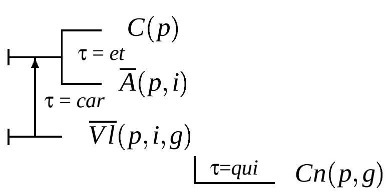
Sous-séquence 1 : Il ne voulait pas laisser l'intérim à ce gouvernement, qu'il ne connaît que trop.

Grize [35, p. 134] propose de traiter la relative explicative « qu'il ne connaît que trop » de la façon suivante. Posons : « que Poher ne pas vouloir laisser l'intérim à ce gouvernement » $=_{df}\overline{VL}(p,i,g)$ et « que Poher ne connaître que trop le gouvernement » $=_{df}Cn(p,g)$. Le fait que $Cn(p,g)$ soit une relative qui explique $\overline{Vl}(p,i,g)$ est symbolisé ainsi :

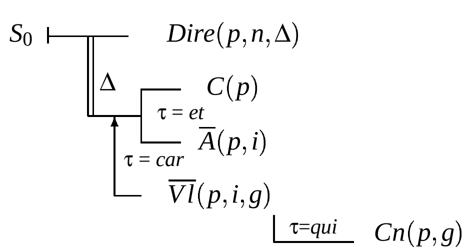


Sous-séquence 2 : S'il l'a fait, c'est qu'il ne voulait pas laisser l'intérim à ce gouvernement, qu'il ne connaît que trop.

L'énoncé « il ne voulait pas laisser l'intérim à ce gouvernement » sert ici d'explication à l'énoncé « il l'a fait ». Grize [35, p.136] suggère de l'indiquer par une simple flèche reliant le premier énoncé au second. Nous avons donc :

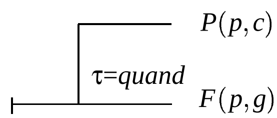


Finalement, la séquence tout entière devient :

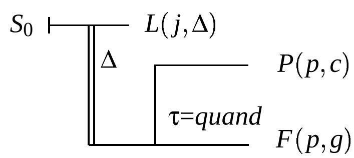


Séquence 4 : Or, j'ai lu dans la presse, il y a trois jours, que M. Poher, présidant le dernier conseil des ministres, avait adressé ses félicitations et ses remerciements à ce même gouvernement.

La proposition circonstancielle « lorsqu'il présidait le dernier conseil des ministres » relève de l'opération τ. Posons : « que Poher présider le dernier conseil des ministres » $=_{df} P(p,c)$; « que Poher féliciter le gouvernement » $=_{df} F(p,g)$. Grize traduit notre proposition circonstancielle par :



Prenons : $j =_{df}$ {je } ; « que je lis dans la presse que » $=_{df} L(j,\Delta)$. Nous obtenons alors pour l'énoncé (5) la représentation suivante :



Il serait bien prétentieux, en guise de conclusion, de porter un jugement d'ensemble sur la logique naturelle de Grize ; elle est encore embryonnaire. Les opérations logico-discursives qu'elle étudie évoquent une série de concepts sur lesquels la logique traditionnelle travaille déjà depuis un moment. Quoique Grize en dise, la proximité des deux types de logique est évidente ; leur confrontation devrait leur être mutuellement bénéfique. Nous pourrions ici regretter que Grize ne dise mot sur certains problèmes que soulève l'application brutale de la logique des prédicats à l'analyse du langage naturel. Dans le cas de la représentation des pronoms anaphoriques, ceci s'explique par le fait que ce problème était encore peu évoqué[18]. Le texte date de 1983. Apportant une solution au problème, la théorie des DRS de Kamp et le calcul dynamique des prédicats de Groenendijk/Stokhof n'existaient pas ou n'étaient que peu connus. Ces travaux datent respectivement de 1981 et de 1991.

1.3.7 Logique de l'illocutoire

Toutes les approches formelles de l'argumentation que nous avons rencontrées laissent une impression d'inachevé, du fait de leur manque apparent de systématicité. Nous nous tournons maintenant vers une théorie qui semble faire exception. Il s'agit de la logique de l'illocutoire, dont la présentation formelle fait l'objet du volume II de *Meaning and Speach Acts*, que D. Vanderveken publia en 1991.

[18]Je remercie le Professeur Paul Gochet pour cette remarque.

Vanderveken part de l'observation, en elle-même banale, qu'une logique de l'illocutoire ne saurait se réduire à une théorie des fonctions de vérité : ce que l'on accomplit, en disant quelque chose, peut être l'acte de poser une question ou celui de donner un ordre. Diverses tentatives ont été faites pour construire des systèmes formels s'appliquant à de telles expressions. L'originalité de Vanderveken est ici de chercher à développer une théorie logique qui, contrairement à celles que proposent ses prédécesseurs, ne se limite pas à un type particulier d'acte de langage. Pour mieux souligner l'ambition de son programme, il qualifie de « générale » la sémantique qu'il nous propose. Il s'agit d'une extension de la logique intensionnelle de Montague [173, 174]. La question se pose de savoir ce qu'une théorie de l'argumentation peut gagner dans cette entreprise. Voici comment Vanderveken répondrait. Aux tenants d'une conception « inférentielle » de l'argumentation, il est a priori tentant d'objecter qu'il existe de nombreux cas où la relation argumentative intervient, non pas entre deux contenus propositionnels, mais entre deux actes illocutoires. Telle est l'objection d'Anscombre et Ducrot [10, p. 11], qui donnent l'exemple du dialogue suivant :

A : Tiens, qu'est-ce qu'il devient, Pierre ?
B : Tu t'intéresses donc à Pierre ?

Si Vanderveken parvient à mener son programme à terme, il aura par là même montré combien ce type d'objection est discutable. Nous pouvons néanmoins nous demander si le détour par une logique de l'illocutoire est nécessaire : dans la mesure où l'enchaînement peut être dit opérer sur des actes, il n'y a peut-être aucune raison de contester son appartenance à l'apophantique.

Les actes illocutoires sont des actions intentionnelles. Comme c'est le cas pour les autres actions humaines, toute tentative de les accomplir peut réussir ou échouer. Par exemple, si un sujet dit à son souverain « Je t'ordonne de sortir », sa tentative d'ordre sera un échec. Aussi, à la notion de conditions de vérité, la logique de l'illocutoire substitue la notion de conditions de succès d'un acte de langage. Elles désignent les conditions qui doivent être réunies pour qu'un acte illocutoire soit réussi. Dans *Meaning and Speach Acts*, Vanderveken limite son étude aux actes illocutoires élémentaires[19]. Voici comment, en substance, il les analyse. A l'instar de Searle, il suppose chacun d'eux constitué d'une force illocutoire F et d'un contenu propositionnel p. Puis il décompose la force illocutoire en six éléments, que nous rappelons brièvement :

1. Le but illocutoire. Vanderveken part du principe qu'il en existe cinq primitifs. Un tel but est soit assertif (représenter comme actuel un état de choses), soit commissif (s'engager à une action future), soit directif (amener autrui à faire quelque chose), soit déclaratif (accomplir une action par le seul fait

[19] Pour un essai d'analyse des actes illocutoires complexes, voir Searle et Vanderveken [229], qui introduisent trois opérateurs, ceux de dénégation, de conjonction et d'implication illocutoires.

de l'énoncer), soit expressif (communiquer un état mental). C'est le composant principal qui détermine les conditions de succès de l'énoncé. Les buts illocutoires renvoient chacun à différentes directions d'ajustement entre les mots et le monde. Ainsi, le but assertif détermine un ajustement des mots au monde, les buts commissif et directif déterminent un ajustement du monde aux mots, le but déclaratif détermine un double ajustement, et but expressif un ajustement vide[20].

2. Le mode d'accomplissement. Il s'agit, tout simplement, du mode sous lequel le but est accompli. Par exemple, le but d'une prière (*request*) est directif ; sa particularité, nous disent Searle et Vanderveken [229, p. 199], est d'être accomplie en laissant une option de refus à celui à qui on s'adresse.

3. La condition sur le contenu propositionnel. Il s'agit de la condition que le contenu propositionnel doit satisfaire, pour que le but puisse être atteint avec succès :

But	Condition sur le contenu propositionnel
Assertif	–
Commissif	p décrit une action future
Directif	p décrit une action future
Déclaratif	–
Expressif	–

4. La condition préparatoire, qui indique ce que la réussite de chaque acte « présuppose » :

But	Condition préparatoire
Assertif	le locuteur a des raisons de croire que p est vrai
Commissif	le locuteur est capable d'accomplir l'action à laquelle il s'engage
Directif	l'auditeur est capable d'accomplir l'action qui lui est imposée
Déclaratif	le locuteur est capable de produire l'état de choses qu'il cherche à produire
Expressif	–

5. La condition de sincérité. En général, dans l'accomplissement de tout acte illocutoire avec un contenu propositionnel, le locuteur exprime une certaine attitude à l'égard de ce contenu propositionnel. Ceci correspond à la condition de sincérité. Nous avons :

[20] L'auteur reprend les analyses de Searle. Une approche différente est proposée par Livet [148]. Suggérant d'analyser un but au moyen de la paire proposition de révision/annulation de révision, il substitue aux quatres buts élémentaires les huit couples suivants : question/réponse, reproche/excuse, ordre/obéissance, soupçon/promesse, incertitude/affi rmation, mise en cause/déclaration (devant un tiers), proposition/refus et lamentation/encouragement.

But	Conditions de sincérité
Assertif	croyance que *p*
Commissif	intention que *p*
Directif	désir que *p*
Déclaratif	désir que *p*
	croyance que déclarer que *p* produira *p*
Expressif	–

6. Le degré de force de la condition de sincérité. Par exemple, « demander » et « implorer » ont la même condition de sincérité, mais pas le même degré de force.

Ainsi, un acte de langage devient un sixtuplet, dont il reste à caractériser formellement chacun des éléments. Pour se faire, Vanderveken se tourne, nous l'avons dit, vers le système de Montague. Il s'agissait essentiellement pour Montague de montrer qu'il est possible de définir une sémantique formelle pour les langues naturelles, selon le principe suivant :

$$\begin{tikzcd}
\text{Langage } L_1 \arrow[d, "\textit{traduction}"] \arrow[dr, "\textit{fonction interprétative induite}"] & \\
\text{Langage } L_2 \arrow[r, "\textit{fonction}", "\textit{interprétative}"'] & \text{univers d'interprétation}
\end{tikzcd}$$

Tout d'abord, on définit une traduction de L_1 (fragment d'une langue naturelle) vers L_2 (logique intensionnelle), Ensuite, on montre comment cette traduction et la sémantique associée à L_2 induisent une sémantique pour L_1. Vanderveken tente de réitérer le procédé, pour la langue naturelle en tant qu'objet d'énonciation. Dans ce qui suit, nous nous intéresserons à ce moment préliminaire qui consiste à définir le langage de la logique intensionnelle de l'illocutoire, et à lui associer une sémantique ainsi qu'une axiomatique.

Il peut être utile de brièvement décrire la logique intensionnelle de Montague[21]. Il s'agit d'une logique modale d'ordre supérieur avec deux types primitifs : le type *e* des entités individuelles et le type *t* des valeurs de vérité. Les types satisfont aux clauses : *e* et *t* sont des types ; si α et β sont des types alors $<\alpha,\beta>$ est un type ; si α est un type alors $<\#,\alpha>$ est un type. Le type $<\alpha,\beta>$ correspond à une fonction dont les arguments sont des éléments de type α et dont le résulat sera un élément de type β. Par exemple, *court* est de type $<e,t>$. L'élément # n'est pas, lui, pris comme un type. Sur un univers d'interprétation, il correspond à un ensemble *I* de mondes possibles. Le type $<\#,\alpha>$ correspond donc à une fonction dont les arguments sont des mondes possibles et donnant comme valeur

[21] Pour un exposé plus complet, voir Gamut [81].

un élément de type α. Le couple $<\#, \alpha>$ désigne ainsi le type des entités intensionnelles. Puis on se donne un alphabet essentiellement constitué de :

– constantes individuelles et prédicatives
– variables d'individus et de prédicats
– opérateurs : $\Box$ (*nécessairement*) ; λ (*lambda*) ; $^\vee$ (*extensionalisation*) ; $^\wedge$ (*intensionalisation*).

Traitant les constantes et les variables comme termes premiers, on engendre les termes complexes à l'aide des clauses :

– si A est un terme de type $<\alpha,\beta>$ et si B est un terme de type α, alors $A(B)$ est un terme de type β
– si x et A sont des termes de type (respectivement) α et β, alors λxA est un terme de type $<\alpha,\beta>$ (une expression de la forme λxA est lue « la fonction qui à x associe A »)
– si A est un terme de type α, alors $^\wedge A$ (« intensionalisation de A ») est un terme de type $<\#,\alpha>$
– si A est un terme de type $<\#,\alpha>$, alors $^\vee A$ (« extensionalisation de A ») est un terme de type α.

Nous avons une sémantique en termes de mondes possibles. En gros, on se donne un ensemble d'indices I et un ensemble d'individus D, puis on définit les dénotations relatives aux objets de différents types. Les conditions de récurrence pour les formules épousent la forme traditionnelle. La notion de validité d'une formule est elle aussi définie de façon habituelle.

1.3.8 Le langage de la logique de l'illocutoire

Type. Tout d'abord, Vanderveken redéfinit la notion de type élémentaire comme suit. Au type e des entités individuelles et au type t des valeurs de vérité, il ajoute le type s des valeurs de succès et le type a des propositions atomiques. La définition récursive de l'ensemble T des types reste, dans son principe, similaire à celle que proposait Montague. T est le plus petit ensemble satisfaisant les trois conditions suivantes :

1. $e, t, s, a \in$T
2. si $\alpha,\beta \in$ T, $<\alpha,\beta> \in$T
3. si $\alpha \in$T, alors $<\#,\alpha> \in$T

Comme chez Montague, # n'est pas, lui, pris comme un type. Sur un univers d'interprétation, il correspond à un ensemble I de contextes. On nous demandera quel type assigner aux principaux concepts auxquels a recours la logique de l'illocutoire. Le tableau suivant donne quelques exemples :

CONCEPT	TYPE	ABRÉVIATION
Acte illocutoire	$<p,<<\#,s>,t>>$	Ω
But illocutoire	$<\#,<p,s>>$	μ
Conditions préparatoires	$<\#,<p,<p,t>>>$	Σ

Exemples de types dérivés (d'après Vanderveken [261, p. 77])

Parcourons ce tableau. Nous voyons que la première ligne mentionne la lettre t. Cela ne doit pas nous troubler. Soient S la valeur du succès et $\bar{S}$ la valeur de l'insuccès. Une expression du type $<<\#,s>,t>$ dénote en fait (la fonction caractéristique d') un ensemble de fonctions définies dans I et à valeurs dans $\{S,\bar{S}\}$ [22]. Ainsi, dire d'un acte illocutoire (*Je t'ordonne de sortir*) qu'il est de type $<p,<<\#,s>,t>>$ revient à le traiter comme une fonction qui, au contenu propositionnel p (*que tu sortes*), associe une fonction f qui prend pour argument un contexte et pour valeur S ou $\bar{S}$. La deuxième ligne dit qu'un but illocutoire est de type $<\#,<p,s>>$ et qu'il est analysé comme une fonction qui, à un contexte d'énonciation, associe une fonction de l'ensemble des propositions dans l'ensemble des valeurs de succès. La troisième ligne dit qu'une condition préparatoire est de type $<\#,<p,<p,t>>>$. Elle est ainsi assimilée à une fonction associant à un contexte et à un contenu propositionnel un ensemble d'autres contenus propositionnels, en l'occurrence ceux que l'atteinte du but présuppose. Il revient à l'utilisateur de spécifier leur nature.

Pour ce qui est des autres types dérivés, nous nous contenterons d'indiquer l'abréviation que Vanderveken utilise pour chacun :

CONCEPT	ABRÉVIATION
Force illocutoire	Φ
Proposition atomique	γ
Proposition	p
Conditions sur le contenu propositionnel	θ
Entier	ι
Mode psychologique	τ
Attitude propositionnelle	ξ
Conditions de sincérité	Ψ

Symboles primitifs. A tout type est associé un ensemble infini de variables ainsi qu'un ensemble infini de constantes non-logiques. Une variable de type α est notée x_α et une constante non-logique de type α est notée c_α. Les symboles syncatégorématiques sont ceux de la logique intensionnelle : =, λ, $^\wedge$ (*intensionalisation*),

[22]En théorie des types, il est d'usage de parler en termes de fonction caractéristique. Par exemple, on dit d'un prédicat unaire qu'il est de type $<e,t>$ et qu'il dénote (la fonction caractéristique d') un ensemble d'individus – ceux-là même auxquels le prédicat s'applique. Comme l'explique Gamut [81, p. 83-84], les locutions «fonction caractéristique de l'ensemble A » et « ensemble A» sont en fait interchangeables.

$^{\vee}$ (*extensionalisation*). Les constantes logiques sont :

- la fonction 1_s, qui nomme le succès ;
- la fonction $\{\}_{ts}$, dont le rôle est de nous faire passer des valeurs de vérité aux valeurs de succès correspondantes ;
- la constante relationnelle $\simeq_{a(at)}$, qui s'applique à deux propositions, lorsque celles-ci ont les mêmes constituants atomiques ;
- la constante $[\ ]_{a(\#t)}$; elle symbolise la fonction qui associe à tel contenu propositionnel atomique ses conditions de vérité (relativement à un contexte) ;
- les constantes π^1_μ, π^2_μ, π^3_μ et π^4_μ, qui désignent respectivement le but illocutoire assertif, engageant, directif et déclaratif ; nous allons voir tout de suite que le but expressif est introduit à l'aide d'une définition ;
- la constante $\gg_{pt}$, qui énumère les présupposés d'un acte ;
- la constante $E_{\xi(\iota s)}$, qui symbolise la propriété d'exprimer tel ou tel état psychologique ; Vanderveken définit le but expressif au moyen de cette constante ; il pose $\pi^5 = {}^{\wedge}(\lambda x_p\{\exists x_\tau\ \exists x_\iota\ (E(x_\tau x_p)x_\iota = 1_s)\})$. S'aidant éventuellement des règles définies ci-après, on vérifiera aisément que cette expression est bien du type des buts illocutoires, à savoir $<\#,<p,s>>$.

Termes. On définit l'ensemble Tm_α des termes de type α au moyen des clauses suivantes :

- Toute variable de type α figure dans Tm_α
- Toute constante de type α figure dans Tm_α
- $A \in \mathrm{Tm}_{<\alpha,\beta>}, B \in \mathrm{Tm}_\alpha \rightarrow A(B) \in \mathrm{Tm}_\beta$
- $A, B \in \mathrm{Tm}_\alpha, \rightarrow (A = B) \in \mathrm{Tm}_t$
- $x \in \mathrm{Tm}_\alpha, A \in \mathrm{Tm}_\beta \rightarrow \lambda x A \in \mathrm{Tm}_{<\alpha,\beta>}$
- $A \in \mathrm{Tm}_\alpha \rightarrow {}^{\wedge}A \in \mathrm{Tm}_{<\#,\alpha>}$
- $A \in \mathrm{Tm}_{<\#,\alpha>} \rightarrow {}^{\vee}A \in \mathrm{Tm}_\alpha$

1.3.9 Sémantique

Les ensembles primitifs. Ils sont au nombre de quatre : l'ensemble des entités individuelles, D ; l'ensemble $\{T, \bar{T}\}$ – où T désigne le vrai et $\bar{T}$ désigne le faux ; l'ensemble $\{S, \bar{S}\}$ – où S désigne le succès et $\bar{S}$ désigne l'insuccès ; l'ensemble des contextes d'énonciations, I.

Ensembles de dénotation. Pour tout type α, l'ensemble des dénotations de type α, noté D_α, est défini par :

1. $D_t = \{T, \bar{T}\}$;
2. $D_s = \{S, \bar{S}\}$;
3. $D_e = D$;

4. $D_{<\alpha,\beta>} = (D_\beta)^{D_\alpha}$ (ensemble des fonctions de D_α dans D_β).
5. $D_{<\#,\alpha>} = (D_\alpha)^I$ (ensemble des fonctions de I dans D_α).

La réunion de tous les ensembles de dénotation constitue l'univers d'interprétation. Seule la clause 2 est propre à la logique de l'illocutoire. Pour plus de simplicité, nous passons sous silence une autre clause propre à l'illocutoire. Il s'agit de la règle qui spécifie la dénotation des constituants propositionnels d'un énoncé.

Fonctions interprétatives élémentaires. Nous appelons ainsi une fonction qui détermine les entités interprétatives à associer aux constantes de différents types. Au niveau de l'illocutoire, Vanderveken introduit tout d'abord quatre fonctions

$$\Pi_1, \Pi_2, \Pi_3, \Pi_4 \in (\{S, \bar{S}\}^{D_p})^I.$$

Elles associent à chaque contexte i une fonction qui associe à son tour à chaque proposition soit la valeur S, soit la valeur $\bar{S}$. Dans le cas du but assertif, Vanderveken fait les hypothèses suivantes :

> Pour que $\Pi_1(i,P) = S$, il faut que P soit membre d'un ensemble X de propositions vérifiant les trois propriétés suivantes : X est minimalement consistant ; X est clos par rapport à ce que l'auteur nomme l'implication « forte » ; X contient un supremum unique par rapport à cette dernière. (Par « implication forte », nous devons entendre ceci : P implique fortement Q si toute évaluation qui rend P vrai rend aussi Q vrai, et si l'ensemble des propositions atomiques qui figurent dans le contenu de Q est inclus dans l'ensemble des propositions atomiques qui figurent dans le contenu de P.).

Vanderveken exige de Π_2, Π_3 et Π_4 qu'elles vérifient une condition quasi identique. Ensuite, deux autres éléments sont introduits. Le premier, $\rangle\rangle \subseteq I \times D_p$, est le «corrélat» sémantique de la constante $\gg$. $(i,P) \in \rangle\rangle$ est lu «la proposition P est présupposée dans le contexte i». Vanderveken suppose $\{P : (i,P) \in \rangle\rangle\}$ clos pour l'implication forte. Le deuxième élément, $\mathbb{E} : D_\xi \rightarrow \{S, \bar{S}\}^{\mathbb{N}}$, est le corrélat sémantique de la constante E.

Assignation de dénotations aux formules. Etant donnés une assignation de valeur σ aux variables et un contexte i, nous pouvons associer à chaque formule A de type α sa dénotation $\|A\|_i^\sigma$ dans i de la façon suivante :

1. si x est une variable de type α, alors $\|x_\alpha\|_i^\sigma = \sigma(x_\alpha)$;
2. si c est une constante de type α, alors $\|c_\alpha\|_i^\sigma = I(c_\alpha)(i)$ – où I désigne une fonction d'interprétation habituelle assignant à toute constante une dénotation de même type ;
3. $||1_s||_i^\sigma = S$;
4. $||\{\}_{ts}||_i^\sigma(u_t) = S$ si et seulement si $u_t = T$;
5. $||A_{\alpha\beta}B_\alpha||_i^\sigma = ||A_{\alpha\beta}||_i^\sigma(||B_\alpha||_i^\sigma)$;

6. $||\lambda x_\alpha A_\beta||_i^\sigma =$ la fonction $h \in D_\beta^{D_\alpha}$ telle que, pour tout $d \in D_\alpha, h(d) = ||A_\beta||_i^{\sigma'[x/d]}$ – où $\sigma'[x/d]$ désigne l'assignation qui diffère au plus de σ en ce que $\sigma'(x) = d$;
7. $||^\wedge D_\alpha||_i^\sigma =$ la fonction $h \in (D_\alpha)^I$ telle que pour tout $i' \in I$, $h(i') = ||A_\alpha||_{i'}^\sigma$;
8. $||^\vee A_\alpha||_i^\sigma = ||A_\alpha||_i^\sigma(i)$;
9. $||A_\alpha = B_\alpha||_i^\sigma = T$ ssi $||A_\alpha||_i^\sigma = ||B_\alpha||_i^\sigma$;
10. $||\pi^k||_i^\sigma = \Pi_k(i)$, pour chaque k tel que $1 \leq k \leq 4$ (i.e. la valeur de Π_k pour l'argument i) ;
11. $|| \gg ||_i^\sigma = \rangle\rangle$;
12. $||E||_i^\sigma = \mathbb{E}$.

Le concept de validité logique se laisse définir assez naturellement. Supposons que nous caractérisions un modèle $\mathfrak{M}$ par la donnée des éléments suivants : un ensemble E d'individus ; un ensemble I de contextes ; une fonction I d'interprétation des constantes non-logiques ; la famille de fonctions $\{\Pi_k | 1 \leq k \leq 4\}$; l'entité $\rangle\rangle$; la fonction $\mathbb{E}$. Nous pouvons considérer une formule A_t comme valide si, quels que soient $\mathfrak{M}$ et l'assignation faite aux variables de A_t, nous avons $||A_t||_i^\sigma = T$ pour tous $i \in I$.

1.3.10 L'axiomatique

Pour obtenir l'axiomatique du système, il faut partir de l'axiomatique de la logique intensionnelle[23] et y adjoindre vingt nouvelles lois primitives, qui se répartissent en cinq catégories. Les deux premiers nouveaux axiomes concernent les valeurs du succès et de l'insuccès. Ils utilisent les constantes 1_t (*vrai*), 0_t (*faux*), 0_s (*insuccès*), qui sont des abréviations de (respectivement) $\lambda x_t x_t = \lambda x_t x_t$, $\lambda x_t x_t = \lambda x_t 1_t$ et $\{0_t\}$. Voici ces deux axiomes :

<u>Succès - insuccès</u>

$$\{1_t\} = 1_s \qquad \text{(Axiome 1)}$$

$$\sim (x_s = 1_s) = (x_s = 0_s) \qquad \text{(Axiome 2)}$$

Il est regrettable que Vanderveken n'explique pas la signification du premier axiome, sur laquelle nous butons. L'axiome 2 nous paraît intéressant. Il stipule que le succès d'un acte de langage n'est pas susceptible de degré. Cela n'implique pas que la satisfaction de cet acte ne soit pas susceptible de degré. Vanderveken [261, p. 102] prend pour exemple l'ordre d'apporter beaucoup de fleurs et la promesse de vérifier quelque chose la plupart du temps. La question de savoir si leurs conditions

[23] voir par exemple Gallin [80].

de félicités sont toutes remplies appelle, admettra-t-on sans trop de difficultés, une réponse en oui-non. Toutefois, certaines logiques non-monotones nous invitent à distinguer trois sortes de « tailles ». Il est clair que, si la quantité de fleurs apportées est petite, l'ordre n'est pas satisfait mais que, si elle est grande, l'ordre est satisfait. Serait-il absurde de dire que, si la quantité de fleurs apportées est de taille moyenne, ni l'ordre n'est satisfait, ni il n'est pas satisfait ? Nous ne chercherons pas à développer cette idée [24]. Le deuxième groupe d'axiomes porte sur la notion de proposition atomique. L'expression "$A_a \simeq B_a$" est lue "A_a et B_a ont les mêmes constituants atomiques"[25].

<u>Propositions atomiques</u>

$$A_a \simeq A_a \tag{Axiome 3}$$
$$A_a \simeq B_a \to B_a \simeq A_a \tag{Axiome 4}$$
$$A_a \simeq B_a \to (B_a \simeq C_a \to A_a \simeq C_a) \tag{Axiome 5}$$

Nous voyons que la relation est réflexive (axiome 3), symétrique (axiome 4) et transitive (axiome 5)[26]. Voyons la troisième catégorie d'axiomes, qui concerne le but illocutoire. Elle comprend huit lois. Les deux premières s'appliquent indifféremment aux quatre buts primitifs. Elles stipulent que, pour que ces buts soient atteints avec succès, il faut qu'ils portent sur une proposition (axiome 7) qui n'est pas contradictoire (axiome 8) :

<u>Buts illocutoires</u>

$$({}^{\vee}\pi^k A_p = 1_s) \to \exists! x_\gamma \, \exists! x_{\gamma t} \; A_p x_\gamma x_{\gamma t} \text{ avec } 1 \leq k \leq 4 \tag{Axiome 7}$$
$$({}^{\vee}\pi^k A_p = 1_s) \to \sim (A_p = (A_p \wedge \sim A_p)) \text{ avec } 1 \leq k \leq 4 \tag{Axiome 8}$$

L'axiome suivant concerne le seul but assertif :

$$({}^{\vee}\pi^1 A_p = 1_s) \to (\exists! x_p \, \forall y_p \, ({}^{\vee}\pi^1 y_p = 1_s) \leftrightarrow (x_p \rightarrowtail y_p)) \tag{Axiome 9}$$

Le symbole $\rightarrowtail$ y désigne la relation d'implication forte que nous avons évoquée tout à l'heure : pour que P implique fortement Q, il faut que les constituants propositionnels de Q soient aussi ceux de P. Nous pouvons interpréter cet axiome comme affirmant un principe de clôture du but assertif sous $\rightarrowtail$ (si nous remplaçons x_p par A_p). Les axiomes (10) et (11) disent la même chose des buts commissif, directif et déclaratif. La loi (19) sera isomorphe aux trois précédentes, mais elle concernera le but expressif. Ce groupe d'axiomes vient (si l'on veut) rendre

[24] D'autant plus qu'il faudrait aussi tenir compte de l'existence ou non d'un lien causal entre l'ordre donné (la promesse faite) et l'action accomplie.

[25] Par simplicité, nous avons passé sous silence la clause de vérité de $\simeq$, ainsi que celle de [...].

[26] Nous passons sur l'axiome 6.

compte du fait que nous pouvons accomplir un acte par implication, c'est-à-dire à travers un autre acte. Il est possible d'illustrer l'intérêt de recourir à une implication forte, plutôt qu'à un implication classique, à partir du paradoxe de Ross[27] : en principe, une énonciation de la forme « Je t'ordonne de poster cette lettre » ne nous engage pas à une énonciation de la forme « Je t'ordonne de poster cette lettre ou de la brûler ». Néanmoins, nous avons rencontré tout à l'heure un exemple qui ici poserait peut-être un problème pour l'implication forte[28]. Il s'agit de la phrase

Si tu me donnes un peu de ta glace, je te donne un peu de la mienne

qui, intuitivement, tend à faire entendre une double implication. Ici, nous avons simultanément « Je suggère que : q si p » et « Je suggère que : p si q ». Cependant, il est inexact de dire qu'il existe une relation d'implication forte entre le contenu des deux actes.

Voici les trois dernières lois concernant la notion de but illocutoire :

$$\pi^k \ddagger \pi^1 = \pi^k \text{ avec } k = 2 \text{ ou } 4 \qquad \text{(Axiome 12)}$$

$$({}^{\vee}\pi^4 A_p = 1_s) \rightarrow ({}^{\vee}A_p = 1_t) \qquad \text{(Axiome 13)}$$

$$\pi^5 \ddagger \pi^k = \pi^5 \text{ avec } 1 \leq k \leq 4 \qquad \text{(Axiome 14)}$$

Le symbole ‡y désigne un opérateur de conjonction des buts illocutoires, défini par :

$$A_\mu \ddagger B_\mu =_{def} \iota x_\mu \exists y_\mu \exists z_\mu \, (y_\mu = A_\mu \wedge z_\mu = B_\mu \wedge \\ \Box \forall x_p \, ({}^{\vee}x_\mu x_p = 1_s) \leftrightarrow ({}^{\vee}y_\mu x_p = 1_s \wedge {}^{\vee}z_\mu x_p = 1_s))$$

La loi (12) affirme l'existence implicite d'un but assertif en tout but commissif ou déclaratif. En vertu de cette loi, lorsque nous promettons de venir ou que nous donnons notre démission, notre propos est aussi de faire savoir que nous viendrons ou de faire savoir que nous démissionnons. Cette loi ne s'applique visiblement pas à la classe des directifs, par exemple à « questionner ». Cette asymétrie semble relativement plausible. Toutefois, il nous paraît intéressant de remarquer qu'une seule et même énonciation peut parfois mettre en jeu des buts illocutoires distincts, comme dans :

John, qui n'a jamais rien compris aux mathématiques élémentaires, a-t-il réellement l'intention de préparer un Doctorat de Mathématiques ?

Ici, une assertive (John n'a jamais rien compris aux mathématiques élémentaires) est enchâssée dans une interrogative. Nous pouvons nous demander si de tels enchâssements ne rendent pas (12) discutable. De son côté, l'axiome (13) précise

[27] Ce paradoxe est emprunté aux logiques déontiques. Voir ci-dessous p. 119.

[28] Voir ci-dessus, p. 20.

que, pour que le but déclaratif soit atteint avec succès, il est nécessaire que son contenu propositionnel soit vrai. En effet, la caractéristique de la classe des déclaratifs est que l'accomplissement réussi de l'acte en question instaure la correspondance voulue entre le contenu propositionnel et le monde. Par exemple, si j'accomplis comme il faut l'acte de déclarer la guerre, alors la guerre est déclarée. Enfin, l'axiome (14) rend le but expressif indépendant des quatre autres buts illocutoires[29]. Nous voyons ici apparaître la raison pour laquelle Vanderveken utilise des types. Nous avons apparemment besoin de ceux-ci, si nous souhaitons pouvoir introduire un axiome qui soit spécifique à un but illocutoire donné.

La troisième et avant-dernière catégorie d'axiomes concerne la notion de présupposé :

Présupposés

$$\sim (\forall x_p)(\gg x_p) \qquad \text{(Axiome 15)}$$

$$\gg A_p \rightarrow ((A_p \rightarrowtail B_p) \rightarrow \gg B_p) \qquad \text{(Axiome 16)}$$

$$\gg A_p \rightarrow \neg(\pi^1 \neg A_p) \qquad \text{(Axiome 17)}$$

L'axiome (15) signifie que l'ensemble des présupposés ne s'identifie pas à l'ensemble total des propositions. L'axiome (16) nous dit que "présuppose" est clos pour l'implication forte. L'axiome (17) stipule que le locuteur ne peut pas présupposer A_p et, dans le même temps, nier que A_p.

Le lecteur néophyte retirera sans doute de cette présentation l'impression que la théorie de Vanderveken est difficile, du fait de la complexité du formalisme utilisé. Cette complexité, du reste, est à l'image de l'ambition du programme : élaborer une sémantique d'ensemble des actes de langage. A l'attitude de Vanderveken, on opposera volontiers l'attitude des logiciens des modalités, qui en général préfèrent étudier chaque type d'actes indépendamment des autres – quitte à s'interroger, dans un deuxième temps, sur leur lien éventuel. Par exemple, il est possible de définir l'acte de poser une question de type « est-ce que p ? » (modalité dite *érotétique*) à partir de l'acte d'ordonner (modalité *déontique*). Le système d'Åqvist [12] ramène ainsi toute expression de la forme

est-ce que p ?

à une expression de la forme

Fais-en sorte que je sache si p ou si non-p !

L'attitude qui consiste à étudier chaque type d'actes indépendamment les uns des autres nous paraît plus prudente. L'analyse d'un impératif à elle seule pose des

[29]Cette interprétation de Vanderveken est un peu troublante, dans la mesure où (14) semble isomorphe à (12).

problèmes, sur lesquels nous reviendrons[30]. Quoi qu'il en soit de cette question, nous ne saurions terminer cette présentation sans indiquer les deux directions dans lesquelles Vanderveken s'engage, une fois l'axiomatique de son système dégagée. Dans un premier temps, l'auteur nous indique comment représenter les actes illocutoires primitifs et comment dériver à partir d'eux tous les autres actes illocutoires, en appliquant différentes opérations sur les composantes de l'acte. Dans un deuxième temps, Vanderveken nous présente les éléments d'une base de traduction d'un fragment de l'anglais sur le langage de la logique de l'illocutoire. Renvoyant à une autre occasion l'étude de ces développements, nous nous contentons d'indiquer une objection qui vient assez naturellement à l'esprit. Celle-ci consiste à faire remarquer que la théorie prend pour objet d'étude les seuls actes illocutoires interprétés littéralement, alors que nos argumentations quotidiennes font intervenir des actes de langages indirects. Par exemple, il arrive qu'une assertion comme

J'ai cassé le vase

ait le sens d'une prière implicite, du type

Je te prie de m'excuser.

Nous avons bien rencontré un groupe d'axiomes rendant compte de la possibilité que nous avons d'accomplir un acte par *implication*, c'est-à-dire en accomplissant un autre acte. Il s'agit des axiomes (9), (10), (11) et (19). Néanmoins, le rapport d'implication que chacun de ces axiomes envisage ne nous autorise pas à passer d'un registre (c'est-à-dire d'un but) illocutoire à l'autre. Ceci les rend inapplicables à notre exemple, où nous passons de l'assertif au directif. Voici très schématiquement comment Searle [228] décrit le mécanisme de dérivation d'un acte de langage indirect. Le destinataire d'une question comme *peux-tu me passer le sel ?* argumente avec lui-même sur les conditions de réussite de l'acte, qu'il met en relation avec les informations d'arrière-plan dont il dispose, au nombre desquelles figure sa croyance en la coopération conversationnelle de l'autre. Il resterait à voir quels réaménagements ou enrichissements la logique de l'illocutoire doit subir, pour être capable de rendre compte du mécanisme de la dérivation. [31] Il nous paraît intéressant de remarquer que Searle insiste sur le caractère probabiliste de la conclusion obtenue. La question se pose de savoir si, modélisant une notion de nécessité contextuelle et révisable, les logiques non-monotones ne nous seraient pas d'un certain secours.

[30] Cf. notre chapitre 3.

[31] La question des actes de langages indirects est en fait abordée par Vanderveken [262].

1.4 Conclusion

Le lecteur retirera de cette présentation l'impression confuse que les recherches dans le domaine de l'argumentation manquent, sinon d'unité, du moins de concertation. Cependant, nous pouvons déjà y dégager quelques fils directeurs très généraux. Par exemple, il est tentant de regrouper les théories de Anscombre/Ducrot, Grice et Searle/Vanderveken sous l'étiquette « théorie de l'argumentation du premier degré », et les théories de Perelman/Olbrechts-Tyteca et Hamblin sous l'étiquette « théorie de l'argumentation du deuxième dégré ». Une théorie du premier degré est l'étude des implications pragmatiques de nos énoncés. Anscombre/Ducrot expliquent celles-ci en termes de propriétés immanentes à la langue, Grice en termes de maximes conversationnelles et Searle/Vanderveken en termes de conditions de succès (actes de langage indirects mis à part). Une théorie de l'argumentation du deuxième degré, quant à elle, serait l'étude des moyens d'obtenir l'adhésion de son auditoire. Perelman/Olbrechts-Tyteca dégagent une série de schèmes argumentatifs, qui jouent soit sur des associations, soit sur des dissociations. Hamblin et ses disciples proposent une série de règles formelles à respecter au cours de la discussion critique. Une théorie du deuxième degré peut-elle faire l'économie d'une théorie de l'argumentation du premier degré ? L'exemple de la notion de présupposition suggère que non. En tant que telle, son analyse relève, dira-t-on, d'une théorie du premier degré. Reste qu'une théorie du second degré peut avoir à utiliser cette notion. En témoigne la théorie de Hamblin. Celle-ci rencontre cette notion, lorsqu'elle cherche à analyser le paralogisme de la question biaisée.

Au cours de cette présentation, nous avons pu constater que le thème de la non-monotonie était récurrent dans les études sur l'argumentation. Voici, pour mémoire, un bref récapitulatif des principaux points d'interface rencontrés : le quasi-syllogisme de Toulmin ; les maximes conversationnelles de Grice comme principes défaisables ; la possibilité de rendre compte des topos en termes de logique graduelle ; le thème de la rétraction des engagements dans la logique dialogique de Hamblin ; la dérivation d'un acte de langage indirect. Ayant ainsi remarqué la présence dans toutes les analyses d'une constante qui est soit évidente, soit sous-jacente, d'une dualité entre ce qui est tenu pour normal et ce qui est accepté comme source de révision, nous en venons tout naturellement à nous demander si ce n'est pas précisément le thème de la révision qui constitue le dénominateur commun aux phénomènes argumentatifs. Il convient donc de se tourner maintenant vers les formalismes non-monotones. Le prochain chapitre est consacré à ceux-ci.

Chapitre 2

Théories de la non-monotonie

2.1 Introduction

Ce chapitre comprend trois sections. L'objet de la première (i.e. section 2.2) est de présenter les logiques non-monotones et les théories de la révision, qui ont été développées à partir des années 1980 dans le cadre de l'Intelligence Artificielle. Le but premier des logiques non-monotones est de formaliser les raisonnements faisant appel à des règles générales sujettes à exception, du type : si ϕ alors normalement ψ. La principale caractéristique de telles règles est de mettre en jeu une relation de conséquence qui n'est pas monotone (ou croissante), en ce sens qu'une extension des prémisses peut conduire au retrait de la conclusion initiale. C'est la raison pour laquelle ces logiques sont appelées « non-monotones ». A dire vrai, elles sont nombreuses et nous ne chercherons pas à les passer toutes en revue. Nous présenterons successivement : la logique des défauts de Reiter ; la logique autoépistémique de Moore ; la théorie de la circonscription de Mc Carthy ; la théorie des modèles préférentiels. Les deux premières utilisent une approche par point-fixe. Les deux autres optent pour une approche à base de minimisation. C'est la raison pour laquelle nous préférons présenter la logique autoépistémique avant la théorie de la circonscription, quoiqu'elle lui soit historiquement postérieure. Nous terminerons cet exposé d'ensemble, en faisant une incursion dans les théories de la révision, qui se sont développées en parallèle des logiques non-monotones. Ces théories mettent au premier plan l'opération de changement de croyances, qu'elles tentent de formaliser. Elles s'y intéressent, notons-le, dans un contexte bien précis : celui de l'élimination d'une inconsistance dans une base de données. Les théories de la révision sont elles aussi nombreuses et nous ne chercherons pas non plus à les passer toutes en revue. Nous axerons la discussion sur la théorie d'Alchourron, Gärdenfors et Makinson, en raison de son caractère fondateur. Cette théorie est communément appelée « théorie AGM ».

Dans les deux autres sections de ce chapitre (i.e. sections 2.3 et 2.4), nous tentons de poser les jalons d'un rapprochement entre l'étude de l'argumentation et celui de la révision ou de la non-monotonie. Nous examinons certaines des applications qui ont été faites de ces formalismes à l'étude de l'argumentation. Nous verrons que ces tentatives d'articulation partent généralement dans trois directions.

2.2 Logiques non-monotones

2.2.1 Logique des défauts

L'une des premières logiques non-monotones qui ait été proposée est la logique des défauts. Nous la devons à Reiter [208]. Il travaille dans le cadre d'une logique des prédicats du premier ordre. Pour plus de simplicité, nous nous placerons dans le cadre de la logique propositionnelle. Les connaissances fournies par l'utilisateur sont ici représentées sous la forme d'une théorie $\Delta = (W, D)$. W est un ensemble de formules propositionnelles. Ce sont les informations que le système doit considérer comme non-révisables. D est un ensemble d'expressions de la forme $\frac{\phi:\psi}{\tau}$, appelées « défauts ». Ce sont les règles d'inférence sujettes à exceptions. Leur lecture est : si ϕ est vrai et si ψ est consistant avec ce que nous savons, alors inférer τ. ϕ est le *prérequis*, ψ est la *justification* et τ est la *conclusion* du défaut.

Partant d'une théorie Δ donnée, nous pouvons inférer de façon consistante un certain nombre d'ensembles de croyances (zéro, un ou plusieurs). Reiter nomme ceux-ci *extensions* de Δ. Il les définit en termes de point fixe. Désignons tout d'abord par $Th(F)$ l'ensemble des formules qui, en logique classique, peuvent être inférées à partir de F. Soit une théorie $\Delta = (W, D)$. Soit $\mathcal{G}$ l'opérateur qui, à un ensemble F de formules, associe le plus petit ensemble $\mathcal{G}(F)$ tel que :

1. $\mathcal{G}(F) \supseteq W$
2. $\mathcal{G}(F) = Th(\mathcal{G}(F))$
3. $\frac{\phi:\psi}{\tau} \in D, \phi \in \mathcal{G}(F), \neg\psi \notin \mathcal{G}(F) \;\mapsto\; \tau \in \mathcal{G}(F)$

Reiter nomme extension de Δ tout ensemble E qui est un point-fixe pour $\mathcal{G}$, c'est-à-dire qui vérifie l'équation $E = \mathcal{G}(E)$. Ainsi, E contient toutes les formules de départ $-$ condition 1. Il est également clos pour la déduction classique $-$ condition 2. Enfin, E contient tous les conséquents des défauts dont le prérequis figure déjà dans l'extension. Plus précisément, E contient les conséquents de ces défauts, sous réserve que la négation de leur justification n'y soit pas $-$ condition 3.

Il est clair que, en règle générale, une théorie peut ne pas avoir d'extension. Il suffit d'imaginer que $W = \emptyset$ et $D = \{\frac{:\neg\phi}{\phi}\}$. Toutefois, Reiter montre qu'il existe une classe particulière de théories avec défauts pour lesquelles l'existence d'extension est néanmoins assurée. Il s'agit de la classe des théories dans lesquelles chaque défaut est *normal*, au sens où son conséquent et sa justification sont identiques. (Par extension, ces théories sont qualifiées de *normales*.). Intuitivement, la caractéristique d'un tel défaut est de ne pas énumérer la liste des exceptions à la règle qu'il énonce $-$ liste que, dans la vie pratique, nous sommes le plus souvent incapables de dresser. Les théories normales n'ont pas pour seule propriété remarquable d'admettre toujours au moins une extension. Si on se limite à de telles théories, on peut définir une théorie de la preuve pour la logique des défauts. Notons $D_i \subseteq D$ un ensemble quelconque de défauts. Notons $CSQ(D_i)$ la conjonction de ses conséquents et $PRQ(D_i)$ la conjonction de ses prérequis. On montre que la formule ϕ est dans une extension de la théorie normale $\Delta = (W, D)$ à condition qu'il existe une preuve de ϕ à partir de Δ, c'est-à-dire une séquence $D_0, \ldots, D_k$ de sous-ensembles finis de D telle que :

1. $W \cup \{CSQ(D_0)\} \vdash \phi$
2. $W \cup \{CSQ(D_i)\} \vdash PRQ(D_{i-1})$ pour $i = 1, 2, ..., k$
3. $D_k = \emptyset$
4. $W \cup \{CSQ(D_i) | 0 \leq i \leq k\}$ est consistant.

Afin d'éclairer le sens de cette définition, prenons le cas limite où chaque D_i (autre que D_k) est un singleton. Les conditions (1)-(3) signifient que, dans ce cas, prouver ϕ revient simplement à montrer que, partant d'éléments de W, nous pouvons mettre nos défauts en chaîne, de façon à obtenir ϕ. Le prérequis du premier défaut utilisé, notons-le, n'est pas nécessairement un élément de W ; il suffit qu'il soit classiquement impliqué par W. Une remarque similaire s'applique au prérequis du énième défaut utilisé ainsi qu'à l'énoncé final ϕ. Ils ne correspondent pas nécessairement au conséquent du dernier défaut utilisé, mais il suffit de pouvoir, sur la base de W, les inférer classiquement du conséquent en question :

$$W \vdash \phi_1 \Rightarrow \psi_1 \overset{via\, W}{\vdash} \phi_2 \Rightarrow \psi_2 \quad \ldots \quad \phi_{n-1} \Rightarrow \psi_{n-1} \overset{via\, W}{\vdash} \phi_n \Rightarrow \psi_n \overset{via\, W}{\vdash} \phi.$$

Ici, $\Rightarrow$ désigne l'inférence par défaut. La condition (4) est simplement destinée à rendre compte du fait que les règles utilisées sont toutes sujettes à exception. Que la réunion de W et de l'ensemble des conséquents des défauts mis en jeu ne soit pas consistante veut dire : l'un au moins de ces défauts n'est pas applicable à la situation présente, telle que W la décrit. Nous voyons ici comment l'on a pu reprocher à la logique des défauts de soulever de sérieux problèmes d'effectivité. Prenons W=$\{f_1, ..., f_n\}$. Montrer que la condition (4) est satisfaite revient évidemment à montrer que la formule suivante n'est pas un théorème dans la logique sous-jacente que l'on utilise :

$$\neg(f_1 \wedge \ldots f_n \wedge \psi_1 \wedge \ldots \psi_m).$$

Or, il est bien connu que la question de la démontrabilité pour la logique du premier ordre est seulement semi-décidable : l'ensemble des formules démontrables est récursivement énumérable ; l'ensemble des formules non-démontrables ne l'est pas. Il en résulte que la question de savoir si une formule donnée figure ou non dans une extension de la théorie $\Delta = (W, D)$ n'est, elle, même pas semi-décidable. Comme le montrent Besnard et al. [32], une stratégie possible consiste à remplacer tout défaut normal

$$\frac{\phi : \psi}{\psi} \tag{2.1}$$

par un défaut sans prérequis dit « libre », de la forme

$$\frac{\top : \phi \rightarrow \psi}{\phi \rightarrow \psi}, \tag{2.2}$$

où $\to$ désigne l'implication matérielle. Nous disposons alors d'une procédure de décision pour les théories munies de défauts de ce type. Cette introduction de l'implication matérielle est paradoxale seulement en apparence. Plaçons-nous dans la situation hypothétique où ϕ figure dans l'extension. Littéralement, le défaut 2.2 stipule que, comme une extension est close pour la déduction classique, nous pouvons y inclure aussi ψ, à condition toutefois que nous n'y rencontrions pas déjà $\phi \wedge \neg\psi$, c'est-à-dire finalement $\neg\psi$. N'est-ce pas là précisément ce qu'affirme 2.1 ? Cela ne veut évidemment pas dire que le comportement des défauts libres ne diffère pas parfois de celui des défauts normaux. Sur ce point, nous renvoyons à e.g. Risch [210, pp. 36-39]. Qu'il nous suffise ici de vérifier que la relation d'inférence générée par une théorie avec défauts libres reste bien non-monotone. Ceci nous confortera définitivement dans l'idée que cet usage de l'implication matérielle n'est pas gênant. Comparons les deux théories suivantes :

$$(\{\phi\}, \{\frac{\phi : \psi}{\psi}\})$$
$$(\{\phi\}, \{\frac{: \phi \to \psi}{\phi \to \psi}\})$$

Notons $|\!\sim$ la relation d'inférence non-monotone. Ce symbole est lu « entraîne normalement ». Eu égard aux besoins de l'exemple, il semble indifférent que $|\!\sim$ soit pris en un sens crédule

$$W \;|\!\sim_{créd} f \quad \leftrightarrow \quad (\exists E\ )(E,\ \text{extension de } \Delta)(f \in E),$$

ou bien sceptique

$$W \;|\!\sim_{scept} f \quad \leftrightarrow \quad (\forall E\ )(E,\ \text{extension de } \Delta)(f \in E).$$

Dans les deux cas, pour les raisons précédemment évoquées, nous avons bien $W \;|\!\sim \psi$ mais nous n'avons pas $W \cup \{\neg\psi\} \;|\!\sim \psi$.

En guise de conclusion, nous indiquons très succinctement certains des développements auxquels a donné lieu la logique des défauts. Reiter et Criscuolo [209] étudient les propriétés de théories dites « semi-normales », dans lesquelles les défauts sont tous de la forme $\frac{\phi:\tau\wedge\neg\psi}{\tau}$ – ce qui veut dire que l'application du défaut est explicitement contrôlée. Ils introduisent ce type de défaut, dans le but de bloquer la transitivité entre défauts normaux. Ils illustrent la caractère non souhaitable de la transitivité, au moyen de l'exemple suivant : si les étudiants sont normalement des adultes, et si les adultes ont habituellement un emploi, il ne s'ensuit pas que les étudiants ont habituellement un emploi. De son côté, Lukaszievicz [152] redéfinit le concept d'extension afin d'en garantir l'existence pour toute théorie avec défauts. Brewka [40] propose une variante assurant la cumulativité de toute théorie avec défauts. Nous verrons un peu plus loin ce que signifie cette propriété.

Touretzky [254], Poole [200] ainsi que Froidevaux et Kayser [78] proposent de compléter la logique des défauts, afin qu'elle tienne compte des priorités entre défauts. Nous verrons quel usage la théorie des actes de langage peut en faire. Lin et Shoham [143], Siegel [234], Siegel et Schwind [235] développent une sémantique modale pour la logique des défauts. Cette sémantique comporte deux opérateurs, l'une exprimant le fait qu'un état de choses est connu, l'autre indiquant le fait qu'il a tout au plus le statut d'hypothèse.

2.2.2 Logique autoépistémique

La logique autoépistémique de Moore [176, 177] est issue pour l'essentiel d'une réflexion sur le système non-monotone de McDermott & Doyle [183] et McDermott [182]. Elle s'écarte de la logique des défauts de Reiter, en ce qu'elle prend appui sur la logique modale conçue comme une logique du *croire*. Moore utilise la modalité du nécessaire, $\Box$, pour désigner la croyance. Du point de vue formel, l'opérateur est soumis aux règles du système de logique modale K45 (terminologie de Chellas [48]) ou encore « S5 faible » (terminologie de Stalnaker [243]). Le système K45 n'est autre que S5 privé du schéma d'axiome de réflexivité,

$$T : (\Box\phi) \to \phi,$$

qu'il paraît naturel d'abandonner, si $\Box$ doit désigner la croyance, prise subjectivement[1]. Voici un exemple typique d'expression que, en logique autoépistémique, nous pouvons rencontrer : $(\neg\Box\neg\phi) \to \phi$; elle est lue « si nous n'avons aucune raison de croire que $\neg\phi$, alors ϕ ». En voici un autre : $(\Box\phi \wedge \neg\Box\neg\psi) \to \psi$; cet énoncé est lu « si nous croyons que ϕ et que nous n'avons aucune raison de penser que $\neg\psi$, alors ψ ». Nos deux formules expriment ce que Reiter nomme des défauts normaux. Elles donnent lieu à un type de raisonnement non-monotone que Moore qualifie de *autoépistémique* pour une raison évidente ; il est intimement lié à une réflexion de l'agent sur son propre savoir, qui est susceptible d'évoluer.

De la logique des défauts à la logique autoépistémique, avons-nous dit, le cadre d'analyse n'est pas le même. Mais nous pressentons déjà que, d'une logique à l'autre, les préoccupations seront identiques. Soit une théorie T. Intuitive-

[1] Rappelons les axiomes de K45. Ils sont au nombre de trois : (K) $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$; (4) $\Box\phi \to \Box\Box\phi$; (5) $\neg\Box\phi \to \Box\neg\Box\phi$. Il est d'usage d'appeler (4) et (5) « axiomes d'introspection positive et négative ». Le premier stipule que, si je crois que ϕ, alors je crois que je crois que ϕ. Le second stipule que, si je ne crois pas que ϕ, alors je crois que je ne crois pas que ϕ. Pour rendre valide (4), nous devons exiger de la relation d'accessibilité entre mondes qu'elle soit transitive. De même, pour rendre valide (5), nous devons supposer la relation d'accessibilité euclidienne : si wRw' et si wRw" alors w'Rw". On vérifiera sans peine qu'une relation transitive et euclidienne n'est pas nécessairement réflexive, laquelle propriété correspond au schéma T. Ceci suffit à montrer l'indépendance de T par rapport à (4) et (5).

ment, elle désigne l'ensemble des croyances initiales de l'agent. Quelles sont les croyances que ce dernier peut inférer de T ? Telle est la question que, à la suite de Reiter, Moore se pose. Celui-ci y répond, en introduisant la notion d'expansion stable d'une théorie, qu'il définit comme suit. Soit $\mathcal{L}$ un langage modal propositionnel. Prenons $\Gamma \subseteq \mathcal{L}$. Posons $\Box\Gamma = \{\Box\phi \mid \phi \in \Gamma\}$ et $\neg\Box\overline{\Gamma} = \{\neg\Box\phi \mid \phi \in \mathcal{L}\ \&\ \phi \notin \Gamma\}$ [2]. S est une expansion stable de T si et seulement si :

$$\mathrm{S} \quad = \quad \{\phi \mid \mathrm{T} \cup \Box\mathrm{S} \cup \neg\Box\overline{\mathrm{S}} \vdash_{K45} \phi\}. \tag{2.3}$$

Nous voyons que (2.3) énonce deux inclusions. Stalnaker [243] disait d'une théorie quelconque qu'elle est stable, si elle contient toutes ses conséquences logiques, qu'elles soient classiques ou « autoépistémiques », c'est-à-dire tirées moyennant l'hypothèse supplémentaire d'une connaissance exhaustive de ce que la théorie contient et ne contient pas. Ceci nous donne l'inclusion :

$$\{\phi \mid \mathrm{T} \cup \Box\mathrm{S} \cup \neg\Box\overline{\mathrm{S}} \vdash_{K45} \phi\} \quad \subseteq \quad \mathrm{S}.$$

S n'est pas seulement un ensemble stable de croyances au sens de Stalnaker. Cette théorie a aussi la propriété remarquable, poursuit Moore [176, p. 85], d'être « fondée » (*grounded*) sur les prémisses T, au sens où elle ne contient pas plus que ce qui peut, de proche en proche, être inféré à partir de T. Ainsi doit-on comprendre l'inclusion :

$$\mathrm{S} \quad \subseteq \quad \{\phi \mid \mathrm{T} \cup \Box\mathrm{S} \cup \neg\Box\overline{\mathrm{S}} \vdash_{K45} \phi\}.$$

Intimement liée au concept de dérivation, la notion de fondement (*groundedness*) est ici à manier avec précaution, compte tenu de la définition de S en termes de point-fixe [3].

Nous avons suggéré que l'expansion stable jouait en logique autoépistémique approximativement le même rôle que l'extension en logique des défauts. En fait, comme l'a montré Konolige [125] et d'autres à sa suite[4], le rapport entre les deux logiques est bien plus étroit que nous ne l'avons laissé paraître. Résumons le résultat célèbre auquel par exemple Konolige parvient et qui, du reste, nécessite que

[2]Conformément à l'usage, $\overline{X}$ désigne le complémentaire de X par rapport à l'ensemble de référence. On distinguera $\neg\Box\overline{\Gamma}$ et $\overline{\Box\Gamma}$. Le premier a pour seuls éléments les énoncés qui sont la négation d'une formule apparaîssant dans $\Box\overline{\Gamma}$. Le second énumère les énoncés qui ne sont pas $\Box\overline{\Gamma}$. Nous allons illustrer dans un instant en quoi différerait l'usage d'une double complémentation.

[3]Procéder par double complémentation, c'est-à-dire substituer, dans (2.3), $\overline{\Box\overline{\mathrm{S}}}$ à $\neg\Box\overline{\mathrm{S}}$ conduirait à un effet « noyade ». Venant de vérifi er qu'une formule propositionnelle p ne fi gure pas dans l'expansion stable, nous serions paradoxalement contraints de l'y introduire. Supposons que $p \notin \mathrm{S}$. Quelle que soit la nature de $\Box\overline{\mathrm{S}}$, nous savons que cet ensemble contient seulement des formules préfi xées de l'opérateur $\Box$ [déf de $\Box\Gamma$]. Cela implique que $p \in \overline{\Box\overline{\mathrm{S}}}$ [déf du complémentaire]. D'où il vient $p \in \mathrm{S}$ [déf de S].

[4]Tel Truszczynski [260].

quelques réaménagements soient apportés à (2.3). Soient $\Delta = (W, D)$ une théorie avec défauts et T une théorie autoépistémique mise sous forme normale dite étendue. Par là, nous devons comprendre que les éléments de T sont tous de la forme $(\Box\phi \wedge \neg\Box\neg\psi) \rightarrow \omega$, où ϕ, ψ et ω sont des énoncés propositionnels. Δ est appelée la transformée de T, et T la transformée de Δ lorsque, pour tous les éléments de W, D et T, nous avons :

1. $\frac{\phi:\psi}{\omega} \in D$ ssi $(\Box\phi \wedge \neg\Box\neg\psi) \rightarrow \omega$
2. $\omega \in W$ ssi $\omega \in \mathrm{T}$

De même que nous pouvons traduire toute théorie autoépistémique sous la forme d'une théorie avec défauts, nous pouvons traduire toute théorie avec défauts sous la forme d'une théorie autoépistémique. Pour plus de simplicité, nous n'évoquerons pas les quelques amendements que Konolige propose d'apporter à (2.3), pour qu'une théorie donnée soit aussi expressive que sa transformée, c'est-à-dire autorise exactement le même nombre de conclusions. Qu'il nous suffise ici d'indiquer que le résultat de Konolige concerne les seules conclusions de nature propositionnelle. Cette restriction s'explique par le fait que la logique de Reiter permet seulement d'inférer des faits.

2.2.3 Théorie de la circonscription

Due dans l'essentiel à Mc Carthy [180, 181], la théorie de la circonscription est habituellement présentée sous sa forme syntaxique. Mc Carthy propose de traduire un énoncé du type « les oiseaux, sauf exceptions, volent » par :

$$\forall x[(oiseau(x) \wedge \neg anormal(x)) \rightarrow vole(x)].$$

Pour déterminer ce qui peut être inféré (par défaut) d'une théorie Δ munie d'énoncés de ce type, Mc Carthy introduit dans Δ un axiome supplémentaire, qu'il nomme « axiome de circonscription ». Le rôle de cet axiome est de réduire au minimum le nombre des exceptions aux règles figurant dans Δ. Notons $\Delta(anormal)$ notre axiomatique initiale. Nous pouvons y remplacer *anormal* par tout autre prédicat p, obtenant ainsi $\Delta(p)$. L'axiome de circonscription peut, en première approximation, être formulé ainsi :

$$(\star)\ \Delta(anormal) \wedge \neg\exists p\,[\Delta(p) \wedge p < anormal],$$

où $p < anormal$ est une abréviation de

$$\forall x(p(x) \rightarrow anormal(x)) \wedge \neg\forall x(anormal(x) \rightarrow p(x)).$$

Intuitivement, l'expression $\neg\exists p\,[\Delta(p) \wedge p < anormal]$ signifie que *anormal* est le plus petit prédicat p pour lequel $\Delta(p)$ existe. A dire vrai, notre formulation $(\star)$

de l'axiome de circonscription n'est pas totalement satisfaisante, pour au moins trois raisons. Tout d'abord, ($\star$) ne rend pas compte du fait qu'il faut généralement minimiser ou circonscrire plusieurs prédicats (car la base peut contenir plusieurs règles défaisables). Ensuite, contrairement à ce que suggère ($\star$), les prédicats p_1, p_2, ..., p_n à circonscrire ne sont pas nécessairement des prédicats d'anormalité (il peut s'agir, par exemple, d'une relation d'égalité). Enfin, ($\star$) ne rend pas compte du fait qu'il faut supposer variable l'extension d'autres prédicats q_1, q_2, ..., q_m (par exemple, la propriété de voler) ; autrement, la circonscription n'aurait pas d'influence sur la valeur qu'ils prennent. Voici une formulation plus satisfaisante du principe de circonscription ; pour plus de simplicité, nous présentons celui-ci sous la forme d'un schéma d'axiome. Dans un ensemble fini Δ de formules fermées du premier ordre, le schéma d'axiome qui réalise la circonscription des prédicats p_1, p_2, ..., p_n en faisant varier les prédicats q_1, q_2, ..., q_m est de la forme :

$$\begin{array}{l} \Delta[\Phi_{p_1}, \ldots \Phi_{p_n}; \varphi_{q_1}, \ldots \varphi_{q_m}] \\ \rightarrow \quad \{ [\forall x(\Phi_{p_1}(x) \rightarrow p_1(x))] \wedge \ldots \wedge [\forall x(\Phi_{p_n}(x) \rightarrow p_n(x))] \\ \qquad \rightarrow \quad [\forall x(p_1(x)) \rightarrow \Phi_{p_1}(x)] \wedge \ldots \wedge [\forall x(p_1(x)) \rightarrow \Phi_{p_1}(x)] \} \end{array}$$

Chacun des Φ_{p_i} (resp. chacun des φ_{q_i}) peut être n'importe quelle formule ouverte. L'énoncé

$$\Delta[\Phi_{p_1}, \ldots \Phi_{p_n}; \varphi_{q_1}, \ldots \varphi_{q_m}]$$

représente la conjonction des éléments de Δ, une fois chaque prédicat p_i (resp. q_i) remplacé dans toutes ses occurrences par Φ_{p_i} (resp. φ_{q_i}). L'ensemble de toutes les instances de ce schéma d'axiome est noté

$$\mathrm{CIRC}_\Delta[p_1, \ldots, p_n; q_1, \ldots, q_m].$$

On dit alors d'une formule f qu'elle est dérivable par circonscription de Δ si et seulement si :

$$\Delta \cup \mathrm{CIRC}_\Delta[p_1, \ldots, p_n; q_1, \ldots, q_m] \quad \vdash \quad f$$

Toutes les instances du schéma de circonscription n'ont pas à être utilisées. Le fait que telle ou telle instance soit ou non « intéressante » dépend de la nature de la formule particulière f que nous cherchons à établir ainsi que de la nature de Δ. Dans la pratique, notre trouvons assez facilement les « bons » prédicats d'instanciation, en nous fiant à notre propre intuition. Reste que l'automatisation de cette recherche des bonnes instances est difficile.

Illustrons la technique de la circonscription au moyen d'un exemple, que nous empruntons à Lifschiftz [144]. Imaginons que Δ contienne les informations suivantes

- Les cubes sont normalement sur la table
- a et b sont deux cubes distincts

– a n'est pas sur la table,

que nous pouvons traduire par :

$$\forall x[(cube(x) \wedge \neg anormal(x)) \rightarrow table(x)] \tag{2.4}$$

$$cube(a) \tag{2.5}$$

$$cube(b) \tag{2.6}$$

$$a \neq b \tag{2.7}$$

$$\neg table(a). \tag{2.8}$$

Circonscrire *anormal* en faisant varier *table* permet de conclure que le cube b est sur la table, i.e.

$$table(b). \tag{2.9}$$

Vérifions-le pas à pas. Notre schéma de circonscription revêt la forme d'une implication dont l'antécédent est

$$\begin{gathered}\forall x\,[\,(\,cube(x) \wedge \neg\Phi_{anormal}(x)\,) \rightarrow \varphi_{table}(x)\,] \\ \wedge\, cube(a) \wedge cube(b) \wedge a \neq b \wedge \neg\varphi_{table}(a)\end{gathered}$$

et dont le conséquent est

$$\forall x\,(\,\Phi_{anormal}(x) \rightarrow anormal(x)\,) \rightarrow \forall x\,(\,anormal(x) \rightarrow \Phi_{anormal}(x)\,).$$

L'intuition conduit ici à poser :

– $\Phi_{anormal}(x) \leftrightarrow (cube(x) \wedge x = a)$

– $\varphi_{table}(x) \leftrightarrow x \neq a$

Ce choix nous donne l'instance suivante du schéma de circonscription :

$$\begin{aligned}&\{\,\forall x\,[\,(\,cube(x) \wedge \neg(cube(x) \wedge x = a)\,) \rightarrow x \neq a\,] \wedge cube(a) \wedge cube(b) \wedge a \neq b \wedge \neg(a \neq a)\,\} \\ \rightarrow\quad &\{\,\forall x[\,(cube(x) \wedge x = a) \rightarrow anormal(x)\,] \rightarrow \forall x[anormal(x) \rightarrow (cube(x) \wedge x = a)]\}\end{aligned}$$

Nous pouvons y supprimer les sous-formules :

$$\forall x\,[\,(\,cube(x) \wedge \neg(cube(x) \wedge x = a)\,) \rightarrow x \neq a\,]$$

$$\neg(a \neq a)$$

En effet, ce sont des tautologies[5]. Nous obtenons :

$$\begin{aligned}&(\,cube(a) \wedge cube(b) \wedge a \neq b\,) \\ \rightarrow\quad &\{\,\forall x[\,(cube(x) \wedge x = a) \rightarrow anormal(x)\,] \rightarrow \forall x[anormal(x) \rightarrow (cube(x) \wedge x = a)]\}\end{aligned}$$

[5]Par contraposition, la première formule est équivalente à $\forall x[\,(\,cube(x) \wedge x = a) \rightarrow (cube(x) \wedge x = a)\,]$.

Examinons l'antécédent de l'implication principale, puis celui de l'implication qui lui est immédiatement subordonnée. Le premier antécédent est la conjonction d'énoncés qui figurent déjà dans Δ. Ainsi :

$$\Delta \cup \mathrm{CIRC}_\Delta[anormal; table] \;\vdash\; cube(a) \wedge cube(b) \wedge a \neq b$$

Quant au second antécédent, qui affirme que *tout cube identique à a est anormal*, de très simples opérations de déduction dans la logique du premier ordre permettent de l'inférer à partir de Δ :

$$\Delta \cup \mathrm{CIRC}_\Delta[anormal; table] \;\vdash\; \forall x[\,(cube(x) \wedge x = a) \to anormal(x)].$$

Par une double application du modus-ponens, nous obtenons :

$$\Delta \cup \mathrm{CIRC}_\Delta[anormal; table] \;\vdash\; \forall x[anormal(x) \to (cube(x) \wedge x = a)].$$

Ainsi, dans notre exemple, l'usage de la circonscription revient à introduire la donnée supplémentaire selon laquelle *le cube a est (sauf preuve du contraire) le seul qui soit anormal*. Munis de cette information supplémentaire, nous en tirons immédiatement l'énoncé *le cube b est sur la table* :

$$\Delta \cup \mathrm{CIRC}_\Delta[anormal; table] \;\vdash\; table(b).$$

Il est tentant de paraphraser la loi « les cubes sont normalement sur la table » par « la plupart des cubes sont sur la table ». Il est aussi tentant de supposer que la locution « la plupart des x » recouvre une collection qui forme la majorité des x, c'est-à-dire plus de la moitié. Nous pourrions alors croire que, dans notre exemple, le fait d'avoir seulement deux cubes joue un rôle évident [6]. Il n'en est rien. Apprenant qu'il existe une multiplicité d'autres cubes, mais ignorant où chacun d'eux se trouve, nous déduisons toujours par circonscription que b est sur la table. Ceci est paradoxal, seulement pour qui oublie que la relation de déduction par circonscription exprime une simple inférence tirée par défaut.

Reposant sur l'introduction et la manipulation d'un axiome supplémentaire, la circonscription semble une approche essentiellement syntaxique. En fait, il est également possible d'en donner une présentation sémantique, en introduisant une relation de préordre sur les modèles de la théorie. Nous ne développerons pas ce point, dans la mesure où cette caractérisation sémantique annonce la théorie des modèles préférentiels, à la présentation de laquelle la section suivante est consacrée.

[6] Si plus de la moitié des cubes est sur la table, alors l'un au moins d'entre eux est sur la table ; ce n'est pas a ; donc c'est b.

2.2.4 La théorie des modèles préférentiels

Gabbay [79], Makinson [157], Kraus et al. [127], Lehmann et Magidor [134] en sont les principaux créateurs. Leur travaux se situent, en un sens, au carrefour de deux traditions en logique. Tout d'abord, conformément à la tradition que Tarski et Gentzen initièrent, ces auteurs mettent au premier plan la relation de conséquence, qu'ils notent $|\sim$ et lisent « entraîne normalement ». Ensuite, pour déterminer les conditions d'évaluation de $|\sim$, ils utilisent une sémantique des mondes possibles, selon la tradition que Kripke inaugura au début des années 1960. Le type d'analyse qu'ils proposent est très proche de celle que Stalnaker [241] et Lewis [140] donnent du conditionnel contrefactuel, ou de celle que Hansson [103] donne de l'obligation conditionnelle. L'idée est, en substance, de remplacer la relation d'accessibilité entre mondes, que Kripke utilisait, par une relation de préférence. A partir de cette première innovation, il devient possible de définir des systèmes d'inférences non-monotones de force croissante. L'un des plus faibles qui soit est le système préférentiel **P**. Essentiellement dû à Kraus et al.[127], ce système est étudié en détail par Lehmann et Magidor [134]. Ceux-ci nomment « assertion conditionnelle » une expression de la forme $\phi \mid\sim \psi$. Nous pouvons la lire « ϕ a normalement pour conséquence ψ ». Pour pouvoir interpréter ce type d'expression, il nous faut nous munir d'un modèle défini comme un triplet $\mathcal{M} = (W, \iota, \prec)$ où :

1. W désigne un ensemble de mondes possibles $w, w', \dots$;
2. ι est une fonction d'évaluation associant à chaque atome propositionnel l'ensemble des mondes dans lequel il est vrai ;
3. $\prec \subseteq W \times W$ est une relation d'ordre strict (transitive et irréflexive) ; intuitivement, $w \prec w'$ signifie « w est plus plausible que w' ».

Soit $[\phi]_{\mathcal{M}}$ l'ensemble des mondes de $\mathcal{M}$ qui vérifient ϕ. Et soit $\min_{\mathcal{M}}(\phi)$ le sous-ensemble de ceux qui sont minimaux sous $\prec$, i.e.

$$\min_{\mathcal{M}}(\phi) \;=\; \{w \in [\phi]_{\mathcal{M}} \mid \neg\exists w'(w' \in [\phi]_{\mathcal{M}} \,\&\, w' \prec w)\}$$

Intuitivement, $\min_{\mathcal{M}}(\phi)$ énumèrent les ϕ-mondes les moins exceptionnels, les plus normaux, etc. De fait, nos auteurs se limitent à une classe particulière de structures, celles dans lesquelles $\prec$ vérifie une propriété qu'ils nomment « smoothness » et qui stipule qu'il n'existe pas de chaîne infinie de ϕ-mondes de plus en plus normaux. Formellement : :

$$\forall w \in ([\phi]_{\mathcal{M}} - \min_{\mathcal{M}}(\phi)) \quad \exists w' \in \min_{\mathcal{M}}(\phi) \text{ tel que } w' \prec w.$$

Les conditions de récurrence pour $|\sim$ sont définies relativement à un modèle. Nous avons :

$$\phi \mid\sim_{\mathcal{M}} \psi \;\;\Leftrightarrow\;\; \min_{\mathcal{M}}(\phi) \subseteq [\psi]_{\mathcal{M}}. \tag{2.10}$$

Lorsque $\phi \mid\sim_{\mathcal{M}} \psi$ se vérifie, on dit de $\mathcal{M}$ qu'il est un modèle de l'assertion conditionnelle $\phi \mid\sim \psi$. Intuitivement, la clause (2.10) signifie que $\mathcal{M}$ est un modèle de l'assertion conditionnelle $\phi \mid\sim \psi$ si et seulement si les plus plausibles des ϕ-mondes que l'univers de $\mathcal{M}$ contient vérifient tous ψ. Soit K un ensemble d'assertions conditionnelles et $\mathcal{A}$ une assertion conditionnelle. $\mathcal{A}$ est appelée « conséquence préférentielle » de K (notation : $\mathcal{A} \in K^p$) si et seulement si tous les modèles de K sont aussi des modèles de $\mathcal{A}$.

Voilà pour l'essentiel de la sémantique du système **P**. Un mot, à présent, sur son axiomatique. On montre que $\mathcal{A}$ compte au nombre des conséquences préférentielles de K, exactement si $\mathcal{A}$ se déduit de K à l'aide des règles suivantes, dont l'ensemble forme le système **P** :

1. Equivalence logique à gauche :
$$\frac{\phi \mid\sim \psi \quad \vdash \phi \leftrightarrow \phi'}{\phi' \mid\sim \psi}$$
2. Affaiblissement à droite :
$$\frac{\phi \mid\sim \psi \quad \vdash \psi \to \psi'}{\phi \mid\sim \psi'}$$
3. Réflexivité :
$$\frac{}{\phi \mid\sim \phi}$$
4. Conjonction :
$$\frac{\phi \mid\sim \psi \quad \phi \mid\sim \psi'}{\phi \mid\sim \psi \wedge \psi'}$$
5. Disjonction :
$$\frac{\phi \mid\sim \psi \quad \phi' \mid\sim \psi}{\phi \vee \phi' \mid\sim \psi}$$
6. Monotonie prudente :
$$\frac{\phi \mid\sim \psi \quad \phi \mid\sim \tau}{\phi \wedge \psi \mid\sim \tau}$$

La règle de monotonie prudente correspond à un affaiblissement du principe de monotonie admis en logique classique. Lorsqu'on ajoute ψ à ϕ, on continue à pouvoir déduire τ, à condition que cela soit un ψ que ϕ permet de déduire non-monotoniquement. Du point de vue sémantique, la validation de la monotonie prudente présuppose l'hypothèse de smoothness. Pour s'en convaincre, il suffit de remarquer que, en présence de cette dernière, nous avons :

$$\min_{\mathcal{M}}(\phi) \subseteq [\psi]_{\mathcal{M}} \quad \to \quad \min_{\mathcal{M}}(\phi \wedge \psi) \subseteq \min_{\mathcal{M}}(\phi).$$

En présence des règles de **P**, nous avons aussi la règle dite de « coupure », qui correspond à un affaiblissement de la transitivité :

$$\frac{\phi \mid\!\sim \psi \quad \phi \wedge \psi \mid\!\sim \tau}{\phi \mid\!\sim \tau}$$

Nous pouvons réunir les lois de monotonie prudente et de coupure en une seule. Nous obtenons le principe dit « de cumulativité » : supposons que $\phi \mid\!\sim \psi$; $\phi \wedge \psi \mid\!\sim \tau$ implique $\phi \mid\!\sim \tau$, et vice-versa.

Des systèmes plus forts sont envisageables. Le plus connu d'entre eux est sans doute le système **R** de Lehmann et Magidor [134]. Il s'agit d'une extension de **P**, que l'on obtient en y introduisant la loi suivante, nommée « règle de monotonie rationnelle » :

$$\frac{\phi \not\mid\!\sim \neg\psi \quad \phi \mid\!\sim \tau}{\phi \wedge \psi \mid\!\sim \tau}$$

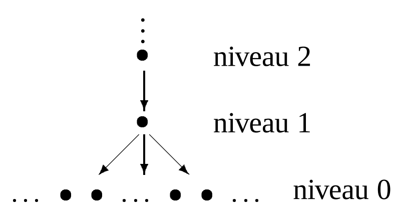
Lorsqu'on ajoute ψ à ϕ, on continue à pouvoir déduire τ, à condition que ψ et ϕ soient (non-monotoniquement) compatibles. Dans cette logique, la relation de préférence $\prec$ vérifie une propriété supplémentaire : elle « range » linéairement les mondes possibles. Ensuite, Bezzazi et al. [34] proposent une extension de **R**, obtenue en introduisant dans **P** la règle :

$$\frac{\phi \mid\!\sim \tau \quad \phi \wedge \psi \not\mid\!\sim \neg\tau}{\phi \wedge \psi \mid\!\sim \tau}$$

Cette règle est équivalente à :

$$\frac{\phi \mid\!\sim \psi \quad \psi \mid\!\sim \tau \quad \phi \not\mid\!\sim \neg\tau}{\phi \mid\!\sim \tau}$$

Les structures pour cette logique sont extrêmement simples. Elles sont de la forme :



Dans ce diagramme, chaque point • représente un monde possible w et les flèches $w\bullet \leftarrow \bullet w'$ dénotent la relation $w \prec w'$. Dans ce type de structure, un embranchement est possible seulement lorsqu'on passe du niveau 1 au niveau 0. (Il faut ici comprendre que plus le niveau augmente, plus le degré de plausibilité diminue.). Cette sorte de structure est qualifiée de quasi-linéaire.

Un compte rendu plus détaillé de la théorie des modèles préférentiels – à laquelle nous faisons ici trop rapidement allusion – nécessiterait des développements formels, dont la préoccupation n'est cependant pas au coeur de notre propos et pour lesquels nous renverrons donc aux publications qui leur ont été consacrées [7]. Qu'il nous suffise d'indiquer l'usage qui en est fait en logique des normes (ou logique déontique). Hansson [103] fut l'un des premiers logiciens à analyser les normes à l'aide d'une sémantique préférentielle. Nous aurons l'occasion de revenir sur ses motivations[8], qui ne nous sont pas essentielles ici. La transition vers le déontique se fait très naturellement. Elle peut être sommairement décrite comme suit. Tout d'abord, aux expressions de la forme $\phi \mid\sim \psi$, on substitue des expressions de la forme $\bigcirc(\psi/\phi)$, qui sont lues « Il est obligatoire que ψ dans la condition où ϕ ». Ensuite, on donne une lecture déontique à la relation de préférence $\prec$ entre mondes possibles : intuitivement, l'expression $w \prec w'$ est désormais interprétée comme signifiant que le monde w est plus parfait que le monde w'. Enfin, on suppose que les conditions de récurrence pour $\bigcirc(-/-)$ sont isomorphes à celles que l'on utilise pour $\mid\sim$. Intuitivement, on dit donc de $\mathcal{M}$ qu'il est un modèle de l'obligation conditionnelle $\bigcirc(\psi/\phi)$ si et seulement si les plus parfaits des ϕ-mondes (que l'univers de $\mathcal{M}$ contient) vérifient ψ. Quoiqu'éclairante, cette présentation de la sémantique de Hansson appelle néanmoins deux précisions. Tout d'abord, il nous faut indiquer que ce logicien travaille, non avec la relation « strictement inférieur à » ($\prec$), mais avec la relation « supérieur ou égal à » ($\succeq$) et que, corrélativement, il définit les conditions de récurrence pour $\bigcirc(-/-)$ en termes de maximalité. Notons $\models$ la relation de satisfaction d'une formule par un monde possible. Nous avons une clause de la forme :

$$w \models \bigcirc(\psi/\phi) \text{ ssi } (\forall w')(wR_\phi w' \rightarrow w' \models \psi),$$

avec

$$wR_\phi w' \text{ ssi : } w' \models \phi \text{ et } (\forall w'')(w'' \models \phi \rightarrow w' \succeq w'').$$

Le propos d'Hansson est seulement de soumettre à notre réflexion un champ d'investigation nouveau. Se plaçant dans une perspective exclusivement sémantique, il étudie trois systèmes :

Système DSDL1 : la relation $\succeq$ est seulement réflexive ;

Système DSDL2 : $\succeq$ est non seulement réflexive mais aussi vérifie une propriété assez voisine de celle de smoothness (pour $\succeq$), quoique plus faible : si $\exists w \in W$ tel que $w \models \phi$, alors $\exists w'$ tel que $wR_\phi w'$. Alchourrón [2] nomme cela la propriété de l'« expansion limite » (*limit expansion*).

Système DSDL3 : outre les deux précédentes propriétés, $\succeq$ vérifie la propriété de transitivité et elle ordonne totalement l'univers W.

Parmi les tentatives intéressantes d'axiomatisation, mentionnons celles d'Åqvist [14, 16, 17]. L'auteur se penche sur le problème de la complétude forte, qu'on peut approximativement formuler comme suit. Soient Δ un ensemble de formules et F une

[7] Outre aux quelques articles déjà mentionnés, nous renvoyons à Makinson [158, 159].

[8] Cf. notre chapitre 3, pp. 116 et ss.

formule. Si (dans la classe des modèles considérés) F est conséquence sémantique de Δ, s'ensuit-il que (dans la logique considérée) F soit démontrable à partir de Δ ? L'auteur reprend et adapte à son propos la méthode du modèle canonique, qui s'est peu à peu imposé en logique modale[9]. La principale innovation d'Åqvist est ici d'introduire dans la syntaxe une série infinie de constantes propositionnelles $\{Q_i\}$ $(1 \leq i \leq \omega)$, chacune desquelles étant le corrélat syntaxique d'un niveau de perfection opt_n $(n = 1, 2, \ldots)$ défini sur l'univers des modèles. Par exemple, nous pouvons tenir $\succeq$ pour primitive, et poser :

$$opt_n = \begin{cases} \{x \in W : (\forall y \in W)(x \succeq y)\} & \text{si } n = 1, \\ \{x \in W - \bigcup\limits_{1 \leq i \leq n-1} opt_i : (\forall y \in W - \bigcup\limits_{1 \leq i \leq n-1} opt_i)(x \succeq y)\} & \text{si } n > 1. \end{cases}$$

Intuitivement, opt_1 regroupe les plus parfaits des éléments de W, qui est ici pris comme un tout. Le bloc opt_2 énumère les plus parfaits des éléments de $W - opt_1$. L'ensemble opt_3 donne la liste des plus parfaits des éléments de $W - (opt_1 \cup opt_2)$, et ainsi de suite. Les conditions de récurrence pour chaque constante $\{Q_i\}$ $(1 \leq i \leq \omega)$ sont alors très simples à formuler :

$w \models Q_i$ ssi $w \in opt_i$ (pour tout entier positif i).

Nous n'entrerons pas dans le détail de la démonstration, qui fait appel à plusieurs lemmes. La série de constantes $\{Q_i\}$ $(1 \leq i \leq \omega)$ joue un rôle crucial dans le raisonnement. Ce sont elles qui permettent de définir la relation $\succeq$ du modèle canonique, sur la construction duquel toute l'argumentation repose.

Voici une dernière remarque sur l'utilisation des sémantiques préférentielles dans le déontique. Elle concerne la notion d'obligation inconditionnelle, qui soulève un problème intéressant. Il est d'usage de considérer $\bigcirc\phi$ comme une abréviation de $\bigcirc(\phi/\top)$, où $\top$ désigne une tautologie quelconque. Ceci n'est pas totalement satisfaisant. Intuitivement, nous souhaiterions qu'une obligation catégorique ne soit pas « isolée » de son contexte, à savoir la situation présente (laquelle ne vérifie pas les seules tautologies du calcul propositionnel). Une variante intéressante nous est suggérée par Alchourrón [2]. Reprenant une idée de von Wright [278], le logicien argentin considère $\bigcirc\phi$ comme étant équivalent à $\bigcirc(\phi/\sigma)$, où σ désigne une constante propositionnelle lue « dans les circonstances présentes ». Du point de vue sémantique, il suppose que l'univers de chaque modèle contient un élément $\star$ spécifique désignant le monde réel et tel que cet élément soit le seul à vérifier la constante σ. Quoique de prime abord plus satisfaisante, cette façon de réintroduire l'obligation inconditionnelle a un inconvénient manifeste. Car, s'il est vrai que $\star$ est l'unique « point » du modèle dans lequel nous pouvons rencontrer la constante σ, alors l'ensemble des σ-mondes les plus parfaits se réduit au singleton

[9]Cette méthode est due (à notre connaissance) à Makinson [155]. En gros, la particularité du modèle canonique est qu'il valide précisément les théorèmes de la logique considérée.

ayant $\star$ pour élément. Dans le monde actuel, on peut ainsi tirer directement de la proposition

Il est vrai que ϕ

la conclusion selon laquelle

Il est obligatoire que ϕ.

Adoptant une optique plus syntaxique, nous souhaiterions pouvoir isoler les principes logiques qui, dans le système du logicien argentin, nous autorise à déduire $\bigcirc\phi$ à partir de ϕ :

$$\frac{\phi}{\therefore \bigcirc\phi}\ (\text{IO})$$

Ce schéma d'inférence peut être utilement rapproché de l'axiome

$$\bigcirc(\phi/\phi) \qquad \text{(Fixité du contexte)}$$

qu'on peut lire

Il est obligatoire que ϕ *dans la condition où* ϕ.

Présent dans le système, ce dernier axiome n'autorise pas en lui-même l'inférence d'un « doit » (*ought*) à partir d'un « est » (*is*). Néanmoins, nous pouvons établir (IO) en invoquant cet axiome, ainsi que deux autres principes : la loi que nous nommerons « axiome de Meredith-Prior » (du nom de ceux qui les premiers l'ont utilisé[10])

$$\phi \rightarrow \Box(\sigma \rightarrow \phi) \tag{2.11}$$

qu'on peut lire

Si ϕ *alors nécessairement les circonstances présentes impliquent* ϕ ;

et l'axiome

$$\Box(\psi \rightarrow \psi') \rightarrow (\bigcirc(\psi/\phi) \rightarrow \bigcirc(\psi'/\phi)), \tag{2.12}$$

qui nous autorise à affaiblir le conséquent d'une obligation conditionnelle. Le raisonnement est extrêmement bref :

[10] Voir Meredith et Prior [170]. Il s'agit de la deuxième des trois lois qui composent l'axiomatique de leur logique pour la succession des états du mondes. Le symbole $\Box$ que nous introduisons ici désigne un opérateur modal de type S5, tel qu'usuellement défi ni.

$$\cfrac{\cfrac{\phi}{\Box(\sigma \rightarrow \phi)} \qquad \cfrac{}{\bigcirc(\sigma/\sigma)}}{\underbrace{\bigcirc(\phi/\sigma)}_{\bigcirc\phi}}$$

DER 1 : *Dérivation d'un « doit » à partir d'un « est »*

Une réaction assez naturelle consiste à imputer la difficulté à la présence de l'axiome $\bigcirc(\phi/\phi)$. Du reste, nombre d'auteurs ont objecté à la sémantique de Hansson le fait qu'elle rende cette loi valide. Nous-mêmes nous aurons l'occasion d'étudier quelques solutions envisageables[11]. Néanmoins, il faut ici admettre que l'axiome de Meredith-Prior n'est pas lui non plus très intuitif. Par exemple, nous pouvons nous demander s'il ne revient pas à identifier la situation présente à l'ensemble des formules booléennes actuellement vraies, abstraction faite de la connaissance que nous en avons. Intuitivement, cela n'est pas très satisfaisant. Nous souhaiterions que, dans l'expression $\bigcirc(\phi/\sigma)$, σ fasse référence aux seules informations qui nous sont disponibles. Comme le suggère Makinson [160], la théorie d'une obligation inconditionnelle ainsi conçue reste à faire.

2.2.5 Théorie de la révision des croyances

Dans les sections précédentes, nous avons passé en revue les principales logiques non-monotones existentes. Il serait maladroit de parler de non-monotonie sans évoquer, même allusivement, les théories de la révision de croyances, qui se sont développées en parallèle des logiques non-monotones. Nous nous permettrons ici de ne faire qu'une timide incursion dans ces théories, dont Alchourrón, Gärdenfors et Makinson élaborèrent les fondements. Le cadre d'analyse qu'ils fixèrent porte le nom de « théorie AGM ». L'essentiel se trouve dans Alchourrón et Makinson [6, 7], Alchourrón, Gärdenfors et Makinson [5], Gärdenfors [84], Gärdenfors et Makinson [82]. Pour une vue d'ensemble, nous renvoyons à Makinson [156], Gärdenfors et Rott [83], Léa Sombe [133] et Livet [149].

2.2.5.1 Trois opérations noyaux

Il est fréquent de distinguer deux grands types d'approches, selon que la révision concerne des théories, ensembles de formules clos sous la conséquence logique, ou leur seule base axiomatique. On qualifie généralement la première approche de « cohérentiste » et la seconde de « fondationaliste ». Les travaux

[11] Ce sera le point de départ de notre chapitre 3, consacré à l'obligation conditionnelle.

de Alchourrón, Gärdenfors et Makinson s'inscrivent dans le cadre de la première approche. Soit K une théorie. A première vue, plusieurs sortes de modifications peuvent lui être apportées. Dans le cas de figure le plus simple, nous nous contenterons d'y introduire un nouvel énoncé ϕ. Alchourrón et al. nomment cette première opération « expansion de K par ϕ » $-$ ce qu'ils notent $K+\phi$. Il peut aussi arriver que nous ayons à retirer ϕ de l'ensemble K. Ils nomment cette seconde opération « contraction de K par ϕ » $-$ ce qu'ils notent $K\dot{-}\phi$. Dans la pratique, procéder à une contraction s'avère nécessaire, toutes les fois que l'ajout d'une nouvelle information introduit une inconsistance. Alchourrón et al. réservent le terme de « révision » pour désigner ce cas de figure précis. Ils définissent la révision de K par ϕ (notation : $K^{\star}\phi$) en termes de contraction et d'expansion :

$$K^{\star}\phi \quad = \quad (K\dot{-}(\neg\phi))+\phi \tag{2.13}$$

Cette équation est communément appelée « identité de Levi ». Elle stipule que réviser K par ϕ revient à contracter K par $\neg\phi$ et à étendre la théorie-résultat par ϕ. D'autres définitions ont, par la suite, été proposées. Nous reviendrons sur ce point. Contentons-nous, pour l'heure, d'une brève remarque. Nous voyons que l'identité de Levi rend la contraction première par rapport à la révision. En fait, rien n'empêche a priori d'imaginer la situation inverse et de supposer la révision première par rapport à la contraction. Nous pourrions, en particulier, nous donner la définition suivante, communément appelée « identité de Harper » :

$$K\dot{-}\phi \quad = \quad K\cap K^{\star}\neg\phi \tag{2.14}$$

Intuitivement, (2.14) paraît moins claire que (2.13). Cela explique que, en général, (2.13) soit plus volontiers adoptée.

2.2.5.2 Enracinement

Notre problème initial ne fait que se déplacer. Analysant la révision en termes de deux autres opérations élémentaires, l'expansion et la contraction, nous devons à présent caractériser formellement ces dernières. Le cas de l'expansion ne pose guère de difficulté. Il suffit de définir $K+\phi$ comme correspondant à la fermeture déductive de $K\cup\{\phi\}$. Le cas de l'opération de contraction est plus délicat. Une difficulté évidente naît lorsque l'énoncé à éliminer est conséquence logique de plusieurs autres énoncés. La procédure doit nous dire lesquels éliminer et lesquels conserver (le changement sur K doit bien sûr être minimal). Proposée par Gärdenfors et Makinson [82], une stratégie possible consiste à supposer donnée une relation d'ordre $\leq_{EE}$ sur les éléments de K, relation dite d'« enracinement épistémique » (angl. *epistemic entrenchment*). Nous pouvons lire $\phi \leq_{EE} \psi$ comme signifiant que ψ est au moins aussi bien enraciné que ne l'est ϕ. La relation $\leq_{EE}$

vérifie en général les propriétés suivantes :

(EE1) $\phi \leq_{EE} \psi, \psi \leq_{EE} \xi \rightarrow \phi \leq_{EE} \xi$ (Transitivité)
(EE2) $\phi \vdash \psi \rightarrow \phi \leq_{EE} \psi$ (Dominance)
(EE3) $\phi \leq_{EE} \phi \wedge \psi$ ou $\psi \leq_{EE} \phi \wedge \psi$ (Conjunctiveness)
(EE4) $\phi \notin K \leftrightarrow \phi \leq_{EE} \psi$ pour tout $\psi \in K$ (Minimalité)
(EE5) $\psi \leq_{EE} \phi$ pour tout $\psi \in K \quad \rightarrow \quad \vdash \phi$ (Maximalité)

Notons $<_{EE}$ la relation d'ordre strict que $\leq_{EE}$ induit, en posant $\phi <_{EE} \psi \leftrightarrow \psi \not\leq_{EE} \phi$. Une première définition envisageable est la suivante[12] :

$$K \dot{-} \phi = \begin{cases} K \cap \{\psi : \phi <_{EE} \psi\} & \text{si } \phi \in K \text{ et } \nvdash \phi, \\ K & \text{sinon} \end{cases} \qquad (\mathrm{C}\dot{-})$$

(C$\dot{-}$) précise tout d'abord que, si ϕ ne figure pas dans K ou si ϕ exprime une loi logique, alors la contraction laisse K intact. La définition stipule ensuite que, lorsque ϕ figure dans K et que cet énoncé n'exprime pas une loi logique, alors le résultat de la contraction de K par ϕ nous donne les énoncés ψ qui sont strictement mieux enracinés que ϕ. Le fait que $<_{EE}$ soit irréflexive garantit que $K \dot{-} \phi$ ne contient plus ϕ. Si, à présent, nous nous tournons vers l'opération de révision K^* que $K \dot{-}$ induit *modulo* l'identité de Levi, nous obtenons la procédure suivante. Pour réviser K par ϕ, il faut commencer par supprimer non seulement $\neg\phi$, mais aussi tout énoncé qui n'est pas mieux enraciné que $\neg\phi$; après quoi, on ajoute ϕ et on ferme déductivement l'ensemble.

Nous pouvons ici nous demander s'il n'y a pas quelque chose d'artificiel dans la définition (C$\dot{-}$). La procédure est essentiellement destinée à nous aider à trancher entre plusieurs contractions ou révisions possibles. Supposons que l'énoncé $\phi := \neg a$ à retirer soit la conclusion (i.e. la conséquence logique) d'un ensemble $b_1, \ldots, b_n$ de prémisses qui, toutes, figurent dans K :

$$\underbrace{b_1, \ldots, b_n}_{\displaystyle \downarrow \atop \displaystyle \neg a}$$

Quelles prémisses conserver, demandions-nous, et quelles prémisses supprimer ? Il peut paraître étrange que, pour y répondre, nous devions comparer $\neg a$ avec chaque prémisse $b_{i\,(1 \leq i \leq n)}$, prise isolément des autres – ce à quoi nous invite (C$\dot{-}$). N'est-ce pas plutôt l'ordre de priorité entre les prémisses *elles-mêmes* qui

[12]Par simplicitié, nous ne donnons pas la définition originale de Gärdenfors et Makinson, mais une variante de Rott [211], qui nous paraît plus facile à manipuler. Pour retrouver la définition originale, il suffit de remplacer dans la définition que nous donnons ici $\phi \not<_{EE} \psi$ par $\phi <_{EE} \phi \vee \psi$.

doit être décisif ? La manière dont Alchourrón et Makinson [6, 7] définissent l'opération de contraction paraît, à cet égard, plus satisfaisante. Voici le principe de la procédure qu'ils suggèrent − elle porte le nom de procédure de « contraction sûre » (*safe contraction*). Soit < une relation d'ordre définie sur K. Dans un premier temps, on détermine les plus petites sous-théories de K qui impliquent l'énoncé ϕ que nous souhaitons retirer[13] :

$$\underbrace{b_1,\ldots,b_n}_{\downarrow \atop \neg a} \qquad \underbrace{c_1,\ldots,c_m}_{\downarrow \atop \neg a} \qquad \underbrace{d_1,\ldots,d_l}_{\downarrow \atop \neg a} \qquad \ldots\ldots$$

Dans un deuxième temps, passant en revue chacun de ces sous-ensembles, on retire les énoncés qui y sont minimaux sous $<$, c'est-à-dire les moins bien enracinés d'entre eux. Dans un troisième et dernier temps, on réunit les énoncés restants − ils sont qualifiés de « sûrs » (*safe*), et on ferme déductivement l'ensemble. Désignons par K/ϕ l'ensemble des $\psi \in K$ tels que ψ n'est minimal dans aucune des plus petites sous-théories de K qui impliquent ϕ. On pose donc :

$$K \ddot{-} \phi = Cn(K/\phi) \tag{C$\ddot{-}$}$$

Ici, $Cn(X)$ désigne l'ensemble des formules qui peuvent être classiquement inférées à partir de X. Il nous paraît utile de comparer la contraction sûre avec la procédure de contraction dite « par choix maximal » (*maxichoice*) ou encore « par intersection partielle » (*partial meet*). Celle-ci fonctionne à rebours : nous nous intéressons aux sous-théories maximales de K qui n'implique pas l'énoncé à retirer. Les arguments qui ne nous orientent pas vers celui-ci deviennent, pour reprendre le mot de Quine, notre principal gibier. La procédure est généralement définie comme suit. Notons $K\perp\phi$ (lecture : K moins ϕ) l'ensemble des sous-théories maximales de K qui n'implique pas ϕ. Soit γ une fonction de sélection associant à chaque proposition ϕ l'ensemble $\gamma(K\perp\phi)$ tel que

$$\begin{cases} \gamma(K\perp\phi) \subseteq K\perp\phi & \text{si } K\perp\phi \neq \emptyset, \\ \gamma(K\perp\phi) = K & \text{sinon.} \end{cases}$$

Intuitivement, il faut interpréter γ comme sélectionnant les « meilleures » sous-théories pour un certain critère. Puis on pose :

$$K \overset{\dots}{-} \phi = Cn(\bigcap \gamma(K\perp\phi)) \tag{C$\overset{\dots}{-}$}$$

La raison pour laquelle on prend l'intersection des éléments de $\gamma(K\perp\phi)$ plutôt que leur réunion peut être expliquée comme suit. Interagissant l'une avec l'autre, deux

[13]Intuitivement, ce sont nos différents arguments à l'appui à de ϕ.

sous-théories présentes dans $\gamma(K \perp \phi)$, disons x_1 et x_2, peuvent de nouveau entraîner ϕ. En règle générale, les données qu'il faut prélever dans x_1 pour restaurer ϕ sont différentes de celles qu'il faut prélever dans x_2. En prenant l'intersection de x_1 et x_2, plutôt que leur réunion, on a ainsi la garantie que ϕ ne réapparaîtra pas.

Nous ne saurions clore cette présentation sans évoquer, même allusivement, ce qu'on a coutume d'appeler les « postulats AGM ». Ceux-ci désignent les propriétés que satisfait en principe toute révision. Des théorèmes de représentation lient ces postulats aux différentes méthodes de révision décrites précédemment. Les postulats traditionnellement envisagés sont :

(K^*1) $K^*\phi$ est un ensemble de croyances (Closure)

(K^*2) $\phi \in K^*\phi$ (Success)

(K^*3) $K^*\phi \subseteq K + \phi$ (Expansion 1)

(K^*4) $\neg\phi \notin K \rightarrow K + \phi \subseteq K^*\phi$ (Expansion 2)

(K^*5) $K^*\phi = K_\perp \leftrightarrow \vdash \neg\phi$ (Consistency preservation)

(K^*6) $\phi \dashv\vdash \psi \rightarrow K^*\phi = K^*\psi$ (Extensionality)

(K^*7) $K^*(\phi \wedge \psi) \subseteq (K^*\phi) + \psi$ (Conjunction 1)

(K^*8) $\neg\psi \notin K^*\phi \rightarrow (K^*\phi) + \psi \subseteq K^*(\phi \wedge \psi)$ (Conjunction 2)

Le premier postulat signifie que $K^*\phi$ reste déductivement clos. Le deuxième postulat stipule que la formule ϕ persiste une fois la révision effectuée. Le troisième et le quatrième postulat ramènent la révision à une simple expansion, lorsque ϕ est consistant avec K. Nous pouvons les combiner et écrire :

(K^*3/K^*4) $\neg\phi \notin K \rightarrow K^*\phi = K + \phi$ (Expansion)

Le cinquième postulat signifie que l'opération de révision préserve la consistance de K. Le sixième requiert que la révision donne un résultat qui ne dépende pas de la forme syntaxique de ϕ. Enfin, les deux derniers recouvrent l'idée qu'en cas d'inconsistance de ϕ avec K, le changement sur K doit être minimal. Plus précisément, ils nous assurent que la révision par la conjonction de deux informations revient à une révision par la première information suivie d'une expansion par la seconde, si celle-ci ne contredit aucune connaissance issue de la première révision.

2.2.5.3 Théorie AGM et dialogue : quelques remarques

En elle-même, la contraction sûre semble plus proche de nos usages argumentatifs : la procédure travaille sur les données sur lesquelles s'appuie l'énoncé à éliminer. Walton et Krabbe [267], disions-nous[14], reprochent au modèle dialogique de Hamblin de ne fournir aucun renseignement sur la façon par laquelle le

[14]cf. page 34.

locuteur doit s'y prendre pour modifier son stock d'engagements. Ils apportent, disions-nous aussi, des éléments de réponses à cette question, qu'il serait intéressant de comparer avec celles qu'apportent la théorie AGM. L'analogue dialogique de la contraction est l'opération de rétraction sur un engagement. Il nous paraît ici significatif de constater que Walton/Krabbe la conçoivent sur le modèle de la contraction sûre, au moins dans ses grandes lignes. Par contre, ils utilisent une opération de révision sensiblement différente. En effet, l'une des règles que Walton/Krabbe [267, p. 150] se donnent stipule que, accusé de contradiction, un locuteur doit restaurer la consistance de son stock d'engagements, en se rétractant sur l'un des deux énoncés en conflit[15]. Cela revient pour le locuteur à − terminologie de Hansson [106] − « semi-réviser » son stock d'engagements, plus qu'à le réviser. Semi-réviser par ϕ, c'est introduire ϕ et contracter le résultat par $\phi \wedge \neg\phi$, sans exiger de la dernière information reçue ϕ qu'elle ait nécessairement la priorité. Hansson, il est vrai, travaille, non avec des théories (ensembles clos de croyances), mais avec des bases (ensembles non-clos). Et il opte pour une approche en termes d'intersection partielle. Hansson nomme « consolidation » l'opération qui consiste à contracter une base $\mathcal{H}$ par $\phi \wedge \neg\phi$. Notons cette opération « Cons $\mathcal{H}$ ». Nous avons une définition de la forme :

$$\text{Cons } \mathcal{H} = \bigcap \gamma(\mathcal{H} \perp \phi \wedge \neg\phi) \tag{2.15}$$

La fonction de sélection γ est définie comme précédemment. Désignons par $\mathcal{H}?\phi$ la procédure de semi-révision par ϕ qui lui est associée. Nous avons :

$$\mathcal{H}?\phi = \text{Cons } [\mathcal{H} \cup \{\phi\}] \tag{2.16}$$

Ici, nous commençons par introduire ϕ, puis nous restaurons la consistance de l'ensemble, en prenant l'intersection des meilleures sous-bases qui sont maximales consistantes. Nous aurons ϕ ou bien $\neg\phi$ comme output ; tout dépend de leur degré d'enracinement respectif. Dans (2.16), il est nécessaire de supposer que $\mathcal{H}$ ne désigne pas une théorie proprement dite, c'est-à-dire un ensemble de croyances clos pour la déduction classique. Autrement, semi-réviser des théories différentes par des formules différentes auraient nécessairement le même résultat.

2.2.5.4 Questions de sémantique

L'une des objections souvent adressée au cadre AGM est qu'il ne permet pas de traiter le problème de l'itération du processus de révision, car rien ne relie deux opérations de révision successives. Nous reviendrons sur ce thème au chapitre 3. En règle générale, les solutions proposées[16] font usage d'une analyse sémantique

[15] Il s'agit de leur règle structurale 3(a).

[16] Voici les plus signifi catives : la théorie des fonctions ordinales conditionnelles de Spohn [240], la révision rangée de Lehmann [135], la théorie des transmutations de Williams [268], l'approche de Darwiche & Pearl [54] et la révision naturelle de Boutilier [37].

particulière de la notion de révision. Il convient de rappeler le principe de cette analyse, du moins pour ce qui est du traitement que, au chapitre 3, nous convoquerons. On exploite ici le fait que, bien qu'initialement introduites pour résoudre des problèmes de gestion de l'inconsistance, les théories de la révision de croyances entretiennent une intime relation avec les théories de la non-monotonie. Pour passer d'un type de compte rendu à l'autre, on s'aide du biconditionnel suivant, qui porte généralement le nom de « test de Ramsey » : $\phi| \sim_K \psi$ si et seulement si $\psi \in K^*\phi$. L'intérêt premier de ce test est de donner *ipso facto* aux théories de la révision une sémantique de type préférentiel. Cette sémantique prend généralement appui sur un système de sphères *à la* Lewis. Soit W l'ensemble des mondes possibles. Désignons par $[K]$ les mondes qui vérifient les énoncés figurant dans K. On dit d'un ensemble \$ de sous-ensembles de W qu'il forme un système de sphères centré sur $[K]$, si \$ désigne une famille de sous-ensembles emboîtés de W (i.e. une famille de sous-ensembles de W linéairement ordonnés par rapport à la relation d'inclusion $\subseteq$) telle que $[K]$ est le $\subseteq$-minimum de \$. Intuitivement, les mondes à l'intérieur des sphères plus petites sont interprétés comme plus étant proches de $[K]$ (la sphère centrale) que les mondes à l'intérieur des sphères plus grandes. Si nous prenons au sérieux la notion de « changement minimal », nous nous attendons à ce que le résultat d'une révision soit aussi proche que possible de la théorie initiale K, sous quelque critère que ce soit. Nous aurons donc $\psi \in K^*\phi$ toutes les fois que ψ apparaît dans chaque monde de la plus petite sphère contenant ϕ. Ainsi, vérifier que $\psi \in K^*\phi$ revient bien à vérifier que la structure est un modèle de l'assertion conditionnelle $\phi \mid\sim \psi$. Il convient d'indiquer au passage que le test de Ramsey, sur lequel nous venons ici de jouer, soulève néanmoins un certain nombre de difficultés. Un théorème (généralement appelé « théorème d'impossibilité de Gärdenfors » $-$ du nom de son auteur) montre que, tel quel, ce test n'est pas compatible avec certains des postulats tout à l'heure évoqués, tels que le postulat (K^*2) de *success* et le couple de postulats (K^*3/K^*4), qui nous recommandent de conserver toutes nos croyances initiales, si l'adjonction de ϕ ne génère aucune inconsistance. Le théorème de Gärdenfors est à l'origine de nombreuses études, qui soit cherchent à reformuler la condition de Ramsey, soit remettent en cause l'un des postulats AGM en jeu. Le lecteur intéressé par cette question consultera avec profit l'étude de Lindström et Rabinowicz [145].

Quoique d'un usage courant, la sémantique que nous venons de présenter n'est pas la seule envisageable. Indiquons-en une autre, à laquelle il nous arrivera plus tard de faire référence[17]. Il s'agit du système DDL (pour *Dynamic Doxastic Logic*) de Lindström et Rabinowicz (voir e.g. [147]). Désignons par la lettre **B** la modalité de la croyance. Puis désignons par $[a]$ l'opérateur de la logique dynamique, lu « une fois l'action a exécutée, il sera nécessairement vrai que ». Lindström et

[17]Voir ci-dessous page 143.

Rabinowicz nous invitent à traiter l'opération AGM de révision comme un mixte de ces deux opérateurs. Aux expressions de la forme $\psi \in K^*\phi$, ils substituent des expressions de la forme

$$[*\phi]\mathbf{B}\psi,$$

dont le sens informel est

Révisant par ϕ, le locuteur croira nécessairement ψ.

Les conditions de vérité sont ici définies relativement à une paire $x = (w(x), h(x))$. L'élement $w(x)$ désigne ce que le monde actuel x contient objectivement. L'élément $h(x)$ renvoie à ce que l'agent croit de x. Typiquement, h a la forme d'un système de sphères ou $-$ dans le langage de Segerberg $-$ d'une hyperthéorie, de sorte que $h(x)$ code par là même une stratégie de révision. Comme dans une logique dynamique standard, on associe une relation binaire $R_a \subseteq W \times W$ à chaque paramètre modal a $-$ ici, à chaque révision. Puis on pose

$$x \models [*\phi]\mathbf{B}\psi \quad \text{ssi } \forall y\ ((x,y) \in R_{*\phi} \to y \models \mathbf{B}\psi), \tag{2.17}$$

c'est-à-dire

$$x \models [*\phi]\mathbf{B}\psi \quad \text{ssi } \forall y\ ((x,y) \in R_{*\phi} \to (z \in \cap h(y) \to z \models \psi)) \tag{2.18}$$

(2.17) précise que l'énoncé $[*\phi]\mathbf{B}\psi$ est vrai dans l'état x si et seulement si $\mathbf{B}\psi$ est vrai dans tout état y qui est accessible à partir de x modulo une révision par ϕ. Explicitant le sens de la clause « $\mathbf{B}\psi$ est vrai dans tout état y », (2.18) ajoute que ψ est vrai tous les mondes z de $\cap h(y)$, i.e. la sphère centrale du système de sphères associé au monde y. Intuitivement, cette sphère centrale regroupe les situations qui sont compatibles avec y.

Une remarque vient à l'esprit. Nous pouvons nous demander pourquoi faire précéder ψ de l'opérateur modal **B** et, corrélativement, scinder la situation présente x en une composante objective et une composante subjective ? Une réponse possible consiste à invoquer l'interprétation parfois défendue selon laquelle, telle qu'initialement conçue, la révision AGM décrit un mécanisme de changement de croyances dans un monde qui lui-même n'évolue pas, plus qu'elle ne modélise un changement affectant le monde lui-même. Cette interprétation fut proposée par Katsuno et Mendelzon. (Ils nomment « mise-à-jour » l'opération décrivant un changement du monde.). Pour justifier l'idée que la révision AGM décrit un mécanisme de changement de croyances, l'exemple de la disjonction est souvent invoqué. Soit la pièce d'un appartement. La seule information dont nous disposons est que, soit la porte est ouverte (p), soit la fenêtre est ouverte (q). Une nouvelle information nous parvient : quelqu'un a fermé la porte ($\neg p$). Ici, $K = Cn(\{p \vee q\})$

et, de toute évidence, $p \notin K$. En vertu du couple de postulats (K^*3/K^*4), nous avons donc :

$$K^*\neg p = Cn(\{q\}).$$

Intuitivement, la révision donne q parce que nous sommes dans un monde statique (i.e. $p \vee q$ reste vrai). Or, il s'avère que Lindström/Rabinowicz introduisent un axiome tout à fait voisin de (K^*3/K^*4) et qu'ils nomment « préservation ». Il s'agit de :

$$\neg\mathbf{B}\neg\phi \to (\mathbf{B}\psi \to [*\phi]\mathbf{B}\psi). \tag{préservation}$$

Intuitivement, cette loi précise que, si la nouvelle information est compatible avec les croyances de l'agent ($\neg\mathbf{B}\neg\phi$), alors celui-ci conserve toutes ses anciennes croyances (avant $\mathbf{B}\psi$; aussitôt après une révision par ϕ, $\mathbf{B}\psi$). Interagissant avec les autres lois du système, cette loi permet effectivement de déduire de

$$\mathbf{B}(p \vee q) \tag{2.19}$$

et de

$$\neg\mathbf{B}p \tag{2.20}$$

la conclusion :

$$[*\neg p]\mathbf{B}q \tag{2.21}$$

Tout d'abord, par application de l'axiome de préservation sur (2.19) et (2.20), il vient :

$$[*\neg p]\mathbf{B}(p \vee q).$$

Ensuite, l'analogue du postulat de *success* nous apprend que

$$[*\neg p]\mathbf{B}\neg p. \tag{succès}$$

Comme $\mathbf{B}$ et $[*\phi]$ obéissent aux lois d'une logique modale normale, nous pouvons former la conjonction des deux résultats, c'est-à-dire écrire

$$[*\neg p](\mathbf{B}(p \vee q) \wedge \mathbf{B}\neg p),$$

puis

$$[*\neg p]\mathbf{B}((p \vee q) \wedge \neg p),$$

qui donne

$$[*\neg p]\mathbf{B}q.$$

Dans cette mesure, l'approche de Lindström/Rabinowicz reste très proche du cadre AGM. Néanmoins, des difficultés nouvelles apparaissent. Par exemple, on montre que, après révision par p,

> la non-croyance (ou l'ignorance) que p et la non-croyance que $\neg p$

donnent

> la croyance que p

jointe à

> la croyance (de deuxième niveau) que p est l'objet d'une non-croyance (de premier niveau).

La vérification procède comme suit. Supposons que

$$\neg \mathbf{B} p \tag{2.22}$$

et que

$$\neg \mathbf{B} \neg p \tag{2.23}$$

D'un côté, par l'axiome de succès :

$$[*p]\mathbf{B} p \tag{2.24}$$

De l'autre, (2.23) signifie que p est consistant avec nos croyances. Supposons que (2.22) soit l'objet d'une croyance (de deuxième niveau). Il vient alors (axiome de préservation)

$$[*p]\mathbf{B}\neg \mathbf{B} p. \tag{2.25}$$

puis

$$[*p]\mathbf{B}(p \wedge \neg \mathbf{B} p). \tag{2.26}$$

On reconnaîtra dans « $p \wedge \neg \mathbf{B} p$ » l'expression du paradoxe de Moore, qui tenait pour bizarre une affirmation du type *Il fait beau et je ne crois pas qu'il fait beau*. Lindström et Rabinowicz [147] esquissent une solution qui, intuitivement, donne droit de cité aux effets rétroactifs d'une révision et qui, formellement, exige que nous passions à une logique modale dite bi-dimensionnelle[18]. Les auteurs disent de la formule $\mathbf{B}(p \wedge \neg \mathbf{B} p)$ que, prise en bloc, elle exprime « ce que, après révision par p, l'agent croit de l'état antérieur à la révision » [147, p. 21]. Une lecture approximative de la formule (2.26) serait « j'aurais dû croire en p - même si à ce moment là je n'y croyais pas ».

[18] Ce type de logique fut mise au point par Segerberg [230]. L'idée essentielle consiste à supposer que l'évaluation d'une formule dans un monde w fait aussi référence à un monde w'.

2.2.6 Conclusion

Tentons de résumer le chemin parcouru. Nous nous sommes tout d'abord tournés vers les principales analyses contemporaines de l'argumentation. Constatant que le thème de la non-monotonie y était récurrent, nous nous sommes ensuite tournés vers les logiques non-monotones et les théories de la révision. Nous nous sommes intéressés à celles-ci en elles-mêmes, car notre souci premier était de tenter de mieux apprécier le principe de ces formalismes. A présent, nous sommes capables d'examiner certaines des applications qui en ont été faites à l'étude de l'argumentation. A notre connaissance, ces tentatives d'articulation partent dans trois directions. Evoquons-les brièvement tour à tour.

Les logiques non-monotones et apparentées furent initialement conçues pour rendre compte de nos argumentations, lorsqu'elles mettent en jeu des connaissances génériques, c'est-à-dire des règles générales sujettes à exception, du type « si x alors normalement y ». Nous pouvons tout d'abord nous demander si elles peuvent rendre compte de formes d'argumentation mettant en jeu des notions plus complexes. Par exemple, Delgrande [56] et Lehmann [136] ont tous les deux consacré une étude au raisonnement fondé sur la notion de prototype. De même, Lewis [141], Shoham [233] puis d'autres à leur suite, se sont demandé comment rendre compte d'une argumentation causale. Dans un même ordre d'idée, Console et al.[52], Marquis [164, 165], Konolige [126], Becher et Boutilier [36] se sont posé la question de savoir comment modéliser l'abduction, forme d'argumentation nous permettant de dériver l'hypothèse expliquant au mieux un fait observé. Enfin, s'intéressant exclusivement au domaine de l'action, de nombreux auteurs ont défendu l'idée que, quelle que puisse être l'analyse correcte d'un plan d'action, d'un désir ou d'une obligation, nous pouvons être sûrs par avance qu'elle fera appel à la théorie des modèles préférentiels. Ainsi de Tan & Pearl [245] et de Hansson [103].

Chercher à s'engager dans la voie que nous venons brièvement d'évoquer est l'idée qui vient le plus immédiatement à l'esprit. A dire vrai, ce n'est pas le seul chemin que nous pouvons prendre. En voici un second, qu'un deuxième groupe de logiciens a voulu emprunter. Leur préoccupation commune est celle de savoir si, rendue non-monotone, la théorie ne peut pas rendre compte de la validité ou de la non-validité de ces inférences que, de prime abord, nous aurions plutôt tendance à considérer comme étant de nature « pragmatique » et, donc, comme n'étant pas affaire de logique. L'exemple de la théorie de la conversation de Grice est significatif. Ce dernier fait de l'annulabilité (*cancellability*) un trait définitoire des implicatures conversationnelles. Un auteur aussi éminent que Levinson en conclut que les implicatures

> « sont très différentes des inférences logiques, et ne peuvent être modélisées au moyen d'une relation sémantique telle que l'implica-

tion » [138, p. 115-116].

Les théories de la non-monotonie suggèrent la conclusion inverse. Reste à tester cette idée sur le cas particulier des implicatures. Tel est l'essentiel du propos de Gazdar [87, 88, 89], Perrault [193], Thomason [247, 248], ainsi que celui de Lascarides, Oberlander et Asher [129, 130, 131]. Gazdar prend pour objet d'étude les implicatures quantitatives, déclenchées par la maxime « soyez aussi informatifs que possible ». Perrault et Thomason étudient les implicatures qualitatives, déclenchées par la maxime « ne dites que ce que vous croyez être vrai et justifié ». Lascarides, Oberlander et Asher prennent pour objet d'étude les implicatures de manière, déclenchées par la maxime « soyez ordonnés ». La notion d'implicature n'est pas le seul exemple de concept pragmatique sur lequel ont essayé de travailler les logiciens de la non-monotonie. Par exemple, Mercer [166, 167, 168, 169] propose d'analyser la notion de présupposition en termes de défaut, afin de résoudre le problème dit de la projection des présupposés. De son côté, Kautz [124] suggère de reconstruire la dérivation searlienne d'un acte de langage indirect en combinant théorie de la reconnaissance des plans et théorie de la circonscription. Nous voyons que, sur les trois degrés que Hansson [104] dégageait dans la pragmatique, seule la pragmatique du premier degré, à savoir l'étude des symboles indexicaux, n'a pas été abordée. Cela est relativement compréhensible, s'il est vrai qu'un symbole indexical désigne une expression systématiquement ambiguë au sens où sa référence varie toujours avec le contexte d'énonciation.

Voici enfin une troisième direction de recherche possible. Nous avons vu que Toulmin [252] reprochait à la logique formelle de ne pas envisager l'argumentation dans son aspect « procédural ». Comme certains commentateurs l'ont souligné, Toulmin n'explicite pas vraiment la nature du nouveau modèle qu'il propose. Une piste possible est la suivante. Reprocher à la logique formelle de ne pas envisager l'argumentation dans son aspect dynamique, c'est simplement lui reprocher de ne pas tenir compte du fait que, s'inscrivant dans un environnement où l'information évolue, toute argumentation passe inévitablement par une série de révisions successives. Décrire formellement cet aspect dynamique de l'argumentation, tel est le programme commun d'un corps de théories récentes groupées sous la désignation générale de théorie des systèmes argumentatifs. Pollock [194, 195, 196, 197, 198] est le principal représentant de ce courant de recherches, qui vise à l'élaboration d'une théorie générale de l'interaction des arguments, en termes d'attaque et de contre-attaque.

Le but des deux prochaines sections est de poser les jalons d'un rapprochement entre l'étude de l'argumentation et celui de la révision. Il serait bien ambitieux de vouloir exposer chacun des travaux que nous venons de mentionner. Nous nous limiterons à ceux qui nous paraissent les plus significatifs. Dans l'exposé de ceux-ci, nous suivrons le plan qui nous a paru le plus naturel pour en faciliter l'accès au lecteur profane. La section 2.3 traitera de l'utilisation de la non-monotonie dans le domaine de la pragmatique. Nous prendrons l'exemple des présupposés et celui des implicatures conversationnelles. La section 2.4 portera sur la théorie des systèmes argumentatifs. Nous axerons notre présentation sur le système fondateur de Pollock et sur celui de Prakken et Sartor [204].

2.3 Pragmatique

2.3.1 Présupposition et défauts

L'idée d'éclairer la notion de présupposition à la lumière de la non-monotonie date de Mercer [166, 167, 168, 169]. Ce style d'analyse s'oppose à celui en termes de trivalence ou de modèles partiels, que la méthode des super-évaluations de van Fraassen [75, 76] porta à maturité. Le principe du compte rendu trivalent est le suivant : pour que « p présuppose q » soit vrai, il faut que q soit encore affirmé par la négation de p. Autrement dit, la règle gouvernant l'usage de l'opérateur « présuppose » est que, si « q » est faux, alors « p » n'est pas faux, mais ni vrai ni faux, parce que hors de propos. Ce qui rend le passage à la non-monotonie nécessaire, c'est l'évidence suivante, qui frappe un certain nombre de linguistes et de pragmaticiens : la relation « présuppose » est défaisable, au même titre que la relation « implique ». Nous ne passerons pas en revue toutes les données linguistiques qui étaient cette hypothèse. Elles sont exposées dans Levinson [138]. Qu'il nous suffise ici d'évoquer le problème connu sous le nom de problème de la *projection des présuppositions*. Dans certains cas, une expression propositionnelle complexe hérite des présuppositions de ses parties, dans d'autres cas elle n'hérite pas de ces dernières. Mercer [167] donne l'exemple des deux phrases disjonctives :

(a) John a cessé de battre le tapis ou Marie a cessé de battre le tapis

(b) John a cessé de battre les oeufs ou Marie a cessé de battre les oeufs,

que nous pouvons traduire par :

$$STOP(BEAT(John, r_1)) \vee STOP(BEAT(Mary, r_1))$$
$$STOP(BEAT(John, e_1)) \vee STOP(BEAT(Mary, e_1))$$

Il faut tout de suite préciser que, tel que nous le comprenons, l'exemple va jouer sur l'hypothèse (que nous croyons réaliste) selon laquelle une disjonction peut être prise tantôt au sens exclusif tantôt (quoique plus rarement) au sens inclusif. Voyons en quoi, bien que possédant la même forme logique, ces deux phrases ne se comportent pas nécessairement de la même façon en ce qui concerne les présuppositions de leurs parties. Dans (a), le premier membre disjoint présuppose que John battait le tapis. Supposons que le symbole « $\mid\!\sim$ » indique la relation « présuppose » − notation que, pour l'instant, nous introduisons uniquement pour suivre ce qui se passe. Nous avons ainsi :

$$\begin{aligned} &STOP(BEAT(John, r_1)) \mid\!\sim \\ &\qquad \exists t.\ DO(BEAT(John, r_1), t) \end{aligned}$$

Le second membre de la disjonction (a) présuppose que Marie battait le tapis :

$$\begin{aligned} &STOP(BEAT(Mary, r_1)) \mid\!\sim \\ &\qquad \exists t.\ DO(BEAT(Mary, r_1), t) \end{aligned}$$

Nos deux présuppositions se « projettent » sur la disjonction tout entière, au sens où nous avons simultanément

$$\begin{aligned} &STOP(BEAT(John, r_1)) \vee STOP(BEAT(Mary, r_1)) \mid\!\sim \\ &\qquad \exists t.\ (DO(BEAT(John, r_1)), t) \end{aligned}$$

et

$$\begin{aligned} &STOP(BEAT(John, r_1)) \vee STOP(BEAT(Mary, r_1)) \mid\!\sim \\ &\qquad \exists t.\ DO(BEAT(Mary, r_1), t) \end{aligned}$$

A première vue, la règle qui est ici en jeu peut être traduite :

$$\frac{\phi_1 \mid\!\sim \psi_1 \quad \phi_2 \mid\!\sim \psi_2}{\phi_1 \vee \phi_2 \mid\!\sim \psi_1 + \psi_2}\ \mathrm{proj}_1$$

Nous utilisons l'expression « $\phi_1 \vee \phi_2 \mid\!\sim \psi_1 + \psi_2$ » comme une abréviation de « $\phi_1 \vee \phi_2 \mid\!\sim \psi_1$ et $\phi_1 \vee \phi_2 \mid\!\sim \psi_2$ ». On reconnaîtra dans le passage de « $\phi_1 \vee \phi_2 \mid\!\sim \psi_1 + \psi_2$ » à « $\phi_1 \vee \phi_2 \mid\!\sim \psi_1 \wedge \psi_2$ » une instance de la règle « et » du système d'inférence non-monotone **C**[19]. Cette forme de projection n'aurait évidemment pas de sens, si les deux membres disjoints ne pouvaient pas être simultanément vrais.

A présent, considérons (b). Nous pouvons penser que la dénotation du terme « oeufs » ne varie pas d'une occurrence à l'autre. Il est également raisonnable de

[19]Un tel passage est légitime, si nous travaillons dans le cadre d' une théorie de l'inférence non-monotone généralement qualifi ée de « sceptique » : seules sont retenues les formules présentes dans toutes les extensions de la théorie. Un tel passage n'est pas légitime, si nous travaillons dans le cadre d'une théorie de l'inférence non-monotone dite « crédule », c'est-à-dire lorsque sont retenues les formules présentes dans au moins une extension de la théorie. Nous souhaitons d'autant moins prendre parti pour l'une ou l'autre de ces options que, comme nous allons le voir, la procédure imaginée par Mercer intègre une forme d'argumentation par cas généralement absente des théories traditionnelles de l'inférence non-monotone.

supposer que des oeufs sont toujours battus par une personne à la fois, c'est-à-dire que la disjonction est ici prise au sens exclusif. Ici, les présuppositions des énoncés disjoints ne se projettent pas de la même façon sur la disjonction ; car nous avons soit

$$STOP(BEAT(John,e_1)) \vee STOP(BEAT(Mary,e_1)) \mid\!\sim \\ \exists t.\, DO(BEAT(John,e_1),t),$$

soit

$$STOP(BEAT(John,e_1)) \vee STOP(BEAT(Mary,e_1)) \mid\!\sim \\ \exists t.\, DO(BEAT(Mary,e_1),t),$$

mais non les deux. Au schéma « proj_1 », se substitue le schéma :

$$\frac{\phi_1 \mid\!\sim \psi_1 \quad \phi_2 \mid\!\sim \psi_2}{\phi_1 \vee \phi_2 \mid\!\sim \psi_1, \psi_2}\ \text{proj}_2$$

Nous utilisons l'expression « $\phi_1 \vee \phi_2 \mid\!\sim \psi_1, \psi_2$ » comme une abréviation de « $\phi_1 \vee \phi_2 \mid\!\sim \psi_1$ ou $\phi_1 \vee \phi_2 \mid\!\sim \psi_2$, mais pas les deux ».

Le problème de la projection des présupposés est généralement présenté comme un argument contre l'analyse en termes de trivalence ou de modèles partiels. Dans notre exemple, nous avons choisi de représenter la relation « présuppose » à l'aide du symbole $\mid\!\sim$. Afin de justifier ce choix, il faut maintenant exposer plus en détail la théorie de Mercer, et montrer comment elle décrit le mécanisme de la projection des présuppositions. Nous illustrerons notre propos, à l'aide du couple de phrases (a)-(b). Comme nous l'avons dit, la théorie de Mercer fait appel à la logique des défauts[20]. Soit un énoncé u. Désignons par KB_H la base de connaissance de l'interlocuteur avant qu'il n'asserte u. KB_H contient non seulement un ensemble de règles strictes, mais aussi un ensemble de défauts normaux, au nombre desquelles figurent les règles génératrices de présuppositions. Dans le cas de (a), nous avons :

- Formule :

  $F_1 : STOP(e) \rightarrow \exists t_1\, \exists t_2.\, t_1 < t_2 \wedge DO(e,t_1) \wedge \neg DO(e,t_2)$

  Lecture : « Pour qu'un évènement e cesse, il faut qu'il existe un instant t_1 au cours duquel e se produit, ainsi qu'un instant t_2 postérieur à t_1 et auquel e ne se produit pas ».

[20] Nous l'avons présentée au paragraphe 2.2.1, page 56.

- Défaut :

$$d : \frac{\neg Stop(e) : \exists t.\, Do(e,t)}{\exists t.\, Do(e,t)}$$

Lecture : « Si e ne cesse pas et s'il est cohérent d'admettre l'existence d'un instant t au cours duquel e se produit, alors on peut conclure qu'un tel instant existe ». Ainsi, préfixer le verbe *cesser* d'une négation ne touche normalement pas les présupposés de *cesser*.

Dans le cas de (b), nous avons aussi :

- Formule :

$$(\exists t_1\ DO(BEAT(x,z_1),t_1) \wedge \exists t_2\ DO(BEAT(y,z_2),t_2) \wedge x \neq y) \rightarrow z_1 \neq z_2$$

Lecture : « Si au moment t_1 il se trouve que x bat z_1 et si au moment t_2 il se trouve que y bat z_2, alors les objets ou les êtres que x et y battent sont distincts l'un de l'autre».

Ce qui est équivalent à :

$$F_2 : (z_1 = z_2 \wedge x \neq y) \rightarrow (\exists t_1\ DO(BEAT(x,z_1),t_1) \rightarrow \neg\exists t_2\ DO(BEAT(y,z_2),t_2))$$

Cela vient traduire le fait que, dans le cas de (b), nous avons la donnée supplémentaire selon laquelle les oeufs sont supposés être battus par une seule personne à la fois. C'est cette formule qui, dans le cas de (b), viendra bloquer l'application du défaut susmentionné (c-a-d la projection du présupposé).

Intuitivement, $Th(CSQ(D))$ - i.e. la clôture déductive de l'ensemble des conséquents des défauts, énumère tous les « candidats » potentiels à la fonction de présupposition. Le test de consistance (avec le contexte) auquel nous devons les soumettre chacun revêt la forme d'une argumentation à partir de cas. Pour plus de clarté, il est préférable de décomposer la procédure en deux moments.

Dans un premier temps – étape simplement préliminaire, il faut ajouter au contexte ce que Gazdar nomme les « implicatures clausales » de u. Elles désignent, rappelons-le [21], ce que l'interlocuteur est en droit de conclure de l'énonciation de u, en se fondant sur l'hypothèse selon laquelle le locuteur a été aussi informatif que possible. Voici, selon Mercer, l'ensemble des implicatures clausales que déclencherait (respectivement) une disjonction et une conjonction :

Enoncé	Implicatures clausales
<1> ϕ *ou* ψ	$P_S(\neg\phi \wedge \psi), P_S(\phi \wedge \neg\psi)$
<2> ϕ *et* ψ	$P_S(\neg\phi \wedge \psi), P_S(\phi \wedge \neg\psi)$

[21]Cf. page 18 et *sqq*.

'P_S' signifie « le fait que ... est compatible avec ce que le locuteur sait » (possibilité épistémique). Nous pouvons justifier <1> de la façon suivante. Qui affirme « ϕ *ou* ψ » ne croit visiblement ni que $\phi \rightarrow \psi$ ni que $\psi \rightarrow \phi$; car, si tel était le cas, il serait plus informatif (coopératif) de simplement affirmer ϕ, ou de simplement affirmer ψ. Nous pouvons justifier <2> d'une façon similaire [22]. Nous voyons que Mercer prend la notion d'implicature clausale en un sens plus fort que ne le faisait Gazdar.

Dans un deuxième temps, il faut envisager tour à tour la réalisation des états de chose que *u* implicite à titre de simple possibilité. Dans le cas d'une disjonction comme (a) et (b), nous sommes ainsi conduits à considérer successivement les deux hypothèses suivantes : $\phi \wedge \neg\psi$ et $\neg\phi \wedge \psi$. Puis, nous remémorant notre liste initiale de « candidats » potentiels à la fonction de présupposé, nous devons garder ceux que chacune de ces hypothèses permet de démontrer par défaut, et uniquement ceux-là [23]. La figure ci-dessous décrit le mécanisme par lequel (b) hérite seulement de

$$\exists t.\, DO(BEAT(John, e_1), t)$$

ou de

$$\exists t.\, DO(BEAT(Mary, e_1), t).$$

On envisage ici le cas où $\phi \wedge \neg\psi$. Le trait en double indique l'application d'une loi stricte. La flèche « $\rightarrow$ » indique l'application d'un défaut. Cette dernière est ici bloquée, ce qu'explicite la croix :

$$\begin{tikzcd}
STOP(BEAT(John,e_1)) \arrow[d, equal, "F_1"'] & & \neg STOP(BEAT(Mary,e_1)) \arrow[d, "d"'] \\
\exists t.\, DO(BEAT(John,e_1),t) \arrow[d, equal, "F_2"'] & {} \arrow[d, phantom, "\perp"] & \exists t.\, DO(\cancel{BE}AT(Mary,e_1),t) \\
\neg\exists t.\, DO(BEAT(Mary,e_1),t) & &
\end{tikzcd}$$

FIG. 2.1 – le présupposé bloqué

Lorsqu'on envisageons le cas où $\neg\phi \wedge \psi$, l'argumentation est similaire, mais inversée.

[22] Supposer que « ϕ *et* ψ » et que « ϕ *ou* ψ » déclenchent les mêmes implicatures clausales n'explique pas pourquoi le locuteur choisit d'utiliser l'une plutôt que l'autre. C'est ici qu'intervient la notion d'implicature scalaire.

[23] Nous passons sous silence trois autres conditions que doivent remplir les candidats ; elles ne sont pas essentielles à notre propos.

Intuitivement, si asymétrie il y a entre (a) et (b) , elle tient apparemment au fait que (a) prend la disjonction au sens inclusif, alors que (b) la prend au sens exclusif. Ceci ne nous semble pas incompatible avec l'idée que, dans nos argumentations quotidiennes, nous la prenons *généralement* au sens exclusif. Car cette dimension est prise en charge, non pas par le clausal, mais par le scalaire qui du reste est défaisable. Cela n'empêchera pas certains de douter de l'idée que le problème de la projection des présupposés puisse constituer un véritable enjeu pour la théorie de l'argumentation. Par exemple, Ducrot [59] soutient l'idée selon laquelle, lorsqu'un énoncé A est enchaîné à un énoncé B, le lien entre A et B ne concerne jamais ce qui est présupposé, mais seulement ce qui est posé par A et par B. Dans cette optique, la fonction essentielle des présupposés est seulement de fixer le cadre de cohérence du discours. De notre côté, nous n'avons pas la prétention de répondre à cette question. Indiquons simplement que nous aurions pu prendre des exemples qui n'utilisent pas la disjonction. Une illustration voisine nous est ainsi fournie par cet ensemble de constructions que Lakoff [128] désigne sous le nom de « qualifications ». Comparons (c) et (d) :

(c) My cousin is not a bachelor
(d) My cousin is a male
(e) My cousin is not a bachelor, because my cousin is a female.

(e) consiste en (c) plus l'adjonction du syntagme qualificateur « because my cousin is female ». Bien que (c) présuppose (d), (e) ne présuppose pas (d).

2.3.2 Implicatures et actes de langage

Nous nous tournons à présent vers la notion d'implicature conversationnelle. On l'utilise parfois pour désigner une conclusion que l'on peut tirer d'une énonciation sans que la relation entre cette conclusion et l'énoncé en question puisse se ramener à une relation logique. L'un des apports les plus manifestes des théories de la non-monotonie est de montrer que cet usage est abusif. Nous avons déjà présenté la théorie de Gazdar, qui prend pour objet d'étude les implicatures quantitatives[24]. Contrairement aux auteurs que nous allons évoquer dans ce chapitre, Gazdar ne se tourne pas vers une logique non-monotone préexistante, comme la logique des défauts ou la théorie de la circonscription. C'est pourquoi nous avons introduit sa théorie au chapitre précédent. Nous axerons ici l'analyse sur la théorie de Perrault [193]. Celle-ci nous paraît intéressante, dans la mesure où elle n'aborde le thème des implicatures (en l'espèce qualitatives) que de biais, à l'occasion d'une réflexion plus générale sur la notion d'acte de langage. L'auteur prend pour point de départ la théorie des actes de langage de Cohen et Levesque,

[24] cf. section 1.2.5, page 16.

qui se veut une théorie de l'interaction rationnelle à part entière. A celle-ci, il reproche (à juste titre, croyons-nous) de ne pas expliciter la façon dont les croyances non-remises en cause par l'accomplissement de l'acte sont transportées de l'état mental précédent dans le nouvel état. Les logiques non-monotones permettent, précisément, de répondre à ce type de question. D'où l'intérêt que souleva l'étude de Perrault. Elle entraîna dans son sillage de nombreux autres travaux, tels que ceux de Appelt et Konolige [11], Sadek [220, 221], Herzig et Longin [109], et Thomason [248].

L'objet premier de Perrault est de modéliser la capacité des locuteurs à redéfinir leurs croyances en présence d'informations nouvellement communiquées au cours du dialogue. Perraul utilise trois sortes de modalités. Tout d'abord, il recourt à une modalité de nature épistémique, $B_{x,t}p$, lue « x croit en t que p ». Elle est supposée vérifier les axiomes du système modal S5 faible[25]. Ensuite, il utilise un opérateur exprimant l'intention, $I_{x,t}p$, lu « x a l'intention en t que p ». Ce dernier est supposé vérifier les axiomes caractéristiques d'une logique modale normale, ainsi que (principe davidsonien de charité oblige) les deux axiomes suivants :

$$I_{x,t}p \rightarrow \neg I_{x,t}\neg p$$
$$I_{x,t}p \rightarrow \neg B_{x,t}\neg p.$$

Le premier exprime l'idée que les intentions d'un agent sont cohérentes entre elles, et le second qu'elles sont également cohérentes avec ses croyances. Enfin, Perrault utilise un opérateur d'action, $DO_{x,t}\alpha$, décrivant l'accomplissement par x de l'acte α en t. En général, l'acte de langage α est lui-même identifié à son contenu propositionnel p – liberté dont Perrault s'autorise, dans la mesure où il se limite aux assertions (*declarative sentences*). Il est, notons-le, des cas où l'opérateur $DO_{x,t}$ sera appliqué à une expression de la forme $Obs(y)$. La formule alors obtenue, $DO_{x,t}Obs(y)$, décrira simplement l'observation par x du comportement de y.

Représentons le locuteur par la lettre S (pour *speaker*) et l'auditeur par la lettre H (pour *hearer*). Au moment $t = 0$ de l'énonciation, nous avons les données suivantes :

$$\begin{aligned} W = \{&DO_{S,0}p, \\ &DO_{S,0}Obs(H), \\ &DO_{H,0}Obs(S), \\ &DO_{H,0}Obs(H), \\ &DO_{S,0}Obs(S)\} \end{aligned}$$

[25] Il correspond au système S5 diminué de l'axiome de réflexivité.

Pour déterminer les effets de l'acte de langage – ici, une assertion de S, Perrault propose de recourir à deux séries de règles. La première contient deux règles strictes (non-défaisables) :

Principe de mémoire : si x croyait p avant l'accomplissement de l'acte de langage, il croit actuellement (après l'accomplissement de cet acte) qu'il croyait que p juste avant. Traduction :

$$B_{x,t}p \rightarrow B_{x,t+1}pB_{x,t}p.$$

Principe d'observabilité : si x accomplit l'acte α et si (à cet instant) y observe x alors (au moment suivant) y croit que x a accompli α. Traduction :

$$DO_{x,t}\alpha \wedge DO_{y,t}Obs(x) \rightarrow B_{y,t+1}DO_{x,t}\alpha$$

Combiné à la règle ci-après, le principe d'observabilité permet à chacun d'acquérir des croyances portant sur les croyances d'autrui.

La seconde série est composée de règles qui sont toutes défaisables :

Règle d'assertion : Si x accomplit un acte de langage de contenu p et que (à cet instant) sa croyance en p est consistante avec le reste de ses croyances, alors x croit que p. Notons $\Rightarrow$ l'implication valable par défaut. Ce postulat de sincérité s'exprime par :

$$DO_{x,t}(p) \Rightarrow B_{x,t}p$$

Naturellement c'est par le biais de cette règle que les implicatures qualitatives sont tirées.

Transfert des croyances : Si x croit que y croit que p et que (à cet instant) sa croyance en p est consistante avec le reste de ses croyances, alors x croit que p. Traduction :

$$B_{x,t}B_{y,t}p \Rightarrow B_{x,t}p$$

Persistance des croyances : si, après l'acte, x croit que, avant l'acte, il croyait que p, alors après l'acte il continue à croire que p – sous réserve que cette croyance soit consistante avec les croyances présentes dans le nouvel état. Traduction :

$$B_{x,t+1}B_{x,t}p \Rightarrow B_{x,t+1}p$$

Nous supposerons ici que l'axiome de mémoire et l'axiome de persistance peuvent être combinés sous la forme suivante :

$$B_{x,t}p \Rightarrow B_{x,t+1}p$$

Fermeture du savoir sous $\Rightarrow$: quels que soient l'agent et l'instant considérés,

Si $p \Rightarrow q$ alors $B_{x,t}p \Rightarrow B_{x,t}q$.

Intentionnalité : les actes de langages sont intentionnels

$$DO_{x,t}\alpha \Rightarrow I_{x,t}DO_{x,t}\alpha$$

Règle d'assertion II :

$$I_{x,t}DO_{x,t}p \Rightarrow B_{x,t}p$$

Clôture II : quels que soient l'agent et l'instant considérés,

Si $B_{x,t}p \Rightarrow B_{x,t}q$ alors $I_{x,t}p \Rightarrow I_{x,t}q$.

La raison pour laquelle Perrault introduit ces règles à titre de règles défaisables est aisée à comprendre. Par exemple – et pour en rester au cas le plus évident, la règle d'assertion, qui n'est autre que la maxime gricéenne de qualité, sera mise en échec si le locuteur croit que l'interlocuteur utilise une figure de rhétorique (ou trope). Pour Grice, l'ironie, la métaphore, la litote et l'hyperbole (en un mot : les tropes) peuvent être analysées à partir d'implicitations résultant de la violation apparente de la maxime de qualité.

Du point de vue formel, Perrault se tourne vers la logique des défauts de Reiter, afin de déterminer l'ensemble des effets d'un acte de langage – but qu'il s'est initialement fixé. En effet, cet ensemble peut être identifié à l'« extension » de la théorie (D,W), ainsi définie par Reiter : E est une extension de (D,W) si et seulement s'il existe une suite E_i telle que $E = \cup E_i$ où :

$$E_0 = W$$
$$E_{i+1} = Th(E_i) \cup \{\beta : \alpha \Rightarrow \beta \in D, \alpha \in E_i, \neg\beta \notin E\}$$

$Th(X)$ désigne l'ensemble des formules qui peuvent être déduites validement de X par les lois de la logique classique. Clos par déduction, E contient tous les conséquents des défauts applicables. A titre d'illustration, reprenons le cas où

$$\begin{aligned} W = \{&DO_{S,0}p, \\ &DO_{S,0}Obs(H), \\ &DO_{H,0}Obs(S), \\ &DO_{H,0}Obs(H), \\ &DO_{S,0}Obs(S)\}. \end{aligned}$$

Dans ce cas, la croyance mutuelle que p, i.e.

$$MB_{S,H,1}p \leftrightarrow (B_{S,1}p \wedge B_{H,1}p \wedge B_{S,1}B_{H,1}p \wedge B_{H,1}B_{S,1}p \wedge \ldots),$$

figurera dans l'extension de la théorie[26]. Ainsi les deux premiers membres de la conjonction sont-ils obtenus :

$$DO_{S,0}p \in W \overset{Assertion}{\rightarrow} B_{S,0}p \in E_1 \overset{Persistance}{\rightarrow} B_{S,1}p \in E_2 \overset{Transfert}{\rightarrow} B_{H,1}p \in E_3$$

Supposons que (tel qu'habituellement en logique des défauts) le symbole « $|\sim$ » indique, non plus la relation « présuppose », mais la relation « implicite qualitativement »[27]. Nous avons ici quelque chose de la forme :

$$\{DO_{S,0}p\} \mid\sim B_{S,0}p. \tag{2.27}$$

Naturellement, $|\sim$ n'est pas monotone. Il suffit de prendre le cas de figure où le locuteur n'est pas sincère, c'est-à-dire que $\neg B_{S,0}p \in W$. Dans ce cas, la dérivation de $B_{S,0}p$ (puis, du reste, celle de $B_{S,1}p$) est évidemment « bloquée » :

$$\{DO_{S,0}p, \neg B_{S,0}p\} \mid\not\sim B_{S,0}p.$$

2.3.3 Extensions multiples

Prenant appui sur une logique des défauts, la théorie de Perrault hérite de certaines de ses particularités. La principale d'entre elles est qu'une théorie peut admettre plusieurs extensions possibles. Supposons, fait ainsi observer Perrault, que le locuteur informe l'auditeur que p, lorsqu'il croit que celui-ci croit que $\neg p$:

i) $DO_{S,0}p$ ii) $B_{S,0}B_{H,0}\neg p$.

Intuitivement, on peut traiter ii) comme une implicature quantitative de i). Ici, la théorie admet deux extensions E' et E'', l'une contenant

iii) $B_{S,1}B_{H,1}\neg p$

et l'autre

iv) $B_{S,1}B_{H,1}p$ [28].

[26] Voir Perrault [193, p. 174]. Il s'agit d'une croyance mutuelle elle-même révisable. A supposer que la croyance mutuelle soit nécessaire à la communication, la question se pose de savoir si celle-ci peut être atteinte. Ce problème est au coeur des préoccupations de certains logiciens, tels que Halpern et Mose − voir e.g. [99]. Ils considèrent le cas d'un système distribué de processeurs − système d'agents en interaction comme un autre. Les variables propositionnelles expriment ici des énoncés du type « le processeur j est dans l'état e », et la séquence d'opérateurs K_iK_j est rendue par « le processeur n°i sait que le processeur n°j sait que ». Il se trouve que la notion de connaissance commune a partie liée avec certaines caractéristiques organisationnelles des systèmes distribués, telles que la synchronie (un système est dit synchrone lorsqu'une horloge interne permet à l'unité centrale d'envoyer des « top » à tous les processeurs simultanément, et asynchrone au contraire s'il n'y a que des temps locaux).

[27] La défi nition fut donnée p. 59. Il est ici indifférent que la relation soit prise au sens crédule ou sceptique.

[28] Lindström et Rabinowicz rencontraient quelque chose de plus ou moins similaire − la dimension dialogique en moins. Voir page 80.

Pour obtenir iii), on joue sur l'hypothèse selon laquelle les croyances de l'interlocuteur persistent normalement à travers le temps. Plus précisément :

1. $B_{S,0}B_{H,0}\neg p$ [Hypothèse]
2. $B_{S,1}B_{H,0}\neg p$ [Persistance sur S]
3. $B_{S,1}B_{H,1}\neg p$ [Persistance sur H, fermeture sous $\Rightarrow$]

Pour obtenir iv), on joue sur l'hypothèse selon laquelle le locuteur dit normalement le vrai, et sur l'hypothèse selon laquelle chaque interactant s'approprie normalement les croyances de l'autre. Plus précisément :

1. $DO_{S,0}p,\quad DO_{S,0}Obs(H),$ $DO_{H,0}Obs(S),\quad DO_{S,0}Obs(S)$ [Hypothèses]
2. $B_{S,1}B_{H,1}DO_{S,0}p$ [Observation - voir [193, p. 174]]
3. $B_{S,1}B_{H,1}B_{S,0}p$ [Assertion, fermeture sous $\Rightarrow$]
4. $B_{S,1}B_{H,1}B_{S,1}p$ [Persistance, fermeture sous $\Rightarrow$]
5. $B_{S,1}B_{H,1}p$ [Transfert, fermeture sous $\Rightarrow$]

iii) et iv) expriment deux croyances incompatibles, dans le sens où chacune est la contradictoire de l'autre (modulo les axiomes de S5 faible). Ainsi, la présence de l'une dans une extension bloque l'introduction de l'autre énoncé. Selon Perrault [193, p. 176], en l'absence de priorité entre l'axiome de transfert et l'axiome de persistance, nous nous attendrions plutôt à ce que la théorie admette une seule extension contenant la formule

$$B_{S,1}(B_{H,1}\neg p \vee B_{H,1}p),$$

indiquant par là que le locuteur reste indécis quant à la question de savoir si l'auditeur a modifié sa croyance concernant p.

Nous nous contenterons de signaler ce dernier problème, et de mentionner une direction selon laquelle on pourrait le traiter. Nous pouvons ici penser que la difficulté tient en partie à ce que la compétence des agents n'est pas prise en compte. Si le locuteur pense que l'auditeur pense que le locuteur est compétent concernant p, le locuteur pensera que l'auditeur s'est très certainement rangé à son avis (le contraire serait irrationnel de sa part). Et si le locuteur pense que l'auditeur pense que le locuteur n'est pas compétent concernant p, le locuteur pensera que l'auditeur n'a vraisemblablement pas changé d'opinion. Deux questions viennent naturellement à l'esprit. Tout d'abord, comment introduire dans le formalisme la notion de compétence ? Une réaction assez naturelle consiste à reprendre la définition que Sadek [221] propose :

$$Comp(x,p) \;=\; (B_{x,t}p \rightarrow p) \wedge (B_{x,t}\neg p \rightarrow \neg p).$$

Ici « $\rightarrow$ » désigne l'implication matérielle. En soi, il semble plus réaliste d'exiger de ce symbole qu'il représente une implication seulement plausible. La raison pour laquelle il n'est pas possible ici d'utiliser « $\Rightarrow$ » est que le formalisme de Reiter n'autorise pas à emboîter un défaut dans une modalité épistémique, i.e. nous ne pouvons pas écrire quelque chose de la forme :

$$B_{x,t}Comp(y,p).$$

Ensuite, la question se pose de savoir comment bloquer lesdites conclusions. Il serait ici tentant de reprendre l'approche de Froidevaux et Kayser [78], qui consiste à utiliser des prédicats d'assertion permettant de raisonner sur les inférences par défaut et de bloquer certaines d'entre elles. Décrite très succintement, la stratégie générale consiste à associer à chaque défaut d_i un prédicat d'assertion $R_i(x)$ signifiant « le défaut d_i peut s'appliquer à x ». Les différents principes régissant la relation interlocutive doivent être reformulés sous la forme de défauts semi-normaux, afin de pouvoir en contrôler explicitement l'application. Par exemple, un principe comme celui de transfert deviendrait [29] :

$$d_1 \quad := \quad \frac{B_{x,t}B_{y,t}p : B_{x,t}p \wedge R_1(x)}{B_{x,t}p}$$

Son application serait contrôlée par la présence d'un défaut du type :

$$d_2 \quad := \quad \frac{B_{x,t}\neg Comp(y,p) : \neg R_1(x) \wedge R_2(x)}{\neg R_1(x)}.$$

Dans notre exemple, une complication évidente vient du fait qu'il faut tenir compte des niveaux supplémentaires de connaissances imbriquées. Il faut aussi prendre garde au fait que le x de $R_i(x)$ est tantôt le locuteur tantôt l'interlocuteur. Aussi, nous ne poursuivrons pas plus avant l'analyse. Qu'il nous suffise d'avoir indiqué le principe d'un traitement possible. Nous voyons que l'idée (apparemment triviale) de traiter une implicature qualitative comme le résultat d'une inférence par défaut se révèle moins facile à mettre en oeuvre qu'il n'y paraît, si du moins nous dépassons le simple cadre du locuteur et si nous tentons d'élargir notre propos à la relation interlocutive.

2.3.4 Quelques développements ultérieurs

Appelt et Konolige, Sadek, Herzig et Longin, Thomasson ont proposé différentes variantes au système de Perrault. Les premiers recourent à une logique

[29] Nous réalignons en conséquence l'écriture du défaut sur la notation originale de Reiter. Cf. ci-dessus, p. 56 et *sqq.*

autoépistémique, les trois suivants à une logique dynamique et le cinquième à une logique intensionnelle enrichie d'un axiome de circonscription. Afin de mieux saisir l'objet de ces différentes théories, il est ici utile de se reporter à la classification des différentes sortes de dialogue que proposent Walton et Krabbe [267, p. 85]. Leur classification s'appuie sur deux critères : nature de la situation initiale ; nature du but poursuivi par les participants. L'origine du dialogue, nous disent-ils, peut être soit un conflit d'opinions, soit l'existence d'un problème, soit un manque d'information. Dans le premier cas, les participants peuvent soit vouloir déterminer laquelle des deux opinions est exacte (dialogue de genre PERSUASION ou DISCUSSION CRITIQUE), soit rechercher l'accord des parties, abstraction faite éventuellement de la valeur de vérité des deux propositions (dialogue de genre NÉGOCIATION). Ensuite, lorsque, à l'origine du dialogue, se trouve un problème, les participants peuvent vouloir soit lui trouver une réponse définitive (genre ENQUÊTE), soit seulement trouver un accord sur la base duquel agir (genre DÉLIBÉRATIF). Enfin, lorsqu'un simple défaut d'information est à l'origine du dialogue, les auteurs parlent du genre QUÊTE D'INFORMATION (*Information seeking*). De toute évidence, la théorie de Perrault et ses variations se présentent comme des modèles du dernier type seulement de dialogue. Elles s'intéressent au changement de croyance, en tant que celui-ci vise à corriger une asymétrie dans la répartition de l'information. Herzig et Longin [109] ont travaillé sur l'exemple d'application suivant. Les deux acteurs sont, d'une part, un distributeur de billets de chemin de fer (le système) et, d'autre part, un voyageur (l'utilisateur du système). Le second fournit un certain nombre de renseignements au premier (destination du voyage, classe de transport) qui, en retour, fournit une réponse à sa requête : quel est le prix du billet ? Le cadre d'analyse est ici une logique multi-modale dynamique. On commence par introduire dans un calcul des prédicats habituels des opérateurs de croyance et d'intention, *Bel* et *Intend*, tous deux indexés par la lettre u (pour « utilisateur ») et par la lettre s (pour « système »). On définit un acte de langage α par la donnée du couple

$$\alpha =< F_{u,s}, A > \text{ ou } < F_{s,u}, A >,$$

dans lequel F désigne la force illocutoire de l'acte (ici, informer) et A le contenu propositionnel. On associe à chaque acte de langage α un opérateur modal *Done*, de manière à former des expressions de la forme $Done_{\alpha}A$, ce qui est lu « l'acte α a été accompli, avant quoi il était vrai que A». Enfin, on définit un dialogue comme une suite $(\alpha_1, ..., \alpha_n)$ d'actes de langage, telle que chaque α_{k+1} transforme l'état S_k du système en un état S_{k+1} :

$$\mathrm{S}_0 \xrightarrow{\alpha_1} \mathrm{S}_1 \xrightarrow{\alpha_2} \cdots \longrightarrow S_{n-1} \xrightarrow{\alpha_n} S_n$$

Ici, le temps est supposé linéaire, aussi bien en direction de l'avenir que du passé. Pour passer d'état S_k à celui qui lui succède, on recourt au principe général suivant

$$S_{k+1} \rightarrow C \text{ ssi } LAWS \vdash (Done_{\alpha_{k+1}} Done_{\alpha_k} \ldots Done_{\alpha_1} S_0) \rightarrow C, \tag{2.28}$$

dans lequel $LAWS$ désigne les lois du domaine, au nombre desquelles figurent des lois gouvernant les actes de langages. Nous ne chercherons pas à les passer toutes en revue. Contentons-nous de mentionner l'une d'elles. Il s'agit du principe :

$$\neg Done_{<INFORM_{i,j},A>} \neg Bel_i \neg A. \tag{2.29}$$

(2.29) signifie que les interactants n'assertent que ce qu'ils croient. La présence de cette loi se comprend, eu égard au contexte d'application qui est celui de nos auteurs. A première vue, nous pourrions faire l'économie de cette loi, si nous cherchions à rendre compte d'un dialogue de type « négotiation ». Néanmoins, une question se pose peut-être, en liaison avec l'opérateur $Done_\alpha$. En tant que telles, les conditions de récurrence pour cet opérateur font référence à l'instant immédiatement précédent, qui est ici supposé unique et dans lequel le locuteur i soit croyait que A soit ne le croyait pas. Or, il arrive souvent que la maxime de sincérité exprime une simple présomption. Par là, nous voulons dire qu'il nous arrive de supposer d'autrui que, à tel moment du dialogue, il dit *normalement* le vrai. Qu'est-ce à dire, sinon que $Done_\alpha$ fait maintenant référence à une pluralité d'instants immédiatement passés, dont un « grand » sous-ensemble seulement contient $Bel_i A$? Ainsi, nous serions conduits à abandonner l'hypothèse (à première vue raisonnable) selon laquelle les évènements passés se situent nécessairement sur une même ligne.

2.4 Systèmes argumentatifs

Aux uns, les travaux que nous venons brièvement de présenter sembleront de portée limitée, tandis que d'autres y verront la réalisation d'une avancée remarquable, dans la mesure où ils confirment l'idée que la théorie logique a quelque chose à dire sur des phénomènes dont on lui refuse traditionnellement l'accès. Nous poursuivons cet essai de panorama, en nous tournant maintenant vers la théorie dite « des systèmes argumentatifs », qui se veut une théorie générale de l'interaction entre arguments. Dans un premier temps, nous présentons le système fondateur de Pollock (§ 2.4.1 à § 2.4.3), ainsi que celui (très suggestif) de Prakken et Sartor (§ 2.4.4 à § 2.4.7). Dans un deuxième temps, nous interrogeons la notion de réinstallation d'un argument, hypothèse sur laquelle ces approches reposent (§ 2.4.8).

2.4.1 Construction du système argumentatif

Voici l'idée essentielle de Pollock. La construction du système argumentatif se fait par les étapes successives suivantes :

1. On se donne une base de connaissance E = (*Input*, *Reason*). *Input* est un ensemble de « faits ». *Reason* désigne un ensemble de raisons, qui peuvent être de nature prima facie. Pour plus de clarté dans l'analyse, Pollock commence par mettre entre parenthèses le fait que celles-ci peuvent être de force différente.
2. On détermine l'ensemble (éventuellement infini) des arguments σ, η,... qui peuvent être construits à partir de E. Réduit à sa plus simple expression, un argument est une suite finie de couples de la forme
$$\sigma_i = <\Gamma, p>,$$
où p désigne la conclusion obtenue à l'étape i et Γ l'ensemble des hypothèses sur lesquelles l'argument se fonde. Désignons par ARG l'ensemble de tous les arguments. Il est défini au moyen de règles, dites « règles de formation ». Dans le cas le plus simple, le locuteur se contentera de tirer les conclusions des faits contenus dans la base (argument dit linéaire). L'ensemble Γ des hypothèses est alors vide, et seules deux règles interviennent. La première (notée **F**) autorise le locuteur à introduire à n'importe quelle étape de l'argumentation ses données de départ. Pollock note $\sigma^\frown x$ le résultat de l'opération consistant à augmenter σ de x. Ainsi cette première règle est-elle énoncée :
   > Soient $p \in$ *Input* et un argument σ. Alors, quel que soit Γ, $\sigma^\frown <\Gamma, p>$ est un argument.

   La seconde (notée **R**) énonce un principe de clôture par rapport aux raisons :
   > si σ est un argument dans lequel figurent $<\Gamma, p_1>$, ... et $<\Gamma, p_n>$, et si $<\{p_1,..., p_n\},q>$ est une raison, alors $\sigma^\frown <\Gamma, q>$ est un argument.

   Quatre autres règles sont introduites par Pollock, afin de rendre compte du raisonnement suppositionnel. Nous reviendrons sur ce point.
3. Ensuite, passant en revue chacun des arguments construits, on détermine ses « défaiseurs ». Pollock admet ici qu'un argument prima facie peut être défait de deux façons très distinctes. Tout d'abord, un argument peut être défait par annulation de sa conclusion ; Pollock qualifie alors son contre-argument de *rebutting defeater* :
   > σ est un *rebutting defeater* de η si η contient une ligne de la forme $<\Delta,q>$, et si σ contient une ligne de la forme $<\Gamma,\neg q>$, où $\Delta \supseteq \Gamma$.

Ensuite, un argument peut être défait par annulation de l'une des raisons prima facie qu'il utilise. Pollock qualifie alors le contre-argument de *undercutting defeater* :

> σ est un *undercutting defeater* de η, si η contient une ligne de la forme $<\Delta,q>$, obtenue à partir de lignes précédentes de la forme $<\Delta, p_1>$, ... et $<\Delta, p_n>$, par application de la règle **R**, et si la dernière ligne de σ est de la forme $<\Gamma,\neg((p_1 \wedge \ldots \wedge p_n) \rightarrow q)>$, où $\Delta \supseteq \Gamma$.

Pollock [196, p. 389] nomme « règle d'inclusion » (*subset rule*) le principe selon lequel l'ensemble Γ des hypothèses sur lequel s'appuie le contre-argument doit être inclus dans l'ensemble Δ des hypothèses sur lequel s'appuie l'argument qu'il défait.

4. Enfin, parmi tous les arguments « bien formés », on en distinguera certains, qui seront dits avalisés ou justifiés (*warranted*), parce qu'ils n'admettent pas de défaiseurs ou parce que leurs défaiseurs sont à leur tour défaits. La procédure à suivre est décrite ci-après.

2.4.2 Conséquence avalisée ou justifiée

Les concepts centraux d'une théorie de l'argumentation ne sont plus, comme en logique propositionnelle ordinaire, ceux de conséquence sémantique et de tautologie, mais ceux de conséquence avalisée ou justifiée (*warranted consequence*) et de proposition avalisée ou justifiée (*warranted proposition*). La seconde est définie comme un cas limite de la première :

> Une proposition p est justifiée, si elle est conséquence justifiée de l'ensemble vide des prémisses.

La notion de conséquence justifiée, quant à elle, est analysée [194, p. 490] en termes d'arguments valides (*in*) et non-valides (*out*) à différents niveaux (*levels*). Tout d'abord,

> Un élément de X (ensemble quelconque d'arguments) est dit valide au niveau 0, s'il ne se défait pas lui-même. Il est dit valide au niveau n+1 si, valide au niveau 0, il n'est défait par aucun autre élément de X valide au niveau n.

A strictement parler, il faut concevoir la notion de validité comme étant relative à l'ensemble X des arguments considérés. Ici, X désigne l'ensemble ARG de tous les arguments engendrables à partir de la base. En effet, Pollock adopte le point de vue d'un argumentateur « idéal », aux ressources cognitives illimitées. Nous avons ensuite la définition suivante :

> Un énoncé p est dit conséquence justifiée (*warranted consequence*) d'un ensemble Γ d'énoncés si et seulement s'il existe un argument σ de paire terminale $<\Gamma, p>$ qui possède la propriété remarquable d'être ultimement non-défait (*ultimately undefeated*) au sens suivant : étant valide à un niveau n quelconque, σ demeure tel à chacun des niveaux m supérieurs à n.

Le principe de la procédure imaginée par Pollock est le suivant. Au niveau 0, nous écartons les éventuels arguments qui s'auto-réfutent. Au niveau 1, nous écartons les arguments qui sont directement attaqués par un autre argument – ne conservant ainsi que les arguments sans défaiseur immédiat. Puis, à partir du niveau 2, nous récupérons progressivement tous les arguments que ces derniers « réinstallent » – en attaquant soit directement, soit indirectement, leurs défaiseurs..

Notons $\Rrightarrow_E$ la relation de conséquence justifiée[30]. En logique ordinaire, la question de savoir si une expression est ou non bien formée et celle de savoir si elle exprime ou non une vérité logique sont indépendantes l'une de l'autre. De même, la question de savoir si un argument est ou non bien formé et celle de savoir s'il est ou non ultimement non-défait (*ultimately undefeated*) sont indépendantes l'une de l'autre. C'est pourquoi, déterminer si $\Rrightarrow_E$ vérifie telle ou telle propriété revient à se poser deux questions :

1. partant d'un argument σ de ligne(s) ..., pouvons-nous construire un argument η de ligne(s) ... – en nous aidant des seules règles de formation du système ?
2. si σ possède la propriété remarquable d'être ultimement non-défait, η la possède-t-il aussi ?

Prenons pour exemple la règle de monotonie :

$$\frac{\Gamma \Rrightarrow_E \ p}{\Gamma \subseteq \Delta \Rrightarrow_E \ p}$$

L'une des originalités du système de Pollock est de chercher à rendre compte du raisonnement suppositionnel – qui consiste à introduire une hypothèse « pour les besoins de l'argument », pour ensuite la décharger en introduisant un conditionnel. Ceci conduit l'auteur à se donner la règle suivante, qu'il nomme *Foreign Adoptions Rule* :

> Si η est un argument ayant pour dernière ligne $<\Gamma, p>$, alors
> $\eta^\frown<\Delta \supseteq \Gamma, p>$ est un argument.

Imaginons que la base de connaissances nous donne les moyens d'engendrer un argument η_1 de paire terminale <{oiseau}, voler>. Cette règle nous autorise à prolonger η_1 de la paire <{oiseau, pingouin}, voler>. Nous avons répondu à la

[30] E désigne la base de connaissances.

première question. Passons à l'examen de la deuxième. Le principe d'inclusion tout à l'heure évoqué implique que, si tout *rebutting defeater* (resp. tout *undercutting defeater*) de η_1 est ipso facto un *rebutting defeater* (resp. un *undercutting defeater*) de $\eta_1^\frown$<Δ ⊇ Γ, p> , la réciproque n'est pas vraie. Supposons donc que la base de connaissances permettent d'engendrer un argument η_2 de paire terminale <{pingouin}, ¬voler> :

η_1 : ... <{oiseau}, voler>
<{oiseau, pingouin}, voler> η_2 : ... <{pingouin}, ¬voler>

En vertu du principe d'inclusion, η_2 attaque la ligne <{oiseau, pingouin}, voler>, mais il n'attaque pas la ligne <{oiseau}, voler>. De la sorte, à supposer que η_1 possède la propriété remarquable d'être ultimement non-défait, $\eta_1^\frown$<{oiseau, pingouin}, voler> ne la possèdera pas pour autant. En résumé, nous n'avons pas :

$$\frac{\{\text{oiseau}\} \Rrightarrow_E \{\text{voler}\}}{\{\text{oiseau, pingouin}\} \Rrightarrow_E \{\text{voler}\}}$$

Nous avons ici un peu de difficulté à comprendre la gêne que Pollock [194] éprouve à l'égard de la *Foreign Adoptions Rule*, sans laquelle le raisonnement suppositionnel ne pourrait évidemment pas démarrer. Embarrassé par le fait que « l'hypothèse nouvelle peut être un défaiseur pour un argument antérieurement construit » (p. 389), Pollock demande finalement (p. 390-391) de limiter sa théorie au raisonnement suppositionnel qu'il nomme « factuel » et dans lequel, restant compatible avec les faits, l'hypothèse envisagée n'obligera pas le locuteur à revenir sur ses précédentes conclusions.

L'une des idées qui vient naturellement à l'esprit consiste à tenter d'examiner le statut des règles d'inférence non-monotone, telles que celles du système **C**, que nous rappelons :

Equivalence logique à gauche	Si $\Gamma \Rrightarrow_E p$ alors $\Gamma' \Rrightarrow_E p$ pourvu que Cl(Γ)= Cl(Γ')
Affaiblissement à droite	$\Gamma \Rrightarrow_E p$ et $p \vdash q$ donnent $\Gamma \Rrightarrow_E q$
Réflexivité	$\Gamma \Rrightarrow_E p$ si $p \in \Gamma$
Conjonction	$\Gamma \Rrightarrow_E p$ et $\Gamma \Rrightarrow_E q$ donnent $\Gamma \Rrightarrow_E p \wedge q$
Monotonie prudente	$\Gamma \Rrightarrow_E p$ et $\Gamma \Rrightarrow_E q$ donnent $\Gamma \cup \{p\} \Rrightarrow_E q$

Nous préférons renoncer à cet examen, pour la raison suivante. Pour pouvoir dire si un argument est ou non avalisé, il faut avoir passé en revue tous ses contre-arguments potentiels. Dans la pratique, il nous paraît généralement très difficile d'être sûr d'avoir pris en compte *tous* les arguments qui peuvent être engendrés à partir de la base.

2.4.3 Défaisablement énumérable

Dans cette section, nous passons à l'examen d'une autre question, sur laquelle Pollock s'attarde plus longuement. Pour qu'un énoncé soit justifié, il est nécessaire que sa négation ne soit pas un théorème logique. Ainsi le système argumentatif que propose Pollock prête-t-il le flanc à l'objection traditionnellement adressée aux logiques non-monotones. Supposant un calcul des prédicats du premier ordre, qui est seulement semi-décidable (les théorèmes peuvent être énumérés récursivement, mais non les non-théorèmes), son système n'est pas décidable, au sens strict du terme. Pollock [198] tente de désamorcer l'objection, en distinguant les ensembles qui sont récursivement énumérables et ceux qui sont « défaisablement énumérables ». A la différence des premiers, les seconds doivent être approchés algorithmiquement non seulement par le bas, mais aussi par le haut. Sans doute l'ensemble des propositions justifiées (désignons-le par la lettre *J*) n'est-il pas récursivement énumérable. Mais − telle est la thèse de Pollock, cet ensemble n'en est pas moins défaisablement énumérable. D'un point de vue formel, Pollock formule cette thèse de la manière suivante. Supposons donnée une suite $\mathrm{S}_0, \mathrm{S}_1, \ldots \mathrm{S}_n$ de sous-ensembles (emboîtés) de ARG. Chaque S_i énumère l'ensemble des arguments construits à une étape (*stage*) donnée i. Supposons donnée une fonction $\Im$ énumérant le sous-ensemble de ceux qui sont justifiés. Elle est définie récursivement par[31] : $\Im(0) = Input = \mathrm{S}_0$; $\Im(n+1) = update(\Im(n))$. Dire que l'ensemble *J* des propositions justifiées est défaisablement énumérable revient à dire que :

$$(\forall x) \text{ si } x \in J \text{ alors } (\exists i)(\forall j > i)\, x \in \Im(j); \tag{2.30}$$

$$(\forall x) \text{ si } x \notin J \text{ alors } (\exists i)(\forall j > i)\, x \notin \Im(j); \tag{2.31}$$

Pour obtenir du récursivement énumérable, il faudrait remplacer (2.31) par

$$(\forall x) \text{ si } x \notin J \text{ alors } (\forall i)\, x \notin \Im(i) \tag{2.32}$$

La conjonction des clauses (2.30) et (2.31) signifie que, programmable, $\Im$ donne de *J* une approximation par le bas et par le haut. Nous allons dans un instant donner des exemples.

Une question vient naturellement à l'esprit. Qu'est-ce qui garantit que les deux types d'approximation, celle par le bas et celle par le haut, converge effectivement vers un même point ? La réponse de Pollock à cette question est double. Pour qu'il y ait bien convergence et que du défaisablement énumérable puisse être atteint, il faut tout d'abord, nous dit-il, que l'ensemble d'inputs n'autorisent pas la construction d'une chaîne infinie d'arguments, chacun venant défaire le précédent :

[31] Ceci est une défi nition donnée par Pollock. D'aucuns, comme Philippe Besnard, la trouveront suspecte (communication personnelle). Notre commentaire se placera à un niveau plus conceptuel que technique.

$$A_1 \xleftarrow{\textit{défait}} A_2 \longleftarrow \cdots \xleftarrow{\textit{défait}} A_m \longleftarrow \cdots$$

FIG. 2.2 – Chaîne infinie de contre-arguments

Qu'il faille que l'ensemble d'inputs n'autorisent pas la construction d'une chaîne infinie d'arguments et de contre-arguments paraît relativement évident. Imaginons que le système argumentatif contienne cette seule chaîne. De toute évidence, $J = \emptyset$, puisque l'ensemble des arguments « in » est alternativement $\{A_1, A_2, ... A_m, ...\}$ ou $\emptyset$, selon que le niveau (*level*) argumentatif dans lequel nous sommes correspond à un nombre pair ou impair. Cela caractérise la situation de l'argumentateur idéal, aux ressources cognitives illimitées. Adoptons à présent le point de vue plus réaliste d'un argumentateur qui construit ses arguments pas à pas. Supposons par exemple qu'il envisage les arguments dans l'ordre suivant : A_1, puis A_2,... puis A_m, ... Comme $J = \emptyset$, on s'attendrait − en vertu de (2.31) − à ce que chaque argument devienne injustifié (*non-warranted*) à partir d'une étape i et à ce qu'il demeure tel à toutes les étapes ultérieures. Le tableau suivant montre qu'il n'en est rien :

Etape (*stage*)	Construit(s)	Justifié(s) $\Im(i)$	Injustifié(s)
1	A_1	A_1	$\emptyset$
2	A_1,A_2	A_2	A_1
3	A_1, A_2A_3	A_1, A_3	A_2
4	A_1, A_2, A_3 A_4	A_2, A_4	A_1,A_3
5	A_1, A_2, A_3 A_4, A_5	A_1, A_3, A_5	A_2,A_4
$\vdots$	$\vdots$	$\vdots$	

A chacune des étapes portant un numéro pair, les arguments de numéro impair jusqu'ici construits sont écartés. Inversement, à chacune des étapes portant un numéro impair, les arguments de numéro pair jusqu'ici construits sont écartés. Aussi, quelle que soit l'étape à laquelle nous nous plaçons, l'ensemble des arguments injustifiés n'est jamais vide. Ici, le défaisablement énumérable n'est pas atteint, parce que la chaîne est infinie et n'est pas cyclique.

Le scénario précédent envisage le cas d'un argument « idéalement » injustifié, et met en échec la clause (2.31). Le scénario que voici porte sur un argument idéalement justifié et il met en échec la clause (2.30). Il s'agit du cas de figure où l'embranchement est lui-même infini [198, p. 24]. Soient un argument

A, ainsi que deux ensembles infinis d'arguments $X = \{A_1, A_2,...,A_n,...\}$ et $Y = \{B_1, B_2,...,B_n,....\}$. Imaginons, nous dit Pollock (p. 22), que A soit défait par chacun des éléments de X et que chacun des éléments de X soit, à son tour, défait par l'élément de Y de même indice que lui :

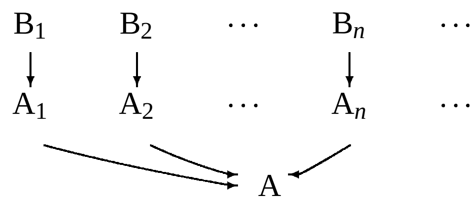


FIG. 2.3 – Embranchement infini

Dans ce cas, l'argument A compte au nombre des arguments justifiés, puisqu'il est réinstallé par chacun des éléments de Y. En vertu de (2.30), l'on s'attendrait à ce que :

$$(\exists i)(\forall j > i)\mathrm{A} \in \Im(j)$$

Mais imaginons que les arguments aient été construits un à un et dans l'ordre suivant : A puis A_1 puis B_1 puis A_2 puis B_2 puis ... puis A_n puis B_n,.... L'on vérifiera aisément que, dans ce cas, le statut de A change indéfiniment, à mesure qu'augmente le nombre des arguments pris en compte :

Etape (*stage*)	Construit(s)	Justifié(s) $\Im(i)$
1	A	A
2	A, A_1	A_1
3	A, A_1, B_1	A, B_1
4	A, A_1, B_1, A_2	A_2, B_1
5	A, A_1, B_1, A_2, B_2	A, B_1, B_2
6	A, A_1, B_1, A_2, B_2, A_3	
$\vdots$	$\vdots$	$\vdots$

A la première étape, seul est pris en compte l'argument A, qui est donc justifié. Les étapes suivantes se répartissent en deux groupes. La caractéristique des étapes portant un numéro pair est de construire un nouveau contre-argument à A, qui n'est donc justifié à aucune de ces étapes. Le propre des étapes portant un numéro impair est d'introduire un contre-argument à ce contre-argument, de sorte que A est justifié à chacune d'elles. Ainsi le statut de A change-t-il indéfiniment. Afin d'empêcher qu'un tel scénario ne se produise, Pollock propose d'exiger de tout système argumentatif qu'il vérifie la propriété suivante :

> « Pour tout énoncé p, il existe un ensemble fini S d'arguments de conclusion p tel que tout argument de conclusion p ne figurant pas dans S 'parasite' l'un au moins des élé-. ments de S, au sens où tous les contre-arguments au second sont aussi des contre-arguments au premier ». [198, p. 24]

Dans l'exemple imaginé par Pollock, l'indécision relative au statut de A est liée à une double possibilité. Nous pouvons, d'une part, introduire indéfiniment un *nouveau* contre-argument à A pour, d'autre part, aussitôt le désamorcer à l'aide d'un argument lui aussi *nouveau*. La contrainte que Pollock propose d'imposer au système argumentatif stipule que, tôt ou tard, il n'y aura pas de nouveau contre-argument envisageable, qui ne soit pas « redondant »[198, p. 23], dans le sens où il peut être désamorcé à l'aide d'un argument précédemment envisagé. Ceci est illustré par la figure suivante :

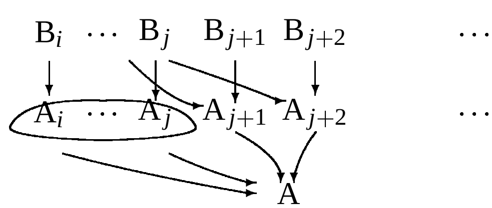


FIG. 2.4 – Forme parasite d'arguments

Dans cette figure, l'ensemble qui a été entouré correspond à S, i.e. S $=\{A_i,..., A_j\}$. En présence de S, les deux nouveaux contre-arguments A_{j+1} et A_{j+2} deviennent superflus (ce sont nos « parasites »), puisqu'ils peuvent être désamorcés à l'aide d'un argument attaquant un membre de S.

Présentant ses deux contraintes comme deux conditions suffisantes, pour que du défaisablement énumérable puisse être atteint, Pollock pose [198, p. 22] la question de savoir si ce sont aussi deux conditions nécessaires. Il semble qu'au moins la seconde ne l'est pas. Reprenant l'exemple que donne Pollock, nous pourrions imaginer que les arguments autres que A soient toujours construits deux par deux dans l'ordre suivant : $\{A\}$, $\{A, A_1, B_1\}$, $\{A, A_1, B_1, A_2, B_2\}$, $\{A, A_1, B_1, A_2, B_2, A_3, B_3\}$, ... Le tableau suivant montre que, si nous modifions l'ordre dans lequel les arguments sont pris en compte, l'indécidabilité relative à A disparaît, sans qu'il soit besoin de supposer une propriété de type « parasite » :

Etape (*stage*)	Construit(s)	Justifié(s) $\Im(i)$
1	A	A
2	A, A_1, B_1	A, B_1
3	A, A_1, B_1, A_2, B_2	A, B_1, B_2
4	A, A_1, B_1, A_2, B_2, A_3, B_3	A, B_1, B_2, B_3
5	A, A_1, B_1, A_2, B_2, A_3, B_3, A_4, B_4	A, B_1, B_2, B_3, B_4
$\vdots$	$\vdots$	$\vdots$

A la première étape, seul est pris en compte l'argument A, qui est donc justifié relativement à cette étape. La caractéristique des étapes ultérieures est d'envisager un nouveau contre-argument à A, en même temps que le contre-argument à ce contre-argument. Ainsi, A est justifié à chacune de ces étapes.

2.4.4 Théorie du dialogue

Cette section et les suivantes portent sur le système de Prakken et Sartor [204]. Sa sémantique est formulée en termes de point fixe. La théorie de la preuve du système épouse un style dialogique[32].

Le duel nous est présenté comme une partie entre un proposant (PRO) et un opposant (OPP), qui prend ici appui sur un système argumentatif donné. Un dialogue est défini comme une suite non-vide de coups (*moves*) M_i de la forme = $(Joueur_i, \mathrm{A}_i)$ et tels que :

1. $Joueur_i$ = PRO lorsque i est pair ; $Joueur_i$= OPP lorsque i est impair
2. Si $Joueur_i$= PRO ($i > 1$) alors A_i défait A_{i-1} mais l'inverse est faux (i.e. A_{i-1} ne défait pas A_i)
3. Si $Joueur_i$= OPP alors A_i défait A_{i-1} mais l'inverse peut être vrai.

Le proposant (PRO) ouvre la partie : choisissant un énoncé à défendre, il argumente en sa faveur (condition 1). L'opposant (OPP) prend la relève et tente de trouver un contre-argument, (conditions 1 et 3). S'il y parvient, le proposant tente de désamorcer le contre-argument de l'opposant (conditions 1 et 2), et ainsi de suite, jusqu'à ce que l'un des deux partenaires ne puisse plus contre-argumenter, auquel cas l'autre partenaire est dit « gagner » la partie dialogique. Il est possible, notons-le, que la réplique de l'opposant défasse celle du proposant, tout en étant défaite par cette dernière. Prakken et Sartor parlent ici de défaite « non-stricte ».

[32]Brewka [39], Dung [65], Vreeswijk [264] et Loui [150] (pour ne citer que les principaux) optent pour une approche tout à fait similaire. Le style d'analyse ici adopté n'a qu'une lointaine analogie avec celui d'une logique dialogique de type Lorenzen. En particulier, les coups joués au cours d'une partie ne suivront pas les règles d'un tableau sémantique.

Par contre, il est nécessaire que la réplique du proposant défasse celle de l'opposant, tout en n'étant pas défaite par celle-ci. Ils parlent ici de défaite « stricte ». En résumé,

$$\mathrm{B} \xrightarrow{\texttt{attaque}} \mathrm{A} \qquad\qquad \mathrm{B} \rightleftarrows \mathrm{A}$$

Défaite stricte *Défaite non-stricte*

Intuitivement, cela signifie essentiellement que le proposant doit nécessairement répondre à l'opposant en utilisant un contre-argument de force supérieure, tandis que l'opposant peut répondre en avançant un contre-argument de force seulement égale. Cette asymétrie s'explique aisément : l'opposant a pour seul rôle d'empêcher le proposant d'établir la vérité de l'énoncé de départ.

Une fois précisées les règles auxquelles les joueurs obéissent, Prakken et Sartor associent un arbre « dialogique » à chaque argument A du système argumentatif. Intuitivement, cet arbre décrit l'ensemble des parties que les partenaires peuvent jouer, si le proposant décide de défendre A. Chaque noeud correspond à un coup possible et chaque branche à une partie dialogique. En outre, il est exigé que :

> Si $Joueur_i$ = PRO, alors les successeurs immédiats de M_i sont tous des contre-arguments à A_i.

Etablir une preuve de A revient alors à s'assurer qu'il existe bien un arbre dialogique de noeud initial A et dont chaque branche décrit une partie gagnée par le proposant. N'étant pas défait, le dernier des arguments que le proposant avance réinstalle tous ceux qu'il a antérieurement défendus et, par là même, l'énoncé A. L'exemple suivant nous servira à illustrer le fonctionnement d'une partie. Supposons donné un système de quatre arguments A, B, C et D tels que, d'une part, B et D attaquent A et, d'autre part, C attaque B. Supposons que le proposant choisisse de défendre A. L'arbre dialogique (ou « de preuve ») à associer à ce système d'arguments est :

$$\begin{array}{ccc} & (\mathrm{PRO,A}) & \\ (\mathrm{OPP,D}) & & (\mathrm{OPP,B}) \\ & & | \\ & & (\mathrm{PRO,C}) \end{array}$$

L'argument A n'est pas justifié. En effet, le proposant gagne la partie seulement si l'opposant opte pour le contre-argument B.

2.4.5 Extension et point fixe

Un mot à présent sur la sémantique, qui est formulée en termes d'extension et de point fixe, au sens de Dung [66]. Un système argumentatif sous-tend un certain nombre de sous-ensembles (zéro, un ou plusieurs ensembles) d'arguments rationnellement acceptables. Ce sont eux qui correspondent aux « extensions » du système. Dung, il est vrai, distingue plusieurs sortes d'extension. Prakken et Sartor s'intéressent ici à l'extension dite « fondée » (*grounded*). Tout d'abord, un argument A est qualifié d'acceptable relativement à un ensemble S d'arguments si et seulement si :

> Quel que soit l'argument B, si B attaque A, alors B est à son tour défait par un élément de S.

Notons « arg » l'ensemble de tous les arguments et $\mathcal{P}(\text{arg})$ l'ensemble de ses parties. Un ensemble cohérent d'arguments est appelé « extension fondée », s'il correspond au plus petit point fixe de la fonction :

$$\begin{aligned} F : \mathcal{P}(\text{arg}) &\longrightarrow \mathcal{P}(\text{arg}) \\ S &\longmapsto F(S) = \{A \in \text{arg} : A \text{ acceptable relativement à } S\} \end{aligned}$$

Intuitivement, $F(S)$ regroupe l'ensemble des arguments que S peut défendre. L'extension « fondée » d'un système argumentatif correspond donc au plus petit ensemble d'arguments capable de se défendre lui-même contre toute attaque extérieure[33]. Comme F est monotone (pour l'inclusion ensembliste) et que tout opérateur monotone admet un plus petit point fixe (qui est unique), nous pouvons en conclure qu'un système argumentatif possède toujours une (et une seule) extension. Une fois celle-ci calculée, les arguments du système se répartissent en trois classes :

Justifi é (*Justifi ed*)	Annulé (*Overruled*)	Défendable (*Defensible*)
Elément de l'extension	Défait par un argument justifi é	Ni justifi é ni annulé

L'approche en termes d'extension fondée définit une forme sceptique de rationalité au sens où, en cas de conflit non-résolu, cette approche reste muette quant à la conclusion à adopter. En cela, elle s'oppose à l'approche en termes d'extension dite « stable » [66], qui définit une forme crédule de rationalité. Egalement définie en termes de point fixe pour un opérateur, l'extension stable se caractérise par le fait que tout argument figurant dans son complémentaire est défait par l'un au moins des arguments qu'elle contient.

Nous sommes maintenant en mesure de comprendre comment s'articulent l'approche dialogique et l'approche monologique en termes d'extension.

[33]Dans le cas particulier où tout argument possède un nombre infi ni de contre-arguments, cette caractérisation sera équivalente à une caractérisation en termes de dialogue. Nous aborderons ce point dans la section suivante.

2.4.6 Equivalence entre les deux approches

Tout d'abord, nous pouvons démontrer que, s'il existe un arbre dialogique de noeud initial M_1=(PRO, A_1=A) dont chaque branche décrit une partie gagnée par le proposant, alors A appartient à l'extension fondée du système argumentatif de départ. Cela se démontre par induction sur le nombre de coups joués au cours de la partie. Considérons une branche quelconque de l'arbre associé à A. Supposons-la de noeud terminal M_n=(PRO, A_n). Il nous faut montrer que :

1. Ultime argument avancé par le proposant, A_n appartient au plus petit point fixe de F ;
2. Précédemment défendu par le proposant, l'argument A_i appartient au plus petit point fixe de F, si l'argument A_{i+2} qu'il avança aussitôt après y appartient aussi.

Le premier point est immédiat. Que l'opposant ne puisse poursuivre la partie signifie que A_n n'a pas de défaiseur. Or un argument n'admettant pas de défaiseur est trivialement posé comme acceptable relativement à n'importe quel ensemble d'arguments,

$$(\forall S)(A_n \in F(S)).$$

Il appartient donc à ce S particulier possédant la propriété remarquable d'être le plus petit point fixe de F.

A présent, vérifions le second point. Supposons que A_{i+2} appartienne au plus petit point fixe de F, disons S' :

$$A_{i+2} \in S'.$$

Par construction, A_i est défait par A_{i+1} qui, à son tour, est défait par A_{i+2}. Cela signifie que A_i compte au nombre des arguments que S' défend,

$$A_i \in F(S'),$$

et donc appartient à S',

$$A_i \in S'. \quad \square$$

Reste à montrer que, inversement, si A appartient à l'extension du système argumentatif, alors il existe un arbre dialogique de noeud initial M_1=(PRO, A_1=A) dont chaque branche décrit une partie gagnée par le proposant. De fait, il est nécessaire de faire une hypothèse supplémentaire, pour pouvoir démontrer cette seconde proposition. Nous devons supposer le système argumentatif « finitaire » (*finitary*), au sens où chacun des arguments qu'il contient possède tout au plus un nombre fini de contre-arguments. Si le système argumentatif est finitaire, alors il est possible de calculer le plus petit point fixe de F pas à pas : on pose

1. $F^1 = F(\emptyset)$
2. $F^{n+1} = F^n \cup F(F^n)$.

Notons *Just* l'ensemble des arguments justifiés. On démontre que :

$$Just = \cup_{i=1}^{\infty}(F^i).$$

Ainsi, dans la pratique, nous calculerons l'extension fondée par application successive de F, en commençant par appliquer F à $\emptyset$. Le procédé est très proche de la caractérisation inductive de Pollock[34] : F^1 donne la liste des arguments sans contre-arguments immédiats ; F^2 rajoute la liste de ceux que les éléments de F^1 réinstallent ; et ainsi de suite.

A présent, supposons que A soit un élément de l'extension du système argumentatif. Pour construire un arbre dialogique de racine (PRO, A) et dont chaque branche représente une stratégie gagnante pour le proposant, on se sert de la suite d'approximations au termes desquelles A a été obtenu. Plus précisément, on s'aide de la sous-suite

$$(\emptyset, F^1, ..., F^i),$$

où F^i désigne le premier ensemble dans lequel A soit apparu. Pour construire ses répliques, le proposant va devoir parcourir « à l'envers » cette sous-suite. Comme A figure dans F^i, tous les contre-arguments A' aux contre-arguments à A figurent dans F^{i-1}. Il peut donc puiser dans cet ensemble sa première réponse à l'opposant. Mais, à leur tour, les contre-arguments A" aux contre-arguments à A' figurent tous dans F^{i-2}. Il puisera donc dans celui-ci sa seconde réplique, et ainsi de suite, jusqu'à atteindre l'ensemble F^1 qui, par construction, regroupe tous les arguments sans défaiseurs immédiats. Ainsi aura-t-il gagné la partie. Désignons par $F^{1 \leq x \leq i}$ un élément quelconque de la suite d'approximations. Indiquons par (PRO, $A' \in F^x$) le fait que le proposant puise sa réplique A' dans F^x. La procédure de construction de l'arbre dialogique peut être résumé ainsi :

> Mettre pour racine le noeud (PRO, A$\in F^{x=i}$)
>
> <u>Tantque</u> $F^x \neq F^1$ <u>faire</u>
> – prolonger tout noeud-parent (PRO,A'$\in F^x$) par autant de noeuds-enfants (OPP, B_1), ...,(OPP,B_m) que A' compte de défaiseurs, puis
> – prolonger chaque noeud (OPP, B_i) ainsi obtenu par autant de noeuds-enfants (PRO, C_1), ...,(PRO,C_k) que F^{x-1} contient de contre-arguments à B_i
>
> <u>FinTantque</u>

[34]Voir notre section 2.4.2.

2.4.7 Charge de la preuve

La force et l'élégance du système de Prakken et Sartor tient au théorème de complétude que nous venons d'évoquer. Si solide que puisse désormais paraître la théorie, on peut se demander de manière sensée si elle n'a pas pris un faux départ. Par exemple, une argumentation juridique oppose un demandeur (π, pour « plaignant ») et un défendeur (δ). Il n'est pas possible d'assimiler le premier au proposant et le second à l'opposant car, dans une argumentation juridique, le partage des rôles n'est pas fixé une fois pour toutes. Tel des deux est en position de proposant, si la charge de la preuve pèse sur lui, et en position d'opposant, dans le cas contraire. Cela conduit Prakken [203] à introduire un concept intermédiaire, celui de rôle dialectique. En voici la définition. Soient M_1,...,M_n une suite finie de coups de la forme M_i = ($Joueur_i$, A_i). Supposons donnée une « allocation » de charge de la preuve. Pour tout *i*, le rôle dialectique du $Joueur_i$, noté *Rôle*($Joueur_i$), est défini récursivement par :

1. Si $i = 1$ alors *Rôle*($Joueur_i$)=PRO
2. Si $i > 1$ alors
   (a) *Rôle*($Joueur_i$)=PRO si quelque sous-argument de A_i a pour conclusion un énoncé dont $Joueur_i$ doit établir la vérité,
   (b) sinon, *Rôle*($Joueur_i$) est l'opposé de *Rôle*($Joueur_{i-1}$).

Un dialogue avec charge de la preuve se présente alors comme une suite non-vide de coups de la forme M_i = ($Joueur_i$, A_i) tels que

1. $Joueur_i$ = π lorsque *i* est pair ; $Joueur_i$= δ lorsque i est impair
2. Si *Rôle*$(Joueur)_i$ = PRO ($i > 1$) alors A_i défait A_{i-1} (l'inverse étant faux)
3. Si *Rôle*$(Joueur)_i$ = OPP alors A_i défait A_{i-1} (l'inverse pouvant être vrai)

Comme précédemment, l'un des joueurs est dit gagner la partie, lorsque son partenaire ne peut plus jouer.

De toute évidence, le principal amendement apporté au modèle initial porte sur la nature des contre-arguments susceptibles d'être utilisés au cours de la partie. Nous avons vu que, aux coups de nombre pair, il fallait utiliser un défaiseur strict, c'est-à-dire un contre-argument de force nécessairement supérieure à l'argument attaqué[35]. Jouant aux coups de nombre pair, le plaignant doit ici utiliser un défaiseur strict seulement si la charge de la preuve pèse sur lui. Comme l'indique Prakken [203, section 5.3], pour réaliste qu'il soit, cet assouplissement a une conséquence fâcheuse : nous perdons le théorème de complétude, qui faisait la force de la théorie. Cette propriété est mise en échec, dans le cas de structures dont les arguments possèdent ce statut particulier qu'est le défendable (*defensible*). Voici un exemple :

[35] Voir plus haut p. 107.

$$C \rightleftarrows B \longrightarrow A$$

Structure avec arguments défendables

Comme précédemment, une flèche de la forme « C→B » veut dire que C attaque B. La fonction F admet quatre points fixes, $\emptyset$,{B},{A, C} et {A, B, C}, dont le plus petit est $\emptyset$. Aussi, les trois arguments ne sont ni (sceptiquement) justifiés ni annulés, mais ils sont défendables. Supposons que le demandeur choisisse de soutenir A. Dans le nouveau modèle, il gagne la partie, alors que dans l'ancien il la perd. La raison en est qu'il ne lui est plus nécessaire de répondre à une attaque à l'aide d'un défaiseur strict. Une fois que le défendeur a joué B, il peut donc jouer C. Nous prenons ici l'exemple d'une structure dont les arguments sont défendables. Cela ne veut pas dire que, dès lors qu'il a ce statut, un argument A correspond à une stratégie gagnante pour le demandeur. Il suffit d'imaginer que A est défait par B, et réciproquement.

2.4.8 Le principe de réinstallation

Nous nous tournons à présent vers une autre objection possible. Est dit « avalisé » l'argument qui est le noeud initial d'un arbre dont chaque branche décrit une partie gagnée par le proposant. N'étant pas défait, le dernier des arguments que le proposant avance réinstalle tous ceux qu'il a antérieurement défendus et, par là même, l'énoncé de départ. Que penser de ce principe « noyau », sur lequel s'appuie la procédure ? Considérons l'exemple suivant[36] :

> Les mollusques [M] sont normalement des coquillages [C_1] ; les céphalopodes [C_2] sont des mollusques mais ne sont normalement pas des coquillages ; les nautiles [N] sont des céphalopodes et sont des coquillages ; Fred [f] est un nautile.

Nous pouvons représenter ceci sous la forme d'un diagramme :

[36] Cet exemple est emprunté à la littérature sur la logique des défauts. Il fut introduit par Fahlman et al. [73].

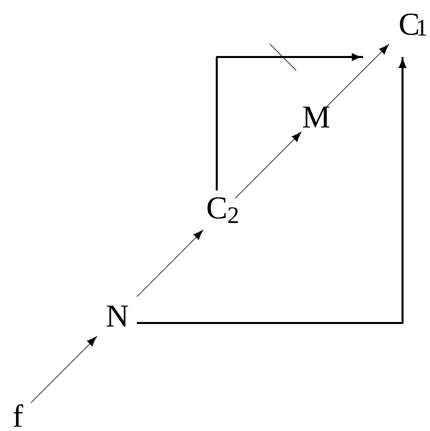


FIG. 2.5 – Fred est-il un coquillage ?

Dans ce diagramme, les liens entre les noeuds doivent être lus « de bas en haut ». Par exemple, le lien positif « f $\rightarrow$ N » représente l'assertion « f est un N » et le lien négatif « $C_2 \not\rightarrow C_1$ » représente l'assertion « C_2 n'est pas un C_1 ». Pour plus de simplicité dans l'écriture, une flèche de la forme $\rightarrow$ (resp. $\not\rightarrow$) désigne un lien aussi bien révisable que strict (c'est-à-dire, non révisable). Ce réseau d'héritage fournit autant d'arguments que les joueurs peuvent avancer. Admettons que le proposant opte tout d'abord pour le chemin A_1= f $\rightarrow$ N $\rightarrow$ C_2 $\rightarrow$ M $\rightarrow$ C_1, qui propose l'assertion f $\rightarrow$ C_1. Opter pour un tel chemin revient à défendre l'idée que f est un coquillage, puisque f est un mollusque. Pour contre-argumenter, l'opposant invoque la donnée selon laquelle f est un céphalopode, c'est-à-dire met en avant le chemin A_2= f $\rightarrow$ N $\rightarrow$ C_2 $\not\rightarrow$ C_1, qui propose l'assertion f $\not\rightarrow$ C_1. Le proposant désamorce l'objection, en invoquant la donnée selon laquelle f est un nautile, c'est-à-dire le chemin A_3 = f $\rightarrow$ N $\rightarrow$ C_1, qui propose l'assertion f $\rightarrow$ C_1. Intuitivement, rien ne nous empêche d'interpréteter un chemin du réseau comme un argument. Une question vient naturellement à l'esprit : quoique réinstallé par la procédure, l'argument (ou chemin) initial A_1 est-il bien pertinent ? La conclusion vers laquelle A_1 nous oriente et les données sur lesquelles il s'appuie sont en elles-mêmes correctes. Voilà apparemment la source de notre embarras : après le contre- argument de l'opposant, la troisième garantie utilisée ne peut plus faire office de passerelle, et autoriser le genre de passage qu'implique l'argument en question. Ainsi, la procédure ne tient pas compte du fait apparemment trivial que (pour parler comme Toulmin) les conditions de réfutation d'une garantie sont, elles aussi, révisables.

Dans l'exemple que nous venons d'aborder, l'argument réinstallé nous oriente vers une conclusion qui reste en elle-même correcte. Comme le suggère Horty [117], il existe également des cas de figures où la procédure déclare victorieux des arguments dont la conclusion est inexacte. Horty propose l'exemple suivant. Imaginons que toutes les assertions suivantes soient vraies :

> La plupart des employés de Microsoft [EM] sont millionnaires [1 M] ; la plupart des personnes qui sont récemment rentrées chez Microsoft [NEM] n'ont même pas un demi- million de dollars en poche [$\frac{1}{2}$ M] ; Beth [b], qui est récemment rentrée chez Microsoft, et qui vient de faire un riche héritage [H], a un demi-million de dollars en poche.

Sous forme de diagramme, cela donne :

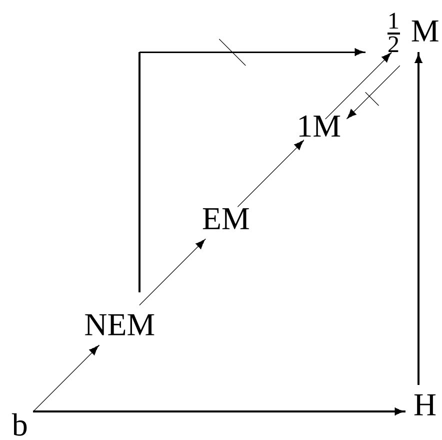


FIG. 2.6 – Beth est-elle millionnaire ?

Le contre-exemple de Horty joue sur le fait que la proposition (nous la désignons par la lettre p) Beth possède un million de dollars est logiquement plus forte que la proposition (nous la représentons par la lettre q) Beth possède un demi-million de dollars. Mais, si p est un argument pour q, il en résulte que ¬q est un argument pour ¬p, même si q n'en est pas un pour p. Imaginons donc que le proposant argumente tout d'abord en faveur de p : il met en avant le chemin $A_1 = \mathrm{b} \rightarrow \mathrm{NEM} \rightarrow \mathrm{EM} \rightarrow \mathrm{1M}$, qui propose l'assertion $\mathrm{b} \rightarrow \mathrm{1M}$. L'opposant contre-attaque en argumentant en faveur de ¬q, c'est-à-dire en attirant l'attention de son interlocuteur sur le chemin $A_2 = \mathrm{b} \rightarrow \mathrm{NEM} \not\rightarrow \frac{1}{2}\mathrm{M} \not\rightarrow \mathrm{1M}$, qui propose l'assertion $\mathrm{b} \not\rightarrow \mathrm{1M}$. Le proposant contre-argumente, en attirant à son tour l'attention de son partenaire sur le chemin $\mathrm{b} \rightarrow \mathrm{H} \rightarrow \frac{1}{2}\mathrm{M}$. Ayant argumenté en faveur de q (Beth possède un demi-million de dollars), le proposant a-t-il pour autant réinstallé la conclusion initiale p (Beth possède un million de dollars) ? Il est évident que non. Ici aussi, la difficulté semble tenir au fait que la procédure ne tient pas compte du fait que l'applicabilité de nos principes d'inférence (ou, si l'on préfère, leurs conditions de réfutation) est une donnée qui évolue en cours d'argumentation. L'opposant contre-attaque, en montrant que telle ou telle garantie G_1 ne s'applique pas au cas considéré. Le proposant désamorce l'objection, en montrant que la garantie G_2 invoquée par son adversaire ne s'applique pas non plus. Ainsi, le fait que, en situation d'incertitude, nos principes d'inférence se formulent d'une seule manière (« normalement les », une étape de la révision) ne devrait pas faire croire que les arguments vont simplement se relier en se réinstallant dès qu'un contre-argument est défait. Nous devons apparemment nous soucier de certaines contraintes qui guident l'itération de nos révisions. Rien ne dit, de plus, que l'on ne puisse faire jouer ces contraintes de diverses manières. Horty indique au passage que, dans l'exemple de Beth, une théorie traditionnelle de l'héritage des propriétés donnerait la conclusion correcte [37]. Il resterait à voir si la théorie des systèmes argumentatifs peut recourir à une approche de ce genre, sans perdre sa spécificité propre.

[37] Il existe une grande variété de mécanismes d'héritage. Horty songe à celui défendu dans [115].

Chapitre 3

L’obligation conditionnelle

3.1 Introduction

3.1.1 Motivations générales

Au chapitre précédent, nous avons constaté que les tentatives d'application des logiques non-monotones au domaine de l'argumentation partaient généralement dans trois directions distinctes :

– analyse de formes particulières d'argumentation ;
– interface sémantique/pragmatique ;
– théorie générale de l'interaction entre arguments.

Dans ce chapitre, notre intérêt va plutôt porter sur la première des directions empruntées. Plus précisément, notre propos concernera l'analyse du concept d'obligation en termes de structures préférentielles[1]. Tout le monde sera d'accord pour dire que les normes tiennent une place importante dans nos argumentations quotidiennes. Elles s'y manifestent à différents niveaux, et sous des formes très variées – jusque sous celle d'exigences et pressions implicites, dont nos conversations sont pleines. Cette observation nous offre un moyen d'aborder de biais la question du rapport entre logique et argumentation. Si la théorie logique limite ses prétentions au fondement des structures mathématiques, elle n'aura évidemment que faire de la considération de l'obligatoire et autres notions apparentées. Si par contre elle étend ses prétentions au fondement de nos argumentations quotidiennes, et n'accepte de renoncer a priori à aucune forme de rationalité, elle ne se reconnaîtra pas le droit de laisser en dehors d'elle la considération de telles notions.

Hansson [103] fut l'un des premiers logiciens à appliquer les sémantiques préférentielles à l'analyse des normes. Il le fit essentiellement dans le but de résoudre certaines des difficultés que soulève la formalisation du concept d'obligation réparatrice (*contrary-to-duty*). Ce concept a donné lieu à une importante controverse. Indiquons très sommairement en quoi les sémantiques préférentielles semblent ici intéressantes. Soient les trois énoncés :

(1a) Il est obligatoire que p ;

(2a) Si p, alors obligatoirement q ;

(3a) Si non-p, alors obligatoirement non-q.

Par exemple : il faut qu'un certain homme aille au secours de ses voisins ; s'il y va, il faut qu'il leur dise qu'il arrive ; s'il n'y va pas, il faut qu'il ne leur dise pas qu'il arrive. Représentons par $\bigcirc\phi$ l'obligation inconditionnelle que ϕ et par $\bigcirc(\psi/\phi)$ l'obligation que ψ dans la condition où ϕ. Nos trois énoncés deviennent :

[1] Nous les avons présenté au paragraphe 2.2.4, page 66. D'autres types d'analyses sont envisageables. Par simplicité, nous allons les passer sous silence. Pour une utilisation de la logique des défauts dans le déontique, cf. Horty [116]. Pour une utilisation de la théorie des systèmes argumentatifs, cf. Prakken [202] ainsi que Royakkers et Dignum [216].

(1b) $\bigcirc p$;

(2b) $\bigcirc(q/p)$;

(3b) $\bigcirc(\neg q/\neg p)$.

A partir de (1b) et (2b), on peut généralement déduire la conclusion selon laquelle :

(4b) $\bigcirc q$.

Imaginons que l'obligation conditionnelle soit définie en termes d'implication matérielle, soit qu'on considère l'expression $\bigcirc(\psi/\phi)$ comme une abréviation de $\phi \to \bigcirc\psi$, soit qu'on la suppose synonyme de $\bigcirc(\phi \to \psi)$. Dans les deux cas, la version déontique du principe de monotonie, i.e.

$$\bigcirc(\psi/\phi_1) \to \bigcirc(\psi/\phi_1 \wedge \phi_2), \qquad \text{(Monotonie déontique)}$$

est valide. De la sorte, à partir de (4b), on peut ensuite déduire la norme conditionnelle

(5b) $\bigcirc(q/\neg p)$

qui, intuitivement, peu difficilement coexister avec (3b). Or, nous nous souvenons comment Hansson analyse l'obligation conditionnelle[2]. Il suppose que, dans un modèle, une relation de préférence ordonne les mondes possibles en fonction de leur degré de perfection, puis il donne à $\bigcirc(-/-)$ des conditions de vérité isomorphes à celles de $|\!\sim$ (« entraîne normalement »). L'inférence de $\bigcirc(\psi/\phi_1 \wedge \phi_2)$ à partir de $\bigcirc(\psi/\phi_1)$ est immédiatement bloquée ; le paradoxe disparaît.

Il existe une autre raison de vouloir appliquer les structures préférentielles à l'analyse formelle du devoir-être. Alchourrón [2] s'en est par exemple fait l'avocat. La plupart des normes auxquelles nous sommes soumis et à propos desquelles nous argumentons sont valables seulement *prima facie* (à première vue). Exemples : tenir ses promesses, dire la vérité, rendre service à qui vous a obligé, distribuer justement, etc. Dans certaines circonstances, relativement exceptionnelles, l'une de ces obligations peut entrer en conflit avec et être « dominée » par une autre obligation, jugée plus forte. Dire d'une obligation que, dans des circonstances plus spécifiques, elle peut être dominée par une obligation plus forte revient précisément à dire qu'elle ne vérifie pas la version déontique du principe de monotonie, i.e.

$$\{\bigcirc(\psi/\phi_1), \neg\bigcirc(\psi/\phi_1 \wedge \phi_2)\} \text{ est logiquement possible.}$$

Pour cette raison, nous ne saurions suivre l'idée autrefois défendue par Hintikka [110], qui suggéra d'analyser l'obligation prima facie en termes d'implication matérielle.

[2] Cf. page 69 et ss.

Toutefois, un style d'analyse autre que celui de Hansson est a priori envisageable. Nous pouvons définir l'obligation conditionnelle à l'aide de l'obligation inconditionnelle (ou monadique) et d'un conditionnel ontique léwisien. Par exemple, Chellas [47] pose :

$$\bigcirc(\psi/\phi) \stackrel{\text{def}}{\longleftrightarrow} \phi \Rightarrow \bigcirc\psi,$$

où $\Rightarrow$ désigne un conditionnel de normalité qui lui-même n'est pas monotone.

Voici enfin une troisième raison qui peut justifier le recours aux modèles préférentiels. Nos argumentations quotidiennes n'invoquent pas seulement comme garanties des obligations réparatrices ou des normes prima facie. En règle générale, elles assortissent aussi nos catégories déontiques de multiples nuances. Par exemple, il nous arrive de qualifier un acte d'héroïque, au sens où son auteur nous a semblé faire plus que ce à quoi il était obligé. Ceci s'avère difficilement exprimable dans une logique déontique opposant brutalement les mondes déontiquement bien faits aux mondes déontiquement mal faits. Nous renvoyons ici à Mc Namara [154] et à Åqvist [18] pour une étude formelle du concept d'acte surérogatoire (*supererogatory*) dans le cadre d'une sémantique préférentielle.

3.1.2 Nos recherches antérieures

Un mot à présent sur nos recherches antérieures. Notre première contribution à la logique des normes fut l'élaboration d'une logique déontique prenant appui sur la logique propositionnelle trivalente de Bochvar[3]. Nous cherchions alors à libérer la logique des normes de certains des paradoxes sur lesquels on considère assez classiquement qu'ont achoppé la plupart des premiers systèmes, à savoir le *paradoxe de Ross*

$$\bigcirc\phi \to \bigcirc(\phi \vee \psi) \tag{3.1}$$

et les *paradoxes de l'obligation dérivée*

$$\bigcirc\phi \to \bigcirc(\psi \to \phi) \tag{3.2}$$

$$\bigcirc\phi \to \bigcirc(\neg\phi \to \psi). \tag{3.3}$$

La logique propositionnelle trivalente de Bochvar conserve toutes les inférences de la logique classique dans la conclusion desquelles ne figurent pas d'atomes qui ne figurent déjà dans les prémisses. Cela est encore vrai dans le déontique. Nous plaçant dans le cas particulier où un opérateur déontique ne figure jamais dans la portée d'un autre opérateur déontique, nous tentions de démontrer, plus précisément, un théorème de la forme :

[3] Cf. notre [188].

PROPOSITION 1
Supposons que l'inférence de $\phi_1, ...\phi_n$ à ψ soit valide en logique déontique bivalente classique. Pour que cette inférence soit également valide en logique déontique trivalente, il faut et il suffit que les deux conditions suivantes soient simultanément vérifiées :

- Si p a une occurrence liée dans ψ (au sens où p apparaît au moins une fois dans la portée de l'opérateur modal) alors p a aussi une occurrence liée dans une prémisse ϕ_i.
- Si p a une occurrence libre dans ψ (au sens où au moins une fois p n'apparaît pas dans la portée de l'opérateur modal) alors p a aussi une occurrence libre dans une prémisse ϕ_i.

Naturellement, il est tentant de chercher à dissiper le malaise que crée la présence des lois (3.1)-(3.3), en invoquant des considérations de nature pragmatique. Par exemple, il ne paraît pas absurde de soutenir que les deux premières au moins ne fournissent jamais qu'une déperdition d'information. Nous pourrions ici recourir à la théorie gazdarienne des échelles quantitatives[4]. S'il est vrai que $\bigcirc\phi$ entraîne $\bigcirc(\phi \vee \psi)$, alors nos deux énoncés forment une échelle quantitative, et toute assertion de la forme $\bigcirc(\phi \vee \psi)$ implicite scalairement que le locuteur ne croit pas que $\bigcirc\phi$. Une remarque similaire s'applique à (3.2). De toute évidence, l'appel à des considérations pragmatiques ne paraît pas possible dans le cas de (3.3), si nous voyons en $\bigcirc(\neg\phi \rightarrow \psi)$ l'expression d'une obligation réparatrice. Celle-ci doit évidemment rester logiquement indépendante. Dans le cas contraire, la contraction de $\{\bigcirc\phi, \bigcirc(\neg\phi \rightarrow \psi)\}$ par $\bigcirc(\neg\phi \rightarrow \psi)$ nous donnerait $\emptyset$. Nous souhaitons pouvoir modifier la seule nature d'une réparation juridique ou morale.

Le type d'approche que nous adoptions est tout à fait courant dans le domaine des logiques épistémiques. Toutefois, en soi, exiger d'une inférence qu'elle n'introduise pas d'information nouvelle paraît relativement draconien. Aussi nous avons décidé de ne pas poursuivre dans cette direction. Nous nous sommes alors tournés vers les sémantiques préférentielles *à la* Hansson et *à la* Lewis [5], car leur démarche nous semblait guidée par des préoccupations assez voisines aux nôtres : remédier à certaines insuffisances de la logique classique.

3.1.3 Aperçu général du chapitre

Ce chapitre comporte trois sections. Les deux premières – sections (3.2) et (3.3)– vont de pair, et doivent être plutôt considérées comme un point de départ.

[4]Cf. notre paragraphe 1.2.5, page 16.

[5]Cf. Lewis [140]. Quoique faisant appel à des systèmes de sphères, la sémantique de Lewis est très proche d'une sémantique préférentielle. Nous sommes dans le cas particulier d'un ordre total sur les mondes possibles.

La section (3.2) porte sur les sémantiques habituelles pour l'obligation conditionnelle. Nous y présentons les premières réactions que, au début de nos recherches, ces sémantiques éveillèrent en nous [6]. Nous avons indiqué tout à l'heure que deux styles d'analyse sont généralement adoptés. Suivant l'exemple de Hansson, certains posent l'opérateur d'obligation dyadique comme primitif, et se contentent de donner une interprétation déontique à la relation de préférence. Suivant l'exemple de Chellas, d'autres définissent l'obligation conditionnelle à l'aide de l'obligation inconditionnelle (ou monadique) et d'un conditionnel léwisien, i.e. $\phi \Rightarrow \bigcirc\psi$. Dans cette section, nous tentons de mieux apprécier en quoi ces deux styles d'analyse se distinguent. En particulier, nous tentons de mieux cerner leur impact respectif sur la logique de l'opérateur d'obligation conditionnelle. Depuis l'étude séminale de Gabbay [79], il est souvent admis que l'opérateur d'inférence non-monotone doit satisfaire deux principes fondamentaux : le principe de *réflexivité* (ou d'identité) et le principe de *cumulativité*. Celui-ci, rappelons-le, se décompose en deux sous-principes : la règle de *coupure*, $\phi \mid\sim \psi$ et $\phi \wedge \psi \mid\sim \tau$ entraînent $\phi \mid\sim \tau$; la règle de *monotonie prudente*, $\phi \mid\sim \psi$ et $\phi \mid\sim \tau$ entraînent $\phi \wedge \psi \mid\sim \tau$. Comme beaucoup l'ont remarqué, la version déontique du principe d'identité, $\bigcirc(\phi/\phi)$, est contre-intuitive. Par contre, nous n'avons a priori (au moins de prime abord) aucune raison de rejeter la version déontique de la cumulativité. Les deux analyses que nous venons d'évoquer manifestent ici une asymétrie. Si nous adoptons la première, le principe d'identité et le principe de cumulativité sont tous les deux valides. Si nous adoptons la seconde, aucun des deux ne l'est. Aussi, dans la section (3.3), nous étudions la question de savoir si ces deux positions sont les seules envisageables. Nous esquissons une analyse de l'obligation conditionnelle, fondée sur l'opérateur *et ensuite* de von Wright [277, 13]. L'une des particularités de cette sémantique est qu'elle élimine la loi d'identité, tout en maintenant une forme de cumulativité. Nous terminons cette section par une série de questions, suggérant que, en l'état, notre problématique de départ manque de solidité[7].

La particularité de la théorie sémantique que nous esquissons dans la section (3.3) est d'introduire du séquentiel dans les analyses habituellement données. Cette idée nous paraît suffisamment originale et intéressante, pour y revenir dans le cadre d'une étude sur l'interface entre la non-monotonie et l'argumentation. Dans la dernière section de ce chapitre − i.e. section (3.4) −, nous tentons de réorienter notre travail dans une autre direction, celle principalement de l'analyse des normes réparatrices. Dans cette section (3.4), nous introduisons et analysons la notion d'échange réparateur au sens de Goffman [93]. Nous montrons que, dans sa première approximation, notre compte rendu séquentiel n'est pas très satisfai-

[6] Nous nous appuierons sur des travaux dont nous avons eu l'occasion de faire état dans [189, 190].

[7] Nous sommes ici infiniment redevables à toute une série de remarques de David Makinson (communication personnelle).

sant. Aussi, nous nous tournons vers un autre style d'analyse. Cherchant à mettre à profit certaines contributions récentes dans le domaine de la révision itérée, nous esquissons un traitement « dynamique » de l'échange réparateur, fondé sur l'idée que la violation d'une obligation peut avoir un impact sur le pré-ordre associé aux prémisses de départ. Nous utilisons, plus précisément, la théorie de la révision naturelle que Boutilier [37] propose.

Les deux théories préférentielles que nous allons présenter mettent au premier plan la notion de séquentialité. Quoiqu'elles trouvent leur point de départ dans la logique des normes, ces théories sémantiques ont vraisemblablement une portée plus générale. En témoigne ce corps de doctrines classiquement désignées (depuis Levinson [138]) par le nom d'analyse conversationnelle (*conversation analysis*). Prenant pour objet d'étude les conversations dites « naturelles » (conversations téléphoniques, interactions parents-enfants, transactions commerciales, etc.), elles mettent au premier plan l'idée qu'une interaction se compose de séquences d'actes de langage, dont certaines sont préférées tandis que d'autres ne le sont pas. Ces séquences minimales dont nos argumentations seraient composées portent le nom de « paires adjacentes ». Le couplet question-réponse offre un exemple, peut-être l'exemple canonique, d'une paire adjacente. A supposer que la théorie logique ait quelque chose d'intéressant à dire concernant une paire adjacente, elle devra nécessairement recourir au concept de normalité. Le principe schegloffien de pertinence conditionnelle d'un deuxième tour rend ceci suffisamment manifeste, pour qu'il soit besoin de vraiment s'y attarder. Dans une paire adjacente, étant donné le premier membre de la paire, un second membre est immédiatement attendu. Malgré tout, il arrive qu'un premier membre d'une seconde paire apparaisse à sa place. Celui-ci est alors interprété comme le préliminaire nécessaire au second membre de la première paire. Exemple :

Q_1 Qui est cette fille ?
Q_2 Tu ne la connais pas ?
R_2 Non
R_1 C'est le nouveau professeur de français.

La séquence Q_1-R_1 forme une première paire adjacente. La séquence Q_2-R_2 en forme une autre. Une fois Q_1 exécuté, R_1 est normalement attendu. Q_2 apparaît à la place.

3.2 Une première incursion

Cette section porte sur les sémantiques de l'obligation conditionnelle déjà existantes. Elles sont nombreuses, et nous allons tenter de faire une première incursion dans le maquis que, vu de l'extérieur, elles forment. Nous prendrons ici pour fil directeur l'étude de David Makinson intitulée *Five Faces of Minimality*[8]. Formaliser, dit-on souvent, conduit à d'inévitables simplifications. Comme le souligne l'auteur, la sémantique de Hansson ne déroge pas à la règle. D'une part, elle rend valide l'axiome d'identité $\bigcirc(\phi/\phi)$, ce qui (comme nous allons le voir) n'est pas très satisfaisant. D'autre part, elle fait totalement abstraction de la dimension du futur, qui constitue pourtant la plus élémentaire des dimensions à laquelle nos argumentations fassent référence. L'une de nos préoccupations va être ici de tenter de clarifier l'éventuel lien entre ces deux réquisits (§ 3.2.3). Cela va nous servir de tremplin vers d'autres considérations (§ 3.2.4). Néanmoins, au préalable, il nous faut dire quelques mots sur la théorie sémantique que nous propose l'auteur de *Five Faces of Minimality* (§ 3.2.2). Nous devons aussi préciser quelque peu les raisons qui peuvent nous conduire à considérer comme gênante la présence de la loi $\bigcirc(\phi/\phi)$. Tel est l'objet du prochain paragraphe (§ 3.2.1). Nous y examinons brièvement le point de vue asymétrique au nôtre que Prakken et Sergot défendent. Ceci nous donnera l'occasion d'évoquer une série d'autres questions qui nous paraissent extrêmement intéressantes.

3.2.1 Fixité du contexte et déconditionnalisation

Reprenant une suggestion initialement faite par Hansson [103], Prakken et Sergot soutiennent que

(i) la validité de $\bigcirc(\phi/\phi)$ n'est pas paradoxale,

si nous supposons que

(ii) *Fixité du contexte* : l'antécédent d'une obligation conditionnelle décrit toujours quelque chose de « fixe » (*settled*) ou d'inévitable.

Pour éclairer le sens de ce principe – et la raison de cette hypothèse, nous croyons utile de nous reporter à la réponse qu'ils proposent de donner à la question : quelle forme de détachement (ou de déconditionnalisation) autorise la sémantique de Hansson ? Nos auteurs imaginent ici d'introduire un opérateur modal $\Box$ de type S5, lu « est inaltérablement vrai ». Puis ils opposent les deux schémas d'inférence :

$$\frac{\bigcirc(\psi/\phi) \qquad \phi}{\bigcirc\psi}\ (\mathrm{FD}) \qquad\qquad \frac{\bigcirc(\psi/\phi) \qquad \Box\phi}{\bigcirc\psi}\ (\mathrm{SFD})$$

L'énoncé $\bigcirc\psi$ est une abréviation de $\bigcirc(\psi/\top)$, où $\top$ désigne une tautologie quelconque. Il est aisé de vérifier que le premier schéma (généralement appelé « principe de détachement factuel ») n'est pas valide, alors que le second (Prakken et

[8]Cf. Makinson [158].

Sergot le nomment « version forte du principe de détachement ») est valide. Dans cette mesure, supposer (ii) revient à dire que

(iii) *Déconditionnalisation* : c'est seulement lorsque la vérité de l'antécédent ϕ est nécessaire que l'obligation inconditionnelle que ψ est détachable – et qu'elle peut conduire à quelque action.

Nous interprétons le fait que la vérité de l'antécédent ϕ soit nécessaire comme voulant dire que ϕ est une donnée « établie », sur la vérité de laquelle l'argumentateur ne peut plus revenir[9]. Nous allons dans un instant tenter d'expliquer le sens de ce principe. Voyons tout d'abord en quoi, pris au sens (iii), (ii) justifie (i). Substituons ϕ à ψ dans (SFD). En présence de l'axiome $\bigcirc(\phi/\phi)$, ce que nous obtenons se laisse immédiatement simplifier en :

$$\frac{\Box\phi}{\bigcirc\phi}$$

Mais nous n'avons pas :

$$\frac{\phi}{\bigcirc\phi}$$

Ainsi, l'axiome d'identité ne permet pas tant d'inférer un *doit* à partir d'un *est*, qu'à partir d'un *nécessaire*. Dans cette mesure, la présence de cette loi est inoffensive ; elle conduit tout au plus le logicien à qualifier d'obligatoires des actions auxquelles l'agent ne saurait se soustraire. Ainsi lisons-nous :

> « $\bigcirc(\phi/\phi)$ dit seulement que ϕ est vrai dans les meilleurs des ϕ-mondes, ce qui n'est pas plus (ni moins) inacceptable que la validité de $\bigcirc\top$ en logique déontique standard... L'affirmation selon laquelle ce qui est inaltérablement vrai est aussi obligatoire n'a rien de particulièrement choquant » [206, p. 245]

Prakken et Sergot citent le cas de l'état tautologiquement vérifié, $\top$, prototype de l'état qui s'impose à nous comme une nécessité. Sans doute, il y a quelque étrangeté à affirmer que ce qui est inaltérablement vrai est obligatoire. Mais cette bizarrerie est de nature pragmatique, et non pas sémantique. Reste à expliquer pourquoi il n'y a guère de sens à communiquer un ordre si celui à qui on s'adresse n'a aucune possibilité d'y déroger. Il nous suffit ici de recourir à la théorie de Gazdar[10]. S'il est vrai que $\Box\phi$ entraîne $\bigcirc\psi$, alors le couple $\langle\Box,\bigcirc\rangle$ forme une échelle quantitative, et toute assertion de la forme $\bigcirc\phi$ implicite scalairement que (le locuteur croit que) $\neg\Box\phi$, et donc que $\Diamond\neg\phi$ (si $\Diamond$ désigne le dual de $\Box$). Ainsi, en recourant à la notion d'implicature scalaire, nous pouvons donner un sens relativement précis à l'affirmation selon laquelle, quoique logiquement impeccable, l'inférence de $\bigcirc\phi$ à partir de $\Box\phi$ est pragmatiquement absurde.

[9] Cette interprétation épistémique de la notion de fi xité du contexte nous est personnelle.
[10] Cf. notre paragraphe 1.2.5, page 16.

Quoiqu'à première vue parfaitement défendable, cette position soulève au moins trois questions, qui l'affaiblissent quelque peu. Tout d'abord, nous pensions trouver dans la théorie de Gazdar de quoi donner un sens plus précis à la notion d'inférence pragmatique. Des obscurités demeurent. Considérons les modalités du permis et de l'interdit. La stratégie la plus immédiate consiste à introduire la première comme la « duale » de l'obligatoire, i.e.

$$P\phi \stackrel{\text{def}}{\longleftrightarrow} \neg \bigcirc \neg\phi,$$

et la seconde comme la négation du permis, i.e.

$$F\phi \stackrel{\text{def}}{\longleftrightarrow} \neg P\phi.$$

Ceci conduit à d'étranges conséquences. De même que nous avons le principe « $\Box\phi$ entraîne $\bigcirc\phi$ », nous avons le principe « $P\phi$ entraîne $\Diamond\phi$ ». Ainsi, s'il est vrai que $\langle\Box, \bigcirc\rangle$ constitue une échelle quantitative, alors $\langle P, \Diamond\rangle$ en constitue une aussi[11]. Dira-t-on que toute assertion de la forme $\Diamond\phi$ implique pragmatiquement que (le locuteur pense que) $\neg P\phi$, c'est-à-dire $F\phi$? Nous voyons que la théorie de Gazdar n'est pas aussi facilement transposable à l'ensemble des modalités déontiques.

A cette première question, s'en ajoute une seconde, qui concerne (SFD). Il faut ici rappeler que les analyses de Prakken et Sergot portent exclusivement sur l'obligation réparatrice. Il faut aussi préciser qu'ils refusent d'assimiler violation et exception. Supposons x dans une situation d'exception (il n'a pas payé ses impôts sur le revenu, parce qu'il n'était pas imposable cette année). Dans ce cas, a-t-on envie de dire, il ne viole pas l'obligation de payer ses impôts et la sanction (verser une majoration de 10%) ne s'applique pas. Naturellement, une obligation réparatrice peut elle-même prendre appui sur un principe prima facie. Nous aborderons brièvement ce point dans un instant. Reprenons tout d'abord (SDF). Appliqué à une obligation réparatrice de la forme $\bigcirc(\psi/\phi)$, ce principe stipule que :

> « c'est seulement si la violation de $\bigcirc\neg\phi$ est inévitable, si $\Box\phi$ est vrai, que l'obligation réparatrice entre pleinement en vigueur, et appartient au contexte $\top$ » [206, p. 241].

Si tel est le sens véritable du principe (SFD), force nous est de reconnaître qu'il y a quelque chose en lui de troublant. Les prémisses de (SFD) sont : $\bigcirc(\psi/\phi)$; $\Box\phi$. A partir de la seconde, nous en inférons déjà que $\bigcirc\phi$. Cela nous semble difficilement conciliable avec l'idée selon laquelle la première prémisse, $\bigcirc(\psi/\phi)$,

[11] Se donner la validité supplémentaire de $\bigcirc\phi \rightarrow P\phi$ reviendrait alors à faire du quadruplet $\langle\Box, \bigcirc, P, \Diamond\rangle$ une échelle quantitative.

exprime une obligation réparatrice. Il nous paraît intéressant de remarquer en passant que, souhaitant rendre compte du fait qu'une obligation réparatrice prend elle-même généralement appui sur un principe prima facie, un auteur comme Belzer [26] introduit une variante qui, semble-t-il, achopppe à une difficulté similaire :

$$\frac{\bigcirc(\psi/\phi) \qquad \Box\phi \qquad U(\psi/\phi)}{\bigcirc\psi}\ (\mathrm{SFD}')$$

La prémisse supplémentaire « $U(\psi/\phi)$ » signifie qu'il n'existe pas de défaiseurs à $\bigcirc(\psi/\phi)$, au sens où nous aurions un énoncé τ tel que $\Box\tau$ et $\neg\bigcirc(\psi/\phi\wedge\tau)$. Dans l'exemple que nous avons pris, τ pourrait désigner le fait d'être au chômage. En elle-même, cette variante semble plus satisfaisante : elle signifie que, pour que l'obligation inconditionnelle soit détachable, il faut aussi au préalable s'être assuré ne pas être dans une situation d'exception. Néanmoins, nous comprenons difficilement comment, dans (SFD′), la prémisse gauche $\bigcirc(\psi/\phi)$ peut exprimer une obligation réparatrice, si $\Box\phi$ entraîne $\bigcirc\phi$.

En résumé, l'argumentation était en deux temps. Pour rendre inoffensive la présence du principe d'identité, on suppose que l'antécédent d'une obligation conditionnelle exprime un trait fixe du contexte. Et, pour expliquer ce postulat, on invoque un principe de déconditionnalisation, qui revêt la forme soit de (SFD) soit de (SFD′). Dans les deux cas, le principe de déconditionnalisation invoqué se révèle ne pas être très clair. Voici la troisième et dernière question qui, selon nous, se pose. Le but final des explications de Prakken/Sergot est de montrer en quoi la présence de la loi $\bigcirc(\phi/\phi)$ ne signifie pas pour autant que nous ayons :

$$\frac{\phi}{\bigcirc\phi}$$

$\bigcirc\phi$ est ici pris comme une abréviation de $\bigcirc(\phi/\top)$, où $\top$ désigne une tautologie quelconque. En elle-même, cette définition de l'obligation inconditionnelle ne paraît pas très satisfaisante[12]. Car, intuitivement, nous souhaiterions que l'obligation catégorique que ϕ ne soit pas « isolée » de son contexte, à savoir la situation présente, qui ne vérifie pas les seules tautologies du calcul propositionnel. La seule variante que nous connaissons est celle d'Alchourrón [2]. Reprenant une idée de von Wright [278], le logicien argentin pose :

$$\bigcirc\phi \stackrel{\text{def}}{\longleftrightarrow} \bigcirc(\phi/\sigma), \tag{3.4}$$

où σ désigne une constante propositionnelle lue « dans les circonstances présentes ». Du point de vue sémantique, il nous suffit de supposer que l'univers

[12] Nous reprenons ici une observation faite plus haut, page 70 et ss.

de chaque modèle contient un élément $\star$ spécifique désignant le monde réel et tel que cet élément soit le seul à vérifier la constante σ. En elle-même, la définition (3.4) paraît plus réaliste. Or il se trouve que, à partir de l'axiome $\bigcirc(\phi/\phi)$, on légitime effectivement l'inférence d'un « doit » (*ought*) à partir d'un « est » (*is*), c'est-à-dire le schéma :

$$\frac{\phi}{\bigcirc\phi}$$

La dérivation fait intervenir deux lois supplémentaires : la loi que nous nommerons « axiome de Meredith-Prior » (du nom de ceux qui, semble-t-il, les premiers l'ont utilisé[13])

$$\phi \to \Box(\sigma \to \phi), \tag{3.5}$$

qu'on peut lire

Si ϕ *alors nécessairement les circonstances présentes impliquent* ϕ ;

et l'axiome

$$\Box(\psi \to \psi') \to (\bigcirc(\psi/\phi) \to \bigcirc(\psi'/\phi)), \tag{3.6}$$

qui nous autorise à affaiblir le conséquent d'une obligation conditionnelle. La vérification est immédiate :

$$\dfrac{\dfrac{\phi}{\Box(\sigma \to \phi)}\,(3.5) \qquad \bigcirc(\sigma/\sigma)}{\bigcirc(\phi/\sigma)}\,(3.6)$$

Dérivation d'un « doit » à partir d'un « est »

Ainsi, l'axiome $\bigcirc(\phi/\phi)$ a bien, semble-t-il, un impact indésirable sur nos obligations inconditionnelles. Cet axiome autorise l'inférence d'un « doit » catégorique à partir d'un « est » de nature *quelconque*. Cherchant à étendre ses prétentions au fondement de nos argumentations quotidiennes, la théorie logique ne se discrédite-t-elle pas aussitôt ?

3.2.2 La dimension du futur

Nous nous tournons ici vers la théorie sémantique que Makinson [158] esquisse. L'auteur suggère, à juste titre croyons-nous, qu'une théorie sémantique de

[13] Voir Meredith et Prior [170].

l'obligation conditionnelle doit à la fois invalider le principe d'identité et tenir compte de la dimension du futur. L'une de nos préoccupations va être de tenter de clarifier l'éventuel lien entre ces deux réquisits. David Makinson procède en deux temps. Dans un premier temps, il commence par s'intéresser au cas de l'obligation conditionnelle indéfaisable, puis il étend la construction au cas de l'obligation conditionnelle défaisable, et pose :

> « $\bigcirc(\psi/\phi)$ est vrai dans le monde w ssi : dans les plus normaux des mondes w' postérieurs à w et qui vérifient ϕ, ψ est vrai dans les plus parfaits des mondes w'' qui sont à leur tour accessibles relativement à w' » [158, p. 373].

ψ vrai dans les plus parfaits des mondes w'' à leur tour accessibles relativement à w', c'est-à-dire :

> « $\bigcirc\psi$ vrai dans le monde w' » [158, p. 372].

Qui est déjà familier avec la sémantique de Hansson remarque aussitôt que la partie droite de la règle d'évaluation de $\bigcirc(/)$ mentionne trois nouveaux ingrédients. Tout d'abord, elle fait usage d'une relation de normalité, que nous noterons $\prec$. Intuitivement $w \prec w'$ signifiera que w est plus plausible que w'. Aucune contrainte particulière n'est imposée à cette relation. Ensuite, la règle fait usage d'une relation temporelle, qui est ici supposée avoir une « logique située entre S4 et S4.3 » [158, p. 373]. Prise au sens de S4, la relation est réflexive et transitive. Prise au sens de S4.3, elle est en outre virtuellement connexe[14]. Enfin, la clause de récurrence fait usage d'une obligation inconditionnelle $\bigcirc$. Nous avons ici $\bigcirc\phi = \bigcirc_H(\phi/\top)$, où la lettre « H » est pour « Hansson ». Nous pourrions ici analyser l'obligation inconditionnelle à l'aide d'une relation d'admissibilité entre mondes possibles, comme dans une logique déontique standard. En fait, eu égard à notre problème de départ $-$ le statut de l'axiome d'identité, opter pour l'une ou l'autre de ces analyses ne fera aucune différence. Notons $\|\phi\|$ l'ensemble des ϕ-mondes, Fut_w l'ensemble des mondes qui sont ultérieurs à w et $\min(X,\prec)$ l'ensemble des $\prec$-minima de X. Notre règle d'évaluation peut être écrite :

DEFINITION 1
$w \models \bigcirc(\psi/\phi)$ ssi $\min(\|\phi\| \cap Fut_w, \prec) \subseteq \|\bigcirc\psi\|$.

Intuitivement, la procédure à suivre pour déterminer la valeur de vérité de $\bigcirc(\psi/\phi)$ dans le monde w est la suivante. Tout d'abord, nous prenons l'intersection de l'ensemble des mondes ultérieurs à w et de l'ensemble des mondes qui vérifient ϕ. Ensuite, passant en revue chacun des éléments qui sont les plus plausibles (les plus normaux,...), nous vérifions s'ils contiennent l'obligation inconditionnelle que ψ. Si la réponse est affirmative, alors $\bigcirc(\psi/\phi)$ est vrai dans le monde w. Si la réponse est négative, alors l'obligation est fausse.

[14] Si wRw' (lecture : w' postérieur à w) et si wRw'' alors : soit $w'Rw''$ soit $w''Rw'$.

3.2.2.1 Lien avec Hansson

La définition 1 opère un double déplacement. Tout d'abord, à la différence de celle de Hansson, elle rend le contenu de l'obligation conditionnelle explicitement orienté-vers-le-futur (*future-orientated*). On projette le contenu ψ de l'obligation, comme l'évènement-condition ϕ, dans le futur de la norme conditionnelle. A la différence de van der Torre [255, p. 25], nous ne croyons pas que la construction proposée exige une lecture de type « antécédent-puis-conséquent ». Car la construction autorise que chaque antécédent-monde soit contemporain avec ses alternatives les meilleures. Elle autorise cela, au moins à titre de possibilité formelle.

Un second déplacement a été opéré. La définition 1 analyse l'obligation conditionnelle en termes d'obligation inconditionnelle et à l'aide d'un «*normal future conditional* » [158, p. 373], que nous écrirons $\rightharpoondown$. Ce conditionnel léwisien est l'homologue « ontique » (ou factuel) de l'obligation conditionnelle hanssonienne. Pour déterminer la valeur de vérité de $\phi \rightharpoondown \psi$, nous devons porter notre attention non pas seulement sur l'ensemble des ϕ-mondes les plus normaux, mais sur le sous-ensemble de ceux qui sont dans le futur. $\bigcirc(\psi/\phi)$ devient équivalent à $\phi \rightharpoondown \bigcirc\psi$, comme l'auteur l'observe [158, p. 373]. Dans la sémantique initiale de Hansson, l'opérateur dyadique d'obligation est considéré comme « primitif » [103, p. 133] et comme n'étant pas définissable à l'aide d'autres opérateurs.

3.2.3 Principe d'identité et dimension du futur

Faisons quelques variations sur notre définition 1, et regardons ce qui se passe :

DEFINITION 2
$w \models \bigcirc(\psi/\phi)$ ssi $\min(\|\phi\| \cap Fut_w, \prec) \subseteq \|\psi\|$.

DEFINITION 3
$w \models \bigcirc(\psi/\phi)$ ssi $\min(\|\phi\|, \prec) \subseteq \| \bigcirc \psi\|$.

La définition 2 est obtenue à partir de la définition 1, en supprimant l'obligation monadique, de sorte que les concepts d'obligation et de conditionnalité fusionnent. La définition 3 est obtenue à partir de la définition 1, en supprimant la dimension du temps. Notons que, comme dans la théorie de Chellas [47, 48] et Mott [178], une telle définition rend $\bigcirc(\psi/\phi)$ équivalent à $\phi \Rightarrow \bigcirc\psi$, où $\Rightarrow$ désigne le conditionnel léwisien, tel qu'habituellement défini. Notons également que supprimer l'opérateur monadique d'obligation ainsi que la dimension du temps nous ramène à la clause de Hansson :

DEFINITION 4
$w \models \bigcirc(\psi/\phi)$ ssi $\min(\|\phi\|, \prec) \subseteq \|\psi\|$.

Nous récapitulons tout ceci sous forme d'un tableau :

	$\bigcirc(/)$ primitif	$\bigcirc(/)$ orienté vers le futur
Definition 4	oui	non
Definition 2	oui	oui
Definition 3	non	non
Definition 1	non	oui

Pour déterminer à quoi imputer l'élimination de $\bigcirc(\phi/\phi)$, il nous suffit de passer en revue chaque combinaison. Voici les résultats du test :

	$\bigcirc(/)$ primitif	$\bigcirc(/)$ orienté vers le futur	$\bigcirc(\phi/\phi)$
Definition 4	oui	non	valide
Definition 2	oui	oui	valide
Definition 3	non	non	non-valide
Definition 1	non	oui	non-valide

Les deux premières lignes nous disent que l'usage de l'opérateur monadique d'obligation est nécessaire pour éliminer le principe d'identité, que la dimension du futur soit ou non prise en compte. Les deux dernières lignes nous disent qu'il s'agit aussi d'une condition suffisante. La définition 4 (clause de Hansson) rend $\bigcirc(\phi/\phi)$ valide parce que

$$\min(\|\phi\|, \prec) \subseteq \|\phi\|.$$

Il en va de même pour la définition 2 :

$$\min(\|\phi\| \cap Fut_w, \prec) \subseteq \|\phi\| \cap Fut_w \subseteq \|\phi\|.$$

Pour les deux autres définitions, la situation est différente, parce que toutes deux expriment l'obligation et la conditionnalité au moyen de deux idiomes distincts. Considérons tout d'abord la définition 3. Nous ne pouvons pas déduire de la proposition

$$\min(\|\phi\|, \prec) \subseteq \|\phi\|,$$

la proposition

$$\min(\|\phi\|, \prec)| \subseteq \|\bigcirc\phi\|,$$

car nous n'avons pas nécessairement

$$\|\phi\| \subseteq \|\bigcirc\psi\|.$$

Cela s'accorde avec nos intuitions. Cette inclusion est fausse, lorsque le modèle ne rend pas valide $\phi \rightarrow \bigcirc\phi$, où $\rightarrow$ désigne l'implication matérielle. Il en va de même pour la définition 1. Nous ne saurions conclure que

$$\min(\|\phi\| \cap Fut_w, \prec) \subseteq \|\bigcirc\phi\|$$

en nous fondant sur la seule prémisse

$$\min(\|\phi\| \cap Fut_w, \prec) \subseteq \|\phi\|.$$

3.2.4 Identité et cumulativité : une asymétrie ?

Cette façon d'éliminer le principe d'identité a initialement été proposée par Chellas [47]. Dans ce qui suit, nous mettons en évidence l'une des conséquences de cette solution, et nous montrons que celle-ci persiste lorsque la dimension du futur est prise en compte. Cela nous conduira à une remarque plus générale concernant les deux sortes d'analyses généralement adoptées en logique déontique. Dans un cas, on pose l'opérateur d'obligation conditionnelle comme terme primitif. Dans l'autre, on définit cet opérateur à l'aide de l'obligation monadique et du conditionnel léwisien.

En règle générale, nous exigeons d'une sémantique qu'elle valide les deux principes suivants :

$$((\phi \Rightarrow \psi) \wedge ((\phi \wedge \psi) \Rightarrow \tau)) \rightarrow (\phi \Rightarrow \tau) \quad \text{(Coupure ou monotonie transitive)}$$
$$((\phi \Rightarrow \psi) \wedge (\phi \Rightarrow \tau)) \rightarrow ((\phi \wedge \psi) \Rightarrow \tau) \quad \text{(Monotonie prudente ou monotonie cumulative)}$$

Leurs homologues déontiques sont (respectivement) :

$$(\bigcirc(\psi/\phi) \wedge \bigcirc(\tau/\phi \wedge \psi)) \rightarrow \bigcirc(\tau/\phi) \quad \text{(Coupure déontique)}$$
$$(\bigcirc(\psi/\phi) \wedge \bigcirc(\tau/\phi)) \rightarrow \bigcirc(\tau/\phi \wedge \psi) \quad \text{(Monotonie prudente déontique)}$$

Intuitivement, la monotonie prudente déontique précise que l'exécution d'une obligation n'a pas d'effet sur nos autres obligations. A notre connaissance, ce principe n'a jamais été discuté dans la littérature. La règle de coupure (plus connue sous le nom de « principe de détachement déontique ») nous dit à quelles conditions mettre nos obligations en chaîne. Ce schéma d'inférence a été discuté dans la littérature, dans le contexte de l'analyse de l'obligation réparatrice, et en liaison avec d'autres principes de détachement, tels que le principe de détachement factuel tout à l'heure évoqué. Prises en elles-mêmes, ces deux règles paraissent intuitives. A supposer que cette impression soit fondée, alors un sentiment d'embarras nous envahit aussitôt. En effet, il est aisé de vérifier que, si l'obligation dyadique est considérée comme primitive, alors nos deux principes sont valides. Pour que la coupure soit valide, aucune contrainte ne doit être imposée à la relation de préférence $\prec$ définie sur les mondes possibles. Par contre, comme les études [127] et [157] l'ont montré, pour que le principe de monotonie prudente soit vérifié, il faut exiger de la relation de préférence qu'elle vérifie la propriété de smoothness (pas de chaîne infinie de mondes de plus en plus normaux)[15]. Néanmoins, rien ne nous empêche de l'exiger aussi de la relation de bonté comparative (i.e. de la relation $\prec$ lue déontiquement). Supposons à présent que l'obligation dyadique soit définie à l'aide de l'obligation monadique et du conditionnel léwisien. Il est aisé de voir que les règles de coupure et de monotonie prudente sont toutes les deux rejetées, que la dimension du futur soit ou non prise en compte. Pour le vérifier, nous supposerons que l'obligation monadique est analysée comme dans une logique déontique standard. Pour autant que nous puissions en juger, nous obtiendrions le même résultat, si elle était définie en termes de bonté comparative.

PROPOSITION 2
Si $\bigcirc(/)$ est pris au sens de la définition 3 ou au sens de la définition 1, alors les deux schémas suivants sont falsifiables :

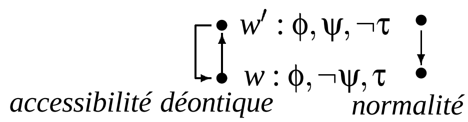
$$(\bigcirc(\psi/\phi)\wedge\bigcirc(\tau/\phi\wedge\psi)) \rightarrow \bigcirc(\tau/\phi) \qquad \text{(Coupure déontique)}$$
$$(\bigcirc(\psi/\phi)\wedge\bigcirc(\tau/\phi)) \rightarrow \bigcirc(\tau/\phi\wedge\psi) \qquad \text{(Monotonie prudente déontique)}$$

DÉMONSTRATION -

Nous commençons par la règle de coupure. Considérons un modèle dans lequel :

- $W = \{w, w'\}$ (nos mondes possibles) ;
- w' est admissible relativement à w et réciproquement (accessibilité déontique) ;
- $\prec = \{(w, w')\}$ (la relation de préférence, lue comme une relation de normalité) ;
- $\|\phi\| = \{w, w'\}$, $\|\psi\| = \{w'\}$ and $\|\tau\| = \{w\}$.

Soit, sous forme de diagramme :



Sur la gauche du schéma, une flèche du type $x \leftarrow y$ veut dire que x est admissible relativement à y. Sur la droite du schéma, elle signifie que x est plus normal que y.

Nous voyons que $\min(\|\phi\|, \prec) = \{w\}$ et que $\|\bigcirc\psi\| = \{w\}$. Ainsi, $w \models \bigcirc(\psi/\phi)$ (par def 3). Nous voyons aussi que $\|\bigcirc\tau\| = \{w'\}$ et que $\min(\|\phi\wedge\psi\|, \prec) = \{w'\}$. De la sorte $w \models \bigcirc(\tau/\phi\wedge\psi)$. Par contre, $w \not\models \bigcirc(\tau/\phi)$, car $w \not\models \bigcirc\tau$.

Pour montrer que, une fois la dimension du futur prise en compte (c'est-à-dire si nous utilisons def 1), la coupure est toujours invalidée, il suffit de reprendre

[15] Ce point fut évoqué lors de notre présentation des sémantiques préférentielles dans le domaine ontique. Formellement, la propriété de smoothness dit que

$$\forall w \in (\|\phi\| - \min(\|\phi\|, \prec)) \quad \exists w' \in \min(\|\phi\|, \prec) \text{ tel que } w' \prec w.$$

Voir ci-dessus page 67.

ce contre-exemple, et de poser $Fut_w = \{w, w'\}$ et $Fut_{w'} = \{w'\}$. Dans ce cas de figure, $\min(\|\phi\| \cap Fut_w, \prec) = \min(\|\phi\|, \prec)$. De même, $\min(\|\phi \wedge \psi\| \cap Fut_{w'}, \prec) = \min(\|\phi \wedge \psi\|, \prec)$. Ceci nous ramène au raisonnement précédent.

Nous pouvons vérifier que la monotonie prudente est invalide, en changeant simplement la valeur de vérité de τ dans les mondes w et w'. □

Il est ici difficile de donner une interprétation aux lettres de proposition qui rende intuitivement claire la raison pour laquelle la monotonie prudente et la coupure sont ici falsifiées. En outre, ces quelques considérations n'épuisent pas la question de savoir ce qui distingue nos deux approches. Qu'il nous suffise de faire remarquer qu'elles donnent un traitement différent à l'axiome d'identité et au principe de cumulativité. Si l'obligation conditionnelle est supposée primitive, les deux principes sont simultanément valides. Si elle n'est pas supposée primitive, aucun des deux ne l'est. Cette observation soulève d'intéressantes questions, telles que :

– D'un simple point de vue conceptuel, qu'est-ce qui nous empêche de développer une sémantique validant la cumulativité, mais pas le principe d'identité ?
– Que se passerait-il si l'obligation dyadique était définie à l'aide d'un conditionnel ontique non-classique autre queléwisien ?

Nous donnerons plus tard des éléments de réponse à la première question[16].

3.3 Une troisième voie

Dans la section précédente, nous avons tenté de clarifier la question du lien éventuel entre la dimension du futur et le principe d'identité. Nous avons vu que, pour éliminer le second, il n'est pas nécessaire d'introduire la première. Cela nous a amené à faire une observation plus générale concernant les deux principaux types d'analyses de l'obligation conditionnelle que nous rencontrons dans la littérature. Les unes prennent l'opérateur dyadique d'obligation comme premier terme non-défini. Les autres définissent celui-ci à l'aide de l'opérateur monadique d'obligation et d'un conditionnel ontique non-monotone. Dans le premier cas, le principe d'identité et le principe de cumulativité sont tous les deux valides. Dans le second cas, aucun des deux ne l'est.

Dans cette section, nous essayons de montrer qu'une troisième position est néanmoins envisageable, et qu'il est possible d'éliminer le principe d'identité tout en maintenant une forme de cumulativité. Nous allons faire ici un pas de plus dans notre investigation du lien entre obligation et temps. Jusqu'ici, nous avons exploré la seule possibilité de prendre en compte la dimension du futur. Que se passerait-il

[16] Voir ci-dessous p. 154.

si nous exigions aussi de l'évènement-condition qu'il soit (temporellement) antérieur à l'action obligatoire ?

Notre fil directeur est le suivant. Abstraction faite de la question du temps, nous pouvons définir l'obligation conditionnelle $\bigcirc(\psi/\phi)$ par $\phi \wedge \psi \gg \phi \wedge \neg\psi$, où $\gg$ désigne un opérateur de préférence (ou prohairétique) convenablement défini[17]. Pour introduire un ordre de succession entre l'antécédent et le conséquent d'une obligation conditionnelle, il suffit de définir celle-ci par $\phi;\psi \gg \phi;\neg\psi$. Ici, « ; » désigne l'opérateur *et ensuite* de von Wright [277, 13] – opérateur qu'il nous arrivera de lire *et au tour suivant*. Les notions de changement et de préférence fonctionnent ici comme notions primitives. Il conviendra évidemment d'en préciser l'analyse.

Voici le plan de cette section. Tout d'abord, nous envisageons tour à tour les notions de changement et de préférence (§ 3.3.1). Puis nous introduisons notre conditionnel de type séquentiel (§ 3.3.2), et montrons comment reformuler son compte rendu en termes de minimalité (§ 3.3.3). Ensuite, nous étudions la logique de l'opérateur (§ 3.3.4). Enfin, nous présentons et tentons de discuter trois objections qui peuvent être élevées contre notre analyse (§ 3.3.5). La principale consiste à demander en quoi la cumulativité est souhaitable.

3.3.1 Notions primitives

Soit $M = (Tree, <)$ une structure temporelle arborescente, où *Tree* est un ensemble non-vide d'instants auquel $<$ donne la forme d'une structure ramifiée vers le futur. Admettons que les règles d'évaluation soient de la forme :

$M, m/h \models \phi$: « dans le modèle M, la formule ϕ est vraie au moment m de l'histoire h ».

Supposant que le temps forme une suite discrète[18] et qu'il est infini, nous posons :

DEFINITION 5 (*Clause primitive - et au tour suivant*)
$M, m/h \models \phi;\psi$ ssi $M, m/h \models \phi$ et $M, m+1/h \models \psi$.

[17] Il est aussi possible de faire l'inverse, c'est-à-dire de définir l'opérateur d'obligation conditionnelle à l'aide de l'opérateur prohairétique $\gg$. Mais ceci cadrerait difficilement avec le type d'analyse que nous cherchons.

[18] L'usage des entiers naturels comme moments du temps restreint l'applicabilité de notre sémantique au raisonnement de sens commun. Nous laissons ce point de côté.

L'axiomatique que von Wright retenait est ici conservée dans son intégralité. Outre les règles et les axiomes du calcul propositionnel, nous avons :

$$\vdash [(\phi_1 \vee \phi_2);(\phi_3 \vee \phi_4)] \leftrightarrow [(\phi_1;\phi_3) \vee (\phi_1;\phi_4) \vee (\phi_2;\phi_3) \vee (\phi_2;\phi_4)] \tag{T0}$$

$$\vdash ((\phi_1;\phi_2) \wedge (\phi_1;\phi_3)) \leftrightarrow (\phi_1;(\phi_2 \wedge \phi_3)) \tag{T1}$$

$$\vdash \phi \leftrightarrow (\phi;(\psi \vee \neg\psi)) \tag{T2}$$

$$\vdash \neg(\phi;(\psi \wedge \neg\psi)) \tag{T3}$$

Le premier axiome met en relief la propriété de distributivité du connecteur *et ensuite* sur le connecteur *ou*. Le deuxième axiome (que von Wright nomme « axiome de coordination ») peut prêter à confusion. Dans le système du logicien finlandais, il nous garantit la linéarité du futur. Sa présence s'explique ici par le fait que nous avons choisi de travailler avec des structures dites « ockhamistes » : les conditions de vérité des énoncés sont relatives à des paires de la forme <moment, histoire>. Nous rencontrerons plus tard une raison d'abandonner cet axiome. Le troisième axiome (que von Wright baptise « loi de redondance ») nous paraît accessoire. C'est lui qui suppose l'infinité du temps en direction de l'avenir, hypothèse que nous pourrions éventuellement abandonner. Quant au dernier axiome (von Wright le nomme « axiome d'impossibilité »), il nous garantit que, à l'instant suivant, les contradictions ne commenceront pas tout à coup à être vraies. Voici un échantillon des théorèmes que, sur la base de cette axiomatique, nous pouvons démontrer :

$$\vdash (\phi_1;\phi_2) \vee (\phi_1;\neg\phi_2) \vee (\neg\phi_1;\phi_2) \vee (\neg\phi_1;\neg\phi_2) \tag{T4}$$

$$\vdash \neg((\phi \wedge \neg\phi);\psi) \tag{T5}$$

$$\vdash (\phi;\psi) \rightarrow \phi \tag{T6}$$

$$\vdash (\phi_1 \wedge (\phi_2;\phi_3)) \leftrightarrow ((\phi_1 \wedge \phi_2);\phi_3) \tag{T7}$$

$$\vdash ((\phi_1;\phi_2);\phi_3) \leftrightarrow (\phi_1;(\phi_2 \wedge \phi_3)) \tag{T8}$$

$$\text{Si } \vdash \psi \rightarrow \psi\prime \text{ alors } \vdash (\phi;\psi) \rightarrow (\phi;\psi\prime) \tag{T9}$$

$$\vdash ((\phi_1;\phi_2);\phi_3) \leftrightarrow ((\phi_1;\phi_2);(\phi_2 \wedge \phi_3)) \tag{T10}$$

$$\vdash ((\phi_1;\phi_2);\neg(\phi_2 \wedge \phi_3)) \leftrightarrow ((\phi_1;\phi_2);\neg\phi_3) \tag{T11}$$

$$\vdash (\phi_1;\phi_2) \rightarrow [(\phi_1;(\phi_2 \wedge \phi_3)) \vee (\phi_1;(\phi_2 \wedge \neg\phi_3))] \tag{T12}$$

Le lecteur trouvera dans von Wright [277] les démonstrations de (T4)-(T8). Nous donnons les autres :

• Pour (T9). Supposons que $\vdash \psi \rightarrow \psi\prime$. Dans ce cas, $\vdash \top \leftrightarrow (\neg\psi \vee \psi\prime)$ [calcul propositionnel]. De la sorte $\vdash ((\phi;\psi);\top) \leftrightarrow ((\phi;\psi);(\neg\psi \vee \psi\prime))$ [remplacement des équivalents]. Le raisonnement est ensuite le suivant :

$$\begin{array}{rll} \vdash (\phi;\psi) & \leftrightarrow (\phi;\psi);\top & \text{Par (T2)} \\ & \leftrightarrow (\phi;\psi);(\neg\psi \vee \psi\prime) & \\ & \leftrightarrow ((\phi;\psi);\neg\psi) \vee ((\phi;\psi);\psi\prime) & \text{Par (T0)} \\ & \leftrightarrow (\phi;(\psi \wedge \neg\psi)) \vee (\phi;(\psi \wedge \psi\prime)) & \text{Par (T8)} \\ & \leftrightarrow \phi;(\psi \wedge \psi\prime) & \text{Par (T3)} \\ & \rightarrow (\phi;\psi\prime) & \text{Par (T1)} \end{array}$$

• Pour (T10). En substituant ϕ_2 à ϕ_3 dans (T8), nous voyons immédiatement que $(\phi_1;\phi_2);\phi_2)$ est équivalent à $(\phi_1;(\phi_2 \wedge \phi_2))$, c'est-à-dire à $\phi_1;\phi_2$. Il vient donc :

$$\begin{array}{rll} ((\phi_1;\phi_2);\phi_2);\phi_3 & \leftrightarrow (\phi_1;\phi_2);(\phi_2 \wedge \phi_3) & \text{Par (T8)} \\ ((\phi_1;\phi_2);\phi_3) & \leftrightarrow (\phi_1;\phi_2);(\phi_2 \wedge \phi_3) & \text{Rempl. des équiv.} \end{array}$$

• Pour (T11).

$$\begin{array}{rll} (\phi_1;\phi_2);\neg(\phi_2 \wedge \phi_3) & \leftrightarrow (\phi_1;\phi_2);(\phi_2 \wedge \neg(\phi_2 \wedge \phi_3)) & \text{Par (T10)} \\ & \leftrightarrow (\phi_1;\phi_2);(\phi_2 \wedge \neg\phi_3). & \text{Rempl. des équiv.} \\ & \leftrightarrow (\phi_1;\phi_2);\neg\phi_3. & \text{Par (T10)} \end{array}$$

• Pour (T12).

$$\begin{array}{rll} (\phi_1;\phi_2) & \rightarrow (\phi_1;\phi_2);(\phi_3 \vee \neg\phi_3) & \text{Par (T2)} \\ & \rightarrow ((\phi_1;\phi_2);\phi_3) \vee ((\phi_1;\phi_2);\neg\phi_3) & \text{Par (T0)} \\ & \rightarrow (\phi_1;(\phi_2 \wedge \phi_3)) \vee (\phi_1;(\phi_2 \wedge \neg\phi_3)) & \text{Par (T8)} \end{array}$$

Sur cette base, nous pouvons maintenant nous tourner vers la composante prohairétique du modèle. Notons $\mathbf{H}$ l'ensemble des histoires. Associons à l'opérateur $\gg$ une relation de préférence $\preceq \subseteq \mathbf{H} \times \mathbf{H}$. Prenons garde que, si l'opérateur $\gg$ s'applique à des formules, la relation $\preceq$ compare des histoires – ingrédients du modèle. Ainsi, l'élément $\preceq$ est le « corrélat » sémantique du symbole $\gg$. Nous suivrons ici la tradition (issue de Lewis) qui consiste à interpréter $h \preceq h'$ (plutôt que $h \succeq h'$) comme signifiant que h est au moins aussi préférable que h'. Nous supposerons aussi que $\preceq$ est un pré-ordre partiel (i.e. une relation réflexive et transitive)[19], et introduirons $h \prec h'$ comme une abréviation de $h \preceq h' \wedge h' \not\preceq h$. Nous

[19]Ici, nous n'exigeons pas de la relation de préférence qu'elle vérifi e la propriété de smoothness (également nommée "stopperedness"). Généralement admise dans les sémantiques préférentielles, cette hypothèse écarte la possibilité d'une chaîne infi nie de mondes (ici, des histoires) de plus en plus parfaits. Nous envisagerons cette hypothèse seulement plus tard, au paragraphe 3.3.3, page 144 et ss.

laissons au lecteur le soin de vérifier que, si $\preceq$ est réflexive, alors $\prec$ est irréflexive, et que si $\preceq$ est transitive, alors $\prec$ l'est aussi. Soit H_m l'ensemble des histoires qui passent à travers un point m. Nous suggérons d'évaluer l'opérateur prohairétique à la façon de Halpern [100] :

DEFINITION 6 (*Clause primitive - préférence)*
$m/h \models \phi \gg \psi$ ssi :

$$\begin{aligned}&(\exists h' \in H_m)(m/h' \models \phi) \ \& \\ &\qquad (\forall h \in H_m)\Big(m/h \models \psi \rightarrow \\ &\qquad\qquad (\exists h' \in H_m)\big(m/h' \models \phi \ \& \ h' \prec h \ \& \\ &\qquad\qquad\qquad \neg(\exists h'' \in H_m)(m/h'' \models \psi \ \& \ h'' \prec h')\big)\Big)\end{aligned}$$

Intuitivement : $\phi \gg \psi$ est vrai en m/h si et seulement si les deux conditions suivantes sont satisfaites :

1. Quelque histoire vérifie ϕ en m ;
2. Chaque histoire vérifiant ψ en m est 'minimisée' par quelque histoire vérifiant ϕ en m, laquelle histoire a cette propriété remarquable de n'être elle-même minimisée par aucune histoire vérifiant ψ en m.

Si, comme Halpern, nous tenons $\Box\phi$ pour une abréviation de $\neg(\neg\phi \gg \bot)$, nous obtenons pour $\Box$ une clause qui apparente cet opérateur à ce qu'on a coutume d'appeler la « nécessité historique » :

DEFINITION 7 (*Nécessité historique)*
$m/h \models \Box\phi$ ssi : $(\forall h' \in H_m)(m/h' \models \phi)$.
Intuitivement : $\Box\phi$ est vrai en m/h si et seulement si chaque histoire h' qui traverse m vérifie ϕ en m.

On vérifiera sans peine que l'opérateur prohairétique $\gg$ se conforme aux quatre lois suivantes, dont l'ensemble forme le système AX de Halpern [100] :

$$\vdash \neg(\phi \gg \phi) \tag{P0}$$
$$\vdash ((\phi_1 \vee \phi_2) \gg \phi_3) \wedge ((\phi_1 \vee \phi_3) \gg \phi_2) \rightarrow (\phi_1 \gg (\phi_2 \vee \phi_3)) \tag{P1}$$
$$\vdash (\Box(\phi \rightarrow \phi') \wedge \Box(\psi' \rightarrow \psi) \wedge (\phi \gg \psi)) \rightarrow \phi' \gg \psi' \tag{P2}$$
$$\vdash \Box\phi \text{ pour toute loi logique} \tag{P3}$$

La loi (P0) signifie que $\gg$ vérifie la propriété d'irréflexivité. (P3) exprime un principe de nécessitation. La loi (P2) (qu'Halpern baptise « axiome d'ordre ») précise que, dans une expression de la forme $\phi \gg \psi$, nous avons le droit d'affaiblir la formule à gauche, et celui de renforcer la formule à droite. Enfin, (P1) (qu'Halpern nomme « principe de qualitativité ») peut être compris ainsi. Dans $(\phi_1 \vee \phi_2) \gg \phi_3$, nous pouvons permuter $\vee$ et $\gg$, à condition d'intervertir l'ordre de subordination entre les deux foncteurs, et si ϕ_2 et ϕ_3 sont eux-mêmes interchangeables.

Bien avant Halpern, Lewis [140] et ses disciples avaient imaginé d'établir la logique prohairétique sur une base apparemment similaire. Une première différence entre le système de Lewis et le présent système d'Halpern est que la syntaxe du premier utilise comme terme premier, non pas le foncteur de préférence stricte que nous avons noté « $\gg$ », mais un foncteur de préférence non-stricte que nous noterons :

$$\phi \geq \psi$$

expression que Lewis lisait :

> ϕ *est au moins aussi bon (possible) que* ψ.

Cette première différence paraît elle-même sans gravité : il suffirait en effet de munir le système d'Halpern de ce nouveau connecteur, puis de reformuler l'axiomatique en conséquence, pour que cette différence s'estompe. Cet exercice n'est pas trivial ; il ne paraît pas irréalisable. Une seconde différence est cette fois plus importante : tandis que le système de Lewis exclut a priori que deux états de choses puissent ne pas être comparables, celui d'Halpern ne l'exclut pas. Ceci explique pourquoi nous nous sommes immédiatement tournés vers le second. Mais une difficulté surgit aussitôt. D'une part, pour démontrer la complétude de son système, Halpern utilise une méthode qui ne paraît pas si aisément transposable au cas temporel[20]. D'autre part, l'auteur [100, p. 9] conjecture que la méthode du modèle canonique (qui est couramment utilisée dans les logiques modales en général, et dans les logiques temporelles en particulier) ne nous est ici d'aucun secours. Persévérant malgré tout dans l'usage que nous avons initialement choisi, devons-nous donc renoncer à tout espoir de pouvoir démontrer un théorème de complétude pour une logique articulant changement et préférence ? L'observation suivante laisse présager que non. Imaginons que nous travaillions avec le foncteur de préférence non-stricte $\geq$, et que nous analysions celui-ci à la façon de Lewis :

$$w \models \phi \geq \psi \text{ ssi } (\forall w_1 \models \psi)(\exists w_2 \models \phi)(w_2 \preceq_w w_1) \tag{3.7}$$

[20]Cette méthode ne porte pas (à notre connaissance) de nom particulier. Halpern raisonne ainsi. Il faut ici montrer la validité de l'implication suivante : si ϕ est consistant (au sens où sa négation n'est pas démontrable) alors ϕ est réalisable dans (un monde d') une structure. Se plaçant dans le cas particulier où un opérateur prohairétique ne fi gure jamais dans la portée d'un autre opérateur prohairétique, l'auteur part de l'idée que ϕ est logiquement équivalent à la disjonction de ses atomes. Par atome de ϕ, il entend une conjonction de la forme $C_1 \wedge \ldots \wedge C_m$, où chaque $C_i (1 \leq i \leq m)$ est soit une formule prohairétique élémentaire (sous-formule de ϕ) soit sa négation. Les lois du calcul propositionnel nous assureraient que si ϕ est consistant, alors au moins un atome de ϕ est consistant. Nommons σ cet atome. Le nerf de l'argument consiste à construire une structure dans laquelle σ (et, par là même, ϕ) est réalisable.

Intuitivement : ϕ est au moins aussi préféré que ψ (dans le monde w) si et seulement si pour tout ψ-monde w_1 il existe un ϕ-monde w_2 qui est au moins aussi préféré que w_1. Imaginons que $\preceq_w$ soit seulement réflexive et transitive [21]. Il nous paraît possible de montrer, en utilisant la méthode du modèle canonique, que cette sémantique est adéquate pour le système (nommons-le L) comportant les deux seuls axiomes suivants, déjà envisagés par Lewis[22] :

$$\vdash (\phi \geq \psi \wedge \psi \geq \xi) \rightarrow \phi \geq \xi \tag{A1}$$

$$\text{Si } \vdash \phi \rightarrow (\psi_1 \vee \ldots \vee \psi_n) \text{ alors } \vdash (\psi_1 \geq \phi) \vee \ldots \vee (\psi_n \geq \phi) \tag{R1}$$

(A1) signifie que $\geq$ est transitive. (R1) implique que $\geq$ est réflexive. Ce système est probablement plus faible que celui d'Halpern. Vraisemblablement aussi, en ajoutant de nouveaux axiomes au premier, nous devrions pouvoir obtenir un système assez voisin de AX, tel que sa relative simplicité le rende plus transparent à une interprétation. Nous ignorons si un tel théorème de complétude existe déjà dans la littérature. Aussi, nous en donnons la démonstration, afin que le lecteur puisse lui-même vérifier si notre conjecture est exacte[23]. Définissons un modèle canonique pour L comme étant un triplet de la forme $(W, \{\preceq_w : w \in W\}, m)$, où :

1. W est l'ensemble des ensembles L-maximaux consistants ;
2. Soit $/\phi/w = \{w' \in W : \psi \in w' \rightarrow \psi \geq \phi \in w\}$. Alors :
   $w_1 \preceq_w w_2$ ssi : $\forall\psi$ tel que $w_2 \in /\psi/w$, $\exists\phi$ tel que $w_1 \in /\phi/w$ et $\phi \geq \psi \in w$;
3. m est la valuation définie par : $w \in m(p) \leftrightarrow p \in w$.

Nous empruntons à Goble [90] cette façon de définir $\preceq_w$ dans le modèle canonique.

LEMME 1
$\preceq_w$ est réflexive et transitive.

DÉMONSTRATION -
$\preceq_w$ est réflexive. Ceci est immédiat par (R1), en présence de laquelle $\vdash \phi \geq \phi$.

[21] Comme précédemment, nous n'exigeons pas de la relation de préférence qu'elle vérifie la propriété de smoothness. Cf. supra, note 19, page 136.

[22] Tous deux apparaissent dans l'axiomatique du plus faible des systèmes que Lewis ait étudiés, à savoir le système V. On obtient ce dernier, par simple adjonction de l'axiome

$$\vdash (\phi \geq \psi) \vee \psi \geq \phi), \tag{A2}$$

dont l'admission nous obligerait à nous écarter du cours de la rationalité ordinaire. Celle-ci admet fort bien que deux termes puissent ne pas être comparables.

[23] Nous ne donnerons qu'une esquisse, en nous arrêtant aux points qui nous semblent essentiels. Le lecteur intéressé trouvera dans Hughes et Cresswell [119] ainsi que dans Gochet et al. [92] une étude d'ensemble de la méthode.

$\preceq_w$ est transitive. Prenons ϕ_3 tel que $w_3 \in /\phi_3/w$. De $w_2 \preceq_w w_3$, il vient $\exists\phi_2$ tel que $w_2 \in /\phi_2/w$ et $\phi_2 \geq \phi_3 \in w$. De $w_1 \preceq_w w_2$, il vient $\exists\phi_1$ tel que $w_1 \in /\phi_1/w$ et $\phi_1 \geq \phi_2 \in w$. Par (A1), $\phi_1 \geq \phi_3 \in w$, c'est-à-dire que $w_1 \preceq_w w_3$.

Que $\preceq_w$ puisse ne pas totalement ordonner W vient de ce que $\nvdash$ A2. □

LEMME 2
Soit w un ensemble L-maximal consistant. Soit $X = \{\neg\phi : \neg(\phi \geq \psi) \in w\}$. L'ensemble $X\prime = X \cup \{\psi\}$ est L-consistant.

DÉMONSTRATION -
De $X\prime \vdash \bot$, nous tirons $\exists\phi_1,\dots,\phi_n \in X$ tels que $\phi_1,\dots,\phi_n \vdash \neg\psi$, c'est-à-dire tels que $\vdash \psi \rightarrow (\neg\phi_1 \vee \ldots \vee \neg\phi_n)$. Par (R1), il vient $(\neg\phi_1 \geq \psi) \vee \ldots \vee (\neg\phi_n \geq \phi) \in w$. Comme w est maximal, nous en inférons que $\exists i$ tel que $(\neg\phi_i \geq \psi) \in w$. De $\phi_i \in X$, nous tirons $\neg(\neg\phi_i \geq \psi) \in w$, c'est-à-dire $(\neg\phi_i \geq \psi) \notin w$. Contradiction. □

LEMME 3 (*Propriété du modèle canonique*)
$w \models \phi$ si et seulement si $\phi \in w$.

DÉMONSTRATION -
Par induction sur la profondeur de ϕ. Nous le vérifions pour le seul cas où ϕ est de la forme $\phi_1 \geq \phi_2$:

– *De droite à gauche*. Supposons que $\phi_1 \geq \phi_2 \in w$. Prenons w_2 tel que $w_2 \models \phi_2$. Nous devons montrer que $\exists w_1$ tel que $w_1 \models \phi_1$ et $w_1 \preceq_w w_2$. Soit $X = \{\neg\phi : \neg(\phi \geq \phi_1) \in w\}$. L'ensemble $X\prime = X \cup \{\phi_1\}$ est L-consistant (lemme 2). Mais alors, il existe un ensemble L-maximal consistant contenant $X\prime$ (lemme de Lindenbaum). Nommons-le w_1. De $\phi_1 \in w_1$, nous tirons $w_1 \models \phi_1$ (hypothèse inductive). Nous souhaitons montrer que $w_1 \preceq_w w_2$. Prenons ψ tel que $w_2 \in /\psi/w$. De $w_2 \in /\psi/w$ et $\phi_2 \in w_2$ – hyp. inductive, nous tirons $\phi_2 \geq \psi \in w$, puis $\phi_1 \geq \psi \in w$ (axiome A1). Reste à vérifier que $w_1 \in /\phi_1/w$. Supposons que $\phi \in w_1$ mais que $\phi \geq \phi_1 \notin w$. Dans ce cas, $\neg(\phi \geq \phi_1) \in w$, si bien que $\neg\phi \in w_1$. Contradiction.
– *De gauche à droite.* Par contraposition. Supposons que $\phi_1 \geq \phi_2 \notin w$. Soit w_1 l'une des extensions L-maximales consistantes de

$$X\prime = \{\neg\phi : \neg(\phi \geq \phi_2) \in w\} \cup \{\phi_2\}.$$

De $\phi_2 \in w_1$, on tire $w_1 \models \phi_2$ – hyp. inductive. Prenons w_2 tel que $w_2 \models \phi_1$. On a déjà $\phi_1 \in w_2$ – hyp. inductive. Supposons *a contrario* que $w_2 \preceq_w w_1$, c'est-à-dire que

$$\forall\psi \text{ tel que } w_1 \in /\psi/w, \exists\phi \text{ tel que } w_2 \in /\phi/w \text{ et } \phi \geq \psi \in w \tag{3.8}$$

Un raisonnement similaire à celui que nous avons tenu lorsque nous argumentions de droite à gauche permet de montrer que $w_1 \in /\phi_2/w$. Il vient donc – par (3.8) :

$$\exists\phi \text{ tel que } w_2 \in /\phi/w \text{ et } \phi \geq \phi_2 \in w \tag{3.9}$$

Or, le fait que $w_2 \in /\phi/w$ veut dire que $\forall\psi'$ si $\psi' \in w_2$, alors $\psi' \geq \phi \in w$. Ainsi, $\phi_1 \geq \phi \in w$ (puisque $\phi_2 \in w_1$). D'où $\phi_1 \geq \phi_2 \in w$. Cela contredit l'hypothèse de départ. Conclusion : $w \not\models \phi_1 \geq \phi_2$.

□

PROPOSITION 3 (*Complétude*)
Si $\models \phi$ alors $\vdash \phi$

DÉMONSTRATION -
Selon le schéma habituel. □

3.3.2 Un conditionnel de type séquentiel

Nous sommes maintenant en mesure d'introduire un conditionnel de type séquentiel. Pour plus de simplicité, nous supposerons l'opérateur prohairétique sous-jacent défini à la façon de Halpern. Nous posons :

DEFINITION 8 (*Clause dérivée)*
$m/h \models \bigcirc(\psi/\phi)$ ssi :

$$\begin{aligned}
&(\exists h' \in H_m)[\, m/h' \models \phi \ \& m+1/h' \models \psi\,]\ \& \\
&\quad (\forall h \in H_m)\Big(\,[\, m/h \models \phi \ \& m+1/h \not\models \psi\,] \to \\
&\qquad (\exists h' \in H_m)\big[m/h' \models \phi \ \& m+1/h' \models \psi \ \& h' \prec h\ \& \\
&\qquad\qquad \neg(\exists h'' \in H_m)(m/h'' \models \phi \ \& m+1/h'' \not\models \psi \ \& h'' \prec h')\big]\Big)
\end{aligned}$$

Intuitivement : $\bigcirc(\psi/\phi)$ est vrai en m/h si et seulement si les deux conditions suivantes sont toutes deux vérifiées :

1. Quelque histoire rend vrai « maintenant ϕ et ensuite ψ » (non-vacuité) ;
2. Chaque histoire qui vérifie « maintenant ϕ et ensuite non-ψ » est minimisée par quelque histoire vérifiant « ϕ et ensuite ψ », laquelle histoire a cette propriété remarquable de n'être elle-même minimisée par aucune histoire vérifiant « maintenant ϕ et ensuite non-ψ ».

EXEMPLE 1
A titre d'illustration, nous prendrons l'exemple d'un blâme. Un énoncé comme

(a) *Tu n'aurais pas dû poser cette question à nouveau !*

paraît difficilement exprimable dans une sémantique préférentielle habituelle. Pour les besoins de l'analyse, rien ne nous empêche d'introduire un nouvel opérateur séquentiel, que nous noterons, pour le distinguer du précédent :

$$\phi \mathrm{Y} \psi,$$

et affirmant cette fois, outre l'actualité de ϕ, l'avènement de ψ à l'instant immédiatement passé. Nous suggérons de traduire (a) par

$$\phi \mathrm{Y}(\phi \wedge \bigcirc(\neg\phi/\phi)).$$

Lecture :

> *Maintenant ϕ et à l'instant immédiatement passé : ϕ et si ϕ alors obligatoirement $\neg\phi$ aussitôt après.*

Le diagramme ramifié ci-dessous donne un modèle de (a) :

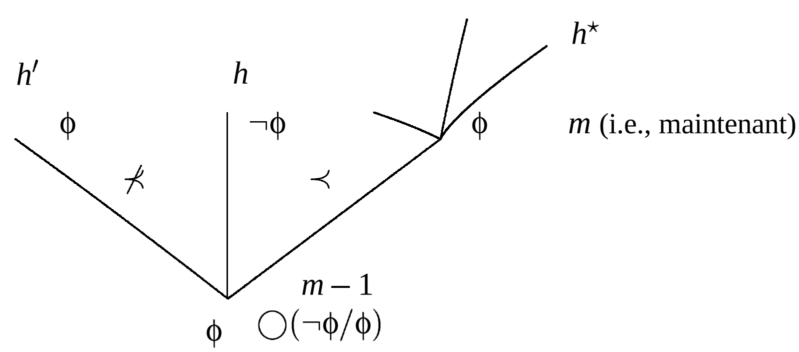


FIG. 3.1 – Exemple du blâme

$h^\star$ désigne ici l'évolution effective de la situation. Supposons que nous soyons en m. Nous avons $m/h^\star \models \phi$ (pour ϕ :=« poser cette question »). Plaçons-nous en $m-1$. Nous avons $m-1/h^\star \models \phi$. Nous avons aussi $m-1/h^\star \models \bigcirc(\neg\phi/\phi)$: l'histoire $h^\star$, qui en m vérifie ϕ, est minimisée par l'histoire h, laquelle en m falsifie ϕ ; à son tour, l'histoire h n'est minimisée par aucune histoire h' vérifiant ϕ en m. Cela donne :

$$m/h^\star \models \phi \mathrm{Y}(\phi \wedge \bigcirc(\neg\phi/\phi)).$$

Nous avons ici un contre-exemple au principe d'identité, $\bigcirc(\phi/\phi)$. Une condition ϕ devient modifiable.

REMARQUE 1
L'idée d'exiger de l'évènement en position d'antécédent qu'il soit antérieur à l'évènement en position de conséquent nous a été directement suggérée par la lecture de Spohn [239] et Alchourrón [3], qui déclarent voir en elle un moyen évident d'éliminer la validité de l'axiome $\bigcirc(\phi/\phi)$. A cet égard, notre contribution aura été d'avoir cherché à donner un contenu formel à leur idée. Comme l'indique David Makinson [160], un conditionnel peut exhiber bien d'autres structures ABC (« antecedent-before-consequent »). Nous nous intéressons ici à l'une des plus simples qu'il mentionne, au moins pour commencer. La formule $(\phi;\psi) \gg (\phi;\neg\psi)$ doit être lue : « si ϕ est vrai à l'instant présent, alors ψ doit être le cas aussitôt après ». Nous avons à notre disposition un moyen relativement simple d'exprimer d'autres types de structures ABC. Il nous suffirait d'introduire des modalités temporelles autres que l'opérateur binaire *et ensuite*. Pour plus de simplicité, nous ne ferons pas l'inventaire des différentes types de structures ABC que nous pourrions au juste exprimer en utilisant ce procédé. Nous ne dresserons pas non plus la liste de celles qui, inévitablement, nous échapperaient.

REMARQUE 2
La construction que nous avons proposée présente quelques affinités avec celle qu'étudient Lindstrom et Rabinowicz dans le domaine ontique et, plus précisément, dans le domaine de la révision. Nous avons décrit le principe de leur approche dans un précédent chapitre[24]. Ils étudient la logique d'expressions de la forme

$$[*\phi]\mathbf{B}\psi,$$

dont le sens informel est

Révisant par ϕ, le locuteur croira nécessairement ψ.

Dans [146], les auteurs suggèrent d'introduire le conditionnel ontique (notons-le >) au moyen de la définition :

$$\mathbf{B}(\phi > \psi) \overset{\text{def}}{\longleftrightarrow} \mathbf{B}[*\phi]\mathbf{B}\psi.$$

Néanmoins, comme leur opérateur de révision vérifie le principe de succès, c'est-à-dire la loi

$$\vdash [*\phi]\mathbf{B}\phi,$$

et que l'opérateur de croyance **B** est un opérateur normal qui obéit à la règle

Si ϕ est une tautologie alors $\vdash \mathbf{B}\phi$,

il en résulte immédiatement que nous avons la loi :

$$\vdash \mathbf{B}(\phi > \phi).$$

[24]Voir plus haut, page 78.

3.3.3 Minimalité

L'analyse esquissée dans ce début de chapitre soulève un certain nombre de questions, dont voici les deux principales :

1. Est-il possible d'établir un lien entre la présente approche et l'approche traditionnellement faite en termes de minimalité ?
2. Que deviennent les autres lois du système non-monotone **P** ? En particulier, qu'advient-il de la cumulativité ?

Nous les envisageons tour à tour. Tout d'abord, nous montrons que, quoique la notion de minimalité ne soit pas directement intervenue dans notre propos, néanmoins nous y faisions implicitement référence. Dans ce qui suit, $\|\phi\|_m$ regroupe, sur l'ensemble des histoires h qui passent par m, le sous-ensemble de celles qui vérifient ϕ (au moment m).

DEFINITION 9
$\min(\|\phi\|_m, \preceq) = \{h \in \|\phi\|_m : \forall h' \in \|\phi\|_m \text{ si } h' \preceq h \text{ alors } h \preceq h'\}$

DEFINITION 10
$\preceq$ est « localement » stoppered, si pour toute formule ϕ et tout $h \in H_m$, si $h \in \|\phi\|_m$ alors il existe $h' \in \|\phi\|_m$ tel que $h' \preceq h$, et $h' \in \min(\|\phi\|_m, \preceq)$.

LEMME 4
Si H_m est fini, alors $\preceq$ est « localement » stoppered.

DÉMONSTRATION -
Immédiat en procédant par l'absurde. Nous laissons au lecteur le soin de le vérifier. □

PROPOSITION 4 (*minimalité)*
Supposons H_m fini. Prenons ϕ historiquement consistant, i.e. $\|\phi\|_m \neq \emptyset$. Les deux énoncés suivants sont équivalents :

(1) $(\exists h' \in H_m)[\, m/h' \models \phi \,\&\, m+1/h' \models \psi \,]$ &

$$(\forall h \in H_m)\Big([\, m/h \models \phi \;\; \& m+1/h \not\models \psi \,] \rightarrow \\ (\exists h' \in H_m)\big[m/h' \models \phi \;\; \& \, m+1/h' \models \psi \;\; \& \, h' \prec h \, \& \\ \neg(\exists h'' \in H_m)(m/h'' \models \phi \;\; \& \, m+1/h'' \not\models \psi \;\; \& \, h'' \prec h')\big]\Big)$$

(2) $\min(\|\phi\|_m, \preceq) \subseteq \|N\psi\|_m$,
où N dénote l'opérateur unaire *ensuite*, tel qu'usuellement défini

DÉMONSTRATION -

• (1) implique (2) :

Soit $h_1 \in \min(\|\phi\|_m, \preceq)$. Par définition, $h_1 \in \|\phi\|_m$. Imaginons que $h_1 \notin \|N\psi\|_m$. Par (1), $\exists h_2$ tel que $m/h_2 \models \phi; \psi$ et $h_2 \prec h_1$. Mais ceci n'est pas compatible avec la minimalité supposée de h_1. D'une part, si $h_2 \prec h_1$ alors $h_2 \preceq h_1$. Et si $h_2 \preceq h_1$ alors $h_1 \preceq h_2$ – def du min. D'autre part, si $h_2 \prec h_1$ alors $h_1 \not\preceq h_2$. Contradiction. Donc $h_1 \in \|N\psi\|_m$.

• (2) implique (1) :

Soit $h_1 \in \|\phi\|_m$ – ϕ localement consistant. Supposons H_m fini. Dans ce cas $\exists h_2 \in \|\phi\|_m$ tel que $h_2 \preceq h_1$ et $h_2 \in \min(\|\phi\|_m, \preceq)$ [Lemme 4]. Par def du min, $m/h_2 \models \phi$. Par (2), $m+1/h_2 \models \psi$. Ainsi, le conjoint gauche de (1) est vrai. Reste à établir le conjoint droit de (1). On procède ici en deux temps :

– Prenons h_1 tel que $m/h_1 \models \phi$ et $m+1/h_1 \not\models \psi$. De $m/h_1 \models \phi$, nous tirons $\exists h_2 \in \|\phi\|_m$ tel que $h_2 \preceq h_1$ et $h_2 \in \min(\|\phi\|_m, \preceq)$ – Lemme 4. Par (2), $m+1/h_2 \models \psi$. Reste à montrer que, non seulement $h_2 \preceq h_1$, mais aussi $h_2 \prec h_1$. Admettons *a contrario* que $h_1 \preceq h_2$. Interagissant avec le fait que $m/h_1 \models \phi$ et que $h_2 \in \min(\|\phi\|_m, \preceq)$, cela donnerait (par la seule transitivité de $\preceq$) $h_1 \in \min(\|\phi\|_m, \preceq)$. Par (2), nous aurions alors $m+1/h_1 \models \psi$ - une contradiction. Donc $h_1 \not\preceq h_2$, comme escompté.
– Prenons h_3 tel que $m/h_3 \models \phi$ et $m+1/h_3 \not\models \psi$. Admettons *a contrario* que $h_3 \prec h_2$, i.e. $h_3 \preceq h_2$ et $h_2 \not\preceq h_3$. De $m/h_3 \models \phi$, $h_3 \preceq h_2$ et $h_2 \in \min(\|\phi\|_m, \preceq)$, il vient $h_2 \preceq h_3$. – une contradiction. Donc $h_3 \not\prec h_2$, comme escompté.

□

Ainsi, les conditions de vérité pour notre opérateur $\bigcirc(/)$ sont reformulables en termes de minimalité. Cette constatation nous incite ici à nous risquer à une brève incursion dans le système DARB d'Åqvist et Hoepelman [19] et Åqvist [15]. Comme la sémantique que les auteurs utilisent diffère sensiblement de la nôtre, nous resterons sur un plan informel. Dans DARB, l'obligation conditionnelle est désignée par Shall(ψ/ϕ). Cette expression est lue « si ϕ est maintenant le cas alors il est obligatoire que ψ soit maintenant le cas ». Les conditions de récurrence pour cet opérateur dyadique sont formulées à la façon de Hansson, c'est-à-dire au moyen d'une relation de bonté comparative. Dans DARB, la locution « ensuite ϕ » est représentée par $\oplus\phi$. Leur système nous autorise à écrire des expressions du type

$$\text{Shall}(\oplus\psi/\phi), \tag{3.10}$$

lesquelles semblent capter exactement le sens de notre opérateur

$$\bigcirc(\psi/\phi).$$

Quoique le système DARB analyse la norme séquentielle d'une façon apparemment analogue, d'importantes différences subsistent pourtant. Voici, selon nous, la plus significative. Les auteurs représentent l'obligation inconditionnelle par SHALLϕ, ce qui peut être lu « il est obligatoire que (maintenant) ϕ ». SHALLϕ nous est présenté comme une abréviation de $\text{Shall}(\phi/\top)$ − axiomes A39 et A41 dans [19, p.201]. Notons $\rightarrow$ l'implication matérielle. Il nous paraît intéressant de remarquer que, quoiqu' Åqvist et son collaborateur définissent les conditions de vérité de leurs foncteurs normatifs en termes de minimalité, nous avons les équivalences suivantes − théorèmes T48 et T49 dans [19, p. 217] :

$$\text{Shall}(\psi/\phi) \leftrightarrow (\phi \rightarrow \text{SHALL}\psi) \text{ si } \phi \text{ ne contient pas } \oplus \tag{3.11}$$

$$\leftrightarrow \text{SHALL}(\phi \rightarrow \psi) \text{ si } \phi \text{ ne contient pas } \oplus \tag{3.12}$$

Ainsi, dans DARB toutes les fois où l'antécédant désigne une formule propositionnelle ordinaire, renvoyant au présent de l'énonciation, la norme séquentielle $\text{Shall}(\oplus\psi/\phi)$ se prête à une reformulation en termes d'implication matérielle, qui lui est substituable sans qu'il n'y ait déperdition d'information. Comme nous, les auteurs notent $\Box$ l'opérateur de nécessité historique ; il est défini par quantification universelle sur l'ensemble des branches traversant le moment auquel l'évaluation est faite. Du point de vue axiomatique, (3.11) et (3.12) viennent de ce que dans DARB nous avons la loi − il s'agit du théorème 2 dans [19, p. 206] −

$$\phi \rightarrow \Box\phi \text{ si } \phi \text{ ne contient pas } \oplus, \tag{3.13}$$

Sauf hypothèse supplémentaire, notre sémantique ne valide pas ces lois.

3.3.4 Quelques lois logiques

Fermant cet aparté sur le système DARB, nous revenons à notre cadre d'analyse et examinons dans cette section ce que deviennent les lois caractéristiques du système préférentiel **P**. Nous avons vu que nous perdons le principe d'identité, $\bigcirc(\phi/\phi)$. Qu'en est-il des autres lois du système ? Nous allons ici les désigner en utilisant des dénominations qui nous sont propres. Le lecteur verra à quelle loi du système **P** il est fait allusion, en se reportant à la présentation que nous avons donnée de ce système au début de notre étude[25]. Dans une expression de la forme $\bigcirc(\psi/\phi)$, on considérera ϕ comme la pré-condition et ψ comme la post-condition.

PROPOSITION 5
Nous avons :

[25] Voir p. 66.

1. Remplacement de la précondition :
$$\frac{\phi \leftrightarrow \phi'}{\bigcirc(\psi/\phi) \leftrightarrow \bigcirc(\psi/\phi')} \tag{Remp}$$
2. Affaiblissement de la post-condition :
$$\frac{\psi \rightarrow \psi'}{\bigcirc(\psi/\phi) \rightarrow \bigcirc(\psi'/\phi)} \tag{Affai}$$
3. Synchronisation des post-conditions :
$$(\bigcirc(\phi_1/\phi) \wedge \bigcirc(\phi_2/\phi)) \rightarrow \bigcirc(\phi_1 \wedge \phi_2/\phi) \tag{Synch1}$$
4. Synchronisation des pré-conditions :
$$(\bigcirc(\phi/\phi_1) \wedge \bigcirc(\phi/\phi_2)) \rightarrow \bigcirc(\phi/\phi_1 \vee \phi_2) \tag{Synch2}$$
5. Forme restrictive de post-renforcement de la pré-condition :
$$(\bigcirc(\phi_1/\phi) \wedge \bigcirc(\phi_2/\phi)) \rightarrow \bigcirc(\phi_2/\phi;\phi_1) \tag{Renf}$$
6. Triangulation :
$$(\bigcirc(\phi_1/\phi) \wedge \bigcirc(\phi_2/\phi;\phi_1)) \rightarrow \bigcirc(\phi_2/\phi) \tag{Tri}$$

Le loi (Renf) est une variation sur le principe de monotonie prudente, qui apporte une restriction sur le renforcement de l'antécédant. Un renforcement ordinaire est de la forme «avant ϕ, maintenant $\phi \wedge \phi_1$». Il est ici de la forme «avant ϕ, maintenant $\phi;\phi_1$». C'est pourquoi nous parlons de *post-renforcement*. La loi (Tri), que nous nommons (faute d'un meilleur terme) « triangulation », est une variation sur le principe de coupure, qui énonce une forme restreinte de transitivité. Prises ensembles, les lois (Renf) et (Tri) donnent une variation sur le principe de cumulativité. Dans la formule $\bigcirc(\phi_2/\phi;\phi_1)$, l'énoncé ϕ se réfère au présent de l'énonciation, tandis que les énoncés ϕ_1 et ϕ_2 se réfèrent tous deux à l'instant suivant. Littéralement, $\bigcirc(\phi_2/\phi;\phi_1)$ nous dit que
$$(\phi;\phi_1);\phi_2 \gg (\phi;\phi_1);\neg\phi_2.$$
Nous pouvons rapprocher cette expression d'une autre expression qui lui est logiquement équivalente. Il s'agit de :
$$\phi;(\phi_1 \wedge \phi_2) \gg \phi;(\phi_1 \wedge \neg\phi_2).$$

Cette deuxième expression nous dit que, étant donnée la vérification de ϕ au moment présent, le fait que ϕ_1 soit vrai en même temps que ϕ_2 à l'instant suivant est préférable au fait que ϕ_1 soit vrai mais pas ϕ_2.

Nous notons au passage que, dans le système DARB, (Renf) et (Tri) sont toutes les deux dérivables − par les théorèmes T31 et T32 dans [19, p. 213]. Néanmoins, nous nous souvenons de cette particularité du système DARB : dans le cas particulier où l'antécédent ne contient pas d'occurrences de l'opérateur unaire $\oplus$ (et ensuite), une norme séquentielle se laisse paraphraser en termes d'implication matérielle. De la sorte, dans ce système, nous retrouvons (trivialement) la monotonie prudente dans sa forme initiale, où « $\wedge$ » remplace « ; ».

DÉMONSTRATION Pour vérifier notre groupe de lois, nous procédons par voie axiomatique, en faisant appel aux lois que vérifient les opérateurs « $\gg$ » et « ; ». Nous omettons la première étape, qui consiste à remplacer chaque $\bigcirc(\phi/\psi)$ par $(\phi;\psi) \gg (\phi;\neg\psi)$. Même remarque pour l'étape finale, qui consiste à faire le remplacement inverse.

- Pour (Remp). Supposons que $\phi \leftrightarrow \phi\prime$ soit valide. Dans ce cas, les formules $(\phi;\psi) \rightarrow (\phi\prime;\psi)$ et $(\phi\prime;\neg\psi) \rightarrow (\phi;\neg\psi)$ sont valides. Les formules $\Box((\phi;\psi) \rightarrow (\phi\prime;\psi))$ et $\Box((\phi\prime;\neg\psi) \rightarrow (\phi;\neg\psi))$ le sont aussi [loi P3]. Il en résulte que nous avons aussi la loi [par P2] :

$$((\phi;\psi) \gg (\phi;\neg\psi)) \rightarrow ((\phi\prime;\psi) \gg (\phi\prime;\neg\psi)).$$

  La validité de la converse se prouve en usant d'une argumentation similaire.
- Pour (Affai). Le raisonnement est similaire − en utilisant T9, P3 et P2.
- Pour (Synch1).
  $\big(((\phi;\phi_1) \gg (\phi;\neg\phi_1)) \wedge ((\phi;\phi_2) \gg (\phi;\neg\phi_2))\big)$

$$\begin{aligned}
&\rightarrow \Big(\big[(\phi;(\phi_1 \wedge \phi_2)) \vee (\phi;(\phi_1 \wedge \neg\phi_2))\big] \gg \big[(\phi;(\neg\phi_1 \wedge \phi_2)) \vee (\phi;(\neg\phi_1 \wedge \neg\phi_2))\big]\Big)\\
&\quad \wedge \Big(\big[(\phi;(\phi_1 \wedge \phi_2)) \vee (\phi;(\neg\phi_1 \wedge \phi_2)) \vee (\phi;(\neg\phi_1 \wedge \neg\phi_2))\big] \gg (\phi;(\phi_1 \wedge \neg\phi_2))\Big)\\
&\qquad\qquad [\text{par T12, T1, P3 et P2}]\\
&\rightarrow (\phi;(\phi_1 \wedge \phi_2)) \gg \big((\phi;(\phi_1 \wedge \neg\phi_2)) \vee (\phi;(\neg\phi_1 \wedge \phi_2)) \vee (\phi;(\neg\phi_1 \wedge \neg\phi_2))\big) \quad [\text{par P1}]\\
&\rightarrow (\phi;(\phi_1 \wedge \phi_2)) \gg \big[\phi;\big((\phi_1 \wedge \neg\phi_2) \vee (\neg\phi_1 \wedge \phi_2) \vee (\neg\phi_1 \wedge \neg\phi_2)\big)\big] \quad [\text{par T0}]\\
&\rightarrow (\phi;(\phi_1 \wedge \phi_2)) \gg (\phi;\neg(\phi_1 \wedge \phi_2)) \quad [\text{calcul prop. , P3 et P2}]
\end{aligned}$$

- Pour (Synch2).

$$\begin{aligned}
&\big(((\phi_1;\phi) \gg (\phi_1;\neg\phi)) \wedge ((\phi_2;\phi) \gg (\phi_2;\neg\phi))\big)\\
&\qquad \to \Big(\big(((\phi_1 \vee \phi_2);\phi) \vee (\phi_2;\neg\phi)\big) \gg (\phi_1;\neg\phi)\Big)\\
&\qquad\qquad \wedge \Big(\big(((\phi_1 \vee \phi_2);\phi) \vee (\phi_1;\neg\phi)\big) \gg (\phi_2;\neg\phi)\Big) \;\; [\text{par T0}, \text{P3}, \text{P2}]\\
&\qquad \to ((\phi_1 \vee \phi_2);\phi) \gg \big((\phi_1;\neg\phi) \vee (\phi_2;\neg\phi)\big) \;\; [\text{par P1}]\\
&\qquad \to ((\phi_1 \vee \phi_2);\phi) \gg ((\phi_1 \vee \phi_2);\neg\phi) \;\; [\text{par T0}, \text{P3}, \text{P2}].
\end{aligned}$$

- Pour (Renf).

$$\begin{aligned}
&\big(((\phi;\phi_1) \gg (\phi;\neg\phi_1)) \wedge ((\phi;\phi_2) \gg (\phi;\neg\phi_2))\big)\\
&\qquad \to \Big(\big((\phi;(\phi_1 \wedge \phi_2)) \vee (\phi;(\phi_1 \wedge \neg\phi_2))\big) \gg (\phi;(\neg\phi_1 \wedge \phi_2))\Big)\\
&\qquad\qquad \wedge \Big(\big((\phi;(\phi_1 \wedge \phi_2)) \vee (\phi;(\neg\phi_1 \wedge \phi_2))\big) \gg (\phi;(\phi_1 \wedge \neg\phi_2))\Big)\\
&\qquad\qquad\quad [\text{par T12}, \text{T1}, \text{P3}, \text{P2}]\\
&\qquad \to (\phi;(\phi_1 \wedge \phi_2)) \gg \big((\phi;(\phi_1 \wedge \neg\phi_2)) \vee (\phi;(\neg\phi_1 \wedge \phi_2))\big) \;\; [\text{par P1}]\\
&\qquad \to (\phi;(\phi_1 \wedge \phi_2)) \gg (\phi;(\phi_1 \wedge \neg\phi_2)) \;\; [\text{P3}, \text{P2}]\\
&\qquad \to ((\phi;\phi_1);\phi_2) \gg ((\phi;\phi_1);\neg\phi_2) \;\; [\text{T8}, \text{P3}, \text{P2}]
\end{aligned}$$

- Pour (Tri).

$$\begin{aligned}
&\big((\phi;\phi_1) \gg (\phi;\neg\phi_1)\big) \wedge \big((\phi;\phi_1);\phi_2 \gg (\phi;\phi_1);\neg\phi_2\big)\\
&\qquad \to \big(((\phi;\phi_1) \gg (\phi;\neg\phi_1)) \wedge (\phi;(\phi_1 \wedge \phi_2) \gg \phi;(\phi_1 \wedge \neg\phi_2))\big) \;\; [\text{T8}, \text{P3}, \text{P2}]\\
&\qquad \to \Big(\big((\phi;(\phi_1 \wedge \phi_2)) \vee (\phi;(\phi_1 \wedge \neg\phi_2))\big) \gg (\phi;(\neg\phi_1 \wedge \neg\phi_2))\Big)\\
&\qquad\qquad \wedge \Big(\big((\phi;(\phi_1 \wedge \phi_2)) \vee (\phi;(\neg\phi_1 \wedge \neg\phi_2))\big) \gg (\phi;(\phi_1 \wedge \neg\phi_2))\Big)\\
&\qquad\qquad\quad [\text{T12}, \text{T1}, \text{P3}, \text{P2}]\\
&\qquad \to (\phi;(\phi_1 \wedge \phi_2)) \gg \big((\phi;(\phi_1 \wedge \neg\phi_2)) \vee (\phi;(\neg\phi_1 \wedge \neg\phi_2))\big) \;\; [\text{P1}]\\
&\qquad \to (\phi;\phi_2) \gg (\phi;\neg\phi_2) \;\; [\text{T1}, \text{T12}, \text{P3}, \text{P2}]
\end{aligned}$$

□

3.3.5 Questions ouvertes

Tentons d'apprécier le chemin parcouru jusqu'ici. Nous sommes partis de l'idée que les analyses de l'obligation conditionnelle tendent à se répartir en deux catégories : celles qui prennent l'opérateur dyadique d'obligation comme premier terme non-défini, et celles qui définissent celui-ci à l'aide de l'opérateur monadique d'obligation et d'un conditionnel ontique non-monotone. Dans le premier cas, le principe d'identité et le principe de cumulativité sont tous les deux valides. Dans le second cas, aucun des deux ne l'est. Nous avons remarqué que l'analyse que propose Makinson relève de la deuxième catégorie. L'originalité de sa sémantique est qu'elle introduit la dimension du futur. Mais un examen attentif révèle que ce n'est pas la prise en compte de cette dimension qui permet l'élimination du principe d'identité. Nous nous sommes alors demandés si une voie intermédiaire n'était pas envisageable, et s'il n'était pas possible d'éliminer le principe d'identité tout en maintenant une forme de cumulativité, en recourant à une logique du changement. Dans ce qui suit, nous présentons et tentons de discuter trois objections qui peuvent être élevées contre cette analyse. Les questions qu'elles soulèvent nous paraissent suffisamment difficiles pour que nous nous permettions de les laisser ouvertes.

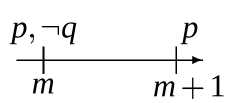
QUESTION 1
Dans l'hypothèse où cette position intermédiaire serait souhaitable, nous pouvons nous étonner que le principe d'identité puisse réapparaître sous la forme[26]

$$\bigcirc(\phi_3/\phi_1;\phi_2) \to \bigcirc(\phi_2 \wedge \phi_3/\phi_1;\phi_2). \tag{3.14}$$

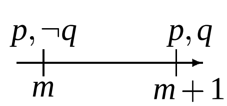
Par exemple, si dans (3.14) nous substituons $p \wedge \neg q$, p et q à (respectivement) ϕ_1, ϕ_2 et ϕ_3, nous obtenons un principe de la forme :

$$\bigcirc(q/(p \wedge \neg q);p) \to \bigcirc(p \wedge q/(p \wedge \neg q);p). \tag{3.15}$$

(3.15) nous paraît contre-intuitive. En effet, l'antécédent $\bigcirc(q/(p \wedge \neg q);p)$ nous dit que, si q est maintenant faux et que p continue à être vrai, i.e.



alors il est obligatoire qu'il commence que q, i.e.



[26]Ceci est une remarque de Leon van der Torre et du rapporteur de la revue *Fundamenta Informaticae*.

La formule $\bigcirc(p \wedge q/(p \wedge \neg q); p)$ rend obligatoire le fait que p continue à être vrai. Ceci est paradoxal, car nous pourrions être dans un contexte où $\bigcirc(\neg p/p)$. Intuitivement, nous ne souhaitons pas pouvoir déduire de la proposition selon laquelle :

> si l'on continue à fumer, alors il est recommandé de faire du sport,

la conclusion selon laquelle :

> il est recommandé de continuer à fumer.

Sauf erreur de notre part, dans le système DARB d'Åqvist, l'implication (3.14) est également dérivable[27].

Pour notre part, nous dissiperions volontiers le paradoxe, en rapprochant (3.14) de deux autres lois de logique temporelle. La première correspond à la partie implicative « de gauche à droite » du biconditionnel (T10) tout à l'heure démontré :

$$\vdash ((\phi_1;\phi_2);\phi_3) \rightarrow (\phi_1;\phi_2);(\phi_2 \wedge \phi_3). \tag{3.16}$$

La seconde est la partie implicative « de gauche à droite » de l'équivalence (T11) ; elle fonctionne un peu comme la duale de la précédente implication :

$$\vdash ((\phi_1;\phi_2);\neg(\phi_2 \wedge \phi_3)) \rightarrow (\phi_1;\phi_2);\neg\phi_3. \tag{3.17}$$

Du point de vue axiomatique, nous tirons facilement (3.14) à partir du couple (3.16) et (3.17). Il suffit d'invoquer l'axiome prohairétique (P2). Concentrons-nous sur (3.16). L'antécédent de cette implication enchâsse une séquence dans une autre. Nous remarquons immédiatement que cette implication est correcte, uniquement parce que la post-condition de la séquence enchâssée et celle de la séquence enchâssante sont supposées s'inscrire dans le futur évidemment unique d'une route privilégiée, celle-là même qui (dirons-nous) porte l'évolution réelle du moment présent. Une remarque similaire s'applique à (3.17). Ainsi, la difficulté vient seulement de ce que, bien que travaillant avec des structures ramifiées vers le futur, nous avons naïvement admis que l'opérateur *et ensuite* obéit aux lois d'une logique du temps linéaire [28]. Or, certaines logiques du temps arborescent, comme le système UB de Ben-Ari et al. [27], s'attaquent directement à cette hypothèse. Elles nous invitent à distinguer deux séries d'opérateurs temporels. Précédés de $\forall$, les premiers expriment des propriétés vérifiées sur toutes les branches de l'arbre.

---

[27] En utilisant par exemple la règle R4 et les théorèmes T24/T20 donnés dans [19], et en les appliquant à la loi $(\phi_1 \wedge \oplus\phi_2) \rightarrow (\oplus\phi_3 \rightarrow (\oplus\phi_3 \wedge \oplus\phi_2))$, elle-même dérivable à partir de la loi $(\oplus\phi_2 \wedge \oplus\phi_3) \rightarrow (\oplus\phi_3 \wedge \oplus\phi_2)$.

[28] Comme le rappelle e.g. Zanardo [280], cette hypothèse est implicitement faite par toute sémantique formulée en termes de <moment, histoire>. La notion d'indéterminisme est prise en charge par un opérateur de possibilité historique (le dual de notre $\Box$), qui est défi ni par quantifi cation existentielle sur l'ensemble des branches traversant le moment présent *m*.

Précédés de $\exists$, les seconds expriment des propriétés vérifiées sur un chemin de l'arbre. Pour éliminer (3.16) et (3.17) – et par là même (3.14), il nous suffirait visiblement d'introduire les définitions :

$$\text{Def1.}\quad \bigcirc(\phi_2/\phi_1) \overset{\text{def}}{\longleftrightarrow} (\phi_1 \wedge \exists \oplus \phi_2) \gg (\phi_1 \wedge \exists \oplus \neg\phi_2) \quad (\oplus\text{= et ensuite})$$

$$\text{Def2.}\quad \phi_1;\phi_2 \overset{\text{def}}{\longleftrightarrow} \phi_1 \wedge \exists \oplus \phi_2$$

Nous notons ici que l'infinité du temps en direction de l'avenir s'exprime sous la forme de la loi

$$(\exists \oplus \phi) \vee (\exists \oplus \neg\phi), \tag{3.18}$$

en présence de laquelle nous dérivons

$$(\phi_1 \wedge \exists \oplus \phi_2) \rightarrow (\phi_1 \wedge \exists \oplus \phi_2 \wedge \exists \oplus \phi_3) \vee (\phi_1 \wedge \exists \oplus \phi_2 \wedge \exists \oplus \neg\phi_3). \tag{3.19}$$

Rapportée à notre problématique de départ, la définition (Def1) présente un vif intérêt. Car, du seul point de vue axiomatique, il nous suffit d'avoir (3.19) pour pouvoir donner de (Renf) et (Tri) une démonstration analogue à la précédente :

• Pour (Renf) :

$$\begin{aligned}
&\big(((\phi \wedge \exists \oplus \phi_1) \gg (\phi \wedge \exists \oplus \neg\phi_1)) \wedge ((\phi \wedge \exists \oplus \phi_2) \gg (\phi \wedge \exists \oplus \neg\phi_2))\big)\\
&\quad \rightarrow \Big(\big((\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \phi_2) \vee (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \neg\phi_2)\big) \gg (\phi \wedge \exists \oplus \neg\phi_1 \wedge \exists \oplus \phi_2)\Big)\\
&\qquad \wedge \Big(\big((\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \phi_2) \vee (\phi \wedge \exists \oplus \neg\phi_1 \wedge \exists \oplus \phi_2)\big) \gg (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \neg\phi_2)\Big)\\
&\qquad\quad [\text{par } 3.19, \text{P3}, \text{P2}]\\
&\quad \rightarrow (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \phi_2) \gg \big((\phi \wedge \exists \oplus \neg\phi_1 \wedge \exists \oplus \phi_2) \vee (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \neg\phi_2)\big) \quad [\text{par P1}]\\
&\quad \rightarrow (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \phi_2) \gg (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \neg\phi_2) \quad [\text{P3}, \text{P2}]
\end{aligned}$$

• Pour (Tri) :

$$\begin{aligned}
&((\phi \wedge \exists \oplus \phi_1) \gg (\phi \wedge \exists \oplus \neg\phi_1)) \wedge ((\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \phi_2) \gg (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \neg\phi_2))\\
&\quad \rightarrow \Big(\big((\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \phi_2) \vee (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \neg\phi_2)\big) \gg (\phi \wedge \exists \oplus \neg\phi_1 \wedge \exists \oplus \neg\phi_2)\Big)\\
&\qquad \wedge \Big(\big((\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \phi_2) \vee (\phi \wedge \exists \oplus \neg\phi_1 \wedge \exists \oplus \neg\phi_2)\big) \gg (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \neg\phi_2)\Big)\\
&\qquad\quad [3.19, \text{P3}, \text{P2}]\\
&\quad \rightarrow (\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \phi_2) \gg \big((\phi \wedge \exists \oplus \phi_1 \wedge \exists \oplus \neg\phi_2) \vee (\phi \wedge \exists \oplus \neg\phi_1 \wedge \exists \oplus \neg\phi_2)\big) \quad [\text{P1}]\\
&\quad \rightarrow (\phi \wedge \exists \oplus \phi_2) \gg (\phi \wedge \exists \oplus \neg\phi_2) \quad [3.19, \text{P3}, \text{P2}]
\end{aligned}$$

QUESTION 2

Nous nous tournons vers une deuxième question. Celle-ci consiste à se demander si notre analyse n'occulte pas une distinction essentielle, celle entre obligation prima facie (défaisable , sujette à exceptions) et obligation actuelle (indéfaisable, non-sujette à exceptions. Cette distinction semble d'une importance déterminante pour notre propos, car nous nous attendrions surtout du premier type d'obligation qu'il satisfasse la cumulativité. Or, en elle-même, la notion de prima facie renvoie à la notion de normalité qui, du point de vue sémantique, est généralement prise en charge par une relation binaire (notons-là $\prec_N$) spécifique définie sur les mondes possibles. Deux sous-questions viennent à l'esprit :

1. Nous avons renvoyé dos à dos Hansson et Chellas. Le premier prend l'opérateur dyadique d'obligation comme premier terme non-défini, et le second définit celui-ci à l'aide de l'opérateur monadique d'obligation et d'un conditionnel ontique non-monotone. Les renvoyer ainsi dos à dos a-t-il bien un sens ? A la différence de la sémantique de Chellas, la sémantique de Hansson ne traite-t-elle pas l'obligation conditionnelle comme indéfaisable (puisque n'utilisant pas d'idiome propre pour la normalité, i.e. $\prec_N$) ? Bref, notre angle d'attaque n'est-il pas en porte-à-faux ?
2. Pris au sens de notre définition 8, l'opérateur séquentiel $\bigcirc(\psi/\phi)$ peut-il réellement être interprété comme l'expression d'une obligation prima facie ?

Commençons par la deuxième question. Notre définition 8 traite le « est » de « Il est obligatoire que » comme un présent proprement dit. Elle suppose également qu'il y ait un laps de temps entre l'exécution de l'obligation et le « est » de « Il est obligatoire que ». Selon van Eck [67, 68], ce seul fait nous autorise à interpréter $\bigcirc(\psi/\phi)$ comme l'expression d'une obligation prima facie. L'instant auquel ψ fait référence est situé dans le futur de l'instant auquel l'obligation conditionnelle est valide. Dans ce laps de temps, des éléments divers peuvent advenir qui libèrent le locuteur de l'obligation qu'il a contractée. En particulier, il se peut qu'entre temps une obligation contraire et plus importante naisse et lui prenne le pas. Si (aujourd'hui) Jones promet à Marie d'emmener les enfants faire (demain) un pique-nique, alors (aujourd'hui) Jones se met dans l'obligation d'emmener les enfants faire (demain) un pique-nique. Il se peut toutefois que, le lendemain matin venu, Jones soit appelé pour une urgence à l'hôpital.

Quoiqu'à première vue séduisante, cette position n'est pas totalement satisfaisante, et soulève plus de problèmes qu'elle n'en résout. Van Eck [67, p. 174-175] suggère que, pour pouvoir dire si une obligation est ou non de l'ordre du prima facie, il faut et il suffit de comparer les indices temporels affectés respectivement à l'opérateur d'obligation et à l'action auquel il s'applique : typiquement, une obligation (inconditionnelle) prima facie serait de la forme

« (au temps t) il faut que (au temps t'>t) »,

tandis qu'une obligation (inconditionnelle) actuelle serait de la forme

« (au temps t) il faut que (au temps t) ».

Ceci semble discutable. Telle ou telle obligation ne peut-elle pas admettre un écart dans le temps entre le moment où elle naît et le moment où elle stipule que l'action doit être exécutée, et cependant être traitée par l'agent comme indéfaisable et comme n'étant pas dominable par une obligation contraire ? Inversement, telle ou telle norme ne peut-elle pas exclure tout écart dans le temps, et cependant être considérée par l'agent comme susceptible d'exceptions ?

Ainsi, la question de savoir si un énoncé en « doit » exprime ou non une obligation prima facie et celle de savoir s'il exclut ou non tout écart dans le temps nous semblent indépendantes l'une de l'autre. Passons alors au premier problème : à la différence de la sémantique de Chellas, la sémantique de Hansson ne traite-t-elle pas l'obligation conditionnelle comme indéfaisable ? Cette question appelle, croyons-nous, une réponse positive, pour la raison suivante. De même qu'une norme peut être traitée comme étant susceptible d'exceptions, de même elle peut être traitée comme n'étant pas susceptible d'exceptions. Ayant admis cela, supposons que nous souscrivions à l'idée selon laquelle, prise au sens de Hansson, l'obligation conditionnelle exprime une norme du premier type. La question est : comment représenter une norme du second type ? Apparemment, nous n'avons pas d'autres possibilités que d'utiliser l'implication matérielle, soit que nous écrivions $\bigcirc(\phi \rightarrow \psi)$ soit que nous écrivions $\phi \rightarrow \bigcirc\psi$. Or, comme nous l'avons rappelé[29], lorsque le dit énoncé exprime de surcroît une norme réparatrice, nous nous heurtons à un paradoxe, que nous options pour l'une ou l'autre de ces représentations formelles. Cela n'est pas satisfaisant. Nous rejoignons ici une idée chère à David Makinson [158] : le rejet de la monotonie (celui-là même qui nous est nécessaire pour bloquer le paradoxe de l'obligation réparatrice) n'est pas constitutif de la notion de défaisabilité, mais en est seulement un produit dérivé. Aussi, nous avons maintenant l'impression que notre commentaire de *Five Faces of Minimality* était en porte-à-faux. Nous avons axé l'analyse sur le cas de l'obligation conditionnelle défaisable, alors que la dernière section de *Five Faces* envisage celle-ci seulement dans un deuxième temps, après avoir considéré le cas de l'obligation conditionnelle indéfaisable.

QUESTION 3

Nous nous tournons vers une troisième question possible. Elle consiste à se demander si, compte tenu de la seule élimination du principe d'identité, il est même

[29] Voir ci-dessus, p. 117 et ss.

raisonnable d'exiger de l'obligation (prima facie ou autre) qu'elle vérifie le principe de cumulativité. La dérivation que nous donnons ci-dessous suggère en effet que, dès lors que nous supposons que l'antécédent de l'obligation conditionnelle décrit un trait modifiable du contexte, l'application du principe de coupure (et donc la cumulativité) permet de dériver des conclusions apparemment discutables. Ces conclusions sont visiblement discutables, que nous ajoutions ou non une clause *ceteris paribus* au « est » de « Il est obligatoire que ». Utilisons ici le symbolisme de la logique entrée/sortie (*input/output*). Dans une paire de la forme (a,x), on considère le corps a comme une entrée décrivant la situation présente et la tête x comme une sortie énonçant ce qui est désirable, obligatoire, etc.[30] Considérons la dérivation suivante :

$$\dfrac{(a, \neg a \vee b) \qquad \dfrac{(a \wedge b, c)}{(a \wedge (\neg a \vee b), c)}}{(a, c)}$$

Le « second best » promu

La prémisse gauche, $(a, \neg a \vee b)$, stipule que, si a est le cas, alors il faut soit supprimer a soit produire b. Par exemple : « Jones doit arrêter de fumer ou bien (faute de mieux) se mettre au sport ». La prémisse droite, $(a \wedge b, c)$, stipule que, si a et b sont le cas, alors c doit être le cas. Par exemple : « Si Jones se met au sport, il doit s'y mettre sérieusement ». A partir de la prémisse droite, nous déduisons $(a \wedge (\neg a \vee b), c)$, par simple remplacement des équivalents[31]. Puis, de $(a, \neg a \vee b)$ et $(a \wedge (\neg a \vee b), c)$, nous tirons (a,c) par application du principe de coupure. Intuitivement, ceci semble paradoxal : notre disjonction initiale est « écrasée » ; un état qui venait en seconde position seulement passe subrepticement au rang supérieur. La dérivation de (a,c) n'est paradoxale que pour autant que la norme $(a, \neg a \vee b)$ laisse à son destinataire la possibilité de choisir entre $\neg a$ et b. Ainsi, la difficulté se dissiperait si (a,a) était un axiome : devenant réductible à (a,b), la paire $(a, \neg a \vee b)$ n'offrirait plus qu'un choix apparent.

Nous pouvons ici nous demander ce qui se passerait si, dans la dérivation, nous remplacions le signe « $\wedge$ » par le signe « ; ». Une remarque vient aussitôt à l'esprit. Alors que

$$a \wedge b$$

[30] Nous ne ferons ici qu'une timide incursion dans la théorie de l'input/ouput. Pour un aperçu plus avant, cf. Makinson et van der Torre [161, 162].

[31] Dans une sémantique de l'obligation conditionnelle *à la* Chellas, ceci nous est possible, si nous analysons le conditionnel ontique sous-jacent ($\Rightarrow$) en termes de minimalité.

est équivalent à

$$a \wedge (\neg a \vee b),$$

l'énoncé

$$a\,;b \tag{3.20}$$

n'est pas équivalent à

$$a;(\neg a \vee b), \tag{3.21}$$

puisque ce futur proche, dans lequel $\neg a \vee b$ et b sont supposés s'inscrire, ne contient pas nécessairement a. Plus précisément, (3.20) implique (3.21), mais la réciproque est fausse. Ainsi, si nous recourons au type d'analyse séquentielle que nous avons suggéré, notre exemple ne marche plus si aisément. Supposons que nous remplacions les paires $(a, \neg a \vee b)$ et $(a \wedge b, c)$ par (respectivement) $\bigcirc(\neg a \vee b/a)$ et $\bigcirc(c/a;b)$. Dans la dérivation suivante, les pointillés indiquent à quelle étape le raisonnement est bloqué :

$$\dfrac{\bigcirc(\neg a \vee b/a) \qquad \begin{array}{c}\bigcirc(c/a;b)\\ \text{-----------}\\ \bigcirc(c/(a;(\neg a \vee b))\end{array}}{\bigcirc(c/a)}$$

De la proposition selon laquelle $\bigcirc(c/a;b)$, nous ne pouvons pas déduire la conclusion selon laquelle $\bigcirc(c/a;(\neg a \vee b))$, pour la raison que nous venons d'invoquer. Par contre, de $\bigcirc(\neg a \vee b/a)$ et de $\bigcirc(c/a;(\neg a \vee b))$, nous continuerions à pouvoir déduire que $\bigcirc(c/a)$ – ceci, en vertu du principe (Tri) que nous avons rencontré dans l'une des sections précédentes.

Néanmoins, cela ne nous dit pas en quoi la cumulativité est souhaitable dans le déontique. Faute d'arguments vraiment convaincants, dans la section suivante nous tentons de réorienter notre travail dans une autre direction.

3.4 L'échange réparateur

3.4.1 Introduction

Nous nous tournons maintenant vers ce qui va constituer la nouvelle charpente de ce chapitre. Toute logique déontique, nous l'avons dit, est confrontée au problème suivant : elle doit rendre compte de manière adéquate de nos argumentations, lorsqu'elles mettent en jeu ce qu'on appelle des « normes réparatrices » (*contrary-to-duty*). Chisholm [49] le premier attira notre attention sur les difficultés liées à cette notion. Elles faisaient principalement objection à la logique des normes de von Wright [275]. Nous axerons ici notre propos sur deux solutions qui ont été proposées et qui, généralement, sont présentées comme distinctes, voire opposées. Historiquement, la première solution vient (à notre connaissance) d'Åqvist [13]. Elle consiste à recourir à une logique temporelle. Danielsson [53] et Hansson [103] sont tous les deux à l'origine de la seconde solution qui va nous intéresser. Elle consiste à recourir à une sémantique préférentielle. La seconde approche vient aider la première, dit-on souvent, lorsque les obligations en jeu se réfèrent à la même unité de temps, comme dans l'exemple souvent cité : *ne tue pas* ; *si tu tues, fais-le gentiment*. L'expression « se réfèrent à » est équivoque. Nous dirons en quel sens nous la prenons.

Certains demanderont s'il est possible d'intégrer ces deux traitements apparemment complémentaires dans un même formalisme, du genre de celui que nous venons de présenter. A dire vrai, la question des normes réparatrices nous paraît suffisamment difficile et obscure, pour que nous nous permettions de ne pas suivre cette première piste possible. Dans un premier temps, nous décrirons et analyserons le principe sous-jacent au compte rendu de type temporel (§ 3.4.2 et § 3.4.3). Dans un deuxième temps, bifurquant quelque peu, nous tenterons une analyse préférentielle qui nous est propre et qui sera basée sur la notion de révision itérée (§ 3.4.4 à § 3.4.7). Nous souhaiterions montrer que, contrairement aux apparences, il y a sans doute une affinité profonde ente les deux types de comptes rendus. En gros, la seconde solution conserve l'esprit de la première, tout en évitant certaines de ses obscurcités, liées au cas où la réparation précède la violation.

Nous ne nous sentons pas suffisamment au clair sur les tenants et les aboutissants des comptes rendus préférentiels déjà existants. C'est pourquoi nous avons choisi de ne pas les évoquer. Il en est de même des exemples cités dans la littérature. Nous ferons plutôt porter la discussion sur ce que Goffman [93] nomme l'« échange réparateur » (*remedial interchange*), qui pour l'essentiel renvoie au mécanisme de réparation d'une offense territoriale. Admettant que les réserves de type territorial constituent la revendication principale des individus en groupe, Goffman fait de l'échange réparateur la pierre angulaire de nos interactions en face à face. Il nous paraît intéressant de remarquer que, s'inspirant directement de Goffman, un certain nombre d'auteurs tendent aujourd'hui à adopter un modèle de l'interaction conversationnelle dans lequel ce genre d'échange tient une place essentielle. C'est le cas de Brown et Levinson [42], ou bien encore de Owen [187]. Nous n'avons pas ici l'ambition de discuter du détail de leur modèle. Nous souhaitons simplement montrer en quoi il peut être parfois utile de confronter deux champs d'étude jusque-là considérés comme indépendants. D'un côté, en règle générale, les auteurs auxquels nous venons de faire allusion ambitionnent de clarifier la logique de nos interactions quotidiennes. De l'autre côté, ce que la logique déontique nous enseigne, c'est que (du point de vue logique) la notion d'activité réparatrice ne va pas de soi.

Prenant pour fil directeur l'étude que, au chapitre 4 des *Relations en Public*, Goffman consacre à l'échange réparateur, nous avons choisi d'introduire pas à pas les principales composantes de son analyse.

3.4.2 L'obligation postdatée

Goffman fait l'hypothèse que les réserves de type territorial constituent la revendication principale des individus en interaction. Telle que décrite au chapitre 2 des *Relations en Public*, la notion de « territoire du moi » est prise en un sens relativement large. Celle-ci désigne à la fois :

– l'espace personnel (la portion d'espace qui entoure un individu) ;
– la place (espace bien délimité auquel l'individu peut avoir droit temporairement et dont la possession est basée sur le principe du « tout ou rien ») ;
– le territoire de la possession (tout ensemble d'objets identifiables au moi) ;
– les réserves d'information (l'ensemble de faits qui concernent l'individu et dont celui-ci entend contrôler l'accès lorsqu'il se trouve en présence d'autrui) ;
– les domaines réservés de la conversation (le droit qu'a l'individu d'exercer un certain contrôle sur qui peut lui adresser la parole et quand).

Pendant le déroulement d'une interaction, les participants sont amenés à accomplir un certain nombre d'actes, qui (pour la plupart) constituent des menaces d'intrusion (ou offense) territoriale[32]. Telle est la fonction de l'échange réparateur, que d'en neutraliser les effets négatifs survenus dans l'interaction. L'activité réparatrice, nous dit Goffman [93, chap. 4], revêt trois formes principales : la justification ; l'excuse ; la prière. Nous ne parlerons ici que des deux dernières. Voici un exemple :

> *A* marche sur les pieds de *B*
> *A* : – Excusez-moi.
> *B* : – Pas de quoi.

Mettons provisoirement entre parenthèses la réaction de *B*, permettant de signaler à *A* que l'incident est clos et que l'équilibre rituel est restauré. Oublions jusqu'aux excuses que *A* présente à *B*. L'exemple met visiblement en jeu deux énoncés normatifs que, en première approximation, nous écrirons :

(i) il est obligatoire que $\neg o$ (o = « offense ») ;
(ii) si o alors obligatoirement e (e = « excuse »).

[32] Pour un premier inventaire des actes de langage qui impliquent une intrusion territoriale, voir Brown et Levinson [42, § 3.2]

Nous ajoutons les données 'factuelles' suivantes :

(iii) e entraîne nécessairement (présuppose qu'il y ait eu antérieurement) o ;
(iv) o.

Si nous recourons à la formalisation que Chisholm lui-même proposait, nous obtenons :

$$\text{SHALL}\neg o \quad (1a)$$
$$o \rightarrow \text{ SHALL } e \quad (1b)$$
$$\Box(e \rightarrow o) \quad (1c)$$
$$o \quad (1d)$$

Nous ne voulons pas que (i) entraîne (ii). C'est pourquoi (1b) enchâsse l'opérateur déontique dans l'opérateur d'implication matérielle, et non l'inverse[33]. Une fois ceci accepté, nous voyons aussitôt la nature de la difficulté. D'une part, à partir de (1b) et (1d) on dérive par application du modus-ponens

$$\text{SHALL } e. \quad (1e)$$

D'autre part, (1c) entraîne logiquement $\Box(\neg o \rightarrow \neg e)$. Notons $\models$ la relation de conséquence sémantique dite "locale", définie par quantification universelle sur les mondes possibles. On a :

$$\Box(e \rightarrow o) \models \Box(\neg o \rightarrow \neg e), \quad \text{(Contraposition)}$$

D'autre part, dans une logique déontique standard, l'obligation est fermée sous l'implication stricte, de sorte qu'on peut toujours « affaiblir » le contenu d'une obligation inconditionnelle. En d'autres termes, on a :

$$\Box(\phi \rightarrow \psi), \text{SHALL}\phi \models \text{SHALL}\psi \quad \text{(Affaiblissement)}$$

A partir de (1a) et (1c), on dérive donc aussi :

[33] Supposons que, à la place de (1b), nous ayons un énoncé du type (1b') SHALL $(o \rightarrow e)$. Dans une logique déontique standard, SHALL$\neg o$ entraîne logiquement SHALL $(o \rightarrow e)$ – ce qui, intuitivement, est choquant.

Ceci est l'argument traditionnellement avancé afi n de justifi er le recours à (1b) plutôt qu'à (1b'). Cet argument ne nous semble pas entièrement convaincant. Rien ne nous empêche de nous placer dans le cas de fi gure où l'agent se conforme à son devoir initial (1a). Ceci reviendrait à remplacer la quatrième et dernière prémisse par la nouvelle prémisse – nommons-là (1d') – stipulant que $\neg o$ est le cas. La norme réparatrice (1b) compte au nombre des conséquences logiques de (1d') – ce qui, intuitivement, est également choquant. L'idée de Goble [91] de partir d'une logique de la relevance nous paraît ici séduisante.

$$\dfrac{\text{SHALL}\neg o \qquad \dfrac{\Box(e \to o)}{\Box(\neg o \to \neg e)}}{\underbrace{\text{SHALL}\neg e}_{(1f)}}$$

Ainsi, à partir de nos normes initiales, nous déduisons deux normes catégoriques qui se révèlent antithétiques.

Les logiques temporelles résolvent assez naturellement cette tension, en distinguant un avant et un après la violation. En gros, lorsque la violation est commise, l'obligation réparatrice prend le pas. Voici une deuxième formalisation ; elle est directement inspirée d'Åqvist [13]. Comme précédemment, le foncteur $\oplus$ est lu « à l'instant prochain ». L'opérateur $\Box$ exprime maintenant la nécessité historique. Le quadruplet (1a)-(1d) devient :

$$\text{SHALL} \oplus \neg o \tag{2a}$$
$$\oplus (o \to \text{ SHALL } \oplus e) \tag{2b}$$
$$\Box((\oplus \oplus e) \to \oplus o) \tag{2c}$$
$$\oplus o \tag{2d}$$

Nous nous plaçons ici juste avant la violation. Jouant sur les propriétés de l'opérateur $\oplus$ et usant d'une argumentation similaire à la précédente, à partir de (2a) et (2c), on dérive

$$\dfrac{\text{SHALL} \oplus \neg o \qquad \dfrac{\Box((\oplus \oplus e) \to \oplus o)}{\Box((\oplus \neg o) \to \oplus \oplus \neg e)}}{\underbrace{\text{SHALL} \oplus \oplus \neg e,}_{(2e)}}$$

conclusion qu'on lira :

Il est maintenant obligatoire que non-e à l'instant post-prochain.

A partir de (2b) et (2d), on dérive (par distribution de $\oplus$ sur $\to$ puis modus-ponens)

$$\dfrac{\oplus o \qquad \dfrac{\oplus(o \to \text{ SHALL } \oplus e)}{(\oplus o \to \oplus \text{ SHALL } \oplus e)}}{\underbrace{\oplus \text{SHALL} \oplus e,}_{(2f)}}$$

proposition qu'on lira :

A l'instant prochain il sera obligatoire que ensuite e.

La différence entre les représentations (2e) et (2f) tient à l'ordre dans lequel les opérateurs sont emboîtés. Pour en éclairer la signification, nous pouvons nous reporter au schéma ci-dessous. Dans ce schéma, nous supposons que l'opérateur monadique d'obligation SHALLϕ est pris pour une abréviation de $\phi \gg \neg\phi$, et que l'opérateur $\gg$ à son tour est pris au sens de la définition 7[34]. Voici un modèle typique de nos prémisses (2a)-(2d) :

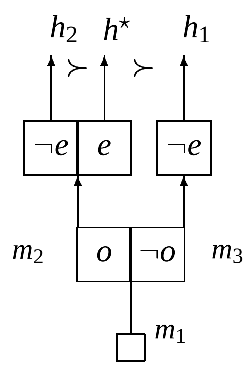


FIG. 3.2 – Scénario de l'excuse (compte rendu temporel)

Ici, $h^\star$ désigne l'évolution réelle de la relation en face à face. La relation $\succ$ compare les différents scénarios envisageables. Par exemple, $h_1 \prec h^\star$ précise que la séquence h_1 est déontiquement préférable à la séquence $h^\star$. Le moment m_1 désigne le moment présent. L'intrusion territoriale n'a pas encore été commise. Au moment m_1 de l'histoire $h^\star$, nous avons

$$m_1/h^\star \models \text{SHALL} \oplus \oplus \neg e.$$

En effet, $h_1 \prec h^\star$. Au moment m_2 de l'histoire $h^\star$, nous avons

$$m_2/h^\star \models \text{SHALL} \oplus e.$$

D'une part, $h^\star \prec h_2$. D'autre part, comme h_1 ne passe pas par m_2 ($h_1 \notin H_{m_2}$), nous ne la prenons plus en considération[35]. Cela implique à son tour que, au moment m_1 de l'histoire $h^\star$, nous avons :

$$m_1/h^\star \models \oplus\text{SHALL} \oplus e.$$

En ce sens, une fois que la violation est commise (instant prochain), l'obligation réparatrice prend le dessus.

[34] Voir supra page 137.

[35] En accord avec nos définitions 7 et 8, données aux pages (respectivement) 137 et 141.

3.4.3 La prière

Nous notons ici que, tentant de développer une approche exclusivement préférentielle, certains logiciens déontiques [205, 206] multiplient les exemples dans lesquels la date à laquelle l'obligation est violée se confond avec celle à laquelle l'obligation réparatrice est ou n'est pas satisfaite. Aussi, nous pouvons nous demander s'il fut essentiel à l'argument que les actions o et e soient exécutées l'une à la suite de l'autre, plutôt que parallèlement l'une à l'autre. Ceci nous amène à cette autre forme de l'activité réparatrice qu'est, selon Goffman, la prière (*request*). Cette dernière

> « consiste à demander à l'offensé [ ] la permission de se livrer à ce qu'il pourrait considérer comme une violation de ses droits » [93, p. 117].

Or,

> « s'il est caractéristique de [ ] voir arriver [les excuses] après l'évènement, [ ] les prières [ ] se placent typiquement avant l'évènement suspect, ou, au plus tard, au cours de ses premières phases » [93, p. 117].

Exemple : s'apprêtant à remplir un chèque, et apercevant le stylo de tel ou tel près de lui, un individu donné se saisit aussitôt du stylo, tout en demandant :

> *Puis-je vous emprunter votre stylo ?*

Pour notre part, nous pensons que le précédent schéma d'analyse continue à s'appliquer. Ceci ressortira mieux, si nous restons dans le cadre d'une logique déontico-temporelle prenant appui sur la logique classique. Désignons par la lettre p la réparation par la prière. Nos énoncés de départ sont : il est obligatoire que $\neg o$; si o alors obligatoirement p ; p entraîne (présuppose qu'il y ait) o ; o. Nous supposons ici que les actions o et p sont accomplies en parallèle. Transcription :

$$\text{SHALL} \oplus \neg o \qquad (3a)$$

$$\oplus (o \rightarrow \text{SHALL}\, p) \qquad (3b)$$

$$\Box \oplus (p \rightarrow o) \qquad (3c)$$

$$\oplus o \qquad (3d)$$

De (3a) et de (3c), il vient :

$$\text{SHALL} \oplus \neg p. \qquad (3e)$$

De (3b) et de (3d), il vient :

$$\oplus\ \text{SHALL}\, p. \qquad (3f)$$

Qui accepte cette analyse logique dira que c'est seulement au moment de la violation (ou, si l'on préfère, au cours de ses premières phases) que l'obligation catégorique de prier est détachée et qu'elle conduit à l'action. En résumé, le fait que l'offense et la réparation soient accomplies en parallèle l'une de l'autre n'interdit pas de raisonner en termes d'un avant et d'un après l'intrusion. Et le fait que certains logiciens multiplient les exemples de scénarios en parallèle ne doit pas nous égarer.

En résumé, dans le cas de l'excuse comme de la prière, le nerf de l'argument consiste à (dirons-nous) *postdater* la force prescriptive de la norme réparatrice. Plus précisément, elle consiste à supposer que la date à laquelle l'obligation réparatrice entre en vigueur coïncide avec l'évènement suspect. Il faut reconnaître que cette position semble parfaitement défendable. Elle témoigne même d'une certaine rigueur. Parce qu'il se peut qu'il ne commette pas d'intrusion, l'offenseur virtuel ne se considérera pas dans l'obligation de réparer, tant que l'intrusion n'a pas effectivement eu lieu. Dire ce genre de choses revient peu ou prou à supposer que les conditions de vérité d'un énoncé sont relatives non pas seulement au présent, mais aussi à ce que le futur est censé être à la lumière de l'information dont nous disposons à ce moment-là. De ce point de vue-là, nous avions finalement raison de travailler avec des règles d'évaluation ayant la forme suivante :

$$m/h \models \phi : \text{« la formule } \phi \text{ est vraie au moment } m \text{ de l'histoire } h \text{ ».}$$

Néanmoins, une question se pose ici. Nous nous souvenons que, dans le cas de la prière, la réparation peut précéder l'évènement suspect. Nous voyons difficilement comment intégrer cette éventualité à l'analyse, si nous supposons que la date à laquelle l'obligation réparatrice prend effet ne précède jamais l'évènement suspect ? Pour illustrer la difficulté, recourons à l'opérateur « miroir » de $\oplus$, noté $\ominus$ et lu *Il s'est trouvé à l'instant immédiatement passé que*. Envisageant le cas de figure où l'offense et la prière étaient accomplies en parallèle l'une de l'autre, nous écrivions ceci :

$$\mathrm{SHALL} \oplus \neg o \tag{3a}$$

$$\oplus (o \to \mathrm{SHALL}\, p) \tag{3b}$$

$$\Box \oplus (p \to o) \tag{3c}$$

$$\oplus o \tag{3d}$$

Comment reformuler ces énoncés, lorsque la prière précède l'offense ? Visiblement, il n'est nécessaire ni de modifier (3a) ni de modifier (3d). Pour ce qui est de

(3c), il suffit visiblement de placer $\ominus$ devant la lettre p. Ceci donnerait déjà :

$$\begin{array}{ll}
\mathrm{SHALL}\oplus\neg o & \text{(3a)} \\
- & \text{(3b)} \\
\Box\oplus(\ominus p \to o) & \text{(3c')} \\
\oplus o & \text{(3d)}
\end{array}$$

Faisant appel à des lois familières en logique temporelle, nous pouvons simplifier (3c'), et écrire :

$$\begin{array}{ll}
\mathrm{SHALL}\oplus\neg o & \text{(3a)} \\
- & \text{(3b)} \\
\Box(p \to \oplus o) & \text{(3c')} \\
\oplus o & \text{(3d)}
\end{array}$$

On montre que, à partir de (3a) et (3c'), on obtient toujours la conclusion voulue :

$$\dfrac{\mathrm{SHALL}\oplus\neg o \qquad \dfrac{\dfrac{\Box(p \to \oplus o)}{\Box(\neg\oplus o \to \neg p)}}{\Box(\oplus\neg o \to \neg p)}}{\underbrace{\mathrm{SHALL}\neg p,}_{\text{(3e')}}}$$

Il reste à reformuler la prémisse (3b). Ici apparaît une difficulté. Une première possibilité évidente consiste à écrire :

$$\oplus(o \to (\mathrm{SHALL}\ominus p)) \qquad \text{(3b')}$$

Mais voici ce que nous détachons à partir de (3b') et de (3d) :

$$\oplus\,\mathrm{SHALL}\ominus p \qquad \text{(3f')}$$

Lecture :

> *A l'instant prochain il sera obligatoire que : p à l'instant immédiatement passé.*

(3f') n'a visiblement pas de sens. Dans (3f'), le contenu de la norme porte (au moment où la norme prend effet) sur un évènement passé, qu'il n'est plus au pouvoir de l'interactant de modifier[36]. La traduction (3f') est problématique, parce qu'elle emboîte $\ominus$ dans SHALL. Nous pourrions donc essayer l'emboîtement inverse, et remplacer (3b') par :

$$\oplus(o \rightarrow (\ominus \text{SHALL } p)) \qquad (3b").$$

A partir de (3b") et de (3d), nous détacherions

$$\oplus \ominus \text{SHALL} p$$

qui, en règle générale, se révèle logiquement équivalent à

$$\text{SHALL} p \qquad (3f")$$

Nous retombons sur la difficulté de départ. (3f") peut difficilement coexister avec (3e') que nous avons déjà.

Que l'approche temporelle classique ait des difficultés à traiter le cas de figure où la réparation précède l'offense a parfois été dénoncé [102, 258]. Ici, ces difficultés prennent la forme d'un problème concernant l'enchâssement d'opérateurs modaux de nature distincte. Une piste possible consisterait à essayer d'utiliser une logique temporelle pluri-dimensionnelle. Conçu pour l'étude linguistique des temps, ce type de formalisme exploite la possibilité de distinguer trois temps : celui du locuteur, celui de l'évènement décrit, et un temps de référence. Au paragraphe (3.4.7), nous étudierons une autre piste possible, qui nous semble plus simple. Celle-ci présuppose le recours à des outils d'analyse empruntés à la littérature sur les conditionnels. Les paragraphes (3.4.4) à (3.4.6) décrivent ces outils.

3.4.4 Principe de la nouvelle analyse

Ne souhaitant sacrifier aucune des deux normes inconditionnelles, certains auteurs ont eu l'idée de distinguer deux sens du mot « obligation » (voir e.g. [43]). Certains même ont suggéré de rendre compte de cette distinction dans les termes d'une sémantique préférentielle (voir [50]). Nous nous tournons vers une autre piste envisageable, celle de la révision itérée. Aussi bien, les difficultés liées à la notion d'échange réparateur montrent seulement que la prise en compte d'une donnée factuelle (une violation) a une incidence directe sur le pré-ordre associé aux prémisses de départ. Nous allons essayer de tester cette hypothèse, en nous aidant de la théorie de la révision naturelle de Boutilier [37]. Comme nous allons le voir, le principe en est relativement simple, du moins pour l'usage que nous en

[36]Il nous paraît ici intéressant de remarquer que, dans le système de Chellas [46, p. 127], un impératif de la forme SHALL $\ominus p$ se révèle équivalent à $\ominus p$.

ferons. En gros, lorsque l'intrusion territoriale est commise, les interactants permutent l'état le meilleur possible et le « second état » le meilleur possible pour la poursuite de l'interaction. Ils opèrent cet ajustement tout en modifiant au minimum l'ancien pré-ordre, afin de mener le cycle réparateur jusqu'à son terme. Nous appellerons cela l'effet « permutation ». Certaines études sur les conditionnels et sur la révision itérée[37] ont révélé que, dans le domaine ontique, la prise en compte d'une information nouvelle pouvait éventuellement avoir des effets autres que celui-là. Nous ne présupposerons ici aucune familiarité avec ces études.

Le type de compte rendu que nous allons esquisser s'apparente au compte rendu temporel, dans la mesure où il nous faut distinguer un avant et un après la violation. Quoique l'analyse contienne du séquentiel, nous n'avons pas besoin d'introduire explicitement cet élément dans le formalisme, au moins dans un premier temps. Nous nous plaçons donc désormais dans le cadre d'une sémantique préférentielle ordinaire. La plupart des notations que nous allons désormais utiliser sont empruntées aux sections (2.2.4) et (2.2.5). Nous nous permettons d'y renvoyer le lecteur qui hésiterait sur la signification à donner à un symbole.

3.4.5 Un aparté sur la condition de Ramsey

Nous ouvrons ici un aparté. Ce que recouvre exactement l'idée de fonder la notion d'obligation sur celle de révision n'est pas tout à fait transparent. Considérons le test de Ramsey, à la lumière duquel les théoriciens de la révision ont l'habitude d'éclairer la procédure d'évaluation d'un conditionnel contraire-aux-faits (*contrary-to-fact*) :

$$\phi > \psi \in K \text{ ssi } \psi \in K^{\star}\phi \tag{>-RT}$$

(>-RT) stipule que, pour savoir si la proposition conditionnelle $\phi > \psi$ est acceptable dans un état de croyance donné K, il faut regarder si le changement minimal de K nécessaire pour y introduire ϕ requiert aussi l'acceptation de ψ. Appliqué à une norme réparatrice (*contrary-to-duty*), le test devient :

$$\bigcirc(\psi/\phi) \in K \text{ ssi } \psi \in K^{\star}\phi$$

Tel quel, ce test semble peu intelligible, à moins de supposer que, d'une occurrence à l'autre, K change de nature. A gauche du biconditionnel, il désigne ce que nous croyons être vrai dans le monde réel. A droite du biconditionnel, il désigne plutôt (pour parler comme Lewis) une sphère d' « obligativité ». Aux frontières délimitées par les obligations non-réparatrices vérifiées dans le monde réel, cette

[37] Nous faisons ici allusion à la théorie des fonctions ordinales conditionnelles de Spohn [240], la révision rangée de Lehmann [135], la théorie des transmutations de Williams [268] et l'approche de Darwiche & Pearl [54].

zone est la seule avec laquelle ϕ ne soit pas compatible. Il nous faudrait donc concevoir un test de Ramsey en deux temps :

$$\bigcirc(\psi/\phi) \in K \text{ ssi } \psi \in (K^{\star}\top)^{\star}\phi \qquad (\bigcirc\text{-RT})$$

C'est la précarité essentielle à toute norme, de quelque nature qu'elle soit, qui nous oblige à réviser tout d'abord K. Ceci nous donne l'opération « $K^{\star}\top$ ». Nous appellerons celle-ci *idéalisation*. La seconde révision, celle de $K^{\star}\top$ par ϕ, nous l'appellerons (pour reprendre le terme consacré) *sous-idéalisation*. Elle consiste à effectuer sur la sphère d'obligativité les changements minimaux pour y introduire le contenu d'une interdiction. Lorsque, dans le cas de l'idéalisation, nous parlons de révision, nous prenons ce terme en un sens quelque peu approximatif. Personne ne doutera du fait qu'un règne des fins kantien (monde dans lequel l'agent se conformerait à toutes ses obligations) soit hors de portée. Aussi, notre test de Ramsey tournera à vide, à moins de recourir à (quelque chose comme) l'approche dite *limite*. Celle-ci exige que nous gardions un oeil non pas sur les mondes les plus parfaits – qui peut-être n'existent pas – mais sur les mondes « suffisamment » parfaits. Par simplicité, nous n'utiliserons pas une analyse de ce type, quoiqu'elle paraisse beaucoup plus satisfaisante. Le lecteur trouvera dans Schlechta [226] un traitement complet de cette approche. En substance, l'idée est de travailler avec des segments initiaux δ_X. Etant donné un ensemble X, $\delta_X \subseteq X$ est appelé un segment initial si, d'une part, tout élément de X est minimisé par un élément de δ_X

$$\forall w \in X \;\; \exists w' \in \delta_X \left(w' = w \vee w' \prec w\right) \qquad (\delta 1)$$

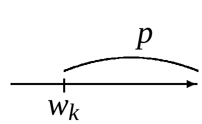
et si, d'autre part, δ_X contient tous les minorants de ses éléments

$$\forall w \in X \;\; \forall w' \in \delta_X \left(w \prec w' \to w \in \delta_X\right). \qquad (\delta 2)$$

Etant donné un modèle $\mathcal{M}$, nous pourrions alors poser (quelque chose comme) :

$$K^{\star}\phi^{\;\mathcal{M}} = \{\psi : \exists\delta_{[\phi]_{\mathcal{M}}} \text{ tel que } (\forall w \in \delta_{[\phi]_{\mathcal{M}}})\ (w \models_{\mathcal{M}} \psi)\} \qquad (3.22)$$

Intuitivement : le résultat de la révision de K par ϕ donne l'ensemble des ψ pour lesquels il existe un segment initial de ϕ-mondes dont les éléments vérifient ψ. L'exemple suivant nous convaincra de l'utilité de cette variation. Supposons que l'univers W du modèle $\mathcal{M}$ considéré soit $\{w_i : i < \omega\}$. Posons : $w_j \prec w_i$ ssi $j > i$. Imaginons que $\exists k$ tel que $(\forall l)(k \le l < \omega \to w_l \models_{\mathcal{M}} p$:



Dans ce cas de figure, la procédure tourne à vide, si l'on fait appel à la définition traditionnelle en termes de minimalité. Aucun monde n'est plus parfait que tous les autres, si bien que p et $\neg p$ apparaissent tous les deux dans la sphère d'obligativité. Avec (3.22), seul p y figure.

Pour justifier notre démarche (comprendre l'obligation comme une révision), nous avons évoqué la possibilité de recourir à un test de Ramsey fondé sur le mouvement (idéalisation, sous-idéalisation). Nous avons aussi suggéré l'idée d'analyser ces mouvements dans les termes de l'approche limite. L'étude d'un tel test de Ramsey reste sans doute à faire. Nous devons la remettre à plus tard. Soulignons simplement que, sous certaines conditions facilement identifiables, la procédure donne de l'obligation inconditionnelle et de l'obligation conditionnelle conforme-au-devoir (*according-to-duty*) un compte rendu tout à fait traditionnel. Ceci, croyons-nous, augmente l'attrait du test ici proposé. Voyons tout d'abord le cas de l'obligation inconditionnelle. Généralement, on l'introduit par la définition $\bigcirc\phi \stackrel{\text{def}}{\longleftrightarrow} \bigcirc(\phi/\top)$. Généralement aussi, l'opération de révision ne donne pas un ensemble inconsistant de formules, et des révisions par des énoncés logiquement équivalents donnent des résultats identiques. Pour que, appliqué à une norme inconditionnelle, ($\bigcirc$-RT) se laisse immédiatement 'simplifier' en

$$\bigcirc\phi \in K \text{ ssi } \phi \in K^\star\top,$$

il suffit visiblement que la révision itérée vérifie la propriété :

$$\text{si } \neg\phi' \notin K^\star\phi \text{ alors } (K^\star\phi)^\star\phi' = K^\star(\phi \wedge \phi'). \tag{3.23}$$

Intuitivement, (3.23) dit que, si l'input d'une nième révision n'est pas inconsistant avec l'output de la révision à laquelle elle fait suite, alors cette nième révision n'apporte intrinsèquement aucun élément nouveau ; réviser directement par $\phi \wedge \phi'$ donne le même résultat[38]. De la sorte :

$$\begin{aligned} \bigcirc\phi \in K \text{ ssi } & \phi \in (K^\star\top)^\star\top && \text{par } (\bigcirc\text{-RT}) \\ \text{ssi } & \phi \in K^\star(\top \wedge \top) && \text{par (3.23) si } \bot \notin K^\star\top \\ \text{ssi } & \phi \in K^\star\top \end{aligned}$$

Une remarque similaire s'applique à l'obligation conditionnelle conforme-au-devoir. Soit ϕ son antécédent. Intuitivement, nous sommes dans le cas de figure où $\phi \in$

[38]Cette propriété doit-elle être rapprochée de la condition AGM de préservation, i.e.

$$\text{Si } \neg\phi \notin K \text{ alors } K \subseteq K^\star\phi, \tag{3.24}$$

qui nous recommande de laisser K tel quel, si l'information introduite ne génère aucune contradiction ?

$K^\star\top$, de sorte que $\neg\phi \notin K^\star\top$. Ainsi, appliqué à une norme conforme-au-devoir, ($\bigcirc$-RT) se laisse maintenant 'simplifier' en :

$$\begin{aligned}\bigcirc(\psi/\phi) \in K \text{ ssi } & \psi \in (K^\star\top)^\star\phi && \text{par } (\bigcirc\text{-RT}) \\ \text{ssi } & \psi \in K^\star(\top \wedge \phi) && \text{par (3.23) si } \neg\phi \notin K^\star\top \\ \text{ssi } & \psi \in K^\star\phi &&\end{aligned}$$

Il nous paraît intéressant de remarquer que la procédure de révision itérée que, dans le prochain paragraphe, nous allons utiliser vérifie cette propriété[39].

3.4.6 L'effet « permutation »

Fermons cet aparté sur le test de Ramsey, et revenons à l'échange réparateur. Jusqu'à présent, nous avons fait comme si l'échange s'achevait avec la réparation. De fait, observe Goffman, il contient généralement trois mouvements supplémentaires. Ici, l'intérêt se déplace de la violation de la norme à la façon dont les interactants traitent l'infraction. Une fois l'offense réparée, la victime signale que la réparation est suffisante − mouvement dit de satisfaction (*relief*). Une fois ceci signalé, l'offenseur se trouve placé dans l'obligation de manifester de la gratitude − mouvement dit d'appréciation. Après quoi, l'offensé manifeste une appréciation de l'appréciation qu'on lui manifeste − minimisation. Le plus souvent, ceci mène l'échange à sa fin : l'équilibre territorial a été restauré. Exemple :

réparation	A :	« Puis-je me servir de votre téléphone pour appeler en ville ? »
satisfaction	B :	« Bien sûr, allez-y. »
appréciation	A :	« C'est très aimable de votre part. »
minimisation	B :	« Ce n'est rien. » (d'après Goffman [93, p. 142])

Dans cet exemple, la réparation est du type prière, et elle précède l'offense. Néanmoins, dans un premier temps, nous ferons abstraction de toute considération temporelle, afin que le principe de notre analyse apparaisse plus clairement. Nous devons insister ici sur le fait que, selon Goffman,

> « un trait fondamental du rituel réparateur est [ ] qu'il constitue une façon *obligatoire* de traiter les déviations occasionnelles, un moyen de maintenir les accords courants face à des infractions sans malice. » [93, p. 161]

Si cette observation est exacte, alors tenter de rendre compte du cycle réparateur complet en mettant au premier plan les normes qui sous-tendent l'échange ne nous paraît pas déphasé.

[39] Voir Boutilier [37, Théorème 14].

Voici les données dont nous disposons, avant que le cycle réparateur ne démarre :

	Normes	Données factuelles
(I)	$\bigcirc\neg o$	
(II)	$\bigcirc(p/o)$	$\Box(p \to o)$
	$\bigcirc(s/o \wedge p)$	$\Box(s \to (o \wedge p))$
(III)	$\bigcirc(a/o \wedge p \wedge s)$	$\Box(a \to (o \wedge p \wedge s))$
	$\bigcirc(m/o \wedge p \wedge s \wedge a)$	$\Box(m \to (o \wedge p \wedge s \wedge a))$

Le groupe (II) d'énoncés correspond à l'enchaînement (réparation, satisfaction), et le groupe (III) correspond à l'enchaînement (appréciation, minimisation). Dans la colonne droite du tableau, nous avons placé une série de lois qui nous paraissent plausibles et qui réduisent le nombre de reprises possibles. Pour les besoins de l'analyse, nous utiliserons l'« appareillage » sémantique du système **P**. Pour plus de clarté, nous en rappelons les principales composantes. Un modèle est défini comme un triplet $\mathcal{M} = (W, \iota, \prec)$ où :

1. W désigne un ensemble de mondes possibles $w, w', \ldots$;
2. ι est une fonction d'évaluation associant à chaque atome propositionnel l'ensemble des mondes dans lequel il est vrai ;
3. $\prec \subseteq W \times W$ est une relation d'ordre strict (transitive et irréflexive) ; intuitivement, $w \prec w'$ signifie « w est plus parfait que w' ».

Soit $[\phi]_{\mathcal{M}}$ l'ensemble des mondes de $\mathcal{M}$ qui vérifient ϕ. Et soit $\min_{\mathcal{M}}(\phi)$ le sous-ensemble de ceux qui sont minimaux sous $\prec$, i.e.

$$\min_{\mathcal{M}}(\phi) \;=\; \{w \in [\phi]_{\mathcal{M}} \mid \neg\exists w'(w' \in [\phi]_{\mathcal{M}} \,\&\, w' \prec w)\}.$$

Intuitivement, $\min_{\mathcal{M}}(\phi)$ énumère les ϕ-mondes les plus parfaits. Les conditions de récurrence pour l'obligation conditionnelle sont définies relativement à un modèle. Nous avons :

$$\mathcal{M} \models \bigcirc(\psi/\phi) \quad \Leftrightarrow \quad \min_{\mathcal{M}}(\phi) \subseteq [\psi]_{\mathcal{M}}. \tag{3.25}$$

Lorsque $\mathcal{M} \models \bigcirc(\psi/\phi)$ se vérifie, on dit de $\mathcal{M}$ qu'il est un modèle de l'obligation conditionnelle $\bigcirc(\psi/\phi)$. Intuitivement, la clause (3.25) dit que $\mathcal{M}$ est un modèle de $\bigcirc(\psi/\phi)$ si et seulement si les plus parfaits des ϕ-mondes que l'univers de $\mathcal{M}$ contient vérifient tous ψ. Enfin, nous conformant à l'usage, nous considérerons une norme inconditionnelle $\bigcirc\phi$ comme une abréviation de $\bigcirc(\phi/\top)$.

Revenons maintenant à notre exemple. Soit $\mathcal{M}$ un modèle de nos prémisses. Typiquement l'univers W de $\mathcal{M}$ contiendra les éléments suivants :

$$\begin{array}{ll}
w_1 : \neg o, \neg p, \neg s, \neg a, \neg m & w_2 : o, p, s, a, m \\
w_3 : o, p, s, a, \neg m & w_4 : o, p, s, \neg a, \neg m \\
w_5 : o, p, \neg s, \neg a, \neg m & w_6 : o, \neg p, \neg s, \neg a, \neg m.
\end{array}$$

L'ensemble $w_1 - w_6$ épuise le champ des possibilités, compte tenu de nos « données factuelles ». Voici comment disposer les éléments de W :

$$\begin{tikzcd}[row sep=small]
\bullet w_i \text{ (pour tout } i = 3, \ldots, 6) \arrow[d] \\
\bullet w_2 \arrow[d] \\
\bullet w_1
\end{tikzcd}$$

FIG. 3.3 – Pré-ordre avant l'offense (modèle $\mathcal{M}$)

Dans ce diagramme, les flèches $w\bullet \leftarrow \bullet w'$ dénotent la relation $w \prec w'$ (lecture : w minimise w'). $\mathcal{M}_1$ satisfait l'obligation initiale $\bigcirc\neg o$, puisque tous les mondes gravitent vers w_1, qui désigne l'état le meilleur (*best*) possible pour la poursuite de l'interaction. $\mathcal{M}_1$ satisfait également les prémisses normatives restantes, puisque w_2 (le « deuxième état » le meilleur possible – *2nd best*) minimise w_3, w_4, w_5 et w_6. D'un point de vue formel, nous n'avons pas besoin d'affiner d'avantage l'ordonnancement à l'intérieur de l'ensemble $w_3 - w_6$. Pour plus de simplicité, nous supposerons que, à l'intérieur de cet ensemble, aucun monde ne minimise l'autre[40].

Le modèle $\mathcal{M}$ décrit ce qui se passe avant l'intrusion territoriale. Nous voyons que, lors de cette phase, nous n'avons pas

$$\mathcal{M} \models \bigcirc p,$$

mais plutôt

$$\mathcal{M} \models \bigcirc\neg p.$$

Ceci correspond à l'étape où, nous plaçant dans le cadre d'une logique temporelle, nous détachions (3e) à partir de (3a) et (3c). Une remarque similaire s'applique aux actions *s*, *a* et *m*. Aucune n'est inconditionnellement obligatoire.

[40] Intuitivement, cela n'est pas très satisfaisant. En particulier, Prakken et Sergot défendent l'idée apparemment plausible selon laquelle "it is better to fulfil an obligation from a more ideal context and violate one from a less ideal context than the other around" [206]. Nous devrions donc avoir $w_3 \prec w_4 \prec w_5 \prec w_6$. Le fait est que, comme les auteurs l'observent, dans la sémantique initiale de Hansson-Lewis, rien n'exige un tel ordonnancement. La stratégie pour laquelle Prakken et Sergot optent consiste à utiliser la règle lexicographique de Ryan [217]. Nous n'introduirons pas cette règle, afin que le principe de notre nouvelle analyse ressorte plus clairement.

Essayons maintenant d'analyser ce qui se passe au moment de l'offense. Notre hypothèse de travail est que, en présence de cette nouvelle donnée, les interactants réajustent le pré-ordre associé aux obligations initiales. Nous passons ainsi (pour utiliser un vocabulaire fréquemment utilisé) à une structure « sous-idéale ». Soit $\mathcal{M} = \langle W, \iota, \prec \rangle$ notre modèle de départ. Notons $\mathcal{M}^\star\phi$ le modèle obtenu après 'révision' (ou sous-idéalisation) par ϕ. Typiquement, $\mathcal{M}^\star\phi$ sera de la forme $\langle W, \iota, \prec' \rangle$, où $\prec'$ désigne le pré-ordre obtenu au moyen des deux principes suivants :

(P1) si $w_1 \in \min_{\mathcal{M}}(\phi)$ alors : $w_1 \prec' w_2$ pour tout $w_2 \in W$ et

$$w_2 \prec' w_1 \text{ ssi } w_2 \in \min_{\mathcal{M}}(\phi);$$

(P2) si $w_1, w_2 \notin \min_{\mathcal{M}}(\phi)$ alors : $w_1 \prec' w_2$ ssi $w_1 \prec w_2$.

(P1) signifie que les mondes qui étaient minimaux dans $[\phi]_{\mathcal{M}}$ sont à présent minimaux dans W. (P2) signifie que la disposition des autres mondes possibles ne change pas. Nous empruntons ces deux règles à la théorie de la révision naturelle de Boutilier [37]. Un examen attentif révèle que, au moment de l'intrusion territoriale, ces deux règles (P1)-(P2) amènent les interactants à permuter l'état le meilleur possible et le « second état » le meilleur possible. Pour le vérifier, substituons o à ϕ. Le diagramme 3.4 montre que, pour construire $\mathcal{M}^\star o$, il suffit d'intervertir w_1 et w_2 dans $\mathcal{M}$:

$$\begin{tikzcd}[row sep=small] \bullet w_i \text{ (pour tout } i = 3, \ldots, 6) \arrow[d] \\ \bullet w_1 \arrow[d] \\ \bullet w_2 \end{tikzcd}$$

FIG. 3.4 – Pré-ordre au moment de l'offense (modèle $\mathcal{M}^\star o$)

Le monde w_2 était minimal dans $[o]_{\mathcal{M}_1}$. Donc, par (P1), tous les mondes maintenant gravitent vers w_2. Considérons l'ensemble $w_3 - w_6$. Aucun de ses éléments ne figurait dans $\min_{\mathcal{M}_1}(o)$. Donc, par (P2), à l'intérieur de cet ensemble, l'ordonnancement reste le même. Pour expliquer pourquoi w_1 doit être placé « sous » le bloc $w_3 - w_6$, il suffit de jouer sur (P2).

Il nous paraît intéressant de remarquer que, à cet instant de l'interaction, nous avons

$$\mathcal{M}^\star o \models \bigcirc p,$$

mais pas

$$\mathcal{M}^\star o \models \bigcirc \neg p,$$

Ainsi, faisant permuter l'état le meilleur et le « second état » le meilleur, les interactants forment le jugement inconditionnel adéquat. Ceci correspond à l'étape où, nous plaçant encore dans le cadre d'une logique temporelle, nous déduisions (3f) à partir de (3b) et (3d). Une remarque similaire s'applique aux actions s, a et m. Toutes deviennent catégoriquement obligatoires.

Plusieurs questions viennent à l'esprit, que nous pouvons ici seulement effleurer. Tout d'abord, nous nous demandons ce qui se serait passé si la nouvelle information avait été, non pas o, mais $\neg o$. La réponse semble immédiate. Dans ce cas, il n'y aurait eu aucun changement :

$$\mathcal{M}^\star \neg o = \mathcal{M}.$$

Ensuite, nous voudrions savoir comment la théorie traite les mouvements ultérieurs de l'échange (satisfaction, appréciation et minimisation). La réponse semble elle aussi immédiate ; nous avons apparemment :

$$\mathcal{M}^\star o = (\mathcal{M}^\star o)^\star p = ((\mathcal{M}^\star o)^\star p)^\star s = (((\mathcal{M}^\star o)^\star p)^\star s)^\star a.$$

Enfin, dans un souci de complétude, nous souhaiterions nous assurer que la procédure donne toujours le résultat correct. Il n'est évidemment pas possible de passer en revue toutes les reprises possibles. Contentons-nous de mentionner le scénario intéressant où l'offenseur n'adresse aucune prière. Ce cas de figure nous explique très simplement pourquoi accomplir deux révisions successives ne revient pas toujours à réviser par une conjonction. En effet, pour obtenir $\mathcal{M}^\star(o \wedge \neg p)$, il faut partir de $\mathcal{M}$, et repositionner les mondes de manière à ce que tous gravitent vers w_6, car seul ce monde contient $o \wedge \neg p$. Pour obtenir $(\mathcal{M}^\star o)^\star \neg p$, on part de $\mathcal{M}^\star o$ et on fait en sorte que tous les mondes gravitent vers w_1. On remarque que $(\mathcal{M}^\star o)^\star \neg p \models \bigcirc \neg o$ mais que $\mathcal{M}^\star(o \wedge \neg p) \models \bigcirc o$. Qu'en est-il de la loi (3.23) tout à l'heure évoquée[41] ? Nous voyons qu'elle n'est pas remise en cause, puisque nous ne sommes pas dans le cas de figure où $p \notin K^\star o$.

3.4.7 Extension au temporel

L'objet de ce paragraphe est de réintégrer le précédent schéma d'analyse dans une logique du temps arborescent du type de celle que nous avons initialement esquissée. Cette réintégration nous aidera à traiter le cas où la réparation précéde l'évènement suspect – éventualité que la réparation par la prière nous oblige

[41] Voir supra p. 168.

à considérer. Nous allons travailler ici avec les foncteurs unaires $\oplus$ (*ensuite*) et $\Box$ (*historiquement nécessaire*). Notre foncteur dyadique d'obligation $\bigcirc(/)$ ne sera pas séquentiel. Nous continuons à traiter $\bigcirc\phi$ comme étant synonyme de $\bigcirc(\phi/\top)$.

Du point de vue sémantique, l'extension au temporel se fait tout naturellement. Définissons un modèle comme un triplet de la forme $\mathfrak{M} = ((Tree, <), \prec)$ où :

1. $(Tree, <)$ désigne une structure temporelle arborescente. $Tree$ est un ensemble non-vide d'instants $m_1, m_2, ...$ auquel $<$ (la relation temporelle) donne la forme d'une structure ramifiée vers le futur. Une chaîne maximale h de moments est nommée « histoire ». On note $\mathbf{H}$ l'ensemble de toutes les histoires et H_m le sous-ensemble de celles qui travervent le point m ;
2. ι est une fonction d'évaluation associant à chaque atome propositionnel l'ensemble des moments dans lequel il est vrai ;
3. $\prec \subseteq \mathbf{H} \times \mathbf{H}$ est une relation d'ordre strict (transitive et irréflexive) ; intuitivement, $h \prec h'$ signifie « h est plus parfaite que (minimise) h' ».

Les règles d'évaluation sont de la forme :

$\mathfrak{M}, m/h \models \phi$: « dans $\mathfrak{M}$, la formule ϕ est vraie au moment m de l'histoire h ».

On note $[\phi]_m$ l'ensemble des histoires qui vérifient ϕ en m, et $\min([\phi]_m)$ le sous-ensemble de celles qui sont minimales sous $\prec$, i.e.

$$\min([\phi]_m) \;=\; \{h \in [\phi]_m \mid \neg\exists h'(h' \in [\phi]_m \,\&\, h' \prec h)\}.$$

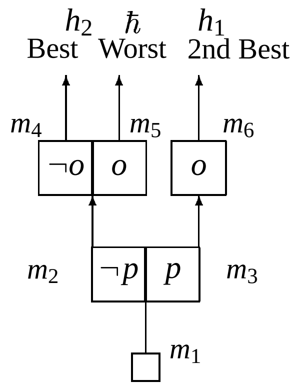
On pose :

$$\mathfrak{M}, m/h \models \oplus\phi \Leftrightarrow \mathfrak{M}, m+1/h \models \phi \qquad \mathfrak{M}, m/h \models \Box\phi \Leftrightarrow (\forall h' \in H_m)(\mathfrak{M}, m/h' \models \phi)$$
$$\mathfrak{M}, m/h \models \bigcirc(\psi/\phi) \Leftrightarrow \min([\phi]_m) \subseteq [\psi]_m.$$

Pour plus de simplicité, nous passons sous silence les mouvements qui succèdent à la prière. Il est tentant de représenter ainsi nos prémisses :

	Normes	Donnée factuelle
(I)	$\bigcirc\oplus\oplus\neg o$	
(II)	$\bigcirc(\oplus p/\oplus\oplus o)$	$\Box(\oplus p \rightarrow \oplus\oplus o)$

Soit, sous forme de diagramme :



$$\hbar = \{m_1, m_2, m_5\}$$
$$h_1 = \{m_1, m_3, m_6\}$$
$$h_2 = \{m_1, m_2, m_4\}$$

FIG. 3.5 – Avant l'anticipation de l'offense (modèle $\mathfrak{M}$)

Dans ce schéma, $\hbar$ désigne l'évolution réelle de l'interaction. Nous avons (la clôture transitive de) $h_2 \prec h_1 \prec \hbar$. Plaçons-nous au point m_1 de $\hbar$. Il est aisé de voir que nos prémisses normatives sont vérifiées en ce point. Nous avons bien :

- $\mathfrak{M}, m_1/\hbar \models \bigcirc \oplus \oplus \neg o$, puisque h_2 minimise les deux autres branches ;
- $\mathfrak{M}, m_1/\hbar \models \bigcirc(\oplus p / \oplus \oplus o)$, puisque h_1 minimise $\hbar$.

A cette étape de l'interaction, où (dirons-nous) l'offense territoriale n'a pas encore été anticipée, nous avons

$$\mathfrak{M}, m_1/\hbar \models \bigcirc \oplus \neg p,$$

et non pas

$$\mathfrak{M}, m_1/\hbar \models \bigcirc \oplus p.$$

A présent, imaginons que les interactants anticipent de concert sur l'offense. Comme précédemment, nous supposerons que l'ajout de cette information a un effet direct sur l'ordonnancement des branches de la structure. Notons $\mathfrak{M}^{\circledast^m}\phi$ le modèle obtenu après 'révision' (au moment m) par ϕ. Désignons par $\prec'$ la relation de préférence qui lui est désormais associée. Posons :

(P1) Si $h \in \min([\phi]_m)$ alors $h \prec' h'$ pour tout $h' \in H_m$ et

$$h' \prec' h \text{ ssi } h' \in \min([\phi]_m);$$

(P2) Si $h, h' \notin \min([\phi]_m)$ alors : $h \prec' h'$ ssi $h \prec h'$.

Substituons $\oplus \oplus o$ à ϕ, et m_1 à m. Dans notre modèle initial, nous avons

$$\min([\oplus \oplus o]_{m_1}) = \{h_1\}.$$

Nous en inférons que $h_1 \prec' \hbar$ et que $h_1 \prec' h_2$ [par P1], et que $h_2 \prec' \hbar$ [par P2]. Qu'est-ce à dire, sinon que le scénario le meilleur et le deuxième scénario le meilleur permutent eux aussi dans ce cas de figure ? Sous forme de diagramme :

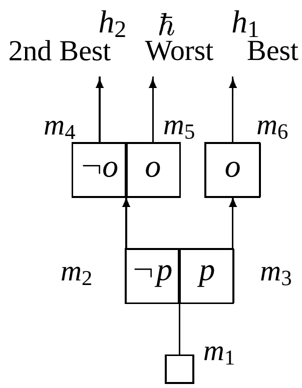


FIG. 3.6 – L'offense anticipée ($\mathfrak{M}^{\circledast^{m_1}} \oplus \oplus o$)

A cette étape de l'interaction, nous avons

$$\mathfrak{M}^{\circledast^{m_1}} \oplus \oplus o, m_1/\hbar \models \bigcirc \oplus p,$$

et non plus

$$\mathfrak{M}^{\circledast^{m_1}} \oplus \oplus o, m_1/\hbar \models \bigcirc \oplus \neg p.$$

Ainsi évitons-nous certaines des obscurités manifestes du compte rendu temporel traditionnel. Par rapport à celui-ci, nous continuons à raisonner en termes d'un avant et d'un après la violation. Néanmoins, nous cessons de « reculer » dans le futur la date à laquelle l'obligation réparatrice entre en vigueur. Cela nous permet de traiter le cas où la réparation précède l'offense.

3.5 Conclusion et recherches à venir

Il est temps de conclure. Ayant constaté que les recherches portant sur l'interface entre le domaine de l'argumentation et celui de la non-monotonie partent généralement dans trois directions distinctes, nous avons finalement essayé de trouver notre voie propre à l'intérieur de la première des trois. Axant notre réflexion sur les sémantiques préférentielles pour l'obligation conditionnelle, nous sommes partis dans deux directions.

Dans un premier temps, nous avons pris pour objet d'étude les sémantiques en question. Nous avons commencé par noter quelques faits saillants les concernant. En particulier, nous avons remarqué qu'elles tendent à se répartir en deux catégories : celles qui prennent l'opérateur dyadique d'obligation comme premier terme non-défini, et celles qui définissent celui-ci à l'aide de l'opérateur monadique d'obligation et d'un conditionnel ontique non-monotone. Dans le premier cas, le principe d'identité et le principe de cumulativité se révèlent tous les deux valides. Dans le second cas, aucun des deux ne l'est. Nous avons aussi remarqué que l'analyse intéressante que propose Makinson relève de la deuxième catégorie. L'originalité de sa sémantique est qu'elle introduit la dimension du futur. Mais un examen attentif révèle que ce n'est pas la prise en compte de cette dimension qui permet l'élimination du principe d'identité. Nous nous sommes alors demandés si une voie intermédiaire n'était pas envisageable, et s'il n'était pas possible d'éliminer le principe d'identité tout en maintenant une forme de cumulativité, en recourant à une logique du changement.

Dans un deuxième temps, nous avons axé notre enquête sur le thème des interactions conversationnelles. Notre propos était ici de montrer que lesdites sémantiques de l'obligation, et la logique déontique en général, peuvent contribuer à l'étude de ces dernières. S'inscrivant dans la tradition de Goffman, qui fait de l'échange réparateur l'un des noyaux de l'interaction conversationnelle, un certain nombre d'auteurs tendent à adopter un modèle d'analyse dans lequel ce type d'échange tient une place essentielle. Ces mêmes auteurs prétendent mettre à jour la logique d'une interaction. Une telle affirmation se révèle quelque peu troublante. En effet, ce que la logique déontique nous apprend, c'est que l'analyse logique de l'activité réparatrice soulève des difficultés manifestes, connues sous le nom de « paradoxe de Chisholm » (du nom de celui qui, le premier, les aborda). Ayant rappelé en quoi consiste le paradoxe, nous nous sommes tout d'abord tournés vers une solution souvent proposée. Elle consiste à distinguer un avant et un après l'offense territoriale. Nous avons illustré le principe de cette solution en prenant l'exemple de l'excuse. Ensuite, constatant que ce type de compte rendu s'applique plus difficilement à la réparation par la prière, nous nous sommes tournés vers un autre style d'analyse, en termes de révision. L'idée de base est relativement simple et assez plausible. Elle consiste à supposer que, au moment de l'offense, les interactants révisent minimalement l'ordonnancement des mondes possibles, de manière à formuler le jugement inconditionnel adéquat. Un examen attentif révèle que, au moment de l'infraction, l'état le meilleur et le « deuxième état » le meilleur permutent. Cela semble très proche de l'opération de révision naturelle, que Boutilier étudie dans le domaine ontique.

Ainsi, nous avons finalement axé la réflexion sur les difficultés auxquelles l'analyste se heurte, lorsqu'il tente de rendre compte d'une interaction en face à face. Au terme de notre enquête, il nous est apparu que ces difficultés s'estompent en partie, si l'on adopte une conception moins statique de la logique, et si l'on met au premier plan l'opération de révision et sa dynamique propre. Il va de soi que le présent chapitre ne pouvait que donner une idée de ce que devraient être les procédures de révision liées à une théorie des obligations. Nulle part, nous n'avons cherché à définir strictement ces procédures, dont la seule rédaction poserait déjà bien des problèmes que nous ne prétendons pas aborder. Demeurant à distance de ces problèmes, nous souhaitions simplement donner un nouvel éclairage à une difficulté maintes fois ressassée. Car enfin, si celle-ci n'a pas été préalablement résolue, l'idée d'étendre les prétentions de la logique au fondement de nos argumentations quotidiennes risque d'inspirer le reproche, auquel nous serions sensibles, d'être peu crédible. Nous sommes de ceux que les projets trop vastes et trop ambitieux mettent mal à l'aise, et qui donc privilégient l'austérité et la retenue du propos. En guise de conclusion, nous indiquons rapidement certaines des pistes que nous souhaiterions explorer dans un futur proche, en rapport avec notre compte rendu en termes de révision.

Tout d'abord, nous avons oublié de mentionner qu'une idée visiblement analogue est développée par van der Torre et Tan [259]. Nous avons découvert celle-ci seulement après coup. Se plaçant à l'intérieur de la sémantique de la mise à jour de Veltman [263], ces auteurs distinguent deux sortes de mise à jour. La première n'est pas au coeur de notre propos ; la seconde fait plutôt songer à la procédure de révision rangée de Lehmann [135]. Expliquons-nous brièvement. A l'hypothèse traditionnelle selon laquelle saisir la signification d'une phrase revient à connaître les conditions sous lesquelles cette phrase est vraie, la sémantique de la mise à jour substitue le principe selon lequel cela consiste à connaître quel(s) changement(s) cette phrase provoque chez celui qui accepte les données dont elle est porteuse. Ainsi, dans le cas du déontique,

> « Il est obligatoire que p »

sera paraphrasé comme

> « La mise à jour d'un univers dans lequel les mondes sont d'égale valeur (degré zéro d'information) donne un univers dans lequel les p-mondes sont strictement préférés aux $\neg p$-mondes ».

Cette première opération de mise à jour, nos auteurs la nomment « opération de réduction ». En voici la raison. Que w et w' soient d'égale valeur se traduit par le fait que $w \preceq w'$ et que $w' \preceq w$. Que w soit strictement préféré à w' se traduit par le fait que $w \preceq w'$ et que $w' \not\preceq w$. Ainsi, nous partons de $\preceq = W \times W$, puis nous réduisons cet ensemble, en y supprimant telle ou telle paire[42]. L'information à incorporer exprime ici une obligation. Il faut à présent envisager le cas où elle exprime un fait et, plus précisément, une violation. Nous sommes ici face à une mise à jour d'un genre nouveau, que nos auteurs nomment « zoom avant » (*zoom in*), car ils supposent qu'elle donne tout au plus une sous-structure de la structure initiale. Désormais, l'univers a pour seuls et uniques éléments les mondes vérifiant le fait à incorporer (une violation). Il y a une certaine analogie entre la mise à jour-zoom avant et la révision rangée de Lehmann. En gros, un modèle rangé peut être décrit comme une succession de strates d'interprétations de plus en plus imparfaites. Ainsi, pour rester sur l'exemple du cycle réparateur, nous aurons typiquement une structure de la forme :

$w_6\bullet$	o	$\neg p$	$\neg s$	$\neg a$	$\neg m$
$\curlyvee$					
$w_5\bullet$	o	p	$\neg s$	$\neg a$	$\neg m$
$\curlyvee$					
$w_4\bullet$	o	p	s	$\neg a$	$\neg m$
$\curlyvee$					
$w_3\bullet$	o	p	s	a	$\neg m$
$\curlyvee$					
$w_2\bullet$	o	p	s	a	m
$\curlyvee$					
$w_1\bullet$	$\neg o$	$\neg p$	$\neg s$	$\neg a$	$\neg m$

FIG. 3.7 – Modèle rangé avant l'offense

Supposons que « $p(\Phi)$ » énumère les interprétations de la strate la plus proche qui sont des modèles de Φ. Puis désignons par « $\Phi \bullet \phi$ » la mise à jour de Φ par ϕ. Voici la procédure qu'utilise Lehmann. S'il y a des modèles communs à $p(\Phi)$ et à ϕ, alors nous restreignons le contenu des obligations, $p(\Phi \bullet \phi)$, à ces modèles communs. Mais si $p(\Phi)$ et ϕ n'ont pas de modèles en commun, alors nous changeons de strate jusqu'à en trouver une qui contient des modèles de la nouvelle information. Avant que l'intrusion territoriale ne soit commise, les interactants prennent

[42] Il est d'usage d'exiger de $\preceq$ qu'elle soit transitive. Ceci pose un problème, lorsque l'information nouvelle est une obligation conditionnelle. Voir van der Torre et Tan [259, p. 86].

donc la strate de rang 1 pour point de repère. Lorsque l'offense est accomplie, c'est-à-dire lorsqu'ils introduisent o dans $\Phi = \ldots \bullet \neg o$, ils mettent aussitôt cette strate à l'arrière-plan, et prennent la strate de rang 2 pour nouveau point de repère, tout en gardant un oeil sur les strates de rang supérieur. Ainsi, tout se passe comme s'ils effectuaient un zoom avant sur l'ensemble des o-mondes, tout en se laissant la possibilité d'annuler ultérieurement cette opération par un zoom arrière − possibilité que la théorie de van der Torre et Tan semble exclure :

$$
\begin{array}{|lccccc|}
\hline
w_6\bullet & o & \neg p & \neg s & \neg a & \neg m \\
\ \ \curlyvee & & & & & \\
w_5\bullet & o & p & \neg s & \neg a & \neg m \\
\ \ \curlyvee & & & & & \\
w_4\bullet & o & p & s & \neg a & \neg m \\
\ \ \curlyvee & & & & & \\
w_3\bullet & o & p & s & a & \neg m \\
\ \ \curlyvee & & & & & \\
w_2\bullet & o & p & s & a & m \\
\hline
\Uparrow \curlyvee & & & \Uparrow & & \Uparrow \\
\end{array}
$$
$$
\begin{array}{lccccc}
w_1\bullet & \neg o & \neg p & \neg s & \neg a & \neg m
\end{array}
$$

FIG. 3.8 – Modèle rangé $(\ldots \bullet \neg o) \bullet o$ et zoom avant

Une étude comparative des deux types d'approches (celle en termes de permutation et celle en termes de zoom avant) reste à faire.

Passons à la deuxième direction de recherches possible. Jusqu'à présent, nous nous sommes principalement intéressés à la question de savoir comment articuler la dimension du devoir-être à celle du temps. De même que nous n'avons pas pris en compte l'individu sur lequel porte l'obligation, nous avons mis entre parenthèse les interférences possibles avec la théorie des actes de langages. Il semble relativement évident que toute vue réaliste du fonctionnement d'une séquence réparatrice doit prendre en compte cette dimension. Il serait ici intéressant d'étudier le travail de Belnap et Green [25], qui se proposent d'articuler la théorie des actes de langages à une logique temporelle arborescente. De même, il serait instructif d'examiner, dans cet esprit, l'étude de Jones [122], dans laquelle nous ne ferons ici qu'une timide incursion. Recourant aux outils que les logiques modales ont mis à notre disposition, ce dernier propose une théorie de la communication, qui accorde une grande importance au composant déontique. Cela nous permettrait de prendre (par exemple) en compte les superpositions qui s'ajoutent au cycle réparateur. Goffman cite le cas où l'on détourne un échange en répondant, au sens littéral d'un énoncé et non au mouvement unanimement perçu :

A : « Est-ce que tu peux me dire l'heure ? »
B : « Oui, et toi ? »

A : « Est-ce que tu as l'heure ? »
B : « Oui ; j'ai aussi les minutes » (d'après [93, p. 166])

Ici, un ordre se superpose à une question. Pour en rendre compte, il est tentant d'utiliser ce que Jones nomme « règle de signalisation », de la forme :

$$(E_s x \wedge Z \wedge \bigcirc_a E_s x) \rightarrow E_s y \qquad \text{(Superposition)}$$

La lettre E désigne un opérateur d'action. « $E_s x$ » est lu « s (le locuteur) fait en sorte que x ». Ici, x désigne une question. L'analyse traditionnelle l'assimilera à un impératif de la forme

$$\text{Shall} E_a (K_s p \vee K_s \neg p),$$

ce qu'on peut lire :

Que a fasse en sorte que s sache si p ou non !

La lettre y à son tour est une abréviation pour l'impératif

$$\text{Shall} E_a q$$

ce qui peut être lu :

Que a fasse en sorte que q !

« $\bigcirc_a E_s x$ » désigne la composante déontique que, selon Jones, recèlerait toute communication. Il lit l'énoncé « $\bigcirc_a E_s x$ » comme signifiant « le fait que s fasse x est optimal pour a, dans le sens que x optimise l'intérêt que a a à être correctement informé ». La lettre Z décrit les caractéristiques du contexte et le symbole $\rightarrow$ désigne l'implication matérielle. Nous ne pensons pas que l'usage de cette sorte d'implication soit paradoxal. Comme Goodman, nous interprétons Z comme une prémisse « cachée » autorisant l'inférence (classique) de $E_s y$ à partir de $E_s x$ et de $\bigcirc_a E_s x$. Ceci rend l'expression « $(\ldots \wedge Z) \rightarrow \ldots$ » synonyme de « ... entraîne normalement... ». Dans le contexte de l'échange réparateur, nous relirons la phrase « le fait que s fasse x est optimal pour a » comme signifiant « le fait que s fasse x minimise l'offense qui est faite à a ». Cette lecture de $\bigcirc_a$ est directement empruntée à Brown et à Levinson [42], qui présentent l'indirection comme l'un des moyens possibles d'atténuer les effets négatifs de l'intrusion territoriale. Nous voyons ici en quel sens deux actes de langage coexistent dans une même réplique du cycle réparateur.

Cette allusion à Brown et à Levinson nous donne l'occasion d'évoquer une autre direction de recherche possible. Ceux-ci fondent leur approche sur le concept de face. Tout interactant en possèderait deux : une face négative, qui renvoie aux territoires du moi de Goffman ; une face positive, qui correspond à l'ensemble des images valorisantes qui les interlocuteurs construisent et tentent d'imposer dans l'interaction. Dans tout échange verbal, ce sont donc quatre faces qui se trouvent en présence, et qui sont potentiellement menacées. Les auteurs parlent ici de *Face Threatening Acts*, ce qu'ils notent « *FTA(s)* ». Certains actes de langage menacent la face négative de celui qui les accomplit. D'autres menacent sa face positive. Même remarque pour celui envers lequel ils sont accomplis. L'essentiel du travail de Brown et Levinson consiste à étudier comment « dans le contexte d'une vulnérabilité mutuelle des faces, un agent rationnel cherchera à éviter ces actes menaçant pour la face, ou bien emploiera certaines stratégies pour minimiser la menace » [42, p. 68]. Il n'entre pas dans le projet de cette conclusion d'analyser dans le détail les diverses stratégies qu'ils dégagent. Quant à retracer les épisodes des débats auxquels leur étude donna lieu, l'épaisseur et la complexité du dossier rendent cette tâche incompatible avec les limites qui nous sont imparties. Néanmoins, nous pensons qu'un logicien moderne pourrait trouver une source d'inspiration dans de tels débats. Par exemple, après avoir esquissé un modèle du raisonnement pratique qui est centré sur la relation entre la fin et les moyens, Brown et Levinson [42, p. 90] suggèrent qu'un tel modèle serait utile pour comprendre la façon dont les interactants s'y prennent pour choisir telle ou telle stratégie. Il serait intéressant d'étudier ce problème à la lumière des récents développements de la logique.

Voici une dernière piste de recherche possible. L'analyse que nous avons esquissée rend compte tout au plus d'un certain type d'interaction verbale. Elle porte sur l'amont, ou l'aval, de nos argumentations quotidiennes. Pouvons-nous donner à notre approche une plus grande généralité ? Le temps nous a également manqué, pour étudier sérieusement les travaux menés en analyse de la conversation. Les plus connus portent sur le thème de la gestion des tours de parole [219]. Ils portent aussi sur les thèmes connexes de l'ouverture [222] et de la clôture d'une conversation [223], ainsi que sur celui de la suspension des procédures de tours de parole [218]. C'est ici, croyons-nous, que l'utilisation d'un cadre d'analyse spécifiquement temporel, du genre de celui qui nous servit de point de départ, a le plus de chance de porter des fruits. Comme on le sait, les logiques du temps arborescent permettent d'étudier le comportement des programmes non-déterministes. Serait-il absurde d'attendre de celles-ci qu'elles nous aident à clarifier l'organisation séquentielle d'une conversation ?

Index

Table des figures

Bibliographie

[1] Abbott B., "Some problems in giving an adequate model-theoretic account of cause", *Berkeley Studies in Syntax and Semantics*, 1.

[2] Alchourrón C. E., 1993, "Philosophical foundations of deontic logic and the logic of defeasible conditional obligations". Dans J.-J.Ch. Meyer et R. J. Wieringa (eds), *Deontic Logic in Computer Science : Normative System Specification*, John Wiley & Sons, p. 43-83.

[3] Alchourrón C. E., 1993, "Defeasible logics : Demarcation and affinities". Dans G. Crocco, L. Farinas del Cerro et A. Herzig (eds), *Conditionals : from Philosophy to Computer Science*, Oxford : Clarendon Press, p. 67-103.

[4] Alchourrón C.E. et Bulygin E., "The expressive conception of norms". Dans R. Hilpinen (ed), *New Studies in Deontic Logic*, Reidel, Dordrecht, p. 95-124.

[5] Alchourrón C.E., Gärdenfors P. et Makinson D., 1985, "On the logic of theory change : partial meet contraction and revision functions", *Journal of Symbolic Logic*, 50, p. 510-530.

[6] Alchourrón C.E. et Makinson D., 1981, "Hierarchies of regulations and their logic". Dans R. Hilpinen (ed), *New Studies in Deontic Logic*, Reidel, Dordrecht, p. 125-148.

[7] Alchourrón C.E. et Makinson D., 1985, On the logic of theory change : Safe contraction", *Studia Logica*, 44, p. 405-422.

[8] Anscombre J.C., 1989, "Théorie de l'argumentation, topoï et structuration discursive", *Revue québecoise de linguistique*, 18/1, p. 13-56.

[9] Anscombre J.C. et Ducrot O., 1978-1979, "Lois logiques et lois argumentatives", *Le Français moderne*, 46 & 47, p. 347-357 & 37-52.

[10] Anscombre J.C. et Ducrot O., 1983, *L'argumentation dans la langue*, Bruxelles, Mardaga.

[11] Appelt D. et Konolige K., 1989, "A nonmonotonic logic for reasoning about speech acts and belief revision". Dans M. Reinfrank, J. De Kleer,M. Ginsberg et E. Sandewall (eds), *Proceedings of the SecondInternational Workshop on Non-Monotonic Reasoning*, volume 346 de LNAI,Springer-Verlag, p. 164-175.

[12] Åqvist L., 1965, "A new approach to the logical theory of interrogatives", *Filosofiska Studier*, Uppsala.

[13] Åqvist L., 1966, "*Next* and *Ought*. Alternative foundations for von Wright's tense logic, with an application to deontic logic", *Logique et Analyse*, 9, p. 231-251.

[14] Åqvist L., 1987, *Introduction to Deontic Logic and the Theory of Normative Systems*, Napoli, Bibliopolis.

[15] Åqvist L., 1991, "Deontic tense logic : restricted equivalence of certain forms of conditional obligation and a solution to Chisholm's paradox". Dans G. Schurz et G.J.W. Dorn (eds), *Advances in Scientific Philosophy*, Amsterdam-Atlanta, GA : Rodopi, 1991, p. 127-141.

[16] Åqvist L., 1993, "A completeness theorem in deontic logic with systematic frame constants", *Logique & Analyse*, 141-142, p. 177-192.

[17] Åqvist L., 1997, "Systematic frame constants in defeasible deontic logic. A new form of Andersonian reduction". Dans D. Nute (ed), *Defeasible Deontic Logic*, Dordrecht, Kluwer, p. 59-77.

[18] Åqvist L., 1999, "Supererogation and offence in deontic logic : an analysis within systems of alethic modal logic with levels of perfection". Dans R. Sliwinski (ed.), *Philosphical crumbs. Essays dedicated to Ann-Mari Henschen-Dahlquist on the occasion of her seventy-fith birthday*, Uppsala University, p. 261-276.

[19] Åqvist L. et Hoepelman J., 1981, "Some theorems about a ‘tree" system of deontic tense logic". Dans R. Hilpinen (ed), *New Studies in Deontic Logic*, Reidel, Dordrecht, p. 187-221.

[20] Asher N. et Morreau M., 1991, "Commonsense entailment : a modal theory ofnonmonotonic reasoning". Dans J. Mylopoulos et R. Reiter (eds), *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, MorganKaufmann, Los Altos, CA, p. 387-392.

[21] Ashley K.D., 1990, *Modeling legal argument : reasoning with cases and hypotheticals*. Cambridge, MA : MIT Press.

[22] Atlas J.D., 1977, "Negation, ambiguity, and presupposition", *Linguistics and Philosophy*, 1, p. 321-336.

[23] Bellman R.E. et Zadeh L.A., 1977, "Local and fuzzy logics". Dans J.M. Dunn et G. Epstein (eds), *Modern Uses of Multiple-Valued Logic*, Dordrecht, Reidel, p. 105-165.

[24] Belnap N.D., 1963, *An analysis of questions,* Preliminary report, Santa Monica : System Development Corporation.

[25] Belnap N.D. et Green M., 1994, "Indeterminism and the thin red line", *Philosophical Perspectives*, 8, Logic & Language.

[26] Belzer M., 1987, "Legal reasoning in 3-D", *Proceedings of the 1st International Conference on AI and Law*, p. 155-163.

[27] Ben-Ari M. et al., 1983, "The temporal logic of branching time", *Acta Informaticae*, 20, p. 207-209.

[28] Bench-Capon T.J.M., 1988, "Specification and implementation of Toulmin-dialogue game". Dans J.C. Hage et al. (eds), *Legal Knowledge Based Systems :JURIX 1998 : The Eleventh Conference,* Nijmegen : GNI, p. 5-19.

[29] van Benthem J., 1991, *Language in action. Categories, lambdas and dynamic logics*, North Holland, Amsterdam.

[30] van Benthem J., 1995, "Temporal logic". Dans D. Gabbay, C.J. Hogger et J.A. Robinson (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. IV, Oxford : Clarendon Press, p. 241-350.

[31] Besnard P., 1989, *An introduction to Default Logic*. Springer Verlag.

[32] Besnard P., Quiniou R. et Quinton P., 1983, "A theorem prover for a decidable subset of default logic", *AAAI83,* p. 27-30.

[33] Besnard P. et Siegel P., 1988, "The preferential models approach to non monotonic logic". Dans P. Smets, A. Mamdani, D. Dubois et H. Prade (eds), *Non-standard logics for automated reasoning*, London : Academic Press, p. 137-161.

[34] Bezzazi H., Makinson D. et Pino Perez R., 1997, "Beyond rational monotony : some strong non-horn rules for non-monotonic inference relations", *Journal of Logic and Computation*, 7, p. 605-631.

[35] Borel M.J., Grize J. et Miéville D., 1983, *Essai de logique naturelle*.

[36] Boutilier C. et Becher V., 1995, "Abduction as belief revision", *Artificial Intelligence*,77, p. 43-94.

[37] Boutilier C., 1996, "Iterated revision and minimal change of conditional beliefs", *Journal of Philosophical Logic*, 25(3), p. 262-305.

[38] Brandom R. B., 1994, *Making it Explicit*, London, Harvard University Press.

[39] Brewka G., 1994, "A logical reconstruction of Rescher's theory of formal disputation based on default logic", *Proc.of the 11th European Conference on Artificial Intelligence,* p. 366-370.

[40] Brewka G., 1991, "Cumulative default logic : in defense of non-monotonic inference rules", *Artificial Intelligence,* 50, p. 183-205.

[41] Brown M. A., 2000, "Conditional obligation and positive permission for agents in time", *Nordic Journal of Philosophical Logic*, 5, p. 83-111.

[42] Brown P. et Levinson S.C., 1987, *Politeness. Some universals in language use*. Cambridge, Cambridge University Press.

[43] Carmo J. et Jones A.J.I., 1997, "A new approach to contrary-to-duty obligations". Dans D. Nute (ed), *Defeasible Deontic Logic*, Dordrecht : Kluwer, p. 317-344.

[44] Castañeda H.N., 1959, "On a proposed revolution in logic", *Mind*, p. 279-292.

[45] Castañeda H.N., 1981, "The paradoxes of deontic logic : the simplest solution to all of them in one fell swoop". Dans R. Hilpinen (ed), *New Studies in Deontic Logic*, Reidel, Dordrecht, p. 378-385.

[46] Chellas B. F., 1971, "Imperatives", *Theoria*, 37, p. 114-129.

[47] Chellas B. F., 1974, "Conditional obligation". Dans S. Stendlund (ed), *Logical Theory and Semantic Analysis*, Dordrecht : D. Reidel, p. 23-33.

[48] Chellas B. F. 1980,*Modal Logic : an Introduction*, Cambridge University Press.

[49] Chisholm R.M., 1963, "Contrary-to-duty imperatives and deontic logic", *Analysis*, 24, p. 33-36.

[50] Cholvy L. et Garion C., 1999, "An attempt to adapt a logic of conditional preferences for reasoning with contrary-to-duties". Dans R. Demolombe et R. Hilpinen (eds), *Proceedings of the 5th International Workshop on Deontic Logic*, Janvier 2000, Toulouse.

[51] Cohen P.R. et Levesque H.J., 1985, "Speech acts and rationality". Dans *Proceedings of the Twenty-third Annual Meeting*, Association for Computational Linguistics, Chicago, IL.

[52] Console L., Theseider Dupré D. et Torasso P., 1991, "On the relationship between abduction and deduction", *Journal of Logic and Computation*, 1, 5, p. 661-690.

[53] Danielsson S., 1968, *Preference and Obligations. Studies in the Logic of Ethics.* Filosofiska föreningen, Uppsala.

[54] Darwiche A. et Pearl J., 1994, "On the logic of iterated belief revision". Dans *Theoretical Aspects of Reasoning about Knowledge : Proceedings of the 1994 Conference*, San Mateo : Morgan Kaufmann Publishers, p. 5-23.

[55] Davidson D., 1970, "How is weakness of the will possible". Dans J. Feinberg (ed), *Moral Concepts*, Oxford University Press. Cet article a été réédité dans Davidson D., 1980, *Essays on Actions and Events*, Oxford University Press, p. 21-42.

[56] Delgrande J.P., 1987, "A first-order conditional logic for prototypical properties", *Artificial Intelligence*, 33, p. 105-130.

[57] Dignum F., Meyer J.-J.Ch. et Wieringa R. J., 1996, "Free choice and contextually permitted action", *Studia Logica*, 57, p. 193-220.

[58] Doyle J., 1972, "A truth-maintenance system", *Artificial Intelligence*, 12, p. 231-272.

[59] Ducrot O., 1972, *Dire et ne pas dire*, Paris, Hermann.

[60] Ducrot O., 1973, *La preuve et le dire*, Tours, Mame.

[61] Ducrot O., 1980, *Les échelles argumentatives*, Paris, Minuit.

[62] Ducrot O. et al., 1980, *Les mots du discours*, Paris, Minuit.

[63] Ducrot O., 1983, "Opérateurs argumentatifs et visée argumentative", *Cahiers de Linguistique Française*, Genève, 5, p. 7-36.

[64] Ducrot O., 1984,*Le dire et le dit*, Paris, Minuit.

[65] Dung P.M., 1992, "Logic programming and as dialogue games", Technical Report, Division of Computer Science, Asian Institute of Technology, Bangkok.

[66] Dung P.M., 1995, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games", *Artificial Intelligence*, 77, p. 321-357.

[67] van Eck J. A., 1981, "A logical analysis of the notions of prima facie obligations and actual obligation". Dans E. Morscher et R. Stranzinger (eds), *Ethics. Proceedings of the Fifth International Wittgenstein Symposium*, Hölder-Pichler-Tempsky, Vienna, p. 172-175.

[68] van Eck J. A., 1982, "A system of temporally relative modal and deontic predicate logic and its philosophical applications", *Logique et Analyse*, 100, p. 249-381.

[69] van Eemeren F.H. et Grootendorst R., 1984, *Speech acts in argumentative discussions*, Dordrecht, Cinnaminson : Foris/Berlin : Mouton de Gruyter.

[70] van Eemeren F.H, Grootendorst R., et F.S. Henkemans, 1996, *Fundamentals of argumentation theory : A handbook of historical backgrounds and contemporary developments*, Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey.

[71] van Eemeren F.H. et Grootendorst R., 1996, *La nouvelle dialectique*, trad. fr. coord. par Ch. Plantin, Editions Kimé, Paris. Dordrecht, Cinnaminson : Foris/Berlin : Mouton de Gruyter.

[72] Engel P., 1989,*La norme du vrai*, Paris, Gallimard.

[73] Fahlman S.E., Touretzky D.S. et van Roggen W., 1981, "Cancellation in a parallel semantic network", *Proc. Int. Joint Conf. Artif. Intell. 81*, Vancouver, p. 257-263.

[74] Fouqueré C. et Vauzeilles J., 1993, "Taxonomic linear theories", *ECSQA-RU'93*, Springer-Verlag, p. 121-128.

[75] van Fraassen B., 1966, "Singular terms, truth-value gaps and free logic", *Journal of Philosophy*, 63, p. 481-495.

[76] van Fraassen B., 1968, "Presupposition, implication and self-reference", *Journal ofPhilosophy*, 65, p. 136-152.

[77] Froidevaux C., 1985,"ISA hierarchies with exceptions", *Actes 5ème Congrès AFCET Reconnaissance des Formes et Intelligence Artificielle*, Grenoble, p. 1127-1138.

[78] Froidevaux C. et Kayser D., 1988, "Inheritance in semantic networks and default logic". Dans P. Smets, A. Mamdani, D. Dubois et H. Prade (eds), *Non-standard logics for automated reasoning*, Academic Press, London, p. 179-212.

[79] Gabbay D.M., 1985, "Theoretical foundations for non-monotonic reasoning in expert systems", *Proceedings NATO*, Advanced Study Institute on Logics and Models of Concurrent Systems, p. 439-457.

[80] Gallin D., 1975, *Intensional and Higher-Order Modal Logic*, Elsevier, Amsterdam.

[81] Gamut L.T.F., 1984, *Logic, language, and Meaning*, vol. II (Intensional Logic and Logical Grammar), The University of Chicago Press, Chicago and London.

[82] Gärdenfors P. et Makinson D., 1988, "Revision of knowledge system using epistemic entrenchment", *TARK'88*, Morgan Kaufmann, p. 83-97.

[83] Gärdenfors P. et Rott H., 1995, "Belief revision", dans D. Gabbay et J.A. Hogger (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol.4, Clarendon Press, Oxford, p. 36-132.

[84] Gärdenfors P., 1988, *Knowledge in Flux : Modeling the Dynamics of Epistemic States*, MIT Press, Cambridge.

[85] Gardies J.L., 1979, *Essai sur les logiques des modalités*, Paris, PUF.

[86] Gardies J.L., 1994, *Les fondements sémantiques du discours naturel*, Paris, Vrin.

[87] Gazdar G., 1977, *Implicature, Presupposition and Logical Form*, Bloomington, IULC.

[88] Gazdar G., 1979, *Pragmatics. Implicature, Presupposition and Logical Form*, New York, Academic Press.

[89] Gazdar G., 1979, "A solution to the projection problem". Dans C.-K. Oh et D.A. Dinneen (eds), *Syntax and Semantics 11 : Presuppositions*, New York, Academic Press, p. 57-89.

[90] Goble L., 1990, "A logic of *Good*, *Should* and *Would* - Part II", *Journal of Philosophical Logic*, 19, p. 253-276.

[91] Goble L., 1999, "Deontic logic with relevance". Dans P. Mc Namara and H. Prakken (eds), *Norms, Logics and Information Systems (ΔEON'98)*, Amsterdam, IOS, p. 331-346,

[92] Gochet P., Gribomont P. et Thayse A., 2000, *Logique*, vol.3, Méthodes pour l'intelligence artificielle, Paris, Hermes Science Publications.

[93] Goffman E., 1973, *La mise en scène de la vie quotidienne*, vol.2, Les relations en public, Paris, Ed. de Minuit.

[94] Gordon T.F. et Karacapilidis N., "The Zeno argumentation framework", *Workshop on ComputationalDialectics*, FAPR'96.

[95] Grice H.P., 1975, "Logic and conversation". Dans P. Cole et J.L. Morgan (eds), *Syntax and Semantics 3 : Speech Acts*, New York, Academic Press, p. 41-58

[96] Grice H.P., 1979, "Logique et conversation", *Communications*, 30, p 57-72.

[97] Grize J., 1982, *De la logique à l'argumentation*, Librairie Droz, Genève.

[98] Grize J. et Piéraut-le Bonniec G., 1983, *La contradiction*, Paris, Presses Universitaires de France.

[99] Halpern J. Y. et Moses Y., 1990, "Knowledge and common knowledge in a distributed environment", *Journal of the ACM*, 37 :3, p. 549-587.

[100] Halpern J. Y., 1997, "Defining relative likelihood in partially-ordered preferential structures", *Journal of Artificial Intelligence Research*, 7, p. 1-24.

[101] Hambin C.L., 1970, *Fallacies*, London : Methuen.

[102] Hansen J., 1999, "On relations between Åqvist's G and Van Eck's deontic temporal logic". Dans P. Mc Namara et H. Prakken (eds), *Norms, Logics and Information Systems (ΔEON'98)*, Amsterdam, IOS, p. 127-144.

[103] Hansson B., 1969, "An analysis of some deontic logics", *Nous*, 3, p. 373-398. Réédité dans R. Hilpinen (ed), 1971, *Deontic Logic : Introductory and Systematic Readings*, D. Reidel Publ. Co., Dordrecht, p. 121-147.

[104] Hansson B., 1974, "A program for pragmatics". Dans S. Stendlund (ed), *Logical Theory and Semantic Analysis*, Dordrecht : D. Reidel.

[105] Hansson S.O., 1991, *Belief base dynamics*, Ph.D. Thesis, Uppsala University.

[106] Hansson S.O., 1997, "Semi-revision", *Journal of Applied Non-Classical Logics*, 7, p. 151-175.

[107] Harrah D., 1984, "The logic of questions". Dans D. Gabbay et Guenthner (eds), *Handbook of Philosophical Logic*, vol. II, Reidel.,p. 715-764.

[108] Hempel C., 1965, *Aspects of Scientific Explanation*, Free Press, New York.

[109] Herzig A. et Longin, D., 2000, "Belief dynamics in cooperative dialogues", *Journal of Semantics*. A paraître.

[110] Hintikka J., 1971, "Some main problems of deontic logic". Dans R. Hilpinen (ed), 1971, *Deontic Logic : Introductory and Systematic Readings*, D. Reidel Publ. Co., Dordrecht, p. 59-104.

[111] Hintikka J., 1976, "The semantics of questions and the questions of semantics", *Acta Philosophica Fennica*, vol. 26, fasc. 4.

[112] Hintikka J., 1987, "The fallacy of fallacies" *Argumentation*, 1, p. 211-238.

[113] Horn L.R., 1984, "Towards a new taxonomy for pragmatic inference Q-based and R-based implicature". Dans D. Schiffrin (ed), *Meaning, Form and Use in Context (GURT84)*, Washington, Georgetown University Press.

[114] Horn L.R., 1988, "Pragmatic theory". Dans F. Newmeyer (ed), *Linguistics : The Cambridge Survey*, vol. 1, Cambridge, Cambridge University Press, p. 113-145.

[115] Horty J., Thomason R. et Touretzky D., 1990, "A skeptical theory of inheritance in nonmonotonic semantics networks", *Artificial Intelligence*, 42, p. 311-348.

[116] Horty J., 1997, "Nonmonotonic foundations for deontic logic". Dans D. Nute (ed), *Defeasible Deontic Logic*, Dordrecht, Kluwer, p. 17-58.

[117] Horty J., 2001, "Argument construction and reinstatement in logics for defeasible reasoning", *Artificial Intelligence and Law*, 9, p. 1-28.

[118] Huang Z. et Masuch M., 1995, "The logic of permission and obligation in the framework of ALX.3 : How to avoid the paradoxes of deontic logics", *Logique et Analyse*, 149, p. 55-74.

[119] Hughes G.E. et Cresswell M.J., 1996, *A New Introduction to Modal Logic*, London, Routledge.

[120] Jackson F., 1985, "On the semantics and logic of obligation", *Mind*, 94, p. 177-196.

[121] Jayez J., 1988, *L'inférence en Langue Naturelle*, Paris, Hermès.

[122] Jones A., 1990, "Towards a formal theory of communication and speech acts". Dans P.R. Cohen, J. Morgan et M.E. Pollack (eds), *Intentions in Communication*, Cambridge, MIT Press, p. 141-160.

[123] Karttunen L. et Peters S., 1979, "Conventional implicature". In C.K. Oh et D.A. Dinneen (eds), *Syntax and Semantics 11 : Presupposition*, New York, Academic Press, p. 1-56.

[124] Kautz H., "A circumscriptive theory of plan recognition". Dans P.R. Cohen, J. Morgan et M.E. Pollack (eds), *Intentions in Communication*, Cambridge, MIT Press, p. 105-133.

[125] Konolige K., 1988, "On the relation between default and autoepistemic", *Artificial Intelligence*, 35, p. 343-382.

[126] Konolige K., 1992, "Abduction versus closure in clausal theories", *Artificial Intelligence*, p. 255-272.

[127] Kraus S., Lehmann D., Magidor M., 1990, "Non-monotonic reasoning, preferential models and cumulative logics", *Artificial Intelligence*, 44, p. 106-207.

[128] Lakoff G., 1970, "Linguistics and natural logic", *Synthese*, 22, N°1-2 ; trad. fr. J. Milner et J. Sampy, *Linguistique et logique naturelle*, Paris, Klincksieck, 1976.

[129] Lascarides A. et Asher N., 1991, "Discourse relations and commonsense entailment", *Proceedings of the 29th Meeting of the Association for Computational Linguistics*, p. 55-63.

[130] Lascarides A. et Asher N., 1993, "Temporal interpretation, discourse relations and commonsense entailment", *Linguistics and Philosophy*, p. 437-494.

[131] Lascarides A. et Oberlander J., 1991, "Temporal coherence and defeasible knowledge", *Proceedings of the Workshop on Discourse Coherence*, Edinburg.

[132] Lea Sombe, 1988, "Raisonnements sur des informations incomplètes en intelligence artificielle", *Revue d'Intelligence Artificielle*, vol. 2, n°3-4.

[133] Lea Sombe, 1992, "Révision de base de connaissances", *4èmes Journées Nationales du Programme de Recherche Coordonnées en Intelligence Artificielle (PRC-IA)*, Marseille, Technopole de Château-Combert, p. 207-238.

[134] Lehmann D. et Magidor M., 1992, "What does a conditional knowledge base entail ?", *Artificial Intelligence*, 55, p. 1-60.

[135] Lehmann D., 1995, "Belief revision, revised", *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, p. 1534-1540.

[136] Lehmann D., 1998, "Stereotypical reasoning : logical properties", *Logic Journal of the Interest Group in Pure and Applied Logics*, 6(1), p. 49-58.

[137] Lenzen W., 1995, "On the semantics and pragmatics of epistemic attitudes". Dans Laux et H.Wansing (eds), *Knowledge and Belief in Philosophy and Artificial Intelligence*, Berlin, Akademie Verlag, p.181-197.

[138] Levinson S.C., 1983, *Pragmatics*, Cambridge University Press.

[139] Levinson S.C., 1987, "Minimization and conversational inference". Dans J. Verschueren et M. Bertucelli-Papi (eds), *The Pragmatic Perspective*, Amsterdam, John Benjamins, p. 61-129.

[140] Lewis D., 1973, *Counterfactuals*, Basil Blackwell.

[141] Lewis D., 1973, "Causation", *Journal of Philosophy*, 70, p. 556-567

[142] Lewis D., 1974, "Semantical analysis for dyadic deontic logic". Dans S. Stendlund (ed.), *Logical Theory and Semantic Analysis*, Dordrecht : D. Reidel, p. 1-14.

[143] Lin F. et Shoham Y., 1990, "Epistemic semantics for fix-point nonmonotonic logics", Technical report STAN-CS-90-1322, Dept. of Computer Science, Standford University.

[144] Lifschiftz V., 1994, "Circumscription". Dans D. Gabbay, C. Hogger et Robinson (eds), *Handbook of logic in artificial intelligence*, vol. III,Clarendon Press, Oxford, p. 297-352.

[145] Lindström S. et Rabinowicz W., 1993, "The Ramsey test revisited". Dans G. Crocco, L. Farinas del Cerro et A. Herzig (eds), *Conditionals : from Philosophy to Computer Science*, Oxford : Clarendon Press, p. 148-191.

[146] Lindström S. et Rabinowicz W., 1997, "Extending dynamic doxastic logic : accommodating iterated beliefs and Ramsey conditionals within DDL". Dans Lindahl, Needham et Sliwinski (eds), *For Good Measure*, Uppsala Philosophical Studies 46, p. 126-153.

[147] Lindström S. et Rabinowicz W., 1999, "Belief Change for Introspective Agents". Dans *Spinning Ideas, Electronic Essays Dedicated to Peter Gärdenfors on His Fiftieth Birthday*, http ://www.lucs.lu.se/spinning/.

[148] Livet P., 2000, "Dynamiques ethnométhodologiques". Dans de Formel M., Ogien A. et Quéré L. (eds), *L'ethnométhodologie*, Colloque de Cerisy.

[149] Livet P. (ed.), 2002, *Révision des croyances*, Paris, Hermès, 2002.

[150] Loui R.P., 1998, "Process and policy : resource-bounded non-demonstrative reasoning", *Computational Intelligence*, 14, p. 1-38.

[151] Loewer B. et Belzer M., 1983, "Dyadic deontic detachment", *Synthese*, 54, p. 295-318.

[152] Lukaszievicz W., 1988, "Considerations on default logic : an alternative approach", *Computational Intelligence*, 4, p. 1-6.

[153] Maibaum T., 1993, "Temporal reasoning over deontic specifications". Dans J.-J.Ch. Meyer et R. Wieringa (eds), *Deontic Logic in Computer Science : Normative System Specification*, John Wiley & Sons, p. 141-202

[154] Mc Namara P.M., 1994, "Doing well enough : toward a logic for common sense morality". Dans A.J.I. Jones et M. Sergot (eds), *Second International Workshop on Deontic Logic in Computer Science (DEON'94)*, Oslo 6-8 January 1994, p. 165-197. Egalement *Studia Logica*, 1996, 57, p. 167-192.

[155] Makinson D., 1966, "On some completeness theorems in modal logic", *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 12, p. 379-384.

[156] Makinson D., 1985, "How to give it up : A survey of some formal aspects of the logic of theory change", *Synthese*, 62, p. 347-363.

[157] Makinson D., 1988, "General theory of cumulative inference". Dans M. Reinfran, J. de Cleer, M.L. Ginsberg et E. Sandewall (eds), *Nonmonotonic Reasoning*, Springer-Verlag, p. 1-18.

[158] Makinson D., 1993, "Five faces of minimality", *Studia Logica*, 52, p. 339-379.

[159] Makinson D., 1994, "General patterns in nonmonotonic reasoning". Dans D. Gabbay, C. Hogger et Robinson (eds), *Handbook of logic in artificial intelligence*, vol. III, p. 35-110.

[160] Makinson D., 1999, "On a fundamental problem of deontic logic". Dans P. Mc Namara et H. Prakken (eds), *Norms, Logics and Information Systems (ΔEON'98)*, Amsterdam, IOS, p. 29-53.

[161] Makinson D. et van der Torre L., 2000, "Input/output logics", *Journal of Philosophical Logic*, 29, p. 383-408.

[162] Makinson D. et van der Torre L., 2001, "Constraints for input/output logics", *Journal of Philosophical Logic*, 30, p.155-185.

[163] Makinson D. et Gärdenfors P., 1991, "Relations between the logic of theory change and non-monotonic logic". Dans A. Fuhrmann et M. Morreau (eds), *The Logic of Theory Change*, Lecture notes in Artificial Intelligence, vol. 465, Springer Verlag, Berlin, p. 185-205.

[164] Marquis P., 1991, *Contribution à l'étude des méthodes de construction d'hypothèses en intelligence artificielle*, Thèse de doctorat de l'université de Nancy I.

[165] Marquis P., 1999, "Sur les preuves non déductives en intelligence artificielle". Dans J. Sallantin et J.J. Szczeciniarz (eds), *Le concept de preuve à la lumière de l'intelligence artificielle*, P.U.F., p. 138-158.

[166] Mercer R., 1987, *A default logic approach to the derivation of natural language presuppositions*, Doctoral Thesis, Department of Computing Science, University of British Columbia. (Egalement disponible sous forme de rapport technique : Technical report TR 87-35, Department of Computing Science, University of British Columbia.).

[167] Mercer R., 1988, "Using defaut logic to derive natural language presuppositions", *Proceedings of the Seventh Biennal Conference of the Canadian Society for Computational Studies of Intelligence*, Morgan Kauffmann, p. 14-21.

[168] Mercer R., 1990, "Deriving natural language presuppositions from complex conditionals", *Proceedings of the Eighth Biennial Conference of the Canadian Society for Computational Studies of Intelligence*, p. 114-120.

[169] Mercer R. et Reiter R., 1982, "The representation of presuppositions using defaults", *Proceedings of the Fourth National Conference of the Canadian Society for Computational Studies of Intelligence*, Saskatoon.

[170] Meredith C.A. et Prior A.N., 1965, "Modal logic with functorial variables and a contingent constant", *Notre Dame Journal of Formal Logic*, vol. VI, N°2, p. 99-109.

[171] van der Meyden R., 1996, "The dynamic logic of permission", *Journal of Logic and Computation*, 6, 3, p. 465-479.

[172] Meyer J.-J.Ch., 1988, "A different approach to deontic logic : deontic logic viewed as a variant of dynamic logic", *Notre Dame Journal of Formal Logic*, 29, p. 109-136.

[173] Montague R., 1970, "Universal grammar", *Theoria*, 36, p. 373-398.

[174] Montague R., 1973, "The proper treatment of quantification in ordinary English". Dans K.J.J. Hintikkaet al. (eds), *Approachesto natural language*, Dordrecht, p. 221-242.

[175] Moore R., 1984, "Possible-worlds semantics for auto-epistemic logic", *AAAI'84*, New Paltz, New York, p. 344-354.

[176] Moore R., 1985, "Semantical considerations on nonmonotonic logic", *Artificial Intelligence*, 25, p. 75-94.

[177] Moore R., 1985, "Autoepistemic logic". Dans P. Smets, A. Mamdani, D. Dubois et H. Prade (eds), *Non-standard Logics for Automated Reasoning*, Academic Press, Londres, p. 105-136.

[178] Mott P. L., 1973, "On Chisholm's paradox", *Journal of Philosophical Logic*, 2, p. 197-211.

[179] Mc Cawley J.D., 1981, *Everything that linguists have always wanted to know about logic (but were ashamed to ask)*, Chicago University Press.

[180] Mc Carthy J., 1980, "Circumscription - A form of nonmonotonic reasoning", *Artificial Intelligence*, 13, p. 27-39.

[181] Mc Carthy J., 1986, "Applications of circumscription to formalizing commonsense", *Artificial Intelligence*, 28, p. 89-116.

[182] McDermott D., 1982, "Nonmonotonic Logic II", *J. of the A.C.M.*, 29, p. 34-57.

[183] McDermott D. et Doyle J., 1980, "Nonmonotonic Logic I", *Artificial Intelligence*, 13, p. 41-72.

[184] Mackenzie J.D., "Question-begging in non-cumulative systems", 1979, *Journal of Philosophical logic*, 8, p. 117-133.

[185] Nute D., 1997, "Apparent obligation". Dans D. Nute (ed), *Defeasible Deontic Logic*, Dordrecht, Kluwer, p. 288-315.

[186] Nute D. et Yu X., 1997, "Introduction". Dans D. Nute (ed), *Defeasible Deontic Logic*, Dordrecht, Kluwer, p. 1-16.

[187] Owen M., 1983, *Apologies and Remedial Interchanges*, Berlin : Mouton.

[188] Parent X., 1998, "A three-valued approach to some of the deontic paradoxes", manuscrit.

[189] Parent X., 2000, "Defeasible conditional obligation : some remarks". Dans R. Demolombe et R. Hilpinen (eds), *Proceedings of the 5th International Workshop on Deontic Logic*, Janvier 2000, Toulouse.

[190] Parent X., 2001, "Cumulativity, identity and time in deontic logic", *Fundamenta Informaticae*, 48, p.237-252.

[191] Pearl J., 1988, "Embracing causality in default reasoning", *Artificial Intelligence*, 35, p. 259-271.

[192] Perelman C. et Olbrechts-tyteca L., 1951, *Traité de l'argumentation*, Editions de l'Université de Bruxelles, Bruxelles.

[193] Perrault C.R., 1990, "An application of default logic to speech act theory". Dans P.R. Cohen, J. Morgan et M.E. Pollack (eds), *Intentions in Communication*, Cambridge, MIT Press, p. 161-185.

[194] Pollock J., 1987, "Defeasible reasoning", *Cognitive Science*, 11, p. 481-518.

[195] Pollock J. , 1989, "OSCAR : a general theory of rationality", 1989, *Journal of Experimental and Theoretical Artificial Intelligence*, 1, p. 209-226.

[196] Pollock J., 1991, "Self-defeating arguments", *Minds and Machines*, 1, p. 367-392.

[197] Pollock J., 1991, "A theory of defeasible reasoning", *International Journal of Intelligent Systems*, Vol. 6, p. 33-54.

[198] Pollock J., 1992, "How to reason defeasibly", *Artificial Intelligence*, 57, p. 1-42

[199] Pollock J., 1996, "Implementing defeasible reasoning", *Workshop on ComputationalDialectics*, FAPR'96.

[200] Poole D., 1985, "On the comparison of theories : preferring the most specific explanation", *IJCAI'85*, Los Angeles, p. 144-147

[201] Prade H., 1988, "Raisonner avec des règles d'inférence graduelle", *Revue d'Intelligence Artificielle*, Vol. 2, n°2, p. 29-44.

[202] Prakken H., 1996, "Two approaches to the formalisation of defeasible deontic logic", *Studia Logica*, 57, p. 73-90.

[203] Prakken H., 1999, "On formalising burden of proof in legal argument". Dans H.J. van den Herik et als. (eds), *Legal Knowledge Based Systems, JURIX 1999*, Nijmegen : GNI, 1999, p. 85-97.

[204] Prakken H. et Sartor G., 1997, "Argument-based extended logic programming with defeasible priorities", *Journal of Applied Non-Classical Logics*, vol. 7, n° 1-2, p. 25-75.

[205] Prakken H. et Sergot M. J., 1996, "Contrary-to-duty obligations", *Studia Logica*, 57, p. 91-115.

[206] Prakken H. et Sergot M. J., 1997, "Dyadic deontic logic and contrary-to-duty obligations". Dans D. Nute (ed), *Defeasible Deontic Logic*, Dordrecht, Kluwer, p. 223-262.

[207] Prakken H. et Vreeswijk G.A.W., 2000, "Logical systems for defeasible argumentation". Dans D. Gabbay (ed), *Handbook of Philosophical Logic*, seconde édition. Dordrecht, Kluwer Academic Publishers.

[208] Reiter R., 1980, "A logic for default reasoning", *Artificial Intelligence*, 13, p. 81-132.

[209] Reiter R. et Criscuolo G., 1981, "On interacting defaults", *IJCAI-81*, p. 418-420.

[210] Risch V., 1993, *Les tableaux analytiques au service des logiques des défauts*, Thèse de doctorat, Informatique et Mathématique, Université d'Aix-Marseille II.

[211] Rott H., 1991, "Two methods of constructing contractions and revisions of knowledge systems", *Journal of philosophical logic*, 20, p. 149-173.

[212] Rott H., 1992, "Preferential belief change using generalized epistemic entrenchment", *Journal of Logic, Language and Information*, 1, p. 45-78.

[213] Rott H., 1992, "On the logic of theory change : More maps between different kinds of contraction functions". Dans P. Gärdenfors (ed), *Belief Revision*, Cambridge, Cambridge University Press, p. 122-144.

[214] Rott H., 2000, "Just Because : Taking belief bases seriously", Institute for Logic, Language and Information, Amsterdam, Rapport technique LP-1998-13.

[215] Rott H. et Pagnucco M., 2000, "Severe withdrawal (and recovery)", *Journal of Philosophical Logic*, 28, p. 501-457.

[216] Royakkers L. et Dignum F., 1997, "Defeasible reasoning with legal rules", D. Nute (ed), *Defeasible Deontic Logic*, Dordrecht, Kluwer, p. 263-286.

[217] Ryan M., 1992, "Representing Defaults as Sentences with Reduced Priority", *KR'92*, Morgan Kaufmann.

[218] Sacks H., 1974, "An analysis of the course of a joke's telling in conversation". Dans R. Bauman et J. Sherzer (eds), *Explorations in the Ethnography of Speaking*, Cambridge, Cambridge University Press, p. 337-353.

[219] Sacks H., Schegloff E.A. et Jefferson G., 1974, "A simplest systematics for the organisation of turn-taking in conversation", *Language*, 50, p. 696-735.

[220] Sadek M. D., 1991, *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication*, Thèse de doctorat en Informatique, Université de Rennes I, Rennes, France.

[221] Sadek M.D., 1994, "Towards a theory of belief reconstruction : Application to communication", *Speech Communication*, 15, p. 251-263.

[222] Schegloff E.A., 1968, "Sequencing in conversational openings", *American Anthropologist*, 70, p. 1075-1095.

[223] Schegloff E.A. et Sacks H., 1973, "Opening up closings", *Semiotica*, 7, p. 289-327.

[224] Schelchta K., 1995, "Defaults as generalized quantifiers", *Journal of Logic and Computation*, 92, 4, pp. 1-22.

[225] Schlechta K., 1997, "Filters and partial orders", *Journal of the Interest Group in Pure and Applied Logics*, vol. 5, n°5, p. 753-772.

[226] Schlechta K., 2000, "Representation results for limit preferential structures", en soumission.

[227] Schlechta K., 2000, "New techniques and completeness results for preferential structures", *Journal of Symbolic Logic*, vol. 65, n°2, p. 719-745.

[228] Searle J., 1975, "Indirect speech acts". Dans P. Cole et J.L. Morgan (eds), *Syntax andSemantics 3 : Speech Acts*, New York, Academic Press, p. 59-82.

[229] Searle J. et Vanderveken D., 1985, *Foundations of illocutionary logic*, Cambridge University Press.

[230] Segerberg K., 1973, "Two-dimentional modal logic", *Journal of Philosophical Logic*, 2, p. 77-96.

[231] Shoham Y., 1987, "A semantical approach to non-monotonic logic", *Proc. of the 10th Int. Conf. on Artificial Intelligence (IJCAI87)*, Milan, p. 388-392.

[232] Shoham Y., 1988, *Reasoning about change*, MIT Press.

[233] Shoham Y., 1990, "Non-monotonic reasoning and causation", *Cognitive Science*, 14, p. 213-252.

[234] Siegel P., 1991, "A modal logic for non-monotonic logic", *Proceedings of the D.R.U.M.S. Workshop*, Marseille, France, Février 1990.

[235] Siegel P. et Schwind C., "Hypothesis theory for nonmonotonic logic", *Proceedings of the 1st International Workshop on Nonstandard Queries and Answers*, Toulouse, France, Juillet 1991.

[236] Simon, 1991, "Non-monotonic reasoning and causation : comment", *Cognitive Science*, 15, p. 293-300.

[237] Sosa E. (ed.), 1975, *Causation and conditionals*, Oxford University Press.

[238] Sperber D. et Wilson D., 1986, *Relevance. Communication and Cognition*, Oxford, Blackwell.

[239] Spohn W., 1975, "An analysis of Hansson's dyadic deontic logic", *Journal of Philosophical Logic*, 4, p. 237-252.

[240] Spohn W., 1988, "Ordinal conditional functions : a dynamic theory of epistemic states". Dans W. Harper et B. Skyrms (eds), *Causation in Decision, Belief Change and Statistics*, vol. 2, Reidel, Dordrecht, p. 105-134.

[241] Stalnaker R. C., 1968, "A theory of conditionals". Dans W.L. Harper, R. C. Stalnaker et G. Pearce (eds), *Ifs*, Basil Blackwell Publisher, p. 41-55.

[242] Stalnaker R.C., 1973, "Presuppositions", *Journal of Philosophical Logic*, 2, p. 447-457

[243] Stalnaker R.C., 1980, "A note on nonmonotonic modal logic", Department of Philosophy, Cornell University.

[244] Suppes P., 1970, *A probabilistic theory of causation*, North Holland.

[245] Tan S.-W. et Pearl J., 1994, "Specification and evaluation of preferences under uncertainty". Dans *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, p. 530-539.

[246] Thomason R.H., 1984, "Combinations of tense and modality". Dans D. Gabbay and F. Guenthner (eds), *Handbook of Philosophical Logic*, vol. III, Dordrecht, D. Reidel, p. 135-165.

[247] Thomason R.H., 1990, "Accommodation, meaning, and implicature". Dans P.R. Cohen, J. Morgan et M.E. Pollack (eds),*Intentions in Communication*, Cambridge, MIT Press, p. 326-363.

[248] Thomason R.H., 1990, "Propagating epistemic coordination through mutual defaults", *TARK'90*, San Mateo, Morgan Kaufmann, p. 29-39.

[249] Thomason R.H. et Horty J.F., 1989, "Logics for inheritance theory". Dans M. Reinfrank et al. (eds), *Non-monotonic Reasoning (2nd International Workshop)*, Lectures Notes in Artificial Intelligence 346, Berlin, Springer, 220-237.

[250] Tomberlin J.E., 1981, "Contrary-to-duty imperatives and conditional obligation", *Noûs*, 16, p. 357-375.

[251] Tomberlin J. E., 1989, "Deontic paradox and conditional obligation", *Philosophy and Phenomenological Research*, 50, p. 107-114.

[252] Toulmin S.E., 1958, *The uses of argument*, Cambridge, Cambridge University Press.

[253] Toulmin S.E., 1993, *Les usages de l'argumentation*, trad. fr. Ph. De Brabanter, Paris, Presses Universitaires de France.

[254] Touretzky D., 1984, "Implicit ordering of defaults in inheritance systems", *AAAI'84*, p. 322-325.

[255] van der Torre L.W.N., 1997, *Reasoning about Obligations*, PhD thesis, Eramus University Rotterdam.

[256] van der Torre L.W.N., 1999, "The logic of reusable propositional output with the fulfilment constraint". Dans D. Basin et al. (eds), *Labelled Deduction*, Applied Logic Series, Kluwer, p. 247-268.

[257] van der Torre L.W.N. et Tan, Y.-H., 1997, "The many faces of defeasibility in defeasible deontic logic". Dans D. Nute (ed), *Defeasible Deontic Logic*, Dordrecht : Kluwer, p. 79-121.

[258] van der Torre L.W.N. and Tan Y.-H., 1998, "The temporal analysis of Chisholm's paradox", *AAAI-98*, Madison, Wisconsin, p. 650-655.

[259] van der Torre L.W.N. and Tan Y.-H., 1999, "An update semantics for deontic reasoning". Dans P. Mc Namara et H. Prakken (eds), *Norms, Logics and Information Systems (ΔEON'98)*, Amsterdam, IOS, p. 73-90.

[260] Truszczynski M., 1991, "Modal interpretations of default logic", *Proceedings of the International Joint Conference on Artificial Intelligence*, Sydney, Australia.

[261] Vanderveken D., 1991, *Meaning and Speech Acts 2 : Formal semantics of success and satisfaction*, Cambridge University Press.

[262] Vanderveken D., 1991, "Non-literal Speech-acts and conversational maxims". Dans E. Lepore et R. Van Gulick, *John Searle and his Critics*, Blackwell, p. 371-384.

[263] Veltman F., 1996, "Defaults in update semantics", *Journal of Philosophical Logic*, 25, p. 221-261.

[264] Vreeswijk G., 1993, *Studies in Defeasible Argumentation*, Ph.D. dissertation, University of Amsterdam.

[265] Wainer J., 1991, *Uses of non-monotonic logic in natural language understanding : Generalized implicatures*, Ph.D. Dissertation, The Pennsylvania State University, State College, PA.

[266] Walton, D.N. et Batten L.M., 1984, "Games, graphs and circular arguments" *Logique et Analyse*, 105, p. 133-164.

[267] Walton, D.N. et Krabbe E.C.W., 1995, *Commitment in Dialogue*, State University of New York Press.

[268] Williams, M., 1994, "Transmutations of knowledge systems". Dans *KR'94*, p. 619-629.

[269] Wilson D, 1975, *Presuppositions and Non-Truth Conditional Semantics*. New York, Academic Press.

[270] Wilson D. et Sperber D., 1979, "Remarques sur l'interprétation des énoncés selon Paul Grice", *Communications*, 30, p. 80-94.

[271] Woods J. et Walton D., 1977, "*Post hoc, ergo propter hoc*", *Review of Metaphysics*, 30, p. 569-593.

[272] Woods J. et Walton D., 1978, "The Fallacy of *ad ignorantiam*", *Dialectica*, 32, p. 87-99.

[273] Woods J. et Walton D., 1979, "Question-Begging and cumulativeness in dialectical games", *Nous*, 16, p. 585-606.

[274] Woods J. et Walton D., 1979, "Equivocation and practical logic", *Ratio*, 21, p. 31-43.

[275] von Wright G. H., 1951, "Deontic logic", *Mind*, 60, p. 58-74.

[276] von Wright G. H., 1964, "A new system of deontic logic", *Danish Yearbook of Philosophy*, **1**, p. 173-182. Réédité dans R. Hilpinen (ed), *Deontic Logic : Introductory and Systematic Readings*, Dordrecht : D. Reidel, 1971, p. 105-120.

[277] von Wright G. H., 1965, "And next", *Acta Philosophica Fennica*, 18, p. 293-304.

[278] von Wright G. H., 1968, *An Essay on Deontic Logic and the General Theory of Action*, North-Holland : Amsterdam.

[279] Zadeh L.A., 1982, "A note on the theory of prototypes", *Cognition*, 12, p. 291-297.

[280] Zanardo A., 1996, "Branching-time logic with quantification over branches", *Journal of Symbolic Logic*, 61, p. 1-39.

Table des matières

ABSTRACT
Over the past decades, argumentation theorists have been increasingly interested in the study of formal logic, and the logical interest in argumentation has also increased. Among the numerous factors that have contributed to this, one might mention the emergence of so-called non-monotonic logics in the late 1980s. Such logics aim to capture a notion of necessity that is contextual, hence providing a more flexible tool for the analysis of argumentative practices. Current research programs in this area tend to fall into three main groups : those focusing on specific argumentative schemes, those dealing with the interface between semantics and pragmatics, and those developing a general theory of how arguments interact.

This dissertation is an attempt to discuss some aspects of these programs. It is also an attempt to appreciate the extent to which deontic logic can be relevant to the study of argumentation. I focus here on the preference-based semantics for conditional obligation as initiated by Hansson and Lewis. The basic idea is to replace the Kripke-type relation of accessibility by a preference relation. I also focus on the notion of remedial interchange that was first introduced by Goffman in the literature on conversational interaction. My emphasis is not on new formal results, but rather on introducing a way of thinking about contrary-to-duty (reparational) obligation that is slightly different from more familiar ones. Close examination reveals that revision has an important role to play, just as aspects of time.

TITLE
Non-monotonic logics and modes of argumentation

KEY-WORDS
Non-monotonic logic, argumentation, deontic logic, preference-based semantics, conversational interaction, remedial interchange, conditional obligation, temporal logic.

LABORATORY
Centre d'Epistémologie et d'Ergologie Comparative CEPERC - UMR 6059 CNRS
29 av Robert Schumann 13621 Aix-en-Provence

RESUME
L'intérêt des logiciens pour l'étude de l'argumentation s'est considérablement accru durant la dernière décennie. Inversement, l'intérêt des théoriciens de l'argumentation pour l'étude des systèmes formels a lui aussi beaucoup grandi. La naissance des logiques non-monotones et des théories de la révision à la fin des années 1980 a largement contribué à ce rapprochement entre les deux domaines d'étude. Modélisant une notion de nécessité contextuelle et révisable, ces formalismes fournissent un outil qui semble particulièrement bien convenir à l'analyse du discours argumentatif. Les programmes de recherches tendent à se répartir en trois catégories. Certains étudient des schémas d'argumentation spécifiques. D'autres s'intéressent à l'interface entre la sémantique et la pragmatique. D'autres enfin cherchent à élaborer une théorie générale de l'interaction entre arguments. L'un des objectifs de ce travail est de décrire, dans ses grandes lignes, les composantes de ces programmes, ainsi que leurs implications. Nous tentons également de voir si, et dans quelle mesure, la logique déontique peut contribuer à l'étude de l'argumentation. Nous axons l'analyse sur les sémantiques de l'obligation conditionnelle, qui entretiennent une étroite affinité avec les sémantiques non-monotones. L'idée essentielle consiste ici à remplacer la relation d'accessiblité, que Kripke utilisait, par une relation de préférence. Nous tentons, pour l'essentiel, de donner un éclairage nouveau à certaines des difficultés auxquelles la théorie logique se heurte, lorsqu'elle cherche à rendre compte de ce que les théoriciens de la conversation ont coutume de nommer « l'échange réparateur ». Un examen attentif révèle que la révision y joue un rôle, de même que la dimension du temps.

TITRE
Logiques non-monotones et modes d'argumentation

FORMATION DOCTORALE
Philosophie

MOTS-CLEFS
Logique non-monotone, argumentation, logique déontique, sémantique préférentielle, interaction conversationnelle, échange réparateur, obligation conditionnelle, logique temporelle.

LABORATOIRE
Centre d'Epistémologie et d'Ergologie Comparative CEPERC - UMR 6059 CNRS
29 av Robert Schumann 13621 Aix-en-Provence