No d'ordre : H209

USTL

LIFL Laboratoire d'Informatique Fondamentale de Lille

CNRS

Modèles et calcul du raisonnement dynamique

Mémoire d'Habilitation à diriger des recherches en Sciences Mathématiques présenté par

Ramón Pino Pérez

Le 4 février 1999

Jury :

Rapporteurs :

Philippe Besnard,	Directeur de recherches au CNRS, IRIT, Univ. de Toulouse III
Michael Freund,	Maître de conférences-HDR à l'Université d'Orléans
Karl Schlechta,	Professeur à l'Université d'Aix-Marseille I

Autres membres :

Jean-Paul Delahaye,	Professeur à l'Université de Lille I
Didier Dubois,	Directeur de recherches au CNRS IRIT, Univ. de Toulouse III
Eric Grégoire,	Professeur à l'Université de l'Artois

A Charlotte et Juan.

Demos tiempo al tiempo:
para que el vaso rebose
hay que llenarlo primero.

Antonio Machado

Table des matières

Remerciements

Je voudrais remercier vivement Hassan Bezzazi, Stéphane Janot, Sébastien Konieczny, David Makinson et Carlos Uzcátegui. Le travail en collaboration avec eux a été décisif pour les résultats présentés dans ce mémoire. Les discussions fructueuses, animées et pleines d'enthousiasme ont fait avancer les idées rapidement. Travailler avec eux a été une vraie joie.

A Michael Freund et à Karl Schlechta je dois un triple remerciement. D'abord pour l'intérêt qu'ils ont porté à mes travaux. Ensuite pour les critiques et remarques pertinentes qu'ils ont faites de certains de mes manuscrits. Cela m'a aidé à comprendre mieux certains aspects de mon travail, à améliorer certains résultats et m'a donné des nouvelles pistes pour poursuivre mes recherches. Enfin je les remercie d'avoir eu la gentillesse d'accepter d'être rapporteurs de ce mémoire et de leur participation au jury.

Un grand merci à Phillipe Besnard pour s'être intéressé à mes travaux, d'avoir accepté d'être rapporteur et de sa participation au jury.

A Didier Dubois je le remercie pour ses conseils et aussi d'être membre du Jury. Les discussions que nous avons eues ont été brèves mais elles m'ont beaucoup éclairé sur le rôle central de la logique préférentielle dans plusieurs aspects de l'intelligence artificielle.

A Eric Grégoire je le remercie pour sa participation au jury. Son approche pratique du domaine m'a beaucoup appris et suggéré de nouvelles voies de recherche.

A Jean-Paul Delahaye va ma reconnaissance sans limites. Il m'a fait l'honneur de m'accueillir dans son équipe, m'a donné toute sa confiance en me laissant pleine liberté dans mes recherches et m'a encouragé à poursuivre dans les moments de doute.

Envers David Makinson j'ai une dette infinie. Pour lui va un remerciement très spécial. Il s'est intéressé à mes premiers travaux dans le domaine. Ses encouragements ont été décisifs pour continuer dans la voie de mes recherches. Ses lectures attentives, ses critiques pertinentes et ses indications bibliographiques m'ont été d'une aide inappréciable.

A Daniel Lehmann je dois aussi des remerciements. C'est en écoutant une conférence passionante sur l'inférence non monotone qu'il a faite à Keele durant le Logic Colloquium'93 que j'ai été séduit par le domaine. Il m'a donné des références bibliographiques très pertinentes pour un débutant. Ensuite il a lu attentivement certains de mes travaux, en faisant des remarques très justes qui m'ont aidé à améliorer ou à corriger certains résultats.

Je ne voudrais pas oublier Jorge Lobo. Merci à lui pour ses lectures attentives, ses remarques judicieuses et pour les agréables discussions que nous avons eues.

Merci aux collegues du LIFL et de l'EUDIL qui par leur gentillesse et leur joie ont fait que le travail de ces années lilloises ait été très agréable.

Préface

J'ai longtemps hésité sur la forme que je devais donner à ce mémoire. Devais-je présenter tous mes travaux ou seulement une partie d'entre eux? Comme la logique a été une constante tout au long de mes recherches ce mémoire a failli s'appeler De la logique du calcul au calcul de la logique, *la première partie de ce titre faisant référence à la période que j'ai dédiée à l'étude du λ-calcul comme modèle logique de la programmation fonctionnelle et la deuxième partie faisant allusion aux logiques du raisonnement de la vie courante.*

Pourtant j'ai choisi de présenter in extenso seulement les travaux concernant la dernière période de mon activité scientifique fondamentalement pour les raisons suivantes :

1. *Garder un maximum de cohérence. Les chapitres présentés ici gardent une unité dans la thématique.*

2. *A cause des perspectives. Ces chapitres posent des questions et ouvrent la voie à des développements futurs qui semblent assez prometteurs.*

Ce mémoire est donc essentiellement la réunion de 5 articles dans le domaine des logiques non monotones et la théorie de la révision des croyances. Ces travaux correspondent aux 4 dernières années des mes recherches. Dans l'appendice D on trouvera un historique des publications constituant le cœur de ce travail.

Bien que j'aie delibéremment omis de présenter in extenso mes recherches concernant le λ-calcul et la partialité on trouvera dans le chapitre 1 une rapide trace de mon itinéraire scientifique. Dans l'appendice E on pourra trouver mon Curriculum Vitae dans une forme standard.

Chapitre 1

Itinéraire scientifique

Les origines

Vers la fin des mes études de la licence de mathématiques en 1978 à Caracas, une question revenait sans cesse qui se déclinait sous plusieurs formes : quelle était la validité de toutes les théories étudiées? Quelle était la validité des raisonnements employés? Ainsi pour essayer de trouver une réponse à ces questions de fondement je me suis tourné naturellement sur la logique. Ceci avec autant plus de force que je n'avais pas eu l'occasion de suivre des cours de logique et qu'il me semblait important de comprendre certains problèmes dont je n'avais que vaguement entendu parler : la cardinalité du continuum et l'incomplétude de l'arithmétique. J'ai dû commencer par le commencement guidé par le Dr. Carlos Di Prisco, maître patient et exemplaire à qui je dois le goût pour la recherche.

J'ai ensuite suivi des études de troisième cycle à l'Université de Paris 7. D'abord un DEA de logique sous la direction du Professeur Gabriel Sabbagh (1980) et ensuite j'ai preparé une thèse de troisième cycle sous la direction de Jean-Pierre Ressayre. Je me suis intéressé à l'étude d'une hiérarchie de théories des ensembles faibles où le théorème de récursion n'est valide que jusqu'à une certaine complexité des formules. Nous avons établi des résultats d'existence de certaines extensions élémentaires [94] et des résultats d'indépendance [87] à l'aide des indicatrices dans le style de Kirby et Paris. Ces résultats et quelques autres sont réunis dans mon mémoire de thèse soutenue en 1983 [86].

La période allant de 1983 à 1985 fut une époque de mutation dans ma recherche. En effet je travaillais à cette période comme professeur à l'Université Simón Bolívar à Caracas. De par ma spécialité en logique j'ai dû enseigner un cours sur les fondements logiques de l'intelligence artificielle aux étudiants du Mastère en Informatique. Le contenu de ce cours portait sur le λ-calcul, le principe de résolution de Robinson et les logiques modales. Ces thèmes qui ne m'etaient pas familiers je les découvrai avec beaucoup de plaisir. A ce moment-là j'ai voulu approfondir davantage sur les fondements de l'informatique, des langages de programmation et du calcul.

La logique du calcul

En 1985 je retourne à Paris avec l'intention de travailler avec des spécialistes des langages fonctionnels. A Paris il y avait à l'époque une équipe (FORMEL) de renommée internationale réunissant des chercheurs de l'Inria (G. Huet), de Paris 7 (G. Cousineau) et de l'Ecole Normale

Supérieure (P.-L. Curien) dont le thème était justement les langages fonctionnels tant d'un point de vue pratique que théorique. Je commençai à travailler avec Pierre-Louis Curien autour du λ-calcul comme un outil pour raisonner sur l'équivalence des programmes. En 1986 je présentai un travail sur le λ-calcul polymorphe comme mémoire de DEA d'Informatique[1] [88].

Un des problèmes importants quand on raisonne sur des programmes (et sur des calculs) est de savoir quand deux programmes doivent être considérés comme égaux. Nous pensons que le comportement entrée-sortie ne suffit pas. En particulier il y a des programmes qui ne terminent pas et qui ont des comportements assez différents. Par exemple un programme peut boucler sans saturer la mémoire tandis qu'un autre la saturera. Le λ-calcul comme paradigme des langages fonctionnels permet de comprendre de façon assez claire ces phénomènes. Nous nous sommes penchés sur l'étude des extensions du λ-calcul qui permettent d'étudier l'équivalence de programmes de manière assez fine. Nous avons ainsi étudié le λ-calcul partiel et des λ-calculs partiels typés.

L'idée essentiel du λ-calcul partiel est de considérer qu'un programme est un couple : un λ-terme plus une mémoire qui sert à garder la trace du calcul. Nous introduisons des règles pour manipuler la mémoire. En particulier lorsqu'un programme termine la mémoire doit être vide. Le calcul que nous proposons est proche de celui de Moggi [75].

Nous montrons que ce calcul a une présentation équationnelle (sémantique algébrique) [89]. Nous montrons aussi que la partie du calcul qui concerne la manipulation de la mémoire est décidable [90]. Nous remarquons qu'aucune technique standard ne marche dans ce cas et que ce problème était ouvert depuis plusieurs années.

Une autre question que nous avons résolue est celle de savoir quel est le modèle opérationnel standard de ce calcul. Nous prouvons que c'est le λ-calcul par valeur paresseux. C'est un résultat qui utilise un lemme de contexte loin d'être trivial. Nous donnons une preuve de ce résultat dans [91].

Ces résultats constituent l'essentiel de mon mémoire de thèse en informatique [92].

En 1988 je commence à travailler à l'Université de Lille I. A partir de ce moment je commence à diriger une petite équipe au sein du Laboratoire d'Informatique Fondamentale de Lille autour des thèmes en rapport avec le λ-calcul tout en conservant les rapports que j'avais avec l'équipe FORMEL, notamment la participation à un projet européen sur le λ-calcul typé.

J'ai dirigé pendant la période de 1988 à 1993 trois mémoires de DEA et une thèse. D'abord en 89 le mémoire de Gabriel Desmet sur des implantations du λ-calcul par valeur (dont une stratégie s'avérera plus tard être le modèle naturel du λ-calcul partiel). Ensuite les mémoires d'Olivier Dubuisson et Thiery Peltier : une étude théorique et pratique du λ-calcul avec substitutions explicites (un raffinement du λ-calcul partiel introduit par Curien). Finalement la thèse de Christian Even dont la problématique générale était l'étude des liens entre le λ-calcul partiel, les algèbres combinatoires partielles et la théorie des catégories, ainsi que la confluence du calcul.

Parmi les résultats saillants de cette période notons l'équivalence entre le λ-calcul partiel et les algèbres combinatoires partielles, question soulevée par Moggi et ouverte jusqu'à notre travail [25]. La confluence du λ-calcul partiel paresseux [93] et l'extension du théorème de Scott-Koymans au cadre partiel [26]

[1] J'ai aussi dû passer quelques examens!

Le calcul de la logique

A partir de la rentrée 93 suite à une politique du LIFL consistant à renforcer les équipes de taille importante j'intègre l'équipe Méthéol (Méthodes et outils logiques pour la programmation) dirigée par Jean-Paul Delahaye. Je m'intéresse particulièrement à des aspects liés au raisonnement non monotone. Depuis lors je dirige un groupe de travail[2] intégré en plus de moi même par Hassan Bezzazi et Stéphane Janot, Maîtres de conférences à l'Université de Lille II et Lille I respectivement, de Sébastien Konieczny, thésard BDI CNRS-Région, dont je dirige la thèse. En plus de ce "noyau dur", divers étudiants de DEA ont travaillé avec nous. Ainsi j'ai dirigé les mémoires de DEA de :

- Omar Malah. Sur l'algorithme de calcul de la clôture rationnelle et ses implantations. Université de Lille 1. 1994.
- Sébastien Konieczny. Sur une archictecture pour la coopération en utilisant des opérateurs de changements de croyances. Université de Lille 1. 1996.
- Christophe Parent. Une étude des algorithmes pour la recherche des explications. Université de Lille 1. 1997.

Pendant cette période nous avons eu des collaborations internationales et nationales. En particulier avec le Professeur Carlos Uzcátegui de l'Université des Andes (Mérida, Venezuela), avec le Professeur Jorge Lobo de Bell Labs (New-York, anciennement Professeur à l'université d'Illinois), avec le Dr. David Makinson. Nous participons au nouveau PRC I3, groupe Modèles du raisonnement. Nous participons au programme de la région Ganymède dont le thème central est la coopération. Nous avons aussi des rapports avec le Centre de Recherches en Informatique de Lens (CRIL) qui participe également au programme Ganymède.

Mes travaux de cette période sont le cœur de ce mémoire. Ils sont présentés dans le chapitre suivant.

[2]voir l'appendice F et aussi http://www.lifl.fr/GNOM

Chapitre 2

Présentation

2.1 Généralités

Ce mémoire réunit quelques contributions au domaine de l'intelligence artificielle. Particulièrement nous nous intéressons aux fondements logiques du "raisonnement du sens commun". Nous aborderons ici quelques aspects d'essentiellement deux domaines :

- Raisonnement non monotone
- Théorie du changement de la connaissance

Dans beaucoup de situations de la vie courante nous manipulons des informations incomplètes, approximatives voire contradictoires et pourtant les conclusions que nous pouvons extraire sont cohérentes. De même nous devons incorporer de nouvelles informations, qui pourraient contredire notre connaissance préalable, tout en gardant la cohérence de l'ensemble actuel des connaissances. Illustrons ces deux types de situations avec des exemples très simples.

Exemple 2.1 Supposons qu'un agent rationnel est en train de raisonner sur les *bons cafés*. Cet agent aime les cafés sucrés. Ainsi il pense que si l'on ajoute du *sucre* au café il sera *bon*. Or cet agent n'a pas le goût trop denaturé et lorsque l'on ajoute du sucre et du *poivre* il va trouver que le café n'est pas bon. Il est très facile de noter que classiquement on obtient des contradictions lorsque l'on est en présence d'un café sucré et poivré en même temps car il sera bon et mauvais en même temps. Pourtant l'agent rationnel de notre exemple ne conclut pas n'importe quoi car "sa logique" n'est pas classique.

Pour modéliser ce genre de raisonnement dans lequel les conséquences ne sont pas forcément préservées par ajout de nouvelles hypothèses ont surgi des formalismes dit non monotones.

Exemple 2.2 Supposons maintenant un autre agent avec un état de connaissances du monde qui se résume à deux faits : le café est *bon*, le café est *sucré*; et à une règle : si un café est sucré et poivré alors il n'est pas bon. Or peu après il apprend que le café est poivré. Comment peut-il changer sa connaissance dans le but d'incorporer cette nouvelle information, garder la cohérence et en même temps perdre le moins possible des informations qu'il possedait ? Plusieurs solutions peuvent se présenter, par exemple le nouvel état des connaissances pourrait être que le café est sucré, bon et poivré (pas de règle) ou bien que le café est bon, poivré et la règle ou bien d'autres alternatives.

Modéliser la rationalité de ces mécanismes de prise en compte de nouvelles informations est le but de la théorie de changement.

2.1.1 Sur le raisonnement dynamique

Il est bien connu que la logique classique est monotone, *i.e.* lorsque l'on considère deux ensembles d'hypothèses Σ et Σ' si Σ est contenu dans Σ' alors les conséquences de Σ seront aussi contenues dans les conséquences de Σ'. Si nous dénotons par $Cn(\Gamma)$ l'ensemble des conséquences classiques de Γ la propriété précédente peut s'ecrire :

$$\Sigma \subseteq \Sigma' \Rightarrow Cn(\Sigma) \subseteq Cn(\Sigma')$$

ce qui veut dire que l'opérateur Cn est monotone par rapport à l'inclusion. Autrement dit si l'on ajoute des hypothèses toute conséquence ancienne doit rester; on ne peut pas changer les inférences par ajout d'une nouvelle information. Dans ce sens la logique classique est statique.

Dans les problèmes qui nous occuppent cette propriété est absente. On doit, au contraire, comme on l'a vu à travers les exemples changer les "vieilles" inférences ou modifier les connaissances. C'est pour cela que nous parlerons des logiques du raisonnement dynamique pour englober les phénomènes que nous étudions. On verra plus tard qu'il y a des raisons plus profondes pour inclure le raisonnement non monotone dans les logiques dynamiques.

2.1.2 Sur notre approche

Nous avons déjà dit que ce qui nous occupe est le raisonnement dynamique, *i.e.* le raisonnement non monotone et la théorie du changement.

Concernant le raisonnement non monotone deux thèmes sont étudiés :

- Les mécanismes de *déduction* de conséquences à partir d'un ensemble des connaissances.
- Les mécanismes de recherche des *explications* à partir d'un ensemble d'informations.

Concernant la théorie du changement de la connaissance deux thèmes sont aussi étudiés :

- Coment changer l'information de façon cohérente lorsqu'une nouvelle information arrive avec *priorité* pour la nouvelle information.
- Comment extraire une information cohérente de plusieurs sources d'information *également* prioritaires.

Nous utilisons une approche logique dans l'étude des problèmes qui nous concernent. En particulier notre langage sera celui de la logique propositionnelle. Néanmoins ce n'est pas dans les buts de ce mémoire de prouver la pertinence des approches logiques à l'intelligence artificielle. Depuis bientôt 40 ans beaucoup de travaux ont été faits [77] montrant que les approches logiques sont bien féconds en au moins trois aspects :

- en tant qu'outil d'analyse,
- en tant qu'outil de représentation et
- en tant qu'outil de calcul.

Nous pensons que notre approche, qui continue d'autres approches logiques, a des apports dans ces trois aspects.

Beaucoup de cadres logiques ont été proposés pour résoudre les problèmes liés à la non monotonie. On peut citer les logiques non monotones modales [74, 76, 11], les logiques des défauts [97, 6, 72], la circonscription [73], les systèmes d'héritage [111], les logiques préférentielles [49], des logiques probabilistes [83, 38], des logiques possibilistes [21]. Une étude comparative de ces formalismes est hors des buts de ce mémoire. On peut consulter [13] pour une étude comparative de la plupart de ces logiques.

Pour ce qui concerne les parties de notre travail ayant trait au raisonnement non monotone nous avons choisi l'approche de l'étude systématique des relations binaires $\mid\!\sim$ entre des formules vues comme des relations de conséquence non monotones. Cette étude est faite via les propriétés structurelles de la relation, les "règles d'inférence". Ceci est l'approche des relations de conséquence non monotones suivant la tradition de Kraus, Lehmann et Magidor [49, 55] et Makinson [67, 69].

Pour ce qui concerne les parties de notre travail ayant trait à la théorie du changement nous avons choisi le paradigme de Alchourrón, Gärdenfors et Makinson [1].

Une des caractéristiques communes à ces deux approches réside dans l'étude abstraite des propriétés syntaxiques assez générales aussi bien pour les relations de conséquence non monotones que pour les mécanismes de changement de la connaissance. Cela permet d'analyser des classes des relations de conséquence, de les comparer en établissant les propriétés communes, les différences, voire certaines hierarchies. Il en va de même pour les mécanismes de changement. Un point intéressant qui découle évidemment de la géneralité de ces approches est qu'elles peuvent être utilisées pour extraire des propriétés d'une logique non monotone particulière ou d'un mécanisme de changement particulier. Ainsi par exemple supposons que l'on considère une relation de conséquence $\mid\!\sim$ associée à une logique non monotone particulière. Supposons aussi que l'on arrive à prouver que la relation $\mid\!\sim$ satisfait les règles préférentielles (voir chapitre 3) alors on aura automatiquement certaines règles nouvelles pour la logique que l'on étudie, *v.g.* on pourra dire que toute règle dérivée de la logique préférentielle sera valable pour la logique que l'on considère. Similairement si l'on considère un opérateur de changement dont on prouve qu'il est un opérateur de révision on aura automatiquement que l'opérateur particulier satisfait toute propriété associée aux opérateurs de révision. Bien que cette observation puisse paraître triviale elle est fort utile lorsque la logique non monotone que l'on étudie ou l'opérateur ont des définitions complexes et sont d'une manipulation ardue.

Une autre caractéristique commune des deux approches est qu'elles possèdent des sémantiques très propres et d'une certaine façon voisines. La proximité des sémantiques est une des voies qui permet de voir d'une manière simple les rapports étroits entre les relations de conséquence non monotones et les opérateurs de changement.

Par ailleurs notons que l'uniformité de ces sémantiques permet une étude assez systématique des relations de conséquence ainsi que des opérateurs de changement. En fait on peut penser d'une façon approximative que les sémantiques sont données par des ordres de formes diverses. Or des "déformations" sur les ordres vont sûrement produire des changements dans la syntaxe. Peut-on capturer complètement ce type de "déformations" par des propriétés structurelles ? Réciproquement des changements dans la syntaxe pourront-ils être capturés par des "déformations" de la structure ordonnée ?

De plus la simplicité des sémantiques donne une vision assez claire des propriétés que l'on étudie. En même temps cela aide énormement à la recherche de contre-exemples à des propriétés syntaxiques, voire à les prouver clairement et simplement là où les preuves au niveau de la syntaxe sont obscures et complexes.

Un autre aspect important des sémantiques des deux approches est qu'elles donnent une vision unificatrice du raisonnement dynamique que nous étudions. En effet les sémantiques sont *grosso modo* des ordres sur des modèles (valuations) qui nous disent quels sont les bons modèles d'une formule. Ainsi les conséquences (non monotones) d'une formule α seront les formules vraies classiquement dans les bons modèles de α. De même pour les mécanismes de changement : les formules résultantes du changement seront celles qui sont vraies dans les bons modèles d'une certaine formule (l'information nouvelle pour le cas de la révision). Ces similitudes au niveau de la sémantique permettent d'établir des liens assez précis [70, 31] entre les relations de conséquence et les mécanismes de changement de la connaissance.

2.2 Les problèmes abordés : un peu d'histoire et motivations

Un effort considérable a été fait depuis les années 80 pour trouver d'abord des formalismes qui rendent compte de manière assez précise des problèmes liés au raisonnement dynamique et ensuite pour trouver un cadre unificateur. Le travail de D. Makinson [69] concernant des propriétés communes de plusieurs formalismes non monotones est particulièrement important et nous a beaucoup influencé. Ainsi dans ce mémoire nous abordons l'étude des mécanismes du raisonnement dynamique d'un point de vue général des formalismes : nous examinons les propriétes structurelles d'une logique via les propriétés de la relation de conséquence qu'elle induit. Cela est une tradition qui remonte à Tarski [109] et qui a été developpée très finement par Gentzen [39] avec son calcul des séquents. Cette approche qui s'est révélée très fructueuse en informatique (au moins dans deux domaines : la théorie des types et la mécanisation du raisonnement) a été reprise par Gabbay [35] (à ma connaissance le premier) pour étudier les logiques non monotones. Il propose un ensemble de règles dans le style du calcul des séquents comme étant l'ensemble minimal des règles qu'une logique non monotone doit posséder. Ces règles sont un sous-ensemble de ce qu'aujourd'hui on appelle la logique cumulative [67] (voir chapitre 3).

A la fin des années 80 on a commencé à établir des sémantiques dites préférentielles pour le raisonnement non monotone. Les premiers travaux dans ce sens sont ceux de Shoham [105, 106], ceux de Makinson [67] et Kraus, Lehmann et Magidor [48]. L'idée des structures sémantiques considérées (préférentielles) est intuitivement la suivante : une structure sémantique $\mathcal{M}$ est un ordre $(W, <)$ où W est un ensemble des modèles d'une logique donnée (prenons la logique classique pour simplifier), c'est-à-dire que l'on dispose d'une relation de conséquence sémantique $\models$ telle que l'on peut déterminer si $M \models \alpha$ pour tout M dans W et toute formule α. L'ordre $<$ représente une relation de préférence ou de plausibilité entre structures. Ainsi $M < N$ signifiera que M est préféré à ou plus plausible que N. Finalement on dit que $\mathcal{M}$ satisfait $\alpha \mathrel|\joinrel\sim \beta$ si les modèles préférés de α sont aussi modèles de β, plus précisément si

$$min(\alpha) \subseteq mod(\beta)$$

où

$$min(\alpha) \stackrel{def}{=} \{M \in W : M \models \alpha \text{ et } N < M \Rightarrow N \not\models \alpha\}$$
$$mod(\beta) \stackrel{def}{=} \{N \in W : N \models \beta\}$$

Les premiers résultats de complétude pour ce genre de sémantique sont de Makinson [67] et de Kraus, Lehman et Magidor [49]. Ici il faut dire que les sémantiques considérées ne sont pas exactement les mêmes. Makinson considère une sémantique très proche de celle que l'on vient de décrire plus haut mais sans exiger que la relation $<$ soit un ordre. Par contre Kraus, Lehmann et Magidor considèrent bien des ordres mais pas directement sur des valuations mais sur un ensemble $\mathcal{S}$ de points qu'ils appellent des *états* et une application $\imath$ de cet ensemble d'états dans les valuations. Ainsi les structures KLM (pour Kraus-Lehmann-Magidor) sont de la forme $\langle \mathcal{S}, \imath, < \rangle$ avec $(\mathcal{S}, <)$ un ordre et pour ces structures on définit $min(\alpha)$ et $mod(\beta)$ de la façon suivante

$$min(\alpha) \stackrel{def}{=} \{s \in \mathcal{S} : \imath(s) \models \alpha \text{ et } s' < s \Rightarrow \imath(s') \not\models \alpha\}$$
$$mod(\beta) \stackrel{def}{=} \{s \in \mathcal{S} : \imath(s) \models \beta\}$$

Comme auparavant, une structure KLM satisfera $\alpha \mid\!\sim \beta$ si $min(\alpha) \subseteq mod(\beta)$.

Nous remarquerons que ces deux types d'approches ont leurs sources dans la sémantique des "mondes possibles" de Kripke utilisée pour étudier les logiques modales. Plus précisément la première approche se rapporte à [51] et la deuxième à [50]. Dans [19] des relations entre les deux approches sont étudiées, en particulier comment voir une structure KLM comme une structure MAK (pour Makinson), plus simple, mais avec une relation de plausibilité moins bien structurée. Nous reviendrons sur ce problème dans la section 2.3.2.

Un des aspects importants de l'approche de la sémantique KLM est qu'elle permet une étude systématique des relations de conséquence non monotones. En effet dans [49] est prouvé un théorème de représentation pour les relations préférentielles (voir la section 3.2) : une relation $\mid\!\sim$ est préférentielle ssi il existe un modèle préférentiel $\mathcal{M} = \langle \mathcal{S}, \imath, < \rangle$ qui représente $\mid\!\sim$, *i.e.*

$$\alpha \mid\!\sim \beta \Leftrightarrow min(\alpha) \subseteq \mathrm{mod}(\beta) \tag{2.1}$$

où un modèle préférentiel est une structure KLM avec une propriété additionnelle pour l'ordre $<$, la propriété de "smoothness" (voir 3.2). Ainsi, une question naturelle qui se pose quand on considère une relation de conséquence qui satisfait certaines propriétés est la suivante : quelle sont les propriétés sur $<$ qui caractérisent $\mid\!\sim$ au sens de l'équivalence 2.1 ? Cette question a été étudiée avec succès pour les relations suivantes :

- Les relations rationnelles [55],
- les relations disjonctives rationnelles [30],
- les relations préférentielles qui satisfont la règle de la rationalité de la négation [32],
- les relations rationnelles transitives [9] et
- d'autres relations qui satisfont des règles plus fortes que la monotonie rationnelle [7]

Le chapitre 3 est justement la présentation de ces derniers résultats.

Un des points importants que l'on utilise pour développer des arguments au niveau de la sémantique d'une façon simple est de se rendre compte que beaucoup de relations de conséquence admettent des représentations injectives. Ceci signifie qu'il y a un modèle préférentiel qui les définit où l'on peut identifier un état à une seule valuation. C'est par exemple le cas des relations rationnelles. Ainsi suite à différents théorèmes de représentation [49, 55, 30, 7] nous avons essayé dans le chapitre 4 de donner une méthode uniforme et directe pour la construction d'un modèle qui représente une relation de conséquence. La méthode est uniforme puisque la même construction couvre les cas allant des relations préférentielles avec une propriété additionnelle introduite par Freund [30] jusqu'aux relations rationnelles transitives. La méthode est directe car on définit l'ordre sur les valuations juste avec la relation de conséquence que l'on veut représenter. Ceci est à comparer aux autres méthodes utilisées pour prouver les théorèmes de représentation où l'on définit d'abord une relation d'ordre entre les formules d'où l'on extrait une relation d'ordre sur les valuations.

Sur la logique de l'abduction, c'est-à-dire les mécanismes de recherche des explications à une observation donnée, beaucoup d'efforts ont été faits depuis les travaux de Peirce [84]. En particulier récemment l'approche *relationnelle* (*i.e.* l'étude des propriétés d'une relation binaire entre des formules) de Flach [29] s'est avérée très éclairante. Il a proposé une série de règles qu'une relation abductive doit posséder pour avoir un bon comportement. La justification des postulats qu'il donne est intuitive. Notons aussi qu'il prouve un théorème de représentation pour ce genre de relations, qui reflète le fait qu'il doit y avoir des liens plus profonds entre les relations abductives et les relations de conséquence non monotones. Nous avons poussé l'étude des relations abductives dans cette direction dans le chapitre 5 en montrant qu'il y a une dualité entre les relations de conséquence non monotones et les relations d'abduction. La dualité que l'on trouve peut être interprétée comme un fondement à l'affirmation que le raisonnement sur les explications, l'abduction, est de la *déduction non monotone à l'envers.*

Ceci est à comparer avec le travail de Cialdea-Mayer et Pirri [66] intitulé "Abduction is not Deduction-in-Reverse". Le cadre que l'on considère est plus restreint que le leur. En particulier elles considèrent une relation de préférence pour choisir les bonnes explications. Nous montrons, à l'aide des techniques qui s'inspirent de nos théorèmes de représentation, que cette relation est en fait implicite dans les règles qu'une relation abductive doit posséder.

Concernant les mécanismes de changement de la connaissance, un des aspects fondamentaux de l'approche d'Alchourrón, Gärdenfors et Makinson [1] est qu'elle capture au niveau des postulats l'idée de changement minimal : lorsque K est une théorie (un ensemble clos par déduction logique) qui est notre état de croyances et α est une nouvelle information on notera $K * \alpha$ la théorie qui résulte de réviser K par α selon l'opérateur $*$ (la méthode qui décrit comment changer les connaissances). $K * \alpha$ sera la théorie de $K \cup \{\alpha\}$ lorsque la nouvelle information n'entre pas en conflict avec l'ancienne, *i.e.* si $K \cup \{\alpha\}$ est consistant. Dans le cas contraire, $K * \alpha$ sera la théorie de $K' \cup \{\alpha\}$ avec $K' \subseteq K$ sélectionnée par la méthode qui est supposée choisir le K' le "plus proche possible" de K. Grace aux théorèmes de représentation des opérateurs qui obéissent aux postulats AGM on peut voir qu'il y a une grande variété des notions de "proximité" qui nous permettent de choisir K'. Or dans tous les cas, quelle que soit la méthode pour choisir K' on doit savoir au préalable si $K \cup \{\alpha\}$ est consistent. Cela revient à résoudre une instance du problème de la satisfaction. Il est bien connu que même dans le cas fini la complexité de ce problème est très grande car il est NP-complet. Par ailleurs beaucoup de méthodes pour calculer K' sont aussi de très haute complexité algorithmique (voir [23]). Dans le souci de pallier à ces problèmes de complexité nous avons proposé dans le chapitre 6 quelque méthodes syntaxiques pour réviser non des théories mais des "bases de connaissances" pour lesquelles la façon de représenter la connaissance est très importante. D'autre part nous affaiblissons les mécanismes de déduction ainsi que la forme des formules que nous considérerons pour rendre les calculs traitables. Dans cette "logique restreinte" nous définissons des mécanismes de changement apparentés à la *full meet contraction*. Par ailleurs en nous inspirant des liens entre opérateurs de révision et relations rationnelles, nous définissons d'autres mécanismes syntaxiques de révision toujours dans une "logique" très affaiblie. Nous donnons une méthode générale pour relativiser les postulats AGM pour la révision et ceux de Katsuno et Mendelzon pour la mise-à-jour au cadre logique faible dans lequel on travaille.

La plupart des mécanismes de changement considérés jusqu'à présent concernent le changement d'une information ancienne à la lumière d'une information nouvelle en donnant priorité à la nouvelle information. En fait peu importe de savoir qui est la nouvelle ou la vieille information ce qui est crucial est de savoir quelle information est prioritaire. Or que faire quand les deux informations sont également prioritaires ? Ou encore, que faire si au lieu de deux sources d'information nous disposons de n sources d'information également prioritaires avec $n > 2$? Quelques éléments de réponse ont été donnés par Revesz [99, 100] et par Liberatore and Shaerf [58, 59]. Nous apportons aussi quelques élements de réponse dans le chapitre 7 en utilisant l'approche structurelle AGM, c'est-à-dire en présentant une série de postulats sur la rationalité de l'information produite par plusieurs sources d'information.

2.3 Panoramas

2.3.1 Panorama sur le chapitre 3

Il est bien connu qu'en présence des règles préférentielles, système P (voir section 3.2), la transitivité est équivalente à la monotonie [33]. Une question naturelle est donc la suivante : quelle règle nous permettra d'avoir un peu de transitivité sans avoir la monotonie ? Une première réponse qui vient à l'esprit est d'imposer des conditions sur l'application de la transitivité (de même que la monotonie prudente et la monotonie rationnelle sont des applications restreintes de la monotonie). Dans cette optique une règle qui semble couler de source est la suivante :

$$\mathsf{RT}\quad \frac{\alpha {\mid\!\sim} \beta \quad \beta {\mid\!\sim} \gamma \quad \alpha {\mid\!\not\sim} \neg\gamma}{\alpha {\mid\!\sim} \gamma}$$

Cette règle appelée *transitivité rationnelle* a été introduite dans [9]. Elle nous dit que lorsque l'on est en situation d'appliquer la transitivité on a bien la conclusion attendue à moins que cette conclusion nous apporte une contradiction flagrante.

Notons qu'à la différence des règles préférentielles, la règle précédente est "non-Horn". En fait quand on l'énonce sous une forme positive elle est équivalente à :

$$\frac{\alpha {\mid\!\sim} \beta \quad \beta {\mid\!\sim} \gamma}{\alpha {\mid\!\sim} \gamma \text{ ou } \alpha {\mid\!\sim} \neg\gamma}$$

c'est-à-dire une règle non déterministe.

L'ajout de certaines règles non-Horn au système P a déjà été étudié dans la littérature. Par exemple l'ajout de la monotonie rationnelle (RM) a été étudié par Lehmann et Magidor [55] d'un point de vue sémantique et aussi calculatoire. L'ajout de la règle de la disjonction rationnelle (DR) a été étudié par Freund [30] d'un point de vue sémantique. L'ajout de la négation rationnelle (NR) a été étudié par Freund et Lehmann [32] d'un point de vue sémantique. Makinson dans [69] étudie aussi ces règles d'un point de vue syntaxique. D'ailleurs Makinson étudie aussi d'autres règles non-Horn, notamment la règle de *préservation de la détermination* :

$$\mathsf{DP} \quad \frac{\alpha \mathrel{|\!\sim} \beta \quad \alpha\wedge\gamma \not\mathrel{|\!\sim} \neg\beta}{\alpha\wedge\gamma \mathrel{|\!\sim} \beta}$$

Cette règle est une forme faible de monotonie. Nous prouvons qu'elle est équivalente à RT.

On savait que modulo le système P les implications suivantes sont vraies : DP $\Rightarrow$ RM, RM $\Rightarrow$ DR, DR $\Rightarrow$ NR. De plus que les réciproques des deux dernières implications ne sont pas vraies. Un des buts de ce chapitre est d'établir les relations précises entre ces règles et d'autres règles qui surgissent naturellement quand on considère RT. Un outil important est d'établir un théorème de répresentation correspondant aux relations préférentielles qui satisfont RT (alias DP).

Très naturellement liées à RT nous considérons les règles suivantes appelées respectivement *contraposée rationnelle* et *faible détermination* :

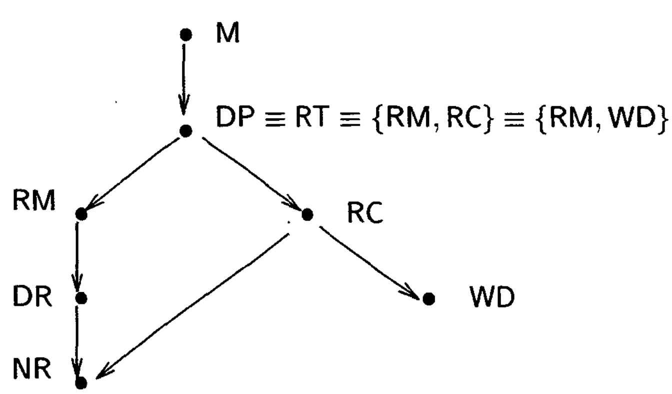
$$\mathsf{RC} \quad \frac{\alpha \mathrel{|\!\sim} \beta \quad \neg\beta \not\mathrel{|\!\sim} \alpha}{\neg\beta \mathrel{|\!\sim} \neg\alpha}$$

$$\mathsf{WD} \quad \frac{\top \mathrel{|\!\sim} \neg\alpha \quad \alpha \not\mathrel{|\!\sim} \beta}{\alpha \mathrel{|\!\sim} \neg\beta}$$

RC fut introduite dans [9] et WD fut formulée par Freund dans des communications avec l'auteur.

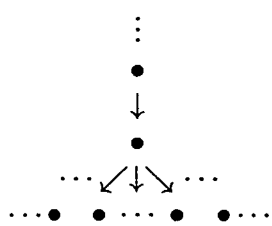
Nous prouvons que chacune de ces règles est équivalente à RT modulo le système P plus la règle RM.

Le premier résultat de ce chapitre peut être résumé dans le diagramme suivant qui nous donne les liens entre ces nouvelles règles :



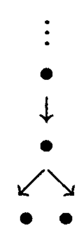
Plus précisément ce diagramme nous dit qu'une règle implique une autre si et seulement si l'on peut suivre un chemin avec les flèches de la première à la dernière. Les parties positives de ce résultat sont prouvées syntaxiquement. Les parties négatives sont prouvées sémantiquement.

Le résultat suivant dans ce chapitre est la preuve d'un théorème de représentation pour les relations préférentielles satisfaisant RT. Elles sont caractérisées par des modèles dits *quasi-linéaires*, *i.e.* l'ordre de l'ensemble d'états est de la forme suivante :



un état qui n'est pas minimal est comparable à tout autre état. Pour prouver ce résultat nous prouvons d'abord par des méthodes purement sémantiques que toute relation rationnelle peut être représentée par un modèle injectif où toute valuation satisfait les conséquences non monotones d'une formule. Ce résultat apparaissait déjà dans le travail de Freund [30] mais prouvé par des méthodes différentes.

Ce type de caracterisation sémantique mène naturellement à considérer certaines structures qui sont des restrictions ou des déformations des ordres quasi-linéaires. Par exemple on considère des ordres *presque-linéaires*, *i.e.* de la forme

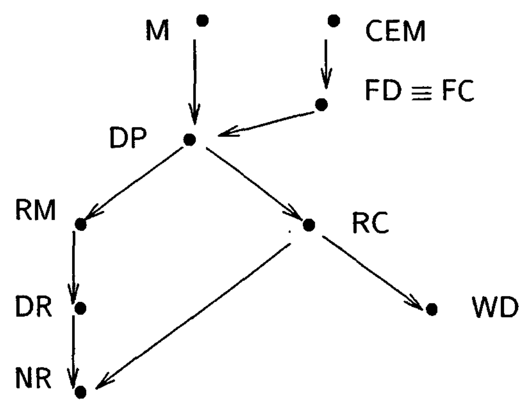


c'est-à-dire un ordre quasi-linéaire avec au plus deux états minimaux. Les modèles préférentiels avec ce type d'ordre sur les états caractérisent les relations satisfaisant les règles suivantes de *fragmentation disjonctive* et de *fragmentation conjonctive* :

$$\text{FD} \quad \frac{\alpha \mid\!\sim \beta\vee\gamma \quad \alpha \not\mid\!\sim \beta \quad \alpha \not\mid\!\sim \gamma}{\neg\beta \mid\!\sim \gamma}$$

$$\text{FC} \quad \frac{\alpha\wedge\beta \mid\!\sim \gamma \quad \alpha \not\mid\!\sim \gamma \quad \beta \not\mid\!\sim \gamma}{\alpha \mid\!\sim \neg\beta}$$

Sachant que les modèles préférentiels où l'ordre est linéaire caracterisent les relations préférentielles satisfaisant la règle du tiers exclu conditionnelle (CEM) (c'est un résultat implicite dans les travaux de Stalnaker et Lewis sur la logique des conditionnels, dont nous donnons une preuve assez simple) et qu'il n' y a pas de rapports entre la monotonie et les règles FD, FC et CEM, nous pouvons compléter notre premier diagramme de la façon suivante :



On doit remarquer que jusqu'à présent toutes les règles utilisées pour augmenter le système P sont non-Horn. Il semblerait que les règles Horn que l'on peut ajouter seraient moins utiles. Néanmoins nous introduisons un groupe des règles Horn plus faible que la monotonie et qui s'avèrent intéressantes. Ce sont les règles de *conjonction insistante*

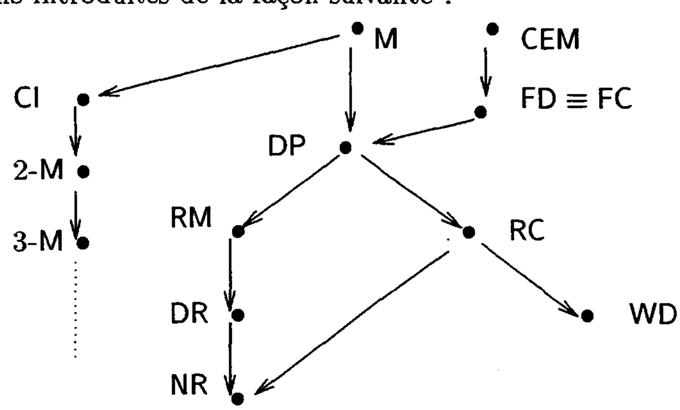
$$\mathsf{CI}\ \frac{\alpha\mid\!\sim\beta \quad \gamma\mid\!\sim\beta}{\alpha\wedge\gamma\mid\!\sim\beta}$$

et de *n-monotonie*

$$n\text{-}\mathsf{M}\ \frac{\alpha_1\mid\!\sim\sigma_1(\phi) \quad \ldots \quad \alpha_1\wedge\ldots\wedge\alpha_n\mid\!\sim\sigma_n(\phi)}{\alpha_1\wedge\ldots\wedge\alpha_n\wedge\alpha_{n+1}\mid\!\sim\sigma_n(\phi)}$$

où $\sigma_i(\phi)$ est ϕ si i est impair et $\neg\phi$ si i est pair et remarquons que la conclusion de la règle utilise σ_n au lieu de σ_{n+1}

Avec ces nouvelles relations nous prouvons que nous pouvons compléter le diagramme de liens entre les notions introduites de la façon suivante :



Nous finissons ce chapitre prouvant que les relations preférentielles qui satisfont la n-monotonie et la monotonie rationnelle se caractérisent par des modèles rangés avec au plus n rangs.

2.3.2 Panorama sur le chapitre 4

Nous avons déjà mentionné l'importance d'avoir des modèles préférentiels représentant une relation de conséquence où la structure est *simple*. Nous avons en particulier utilisé les représentations dites injectives où l'on peut identifier chaque état d'un modèle préférentiel à une valuation unique. Ceci fut exploité dans le chapitre 3. Ainsi une question qui vient naturellement est la suivante : existe-t-il une propriété logique, une règle, qui caractérise les relations préférentielles admettant des représentations injectives ?

Remarquons que Kraus, Lehmann et Magidor [49] ont noté qu'il y a des relations préférentielles qui n'admettent pas de représentations injectives. Schlechta a donné une esquisse de preuve de ce fait dans [104] et nous donnons une preuve in extenso de ce fait y compris dans le cas infini dans la section 3.2.

Un résultat intéressant concernant les relations de conséquence qui admettent des représentations injectives est dû à Freund [30]. Il prouve que, dans le cas fini, une relation préférentielle $\mid\!\sim$ admet une représentation injective si et seulement si elle satisfait la propriété suivante :

$$C(\alpha \vee \beta) \subseteq Cn(C(\alpha) \cup C(\beta))$$

Nous appelons cette propriété W-DR (weak disjunctive rationality) car elle est plus faible que DR qui dit justement

$$C(\alpha \vee \beta) \subseteq C(\alpha) \cup C(\beta)$$

Malheureusement il y a une réelle différence entre le cas fini et le cas infini. En particulier nous montrons que, dans le cas infini, la propriété W-DR ne caractérise pas l'existence de représentations injectives au sens KLM, c'est-à-dire celles où la relation entre les valuations est un ordre strict qui est *smooth*. Néanmoins nous prouvons que cette propriété caractérise les relations préférentielles qui admettent une représentation injective au sens MAK, c'est-à-dire celle où la seule propriété exigée pour la relation entre les mondes est le fait d'être *smooth*.

Lorsque l'on analyse les preuves des théorèmes de représentation pour les relations préférentielles, rationnelles et disjonctives rationnelles dans [49, 55, 30] on remarque que la construction de l'ordre dans les modèles se fait en deux temps : d'abord on définit un ordre sur les formules et ensuite à l'aide de cet ordre on bâtit l'ordre sur les états ou sur les valuations. Or une question qui semble naturelle de se poser est la suivante : étant donnée une relation de conséquence $\mid\!\sim$ y-a-t-il une façon canonique et directe de définir un ordre sur les valuations pour obtenir une représentation ? Nous pensons que la réponse est positive et que les résultats dans ce chapitre sont en grande partie une justification de cela.

Pour définir un tel ordre on remarque d'abord que les seules valuations qui ont de l'information sont les valuations dites *normales*, c'est-à-dire celles pour lesquelles il existe des formules dont la valuation satisfait leurs conséquences non monmonotones. Plus précisément N est normal s'il existe α tel que $N \models C(\alpha)$. Nous proposons un ordre $<_e$ entre les valuations normales, appelé l'ordre essentiel, qui est défini très simplement :

$$M <_e N \quad \text{ssi} \quad \forall \alpha (N \models C(\alpha) \Rightarrow M \not\models \alpha)$$

autrement dit $M <_e N$ si et seulement si la théorie classique de M, dénotée $Th(M)$ (*i.e.* $Th(M) = \{\alpha : M \models \alpha\}$), n'intersecte pas la théorie non monotone de N, dénotée $T(M)$ (*i.e.* $T(M) = \{\alpha : M \models C(\alpha)\}$).

Nous prouvons que cette relation essentielle est en fait la plus générale qui représente une relation de conséquence $\mid\!\sim$ lorsque cette dernière satisfait W-DR.

Pour préciser ces idées fixons une relation de conséquence $\mid\!\sim$. Soit S l'ensemble de valuations normales et soit $\prec$ un ordre strict sur S qui a la propriété d'être *smooth*, *i.e.* pour chaque formule α si $M \in mod(\alpha) \cap S$ alors soit M est minimal dans $mod(\alpha) \cap S$ ou bien il existe un $N \in mod(\alpha) \cap S$ tel que $N \prec M$ et N est minimal dans $mod(\alpha) \cap S$. Avec ces données nous dirons que $(S, \prec)$ est une représentation standard de $\mid\!\sim$ lorsque l'on a l'égalité suivante pour toute formule α :

$$mod(C(\alpha)) = min(mod(\alpha) \cap S, \prec) \tag{2.2}$$

et nous dirons aussi que dans ce cas $\prec$ est un ordre standard représentant $\mid\!\sim$. Nous faisons la distinction entre un ordre standard représentant $\mid\!\sim$ et une *relation standard* $\prec'$ définie sur S représentant $\mid\!\sim$. Dans ce dernier cas la seule chose que l'on exige de $\prec'$ est que l'équation 2.2 soit satisfaite, mais on n'exige pas de propriétés particulières pour $\prec'$. Par exemple on n'exige pas que $\prec'$ soit transitive ou qu'elle soit *smooth.* Notons qu'une relation standard qui représente $\mid\!\sim$ est un modèle MAK lorsqu'elle est *smooth.*

Notre première observation est que pour n'importe quelle relation standard $\prec$ représentant $\mid\!\sim$ on a $\prec \subseteq <_e$. Et le résultat central est que toute relation préférentielle satisfait W-DR si et seulement si $<_e$ est une relation standard qui la représente et qui est *smooth.* En fait, dans le cas fini, $<_e$ est un ordre mais dans le cas infini la relation $<_e$ n'est pas forcément transitive.

Nous prouvons ensuite que pour les relations disjonctives rationnelles (P+DR), rationnelles (P + RM) et rationnelles transitives (P + RT) la paire $(S, <_e)$ est en fait un modèle au sens KLM (c'est-à-dire $<_e$ est un *ordre smooth*) qui les représente. Nous donnons des preuves directes de ces résultats qui ont l'avantage de montrer certaines propriétés de la relation $<_e$. Mais remarquons que l'on peut donner des preuves indirectes en utilisant des résultats de Freund [30]. En effet, Freund définit une relation $<_S$ qui est un ordre standard qui représente une relation disjonctive rationnelle. Et l'on peut prouver que pour les relations disjonctives rationnelles la relation essentielle et l'ordre de Freund coïncident, *i.e.* $<_e = <_S$.

Nous abordons aussi dans ce chapitre l'étude de l'unicité des représentations standards. Plus précisément la question à laquelle nous voudrions répondre est la suivante : quelles propriétés doit avoir une relation de conséquence $\mid\!\sim$ pour que dans le cas où elle admet une représentation standard elle soit unique ?

On sait que dans le cas fini toute représentation standard est unique. Dans le cas infini nous avons dégagé une propriété topologique des représentations qui nous donnera l'unicité.

On considère la topologie suivante sur l'ensemble de valuations : les ouverts de base sont les ensembles $mod(\alpha)$ pour toute formule α. Ensuite, pour une relation de conséquence donnée on considère la topologie induite sur S, l'ensemble de valuations normales. On dira qu'une relation $\prec$ sur S est close vers le bas si pour chaque $N \in S$ l'ensemble $\{M \in S : M \prec N\}$ est un fermé. On montre que la relation essentielle est close vers le bas et que si l'on prend une relation standard $\prec$ représentant $\mid\!\sim$, qui de plus est close vers le bas alors $\prec = <_e$. Comme corollaire on obtient que la relation essentielle est l'unique relation standard close vers le bas représentant une relation préférentielle qui satisfait W-DR.

A l'aide de ces outils topologiques on peut prouver assez facilement que pour une relation rationnelle l'ordre essentiel sera l'unique ordre modulaire qui la représente.

Pour finir ce chapitre nous donnons quelques exemples et contre-exemples qui montrent les problèmes qui se présentent dans le cas infini. Notamment nous construisons une relation préférentielle qui admet un modèle injectif qui ne satisfait pas W-DR. Nous montrons aussi l'existence d'une relation préférentielle avec un modèle standard (en particulier W-DR est satisfaite) mais la relation essentielle n'est pas transitive bien qu'elle la représente au sens de l'équation 2.2.

2.3.3 Panorama sur le chapitre 5

Nos principales motivations dans ce chapitre sont :

- Trouver quelles sont les propriétés logiques du raisonnement sur les explications.
- Etablir des liens entre l'abduction et la déduction (à l'envers).
- Etabir des liens entre différents critères de préférence pour sélectionner des explications et des propriétés purement logiques de l'abduction.

Pour ce faire nous adoptons, suivant Flach [29], le point de vue relationnel qui s'est déjà avéré très puissant et fructueux dans l'étude de l'inférence non monotone. Nous considérons donc des relations binaires entre des formules propositionnelles. Ces relations binaires seront généralement notées $\rhd$. Lorsque $\alpha \rhd \gamma$ nous dirons que α est bien expliqué par γ ou bien que γ est une bonne explication de α. Nous cherchons quelles sont les propriétés structurelles de la relation $\rhd$ pour qu'elle soit une "bonne" relation d'explication. Notons que cette approche a déjà été suivie par Zadrozny [114], Flach [29], Cialdea et Pirri [66] et Aliseda [2]. Une des différences essentielles entre notre approche et les leurs est que leurs justifications des postulats de rationalité (des règles structurelles) est essentiellement intuitive par contre nous donnons des outils logiques pour analyser le bon comportement des postulats.

Nous nous plaçons dans un cadre assez traditionnel du raisonnement abductif dans lequel on assume qu'il y a une relation causale entre une observation que l'on veut expliquer et ses explications. L'idée de base pour modéliser l'abduction est la déduction à l'envers plus d'autres conditions. Plus précisément on suppose que l'on a une théorie de base Σ à la lumière de laquelle on doit expliquer une observation α. Les candidats à être des explications de α seront les formules γ consistantes avec Σ telles que $\Sigma \cup \{\gamma\} \vdash \alpha$. En général tous les candidats à être des explications de α ne sont pas de bonnes explications. Parmi ces candidats les bonnes explications seront justement les γ telles que $\alpha \rhd \gamma$.

A chaque relation d'explication on peut associer naturellement une relation de conséquence : les conséquences d'une formule seront les conclusions de sa meilleure explication. Plus précisément supposons que $\rhd$ est une relation d'explication. On lui associe la relation $\mid\!\sim_{ab}$ de la façon suivante

$$\alpha \mid\!\sim_{ab} \beta \quad \text{si } \Sigma \cup \{\gamma\} \vdash \beta \text{ pour tout } \gamma \text{ tel que } \alpha \rhd \gamma. \tag{2.3}$$

Cette relation a déjà été suggérée par Levesque [56] comme une nouvelle opération déductive pouvant être utile pour faire des "expériments" contrafactuels. Néanmoins la motivation pour introduire cette définition vient de Lobo et Uzcátegui [65]. Ils définissent des relations assez semblables à $\mid\!\sim_{ab}$ pour étudier le raisonnement abductif. L'idée centrale dans la recherche des règles structurelles des relations d'explication sera de considérer les liens entre $\rhd$ et $\mid\!\sim_{ab}$. En particulier nous voulons que les règles satisfaites pour une relation $\rhd$ impliquent un bon comportement de la relation $\mid\!\sim_{ab}$. Par exemple quels postulats sur $\rhd$ permettront que la relation de conséquence associée soit cumulative, préférentielle, disjonctive rationnelle, rationnelle.

Cette approche justifiera d'une façon formelle la plupart des postulats introduits dans d'autres approches et nous permettra de dégager des règles de coupure pour les relations d'abduction. Ces règles de coupure sont en quelque sorte les règles duales de différentes règles de monotonie faible. Notamment nous introduisons les règles de coupure prudente explicative (E-C-Cut), coupure rationnelle explicative (E-R-Cut) et coupure explicative(E-Cut) énoncées par la suite où la notation $\gamma \vdash_\Sigma \alpha$ signifie $\Sigma \cup \{\gamma\} \vdash \alpha$:

E-C-Cut: $$\frac{(\alpha \wedge \beta) \rhd \gamma \ , \ \forall\delta\,[\alpha \rhd \delta \ \Rightarrow \ \delta \vdash_\Sigma \beta\,]}{\alpha \rhd \gamma}$$

E-R-Cut: $$\frac{(\alpha \wedge \beta) \rhd \gamma \ ; \ \exists\delta\,[\alpha \rhd \delta \ \&\ \delta \vdash_\Sigma \beta]}{\alpha \rhd \gamma}$$

E-Cut: $$\frac{(\alpha \wedge \beta) \rhd \gamma}{\beta \rhd \gamma}$$

Ces règles sont le pendant de la monotonie prudente, la monotonie rationnelle et de la monotonie respectivement.

Dans la première partie de ce chapitre nous isolons des ensembles de postulats pour les relations d'explication qui seront en correspondance avec la hiérarchie des relations de conséquence non monotone (*v.g.* cumulative, preférentielle, disjonctive rationnelle, rationnelle). Illustrons ceci avec les règles correspondantes aux relations préférentielles. Supposons que $\rhd$ satisfait la coupure prudente explicative (E-C-Cut) déjà mentionnée plus l'équivalence logique à gauche (LLE), la monotonie prudente explicative (E-CM), et la conjonction à droite (RA) où LLE_Σ, E-CM et RA sont les règles suivantes :

LLE_Σ: $$\frac{\vdash_\Sigma \alpha \leftrightarrow \alpha' \ , \ \alpha \rhd \gamma}{\alpha' \rhd \gamma}$$

E-CM: $$\frac{\alpha \rhd \gamma \ ; \ \gamma \vdash_\Sigma \beta}{(\alpha \wedge \beta) \rhd \gamma}$$

RA: $$\frac{\alpha \rhd \gamma \ ; \ \gamma' \vdash_\Sigma \gamma \ ; \ \gamma' \not\vdash_\Sigma \bot}{\alpha \rhd \gamma'}$$

Alors la relation $\mid\!\sim_{ab}$ associée à $\rhd$ sera préférentielle.

Réciproquement si l'on suppose que l'on est en train de raisonner avec une relation de conséquence $\mid\!\sim$ qui nous donne les conséquences normales d'une observation on peut définir une relation d'explication $\tilde{\rhd}$ associée à $\mid\!\sim$ en posant

$$\alpha \tilde{\rhd} \gamma \overset{def}{\Leftrightarrow} \gamma \not\vdash_\Sigma \bot \ \& \ C(\alpha) \subseteq Cn\,(\Sigma \cup \{\gamma\}) \tag{2.4}$$

On aura alors que la relation d'explication $\tilde{\rhd}$ a de bonnes propriétés si la relation $\mid\!\sim$ est ***adéquate***, c'est-à-dire si elle satisfait

$$C(\alpha) = \bigcap\{Cn\,(\Sigma \cup \{\gamma\}) : \ C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})\} \tag{2.5}$$

Par exemple si $\mid\!\sim$ est préférentielle et adéquate la relation $\tilde{\rhd}$ satisfera les règles LLE_Σ + E-CM + E-C-Cut + RA.

Nous avons vu comment associer une relation de conséquence $\mid\!\sim_{ab}$ à une relation d'explication $\rhd$. Nous avons aussi vu comment associer une relation d'explication $\tilde{\rhd}$ à une relation de conséquence $\mid\!\sim$. Or, une question qui semble naturelle est de savoir quand une relation d'explication est de la forme $\tilde{\rhd}$. Plus précisément lorsque nous partons de $\rhd$ et réalisons le cheminement suivant :

$$\rhd \quad \leadsto \quad \mid\!\sim_{ab} \quad \leadsto \quad \tilde{\rhd}$$

est-ce que $\rhd = \tilde{\rhd}$?

En général cette égalité n'est pas vraie. Lorsque cela arrive on dira que la relation d'explication $\rhd$ est *causale*. D'après les définitions ceci peut simplement s'exprimer de la façon suivante :

$$\alpha \rhd \gamma \text{ iff } C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \gamma) \tag{2.6}$$

Un corollaire assez immédiat de la causalité est que l'on peut obtenir des théorèmes de représentation à la KLM pour les relations d'explication qui satisfont cette propriété.

Nous avons réussi à caractériser la causalité. Dans le cas fini par deux simples règles : la conjonction à droite (RA), déjà mentionnée plus la règle d'affaiblissement à droite (E-RW)

E-RW:
$$\frac{\alpha \rhd \gamma \ ; \ \alpha \rhd \delta}{\alpha \rhd (\gamma \vee \delta)}$$

En contraste dans le cas infini la causalité est caractérisée par RA plus une propriété de nature un peu plus complexe. Nous appelons cette propriété l'axiome de causalité. En termes topologiques cet axiome dit que si la collection de bonnes explications de α est dense au dessous de γ (dans le réticulé des formules) alors γ est une bonne explication de α.

Une des façons les plus naturelles de définir une relation d'explication est de supposer que l'on dispose d'une relation de préférence $\prec$ entre les formules. Alors les bonnes explications d'une observation α seront les éléments $\prec$-minimaux de l'ensemble des candidats à être des explications de α. Rappelons que certains types d'ordre entre les formules caractérisent certaines classes de relations de conséquence. Par exemple des ordres possibilistes [22] et des ordres d'expectatives [37] caractérisent les relations rationnelles, les ordres préférentiels [30] caractérisent les relations préférentielles. Ainsi on peut se demander quels types "d'ordre" caractériseront les différentes classes de relations d'explications que nous avons introduites. L'étude de cette question est un des apports de ce chapitre. En fait nous réussissons à prouver des théorèmes de représentation en termes d'ordres pour certaines classes de relations d'explication. Le corollaire de ces résultats est que dans les propriétés structurelles d'une relation d'explication on peut trouver implicitement l'ordre qui sert à choisir les bonnes explications. Ceci est à comparer avec l'approche de Cialdea-Mayer et Pirri [66] où elles utilisent un ordre explicite pour présenter les propriétés structurelles d'une relation d'explication.

Les idées que nous utilisons dans cette étude s'inspirent de celles utilisées dans le chapitre 4. Deux choses seront importantes :

- Savoir quelles sont les formules qui jouent le rôle des valuations normales.
- Définir un ordre (ou une relation) directement à partir de $\rhd$, qui joue le rôle de la relation essentielle associée à une relation de conséquence.

Concernant le premier point nous avons le concept de formules *admissibles* : ce seront celles qui sont une bonne explication de quelque formule. Plus precisément γ est admissible ssi il existe α telle que $\alpha \rhd \gamma$.

Concernant le deuxième point on définit une relation $\prec_e$ entre les formules de la façon suivante : si γ est admissible et δ n'est pas admissible alors on pose $\gamma \prec_e \delta$; si γ et δ sont admissibles alors on pose $\gamma \prec_e \delta$ ssi $Cn(\Sigma \cup \{\gamma\}) \cap \{\beta : \beta \rhd \delta\} = \emptyset$. Notons l'analogie entre cette définition et celle de $<_e$. Ici l'ensemble $Cn(\Sigma \cup \{\gamma\})$ joue le rôle de $Th(\gamma)$ et l'ensemble $\{\beta : \beta \rhd \delta\}$ joue le rôle de $T(\delta)$ (la théorie non monotone de δ).

A l'aide de ces outils nous prouvons une serie de théorèmes de représentation du type suivant :

Pour une relation d'explication $\rhd$ sont équivalentes :

(i) La relation $\rhd$ satisfait $\mathsf{LLE}_\Sigma + \mathsf{E\text{-}CM} + \mathsf{E\text{-}C\text{-}Cut} + \mathsf{RA} + \mathsf{E\text{-}R\text{-}Cut} + \mathsf{E\text{-}Con}_\Sigma$.

(ii) Il existe une relation de préférence modulaire et *smooth* $\prec$ satisfaisant la "continuation vers le haut"[1] telle que pour tout α on a

$$\alpha \rhd \gamma \quad \text{ssi} \quad \gamma \in min(Expla(\alpha), \prec)$$

Le sens (ii)$\Rightarrow$(i) est une simple vérification. Pour la réciproque on démontre que $\prec_e$ a les propriétés requises.

2.3.4 Panorama sur le chapitre 6

Nous avons déjà évoqué les motivations pour étudier des opérateurs de changement définis d'une manière syntaxique (et en fait très dépendants de la représentation de la connaissance). Elles résident principalement dans le but de diminuer la complexité algorithmique des calculs sous-jacents à la définition des opérateurs.

Les idées que nous manipulons ici sont très simples et elles remontent d'une part aux travaux de Fagin et al. [28, 27], Ginsberg [40] et Nebel [79]. Quand on veut réviser une base B (non nécessairement close par déduction classique) par une nouvelle information ϕ on calcule l'ensemble M des sous-ensembles de B maxiconsistants avec ϕ et le résultat de la révision sera alors

$$\bigcap_{X \in M} Cn(X) \cup \{\phi\}$$

Nous allons faire quelques variations sur cette idée qui consisteront essentiellement dans les points suivants :

- Affaiblir la logique (l'opérateur Cn). La seule règle d'inférence que l'on considère est le modus ponens mais on n'a pas d'axiomes logiques. Ainsi par exemple on pourra inférer β de α et $\alpha \rightarrow \beta$ mais de ce même ensemble on ne pourra pas inférer $\neg\beta \rightarrow \neg\alpha$.
- Restreindre la nature des objets que l'on représente. On ne considérera que des formules (règles) du type $l_1 \wedge l_2 \wedge \cdots \wedge l_n \rightarrow l_{n+1}$ où les l_i sont des littéraux, *i.e.* des variables propositionnelles ou des négations des variables propositionnelles. Ces règles

[1] Cette propriété nous dit que $\forall\, \gamma, \gamma', \delta \;\; (\gamma \not\vdash_\Sigma \bot \;\;\&\;\; \gamma \vdash \gamma' \;\;\&\;\; \delta < \gamma \;\; \Rightarrow \delta < \gamma')$

seront écrites (suivant la tradition de la programmation logique) $l_1, l_2, \ldots, l_n \rightarrow l_{n+1}$. Lorsque $n = 0$, c'est-à-dire la règle n'a pas de prémisses la formule se réduit à un littéral (un fait). Nous définissons l'ensemble de conséquences d'un ensemble P des règles comme le plus petit ensemble de littéraux contenant les littéraux (les faits) de P et qui est clos par modus ponens par rapport aux règles de P. Cet ensemble sera dénoté $C_{fc}(P)$.

Maintenant nous fixons un ensemble de règles P qui sera considéré comme la théorie de base pour notre raisonnement. On suppose que l'on a un ensemble de faits (ou plutôt de multi-ensembles de faits) qui sont notre état actuel de croyances (les multi-ensembles correspondront à une disjonction). Ce que nous voulons faire est de changer (ou de mettre à jour) l'état actuel des connaissances par une nouvelle information dont la nature est très simple : un ensemble de faits. Dans le cas où l'information actuelle est un simple ensemble de faits L le résultat de la mise à jour par L' sera[2] le multi-ensemble

$$\langle L_1 \cup L', \ldots, L_n \cup L' \rangle$$

où $\{L_1, \ldots, L_n\}$ est l'ensemble des sous-ensembles de L qui sont maximaux et consistants avec $P \cup L'$ (on ne trouve pas deux littéraux opposés dans $C_{fc}(L_i \cup P \cup L')$). Ce résultat sera noté $L \diamond_P L'$ et nous appelerons $\diamond_P$ opérateur d'actualisation factuelle.

Dans le cas où l'information actuelle $\mathcal{F}$ est un multi-ensemble $\langle L_1, \ldots, L_n \rangle$ nous définissons $\mathcal{F} \diamond_P L'$ additivement, *i.e.*

$$\mathcal{F} \diamond_P L' = (L_1 \diamond_P L') \sqcup (L_2 \diamond_P L') \cdots \sqcup (L_n \diamond_P L')$$

où $\sqcup$ est la réunion de multi-ensembles[3].

Afin d'analyser les propriétés de rationalité de ce type d'opérateur nous devons donner un sens au multi-ensembles. Nous considérerons ses conséquences (toujours par rapport à un P fixé). Pour cela nous voyons un multi-ensemble comme une disjonction. Plus précisément si $\mathcal{F} = \langle L_1, \ldots, L_n \rangle$ nous posons

$$C_{fc}(\mathcal{F}) = \bigcap_{i=1}^{n} C_{fc}(L_i \cup P)$$

Par ailleurs nous relativisons les postulats AGM pour la révision et ceux de Katsuno et Mendelzon pour la mise à jour au cadre syntaxique restreint dans lequel nous travaillons. Nous prouvons que l'opérateur $\diamond_P$ est en fait un opérateur de mise à jour syntaxique. Ici nous insistons sur le fait que $\mathcal{F} \diamond_P L'$ est bien un multi-ensemble $\langle L_1, \ldots, L_n \rangle$ et non ses conséquences $C_{fc}(\langle L_1, \ldots, L_n \rangle)$. Faire cette distinction est la clé du bon comportement de l'opérateur. Par contre au niveau des postulats le rôle de C_{fc} est crucial car les postulats sont énoncés justement par rapport à la notion de conséquence.

Nous donnons un algorithme pour calculer $\diamond_P$. Mais à nouveau ici nous ne pouvons échapper à la NP-complétude car nous devons, dans une étape de l'algorithme, calculer un ensemble minimal de couverture.

Une autre idée pour définir des opérateurs de changement consiste à hiérarchiser l'information de la base actuelle P. Mais d'où peut-on extraire la hiérarchie ? Nous pensons que la représentation de l'information de B a implicitement une hiérarchie. En effet on peut voir la base P

[2]Ceci est approximatif, pour la définition précise voir la section 6.2.1.

[3]Ceci est à nouveau approximatif, voir la section 6.2.1 pour la définition précise.

comme un ensemble d'informations conditionnelles. Ceci permettra de définir une hiérarchie de la base qui s'inspire directement du calcul de la clôture rationnelle d'une base conditionnelle (voir [55]). Ce qui est important ici est que dû à la nature des objets que l'on manipule on peut faire ce calcul en temps polynomial.

La hiérarchie associée à P est simplement $P_0, P_1, \ldots P_n$ où $P_0 = P$ et P_{i+1} contient les éléments de P_i qui sont exceptionnels au sens de Lehmann et Magidor [55]. L'idée est que plus l'indice i de la hiérarchie est grand plus les objets qui sont dans les premisses de P_i sont rares.

Nous définissons la révision rangée d'une base P par une base P', noté $P \circ_{rk} P'$ comme $P_i \cup P'$ où i est le premier indice tel que $C_{fc}(P_i \cup P')$ est consistant.

Nous prouvons que cet opérateur a de bonnes propriétés structurelles. Notamment il est un opérateur de révision syntaxique. En outre cet opérateur est calculatoirement traitable. Mais nous montrons aussi quelques exemples où l'opérateur $\circ_{rk}$ exclut plus de formules qu'on ne le voudrait.

Pour essayer de réparer ce "mauvais comportement" nous étudions trois variations naturelles de $\circ_{rk}$. Ces opérateurs ont un mauvais comportement au niveau calculatoire (calcul d'ensemble minimal de couverture sous-jacent). Concernant leur rationalité seulement un d'entre eux est un opérateur de révision syntaxique : il est défini à l'aide de fonctions de sélection ("maxichoice") particulières.

Nous terminons ce chapitre en montrant un exemple d'une telle fonction de sélection.

2.3.5 Panorama sur le chapitre 7

Dans ce chapitre nous nous intéressons à une axiomatisation des opérateurs de fusion. L'idée intuitive que nous avons des opérateurs de fusion est la suivante : en entrée on a une "collection" d'informations provenant de sources *distinctes*, en sortie on doit avoir une information cohérente qui soit le plus en accord avec la "collection" d'informations selon certains critères qui peuvent être différents selon les cas : par exemple minimiser l'insatisfaction individuelle ou minimiser l'insatisfaction globale.

Les premières tentatives de fusionner l'information viennent de la communauté de bases de données mais avec un oubli des sources d'information. Par exemple on réunit toutes les informations et comme résultat on donne l'intersection de maxiconsistants. Plus récemment Revesz [99, 100], Lin and Mendelzon [62, 61] et Liberatore et Schaerf [58, 59] proposent des postulats qui essayent de capturer l'idée de fusion sans l'oubli des sources d'information. Revesz a une notion d'opérateurs de *model fitting* qui correspond à une fusion avec des contraintes d'intégrité. Lin et Mendelzon proposent des opérateurs de majorité. A la différence de ces deux approches Liberatore et Schaerf proposent des opérateurs qui prennent en compte seulement deux sources d'information. Leurs opérateurs se définissent de la façon suivante : la fusion de ϕ et ψ sera $(\phi * \psi) \vee (\psi * \phi)$ où $*$ est un opérateur de révision avec certaines propriétés. En particulier le résultat de fusionner ϕ et ψ à la Liberatore et Schaerf sera une formule qui implique toujours $\phi \vee \psi$.

Nous voyons cette dernière restriction comme une limitation car dans beaucoup de situations il sera raisonnable de trouver que le résultat de la fusion n'implique pas cette disjonction d'informations. Par exemple on peut imaginer que l'on est en train de raisonner sur la hauteur d'un avion en vol. Un observateur ϕ croit le voir à 1 Km au dessus du sol; un autre observateur ψ croit le voir à 3 Km au dessus du sol. Dans ce cas il pourrait être raisonnable de penser que l'avion se trouve à 2 Km au dessus du sol ce qui n'implique pas $\phi \vee \psi$.

Nous proposons une axiomatique pour la fusion pure. En particulier nous n'imposons pas ce genre de contraintes (d'intégrité) sur le résultat. Bien que cela puisse être vu comme une faiblesse de notre approche il n'est pas difficile d'imaginer comment généraliser notre approche au cas où l'on doit satisfaire des contraintes d'intégrité.

On peut dire que malgré l'absence de contraintes d'intégrité notre travail continue et étend l'approche de Revesz. En particulier nous introduisons trois postulats qui semblent jouer un rôle important pour pouvoir étudier la nature des opérateurs de fusion. D'abord un postulat d'indépendance de la syntaxe qui tient compte de l'origine des informations; ensuite un postulat d'équité et finalement un postulat de (faible) indépendance de la majorité.

Pour avoir une idée plus claire de nos postulats nous devons préciser un petit peu plus la nature des objets que nous manipulons. Nous considérons que les opérateurs de fusion agissent sur des *ensembles de connaissances* qui sont en fait des **multi-ensembles** de formules, cela correspond aux sources d'information. Le fait de considérer un multi-ensemble est très important car il se pourrait que plusieurs sources d'information véhiculent le même contenu et alors le résultat d'une fusion peut sensiblement changer selon le nombre de ces sources.

Nos opérateurs de fusion agissent donc sur des ensembles de connaissances et en résultat ils nous fournissent une formule.

On dira que deux ensembles de connaisances E_1 et E_2 sont équivalents, $E_1 \leftrightarrow E_2$, ssi il existe une bijection f de E_1 sur E_2 telle que $\vdash f(\phi) \leftrightarrow \phi$. En particulier le nombre de "sources d'information" de E_1 et E_2 est le même.

Les postulats de base d'un opérateur Δ de fusion sont les suivants :

(A1) $\Delta(E)$ est consistant.

(A2) Si $\bigwedge E$ est consistant , alors $\Delta(E) = \bigwedge E$

(A3) Si $E_1 \leftrightarrow E_2$, alors $\vdash \Delta(E_1) \leftrightarrow \Delta(E_2)$

(A4) If $\phi \wedge \phi'$ n'est pas consistant, alors $\Delta(\phi \sqcup \phi') \not\vdash \phi$

(A5) $\Delta(E_1) \wedge \Delta(E_2) \vdash \Delta(E_1 \sqcup E_2)$

(A6) Si $\Delta(E_1) \wedge \Delta(E_2)$ est consistant, alors $\Delta(E_1 \sqcup E_2) \vdash \Delta(E_1) \wedge \Delta(E_2)$

A1, A2 et A3 sont assez intuitifs. A3 est le postulat d'indépendence de la syntaxe. A4 est le postulat d'équité : l'opérateur ne doit pas privilégier une opinion par rapport à une autre lorsque ces opinions se trouvent en opposition. A5 et A6 ensemble expriment que si deux sous-groupes sont d'accord après fusion alors le résultat de la fusion globale est justement ce en quoi les deux sous-groupes sont d'accord.

Nous distinguons ensuite les opérateurs majoritaires des opérateurs qui sont insensibles (faiblement) à la majorité.

Pour les majoritaires le postulat est évident :

$$(M7) \quad \forall \phi \; \exists n \; \Delta(E \sqcup \phi^n) \vdash \phi$$

Pour les insensibles à la majorité il est moins évident car l'axiome qui pourrait paraître naturel

$$(A7') \quad \forall \phi \, \forall n \; \Delta(E \sqcup \phi^n) = \Delta(E \sqcup \phi)$$

est en fait inconsistant avec les postulats de base.

Nous proposons une forme affaiblie :

$$(A7) \quad \forall \phi \, \exists \psi \; \phi \nvdash \psi \; \forall n \; \Delta(\phi \sqcup \psi^n) = \Delta(\phi \sqcup \psi)$$

Ce postulat nous dit que dans beaucoup des cas le résultat ne dépend pas de la fréquence d'une opinion.

On montre facilement qu'en présence du postulat d'équité A7 et M7 ne peuvent être satisfaits en même temps.

Nous prouvons que les postulats de base plus M7 est un ensemble de postulats consistant. De même que les postulats de base plus A7 est un ensemble de postulats consistant. Pour ce faire nous construisons des opérateurs qui ressemblent à certaines fonctions de la théorie de la décision multicritère mais dans un cadre complètement qualitatif (en fait sémantique). En particulier pour le deuxième des résultats nous construisons un opérateur qui est une sorte de *maximin* géneralisé.

Une observation intéressante est la traduction de ces postulats en termes sémantiques. Cela donne une représentation en termes d'ordres qui malgré sa trivialité a le mérite de faire apparaître de manière claire des analogies entre cette approche et la théorie du choix social. Etablir des liens de façon précise entre ces deux théories est une de nos tâches en perspective.

Chapitre 3

Beyond rational monotony: some strong non-Horn rules for nonmonotonic inference relations

Lehmann, Magidor and others have investigated the effects of adding the non-Horn rule of *rational monotony* to the rules for preferential inference in nonmonotonic reasoning. In particular, they have shown that every inference relation satisfying those rules is generated by some ranked preferential model.

We explore the effects of adding a number of other non-Horn rules that are stronger than or incomparable with rational monotony, but which are still weaker than plain monotony. Distinguished among these is a rule of *determinacy preservation*, equivalent to one of *rational transitivity*, for which we establish a representation theorem in terms of *quasi-linear* preferential models. An important tool in the proof of the representation theorem is the following purely semantic result, implicit in work of Freund, but here established by a more direct argument: every ranked preferential model generates the same inference relation as some ranked preferential model that is *collapsed*, in the sense of being both injective and such that each of its states is minimal for some formula.

We also consider certain other non-Horn rules which are incomparable with monotony but are implied by conditional excluded middle, and establish a representation result for a central one among them, which we call *fragmented disjunction*, equivalent to *fragmented conjunction*, in terms of *almost linear* preferential models.

Finally, we consider briefly some curious Horn rules beyond the preferential ones but weaker than monotony, notably those which we call *conjunctive insistence* and *n-monotony*.

Keywords: nonmonotonic reasoning, rational monotony, preferential models.

3.1 Introduction and Overview

The postulates for preferential inference, as formulated by Kraus, Lehmann and Magidor [49] are intended to gather together some properties for inference relations that may be regarded as in principle desirable, even when the inference relations are not monotonic. They are all Horn conditions, that is of the form: if such and such pairs are in the relation, so too is such another pair. Lehmann and Magidor [52] and [55] have also studied the effects of adding to the preferential postulates a further rule, non-Horn in character, called *rational monotony*. As usually formulated with a negative premise, it is: if $\alpha \mathrel{|\!\sim} \beta$ and *not* $\alpha \mathrel{|\!\sim} \neg\gamma$, then $\alpha\wedge\gamma \mathrel{|\!\sim} \beta$.

Equivalently, with positive premises but disjunctive conclusion, it is: if $\alpha \mid\!\sim \beta$ then either $\alpha \wedge \gamma \mid\!\sim \beta$ or $\alpha \mid\!\sim \neg\gamma$. In the two papers mentioned, it is shown that every inference relation satisfying the preferential rules is determined by some model of a certain kind, also called preferential, and that every inference relation satisfying in addition rational monotony is determined by some ranked preferential model.

It is known that rational monotony implies certain other non-Horn conditions of interest, notably *disjunctive rationality*, which in turn implies *negation rationality* - see for example the brief accounts in Makinson [69] and Lehmann and Magidor [55], or the more extensive work in Freund [30] and Freund and Lehmann [32] which provide a semantic characterization of inference relations satisfying these two rules. It is natural to ask whether there are any other rules of interest, stronger than or incomparable with rational monotony, but still weaker than plain monotony.

Makinson [69] drew attention to one such rule, called *determinacy preservation*, showing that it lies between monotony and rational monotony, but without investigating it semantically. Bezzazi and Pino Pérez [9] began a semantic investigation of two other rules, *rational transitivity* and *rational contraposition*. In this paper we study these and related conditions more systematically, establishing interrelations and providing semantic characterizations.

It turns out, as we shall show, that given the preferential rules, rational transitivity and determinacy preservation are equivalent, and are in turn equivalent to the combined force of rational monotony with rational contraposition, as also to the combined force of rational monotony with another rule that we shall consider. Rational transitivity *alias* determinacy preservation thus appears to occupy a rather pivotal position in this region. We show that any inference relation satisfying that rule in addition to preferential ones, is determined by a preferential model that is in a certain sense quasi-linear. The proof makes use of Lehmann and Magidor's representation theorem for rational monotony, but also of an important tool of a purely semantic nature. This is the result, implicit in Freund [30], that every ranked preferential model determines the same inference relation as some ranked preferential model that is *collapsed*, in the sense of being both injective and such that each state is minimal for some formula.

We also consider certain non-Horn rules that are not implied by monotony, and which for this reason are perhaps intuitively less interesting, but which are nevertheless weaker than the well-known rule of *conditional excluded middle* of Stalnaker [107], also called *full determinacy* in Makinson [69]: if not $\alpha \mid\!\sim \beta$ then $\alpha \mid\!\sim \neg\beta$. We isolate two such rules of particular formal interest, which we call *disjunction fragmentation* and *conjunction fragmentation*. We prove that they are equivalent and then we establish a representation theorem for preferential relations satisfying disjunction fragmentation (conjunction fragmentation), using the same semantic tool as for rational transitivity above.

All of the rules so far mentioned as potential additions to those for preferential inference, are non-Horn. Curiously, Horn rules appear to be less plentiful as potential additions. However in a final section we identify some such rules, weaker than monotony but not implied by rational monotony, represent some of them semantically, and raise a number of open questions.

We presume some familiarity with the main lines of at least one of Kraus, Lehmann and Magidor [49], Lehmann and Magidor [55], Makinson [69].

3.2 Background

In this section we recall some basic definitions and results from Kraus, Lehmann and Magidor [49] and Lehman and Magidor [55], which will be used in the paper.

We consider formulae of classical propositional calculus built over a set of elementary formulae denoted *Var* plus two constants $\top$ and $\bot$ (the formulae **true** and **false** respectively). Let $\mathcal{L}$ be the set of formulae. If *Var* is finite we will say that the language $\mathcal{L}$ is finite. Let $\mathcal{U}$ be the set of valuations (or worlds), *i.e.* functions $v : Var \cup \{\top, \bot\} \longrightarrow \{0,1\}$ such that $v(\top) = 1$ and $v(\bot) = 0$. We use lower case letters of the Greek alphabet to denote formulae, and the letters $v, v_1, v_2, \ldots$ to denote worlds. As usual, $\vdash \alpha$ means that α is a tautology and $v \models \alpha$ means that v satisfies α where compound formulae are evaluated using the usual truth-functional rules. We consider certain binary relations between formulae. These relations will be called inference relations and will be written $\mid\!\sim$.

Definition 3.1 *A relation* $\mid\!\sim$ *is said to be preferential iff the following rules hold*

$$\text{REF}\quad \frac{}{\alpha\mid\!\sim\alpha} \qquad\qquad \text{LLE}\quad \frac{\alpha\mid\!\sim\beta \quad \vdash \alpha\leftrightarrow\gamma}{\gamma\mid\!\sim\beta}$$

$$\text{RW}\quad \frac{\alpha\mid\!\sim\beta \quad \vdash \beta\rightarrow\gamma}{\alpha\mid\!\sim\gamma} \qquad\qquad \text{AND}\quad \frac{\alpha\mid\!\sim\beta \quad \alpha\mid\!\sim\gamma}{\alpha\mid\!\sim\beta\wedge\gamma}$$

$$\text{OR}\quad \frac{\alpha\mid\!\sim\gamma \quad \beta\mid\!\sim\gamma}{\alpha\vee\beta\mid\!\sim\gamma} \qquad\qquad \text{CM}\quad \frac{\alpha\mid\!\sim\beta \quad \alpha\mid\!\sim\gamma}{\alpha\wedge\gamma\mid\!\sim\beta}$$

These rules are known as the rules of the system P. The abbreviations above are read as follows: REF -reflexivity, LLE -left logical equivalence, RW -right weakening, CM -cautious monotony. AND and OR are self-explanatory.

A relation $\mid\!\sim$ *is said to be rational iff it is preferential and the following rule (rational monotony) holds*

$$\text{RM}\quad \frac{\alpha\mid\!\sim\beta \quad \alpha\not\mid\!\sim\neg\gamma}{\alpha\wedge\gamma\mid\!\sim\beta}$$

Definition 3.2 *A structure* $\mathcal{M}$ *is defined by a triple* $\langle S, \imath, \prec\rangle$ *where* S *is a set (of arbitrary items, called states),* $\prec$ *is a strict order (*i.e. *transitive and irreflexive) on* S *and* $\imath : S \longrightarrow \mathcal{U}$ *is a total function (the interpretation function). If the function* $\imath$ *is injective the structure is said also to be injective.*

Let $\mathcal{M} = \langle S, \imath, \prec\rangle$ be a structure. We adopt the following notations: if $T \subseteq S$, then $\min(T)$ is the set of all minimal elements of T with respect to $\prec$, *i.e.* $\min(T) = \{t \in T : \neg\exists t' \ (t' \in T \text{ and } t' \prec t)\}$; $\text{mod}_{\mathcal{M}}(\alpha) = \{s \in S : \imath(s) \models \alpha\}$; $\min_{\mathcal{M}}(\alpha)$ denotes $\min(\text{mod}_{\mathcal{M}}(\alpha))$.

Definition 3.3 *A structure* $\mathcal{M} = \langle S, \imath, \prec\rangle$ *is said to be a preferential model iff for any formula* α *the following property (smoothness) holds*

$$\forall s \in mod_{\mathcal{M}}(\alpha) \setminus min_{\mathcal{M}}(\alpha) \ \ \exists s' \in min_{\mathcal{M}}(\alpha) \ \ s' \prec s$$

A structure $\mathcal{M} = \langle S, \imath, \prec \rangle$ is said to be a ranked model iff it is a preferential model and there exists a strict linear order $(\Omega, <)$ and a function $r : S \longrightarrow \Omega$ such that for any $s, s' \in S$, $s \prec s'$ iff $r(s) < r(s')$.

Definition 3.4 *Let $\mathcal{M} = \langle S, \imath, \prec \rangle$ be a preferential model. The inference relation $\mid\!\sim_{\mathcal{M}}$ is defined by the following*

$$\alpha \mid\!\sim_{\mathcal{M}} \beta \Leftrightarrow min_{\mathcal{M}}(\alpha) \subseteq mod_{\mathcal{M}}(\beta)$$

The following (representation) theorem is due to Kraus, Lehmann and Magidor [49].

Theorem 3.5 *$\mid\!\sim$ is a preferential relation iff there is a preferential model $\mathcal{M} = \langle S, \imath, \prec \rangle$ such that $\mid\!\sim_{\mathcal{M}} = \mid\!\sim$. If the language is finite then S can be chosen finite.*

The following (representation) theorem is due to Lehmann and Magidor [55].

Theorem 3.6 *$\mid\!\sim$ is a rational relation iff there is a ranked model $\mathcal{M} = \langle S, \imath, \prec \rangle$ such that $\mid\!\sim_{\mathcal{M}} = \mid\!\sim$. If the language is finite then S can be chosen finite.*

Proposition 3.7 *If $\mid\!\sim$ is a preferential relation then the following rules hold*

$$\mathsf{S}\ \frac{\alpha \wedge \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \beta \rightarrow \gamma} \qquad \mathsf{CUT}\ \frac{\alpha \wedge \beta \mid\!\sim \gamma \quad \alpha \mid\!\sim \beta}{\alpha \mid\!\sim \gamma}$$

If $\mid\!\sim$ is a rational relation then the following rules hold

$$\mathsf{DR}\ \frac{\alpha \vee \beta \mid\!\sim \gamma \quad \alpha \not\mid\!\sim \gamma}{\beta \mid\!\sim \gamma} \qquad \mathsf{NR}\ \frac{\alpha \mid\!\sim \beta \quad \alpha \wedge \gamma \not\mid\!\sim \beta}{\alpha \wedge \neg \gamma \mid\!\sim \beta}$$

For the proofs see [49] for S and CUT and see [69] or [55] for DR and NR. The abbreviations above are read as follows: S -Shoham rule (this abbreviation is taken from [49]; note that this rule corresponds to the hard half of the deduction theorem for classical $\vdash$), DR -disjunctive rationality, NR -negation rationality. The term CUT is self-explanatory, but it should be noted that this form of cut, which plays an important role in nonmonotonic logic, is weaker than the forms of cut usually studied in Gentzen-style formulations of classical and intuitionistic logic. The latter imply transitivity of the inference relation; the former does not.

Notation: If n is a natural number, $\overline{n}$ will denote the set $\{0, 1, \ldots, n\}$ linearly ordered with the natural order $<$. If A is a set, the cardinality of A will be denoted by $|A|$. When $\mathcal{M} = \langle S, \imath, \prec \rangle$ is a preferential model, $u \in S$ and α a formula, if there is no ambiguity we shall write $u \models \alpha$, $\mathrm{mod}(\alpha)$ and $\min(\alpha)$ instead of $\imath(u) \models \alpha$, $\mathrm{mod}_{\mathcal{M}}(\alpha)$ and $\min_{\mathcal{M}}(\alpha)$ respectively.

Observation 3.8 *It is known that there are preferential models whose inference relation is not generated by any injective one. A simple finite example was given* en passant *by Krauss, Lehmann and Magidor at the end of section 5.2 of [49]. The language is assumed to have just two elementary sentences p, q. The states are s_i $(0 \leq i \leq 3)$ with $s_0 < s_2$ and $s_1 < s_3$,* *and* $s_0 \models p\wedge\neg q$, $s_1 \models \neg p\wedge\neg q$, *whilst* $s_2, s_3 \models p\wedge q$. *Kraus, Lehmann and Magidor leave the verification of the example as an exercise; a verification is sketched by Schlechta in section 1 of [104]. Because of its relation with the theme of this paper, we give the verification in full, using moreover an infinite language so as to make it clear that the example is not an artifact of a limited number of elementary sentences.*

Let p_j $(j \in J)$ *be all the other elementary sentences and make them behave just like* p, i.e. *put* $s_i \models p_j$ *iff* $s_i \models p$. *Let* $\mid\!\sim$ *the inference relation determined by this preferential model. Then clearly we have the following: (1)* $(p\wedge q)\vee\neg q \mid\!\sim \neg q$, *whilst (2)* $(p\wedge q)\vee(p\wedge\neg q) \not\mid\!\sim \neg q$ *(witness* s_3*) and (3)* $(p\wedge q)\vee(\neg p\wedge\neg q) \not\mid\!\sim \neg q$ *(witness* s_2*). Moreover (4)* $p\wedge\neg p_j \mid\!\sim \bot$. *We claim that any injective preferential model whose inference relation agrees with this one on (1) and (4), disagrees with it on (2) or (3).*

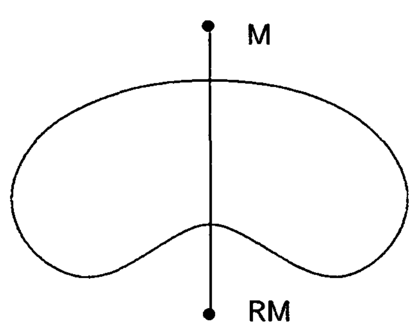
Consider any injective preferential model $\mathcal{M} = \langle S, \imath, <\rangle$ *and suppose that (1) and (4) hold. In the case that* $p\wedge q \mid\!\sim \bot$ *clearly we have* $(p\wedge q)\vee(p\wedge\neg q) \mid\!\sim \neg q$ *and also* $(p\wedge q)\vee(\neg p\wedge\neg q) \mid\!\sim \neg q$, *so we may suppose without loss of generality that* $p\wedge q \not\mid\!\sim \bot$. *Then there is a state* $s \in S$ *with* $s \models p\wedge q$ *so* $s \models (p\wedge q)\vee\neg q$. *By (1)* $s \notin min((p\wedge q)\vee\neg q)$ *so there is a* $t \in S$ *with* $t < s$ *and* $t \in min((p\wedge q)\vee\neg q)$ *so by (1) again* $t \models \neg q$. *Now either* $t \models p$ *or* $t \models \neg p$. *Consider the latter; the argument for the former is similar. Suppose for reductio that (3) holds,* i.e. *there is* $u \in S$ *with* $u \in min((p\wedge q)\vee(\neg p\wedge\neg q))$ *and* $u \models q$. *Then* $u \models p$. *Moreover for all* $j \in J$, *we have* $u \models p_j$, *for otherwise there is* $i \in J$ *such that* $u \models p\wedge\neg p_i$ *so there is* u *with* $u' \leq u$ *and* $u' \in min(p\wedge\neg p_i)$ *contradicting (4). Similarly* $s \models p_j$ *for all* $j \in J$. *Since* $s, u \models p\wedge q\wedge p_j$ *for all* $j \in J$ *we have* $\imath(s) = \imath(u)$ *so by injectivity* $s = u$. *Thus* $s \in min((p\wedge q)\vee(\neg p\wedge\neg q))$, *contradicting* $t < s$ *and* $t \models \neg p\wedge\neg q$.

3.3 Some strong non-Horn conditions

Rational monotony of course is a restricted form of, and thus implied by, plain monotony (M):

$$\mathsf{M}\quad \frac{\alpha \mid\!\sim \beta}{\alpha\wedge\gamma \mid\!\sim \beta}$$

One of our purposes in this paper is to examine some interesting non-Horn conditions, stronger than rational monotony (or in some cases, independent of it) but still weaker than monotony. In other words, we wish to investigate the enclosed area of the following diagram



Four rules that arise in this connection are *determinacy preservation*, *rational transitivity*, *rational contraposition*, and *weak determinacy*.

Determinacy preservation (DP), briefly considered by Makinson [69], is the rule

$$\text{DP}\ \frac{\alpha \mid\!\sim \beta \quad \alpha \wedge \gamma \not\mid\!\sim \neg\beta}{\alpha \wedge \gamma \mid\!\sim \beta}$$

This rule evidently is a weak form of monotony. It can also seen as a weak form of Stalnaker's rule [107] of conditional excluded middle (if $\phi \not\mid\!\sim \psi$ then $\phi \mid\!\sim \psi$): a consequence of a formula is either conserved when we add a new hypothesis or we get the negation of this consequence.

Rational transitivity (RT), introduced by Bezzazi and Pino Pérez [9], is the rule

$$\text{RT}\ \frac{\alpha \mid\!\sim \beta \quad \beta \mid\!\sim \gamma \quad \alpha \not\mid\!\sim \neg\gamma}{\alpha \mid\!\sim \gamma}$$

Obviously this rule is a weak form of transitivity. The intuition behind this rule is the following: when the premises of transitivity hold we get the usual conclusion except when its 'opposite' holds. Note that this rule is also a weak form of conditional excluded middle.

Rational contraposition (RC), also introduced by Bezzazi and Pino Pérez [9], is the rule

$$\text{RC}\ \frac{\alpha \mid\!\sim \beta \quad \neg\beta \not\mid\!\sim \alpha}{\neg\beta \mid\!\sim \neg\alpha}$$

Obviously this rule is a weak form of contraposition. The intuition behind this rule is the following: when the premise of contraposition holds we get the usual conclusion except when its 'opposite' holds. This rule is again a weak form of conditional excluded middle.

Weak determinacy (WD), formulated by Michael Freund in correspondence with the authors, is the rule

$$\text{WD}\ \frac{\top \mid\!\sim \neg\alpha \quad \alpha \not\mid\!\sim \beta}{\alpha \mid\!\sim \neg\beta}$$

This rule says that any formula α that is 'exceptional' in the sense of Lehmann and Magidor [55], *i.e.* such that $\top \mid\!\sim \neg\alpha$, is complete in the sense that for every formula, either it or its negation is a consequence of the exceptional formula. Given the preferential rules, this is a special case of both monotony and conditional excluded middle.

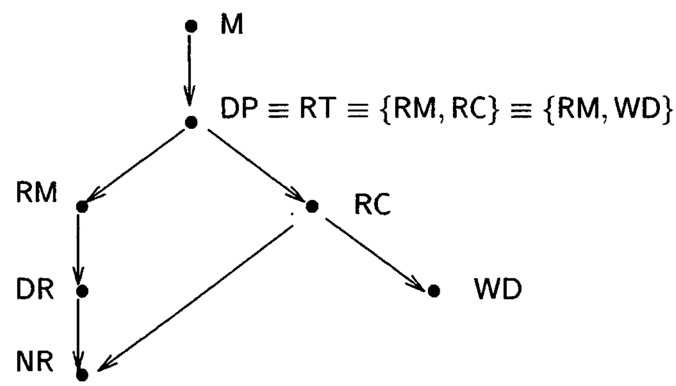
Definition 3.9 *A relation* $\mid\!\sim$ *is said to be determinacy preserving iff it is preferential and the rule DP holds.*

A relation $\mid\!\sim$ *is said to be rational transitive iff it is preferential and the rule RT holds.*

In this section we compare the strength of the rules DP, RT, RC, WD with each other as well as RM on the lower side and M on the upper side. The general picture turns out as follows:

Proposition 3.10 *Given the preferential rules P, the rules DP and RT are equivalent, and are implied by monotony. They are also equivalent to the pair {RM, RC} and also to the pair {RM, WD}. Moreover, given P, RC implies both WD and NR. However given P, none of the following implications hold: RM to WD, RC to DR, WD to NR.*

Recalling from [69] that M quite trivially implies DP but not conversely, and that RM implies DR which implies NR but neither conversely, proposition 3.10 gives us the following diagram, where one condition implies another, given P, iff one can follow arrows from the former to the latter.



This proposition suggests a central role for DP *alias* RT. We verify the components of the proposition separately. The positive parts are first proven syntactically, the negative parts are then established semantically.

Observation 3.11 *P + RT ⇔ P + DP*

Proof *(⇒) Suppose* $\alpha \mid\!\sim \gamma$ *and* $\alpha \wedge \beta \not\mid\!\sim \neg\gamma$*. We want to show* $\alpha\wedge\beta \mid\!\sim \gamma$*. By preferentiality,* $\alpha\wedge\beta \mid\!\sim \alpha$*. Thus we have* $\alpha\wedge\beta \mid\!\sim \alpha$*,* $\alpha \mid\!\sim \gamma$ *and* $\alpha\wedge\beta \not\mid\!\sim \neg\gamma$*. So by RT* $\alpha\wedge\beta \mid\!\sim \gamma$ *as desired.*

(⇐) Suppose $\alpha \mid\!\sim \beta$*,* $\beta \mid\!\sim \gamma$ *and* $\alpha \not\mid\!\sim \neg\gamma$*. We want to show* $\alpha \mid\!\sim \gamma$*. Now,* $\alpha\wedge\beta \not\mid\!\sim \neg\gamma$ *for otherwise since* $\alpha \mid\!\sim \beta$ *we would have by cut that* $\alpha \mid\!\sim \neg\gamma$ *contrary to supposition. Hence since* $\beta \mid\!\sim \gamma$ *we have by DP* $\alpha\wedge\beta \mid\!\sim \gamma$*, and so since* $\alpha \mid\!\sim \beta$ *we have by cut that* $\alpha \mid\!\sim \gamma$ *as desired.* ∎

Observation 3.12 *P + RT ⇒ RM. That is, a rational transitive relation is indeed a rational relation.*

Proof *This is a corollary of observation 3.11 and the fact that P + DP ⇒ RM proven in [69]. Here we give a direct proof. Assume* $\alpha \mid\!\sim \beta$ *and* $\alpha \not\mid\!\sim \neg\gamma$*. We will show* $\alpha\wedge\gamma \mid\!\sim \beta$*. First we show* $\alpha\wedge\gamma \not\mid\!\sim \neg\beta$*. Suppose that it is not true,* i.e. $\alpha\wedge\gamma \mid\!\sim \neg\beta$*. Then, by S,* $\alpha \mid\!\sim \gamma \rightarrow \neg\beta$*. Since* $\alpha \mid\!\sim \beta$*, by AND and RW we get* $\alpha \mid\!\sim \neg\gamma$*, a contradiction. Second, we have* $\alpha\wedge\gamma \mid\!\sim \alpha$ *because of REF and RW. Finally, since* $\alpha\wedge\gamma \mid\!\sim \alpha$*,* $\alpha \mid\!\sim \beta$ *and* $\alpha\wedge\gamma \not\mid\!\sim \neg\beta$*, we conclude using RT.* ∎

Observation 3.13 *P + DP ⇒ RC*

Proof *Suppose* $\alpha \mid\!\sim \beta$*; we want to show that either* $\neg\beta \mid\!\sim \neg\alpha$ *or* $\neg\beta \mid\!\sim \alpha$*.*
Case 1: *Suppose* $\top \not\mid\!\sim \beta$*. Now by preferentiality from* $\alpha \mid\!\sim \beta$ *we have* $\top \mid\!\sim \alpha \rightarrow \beta$*. Hence, applying RM (which we noted follows from DP) we have* $\top\wedge\neg\beta \mid\!\sim \alpha \rightarrow \beta$ *so by preferentiality* $\neg\beta \mid\!\sim \neg\alpha$ *as desired.*
Case 2: *Suppose* $\top \mid\!\sim \beta$*. Then by preferentiality* $\top \mid\!\sim \neg\beta \rightarrow \neg\alpha$ *so by the hypothesis DP either* $\top\wedge\neg\beta \mid\!\sim \neg\beta \rightarrow \neg\alpha$ *or* $\top\wedge\neg\beta \mid\!\sim \neg(\neg\beta \rightarrow \neg\alpha)$*.*

Subcase 2.1: *Suppose* $\top\wedge\neg\beta \mid\!\sim \neg\beta \rightarrow \neg\alpha$*. Then by preferentiality* $\neg\beta \mid\!\sim \neg\alpha$ *as desired.*

Subcase 2.1: *Suppose* $\top\wedge\neg\beta \mid\!\sim \neg(\neg\beta \rightarrow \neg\alpha)$*. Then by preferentiality* $\neg\beta \mid\!\sim \alpha$ *as desired.* ∎

Observation 3.14 $P + RC \Rightarrow WD$

Proof *Let* $\mid\!\sim$ *be an inference relation satisfying the preferential rules, and suppose that it fails WD.*

Since $\mid\!\sim$ *fails WD, there are* α, β *with* $\top\mid\!\sim\neg\alpha$, $\alpha\not\mid\!\sim\beta$, $\alpha\not\mid\!\sim\neg\beta$. *Since* $\top\mid\!\sim\neg\alpha$ *we have by preferential rules that* $\top\mid\!\sim\neg\alpha\vee\neg\beta$, *and so combining this with* $\top\mid\!\sim\neg\alpha$ *again we have by CM that* $\neg\alpha\vee\neg\beta\mid\!\sim\neg\alpha$.

On the other hand, since $\alpha\not\mid\!\sim\beta$ *we have by preferential rules that* $\alpha\not\mid\!\sim\alpha\wedge\beta$ *and so* $\neg\neg\alpha\not\mid\!\sim\neg(\neg\alpha\vee\neg\beta)$. *Also since* $\alpha\not\mid\!\sim\neg\beta$ *we have by preferentiality that* $\alpha\not\mid\!\sim\neg\alpha\vee\neg\beta$ *and so* $\neg\neg\alpha\not\mid\!\sim\neg\alpha\vee\neg\beta$.

Putting these three facts together we see that RC fails. ∎

Observation 3.15 $P + RM + WD \Rightarrow DP$

Proof *Suppose* $\alpha\mid\!\sim\beta$ *and* $\alpha\wedge\gamma\not\mid\!\sim\neg\beta$; *we want to show* $\alpha\wedge\gamma\mid\!\sim\beta$. *If* $\top\mid\!\sim\neg(\alpha\wedge\gamma)$ *then the second hypothesis with WD give us what we want. On the other hand, if* $\top\not\mid\!\sim\neg(\alpha\wedge\gamma)$ *then noting from the first hypothesis that* $\top\mid\!\sim\alpha\rightarrow\beta$ *we get by RM that* $\top\wedge(\alpha\wedge\gamma)\mid\!\sim\alpha\rightarrow\beta$ *so by preferential rules,* $\alpha\wedge\gamma\mid\!\sim\beta$ *as desired.* ∎

Observation 3.16 $P + RM + RC \Rightarrow DP$

Proof *This follows immediately from observations 3.14 and 3.15. For another verification, suppose that* $\alpha\mid\!\sim\beta$. *We want to show that either* $\alpha\wedge\gamma\mid\!\sim\beta$ *or* $\alpha\wedge\gamma\mid\!\sim\neg\beta$. *Now either* $\alpha\mid\!\sim\neg\gamma$ *or* $\alpha\not\mid\!\sim\neg\gamma$.

Case 1: *Suppose* $\alpha\not\mid\!\sim\neg\gamma$. *Then by the hypothesis RM we have* $\alpha\wedge\gamma\mid\!\sim\beta$ *as desired.*

Case 2: *Suppose* $\alpha\mid\!\sim\neg\gamma$. *Then by preferentiality (rule S)* $\top\mid\!\sim\alpha\rightarrow\neg\gamma$ i.e. $\top\mid\!\sim\neg(\alpha\wedge\gamma)$ *so by NR which holds by proposition 3.7 either* $\beta\mid\!\sim\neg(\alpha\wedge\gamma)$ *or* $\neg\beta\mid\!\sim\neg(\alpha\wedge\gamma)$, *and in each of these two subcases RC tells us that either* $\alpha\wedge\gamma\mid\!\sim\beta$ *or* $\alpha\wedge\gamma\mid\!\sim\neg\beta$, *as desired.* ∎

Observation 3.17 $P + RC \Rightarrow NR$

Proof *We have already that* $P + RC$ *implies WD (observation 3.14). So it will be enough to prove the following two facts:*

Fact 1. $P + RC \Rightarrow RC^+$, *where* RC^+ *is the following rule*

$$\frac{\alpha\wedge\beta\mid\!\sim\gamma \quad \alpha\wedge\neg\gamma\not\mid\!\sim\beta}{\alpha\wedge\neg\gamma\mid\!\sim\neg\beta}$$

Fact 2. $P + RC^+ + WD \Rightarrow NR$

Proof of fact 1: *Suppose RC holds, and suppose* $\alpha\wedge\beta\mid\!\sim\gamma$. *Then by preferential rules,* $\alpha\wedge\beta\mid\!\sim\gamma\vee\neg\alpha$, *so by RC either* $\neg(\gamma\vee\neg\alpha)\mid\!\sim\neg(\alpha\wedge\beta)$ *or* $\neg(\gamma\vee\neg\alpha)\mid\!\sim\alpha\wedge\beta$. *In the former case we have by preferential rules* $\alpha\wedge\neg\gamma\mid\!\sim\neg\alpha\vee\neg\beta$, *so by preferential rules again* $\alpha\wedge\neg\gamma\mid\!\sim\neg\beta$ *as desired. In the latter case we have by preferential rules* $\alpha\wedge\neg\gamma\mid\!\sim\alpha\wedge\beta$ *so* $\alpha\wedge\neg\gamma\mid\!\sim\beta$ *as desired.*

Proof of fact 2: *Suppose* RC^+ *and WD hold, and suppose* $\alpha\mid\!\sim\beta$*; we want to show that either* $\alpha\wedge\gamma\mid\!\sim\beta$ *or* $\alpha\wedge\neg\gamma\mid\!\sim\beta$*. Since* $\alpha\mid\!\sim\beta$ *we have by preferential rules that* $\top\mid\!\sim\neg\alpha\vee\beta$*. Hence by WD either* $\neg(\neg\alpha\vee\beta)\mid\!\sim\gamma$ *or* $\neg(\neg\alpha\vee\beta)\mid\!\sim\neg\gamma$, i.e. *either* $\alpha\wedge\neg\beta\mid\!\sim\gamma$ *or* $\alpha\wedge\neg\beta\mid\!\sim\neg\gamma$*.*
Case 1. Suppose $\alpha\wedge\neg\beta\mid\!\sim\gamma$*. Then by* RC^+*, either* $\alpha\wedge\neg\gamma\mid\!\sim\beta$ *or* $\alpha\wedge\neg\gamma\mid\!\sim\neg\beta$*. In the former subcase we are done. In the latter subcase by preferential rules (rule S)* $\alpha\mid\!\sim\gamma\vee\neg\beta$ *which combined with* $\alpha\mid\!\sim\beta$ *gives* $\alpha\mid\!\sim\gamma$*. But this again combined with* $\alpha\mid\!\sim\beta$ *gives, by CM,* $\alpha\wedge\gamma\mid\!\sim\beta$ *and in this subcase we are also done.*
Case 2. Suppose $\alpha\wedge\neg\beta\mid\!\sim\neg\gamma$*. Then by* RC^+*, either* $\alpha\wedge\gamma\mid\!\sim\beta$ *or* $\alpha\wedge\gamma\mid\!\sim\neg\beta$*. In the former subcase we are done. In the latter subcase by preferential rules we have* $\alpha\mid\!\sim\beta\rightarrow\neg\gamma$ *which combined with* $\alpha\mid\!\sim\beta$ *gives again by preferential rules* $\alpha\mid\!\sim\neg\gamma$*. From this, we conclude as above using CM that* $\alpha\wedge\neg\gamma\mid\!\sim\beta$ *and we are also done.* ∎

Given the above positive parts of proposition 3.10, it suffices to show the following negative ones: $\mathsf{P}+\mathsf{RM}\not\Rightarrow\mathsf{WD}$, $\mathsf{P}+\mathsf{RC}\not\Rightarrow\mathsf{DR}$, $\mathsf{P}+\mathsf{WD}\not\Rightarrow\mathsf{NR}$.

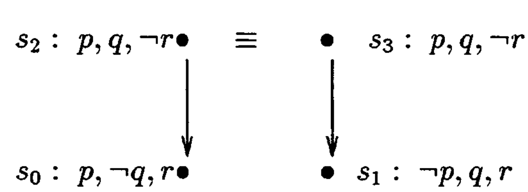
Observation 3.18 *Let* $\mathcal{M}=\langle S,\imath,\prec\rangle$ *be any preferential model with* $S=\{s_0,s_1,s_2,s_3\}$*,* $\prec=\{(s_0,s_2),(s_1,s_3)\}$ *and* $\imath(s_2)=\imath(s_3)$*. Then* $\mid\!\sim_{\mathcal{M}}$ *satisfies RC.*

Proof *Assume* $\alpha\mid\!\sim\beta$ *and* $\neg\beta\not\mid\!\sim\neg\alpha$*. We want to show* $\neg\beta\mid\!\sim\alpha$*. It will be enough to see that* $mod(\alpha)=S$*. Note that* $min(\neg\beta)\cap mod(\alpha)\neq\emptyset$ *because* $\neg\beta\not\mid\!\sim\neg\alpha$*. But neither* s_0 *nor* s_1 *can be in* $min(\neg\beta)\cap mod(\alpha)$ *for otherwise as* s_0 *and* s_1 *are minimals we would have* $\alpha\not\mid\!\sim\beta$ *contradicting our assumption* $\alpha\mid\!\sim\beta$*. So, either* s_2 *or* s_3 *is in* $min(\neg\beta)\cap mod(\alpha)$*, so since* $\imath(s_2)=\imath(s_3)$ *we have both* $s_2,s_3\in mod(\neg\beta)\cap mod(\alpha)$*. This and* $\alpha\mid\!\sim\beta$ *imply by smoothness that* $s_0,s_1\in min(\alpha)$*. Thus* $mod(\alpha)=S$*, as desired.* ∎

Note that the hypothesis $\imath(s_2)=\imath(s_3)$ is necessary in this observation. We can easily find a preferential model $\mathcal{M}=\langle S,\imath,\prec\rangle$ with $S=\{s_0,s_1,s_2,s_3\}$, $\prec=\{(s_0,s_2),(s_1,s_3)\}$ and $\imath(s_2)\neq\imath(s_3)$ which does not satisfy RC.

Observation 3.19 $P+RC\not\Rightarrow DR$

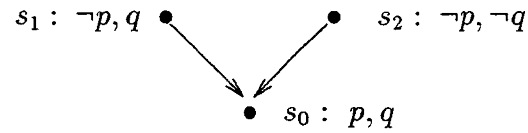
Proof *Consider a model as in the previous observation with* $\imath(s_2)=\imath(s_3)=\{p,q\}$*,* $\imath(s_0)=\{p,r\}$*,* $\imath(s_1)=\{q,r\}$ *(we give the valuations as for a Herbrand model, that is identifying the subset of variables with its characteristic function). Graphically*



By observation 3.18 RC holds in $\mathcal{M}$ *but it is clear that* $p\not\mid\!\sim r$ *(witness* s_3*),* $q\not\mid\!\sim r$ *(witness* s_2*) and* $p\vee q\mid\!\sim r$ *so DR fails.* ∎

Observation 3.20 $P+RM\not\Rightarrow WD$

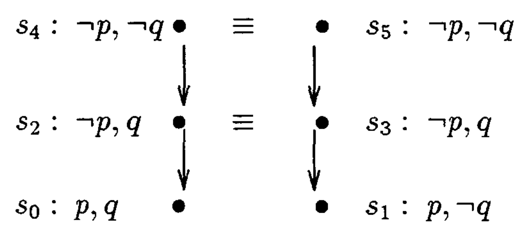
Proof *Let $\mathcal{L}$ be the set of all formulae built from the elementary formulae p and q. Consider the following model $\mathcal{M} = \langle S, \imath, \prec \rangle$ where $S = \{s_0, s_1, s_2\}$, $\prec = \{(s_0, s_1), (s_0, s_2)\}$, $\imath(s_0) = \{p, q\}$, $\imath(s_1) = \{q\}$ and $\imath(s_2) = \emptyset$. Graphically*



It is clear that $\mathcal{M}$ is ranked, so it satisfies RM. However it fails to satisfy WD since $\top \mid\!\sim p$ whilst $\neg p \not\mid\!\sim q$ and $\neg p \not\mid\!\sim \neg q$. ∎

Observation 3.21 *P + WD $\not\Rightarrow$ NR*

Proof *Let $\mathcal{M}$ be the model represented by the following schema:*



i.e. $\mathcal{M} = \langle S, \imath, \prec \rangle$ *with* $S = \{s_0, \ldots, s_5\}$, $\prec$ *the transitive closure of the relation* $\{(s_0, s_2), (s_2, s_4), (s_1, s_3), (s$
$\imath(s_0) = \{p, q\}$, $\imath(s_1) = \{p\}$, $\imath(s_2) = \imath(s_3) = \{q\}$ *and* $\imath(s_4) = \imath(s_5) = \emptyset$.
NR fails in this model because $\top \mid\!\sim p$, $q \not\mid\!\sim p$ (witness s_3), $\neg q \not\mid\!\sim p$ (witness s_4). But WD is satisfied in this model. Suppose that $\top \mid\!\sim \neg\alpha$, $\alpha \not\mid\!\sim \beta$, $\alpha \not\mid\!\sim \neg\beta$. Then there must be $u \in min(\alpha) \cap mod(\neg\beta)$ and $v \in min(\alpha) \cap mod(\beta)$ and $u, v \in \{s_2, s_3, s_4, s_5\}$. But it is clear that each choice of u, v here gives a contradiction. For example if $u = s_2$ and $v = s_5$ then since $\imath(s_2) = \imath(s_3)$ we have $v \notin min(\alpha)$ giving a contradiction. ∎

Observations 3.19 and 3.21 have been established using non-injective preferential models as examples. In the case of 3.19, at least, there is no injective model that does the job. For by 3.17, any injective model of P + RC is an injective model of P + NR, and it has been shown by Freund and Lehmann [32], that every injective model of P + NR is a model of DR.

It is immediate that transitivity (T) of $\mid\!\sim$ implies RT. However, the converse does not hold: given that P + RT $\Leftrightarrow$ P + DP shown above, and the well-known facts (see *e.g.* [69]) that P + T $\Rightarrow$ M whilst P + DP $\not\Rightarrow$ M, we have P + RT $\not\Rightarrow$ T. A direct verification can also be made with an appropriate two-state model (see corollary 3.33).

As already remarked, DP, RT, RC and WD are weakened forms not only of monotony but also of Stalnaker's rule of conditional excluded middle which, unlike the principles so far considered, is not implied by monotony but has figured in philosophical discussion of counterfactuals (*e.g.* [43, 57, 81]). We shall study some other rules in the vicinity of conditional excluded middle in section 3.7.

3.4 Collapsed models

Our goal in section 3.5 will be to prove a representation theorem for P + RT (equivalently P + DP). As a preliminary, we shall show in this section that every ranked preferential model is equivalent (in the sense of generating the same inference relation) to one that is both injective and *parsimonious*, in the sense that every one of its states is minimal in at least one formula. This result is indeed implicit in Freund [30] in the more general case of relations satisfying P + DR, but using different arguments. Our procedure for transforming a ranked model into one with these characteristics is quite straightforward. We proceed in two steps. First, at each level of the model we identify the states of that level that are labelled with the same valuation. Second, we suppress all states that are not minimal in some formula.

Definition 3.22 *A model* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *is said to be horizontally injective iff for all distinct* $s, t \in S$, *if* $s \not\prec t$ *and* $t \not\prec s$ *then* $\imath(s) \neq \imath(t)$.

Note that for ranked models, being horizontally injective actually means injectivity by levels.

Lemma 3.23 *For any ranked model* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *there exists a horizontally injective ranked model* $\mathcal{M}' = \langle S', \imath', \prec' \rangle$ *such that* $\mid\!\sim_{\mathcal{M}} = \mid\!\sim_{\mathcal{M}'}$.

Proof *We define an equivalence relation on* S *as follows*

$$s \equiv s' \Leftrightarrow r(s) = r(s') \text{ and } \imath(s) = \imath(s')$$

Put $S' = S/\equiv$ *(the quotient of* S *by* $\equiv$*). As usual let* $[s]$ *denote the equivalent class of* s. *Define* $r' : S' \longrightarrow \Omega$, $\prec' \subseteq S' \times S'$ *and* $\imath' : S' \longrightarrow \mathcal{U}$ *as follows:* $r'([s]) = r(s)$, $[s] \prec' [s']$ *iff* $s \prec s'$, *and* $\imath'([s]) = \imath(s)$. *It can be easily verified that* $r', \prec'$ *and* $\imath'$ *are well defined,* i.e. *their definition does not depend on the choice of the representative of* $[s]$. *It is also clear that for all* $j \in \Omega$, $\imath'$ *restricted to* $S'_j = \{[s] \in S' : r([s]) = j\}$ *is injective. Notice that*

$$[s] \prec' [s'] \Leftrightarrow s \prec s' \Leftrightarrow r(s) < r(s') \Leftrightarrow r'([s]) < r'([s'])$$

So, the model $\mathcal{M}'$ *defined by* $\mathcal{M}' = \langle S', \imath', \prec' \rangle$ *is a ranked model. Moreover, we have* $\mid\!\sim_{\mathcal{M}} = \mid\!\sim_{\mathcal{M}'}$, *since clearly for all* $s \in S$ *and all* $\alpha, \beta \in \mathcal{L}$, $s \in mod_{\mathcal{M}}(\beta)$ *iff* $[s] \in mod_{\mathcal{M}'}(\beta)$, *and also* $s \in min_{\mathcal{M}}(\alpha)$ *iff* $[s] \in min_{\mathcal{M}'}(\alpha)$. ∎

Definition 3.24 *A model* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *is said to be parsimonious iff for every state* $s \in S$ *there is a formula* α *such that* $s \in min_{\mathcal{M}}(\alpha)$.

Proposition 3.25 *If* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *is a preferential model then there exists a preferential model* $\mathcal{M}' = \langle S', \imath', \prec' \rangle$ *such that* $S' \subseteq S$ *and the following properties hold*

1. $\mathcal{M}'$ *is parsimonious.*

2. *If* $\mathcal{M}$ *is ranked so is* $\mathcal{M}'$.

3. *Whenever* $s, t \in S'$ *with neither* $s \prec' t$ *nor* $t \prec' s$, *if* $\imath'(s) = \imath'(t)$ *then* $\imath(s) = \imath(t)$.

4. $\mid\!\sim_{\mathcal{M}} = \mid\!\sim_{\mathcal{M}'}$

Proof *It is quite simple. It is enough to suppress the states which are not minimals for any formula. Essentially the same trick has been used by Pavlos Peppas [85] in the context of systems-of-spheres models for belief revision. Define* $S' = S \setminus \{s : \neg\exists\alpha \ s \in min_{\mathcal{M}}(\alpha)\}$. *Let* $\imath'$ *and* $\prec'$ *be the restrictions of* $\imath$ *and* $\prec$ *to* S'. *Put* $\mathcal{M}' = \langle S', \imath', \prec' \rangle$. *By definition of* $\mathcal{M}'$, *it is obvious that*

$$min_{\mathcal{M}}(\alpha) = min_{\mathcal{M}'}(\alpha) \qquad (*)$$

Hence smoothness of $\mathcal{M}$ *implies smoothness of* $\mathcal{M}'$. *So,* $\mathcal{M}'$ *is a preferential model which, by its construction, is parsimonious. Clearly if* $\mathcal{M}$ *is ranked so is* $\mathcal{M}'$. *Property 3 is trivially verified by definition of* $\mathcal{M}'$. *Finally, let us verify* $\alpha \mid\!\sim_{\mathcal{M}} \beta \Leftrightarrow \alpha \mid\!\sim_{\mathcal{M}'} \beta$.
($\Leftarrow$): *Suppose that* $min_{\mathcal{M}'}(\alpha) \subseteq mod_{\mathcal{M}'}(\beta)$. *We want to show* $min_{\mathcal{M}}(\alpha) \subseteq mod_{\mathcal{M}}(\beta)$. *This follows from* $(*)$ *and the fact that* $mod_{\mathcal{M}'}(\beta) \subseteq mod_{\mathcal{M}}(\beta)$.
($\Rightarrow$): *Suppose that* $min_{\mathcal{M}}(\alpha) \subseteq mod_{\mathcal{M}}(\beta)$. *By definition of* S' $min_{\mathcal{M}}(\alpha) \subseteq mod_{\mathcal{M}'}(\beta)$ *because* $mod_{\mathcal{M}'}(\beta) = \{s \in mod_{\mathcal{M}}(\beta) : \exists\gamma \ s \in min_{\mathcal{M}}(\gamma)\}$. *So, by* $(*)$, $min_{\mathcal{M}'}(\alpha) \subseteq mod_{\mathcal{M}'}(\beta)$, i.e. $\alpha \mid\!\sim \beta$.
∎

Remark 3.26 *When the ranked model* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *is finite (notice that by theorem 3.6 every rational relation over a finite language is represented by a finite model),* i.e. S *is finite, then* S' *in the previous lemma can be constructed by an algorithm. In order to see this, first remark that when* S *is finite, we can suppose the rank function is of the form* $r : S \longrightarrow \overline{n}$. *Then define* $S'_0 = \{s \in S : r(s) = 0\}$ *and for* $k = 1$ *to* n, $S'_k = \{s \in S : r(s) = k$ *and there exists* α *with* $s \in min_{\mathcal{M}}(\alpha)\}$. *Finally put* $S' = \cup_{k=0}^{n} S'_k$.

Theorem 3.27 (Collapsing) *If* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *is a ranked model then there exists a parsimonious, injective ranked model* $\mathcal{M}' = \langle S', \imath', \prec' \rangle$ *such that* $\mid\!\sim_{\mathcal{M}} = \mid\!\sim_{\mathcal{M}'}$.

Proof *Let* $\mathcal{M}' = \langle S', \imath', \prec' \rangle$ *be the model obtained from* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *by application of lemma 3.23 and then proposition 3.25. Clearly* $\mathcal{M}'$ *is parsimonious and ranked, and also* $\mid\!\sim_{\mathcal{M}} = \mid\!\sim_{\mathcal{M}'}$. *It remains to check injectivity. Now by lemma 3.23 and part (3) of proposition 3.25,* $\mathcal{M}'$ *is horizontally injective. Clearly parsimony implies that* $\mathcal{M}'$ *is 'vertically injective' in the sense that* $s \prec' t$ *implies* $\imath(s) \neq \imath(t)$. *Finally, horizontal and vertical injectivity clearly imply injectivity.* ∎

The model $\mathcal{M}'$ obtained from a ranked model $\mathcal{M}$ by successive application of lemma 3.23, and proposition 3.25 will be called the *collapse* of $\mathcal{M}$. Clearly a model is equal to its collapse iff it is both parsimonious and injective. As a corollary of theorem 3.27 we have the following result:

Corollary 3.28 *Every rational inference relation is generated by some collapsed ranked preferential model.*

Proof *Immediate from the Lehmann-Magidor representation theorem (theorem 3.6) and theorem 3.27.* ∎

Remark 3.29 *(i) In the proof of theorem 3.27 we have applied lemma 3.23 and then proposition 3.25 but the same result is obtained if we reverse the order.*
(ii) It is not hard to see that in the finite case (finite language), if a model is injective then

each state is minimal for some formula, so the model is parsimonious. But this is not true in general for infinite languages. For instance consider an injective ranked model with two levels: one world in the upper level and the rest of the worlds in the lower level; then there is no formula for which the upper world is minimal.

Remark 3.30 *Theorem 3.27 and its corollary 3.28 are implicit in Freund [30] but in reverse order of demonstration and by quite a different strategy. Freund shows that if* $\mid\!\sim$ *is a rational inference relation (indeed more generally, any preferential inference relation satisfying DR) then we can construct its 'associated standard model', which is a model generating* $\mid\!\sim$*, that is ranked (or, under the hypothesis of DR, that is 'filtered' in the sense of his definition 5.1) and has additional properties including the following:*

1. *Every state* u *is '*$\mid\!\sim$*-consistent' in the sense (Freund [30], section 2.1) that there is a formula* α *with* $u \models C(\alpha)$*, where as usual* $C(\alpha) = \{\beta : \alpha \mid\!\sim \beta\}$*;*

2. *The model is 'standard with respect to* $\mid\!\sim$*' in the sense (Freund [30], definition 3.2) that it is injective and for every formula* α *and state* u*,* $u \models C(\alpha)$ *iff* $u \in min(\alpha)$*.*

Property 2 explicitly implies injectivity, and the two properties taken together clearly imply parsimony. Conversely, parsimony and injectivity together imply properties 1 and 2, if we assume that the model is ranked: property 1 is immediate from parsimony recalling that $u \in min(\alpha)$ *immediately implies* $u \models C(\alpha)$ *in every preferential model, whilst to derive property 2 it suffices to show that whenever* $u \notin min(\alpha)$ *then* $u \not\models C(\alpha)$*. Suppose* $u \notin min(\alpha)$*. If* $u \not\models \alpha$ *then we are done, so suppose that* $u \models \alpha$*. Then there is* $v \prec u$ *with* $v \in min(\alpha)$*. By parsimony, there exists* β *such that* $u \in min(\beta)$ *and thus by rankedness for any* $v' \in min(\alpha)$*,* $v' \models \neg\beta$*. Thus* $\neg\beta \in C(\alpha)$ *and so since* $u \models \beta$ *we have* $u \not\models C(\alpha)$*, establishing property 2.*

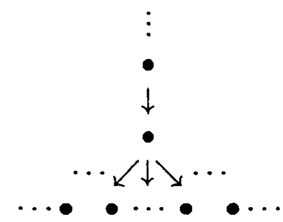
*Evidently, each approach has its advantages, depending in part on the purposes for which is used. Since our approach covers only ranked models and thus rationally monotone inference relations, unless it can be generalized it is useless for Freund's purpose, which is to represent preferential inference relations satisfying DR. On the other hand, it provides a simple and natural way of proving representation theorems for conditions such as rational transitivity that are stronger than rational monotony (*e.g. *theorems 3.38, 3.68 and 3.74 below) and a very direct argument for results of independent model-theoretic interest such as lemma 3.23, proposition 3.25, theorem 3.27 and its corollary 3.28.*

3.5 Representation

The goal of this section is to characterize the ranked models that generate rational transitive relations. Our argument exploits corollary 3.28

Definition 3.31 *A preferential model (not necessarily injective)* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *is said to be* quasi-linear *iff it is ranked and it has at most one state at any level above the lowest. In other words quasi-linear means ranked and whenever* $r \prec s$*,* $r \prec t$ *then either* $s = t$ *or* $s \prec t$ *or* $t \prec s$*.*

Quasi-linear models have the following graphical shape:



Proposition 3.32 *If $\mathcal{M} = \langle S, \imath, \prec\rangle$ is quasi-linear then the relation $|\!\!\sim \; = \; |\!\!\sim_{\mathcal{M}}$ is rational transitive.*

Proof *$\mathcal{M}$ is ranked so $|\!\!\sim$ is a rational relation. We have to prove that $|\!\!\sim$ satisfies RT. So, suppose $\alpha |\!\!\sim \beta$, $\beta |\!\!\sim \gamma$ and $\alpha \not|\!\!\sim \neg\gamma$. We want to show $\alpha |\!\!\sim \gamma$. We consider two cases. First, suppose that min(α) is contained in the lowest level. As also $\alpha |\!\!\sim \beta$, necessarily min(α) $\subseteq$ min(β). But min(β) $\subseteq$ mod(γ) because $\beta |\!\!\sim \gamma$. Therefore, min(α) $\subseteq$ mod(γ),* i.e. *$\alpha |\!\!\sim \gamma$.*

Second, suppose min(α) is not contained in the lowest level. Then min(α) is a singleton because $\mathcal{M}$ is quasi-linear; suppose min(α) $= \{s\}$. Then $s \not\models \neg\gamma$ because $\alpha \not|\!\!\sim \neg\gamma$. Thus $s \models \gamma$, so $\alpha |\!\!\sim \gamma$. ∎

Corollary 3.33 *There are rational transitive inference relations which are not transitive.*

Proof *Consider a language $\mathcal{L}$ built on the propositional variables p, q and r. Define $\mathcal{M} = \langle S, \imath, \prec\rangle$ where $S = \{s_0, s_1\}$, $s_0 \prec s_1$ $\imath(s_0) = \{q, r\}$, $\imath(s_1) = \{p, q\}$. By proposition 3.32 the relation $|\!\!\sim \; = \; |\!\!\sim_{\mathcal{M}}$ is a rational transitive relation. But we can easily verify, $p |\!\!\sim q$, $q |\!\!\sim r$ and $p \not|\!\!\sim r$. So $|\!\!\sim$ is not a transitive relation.* ∎

Observation 3.34 *Suppose that the language is finite and $\mathcal{M} = \langle S, \imath, \prec\rangle$ is an injective ranked model which is not quasi-linear. Then $|\!\!\sim \; = \; |\!\!\sim_{\mathcal{M}}$ does not satisfy RT.*

Proof *As $\mathcal{M}$ is not quasi-linear, necessarily there are three different states s_1, s_2 and s_3 in S such that s_1 is in the lowest level, s_2 and s_3 are in the same level and $s_1 \prec s_i$ for $i = 2, 3$. Let α, β and γ be formulae such that mod(α) $= \{s_2, s_3\}$, mod(β) $= \{s_1, s_2, s_3\}$ and mod(γ) $= \{s_1, s_2\}$. By finiteness and injectivity such α, β and γ clearly exist. Then, it is clear that $\alpha |\!\!\sim \beta$, $\beta |\!\!\sim \gamma$ whilst $\alpha \not|\!\!\sim \neg\gamma$ and $\alpha \not|\!\!\sim \gamma$. Therefore $\mathcal{M}$ does not satisfy RT.* ∎

Remark 3.35 *When the language is infinite the above observation does not hold. This can be seen by the following example. Let $\mathcal{M}$ be a ranked model whose states are worlds, with two levels: v_0 and v_1 in the upper level and the rest of the valuations in the lower level,* i.e. *the order is $v \prec v_i$ for all valuations $v \neq v_i$, $i = 1, 2$. By definition $\mathcal{M}$ is not quasi-linear. However, $|\!\!\sim_{\mathcal{M}}$ satisfies RT (and indeed, satisfies transitivity and monotony) because for any formula α, min$_{\mathcal{M}}(\alpha)$ lies in the lowest level.*

But if instead of injectivity in the observation 3.34 we require that the model $\mathcal{M}$ be collapsed then a similar argument can be used to extend observation 3.34 to the case of infinite languages. More precisely we have the following proposition:

Proposition 3.36 *Suppose $\mathcal{M} = \langle S, \imath, \prec \rangle$ is a collapsed ranked model which is not quasi-linear. Then $\mid\!\sim\, = \,\mid\!\sim_{\mathcal{M}}$ does not satisfy RT.*

Proof *As $\mathcal{M}$ is not quasi-linear, necessarily there are three different states s_1, s_2 and s_3 in S such that s_1 is in the lowest level, s_2 and s_3 are in the same level and $s_1 \prec s_i$ for $i = 2, 3$. We need to find formulae α, β, γ with $\alpha \mid\!\sim \beta$, $\beta \mid\!\sim \gamma$, $\alpha \not\mid\!\sim \neg\gamma$, $\alpha \not\mid\!\sim \gamma$. By parsimony there are formulae ϕ_i ($i = 1, 2, 3$) with $s_i \in min(\phi_i)$ and by injectivity there are formulae ψ_{ij} ($i, j = 1, 2, 3$ and $i \neq j$) with $s_i \in mod(\psi_{ij})$ and $s_j \notin mod(\psi_{ij})$. Put $\alpha = \phi_2 \vee \phi_3$, $\beta = (\phi_2 \vee \phi_3) \vee (\phi_1 \wedge \psi_{12})$ and $\gamma = (\phi_1 \wedge \psi_{12}) \vee (\phi_3 \wedge \psi_{32})$. Then it is clear that $\alpha \vdash \beta$ so $\alpha \mid\!\sim \beta$; and using rankedness $min(\beta) \subseteq mod(\phi_1 \wedge \psi_{12})$ so $\beta \mid\!\sim \gamma$; whilst again using rankedness $s_2 \in min(\alpha)$ but $s_2 \notin mod(\gamma)$ so $\alpha \not\mid\!\sim \gamma$ and finally $s_3 \in min(\alpha)$ but $s_3 \in mod(\gamma)$ so $\alpha \not\mid\!\sim \neg\gamma$.* ∎

Propositions 3.32 and 3.36 immediately imply:

Theorem 3.37 *Let $\mathcal{M}$ be a collapsed ranked model. Then $\mathcal{M}$ is quasi-linear iff $\mid\!\sim_{\mathcal{M}}$ is rational transitive.*

This with corollary 3.28 immediately imply the promised representation theorem for rational transitive relations:

Theorem 3.38 *$\mid\!\sim$ is a rational transitive relation iff there is a quasi-linear model $\mathcal{M}$ such that $\mid\!\sim\, = \,\mid\!\sim_{\mathcal{M}}$.*

Putting together theorem 3.38 and proposition 3.10 we clearly have:

Theorem 3.39 *The following conditions are equivalent for any preferential inference relation $\mid\!\sim$:*

1. *$\mid\!\sim$ is determined by some quasi-linear model.*
2. *$\mid\!\sim$ is determinacy preserving.*
3. *$\mid\!\sim$ is rational transitive.*
4. *$\mid\!\sim$ satisfies both RM and RC.*
5. *$\mid\!\sim$ satisfies both RM and WD.*

Remark 3.40 *The above results leave open the question of representation theorems for the weaker postulate sets P + RC and P + WD. It may be noted that the techniques used above do not appear to carry over in a straightforward way to those systems. Lemma 3.23 (used for theorem 3.27 and thus corollary 3.28 and thus theorem 3.38) is here proven only for ranked preferential models, and even if the less direct techniques of Freund [30] are used (cf. remark 3.30) their scope covers only postulate systems at least as strong as P + DR.*

3.6 Preferential Orderings and Rational Transitivity

After seeing Bezzazi and Pino Pérez [9] Michael Freund (personal communication) conjectured theorem 3.49 below, which characterizes rational transitive relations in terms of properties of their preferential orders defined as in [30]. The purpose of this section is to prove that characterization. This gives us another way to obtain the representation theorem 3.38. The results of the subsequent sections do not depend upon this one.

Definition 3.41 *Let $\mid\!\sim$ be a preferential inference relation. Let α and β be formulae. The preferential order associated with $\mid\!\sim$ is defined by*

$$\alpha < \beta \quad \Leftrightarrow \quad \alpha \vee \beta \mid\!\sim \neg\beta$$

The relation $<$ is not a strict order because irreflexivity does not quite hold (for instance $\bot < \bot$). Nevertheless by tradition we conserve the name of preferential order for it.

The following lemma from [30] helps understand better the meaning of this relation:

Lemma 3.42 *1. $\alpha < \beta \Leftrightarrow \alpha \mid\!\sim \neg\beta$ and $\alpha \vee \beta \mid\!\sim \alpha$*

2. $\alpha \mid\!\sim \beta \Leftrightarrow \alpha < \alpha \wedge \neg\beta$

3. $\alpha < \beta$ iff in every preferential model $\mathcal{M} = \langle S, \imath, \prec \rangle$ defining $\mid\!\sim$ the following property holds: for every element $s \in mod(\beta)$ there exists $t \in mod(\alpha)$ such that $t \prec s$.

It is easy to show the following corollary of point 3 of this lemma:

Lemma 3.43 *Let $\mid\!\sim$ be a rational relation defined by a preferential ranked model $\mathcal{M} = \langle S, \imath, \prec \rangle$ with $r : S \longrightarrow \Omega$ the ranking function (Ω linearly ordered by $\lhd$). For any formula α define its level, $\ell(\alpha)$, as ∞ if $\alpha \mid\!\sim \bot$ and otherwise its level is the unique $a \in \Omega$ such that there exists $s \in min_{\mathcal{M}}(\alpha)$ with $r(s) = a$. Then, the level is well defined and $\alpha < \beta$ iff $\ell(\alpha) \lhd \ell(\beta)$.*

We remark that the relation $<$ of [55] (definition A3, first defined in [52]) is equivalent to that of definition 3.41 in the case of rational inference relations; the idea behind these 'orders' has roots in Lewis [57]. In [30] Freund called *preferential order* any relation $<$ on formulae satisfying the following four properties:

P_0: $\alpha < \bot$
P_1: If $\alpha \vdash \beta$, then
(a) $\alpha < \gamma \;\Rightarrow\; \beta < \gamma$
(b) $\delta < \beta \;\Rightarrow\; \delta < \alpha$
P_2: $\alpha < \gamma$ and $\alpha < \delta$ implies $\alpha < \gamma \vee \delta$
P_3: $\alpha \vee \beta < \beta$ implies $\alpha < \beta$

Freund proves that the 'order' associated with a preferential inference relation by definition 3.41 satisfies these properties. Conversely the inference relation $\mid\!\sim$ associated with a relation $<$ satisfying these properties by putting $\alpha \mid\!\sim \beta$ iff $\alpha < \alpha \wedge \neg\beta$ is a preferential inference relation;

moreover the order associated with this inference relation by definition 3.41 coincides with $<$. Thus $<$ satisfies properties P_0-P_3 iff it is the preferential order associated with some preferential inference relation in the sense of definition 3.41.

We recall the definition of modular relation (see [55]):

Definition 3.44 *A relation $<$ on E is said to be modular iff there exists a linear order $\prec$ on some set Ω and a function $r : E \longrightarrow \Omega$ such that $a < b \Leftrightarrow r(a) \prec r(b)$.*

The following characterization of modularity is well-known and easy to verify.

Lemma 3.45 *An order $<$ on E is modular iff for any $a, b, c \in E$ if a and b are incomparables and $a < c$ then $b < c$.*

The following proposition is due to Freund (personal communication).

Proposition 3.46 *$\mid\!\sim$ is a rational relation iff the preferential order between formulae associated by definition 3.41 with $\mid\!\sim$ is modular over the set of $\mid\!\sim$-consistent formulae,* i.e. *those formulae α with $\alpha \not\mid\!\sim \bot$.*

Proof *The* only if *part follows from the representation theorem 3.6 and lemma 3.43. More precisely, by the representation theorem 3.6 there exists a ranked model $\mathcal{M} = \langle S, \imath, \prec \rangle$ with $r : S \longrightarrow \Omega$ the ranking function (Ω linearly ordered by $\lhd$) such that $\mid\!\sim = \mid\!\sim_{\mathcal{M}}$. Define the function ℓ mapping a formula α to its level $\ell(\alpha)$. This mapping and lemma 3.43 prove that $<$ is modular.*

Conversely, suppose that the preferential order between formulae $<$ is modular. Assume $\alpha \mid\!\sim \beta$ and $\alpha \not\mid\!\sim \neg\gamma$. We want to show that $\alpha\wedge\gamma \mid\!\sim \beta$. By part 2 of lemma 3.42, this last expression is equivalent to $\alpha\wedge\gamma < \alpha\wedge\gamma\wedge\neg\beta$ and the assumptions are equivalent to $\alpha < \alpha\wedge\neg\beta$ and $\alpha \not< \alpha\wedge\gamma$. Note that either $\alpha\wedge\gamma < \alpha$ or $\alpha\wedge\gamma \not< \alpha$. In the first case we use Freund's property P_1 (b) to obtain $\alpha\wedge\gamma < \alpha\wedge\gamma\wedge\neg\beta$. In the second case, α and $\alpha\wedge\gamma$ are incomparables because $\alpha \not< \alpha\wedge\gamma$ using part 2 of lemma 3.42 again. So by modularity, $\alpha\wedge\gamma < \alpha\wedge\neg\beta$ because $\alpha < \alpha\wedge\neg\beta$. Now, as before using the property P_1 (b), we obtain $\alpha\wedge\gamma < \alpha\wedge\gamma\wedge\neg\beta$. ∎

The following lemma will be useful:

Lemma 3.47 *Let $\mid\!\sim$ a preferential relation and $<$ its associated preferential order. For any formulae α and β if $\alpha < \beta$ then $\top < \beta$*

Proof *Note that $\alpha \vdash \top$. Suppose $\alpha < \beta$. Then by P_1 (a) we have $\top < \beta$.* ∎

The *quasi-linear property*, QLP in short, for an order $<$ associated to an inference relation $\mid\!\sim$ is the following property: for any formulae α and β, if $\top < \alpha$ then either $\alpha < \beta$ or $\beta < \alpha$ or α is $\mid\!\sim$-equivalent to β, *i.e.* $\alpha \mid\!\sim \beta$ and $\beta \mid\!\sim \alpha$.

Proposition 3.48 *Let $\mathcal{M} = \langle S, \imath, \prec \rangle$ be a ranked collapsed model and put $\mid\!\sim = \mid\!\sim_{\mathcal{M}}$. If the preferential order associated with $\mid\!\sim$ satisfies the property QLP, then the model $\mathcal{M}$ is quasi-linear.*

Proof *Suppose that $\mathcal{M}$ is not quasi-linear. We want to show that the preferential order does not satisfy QLP. As $\mathcal{M}$ is not quasi-linear, there are three different states s_1, s_2 and s_3 in S such that s_1 is in the lowest level, s_2 and s_3 are in the same level and $s_1 \prec s_i$ for $i = 2, 3$. By parsimony there are formulae ϕ_2 and ϕ_3 with $s_2 \in min(\phi_2)$ and $s_3 \in min(\phi_3)$. By injectivity there exists a formula ζ such that $s_2 \in mod(\zeta)$ and $s_3 \in mod(\neg\zeta)$. Put $\alpha = \phi_2 \wedge \zeta$, et $\beta = \phi_3 \wedge \neg\zeta$. Note that $\alpha \not< \beta$ and $\beta \not< \alpha$ because their minimal states lie in the same level. Moreover, $s_2 \in min_{\mathcal{M}}(\alpha)$ but $s_2 \notin min_{\mathcal{M}}(\beta)$, so α and β are not $\mid\!\sim$-equivalent. But it is clear that, $\top < \alpha$. Hence the preferential order $<$ does not satisfy QLP.* ∎

Note that the argument in this proof "translates" the one of proposition 3.36.

Theorem 3.49 *(Conjectured by Freund, personal communication): Let $\mid\!\sim$ be a preferential relation. Then $\mid\!\sim$ satisfies RT iff the preferential order $<$ associated with $\mid\!\sim$ satisfies the property QLP.*

Proof *The* only if *part is deduced from theorem 3.38 as follows. Suppose that $\mid\!\sim$ is rational transitive. Then by theorem 3.38 there is a quasi-linear model $\mathcal{M}$ such that $\mid\!\sim\, = \,\mid\!\sim_{\mathcal{M}}$. We know that $\mathcal{M}$ is ranked* i.e. $\mathcal{M} = \langle S, \imath, \prec\rangle$ *with $r : S \longrightarrow \Omega$ the ranking function (Ω linearly ordered by $\lhd$) Now, if $\alpha \not< \beta$ and $\beta \not< \alpha$ we have by lemma 3.43 $\ell(\alpha) = \ell(\beta) = a \in \Omega$. But if $\top < \alpha$ then by quasi-linearity and lemma 3.43 there is at most one state $s \in S$ such that $r(s) = a$. So, $min_{\mathcal{M}}(\alpha) = min_{\mathcal{M}}(\beta) = \{s\}$ or $min_{\mathcal{M}}(\alpha) = min_{\mathcal{M}}(\beta) = \emptyset$. Hence, in either case, $min_{\mathcal{M}}(\alpha) = min_{\mathcal{M}}(\beta)$,* i.e. *$\alpha$ is $\mid\!\sim$-equivalent to β.*

Now we prove the if *part. Suppose that $\mid\!\sim$ is a preferential relation which satisfies QLP. We want to show that $\mid\!\sim$ satisfies RT. By theorem 3.38, it will be enough to see that $\mid\!\sim$ is represented by a quasi-linear model. In order to do that, we first show that $<$ is modular. Suppose that $\alpha \not< \beta$, $\beta \not< \alpha$, and $\alpha < \gamma$. We want to show that $\beta < \gamma$. By lemma 3.47, $\top < \gamma$. So, by QLP, $\beta < \gamma$ or $\gamma < \beta$ or γ and β are $\mid\!\sim$-equivalent. But we shall see that the last two cases lead to a contradiction.*
Case 1: *Suppose that $\gamma < \beta$. Then by transitivity of $<$ we have $\alpha < \beta$, a contradiction.*
Case 2: *Suppose that γ and β are $\mid\!\sim$-equivalent. Then, in any model $\mathcal{M}$ representing $\mid\!\sim$ we have $min_{\mathcal{M}}(\gamma) = min_{\mathcal{M}}(\beta)$. So by the lemma 3.42 using $\alpha < \gamma$ we conclude that $\alpha < \beta$. We find again a contradiction.*
Therefore the only possibility is $\beta < \gamma$ as desired. As $<$ is modular, by the proposition 3.46, the relation $\mid\!\sim$ is rational. So there is a ranked model $\mathcal{M}$ representing it, and by theorem 3.27 we can suppose that $\mathcal{M}$ is collapsed. Thus by proposition 3.48, $\mathcal{M}$ is quasi-linear. Therefore $\mid\!\sim$ satisfies RT. ∎

Remark 3.50 *We can give a different proof of theorem 3.49 which does not use the representation theorem 3.38. Moreover this proof provides an alternative argument for theorem 3.38.*

Proof *Here we give only a sketch. The argument uses Freund's notion of 'standard model'. The* only if *part,* i.e. *that RT implies QLP, is proven as follows. Suppose that $\top < \alpha$,* i.e. *$\top \mid\!\sim \neg\alpha$. If $\top \not\mid\!\sim \neg\beta$,* i.e. *$\top \not< \beta$, then β lies at the lowest level. So $\beta < \alpha$. If $\top \mid\!\sim \neg\beta$ we have the following situation: $\alpha \vee \beta \mid\!\sim \top$, $\top \mid\!\sim \neg\alpha$ and $\top \mid\!\sim \neg\beta$. Then, by RT we have $\alpha \vee \beta \mid\!\sim \neg\alpha$, or $\alpha \vee \beta \mid\!\sim \neg\beta$, or (when these two possibilities fail) we have both $\alpha \vee \beta \mid\!\sim \alpha$ and $\alpha \vee \beta \mid\!\sim \beta$. So, $\beta < \alpha$ or $\alpha < \beta$ or α and β are $\mid\!\sim$-equivalents.*

The if *part,* i.e. *that QLP implies RT, is proven as follows. By proposition 3.32 it is enough to show that* $\mid\!\sim$ *is represented by a quasi-linear model. The relation* $\mid\!\sim$ *is rational because* $<$ *is modular as remarked earlier. So, by the theorem 6.3 of [30],* $\mid\!\sim$ *is generated by its associated standard model* (cf. *remark 3.30) which is ranked. Moreover, this canonical model is quasi-linear: suppose the canonical model is not quasi-linear,* i.e. *there are two different worlds, m and n, both in a non-minimal level. We want to show that the condition QLP does not hold. By standardness (the canonical model is standard), there are formulae* α, β *(not necessarily different) such that* $m \models C(\alpha)$, $n \models C(\beta)$, $m \in min(\alpha)$ *and* $n \in min(\beta)$. *But,* $m \neq n$ *implies that there is a formula* γ *such that* $m \models \gamma$ *and* $n \models \neg\gamma$. *Put* $\alpha_1 = \alpha\wedge\gamma$ *and* $\beta_1 = \beta\wedge\neg\gamma$. *It is clear that* $m \in min(\alpha_1)$ *and* $n \in min(\beta_1)$, *so the minimal elements of* α_1 *and* β_1 *are at the same level. Therefore* $\alpha_1 \not< \beta_1$ *and* $\beta_1 \not< \alpha_1$. *But it is also clear that* $m \notin min(\beta_1)$ *so* α_1 *and* β_1 *are not* $\mid\!\sim$*-equivalent. Thus to see that the property QLP does not hold for* α_1 *and* β_1 *it is enough to observe that* $\top < \alpha_1$ *because the minimal elements of* α *are in a level above the lowest one.* ∎

3.7 Some Non-Horn rules incomparable with monotony

We consider some non-Horn rules that are stronger than rational monotony, but are not implied by monotony and for this reason are perhaps less interesting than those we have considered so far. We show how they may be characterized by certain subclasses of quasi-linear models.

Definition 3.51 *A preferential relation* $\mid\!\sim$ *is said to be* completely determinated *iff the following rule holds*

$$\text{CEM}\ \frac{\alpha \not\mid\!\sim \beta}{\alpha \mid\!\sim \neg\beta}$$

In other words for any α *and* β, $\alpha\mid\!\sim\beta$ *or* $\alpha\mid\!\sim\neg\beta$.

This rule is called conditional excluded middle in Stalnaker [107] and also called full determinacy in Makinson [69].

Remark 3.52 *1. CEM* $\Rightarrow$ *DP.*

2. P + M $\not\Rightarrow$ *CEM.*

3. P + CEM $\not\Rightarrow$ *M.*

Proof *1. This is immediate. Note that, as a consequence, by proposition 3.10 (see diagram), CEM + P implies each of RM, DR, NR, RC, WD.*

2. This is well known. Take, for instance, $\mid\!\sim$ *to be the classical consequence relation. This relation obviously satisfies P and M but does not satisfy CEM.*

3. Also well known. To recall: take the preferential structure with just two states, one less than the other. Every model on this structure satisfies P and CEM, whilst an appropriate model on it (e.g. the one used in the proof of corollary 3.33) fails to satisfy M. ∎

Definition 3.53 *A preferential model (not necessarily injective)* $\mathcal{M} = \langle S, \imath, \prec \rangle$ *is said to be* linear *iff it is ranked and has at most one state at each level,* i.e. *iff it is of the shape:*

The following theorem can be extracted from work of Stalnaker and Lewis on the logic of counterfactual conditionals, but we give a direct verification here.

Theorem 3.54 *A preferential inference relation* $\mid\!\sim$ *is completely determinated iff there exists a linear model* $\mathcal{M}$ *such that* $\mid\!\sim_{\mathcal{M}} = \mid\!\sim$.

Proof *The* if *part is evident. We prove the* only if *part. Suppose that* $\mid\!\sim$ *is completely determinated. Then by remark 3.52 and proposition 3.10,* $\mid\!\sim$ *satisfies RM. By the Lehmann-Magidor representation theorem 3.6,* $\mid\!\sim$ *can be represented by a ranked preferential model, which by theorem 3.27 we may suppose collapsed. To show that the model is linear, it suffices to show that there are no two distinct states on the same level. Suppose for reductio that s, t are on the same level and* $s \neq t$. *By parsimony there are formulae* α, β *with* $s \in min(\alpha)$ *and* $t \in min(\beta)$. *By injectivity there is an elementary formula p with* $s \in mod(p)$ *and* $t \notin mod(p)$. *Then clearly, using rankedness of the model, we have* $\{s,t\} \subseteq min(\alpha \vee \beta)$, *so* $\alpha\vee\beta \not\mid\!\sim p$ *and* $\alpha\vee\beta \not\mid\!\sim \neg p$, *contrary to complete determination.* ∎

Definition 3.55 *A preferential model is said to be* almost linear *iff it is ranked and has at most one state at any rank above the lowest and at most two states at the lowest level. In other words, iff it is quasi-linear and has at most two states in the lowest level.*

One may wonder whether these models satisfy any interesting new rules. And if so, whether we can characterize those rules by almost linear models. Both answers are positive, as we shall now show. We consider the following two rules

Fragmented Disjunction:

$$\text{FD} \quad \frac{\alpha \mid\!\sim \beta\vee\gamma \quad \alpha \not\mid\!\sim \beta \quad \alpha \not\mid\!\sim \gamma}{\neg\beta \mid\!\sim \gamma}$$

that is, if $\alpha \mid\!\sim \beta\vee\gamma$ then either $\alpha \mid\!\sim \beta$ or $\alpha \mid\!\sim \gamma$ or $\neg\beta \mid\!\sim \gamma$. Its "dual" rule is

Fragmented Conjuction:

$$\text{FC} \quad \frac{\alpha\wedge\beta \mid\!\sim \gamma \quad \alpha \not\mid\!\sim \gamma \quad \beta \not\mid\!\sim \gamma}{\alpha \mid\!\sim \neg\beta}$$

that is, if $\alpha\wedge\beta \mid\!\sim \gamma$ then either $\alpha \mid\!\sim \gamma$ or $\beta \mid\!\sim \gamma$ or $\alpha \mid\!\sim \neg\beta$.

Proposition 3.56 $P + FD \Rightarrow FC$

Proof *Suppose FC fails,* i.e. *that* $\alpha\wedge\beta \mid\!\sim \gamma$, $\alpha \not\mid\!\sim \gamma$, $\beta \not\mid\!\sim \gamma$ *and* $\alpha \not\mid\!\sim \neg\beta$. *From the first we have (by S and RW)* $\alpha \mid\!\sim \neg\beta\vee\gamma$ *and from the third (by LLE) we have* $\neg\neg\beta \not\mid\!\sim \gamma$. *But these together with the second and fourth show the failure of FD.* ∎

Proposition 3.57 *$P+FC \Rightarrow RM$*

Proof *Suppose $\alpha \mid\!\sim \beta$, $\alpha \not\mid\!\sim \neg\gamma$. We want to show that $\alpha\wedge\gamma \mid\!\sim \beta$. We consider two cases: $\alpha \mid\!\sim \gamma$ and $\alpha \not\mid\!\sim \gamma$. In the first case we have $\alpha\wedge\gamma \mid\!\sim \beta$ by CM. Now consider the case $\alpha \not\mid\!\sim \gamma$. By preferentiality (REF and RW) we have*

$$(\alpha\wedge\beta)\wedge(\alpha\wedge\gamma) \mid\!\sim \alpha\wedge\beta\wedge\gamma \tag{3.1}$$

We cannot have $\alpha\wedge\beta \mid\!\sim \alpha\wedge\beta\wedge\gamma$, otherwise by preferentiality (CUT and RW) we have $\alpha \mid\!\sim \gamma$ a contradiction. Thus

$$\alpha\wedge\beta \not\mid\!\sim \alpha\wedge\beta\wedge\gamma \tag{3.2}$$

We cannot have $\alpha\wedge\beta \mid\!\sim \neg(\alpha\wedge\gamma)$, otherwise by preferentiality (RW) we would have $\alpha\wedge\beta \mid\!\sim \alpha \to \neg\gamma$ and again by preferentiality (CUT,REF,AND and RW) we have $\alpha \mid\!\sim \neg\gamma$ a contradiction. Thus

$$\alpha\wedge\beta \not\mid\!\sim \neg(\alpha\wedge\gamma) \tag{3.3}$$

By FC, it follows from 3.1, 3.2 and 3.3 that $\alpha\wedge\gamma \mid\!\sim \alpha\wedge\beta\wedge\gamma$ so by RW $\alpha\wedge\gamma \mid\!\sim \beta$. ∎

Proposition 3.58 *$P+FC \Rightarrow FD$*

Proof *Suppose that FD fails,* i.e. *$\alpha \mid\!\sim \beta\vee\gamma$, $\alpha \not\mid\!\sim \beta$, $\alpha \not\mid\!\sim \gamma$ and $\neg\beta \not\mid\!\sim \gamma$. From the first two hypotheses (by RM which holds by proposition 3.57) we have $\alpha\wedge\neg\beta \mid\!\sim \beta\vee\gamma$. By REF and RW we have $\alpha\wedge\neg\beta \mid\!\sim \neg\beta$ so by AND and RW we have $\alpha\wedge\neg\beta \mid\!\sim \gamma$. We have also, from the second, $\alpha \not\mid\!\sim \neg\neg\beta$. But these last two together with the third and fourth hypotheses show the failure of FC.* ∎

From propositions 3.56 and 3.58 we have immediately:

Corollary 3.59 *Given P, $FD \Leftrightarrow FC$*

Proposition 3.60 *If $\mathcal{M}$ is an almost linear model then $\mid\!\sim_{\mathcal{M}}$ verifies FD and FC.*

Proof *By corollary 3.59 it suffices to consider FD. Let $\mathcal{M} = \langle S, \imath, \prec\rangle$ be an almost linear model and put $\mid\!\sim\, = \,\mid\!\sim_{\mathcal{M}}$. Suppose that $\mid\!\sim$ does not verify FD, i.e. that $\alpha \mid\!\sim \beta\vee\gamma$, $\alpha \not\mid\!\sim \beta$ $\alpha \not\mid\!\sim \gamma$ and $\neg\beta \not\mid\!\sim \gamma$. From the last three, there are states s_1, s_2 and s_3 such that $s_1 \in min(\alpha) \cap mod(\neg\beta)$, $s_2 \in min(\alpha) \cap mod(\neg\gamma)$, $s_3 \in min(\neg\beta) \cap mod(\neg\gamma)$. Then s_1 and s_2 are on the same level and are distinct (for using $\alpha \mid\!\sim \beta\vee\gamma$ we have $s_1 \in mod(\gamma)$), and so, by quasi-linearity, they are on the lowest level. Since $s_1 \in mod(\neg\beta)$, necessarily $min(\neg\beta)$ is included in the lowest level. So s_3 is also on the lowest level. Since $\alpha \mid\!\sim \beta\vee\gamma$ we have also $s_2 \in mod(\beta)$ and $s_3 \in mod(\neg\alpha)$. Hence s_1, s_2 and s_3 are mutually distinct states on the lowest level, contradicting almost linearity.* ∎

Remark 3.61 *Clearly, P+M does not imply FD; we need only note that classical consequence, which satisfies monotony fails not only CEM as already observed in remark 3.52 but also FD. The same also follows from the fact that are flat models (no states less than any other) that fail FD. For instance consider the model $\mathcal{M} = \langle S, \imath, \prec\rangle$ consisting of just three states s_1, s_2 and s_3 all of the same and hence lowest level. Choose α, β and γ three distinct elementary propositions and put $\imath(s_1) \models \alpha\wedge\neg\beta\wedge\gamma$, $\imath(s_2) \models \alpha\wedge\beta\wedge\neg\gamma$ and $\imath(s_3) \models \neg\alpha\wedge\neg\beta\wedge\neg\gamma$. Then clearly FD fails. Actually we can put this observation in more general form:*

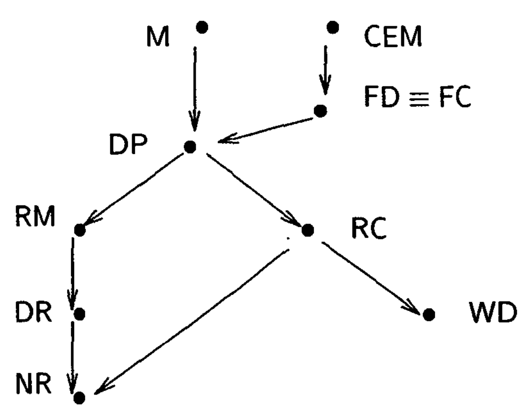
Proposition 3.62 *If $\mathcal{M}$ is an injective quasi-linear model which is not almost linear then FD (and FC) fails.*

Proof *Assume that $\mathcal{M} = \langle S, \imath, \prec\rangle$ is an injective quasi-linear model which is not almost linear. Then there are three different states s_1, s_2 and s_3 on the lowest level. Using injectivity, there are formulae α_1, α_2, α_3 such that $\alpha_i \in mod(s_j)$ iff $i = j$ (for $i, j = 1, 2, 3$). Now trivially $\alpha_1 \vee \alpha_2 \vdash \alpha_1 \vee \alpha_2$ so $\alpha_1 \vee \alpha_2 \mid\!\sim \alpha_1 \vee \alpha_2$. But $\alpha_1 \vee \alpha_2 \not\mid\!\sim \alpha_1$ (witness s_2) and $\alpha_1 \vee \alpha_2 \not\mid\!\sim \alpha_2$ (witness s_1), and $\neg\alpha_1 \not\mid\!\sim \alpha_2$ (witness s_3), so that FD fails.* ∎

We now compare the strength of the rules FD and FC with those implied by monotony that were studied in section 3.3.

Theorem 3.63 *Given P we have CEM ⇒ FC, FD ⇒ DP, but neither converse holds.*

Before proving the theorem, we combine the information that it contains with corollary 3.59, remark 3.61, and proposition 3.10, to get the following diagram:



As before, we verify the positive parts of the theorem syntactically and the negative parts semantically. We begin with the positive parts

Proposition 3.64 *P + CEM ⇒ FC*

Proof *Suppose $\alpha \wedge \beta \mid\!\sim \gamma$ but $\alpha \not\mid\!\sim \gamma$ and $\beta \not\mid\!\sim \gamma$; we want to show $\alpha \mid\!\sim \neg\beta$. Suppose for reductio that $\alpha \not\mid\!\sim \neg\beta$. Then by CEM, $\alpha \mid\!\sim \beta$ and so by the first premise using CUT, $\alpha \mid\!\sim \gamma$ contradicting the second premise.* ∎

Note that we also have an easy semantical proof of this proposition using theorem 3.54 and proposition 3.60.

Proposition 3.65 *P + FD ⇒ DP*

Proof *We have already shown (proposition 3.57 and corollary 3.59) that P + FD ⇒ RM, and we know from theorem 3.10 that P + RM + RC ⇒ DP, so we need only show P + FD ⇒ RC.*

Recall that by preferentiality, whenever $\phi\mid\!\sim\bot$ *then* $\phi\mid\!\sim\psi$ *so by CM* $\phi\wedge\psi\mid\!\sim\bot$ *so by S* $\psi\mid\!\sim\phi\rightarrow\bot$ *so* $\psi\mid\!\sim\neg\phi$.

Assume P + FD. Suppose that RC does not hold, i.e. $\alpha\mid\!\sim\beta$, $\neg\beta\not\mid\!\sim\neg\alpha$ *and* $\neg\beta\not\mid\!\sim\alpha$. *We want to get a contradiction. By supraclassicality (*i.e. *the rule* $\alpha\vdash\beta\Rightarrow\alpha\mid\!\sim\beta$*, derivable from the preferential postulates REF and RW) we have* $\neg\beta\mid\!\sim(\neg\beta\wedge\neg\alpha)\vee(\neg\beta\wedge\alpha)$. *We have also* $\neg\beta\not\mid\!\sim\neg\beta\wedge\neg\alpha$ *and* $\neg\beta\not\mid\!\sim\neg\beta\wedge\alpha$. *So by FD* $\neg(\neg\beta\wedge\neg\alpha)\mid\!\sim\neg\beta\wedge\alpha$. *So by LLE* $\alpha\vee\beta\mid\!\sim\alpha\wedge\neg\beta$. *Thus* $\alpha\vee\beta\mid\!\sim\alpha$ *and* $\alpha\vee\beta\mid\!\sim\neg\beta$. *So, by CM* $(\alpha\vee\beta)\wedge\alpha\mid\!\sim\neg\beta$. *So, by LLE* $\alpha\mid\!\sim\neg\beta$. *But this together with* $\alpha\mid\!\sim\beta$ *implies (by AND)* $\alpha\mid\!\sim\bot$. *By the fact recalled at the beginning of the proof, we get* $\neg\beta\mid\!\sim\neg\alpha$, *a contradiction as desired.* ∎

Proposition 3.66 *P + FC* $\not\Rightarrow$ *CEM*

Proof *Take an almost linear model of only one level, containing two states both of which satisfy* α *but just one of which satisfies* β. *Clearly this fails CEM, but by proposition 3.60 it satisfies FC.* ∎

Proposition 3.67 *P + DP* $\not\Rightarrow$ *FD*

Proof *Immediate form remark 3.61 and the fact that trivially M* $\Rightarrow$ *DP.* ∎

Theorem 3.68 *Let* $\mid\!\sim$ *be a preferential relation. Then* $\mid\!\sim$ *verifies FD (or FC) iff there exists an almost linear model* $\mathcal{M}=\langle S,\imath,\prec\rangle$ *such that* $\mid\!\sim\;=\;\mid\!\sim_{\mathcal{M}}$

Proof *The* if *part is proposition 3.60. We prove the* only if *part. By proposition 3.65 and theorem 3.38, there exists a quasi-linear model* $\mathcal{M}$ *such that* $\mid\!\sim\;=\;\mid\!\sim_{\mathcal{M}}$. *By collapsing, we can assume that* $\mathcal{M}$ *is injective. So by proposition 3.62,* $\mathcal{M}$ *is almost linear.* ∎

3.8 Some Horn rules between Preferential Inference and Monotony

Up to now all the rules studied as potential additions to those of preferential inference, are non-Horn. One may wonder if there are 'interesting' Horn rules beyond those of preferential inference, but still weaker than monotony. One such rule may be called *Conjunctive Insistence*:

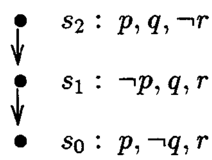
$$\text{CI}\quad\frac{\alpha\mid\!\sim\beta\quad\gamma\mid\!\sim\beta}{\alpha\wedge\gamma\mid\!\sim\beta}$$

Proposition 3.69 *Monotony implies CI but the converse does not hold even if we suppose CEM. The preferential rules plus CEM do not imply CI. Moreover CI does not imply NR or WD.*

Proof *Clearly CI is implied by monotony. To see that monotony is not implied by CI even with CEM recall again the model in the proof of corollary 3.33 and of remark 3.52 (3). We know that this model satisfies P and CEM but fails M. We complete the proof by showing that every model with this structure satisfies CI.*

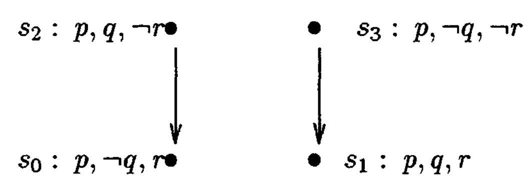
Suppose that $\alpha\wedge\beta\not\mid\!\sim\gamma$ *in a model with this structure; we need to show that either* $\alpha\not\mid\!\sim\gamma$ *or* $\beta\not\mid\!\sim\gamma$. *By the hypothesis there is a state* $s \in min(\alpha\wedge\beta)$ *satisfiying* $\neg\gamma$.
Case 1: Suppose s is of level zero. Then $s \in min(\alpha) \cap min(\beta)$ *and so in fact we have both* $\alpha\not\mid\!\sim\gamma$ *and* $\beta\not\mid\!\sim\gamma$.
Case 2: Suppose s is of level one. Since $s \in min(\alpha\wedge\beta)$ *the unique state* $t \prec s$ *of level zero fails to satisfy at least one of* α, β. *But in this case* $s \in min(\alpha)$ *or* $s \in min(\beta)$ *giving us in this case* $\alpha\not\mid\!\sim\gamma$ *or* $\beta\not\mid\!\sim\gamma$.

CI is not implied by preferential rules plus CEM - consider for instance a ranked preferential structure with just three states at three levels, with an appropriate distribution of truth values as in the following figure:



Clearly here $p\mid\!\sim r$, $q\mid\!\sim r$ *but* $p\wedge q\not\mid\!\sim r$. *Note that this model is also linear, so that indeed P plus CEM does not imply CI.*

We prove now that CI does not imply NR (which is enough to show that CI does not imply any of DR, RM,RC, DP, FD, CEM). Consider the model defined by the following figure:



Clearly this fails NR, for $p\mid\!\sim r$ *whilst* $p\wedge q\not\mid\!\sim r$ *(witness* s_2*) and* $p\wedge\neg q\not\mid\!\sim r$ *(witness* s_3*). However, it satisfies CI. Suppose for reductio that* $\alpha\mid\!\sim\gamma$, $\beta\mid\!\sim\gamma$ *but* $\alpha\wedge\beta\not\mid\!\sim\gamma$. *From the last assumption, there is a* $s \in min(\alpha\wedge\beta)$ *such that* $s \models \neg\gamma$. *Since* $\alpha\mid\!\sim\gamma$, $\beta\mid\!\sim\gamma$, $s \models \alpha$, $s \models \beta$, $s \not\models \gamma$ *there are* $u, u' \prec s$ *with* $u \in min(\alpha)$, $u' \in min(\beta)$. *But there is at most one state less than s, so* $u = u'$ *so* $u \models \alpha\wedge\beta$ *contradicting the minimality of s in* $mod(\alpha\wedge\beta)$.

The same model shows that $P + CI \not\Rightarrow WD$; *we have* $\top\mid\!\sim\neg(p\wedge\neg r)$, $p\wedge\neg r\not\mid\!\sim q$ *(witness* s_3*),* $p\wedge\neg r\not\mid\!\sim\neg q$ *(witness* s_2*).* ∎

We suspect that there are not 'very many' Horn rules which, like CI, are implied by preferential rules with monotony but are not implied by the preferential rules alone. There are some, however, of technical more than conceptual interest. Consider the infinite series of rules of n-monotony ($n \geq 1$), n-M in short, constructed as follows:

$$\text{1-M} \quad \frac{\alpha_1\mid\!\sim\phi}{\alpha_1\wedge\alpha_2\mid\!\sim\phi}$$

$$\text{2-M} \quad \frac{\alpha_1\mid\!\sim\phi \quad \alpha_1\wedge\alpha_2\mid\!\sim\neg\phi}{\alpha_1\wedge\alpha_2\wedge\alpha_3\mid\!\sim\neg\phi}$$

and in general

$$n\text{-M}\quad \frac{\alpha_1 \mid\!\sim \sigma_1(\phi) \quad \ldots \quad \alpha_1\wedge\ldots\wedge\alpha_n \mid\!\sim \sigma_n(\phi)}{\alpha_1\wedge\ldots\wedge\alpha_n\wedge\alpha_{n+1} \mid\!\sim \sigma_n(\phi)}$$

where each $\sigma_i(\phi)$ is either ϕ or $\neg\phi$ according as i is odd or even, and noting that the conclusion-rule uses σ_n rather than σ_{n+1}. This rule is evidently reminiscent of the alternating sequence of statements in the party example in section 1.2 of Lewis [57]: *'If Otto had come, it would have been a lively party; but if both Otto and Anna had come, it would have been a dreary party; but if Waldo had come as well, it would have been lively; but ...'* Of course, that is an infinite list of conditional expressions, and not a list of Horn rules about them. The rule number n is a scheme, one of whose instances in effect takes the first n Lewis statements as its n premises, and puts as conclusion a statement that is like Lewis' statement $n+1$ but with opposite consequent.

Clearly, 1-M is plain monotony. Moreover we have the following:

Observation 3.70 *1. For all n, P+ n-M implies $(n+1)$-M.*

2. For all n, P+ $(n+1)$-M does not imply n-M, even if CEM is also assumed.

3. P+ CI implies 2-M.

4. P+ 2-M does not imply CI, even if FD is also assumed.

Proof *Here we give only an outline.*

1. Simply treat $\alpha_1\wedge\alpha_2$ as a single formula in the premises of the rule of $(n+1)$-M, relabel letters, and apply the rule of n-M to the last n premises.

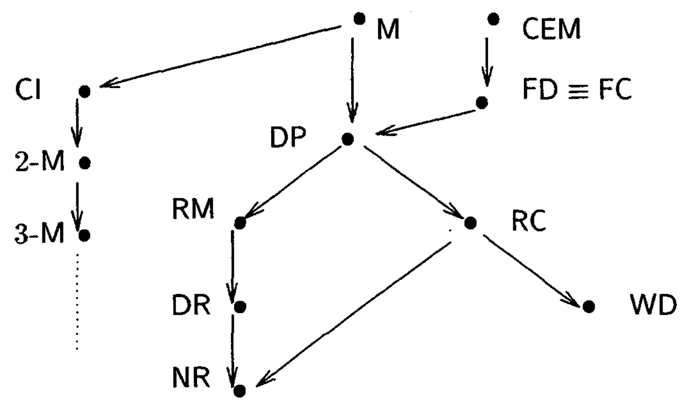
2. It is easy to see that (a) any ranked model with at most n levels satisfies n-M, (b) there is a linear preferential model with $n+1$ levels that fails n-M. These two facts give the desired result, using theorem 3.54.

3. Let $\mid\!\sim$ be a preferential relation that fails 2-M; we want to show that it fails CI. Since it fails 2-M, there are formulae $\alpha_1, \alpha_2, \alpha_3, \phi$ with $\alpha_1 \mid\!\sim \phi$, $\alpha_1\wedge\alpha_2 \mid\!\sim \neg\phi$, $\alpha_1\wedge\alpha_2\wedge\alpha_3 \not\mid\!\sim \neg\phi$. We need to find formulae β, γ, δ with $\beta \mid\!\sim \delta$, $\gamma \mid\!\sim \delta$, $\beta\wedge\gamma \not\mid\!\sim \delta$. Put $\beta = \alpha_1\wedge(\neg\alpha_2\vee\alpha_3)$, $\gamma = \alpha_1\wedge\alpha_2$, $\delta = \neg\alpha_2\vee\neg\phi$. To show that $\beta \mid\!\sim \delta$, i.e. *that $\alpha_1\wedge(\neg\alpha_2\vee\alpha_3) \mid\!\sim \neg\alpha_2\vee\neg\phi$ note that since $\alpha_1\wedge\alpha_2 \mid\!\sim \neg\phi$ we have $\alpha_1 \mid\!\sim \neg\alpha_2\vee\neg\phi$, so since $\alpha_1 \mid\!\sim \phi$ we have $\alpha_1 \mid\!\sim \neg\alpha_2$ by preferential rules, so that by further preferential rules, $\alpha_1 \mid\!\sim \neg\alpha_2\vee\alpha_3$ and also $\alpha_1 \mid\!\sim \neg\alpha_2\vee\neg\phi$, so finally by the preferential rule CM, $\alpha_1\wedge(\neg\alpha_2\vee\alpha_3) \mid\!\sim \neg\alpha_2\vee\neg\phi$ as desired. To show that $\gamma \mid\!\sim \delta$,* i.e. *that $\alpha_1\wedge\alpha_2 \mid\!\sim \neg\alpha_2\vee\neg\phi$ simply apply the preferential rule RW to the assumption $\alpha_1\wedge\alpha_2 \mid\!\sim \neg\phi$. Finally, to show that $\beta\wedge\gamma \not\mid\!\sim \delta$,* i.e. *that $\alpha_1\wedge(\neg\alpha_2\vee\alpha_3)\wedge(\alpha_1\wedge\alpha_2) \not\mid\!\sim \neg\alpha_2\vee\neg\phi$, suppose the contrary and apply LLE to get $\alpha_1\wedge\alpha_2\wedge\alpha_3 \mid\!\sim \neg\alpha_2\vee\neg\phi$ so that $\alpha_1\wedge\alpha_2\wedge\alpha_3 \mid\!\sim \neg\phi$ contrary to hypothesis.*

4. To prove this part, it suffices by proposition 3.60 and 2(a) above, to find an almost linear model with two levels that fails CI. Clearly the following model will do: the language is built over the elementary formulae p, q, r; the lowest level has two states s_1, s_2 with $\imath(s_1) = \{p, r\}$, $\imath(s_2) = \{q, r\}$, whilst the next level has one state only s_3 with $\imath(s_3) = \{p, q\}$, so $p \mid\!\sim r$, $q \mid\!\sim r$, $p\wedge q \not\mid\!\sim r$. ∎

We note that since by proposition 3.69, P + CI $\not\Rightarrow$ NR|WD, points 1 and 3 above tells us that P + n-M $\not\Rightarrow$ NR|WD, whenever $n > 1$.

We may thus extend our diagram as follows:



Contrasting with point 4 of observation 3.70 we have:

Observation 3.71 $P+2\text{-}M+CEM \Rightarrow CI$

Proof *It is possible to verify this semantically, but the following is a direct syntactic proof. Suppose P+2-M+CEM. Let* α, β, ϕ *be formulae. Suppose* $\alpha \mid\!\sim \phi$, $\beta \mid\!\sim \phi$; *we want to show* $\alpha \wedge \beta \mid\!\sim \phi$. *We divide the argument into two cases.*

Case 1. *Suppose* $\alpha \vee \beta \mid\!\sim \alpha \leftrightarrow \phi$ *and* $\alpha \vee \beta \mid\!\sim \beta \leftrightarrow \phi$. *Since* $\alpha \mid\!\sim \phi$, $\beta \mid\!\sim \phi$ *we have by OR* $\alpha \vee \beta \mid\!\sim \phi$, *so by preferentiality,* $\alpha \vee \beta \mid\!\sim \alpha \wedge \beta$. *Now using CM we have* $(\alpha \vee \beta) \wedge (\alpha \wedge \beta) \mid\!\sim \phi$, i.e. $\alpha \wedge \beta \mid\!\sim \phi$ *as desired.*

Case 2. *Suppose* $\alpha \vee \beta \not\mid\!\sim \alpha \leftrightarrow \phi$ *or* $\alpha \vee \beta \not\mid\!\sim \beta \leftrightarrow \phi$. *We consider the former; the latter is similar. By CEM,* $\alpha \vee \beta \mid\!\sim \neg(\alpha \leftrightarrow \phi)$. *But also since* $\alpha \mid\!\sim \phi$ *we clearly have* $\alpha \mid\!\sim \alpha \leftrightarrow \phi$, i.e. $(\alpha \vee \beta) \wedge \alpha \mid\!\sim \alpha \leftrightarrow \phi$. *Hence by 2-M we have* $((\alpha \vee \beta) \wedge \alpha) \wedge \beta \mid\!\sim \alpha \leftrightarrow \phi$, i.e. $\alpha \wedge \beta \mid\!\sim \alpha \leftrightarrow \phi$, *so* $\alpha \wedge \beta \mid\!\sim \phi$ *as desired.* ∎

3.8.1 Semantics for n-Monotony

Let $\mathcal{M} = \langle S, \imath, \prec \rangle$ be a preferential model. Define the *height* of $\mathcal{M}$ to be the maximal length of any chain of states $s_1, \ldots, s_n$ with $s_1 \prec s_2 \prec \cdots \prec s_n$. If there is no maximal length, put the height of the model to be ∞. The following observations generalize points 2(a) and 2(b) in the proof of observation 3.70.

Observation 3.72 *Every preferential model of height* $\leq n$ *satisfies n-M.*

Proof *Consider a preferential model that fails n-M. We show that it has height* $\geq n+1$. *Since n-M fails in the model, there are formulae* $\alpha_1, \ldots \alpha_n, \alpha_{n+1}, \phi$ *with*

$$\begin{array}{ll}
\alpha_1 \mid\!\sim \phi & (1) \\
\alpha_1 \wedge \alpha_2 \mid\!\sim \neg\phi & (2) \\
\vdots & \\
\alpha_1 \wedge \ldots \wedge \alpha_n \mid\!\sim \sigma_n(\phi) & (n) \\
\alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1} \not\mid\!\sim \sigma_n(\phi) & (n+1)
\end{array}$$

where each $\sigma_i(\phi)$ is ϕ or $\neg\phi$ according as i is odd or even. From $(n+1)$ there is a state s_{n+1} with $s_{n+1} \in min(\alpha_1 \wedge \ldots \wedge \alpha_{n+1})$ but $\imath(s_{n+1}) \not\models \sigma_n(\phi)$. But by (n), $s_{n+1} \notin min(\alpha_1 \wedge \ldots \wedge \alpha_n)$, so there is $s_n \prec s_{n+1}$ with $s_n \in min(\alpha_1 \wedge \ldots \wedge \alpha_n)$ and $\imath(s_n) \models \sigma_n(\phi)$. Continuing down like this we get a sequence $s_{n+1} \succ s_n \succ \cdots \succ s_1$ of length $n+1$ of states of the model, so that its height is $\geq n+1$. ∎

Observation 3.73 *Every parsimonious ranked model of height $> n$ fails n-M.*

Proof *Consider any parsimonious ranked model. Let $\mid\!\sim$ be the inference relation generated by the model. Since the model is of height $> n$ there is a sequence of states $s_1, \ldots, s_n, s_{n+1}$ with $s_1 \prec s_2 \prec \cdots \prec s_{n+1}$. By parsimony there exists a sequence of formulae $\gamma_1, \ldots, \gamma_{n+1}$ with $s_i \in min(\gamma_i)$ for $i = 1, \ldots, n+1$. We define the formulae α_i for $i = 1, \ldots, n+1$ and ϕ as follows:*

$$\alpha_i = \bigvee_{k=i}^{n+1} \gamma_k \qquad \phi = \bigvee_{k\geq 0}^{2k\leq n} (\gamma_{2k+1} \wedge \bigwedge_{j=2k+2}^{n+1} \neg\gamma_j)$$

Then we have:

1. *For every $i = 1, \ldots, n$ and for every state u at the same level as s_i, $u \models \neg\gamma_k$ for any $k = i+1, \ldots, n+1$.*

2. *$min(\alpha_i) = min(\gamma_i)$ for any $i = 1, \ldots, n+1$.*

3. *$\vdash \alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_i \leftrightarrow \alpha_i$ for every $i = 1, \ldots, n+1$.*

The points 1 and 2 are easy consequences of rankedness and point 3 is evident by definition of α_i. From points 1 and 2 is easy to see that $\alpha_i \mid\!\sim \sigma_i(\phi)$ for any $i = 1, \ldots, n+1$. Thus we have

$$\begin{array}{l} \alpha_1 \mid\!\sim \phi \\ \alpha_1 \wedge \alpha_2 \mid\!\sim \neg\phi \\ \vdots \\ \alpha_1 \wedge \ldots \wedge \alpha_n \mid\!\sim \sigma_n(\phi) \\ \alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1} \mid\!\sim \sigma_{n+1}(\phi) \end{array}$$

From this it is evident that to show that n-M fails it is enough to see that $\alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1} \not\mid\!\sim \sigma_n(\phi)$. From points 2 and 3, $\alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1}$ is $\mid\!\sim$-consistent and since $\alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1} \mid\!\sim \sigma_{n+1}(\phi)$, necessarily $\alpha_1 \wedge \ldots \wedge \alpha_n \wedge \alpha_{n+1} \not\mid\!\sim \sigma_n(\phi)$, as desired. ∎

Theorem 3.74 *Let $\mid\!\sim$ be any preferential inference relation. Then the following conditions are equivalent:*

1. *$\mid\!\sim$ satisfies n-M and RM (resp. RT, FD).*

2. *$\mid\!\sim$ is generated by some ranked (resp. quasi-linear, almost linear) preferential model of height $\leq n$.*

Proof *For the implication $2 \Rightarrow 1$, apply observation 3.72 together with theorem 3.6 (resp. 3.32, 3.60). For the implication $1 \Rightarrow 2$, apply observation 3.73 together with corollary 3.28 (plus 3.36, 3.62 respectively).* ∎

Conjecture and open questions

We conclude with a conjecture and some open questions.

Conjecture: There are no Horn-rules *with a single premise* which, given the preferential rules, are implied by monotony, but do not imply monotony, and are not implied by the preferential rules alone (recall that CI has two premises, and n-monotony has n premises).

Open questions:

1. Determine whether the non implication $\mathsf{P} + \mathsf{WD} \not\Rightarrow \mathsf{NR}$ can be witnessed by *injective* preferential models (*cf.* observation 3.21)
2. Determine whether the construction used to prove lemma 3.23 can be adapted for a class of preferential models broader than the ranked ones -*e.g.* to the class of all models that are filtered in the sense of Freund [30] (*cf.* the discussion in remark 3.30).
3. Find appropriate classes of preferential models to provide representation theorems for RC, WD, CI (*cf.* the discussion in remark 3.40).

Acknowledgements

We would like to thank Michael Freund for very penetrating comments on several versions of this work and for the unpublished observations cited as personal communications in the text.

Chapitre 4

On representation theorems for nonmonotonic inference relations

One of the main tools in the study of nonmonotonic consequence relations is the representation of such relations in terms of preferential models. In this paper we give an unified and simpler framework to obtain such representation theorems.

4.1 Introduction

A consequence relation $\mid\sim$ is a binary relation between formulas on a classical propositional language. We are interested in nonmonotonic consequence relations, *i.e.* those relations that do not satisfy the monotonicity rule: If $\alpha \mid\sim \gamma$ then $\alpha \wedge \beta \mid\sim \gamma$. Several systems of postulates (cumulative, preferential, rational and others) for classifying nonmonotonic consequence relations has been investigated [49, 55, 37, 33, 30, 7]. One of the main features of these systems is the amount of monotony that is required from the consequence relation. The study of non monotonic reasoning has been motivated by problems arising in artificial intelligence (knowledge representation, belief revision, etc). There is a vast literature concerning nonmonotonity, for the particular approach dealt with in this paper we refer the reader to [69, 49] and the references therein.

An important tool for the study and classification of nonmonotonic consequence relations is the representation of such relations in terms of preferential models. A preferential model $\mathcal{M}$ is a triple $\langle S, \imath, \prec \rangle$, where S is a set of states, $\imath$ is function assigning to each state a valuation and $\prec$ is a strict partial order over S. $\mathcal{M}$ is said to be a model of $\mid\sim$ when $\alpha \mid\sim \beta$ iff $\imath(s) \models \beta$ for all s which are $\prec$-minimal among all states t such that $\imath(t) \models \alpha$ (the details are given in §3.2). A consequence relation $\mid\sim$ is preferential relation if and only if it is of the form $\mid\sim_{\mathcal{M}}$ for some preferential model $\mathcal{M}$ ([49]). If $\mid\sim$ is rational then the model can be found ranked ([55]). Disjunctive relations were studied in [30] and shown to be those relations represented by filtered models. When the relation also satisfies rational transitivity then $\mathcal{M}$ can be found quasi-linear ([9, 7]). These results are referred to as representation theorems and they can be regarded as a sort of a soundness and completeness theorems. These representations, besides providing a semantic interpretation of $\mid\sim$, are also quite useful to establish most properties of $\mid\sim$ by model theoretic arguments instead of proof theoretic ones.

In this paper we give simpler proofs of representation theorems for injective relations. The key idea is the notion of the essential relation $<_e$ (defined in §4.3) associated with a preferential consequence relation $\mid\!\sim$. We will show that if $\mid\!\sim$ is preferential and disjunctive, then $<_e$ is a transitive strict order defined on a set of valuations such that the models of $\{\beta : \alpha\mid\!\sim\beta\}$ are the $<_e$-minimal valuations that satisfy α. In other words, $<_e$ provides a representation of $\mid\!\sim$. We will show also that if $\mid\!\sim$ is disjunctive (resp. rational, rational transitive), then $<_e$ is filtered (resp. ranked, quasi-linear). Most of these results were known but they were proved by quite different means (see [49, 55, 30, 37, 8]). We think our proofs are easier and in a sense "canonical". One interesting feature of our approach is that $<_e$ provides a direct way of "ordering" the valuations without using an auxiliary order over formulas, as is the case of other proofs of representation theorems. Freund introduced a property (that we denote by W-DR) weaker than disjunctiveness. We show that if $\mid\!\sim$ is preferential and satisfies W-DR, then $<_e$ represents $\mid\!\sim$. In §4.5 we address the question of uniqueness of these representations. In particular, we will compare our results with Freund's and show that $<_e$ coincides with Freund's relation when DR holds. We will see in §4.6 that in spite of the fact that in some cases $<_e$ is not transitive, it still provides a good representation of some preferential relations for which other methods do not work. We also present an example showing that W-DR is not a necessary condition for having an injective model.

4.2 Preliminaries

We recall some basic definitions and results from Kraus, Lehmann and Magidor [49], Lehmann and Magidor [55] and Freund [30] which will be used in the paper.

We consider formulas of classical propositional calculus built over a set of variables denoted *Var* plus two constants $\top$ and $\bot$ (the formulas **true** and **false** respectively). Let $\mathcal{L}$ be the set of formulas. If *Var* is finite we will say that the language $\mathcal{L}$ is finite. Let $\mathcal{U}$ be the set of valuations (or worlds), *i.e.* functions $M : Var \cup \{\top, \bot\} \longrightarrow \{0, 1\}$ such that $M(\top) = 1$ and $M(\bot) = 0$. We use lower case letters of the Greek alphabet to denote formulas, and the letters $M, N, P, M_1, M_2, \ldots$ to denote worlds. As usual, $\vdash \alpha$ means that α is a tautology and $M \models \alpha$ means that M satisfies α where compound formulas are evaluated using the usual truth-functional rules. We consider certain binary relations between formulas. These relations will be called consequence relations and will be written $\mid\!\sim$.

Definition 4.1 *A relation $\mid\!\sim$ is said to be cumulative iff the following rules hold*

REF	$\forall\alpha[\ \alpha\mid\!\sim\alpha\ ]$
LLE	$\forall\alpha,\beta,\gamma\ [\ \alpha\mid\!\sim\beta\ \&\ \vdash\alpha\leftrightarrow\gamma\Rightarrow\gamma\mid\!\sim\beta\ ]$
RW	$\forall\alpha,\beta,\gamma\ [\ \alpha\mid\!\sim\beta\ \&\ \vdash\beta\rightarrow\gamma\Rightarrow\alpha\mid\!\sim\gamma\ ]$
CUT	$\forall\alpha,\beta,\gamma\ [\ \alpha\wedge\beta\mid\!\sim\gamma\ \&\ \alpha\mid\!\sim\beta\Rightarrow\alpha\mid\!\sim\gamma\ ]$
CM	$\forall\alpha,\beta,\gamma\ [\ \alpha\mid\!\sim\beta\ \&\ \alpha\mid\!\sim\gamma\Rightarrow\alpha\wedge\gamma\mid\!\sim\beta\ ]$

These rules are known as the rules of the system $\mathcal{C}$. The abbreviations above are read as follows: REF -reflexivity, LLE -left logical equivalence, RW -right weakening, CM -cautious monotony. CUT is self-explanatory, but it should be noted that this form of cut, which plays an important role in nonmonotonic logic, is weaker than the form of cut usually studied in Gentzen-style formulations of classical and intuitionistic logic. The latter implies transitivity of the consequence relation; the former does not.

It is well known [49] that the following rules (And, Reciprocity) are derivable from system C:

AND $\forall\alpha,\beta,\gamma\,[\,\alpha{\mid\!\sim}\beta\ \&\ \alpha{\mid\!\sim}\gamma \Rightarrow \alpha{\mid\!\sim}\beta\wedge\gamma\,]$

RECIP $\forall\alpha,\beta,\gamma\,[\,\alpha{\mid\!\sim}\beta\ \&\ \beta{\mid\!\sim}\alpha\ \&\ \alpha{\mid\!\sim}\gamma \Rightarrow \beta{\mid\!\sim}\gamma\,]$

Definition 4.2 *A relation* ${\mid\!\sim}$ *is said to be preferential iff it is cumulative and satisfies the following rule (or):*

$$OR\quad \forall\alpha,\beta,\gamma\,[\,\alpha{\mid\!\sim}\gamma\ \&\ \beta{\mid\!\sim}\gamma \Rightarrow \alpha\vee\beta{\mid\!\sim}\gamma\,]$$

A relation ${\mid\!\sim}$ *is said to be disjunctive rational iff it is preferential and the following rule (disjunctive rationality) holds*

$$DR\qquad \forall\alpha,\beta,\gamma\,[\,\alpha\vee\beta{\mid\!\sim}\gamma\ \&\ \alpha{\not\mid\!\sim}\gamma \Rightarrow \beta{\mid\!\sim}\gamma\,]$$

A relation ${\mid\!\sim}$ *is said to be rational iff it is preferential and the following rule (rational monotony) holds*

$$RM\qquad \forall\alpha,\beta,\gamma\,[\,\alpha{\mid\!\sim}\beta\ \&\ \alpha{\not\mid\!\sim}\neg\gamma \Rightarrow \alpha\wedge\gamma{\mid\!\sim}\beta\,]$$

It is well known [49, 69] that given the preferential rules (system C plus OR), RM implies DR and also that any preferential relation satisfies the following rule

$$S\qquad \forall\alpha,\beta,\gamma\,[\,\alpha\wedge\beta{\mid\!\sim}\gamma \Rightarrow \alpha{\mid\!\sim}\beta\rightarrow\gamma\,]$$

Let ${\mid\!\sim}$ be a consequence relation. As usual, $C_{\mid\sim}(\alpha) = \{\beta : \alpha{\mid\!\sim}\beta\}$. If there is no ambiguity we shall write $C(\alpha)$ instead of $C_{\mid\sim}(\alpha)$. If U (α) is a set of formulas (a formula) then $Cn(U)$ $(Cn(\alpha))$ will denote the set of classical consequences of U (α).

We recall the definition of preferential models.

Definition 4.3 *A structure* $\mathcal{M}$ *is a triple* $\langle S, \imath, \prec\rangle$ *where* S *is a set (called the set of states),* $\prec$ *is a strict order (*i.e. *transitive and irreflexive) on* S *and* $\imath : S \longrightarrow \mathcal{U}$ *is a function (called the interpretation function).*

Let $\mathcal{M} = \langle S, \imath, \prec\rangle$ be a structure. We adopt the following notations: if $T \subseteq S$, then $\min(T) = \{t \in T : \neg\exists t' \in T,\ t' \prec t\}$, *i.e.* $\min(T)$ is the set of all minimal elements of T with respect to $\prec$; $\mathrm{mod}_{\mathcal{M}}(\alpha) = \{s \in S : \imath(s) \models \alpha\}$; $\min_{\mathcal{M}}(\alpha) = \min(\mathrm{mod}_{\mathcal{M}}(\alpha))$.

Definition 4.4 *Let* $\mathcal{M} = \langle S, \imath, \prec\rangle$ *be a structure and* $T \subseteq S$. *We say that* T *is* ***smooth*** *if it satisfies the following*

$$\forall s \in T \setminus min(T)\quad \exists s' \in min(T)\quad s' \prec s$$

$\mathcal{M}$ *is said to be a* ***preferential model*** *if* $mod_{\mathcal{M}}(\alpha)$ *is smooth for any formula* α.

Each preferential model has an associated consequence relation given by the following:

Definition 4.5 *Let* $\mathcal{M} = \langle S, \imath, \prec\rangle$ *be a preferential model. The consequence relation* ${\mid\!\sim}_{\mathcal{M}}$ *is defined by the following*

$$\alpha{\mid\!\sim}_{\mathcal{M}}\beta \Leftrightarrow min_{\mathcal{M}}(\alpha) \subseteq mod_{\mathcal{M}}(\beta) \tag{4.1}$$

The following representation theorems are one of the basic tools in the study of nonmonotonic consequence relations. The *if* part of them are not difficult to establish. The main subject of this paper consists in providing, for a large class of preferential relations, a 'canonical' way of proving the *only if* part.

Theorem 4.6 *(Krauss, Lehmann and Magidor [49]) A consequence relation* $\mid\sim$ *is preferential iff there is a preferential model* $\mathcal{M}$ *such that* $\mid\sim\, = \,\mid\sim_{\mathcal{M}}$.

A structure $\mathcal{M} = \langle S, \imath, \prec\rangle$ is said to be a *ranked model* if it is a preferential model and there exists a strict linear order $(\Omega, <)$ and a function $r : S \longrightarrow \Omega$ such that for any $s, s' \in S$, $s \prec s'$ iff $r(s) < r(s')$.

Theorem 4.7 *(Lehmann and Magidor [55]) A consequence relation* $\mid\sim$ *is rational iff there is a ranked model* $\mathcal{M}$ *such that* $\mid\sim\, = \,\mid\sim_{\mathcal{M}}$.

In general, it is not easy to grasp the intuition behind the set of states S and the interpretation function $\imath$. A special case, which is intuitively easy to handle, is when the function $\imath$ is injective (in this case, $\mathcal{M}$ is said to be an injective model). If a preferential model is injective one does not need to mention the interpretation function $\imath$, instead one can assume that S is a set of valuations and $\prec$ is a strict smooth order over S, so $\imath$ would be the identity function. In this case the notion of a smooth relation says that for every $M \in S$ and for every formula α if $M \models \alpha$ and M is not in $\min(\mathrm{mod}(\alpha) \cap S, \prec)$, then there is $N \in S$ such that $N \prec M$ and $N \in \min(\mathrm{mod}(\alpha) \cap S, \prec)$ (where the notion of a $\prec$-minimal element is defined as in the paragraph following 4.3). The relation $\prec$ is understood as a preference relation over valuations. Thus (4.1) says that to compute the consequences of a formula α we need to look only at the preferred valuations of α according to $\prec$, i.e. those valuations belonging to $\min(\mathrm{mod}(\alpha) \cap S)$.

Freund [30] studied a family of consequence relations admitting injective models. [1] He observed that one can always assume that S is certain collection of valuations which we define next

Definition 4.8 *Let* $\mid\sim$ *be a consequence relation. A valuation* N *is called* normal *w.r.t.* $\mid\sim$ *if there is a formula* α *such that* $N \models C(\alpha)$.

If there is not ambiguity we shall say that an interpretation is normal instead of normal with respect to $\mid\sim$ (in [30] normal valuations were called $\mid\sim$-consistent). Freund showed (see remark 3.1 in [30]) that if $\mid\sim$ is represented by an injective model then it can also be represented by an injective model where the set S is the collection of all normal valuation w.r.t. $\mid\sim$. We will state his result next

Theorem 4.9 *(Freund [30]) Let* $\mid\sim$ *be a consequence relation and* S *the collection of normal valuation w.r.t* $\mid\sim$. *Then* $\mid\sim$ *is represented by an injective model iff there is a smooth strict order* $\prec$ *over* S *such that*

$$\alpha \mid\sim \beta \Leftrightarrow \mathit{min}(\mathit{mod}(\alpha) \cap S, \prec) \subseteq \mathit{mod}(\beta) \tag{4.2}$$

[1] According to the referee the first study of consequence relations having injective models is due to Satoh [102].

From this point on we will assume without explicitly mention it that an injective model has the corresponding partial order defined on S.

Let us observe that (4.2) can be restated in the following way: $\min(\text{mod}(\alpha) \cap S, \prec) \subseteq \text{mod}(C(\alpha))$. Some consequence relations admit an injective representation where the equality holds. They were called standard in [30], the formal definition is the following

Definition 4.10 *Let $\mid\!\sim$ be a consequence relation and S the collection of normal valuations w.r.t $\mid\!\sim$. We say that $\mid\!\sim$ is represented by a* ***standard model*** *if there is a smooth strict order $\prec$ over S such that*

$$mod(C(\alpha)) = min(mod(\alpha) \cap S, \prec)$$

Such order $\prec$ will be called a ***standard order*** *that represents $\mid\!\sim$.*

4.3 The essential relation and the main representation theorem

It is not difficult to show that if the language is finite the notions of an injective and a standard model coincide (see [30] pag. 236) but this is not the case if the language is infinite (an example will be given in §4.5). Freund characterized some preferential relations that admit a standard representation. In the case of a finite language his characterization is quite easy to state. The following property is called *Weak Disjunctive Rationality*

$$\textsf{W-DR} \quad C(\alpha \vee \beta) \subseteq Cn(C(\alpha) \cup C(\beta))$$

Freund showed that for a finite language, a preferential relation admits an injective (thus standard) model iff it satisfies W-DR. In order to deal with an infinite languages, Freund introduced a property stronger than W-DR which is based in the notion of a trace of a formula.

In this section we will prove the main result of this paper which is a general representation theorem for consequence relations that satisfy W-DR. For that end we will introduce the *essential relation* which plays a key role in the proof. We will show that this relation can be considered the canonical relation that represents a given preferential consequence relation that satisfies W-DR. The essential relation seems easier to handle than the relation defined by Freund. We will see in §4.5 that they are equal under some conditions. However, we will also give an example of a preferential relation $\mid\!\sim$ represented by our relation but not by Freund's. The idea behind the definition of the essential relation seems to be quite general and turns out to be also useful in a different context (see [95]).

Notation: Given a consequence relation $\mid\!\sim$ we will always denote by $S(\mid\!\sim)$ the collection of normal valuation w.r.t $\mid\!\sim$, when there is no ambiguity about which consequence relation is used we will just write S. If M is a valuation, $Th(M)$ will denote the theory of M, *i.e.* $Th(M) = \{\alpha : M \models \alpha\}$. For a fixed consequence relation $\mid\!\sim$ and a valuation M, $T^{\mid\sim}(M)$ will denote the set $\{\alpha : M \models C(\alpha)\}$, *i.e.* a sort of "nonmonotonic theory" of M. If there is no ambiguity we write $T(M)$ instead of $T^{\mid\sim}(M)$.

Definition 4.11 *Let $\mid\!\sim$ be a consequence relation. The* essential relation *is defined by the following: Let N and M normal valuations,*

$$M <_e N \leftrightarrow \forall \alpha\, (N \models C(\alpha) \Rightarrow M \not\models \alpha\,)$$

In other words, $M <_e N$ iff $Th(M) \cap T(N) = \emptyset$.

The essential relation is not in general transitive (we will see an example in §4.6). It is not difficult to show that transitivity of $\prec$ is not necessary in order to get the easy half of 4.6 (but smoothness can not be avoided). This was already observed in [49] (pag. 193) and we state it for later reference.

Lemma 4.12 *([49]) Let $<$ be a binary irreflexive (but not necessarily transitive) smooth relation over a set T of valuations. Define a consequence relation by $\alpha \mid\!\sim \beta$ iff $min(mod(\alpha) \cap T, <) \subseteq mod(\beta)$. Then $\mid\!\sim$ is preferential.* ∎

Since the usual definition of a standard model requires transitivity of the relation, it is quite natural to ask when is $<_e$ transitive. In §4.5 we will show that if $\mid\!\sim$ is disjunctive, then $<_e$ is transitive and in §4.6 we give an example of a preferential relation satisfying W-DR for which $<_e$ is not transitive. However, for a finite language $<_e$ is transitive for $\mid\!\sim$ preferential. We thank the referee for pointing out that in this case our original assumption (W-DR) was superfluous and suggesting lemma 4.13 below.

As we said in the introduction, previous proofs of representation theorems usually have used an order among formulas as a tool to define the preferential model. For a finite language the essential relation is very much related to one of such orders. Let us recall the definition of $<$ given by Freund

$$\alpha < \beta \text{ iff } \alpha \vee \beta \mid\!\sim \neg\beta$$

In the particular case where $\mid\!\sim$ is rational, this relation coincides with the one that was defined in [49] and it is quite related also to the expectation ordering of [37].

Assuming the language is finite we fix for every valuation N a formula γ_N such that $\text{mod}(\gamma_N) = \{N\}$. Observe that a valuation N is normal iff $\gamma_N \not\mid\!\sim \bot$.

Lemma 4.13 *Suppose the language is finite and $\mid\!\sim$ is preferential. Let N and M be normal valuations. Then $M <_e N$ iff $\gamma_M < \gamma_N$. In particular, $<_e$ is transitive.*

Proof: The order $<$ can be characterized in term of a preferential model given by 4.6. Fix a preferential model $\mathcal{M} = \langle T, \imath, \prec\rangle$ such that $\mid\!\sim = \mid\!\sim_{\mathcal{M}}$. Recall that for a given formula α, we denote by $\text{mod}_{\mathcal{M}}(\alpha)$ the set $\{s \in S : \imath(s) \models \alpha\}$. In [30] (see lemma 4.1) was shown that

$$\alpha < \beta \text{ iff for all } t \in \text{mod}_{\mathcal{M}}(\beta) \text{ there is } s \in \text{mod}_{\mathcal{M}}(\alpha) \text{ such that } s \prec t.$$

Notice that $s \in \text{mod}_{\mathcal{M}}(\gamma_N)$ iff $\imath(s) = N$. So it suffices to show the following fact

$$M <_e N \text{ iff for each } t \in T \text{ such } \imath(t) = N \text{ there is } s \in T \text{ such that } s \prec t \text{ and } \imath(s) = M$$

Suppose that $M <_e N$. Let $t \in T$ be such that $\imath(t) = N$. Consider the formula $\alpha = \gamma_N \vee \gamma_M$. Then $t \in \text{mod}_{\mathcal{M}}(\alpha)$ but t can not be minimal in $\text{mod}_{\mathcal{M}}(\alpha)$ otherwise we would have that $N \models C(\alpha)$ and $M \models \alpha$. Therefore there is $s \in \text{mod}_{\mathcal{M}}(\alpha)$ which is minimal and $s \prec t$. Clearly $\imath(s) \in \{N, M\}$ and since $\imath(s) \models C(\alpha)$ then as before $\imath(s) \neq N$.

Suppose now that $M \not<_e N$ and let α be a formula such that $N \models C(\alpha)$ and $M \models \alpha$. Since the language is finite, there is $t \in T$ such that t is minimal in $\text{mod}_{\mathcal{M}}(\alpha)$ and $\imath(t) = N$. Since $M \models \alpha$, then $\imath(s) \neq M$ for all $s \prec t$.

Since $<$ is transitive, then it follows that $<_e$ is also transitive. But this can be verified directly using the previous characterization. ∎

When W-DR holds we will give later a different proof of the previous lemma which does not use 4.6 (see 4.22).

We will see that under the presence of W-DR the relation $<_e$ is smooth and represents $\mid\!\sim$ in the sense that equation in 4.10 holds. For this reason we will use the following notion, which is more permissive than that of a standard model.

Definition 4.14 *Let* $\mid\!\sim$ *a consequence relation and* $\prec$ *a binary relation over* S. *We say that* $\prec$ *is a* ***standard relation*** *that represents* $\mid\!\sim$ *if the following holds*

$$mod(C(\alpha)) = min(mod(\alpha) \cap S, \prec) \tag{4.3}$$

We emphasize that we do not ask the relation to be a strict smooth order, but in most interesting cases the relation will be smooth. We show next that (4.3) implies that $\mid\!\sim$ satisfies W-DR.

Lemma 4.15 *Suppose* $\mid\!\sim$ *is a consequence relation and* $<$ *is a standard relation that represents* $\mid\!\sim$. *Then* $\mid\!\sim$ *satisfies W-DR.*

Proof: Let $N \models C(\alpha) \cup C(\beta)$, we have to show that $N \models C(\alpha \vee \beta)$. Since $<$ is standard and N is normal then from (4.3) we have that $N \in \min(\text{mod}(\alpha)) \cap \min(\text{mod}(\beta))$. It is easy to check that $N \in \min(\text{mod}(\alpha \vee \beta))$. ∎

The following observation shows that the essential relation associated with $\mid\!\sim$ is finer than any standard relation representing $\mid\!\sim$.

Lemma 4.16 *Let* $\mid\!\sim$ *be a consequence relation and* $<$ *a standard relation that represents* $\mid\!\sim$. *Then for all normal valuations* N *and* M, *if* $N < M$, *then* $N <_e M$.

Proof: Suppose N and M are normal valuations such that $N \not<_e M$. That is to say, there is α such that $N \models \alpha$ and $M \models C(\alpha)$. Since $<$ is standard and M is normal then from (4.3) we have that $M \in \min(\text{mod}(\alpha) \cap S, <)$, therefore $N \not< M$. ∎

The following observation is obvious and says that $<_e$ satisfies one half of (4.3) without any hypothesis about $\mid\!\sim$.

Lemma 4.17 *Let* $\mid\!\sim$ *be a consequence relation. If* $M \models C(\alpha)$ *then* $M \in min(mod(\alpha) \cap S, <_e)$. ∎

The following observation is well known [69]

Lemma 4.18 *Let* $\mid\!\sim$ *be a cumulative relation. If* $\alpha \mid\!\sim \beta$ *then* $C(\alpha) = C(\alpha \wedge \beta)$. ∎

Lemma 4.19 *Let* $\mid\!\sim$ *be a preferential relation. If* $M \models \alpha$ *and* $M \models C(\beta)$ *then* $M \models C(\alpha \wedge \beta)$.

Proof: Suppose $\alpha\wedge\beta\mid\!\sim\gamma$. We want to show that $M\models\gamma$. By the S rule, $\beta\mid\!\sim\alpha\rightarrow\gamma$, since $M\models C(\beta)$ then $M\models\alpha\rightarrow\gamma$. Since $M\models\alpha$, then $M\models\gamma$. ∎

Since we are dealing with non monotonic consequence relations we can not expect the set $T(M)$ to be closed under $\wedge$ (not even in the case of a rational consequence relation). On the other hand, in general, $T(M)$ is not closed under $\vee$. The next lemma establish under which condition $T(M)$ is closed under $\vee$.

Lemma 4.20 $\mid\!\sim$ *satisfies W-DR if and only if for any M, $T(M)$ is closed under the connective* $\vee$, i.e. *for any $\beta_1,\beta_2\in T(M)$, $(\beta_1\vee\beta_2)\in T(M)$.*

Proof: Suppose that $\beta_1,\beta_2\in T(M)$, so $M\models C(\beta_i)$ for $i=1,2$. Thus $M\models Cn(C(\beta_1)\cup C(\beta_2))$. By W-DR, $C(\beta_1\vee\beta_2)\subseteq Cn(C(\beta_1)\cup C(\beta_2))$, then we have $M\models C(\beta_1\vee\beta_2)$, *i.e.* $(\beta_1\vee\beta_2)\in T(M)$. The other direction is also straightforward. ∎

The following result is the basic representation theorem in this paper. All others representation theorems that we will show are based on it and will only add that $<_e$ has nicer properties (like being transitive, filtered, modular or quasi-linear) when the preferential relation $\mid\!\sim$ satisfies some extra postulates besides W-DR. This theorem is a generalization of Freund's main representation theorem (see his theorem 4.11 in [30]).

Theorem 4.21 *Let $\mid\!\sim$ be a consequence relation. Then $\mid\!\sim$ is a preferential relation satisfying W-DR if, and only if $<_e$ is a smooth standard relation representing $\mid\!\sim$.*

Proof: The *if* part follows from 4.12 and 4.15. For the *only if* part we start by showing that $<_e$ is irreflexive. If M is normal then there exists α such that $M\models C(\alpha)$, so $Th(M)\cap T(M)\supseteq\{\alpha\}\neq\emptyset$, *i.e.* $M\not<_e M$.

Now we show that $<_e$ is smooth. Let $M\in\mathrm{mod}(\alpha)\cap S$. We want to show that either $M\in\min(\mathrm{mod}(\alpha)\cap S,<_e)$ or there exists $N\in\min(\mathrm{mod}(\alpha)\cap S,<_e)$ with $N<_e M$. We consider two cases: $M\models C(\alpha)$ or $M\not\models C(\alpha)$. In the former case, by lemma 4.17, we have $M\in\min(\mathrm{mod}(\alpha)\cap S,<_e)$. In the latter case define $U=C(\alpha)\cup\{\neg\beta:\beta\in T(M)\}$. We claim that U is consistent. Otherwise by compactness there are $\alpha_1,\ldots,\alpha_m$ in $C(\alpha)$ and $\beta_1,\ldots,\beta_n$ in $T(M)$ such that $\{\alpha_1,\ldots,\alpha_m,\neg\beta_1,\ldots,\neg\beta_n\}\vdash\bot$. Hence $\alpha_1\wedge\cdots\wedge\alpha_m\vdash\beta_1\vee\cdots\vee\beta_n$. Put $\beta=\beta_1\vee\cdots\vee\beta_n$. By AND, $\alpha\mid\!\sim\alpha_1\wedge\cdots\wedge\alpha_m$, so by RW, $\alpha\mid\!\sim\beta$. Hence, by lemma 4.18, $C(\alpha)=C(\alpha\wedge\beta)$. By lemma 4.20, $\beta\in T(M)$. Thus by lemma 4.19, $M\models C(\alpha\wedge\beta)$, *i.e.* $M\models C(\alpha)$, a contradiction. Now consider N such that $N\models U$. By definition of U, $N\models C(\alpha)$ so by lemma 4.17, $N\in\min(\mathrm{mod}(\alpha)\cap S,<_e)$. Also by definition of U it is clear that $N<_e M$.

To see that $<_e$ is a standard relation that represents $\mid\!\sim$ it suffices to show that if $M\in\min(\mathrm{mod}(\alpha)\cap S,<_e)$ then $M\models C(\alpha)$, the other direction is given by 4.17. But this was already shown above, since we have proved that if $M\not\models C(\alpha)$, then $M\notin\min(\mathrm{mod}(\alpha)\cap S,<_e)$. ∎

Remark 4.22 For a finite language and in the presence of W-DR the proof of the transitivity of $<_e$ follows from the previous result. In fact, suppose that $M<_e N$ and $N<_e P$; we want to show that $M<_e P$. Consider the formula $\alpha=\gamma_M\vee\gamma_N\vee\gamma_P$. Note that $\mathrm{mod}(\alpha)=\{M,N,P\}$. By the assumptions, M is the only element of $\{M,N,P\}$ which can be minimal in $\mathrm{mod}(\alpha)$. Therefore by the smoothness of $\mathrm{mod}(\alpha)$, $M<_e P$. ∎

Putting together 4.12, 4.15, 4.21 and 4.22 we obtain the following result which is essentially the same result of Freund (see his theorem 4.13 in [30]) but with a different proof.

Theorem 4.23 *Assume the language is finite. Then* $\mid\!\sim$ *is a preferential relation satisfying W-DR if and only if* $<_e$ *is a standard order that represents* $\mid\!\sim$. ∎

4.4 Disjunctive, Rational and other relations

In this section we will use the main result of §4.3 to give give simple and uniform proofs of representation theorems for Disjunctive, Rational and other consequence relations. We start with those relations satisfying disjunctive rationality DR. The next remark is trivial but useful

Lemma 4.24 *DR is equivalent to saying that* $C(\alpha \vee \beta) \subseteq C(\alpha) \cup C(\beta)$ *for all formulas* α *and* β. *In particular, any consequence relation satisfying DR satisfies W-DR.* ∎

Lemma 4.25 *The following properties are equivalent for a cumulative relation* $\mid\!\sim$:

(i) The relation $\mid\!\sim$ *satisfies DR.*

(ii) For any valuations M, N *and for any formulas* α, β *if* $M \models C(\alpha)$ *and* $N \models C(\beta)$ *then* $M \models C(\alpha \vee \beta)$ *or* $N \models C(\alpha \vee \beta)$.

Proof: $(i \Rightarrow ii)$ Suppose $M \models C(\alpha)$ and $N \models C(\beta)$. For reductio, suppose $M \not\models C(\alpha \vee \beta)$ and $N \not\models C(\alpha \vee \beta)$. Then there are formulas $\gamma_1, \gamma_2 \in C(\alpha \vee \beta)$ such that $M \not\models \gamma_1$ and $N \not\models \gamma_2$. By AND, $\gamma_1 \wedge \gamma_2 \in C(\alpha \vee \beta)$, so by 4.24 $\gamma_1 \wedge \gamma_2 \in C(\alpha)$ or $\gamma_1 \wedge \gamma_2 \in C(\beta)$. But in both cases we get a contradiction because neither M nor N are models of $\gamma_1 \wedge \gamma_2$.

$(ii \Rightarrow i)$ Suppose $\gamma \in C(\alpha \vee \beta)$. We want to show that $\gamma \in C(\alpha)$ or $\gamma \in C(\beta)$. Suppose not. Then there are valuations M, N such that $M \models C(\alpha)$, $N \models C(\beta)$, $M \not\models \gamma$ and $N \not\models \gamma$. By (ii), $M \models C(\alpha \vee \beta)$ or $N \models C(\alpha \vee \beta)$. But in both cases we get a contradiction because neither M nor N are models of γ. ∎

The following relation between valuations was defined in [65]. We came up with the definition of $<_e$ by trying to extend the results in [65] to the case of an infinite language and to a larger class of consequence relations.

Definition 4.26 *Let* $\mid\!\sim$ *a consequence relation. We define the relation* $<_u$ *over the normal valuations by:*

$$M <_u N \leftrightarrow \forall\alpha\forall\beta\, [\, M \models C(\alpha) \;\&\; N \models C(\beta) \Rightarrow M \models C(\alpha \vee \beta) \;\&\; N \not\models C(\alpha \vee \beta)\,]$$

The relation $<_u$ is quite more intuitive and we show next that it is equal to $<_e$ under the presence of DR.

Lemma 4.27 *Let* $\mid\!\sim$ *a disjunctive rational relation. Then* $<_e$ *is equal to* $<_u$.

Proof: ($<_e \subseteq <_u$) Suppose $M <_e N$, $M \models C(\alpha)$, $N \models C(\beta)$. We want to show that $M \models C(\alpha \vee \beta)$ and $N \not\models C(\alpha \vee \beta)$. Since $M \models \alpha$, $M \models \alpha \vee \beta$ and $M <_e N$, then $N \not\models C(\alpha \vee \beta)$. Therefore by proposition 4.25, $M \models C(\alpha \vee \beta)$.

($<_u \subseteq <_e$) Suppose $M <_u N$. We want to show that $Th(M) \cap T(N) = \emptyset$. Suppose not, then there is a formula β such that $M \models \beta$ and $N \models C(\beta)$. Let α be a formula such that $M \models C(\alpha)$. By lemma 4.19, $M \models C(\alpha \wedge \beta)$; and since $M <_u N$, then $N \not\models C((\alpha \wedge \beta) \vee \beta)$. But $\vdash ((\alpha \wedge \beta) \vee \beta) \leftrightarrow \beta$, so $N \not\models C(\beta)$, a contradiction. ∎

Lemma 4.28 *If the relation* $\mid\!\sim$ *is disjunctive rational then* $<_e$ *is transitive.*

Proof: Suppose $N <_e M$ and $M <_e P$ but $N \not<_e P$. Let α be such that $P \models C(\alpha)$ and $N \models \alpha$. Let β be such that $M \models C(\beta)$, then it follows from the definition of $<_e$ that $P \not\models C(\alpha \vee \beta)$ and $M \not\models C(\alpha \vee \beta)$. By 4.25 $\mid\!\sim$ does not satisfy DR. ∎

The following definition is due to Freund [30]

Definition 4.29 *An order* $\prec$ *over valuations is filtered iff for any formula* α *and any valuations* $M, N \in mod(\alpha)$ *such that* $M \notin min(\alpha)$ *and* $N \notin min(\alpha)$ *there exists* $P \in min(\alpha)$ *such that* $P \prec M$ *and* $P \prec N$.

Lemma 4.30 *If* $\mid\!\sim$ *is disjunctive rational then* $<_e$ *is filtered.*

Proof: The argument is very close to that in the proof of the smoothness of $<_e$ (*cf.* proof of proposition 4.21). By hypothesis and lemma 4.17, $M \not\models C(\alpha)$ and $N \not\models C(\alpha)$. Put $U = C(\alpha) \cup \{\neg\beta : \beta \in T(M)\} \cup \{\neg\gamma : \gamma \in T(N)\}$. We claim that U is consistent. Suppose not, then by compactness there are $\alpha_1, \ldots, \alpha_m$ in $C(\alpha)$, $\beta_1, \ldots, \beta_n$ in $T(M)$ and $\gamma_1, \ldots, \gamma_r$ in $T(N)$ such that $\{\alpha_1, \ldots, \alpha_m, \neg\beta_1, \ldots, \neg\beta_n, \neg\gamma_1, \ldots, \neg\gamma_r\} \vdash \bot$. Hence $\alpha_1 \wedge \cdots \wedge \alpha_m \vdash \beta_1 \vee \cdots \vee \beta_n \vee \gamma_1 \vee \cdots \vee \gamma_r$. Put $\beta = \beta_1 \vee \cdots \vee \beta_n$ and $\gamma = \gamma_1 \vee \cdots \vee \gamma_r$. By AND, $\alpha \mid\!\sim \alpha_1 \wedge \cdots \wedge \alpha_m$ and by RW, $\alpha \mid\!\sim \beta \vee \gamma$. By observation 4.20, $\beta \in T(M)$ and $\gamma \in T(N)$. Thus by proposition 4.25 $M \models C(\beta \vee \gamma)$ or $N \models C(\beta \vee \gamma)$. Without lost of generality suppose that $M \models C(\beta \vee \gamma)$ (the other case is similar). By lemma 4.19, $M \models C(\alpha \wedge (\beta \vee \gamma))$ and since $\alpha \mid\!\sim \beta \vee \gamma$, then by lemma 4.18, $C(\alpha) = C(\alpha \wedge (\beta \vee \gamma))$, hence $M \models C(\alpha)$, a contradiction. Hence U is consistent. Let P be a model of U. By definition of U, $P \models C(\alpha)$, $P <_e M$ and $P <_e N$. So by 4.17 $P \in \min(\alpha)$ ∎

Freund [30] has shown that a consequence relation is disjunctive rational if and only if it has a standard filtered model. The next theorem is the hard half of his result with a different proof. The theorem follows from 4.21, 4.28, and 4.30.

Theorem 4.31 *Let* $\mid\!\sim$ *be a disjunctive rational relation. Then* $<_e$ *is a standard filtered order representing* $\mid\!\sim$. ∎

Now we look at the properties that $<_e$ would have in the presence of rational monotony RM. It is not difficult to check the well known fact (see [49]) that any rational relation satisfies DR. Thus, if $\mid\!\sim$ is rational then $<_e$ is filtered and in particular transitive. We have already mentioned that rational relations are represented by ranked models (see 4.7). A preferential model is ranked when the order relation is modular. We recall the definition of modular relation (see [55]):

Definition 4.32 *A relation $<$ on E is said to be modular iff there exists a strict linear order $\prec$ on some set Ω and a function $r : E \longrightarrow \Omega$ such that $a < b \Leftrightarrow r(a) \prec r(b)$.*

The following characterization of modularity is well-known and easy to verify.

Lemma 4.33 *An order $<$ on E is modular iff for any $a, b, c \in E$ if a and b are incomparable and $a < c$ then $b < c$.* ∎

The following result is well known and we include its proof for the sake of completeness.

Lemma 4.34 *Let $\mid\!\sim$ be a rational relation. If $\alpha \not\mid\!\sim \neg\beta$, then $C(\alpha\wedge\beta) = Cn(C(\alpha) \cup \{\beta\})$*

Proof: Let $\delta \in C(\alpha)$ then by RM we have $\delta \in C(\alpha\wedge\beta)$. Thus $Cn(C(\alpha) \cup \{\beta\}) \subseteq C(\alpha\wedge\beta)$. For the other inclusion, if $\alpha\wedge\beta\mid\!\sim\delta$ then by the rule S we have $\alpha\mid\!\sim\beta \rightarrow \delta$. Therefore $\delta \in Cn(C(\alpha) \cup \{\beta\})$. ∎

The next result shows that under the presence of RM it is quite easy to check that $N <_e M$.

Lemma 4.35 *Let $\mid\!\sim$ be a rational relation and N, M be normal models. Then $N <_e M$ if and only if there are α and β formulas such that $N \models C(\alpha)$, $M \models C(\beta)$ and $N \models C(\alpha\vee\beta)$ but $M \not\models C(\alpha\vee\beta)$.*

Proof: The *only if* part comes from 4.27 (recall that rational relations are in particular disjunctive rational). For the *if* part, suppose that such α and β exist, we will show that $N <_u M$. Let γ and δ be any formulas such that $N \models C(\gamma)$ and $M \models C(\delta)$. From proposition 4.25 we get that $\gamma\vee\delta \not\mid\!\sim \neg(\alpha\vee\beta)$ and also $\alpha\vee\beta \not\mid\!\sim \neg(\gamma\vee\delta)$. Hence from lemma 4.34 we get that

$$\begin{aligned} C((\alpha\vee\beta)\wedge(\gamma\vee\delta)) &= Cn(\, C(\gamma\vee\delta) \;\cup\; \{\alpha\vee\beta\}) \\ &= Cn(\, C(\alpha\vee\beta) \;\cup\; \{\gamma\vee\delta\}) \end{aligned}$$

and from this the result follows because $N \models Cn(\, C(\alpha\vee\beta) \;\cup\; \{\gamma\vee\delta\})$ so $N \models C(\gamma\vee\delta)$ and since $M \not\models Cn(\, C(\alpha\vee\beta) \;\cup\; \{\gamma\vee\delta\})$ and $M \models \alpha\vee\beta$, we have $M \not\models C(\gamma\vee\delta)$. ∎

A straightforward consequence of this lemma is the following

Lemma 4.36 *Let $\mid\!\sim$ be a rational relation and N, M be normal models. $N \not<_e M$ and $M \not<_e N$ if and only if N, $M \models C(\gamma\vee\delta)$ for all formulas γ and δ such that $N \models C(\gamma)$, $M \models C(\delta)$.* ∎

Lemma 4.37 *If the relation $\mid\!\sim$ is rational then $<_e$ is modular.*

Proof: Let M, N, P be normal valuations. Suppose $N \not<_e M$, $M \not<_e N$ and $M <_e P$. By 4.33 it suffices to show that $N <_e P$. Let α, β, γ be formulas such that $M \models C(\alpha)$, $N \models C(\beta)$ and $P \models C(\gamma)$. Since M and N are incomparable, by lemma 4.36 we have $M \models C(\alpha\vee\beta)$ and $N \models C(\alpha\vee\beta)$. We claim that $P \not\models C(\alpha\vee\beta\vee\gamma)$ and $N \models C(\alpha\vee\beta\vee\gamma)$, which implies, by lemma 4.35, that $N <_e P$. To prove the claim it suffices (by lemma 4.25) to see that $P \not\models C(\alpha\vee\beta\vee\gamma)$. Since $M <_e P$ and $M \models C(\alpha\vee\beta)$ and $P \models C(\gamma)$, then $P \not\models C(\alpha\vee\beta\vee\gamma)$. ∎

Now putting together 4.31 and 4.37 we get the following well known theorem which has been proved in many different ways ([55, 37, 30]). We will see in §4.5, that $<_e$ is in fact the unique standard modular order that represents a given rational relation.

Theorem 4.38 *If $\mid\!\sim$ is a rational relation then $<_e$ is a standard and modular relation that represents $\mid\!\sim$.*

To finish this section we will comment about a postulate stronger than rational monotony. A relation $\mid\!\sim$ is rational transitive, if it is preferential and the following rule (RT) holds

$$\text{RT} \quad \frac{\alpha\mid\!\sim\beta \quad \beta\mid\!\sim\gamma \quad \alpha\not\mid\!\sim\neg\gamma}{\alpha\mid\!\sim\gamma}$$

It is known that rational transitive consequence relations satisfies RM and that rational transitive consequence relations are represented by 'quasi-linear' standard relations (a relation $<$ is quasi-linear if M is a valuation that is not minimal then for any valuation N different of M we have $N < M$ or $M < N$) (see [9, 7]). If $\mid\!\sim$ is rational transitive then $<_e$ is quasi-linear (this follows from proposition 5.6 of [8]).

4.5 Uniqueness of representation.

In this section we will address the problem of when a consequence relation has a unique representation. We will also compare our relation $<_e$ with that introduced by Freund [30]. In particular, we will show that they coincide if DR holds.

Let us make first some simple observations to put the question in the right setting. By 4.9 we know that an injective model for a consequence relation $\mid\!\sim$ can be assumed to be defined without loss of generality on the set S of all normal valuations w.r.t. $\mid\!\sim$. In other words, there are consequence relations $\mid\!\sim$ that can be represented (as in 4.5) by various order relations defined on different sets of valuations. But there is always at least one such relation defined on the entire set S. It is nothing strange that there are so many representations, just recall that only countable many valuations are needed to define the semantic counterpart $\models$ of the classical entailment relation $\vdash$. Taking these considerations into account, the question we want to address is whether for a given preferential relation $\mid\!\sim$ (admitting an injective model) there is a unique order on S representing $\mid\!\sim$. In this generality, this uniqueness seems to be quite rare when the language is infinite (it holds when it is finite). So we will mainly be interested in the following more restrictive question: if there is a standard model, when is it unique?

It is well known that a subset T of the collection of valuations $\mathcal{U}$ suffices to define the classical relation $\models$ iff T is topologically dense in $\mathcal{U}$ with respect to a natural topology associated with $\mathcal{U}$. This topology turns out to be quite useful in relation with the problems we address in this section. Its use will make some proofs short and simple, and more important, we will show that $<_e$ has a topological property that makes it unique among other standard relation.

We will use the natural topology on the set of valuations coming from the identification of a valuation with the characteristic function of a set of propositional variables. In other words, each valuation N is viewed as a function $N : Var \to \{0,1\}$. The collection of all such functions is usually denoted by $\{0,1\}^{Var}$. This set is endowed with the usual product topology where $\{0,1\}$ is given the discrete topology. We will assume that Var is countable, so $\{0,1\}^{Var}$ is a metric space (in fact, homeomorphic to the classical Cantor space). The topology on $\{0,1\}^{Var}$ is then defined by declaring mod(α) as the basic open sets for every formula α (in fact, mod(α) is also closed). We will regard S as a topological space by using its subspace topology. The well known basic facts about this topology that will be needed in the sequel are stated in the following lemma.

Lemma 4.39 *(i) Let N and N_i with $i \geq 1$ be valuations. The following two conditions are equivalent: (a) N_i converges to N. (b) for all formula α, $N \models \alpha$ if and only if there is a j such that $N_i \models \alpha$ for all $i \geq j$.*

(ii) A set $F \subseteq S$ is closed in S iff given $N_i \in F$ converging to a normal valuation N, then $N \in F$.

(iii) If $F \subseteq S$ is closed in S and $N \in S \setminus F$, then there is a formula α such that $N \models \alpha$ and $P \not\models \alpha$ for all $P \in F$.

(iv) Let C be a set of formulas and $V \subseteq mod(C)$. Then $Th(V) = Cn(C)$ iff V is topologically dense in $mod(C)$ (i.e. for all $M \in mod(C)$ and all formula α with $M \models \alpha$, there is $N \in V$ such that $N \models \alpha$). ∎

It is convenient to have a quick way of checking when an injective representation is in fact standard. The following lemma will be useful.

Lemma 4.40 *Let $<$ be a relation over S representing $\mid\!\sim$.*

(i) If $N \notin min(mod(\alpha) \cap S, <)$ and $N \models C(\alpha)$, then there is a sequence $N_i \in min(mod(\alpha) \cap S, <)$ converging to N.

(ii) $<$ is standard iff $min(mod(\alpha) \cap S, <)$ is topologically closed for all α. In particular, if $min(mod(\alpha) \cap S, <)$ is finite for all α, then $<$ is standard.

Proof: From 4.39(ii) we have that $\mathrm{mod}(C(\alpha))$ is closed and by 4.39(iv) we have that $\min(\mathrm{mod}(\alpha) \cap S, <)$ is dense in $\mathrm{mod}(C(\alpha)$. From this the result follows. ∎

We will introduce next a property that $<_e$ has and in fact it is the unique standard relation (with this property) that represents $\mid\!\sim$.

Definition 4.41 *Let $<$ be a binary relation over S, we will say $<$ is* ***downward-closed*** *is for all N in S the set $\{M \in S : M < N\}$ is (topologically) closed in S.*

Lemma 4.42 *Let $\mid\!\sim$ be a consequence relation. Then $<_e$ is downward-closed.*

Proof: Let N, M, M_i be normal valuations with M_i converging to M. Suppose that $M_i <_e N$ for all i. We will show that $M <_e N$. Let α be a formula such that $N \models C(\alpha)$, then by assumption $M_i \models \neg\alpha$. Since M_i converges to M, then $M \models \neg\alpha$, *i.e.* $M <_e N$. ∎

Lemma 4.43 *Let $\mid\!\sim$ be a consequence relation. Suppose that $<$ is a standard relation that represents $\mid\!\sim$. If $<$ is downward-closed then $< = <_e$.*

Proof: From 4.16 we already know that $< \subseteq <_e$. For the other direction, let N, M be normal valuations such that $M \not< N$. We will show that $M \not<_e N$. Since $F = \{P \in S : P < N\}$ is closed and $M \notin F$, then by 4.39(iii) there is a formula α such that $M \models \alpha$ and $P \not\models \alpha$ for all $P \in F$. Let β be such that $N \models C(\beta)$. It suffices to show that $N \models C(\alpha \vee \beta)$. Since $<$ is standard and represents $\mid\!\sim$, then $N \in \min(\mathrm{mod}(\beta) \cap S, <)$. Hence $P \not\models \beta$ for all $P < N$. On the other hand, by the choice of α, we also have that $P \not\models \alpha$ for all $P < N$. Therefore $N \in \min(\mathrm{mod}(\alpha \vee \beta) \cap S, <)$ and since $<$ represents $\mid\!\sim$ then $N \models C(\alpha \vee \beta)$. ∎

From the previous results we immediately get the following

Theorem 4.44 *Let $\mid\sim$ be a preferential relation satisfying W-DR. Then $<_e$ is the unique downward-closed standard relation that represents $\mid\sim$.* ∎

A valuation $N \in S$ is said to be **isolated** in S, if there is a formula α such that $\text{mod}(\alpha) \cap S = \{N\}$. We will say that S is **discrete** if every $N \in S$ is isolated in S. These notions correspond to the topological notion of an isolated point and discrete space. In particular, every finite set is discrete. In every discrete space the only converging sequences are the eventually constant sequences, therefore every relation over a discrete space is trivially downward-closed. On the other hand, by using the same argument as in the proof of 4.22 it can be easily checked that if S is discrete and $\mid\sim$ satisfies W-DR, then $<_e$ is transitive. Moreover, by 4.40(i) we have also that any injective model defined on a discrete set is necessarily standard. Thus we have the following generalization of an analogous result known for finite languages.

Corollary 4.45 *Let $\mid\sim$ be a preferential consequence relation satisfying W-DR. If the collection of normal valuations is discrete, then $<_e$ is the unique (and in fact standard) order representing $\mid\sim$.* ∎

The following result might be known but it is now quite easy to show

Corollary 4.46 *Let $\mid\sim$ be a rational relation. Then $<_e$ is the unique standard modular order representing $\mid\sim$.*

Proof: It suffices to show that every modular standard order representing $\mid\sim$ is downward-closed. Let $<$ be such modular relation and M, N, N_i be normal valuations with N_i converging to N and $N_i < M$ for all i. Let α, β be formulas such that $M \models C(\alpha)$ and $N \models C(\beta)$. It suffices to show that $M \not\models C(\alpha \vee \beta)$. Since in this case, there must exists a normal valuation $P < M$ such that $P \models \beta$. Since $N \models C(\beta)$ and $<$ is modular, standard and represents $\mid\sim$ then $N < M$. To see that $M \not\models C(\alpha \vee \beta)$ we need to show that $M \notin \min(\text{mod}(\alpha \vee \beta) \cap S, <)$. Since $N \models \beta$ and N_i converges to N, then there is (in fact, infinitely many) i such that $N_i \models \beta$. Since $N_i < M$, then M is not minimal in $\text{mod}(\alpha \vee \beta)$. ∎

We will use the results presented in this section to compare $<_e$ with the relation $<_S$ defined by Freund [30]. Let $\mid\sim$ be a preferential relation. We say that α is $\mid\sim$-consistent if $\alpha \not\mid\sim \bot$. The *trace* of a formula α is denote by α^+ and is defined as the set of all formulas β such that $\alpha \vee \neg\beta \mid\sim \beta$. The relation $<_S$ is defined over S by

$$M <_S N \iff \forall \alpha \mid\sim\text{-consistent } (N \models \alpha^+ \Rightarrow M \not\models \alpha)$$

For $\mid\sim$ preferential, Freund showed that $<_S$ is transitive and irreflexive and also that $C(\alpha) = Cn(\{\alpha\} \cup \alpha^+)$ for all α. Now it is easy to verify that $<_S \subseteq <_e$ and that $<_S$ is a downward-closed relation.

A consequence relation is said to have the (**) property if the following holds for every pair of $\mid\sim$-consistent formulas α and β:

$$C(\alpha \vee \beta) = Cn(\alpha^+ \cup \beta^+ \cup \{\alpha \vee \beta\})$$

The (**) property seems to be tailor-made for getting part (i) of the following result

Theorem 4.47 *(Freund [30]) (i) A preferential relation $\mid\!\sim$ has the (**) property iff $<_S$ is a standard order representing $\mid\!\sim$.*

(ii) Every disjunctive relation has the (**) property.

(iii) The (**) property implies W-DR and they are equivalent when the language is finite.

(iv) DR is strictly stronger than W-DR. ∎

We will show in the next section (see 4.51) that (**) is strictly stronger than W-DR for an infinite language. Since $<_S$ is downward-closed and transitive then from 4.21, 4.44 and the previous theorem we conclude the following

Theorem 4.48 *Let $\mid\!\sim$ be a preferential relation satisfying W-DR. Then $\mid\!\sim$ has the (**) property iff $<_e = <_S$. In particular, if $\mid\!\sim$ has the (**) property, then $<_e$ is transitive.* ∎

4.6 Two examples and final comments

In this section we present two examples and make some final comments. Our first example shows that W-DR is not a necessary condition for having an injective model. In particular, by 4.15, we conclude that the property of having a standard model is strictly stronger than that of having an injective one. This result stands in contrast to what happens when the language is finite (see 4.23). Our second example shows that the (**) property is strictly stronger than W-DR and also that $<_e$ is not necessarily transitive.

Example 4.49 *(A preferential relation not satisfying W-DR and with an injective model)*Let $\{p_1, p_2, \cdots, p_n, \cdots\}$ denote the set of propositional variables. Let P be the valuation identically equal to one, *i.e.* $P \models p_i$ for all i. Let Q be the valuation satisfying $Q \models p_1$ and $Q \models \neg p_i$ for $i > 1$. Let N be the valuation identically equal to zero, that is to say, $N \models \neg p_i$ for all i. Let N_i and M_i be such that $N_i \models \neg p_1 \wedge \cdots \wedge \neg p_i$ and $N_i \models p_j$ for all $j > i$; $M_i \models \neg p_1 \wedge \cdots \wedge \neg p_i \wedge p_{i+1} \wedge \neg p_{i+2}$ and $M_i \models p_j$ for all $j > i+2$. Notice that both sequences converge to N.

We define a strict order $\prec$ over $S = \{N, P, Q, N_i, M_i\}$ by letting $P \prec N$, $Q \prec N$, $P \prec N_i$, $Q \prec M_i$, $N_i \prec N$ and $M_i \prec N$ for all $i \geq 1$. Let $\mid\!\sim$ be the preferential consequence relation defined by $(S, \prec)$. It is easy to check that S is the collection of all normal valuations w.r.t. $\mid\!\sim$. First we prove that every valuation in S is normal. Note that $\min(\text{mod}(\neg p_1) \cap S, \prec) = \{N_i, M_i : i \geq 1\}$ so N_i and M_i are normal for all i and since $\text{mod}(C(\neg p_1))$ is closed then $N \models C(\neg p_1)$. Notice that $N \notin \min(\text{mod}(\neg p_1) \cap S, \prec)$ and therefore $\prec$ is not standard. It is not difficult to see that $P \models C(p_1 \wedge p_2)$ and $Q \models C(p_1 \wedge \neg p_2)$. Conversely, suppose that $R \models C(\alpha)$. We want to show that $R \in S$. We know that $C(\alpha) = Th(\min(\text{mod}(\alpha) \cap S))$. By 4.40 there exists a sequence $R_i \in \min(\text{mod}(\alpha) \cap S)$ converging to R. But it is easy to see that S is closed, so $R \in S$.

We will show that $\mid\!\sim$ does not satisfies W-DR. For this end, it suffices to find two formulas α and β such that $N \models C(\alpha) \cup C(\beta)$ but $N \not\models C(\alpha \vee \beta)$. Let $\alpha = \neg p_1 \vee (p_1 \wedge \neg p_2)$ and $\beta = \neg p_1 \vee (p_1 \wedge p_2)$. It is easy to verify that

$$\begin{aligned}
\min(\text{mod}(\alpha) \cap S, \prec) &= \{Q\} \cup \{N_i : i \geq 1\} \\
\min(\text{mod}(\beta) \cap S, \prec) &= \{P\} \cup \{M_i : i \geq 1\} \\
\min(\text{mod}(\alpha \vee \beta) \cap S, \prec) &= \{P, Q\}.
\end{aligned}$$

Therefore $N \models C(\alpha) \cup C(\beta)$, but $N \not\models C(\alpha \vee \beta)$. ∎

Since having a standard representation is a more restrictive condition we expected that it might imply that in this case $<_e$ should be transitive. In other words, if $\mid\sim$ admits a standard representation (in particular, W-DR holds) then $<_e$ must be transitive (and thus it would be a standard order representing $\mid\sim$). Our second example shows that this is not the case.

Example 4.50 *A preferential relation $\mid\sim$ with a standard model (in particular W-DR holds) and $<_e$ not transitive*Let $\{p_1, p_2, \cdots, p_n, \cdots\}$ denote the set of propositional variables. We will define valuations N, M, P, N_i and M_i (for $i \geq 1$) viewing them as characteristic functions (*i.e.* as sequences of 0 and 1):

M_i	$=$	$< 0, 0, \cdots, 0, 1, 1, 1, \cdots >$	It starts with i ceros and then follows only 1's
N_i	$=$	$< 0, 0, \cdots, 0, 1, 0, 1, \cdots >$	It starts with i ceros, then follows 1, 0 and then only 1's
P	$=$	$< 1, 0, 1, 0, \cdots >$	1,0 periodically repeated.
M	$=$	$< 0, 0, \cdots >$	Only 0's
N	$=$	$< 1, 1, \cdots >$	Only 1's

The order among this valuation is the transitive closure of the following pairs

$$\begin{array}{rcl} N_i & < & M_i \\ N_i & < & N_{i+1} \\ M_i & < & P \\ N & < & M \end{array}$$

In particular we have that $N_i < P$ and also that $N_i < N_j$ and $N_i < M_j$ for all $i < j$. Notice that $M \not< P$. Let $S = \{N, M, P\} \cup \{N_i, M_i : i \geq 1\}$. Since $<$ is clearly wellfounded then it is smooth. Let $\mid\sim$ be the preferential relation defined by $(S, <)$. We claim that S is the collection of normal valuation w.r.t. $\mid\sim$. First, we show that the elements of S are normal. Notice that every valuation isolated in S is clearly normal. Since M is the only not isolated point of S it suffices to check that M is a normal valuation. In fact, it is easy to verify that $M \in \min(\text{mod}(\neg p_1) \cap S, <)$. Conversely, suppose $R \models C(\alpha)$. We want to show that $R \in S$. To see that it is enough to prove that $\min(\text{mod}(\alpha) \cap S, <)$ is finite for every formula α and then we apply 4.40. This also shows that $<$ is standard. Suppose that α uses only the letters $p_1, \cdots, p_s$. We consider two cases: (a) $\min(\text{mod}(\alpha) \cap S, <) \subset \{M, N, P\}$. In this case we are obviously done. (b) Suppose that $N_i \models \alpha$ or $M_i \models \alpha$ for some i. If $N_i \models \alpha$ for some i, then it is easy to verify that

$$\min(\text{mod}(\alpha) \cap S, <) \subset \{M, N\} \cup \{N_j, M_j : j \leq i\} \tag{4.4}$$

and we will be done. Suppose then that $M_i \models \alpha$ for some i. Let $\gamma = \neg p_1 \wedge \neg p_2 \wedge \cdots \wedge \neg p_s$, then $N_i, M_i \models \gamma$ for all $i \geq s$. Observe that if $M_i \models \alpha$ for some $i \geq s$, then $\gamma \vdash \alpha$, thus $N_s \models \alpha$ and therefore by (4.4) we are done. From this it follows that $\min(\text{mod}(\alpha) \cap S, <)$ is finite for all α.

Since $<$ is standard then from 4.15 we know that $\mid\sim$ satisfies W-DR and therefore by 4.21 $<_e$ is also a standard relation representing $\mid\sim$. By 4.16 we have that $< \subseteq <_e$. However, $<_e$ is not transitive. We have that $N <_e M$ (as $N < M$) and we claim that $M <_e P$ but $N \not<_e P$. In fact, it is easy to check that $N, P \models C(p_1)$ and therefore $N \not<_e P$. On the other hand, M_i converges to M, $M_i < P$ and since $<_e$ is downward-closed (by 4.42) then $M <_e P$. ∎

We will see below that in spite of the fact that $<_e$ might not be transitive it provides a very good representation of $\mid\!\sim$ even in some cases where other methods do not work.

Proposition 4.51 *The (**) property is strictly stronger than W-DR. Moreover, there is a preferential relation represented by $<_e$ but not by $<_S$.*

Proof: We will show that the consequence relation $\mid\!\sim$ given in 4.50 does not have the (**) property. Recall that $\mid\!\sim$ was defined by a strict order that in fact is a standard model of $\mid\!\sim$. In particular, $\mid\!\sim$ satisfies W-DR. Since $\mid\!\sim$ is preferential then $<_S$ is transitive. But $<_e$ is not transitive, thus $<_e \neq <_S$. Therefore, by 4.48 $\mid\!\sim$ does not have the (**) property. Moreover, by 4.47 (i) we conclude that $<_S$ does not represent $\mid\!\sim$, but by 4.21 $<_e$ does (even though $(S, <_e)$ is *not* a standard model of $\mid\!\sim$ because it is not transitive). ∎

A final question: is there a postulate that characterize when a preferential relation has an injective model or a standard model? By the example 4.49 we know that W-DR is not a necessary condition to have an injective model. The example 4.50 shows that the (**) property is not a necessary condition (but it is sufficient) to have a standard model. None of our examples have ruled out that W-DR suffices to obtain an standard model. Given a preferential relation $\mid\!\sim$ satisfying W-DR by 4.16 we know that any (if it exists) standard order representing $\mid\!\sim$ has to be contained in $<_e$. Thus we have to remove from $<_e$ some pairs in order to make it transitive. It is quite natural to use the following strategy to get an injective (hopefully standard) model of $\mid\!\sim$: start with $<_e$ and remove all instances of non transitivity and get $<_e^* \subset <_e$. It is quite curious that this process indeed leads to a transitive relation. In principle, one would expect that after a pair is removed other instances of non transitivity might appear. But this is not the case with $<_e$. However, it is not clear that this 'pruned" relation $<_e^*$ still represents $\mid\!\sim$ (we even don't know if $<_e^*$ is smooth). These two families of consequence relations seem so complex that we will not be surprised if there is no such a characterization (at least in terms of the type of postulates used so far to classify consequence relations).

Acknowledgments: Partial support for Carlos Uzcátegui was provided by a CDCHT-ULA (Venezuela) grant. This work was initiated while he was visiting the Laboratoire d'Informatique Fondamentale de Lille (LIFL), France. He would like to thank LIFL for the financial assistance and facilities they provided.

The final version of this paper was done when Ramón Pino Pérez was visiting the Mathematics department of University of Los Andes (Venezuela). He would like to thank the UNESCO's TALVEN program and the Mathematics department of ULA for the partial support they provided.

We would like to thank the anonymous referee for his careful reading of the paper and for the valuable suggestions which have helped us to improve its presentation.

Chapitre 5

Jumping to explanations vs. Jumping to conclusions

Abduction is usually defined as the process of inferring the best explanation of an observation. There are many information processing operations that can be viewed as a search for an explanation. For instance, diagnosis, natural language interpretation and plan recognition. This paper is concerned about the following aspects of abduction: (i) what are the logical properties of abduction when it is regarded as a form of inference?, (ii) how close is abduction to reversed deduction? and (iii) since preference criteria for selecting explanations are so fundamental to abduction, how is (i) and (ii) related to the selection mechanism?

In the logic-based approach to abduction, the background theory is given by a consistent set of formulas Σ. The notion of an explanation is defined by saying that a formula γ (consistent with Σ) is an explanation of α if $\Sigma \cup \{\gamma\} \vdash \alpha$. An explanatory relation is a binary relation $\rhd$ among formulas where the intended meaning of $\alpha \rhd \gamma$ is "γ is a preferred explanation of α". To each explanatory relation is associated a consequence relation $\mid\!\sim_{ab}$ defined as follows: $\alpha \mid\!\sim_{ab} \beta$ if $\Sigma \cup \{\gamma\} \vdash \beta$ for each γ such that $\alpha \rhd \gamma$.

The study of the logical properties of explanatory reasoning is approached by a systematic analysis of $\mid\!\sim_{ab}$. We show that there are rationality postulates for abduction (i.e. constrains on the explanatory relation $\rhd$) that are, in a very precise sense, equivalent to rationality postulates (in the Krauss-Lehmann-Magidor tradition) for non-monotonic reasoning (i.e. for the relation $\mid\!\sim_{ab}$). This tight correspondence between postulates for explanatory reasoning and non-monotonic reasoning will make apparent a strong duality between these two forms of inference. Isolating the postulates and showing this duality are one of the main contributions of the paper. We argued that abduction is reversed non-monotonic reasoning. It is also shown that "good" explanatory relations are given by preference relations over formulas.

Keywords: Abduction; explanatory and non-monotonic reasoning, nonmonotonic consequence relations; preferential orders.

5.1 Introduction

Abduction is usually defined as the process of inferring the best explanation of an observation. There are many information processing operations that can be viewed as a search for an explanation, and thus, as operations that perform some form of abduction. (a) Diagnosis is the typical example of abduction. When a system (an electrical circuit, a trade market or something as complex as a living being) is ill-functioning or not functioning as expected, we seek for explanations that will help to return the system to its normal state. If there is more than one explanation, usually some relevance or simplicity criterion is invoked to guide the selection of the best explanation. (b) We might need to explain an observation (input) α in order to give a meaning to it, because α itself is just a string of symbols. For instance, when reading a text we come across a word α that we do not know, we look up in a dictionary to give a meaning to it. If α has several senses, we select one of them according to the context. (c) We can also use abduction when trying to make a plan to achieve a goal or to decide how to continue an activity. For example, in order to decide what to do after an experiment is made (maybe to confirm or disprove a conjecture), the output data has to be analyzed and then, in the best case, it will be explained by the background theory.

A traditional model of abductive reasoning assumes a deductive relationship between the explanandum (or fact to be explained) and its explanations. The basic idea is to model abduction as reversed deduction plus some additional conditions. In this logic based approach to abduction, the background theory is given by a consistent set of formulas (which will be denoted by Σ) and a formula γ is said to be an explanation of α (w.r.t. Σ) if $\Sigma \cup \{\gamma\}$ entails α. To avoid trivial explanations it is also required that an explanation has to be a formula consistent with Σ. Since abduction is the process of inferring the "best" explanation, this notion of explanation captures only possible or candidate explanations of α. Thus some additional conditions are needed to define the key notion of "preferred explanations". This paper is concerned about the following three aspects of abduction: (i) What are the logical properties of abduction when it is regarded as a form of inference?, (ii) How close is abduction to reversed deduction? and (iii) Since preference criteria for selecting explanations are so fundamental to abduction, how is (i) and (ii) related to the selection mechanism? Let us see these three aspects separately.

(i) Several people have studied the logical properties of abductive reasoning: Zadrozny [114], Flach [29], Cialdea-Pirri [66] and Aliseda [2]. They have approached the problem by isolating rationality postulates or rules that abductive reasoning should conform to. As Zadronzny put it, abduction is an inference process that preserves sets of explanations. The structural properties we are looking for should provide a clear picture of the peculiar features that truly makes abduction a form of logical inference. The following are two basic questions related to this aspect:

a) How explanations can be combined to get new explanations? For instance, if γ is a preferred explanation of α and γ' entails γ, should γ' be considered also a preferred explanation of α? Another example, if γ is a preferred explanation of α and also of β, is γ a preferred explanation of $\alpha \vee \beta$? A related question is: How much a change of the observation affects its explanations? For instance, suppose that γ is a preferred explanation of $\alpha \wedge \beta$. Is γ also a preferred explanation of α?

b) Should changes on the background theory be allowed in other to explain an observation? and how much a change of the background theory affects explanations? For instance, suppose that γ is a preferred explanation of α w.r.t. Σ. Should γ be also a preferred explanation of α but now w.r.t. $\Sigma \cup \{\beta\}$?

Several sources of motivating ideas have been used for isolating the structural properties that will account for these basic questions. First of all, there is a vast literature on different areas of application of abduction: philosophy of science, linguistic, artificial intelligence, computer science, etc. All of them provide a large variety of examples where to look at for regularity patterns. A second source of ideas is, of course, found on the structural properties of logical deduction (both classical and non-classical). It has been studied which of them could be considered valid for explanatory reasoning and how to modify those which are not valid in the context of abduction. For a comprehensive overview of abduction we refer the reader to [2, 82] and also to [114, 29, 66]. The main idea for isolating our rules for explanatory reasoning will be explained in the following.

The examples given at the beginning of the introduction suggest that an important aspect of abduction is the set of conclusions to which the best explanation leads to. In other words, the consequences implied by the best explanation might be, in some cases, as relevant as the explanation itself. These considerations suggest that a measure of the "rationality " of an abductive method is given by the "rationality" of its "abductive consequences". More precisely, we view abduction as a binary relation between an observation and its preferred explanations. Following Flach's approach we work with a binary relation $\alpha \rhd \gamma$ between formulas which is read as saying γ *is a preferred explanation of* α. A *rationality postulate for explanatory reasoning* is a property of $\rhd$ saying that this relation is "well-behaved".

To each explanatory relation $\rhd$ we associate a consequence relation: given an observation α, we infer from α the common consequences of all preferred explanations of α. More formally, we define a consequence relation $\mid\!\sim_{ab}$ by

$$\alpha \mid\!\sim_{ab} \beta \quad \text{if } \Sigma \cup \{\gamma\} \vdash \beta \text{ for every } \gamma \text{ such that } \alpha \rhd \gamma. \tag{5.1}$$

We read $\alpha \mid\!\sim_{ab} \beta$ as "normally, when α is observed then β should also be present". In other words, β is a concomitant feature of any situation where α usually occurs.

The definition of $\mid\!\sim_{ab}$ is quite natural and, in fact, Levesque already suggested the idea of defining such consequence relation as a new deductive operation that would be useful when doing counterfactual experiments (see the concluding remarks of [56]). But the motivation to introduce this definition came from [65] where a consequence relation quite similar to $\mid\!\sim_{ab}$ was used to model abductive reasoning. Moreover, the results of [65] shows that $\mid\!\sim_{ab}$ has very nice formal properties. The key idea to isolate the postulates for explanatory reasoning introduced in §5.2 is based in the interplay between $\rhd$ and $\mid\!\sim_{ab}$. We would like $\mid\!\sim_{ab}$ to be a bona fide consequence relation and for this end we have searched for postulates for $\rhd$ mainly guided by the well known rationality postulates for consequence relations studied by Krauss, Lehmann and Magidor [49], Gärdenfors and Makinson [37] and many others [49].

We think that the use of the KLM methodology for isolating the postulates is not only an heuristic device but it also provides a fair enough justification for the postulates. The results of our analysis will give a formal justification for most of the postulates introduced by previous approaches and, in addition, it will shed new light on some aspects of abduction that we think have not been studied (this will be clarified in the following paragraphs).

In relation to b) it is clear that these questions implicitly have the assumption that the background theory is also a parameter and thus that abduction is a ternary relation. This issue was addressed by Cialdea-Pirri and Aliseda who presented rules that allows some changes on Σ. However, they considered only changes that consists of adding new formulas to Σ. This restriction is quite natural, since more substantial changes (like contracting or revising Σ) are not a trivial matter as it is by now well known from the theory of belief revision developed by Gärdenfors and others [1, 36]. In this paper the background theory will be fixed and therefore only formulas consistent with Σ can be explained. This can be considered a weakness since it has been argued that the more interesting observation are those which are not consistent with the theory ("surprising observations"). Boutilier and Becher [12] have presented a view of abduction based on the AGM theory for belief revision [1] by exploting the idea that observations inconsistent with the background theory can be explained by revising the theory in order to make the observation either true or at least possible. At a first glance our approach seems to be incompatible with the belief revision approach. The incompatibility occurs because in belief revision Σ is considered a belief set and therefore as something defeasible, but we will give Σ the role of a system description which is independent of the beliefs of the agent. The agent's believes are about which parts of the system are responsible for the observation but not about how the system is built. In other words, Σ represents the known laws of the world and base on them we explain an observation [1]. In spite of all this apparent differences, we will show in §5.4 that our approach also has an "epistemic" reading in the sense of belief revision.

(ii) Zadronzny, Cialdea-Pirri and Aliseda argued that abduction is a different form of reasoning and should not be reduced to reversed deduction. Flach's postulates reduces explanatory reasoning to reversed deduction (essentially because he did not include preference in his formalism). Nevertheless, his result goes in a direction similar to ours. The exact relationship between abduction and reversed deduction is however vague and, to our knowledge, has not being clarified in a formal way. We will say that an explanatory relation is *causal* if the following condition holds

$$\alpha \rhd \gamma \text{ iff } C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \gamma) \tag{5.2}$$

Where $C_{ab}(\alpha) = \{\beta : \alpha \mid\!\sim_{ab} \beta\}$ and $\mid\!\sim_{ab}$ is defined as in (5.1) and $Cn(X)$ is the set of classical consequences of X (for X a set of formulas or a formula).We will argue in §5.3 that (5.2) can formally be regarded as saying that $\rhd$ and $\mid\!\sim_{ab}$ are dual objects and therefore that causal explanatory reasoning is non-monotonic reasoning-in-reverse. We will see several example of explanatory relations based on belief revision which are not causal (in our sense). These examples will show that the main feature of causal explanatory relations is that they are based on a non defeasible notion of explanation (as opposite to those notions based on belief).

(iii) As we have said the most distinct feature of abduction is the emphasis it makes on preferred explanations rather than possible explanations. Most formalism we have mentioned include the notion of preference as an external requirement. Preference criteria for selecting the best explanation are regarded as qualitative properties (a sort of a simplicity criteria [2]) which are not reducible to logical ones. Moreover, in those formalism, the preference relation (for instance an order over formulas) is explicitly mentioned in the postulates that intend to capture the notion of "best" explanation. Cialdea and Pirri's approach tries to use preference criteria for selecting explanations based on logic but their results does not fully accomplish this goal since the preference relation has to be represented in a separated theory. We will show that preference criteria are implicit in the logical properties of abduction and therefore they do not need to be explicitly included as part of the postulates. In other words, the structural properties of explanatory reasoning implicitly include an order encoding which are the preferred explanations. More formally, we will show that (under some conditions) for every explanatory relation $\rhd$ there is an order relation $\prec$ such that $\alpha \rhd \gamma$ iff γ is a $\prec$-minimal explanation of α. The properties of $\prec$ are studied and shown to have a close connection with the postulates satisfied by $\rhd$.

The paper is organized as follows. In §5.2 we will introduce and study the postulates for explanatory relations. In §5.3 we will present the results showing the tight relationship between our postulates and the rationality postulates for consequence relations in the KLM style. Also,

[1] A different but related problem is to repair Σ after some unexplainable fact is observed (or when the explanation are shown to be incorrect by other means). We think this problem is very close related with inductive reasoning and deserve a separated study.

[2] Occam's razor: "*Entia praeter necessitatem non sunt multiplicanda.*"

we will introduce the notion of causal explanatory relation and show that they are the formal counterpart of non-monotonic consequence relations. In §5.4 we will see how our approach is viewed from the belief revision perspective. In §5.5 we will show the results concerning preference relations for defining explanatory reasoning. In §5.6 we will make precise comments about the work of Flach, Cialdea-Pirri and Aliseda. In §5.7 we will make some final remarks. Lists of the main postulates for consequence relations and explanatory relations used in the paper will be found in appendixes A.1 and A.2 respectively. A summary of the main results from sections §5.2 and §5.3 will be given in appendix B. The proofs will be given in section 5.8.

A preliminary version of this paper appeared as a technical report of LIFL 1997 (Lille, France) and part of it was presented at WOLLIC97 (Brazil) and at NMR'98 (Italy).

5.2 Reasoning with explanations

The *background theory* denoted by Σ, will be a consistent set of formulas in a classical propositional language. We will use the following notation: $\alpha \vdash_\Sigma \beta$ when $\Sigma \cup \{\alpha\} \vdash \beta$.[3] We could have avoid the use of Σ and $\vdash_\Sigma$ and instead use a semantic entailment relations $\models$ satisfying the standard requirements (like compactness and the properties of $\vee$ and $\wedge$). This way the background theory would be taken for granted and the notion of explanation would be somewhat elliptical. But we have chosen to keep Σ for several reasons. First of all, because it is customary in most presentation of abduction to have a background theory. Secondly, because many examples are naturally presented with a background theory that constrains the notion of explanation. And third, because by keeping Σ we leave open the question regarding the properties of abduction when the background theory is also considered a parameter.

We now introduce the notion of an explanation of a formula with respect to Σ.

Definition 5.1 *For every formula α, the collection of explanations of α w.r.t. Σ is denoted by $Expla(\alpha)$ and is defined as follows:*

$$Expla(\alpha) = \{\gamma : \gamma \not\vdash_\Sigma \bot \ \ \& \ \ \gamma \vdash_\Sigma \alpha\}$$

Notice that we have ruled out trivial explanations by asking that γ has to be consistent with Σ. We are interested in studying the relation "γ is a preferred explanation of α", which will be denoted by $\alpha \rhd \gamma$. In explanatory reasoning the input is an observation and the output is an explanation, that is the reason to write $\alpha \rhd \gamma$ with α as input and γ as output. Our next definition capture the ideas mentioned in the introduction.

Definition 5.2 *Let Σ be a background theory. An **explanatory relation** for Σ will be any binary relation $\rhd$ such that for every α and γ,*

$$\alpha \rhd \gamma \quad \Rightarrow \quad \gamma \not\vdash_\Sigma \bot \ \ \textit{and} \ \gamma \vdash_\Sigma \alpha$$

We read $\alpha \rhd \gamma$ as saying that γ is a preferred explanation (with respect to Σ) of α. The associated consequence relation is defined as follows

$$\alpha {\mid\!\sim}_{ab} \beta \ \ \stackrel{def}{\Leftrightarrow} \ \ \gamma \vdash_\Sigma \beta \textit{ for all } \gamma \textit{ such that } \alpha \rhd \gamma.$$

[3] Readers familiar with [65] should note that in that paper $\vdash_\Sigma$ denotes a different relation.

We read $\alpha \mid\!\sim_{ab} \beta$ *as "normally, when* α *is observed then* β *should also be present". The collection of all abductive consequence of an observation* $C_{ab}(\alpha)$ *is defined as follows*

$$C_{ab}(\alpha) = \{\beta : \alpha \mid\!\sim_{ab} \beta\}$$

As we said in the introduction our initial and motivating idea was that $\mid\!\sim_{ab}$ can be used heuristically to isolate the logical properties of explanatory relations. These properties will be called *postulates for explanatory reasoning.* We would like $\mid\!\sim_{ab}$ to be a bona fide consequence relation and for this end we have searched for the postulates mainly guided by the well known KLM rationality postulates for consequence relations [49] (a list of the main postulates for consequence relations is given in appendix A.1) The first thing we need is, of course, that $\mid\!\sim_{ab}$ has to be reflexive, *i.e.* $\alpha \mid\!\sim_{ab} \alpha$ for all α. This is obvious from the fact that when $\alpha \rhd \gamma$ then $\gamma \vdash_\Sigma \alpha$. Notice also that if $\alpha \vdash_\Sigma \beta$ then $\alpha \mid\!\sim_{ab} \beta$. In particular, if $\alpha \vdash_\Sigma \bot$, then $\alpha \mid\!\sim_{ab} \bot$.

A very natural assumption is to consider that explanatory relations are independent of the syntax. In our context this is expressed by the rules *Left Logical Equivalence* (LLE_Σ) and *Right Logical Equivalence* (RLE_Σ). Notice that these rules are somewhat stronger than the usual rules for consequence relations, since our notion of logical equivalence uses $\vdash_\Sigma$ instead of $\vdash$.

LLE$_\Sigma$: $$\frac{\vdash_\Sigma \alpha \leftrightarrow \alpha' \ , \ \alpha \rhd \gamma}{\alpha' \rhd \gamma}$$

RLE$_\Sigma$: $$\frac{\vdash_\Sigma \gamma \leftrightarrow \gamma' \ ; \ \alpha \rhd \gamma}{\alpha \rhd \gamma'}$$

Next we introduce a postulate called *Explanatory Cautious Monotony* (E-CM), since it has the form of a monotonicity rule on the left.

E-CM: $$\frac{\alpha \rhd \gamma \ ; \ \gamma \vdash_\Sigma \beta}{(\alpha \wedge \beta) \rhd \gamma}$$

This rule says that a preferred explanation γ of a simple observation α will be a preferred explanation of any observation more complex than α (like $\alpha \wedge \beta$) which is also entailed by γ. This seems quite natural because if we have decided that γ is a preferred explanation of α and we know further that γ implies β, then based on a larger set of observations (like $\alpha \wedge \beta$) it is reasonable to think that γ is a preferred explanation of $\alpha \wedge \beta$ (this situation will be natural when prefered explanations are choosen used some 'orders' between the explanations, cf. 5.41).

Now we will introduce the *Explanatory Cut rules.* These rules play an important role in our setting and, as we will see, there is a duality between monotony rules for consequence relations and cut rules for explanatory reasoning. Explanatory Cut rules relate the preferred explanations of an observation $\alpha \wedge \beta$ and the preferred explanations of α. If we have certain complex observation (like $\alpha \wedge \beta$), then we might have an explanation for it which is not a preferred explanation for a simpler observation (like α). The observation of two facts (symptoms) together or simultaneously "forces" to select an explanation which might not be considered a preferred explanation when only one of the facts is observed. A Cut rule will say that a preferred explanation of the more complex observation ($\alpha \wedge \beta$) might also be, in some cases, a preferred explanation of the simpler or incomplete observation (α). In other words, Cut rules allow to keep a preferred explanation even when the set of observations is not longer complete.

One could get an idea of the usefulness of an Explanatory Cut rule by looking at a diagnosis process: if we know a fairly complete list of a patient's symptoms, then we might be able to decide which is the most likely illness that caused them. However, what if we know only few of the symptoms? An Explanatory Cut rule says that in some cases this incomplete information suffices. These rules are the key fact for encoding preference criteria.

We will consider in this paper three Cut rules: Explanatory Cautious Cut, Explanatory Rational Cut and Explanatory Cut. We start with *Explanatory Cautious Cut* which is the weakest of all Cut rules.

E-C-Cut:
$$\frac{(\alpha \wedge \beta) \rhd \gamma \ , \ \forall \delta \, [\alpha \rhd \delta \ \Rightarrow \ \delta \vdash_\Sigma \beta\,]}{\alpha \rhd \gamma}$$

The meaning of E-C-Cut is more easily grasp by analyzing its contrapositive: suppose $(\alpha \wedge \beta) \rhd \gamma$ and $\alpha \not\rhd \gamma$, then there exists δ such that $\alpha \rhd \delta$ and $\delta \not\vdash_\Sigma \beta$. In particular, it says that if we are able to find a good explanation for $\alpha \wedge \beta$, then we should also be able to find a good explanation for α (but maybe a different one).

The following observation shows why we have imposed some restrictions in the hypothesis of E-C-Cut. Suppose E-CM and also that the following rule hold:

$$\frac{(\alpha \wedge \beta) \rhd \gamma}{\alpha \rhd \gamma} \tag{5.3}$$

Then we also have

$$\frac{\alpha \rhd \gamma \, , \alpha \vdash_\Sigma \beta}{\beta \rhd \gamma} \tag{5.4}$$

In fact, it follows from E-CM and the hypothesis of (5.4) that $(\alpha \wedge \beta) \rhd \gamma$, hence from the proposed rule (5.3) we obtain $\beta \rhd \gamma$. Notice also that (5.4) clearly implies (5.3).

We consider the rule (5.4) to be too strong to model the relation "γ is a preferred explanation of α". When γ is a *preferred* explanation of α, and α is a more specific observation than β (*i.e.* $\alpha \vdash \beta$), then the *preferred* explanations of β might not include γ, because we might need "less" to explain β than to explain α (an extreme case is when β is a consequence of Σ). We will present examples of natural explanatory relations which does not satisfy (5.3) and we will see that (5.3) implies $\mid\!\sim_{ab}$ to be monotonic.

Among our cut rules, this rule is essentially the only Cut rule we have seen in the literature. In most cases, it is presented by requiring that $\rhd$ being transitive.

In general $\rhd$ is not reflexive, because a formula might not be a *preferred* explanation of itself (this was already noticed in [66, 29]), however there is a version of reflexivity that holds in most cases.

E-Reflexivity:
$$\frac{\alpha \rhd \gamma}{\gamma \rhd \gamma}$$

Suppose that E-CM and E-C-Cut hold. Let $\alpha \rhd \gamma$, then by E-CM we have $(\gamma \wedge \alpha) \rhd \gamma$. It is easy to check that the hypothesis of E-C-Cut are satisfied and hence $\gamma \rhd \gamma$. So we have shown the following

Proposition 5.3 *Let* $\rhd$ *an explanatory relation satisfying E-CM and E-C-Cut. Then E-Reflexivity holds.*

The following result shows that the postulates for explanatory relations considered so far are the counterpart of cumulative consequence relations, *i.e.* relations satisfying the following rules:REF (reflexivity) $\alpha \mid\!\sim \alpha$LLE (left logical equivalence) $\alpha \mid\!\sim \beta \;\&\; \vdash \alpha \leftrightarrow \gamma \Rightarrow \gamma \mid\!\sim \beta$RW (right weakening) $\alpha \mid\!\sim \beta \;\&\; \vdash \beta \rightarrow \gamma \Rightarrow \alpha \mid\!\sim \gamma$ CUT $\alpha \wedge \beta \mid\!\sim \gamma \;\&\; \alpha \mid\!\sim \beta \Rightarrow \alpha \mid\!\sim \gamma$CM (cautious monotony) $\alpha \mid\!\sim \beta \;\&\; \alpha \mid\!\sim \gamma \Rightarrow \alpha \wedge \gamma \mid\!\sim \beta$

Theorem 5.4 *Suppose* $\rhd$ *satisfies* LLE_Σ*, E-CM and E-C-Cut, then* $\mid\!\sim_{ab}$ *is cumulative.*

Now will address the problem of how explanatory relations treat disjunctions. We will start by analyzing the right side. In [56] it was argued that if α has more than one preferred explanation, then the disjunction of all of them is the explanation that fully and non-trivially accounts for α. The consequence relation $\mid\!\sim_{ab}$ is capturing this intuition, since to compute the abductive consequences of α is irrelevant whether the collection of preferred explanations of α is closed under disjunctions. These considerations suggest the following postulate.

E-RW:

$$\frac{\alpha \rhd \gamma \;;\; \alpha \rhd \delta}{\alpha \rhd (\gamma \vee \delta)}$$

E-RW will be called *Explanatory Right Weakening.* It is the only rule that we will consider that allows to weakening a preferred explanation.

The next postulate is *Right Or*

ROR: If $\alpha \rhd (\gamma \vee \rho)$ then either $\alpha \rhd \gamma$ or $\alpha \rhd \rho$.

A stronger version of this postulates is the following

E-Disj:

$$\frac{\alpha \rhd (\gamma \vee \rho) \;;\; \gamma \not\vdash_\Sigma \bot}{\alpha \rhd \gamma}$$

Notice that E-Disj is the converse of E-RW.

The following postulate is the key fact to obtain that $\mid\!\sim_{ab}$ is preferential, *i.e.* if in addition to cumulative rules $\mid\!\sim_{ab}$ satisfies the rule Or: for any formulas α, β and γ if $\alpha \mid\!\sim_{ab} \gamma$ and $\beta \mid\!\sim_{ab} \gamma$ then $\alpha \vee \beta \mid\!\sim_{ab} \gamma$. This postulate will be called *Right And* (RA) since it gives some amount of monotony on the right. A similar postulate has been considered by Flach in [29].

RA:

$$\frac{\alpha \rhd \gamma \;;\; \gamma' \vdash_\Sigma \gamma \;;\; \gamma' \not\vdash_\Sigma \bot}{\alpha \rhd \gamma'}$$

RA says that any explanation more "complete"(stronger, more specific) than a preferred explanation of α is also a preferred explanation of α. This postulates says that explananda are not defeasible. The importance of RA is shown in the following

Proposition 5.5 *Let $\rhd$ be an explanatory relation satisfying RA. Then*

(i) $\rhd$ satisfies RLE_Σ and ROR.

(ii) if $\alpha \rhd \gamma$ and $\gamma \rhd \delta$, then $\alpha \rhd \delta$. In other words, $\rhd$ is transitive.

(iii) Suppose $\rhd$ also satisfies E-CM. If $\alpha \rhd \gamma$ and $\gamma \not\vdash_\Sigma \neg\beta$, then there is $\gamma' \vdash_\Sigma \gamma$ such that $\alpha \rhd \gamma'$ and $\gamma' \vdash_\Sigma \beta$, hence $(\alpha \wedge \beta) \rhd \gamma'$. In words, if γ is a preferred explanation of α and γ does not ruled out β, then we can extend γ to a preferred explanation of α and β.

(iv) E-Disj together with RLE_Σ is equivalent to RA.

Definition 5.6 *An explanatory relation is said to be* E-preferential *if satisfies LLE_Σ, E-CM, E-C-Cut and RA.*

The following result shows the relationship between the preferred explanation of a disjunction $\alpha \vee \beta$ and the preferred explanations of α and β.

Proposition 5.7 *Let $\rhd$ be a E-preferential explanatory relation. Then*

$$\{\gamma : (\alpha \vee \beta) \rhd \gamma\} \subseteq \{\gamma : \alpha \rhd \gamma\} \cup \{\gamma : \beta \rhd \gamma\} \cup \{\gamma : \vdash_\Sigma \gamma \leftrightarrow (\gamma_1 \vee \gamma_2),\ \alpha \rhd \gamma_1, \beta \rhd \gamma_2\}$$

It is interesting to observe the analogy between 5.7 and the fact that for a preferential consequence relation $C(\alpha) \cap C(\beta) \subseteq C(\alpha \vee \beta)$. In other words, the sets $\{\gamma :\ \alpha \rhd \gamma\}$ and $C(\alpha)$ seem to play dual roles.

The following result is a straightforward consequence of 5.7.

Corollary 5.8 *Suppose that $\rhd$ is a E-preferential explanatory relation then the following holds: if $(\alpha \vee \beta) \rhd \gamma$ and $\alpha \not\rhd \gamma$, then there is γ' such that $\beta \rhd (\gamma \wedge \gamma')$.* □

The next result says that E-preferential explanatory relations captures our initial motivation for introducing the postulates.

Theorem 5.9 *If $\rhd$ is an E-preferential explanatory relation, then $\mid\!\sim_{ab}$ is preferential.*

We will continue using the properties of $\mid\!\sim_{ab}$ as a guideline for isolating rationality postulates for abduction. We will consider next the following postulates: *Weak Disjunctive Rationality* (W-DR): $C(\alpha \vee \beta) \subseteq Cn(C(\alpha) \cup C(\beta))$, for every formulas α and β;*Disjunctive Rationality* (DR): if $\alpha \vee \beta \mid\!\sim \rho$ then either $\alpha \mid\!\sim \rho$ or $\beta \mid\!\sim \rho$ for any α, β and ρ;and*Rational Monotony* (RM): .These rules has been studied both from a semantics point of view [30, 55] and a syntactical point of view [69]. The new postulates for $\rhd$ will be related to properties satisfied by the preferred explanations of a disjunctive formula. Which is not surprising, since W-DR, DR and RM impose constrains to the set of consequences of a disjunctive formula.

The next postulate is *Left Or*

LOR:
$$\frac{\alpha \rhd \gamma \ ;\ \beta \rhd \gamma}{(\alpha \vee \beta) \rhd \gamma}$$

The intuition behind LOR is the following. Suppose that when we observe either α or β (no matter which one) we are willing to accept that γ is a very likely explanation for both of them. Now we are told that one of them is observed (but maybe it is not known which one). Is it rational to conclude that γ is still a very likely explanation of that observation (*i.e.* a very likely explanation of $\alpha \vee \beta$)? The new postulate implies that the answer is yes. It is interesting to notice that LOR was considered by Flach and Aliseda as a principle for confirmatory induction rather than for explanatory inference.

The dual of LOR is *Weak Disjunctive Rationality* (W-DR).Freund [30] proved that, in the case of finite languages, a preferential relation satisfies W-DR iff it can be represented by an injective preferential model, *i.e.* if there is a partial order $\leq$ over the collection of valuations of the language such that $mod(C(\alpha))$ is given by the $\leq$-minimal model of $mod(\alpha)$, where if A is a set of formulas or a formula $mod(A)$ is the set of models of A.

Theorem 5.10 *Suppose the language is finite and let $\rhd$ be an E-preferential explanatory relation that satisfies LOR. Then $\mid\!\sim_{ab}$ is preferential and satisfies W-DR.*

Remark: We don't know if theorem 5.10 holds when the language is infinite.

It is easy to see that DR is equivalent to saying that for every α and β, $C(\alpha \vee \beta) \subseteq C(\alpha) \cup C(\beta)$. Hence, DR is stronger than W-DR. The corresponding postulate for explanatory relations is the following:

E-DR:
$$\frac{\alpha \rhd \gamma \; ; \; \beta \rhd \delta}{(\alpha \vee \beta) \rhd \gamma \text{ or } (\alpha \vee \beta) \rhd \delta}$$

Proposition 5.11 *Let $\rhd$ be an explanatory relation satisfying E-DR. Then $\rhd$ satisfies LOR and $\mid\!\sim_{ab}$ satisfies DR.*

As a corollary of 5.9 and 5.11 we have

Theorem 5.12 *Let $\rhd$ be an E-preferential explanatory relation that satisfies E-DR. Then $\mid\!\sim_{ab}$ is preferential and satisfies DR.* □

A relation $\mid\!\sim$ is called *Rational* if it is preferential and satisfies Rational Monotony (RM).The corresponding postulate for abduction has the form of a Cut rule which is stronger than E-C-Cut. We will call it *Explanatory Rational Cut* (E-R-Cut).

E-R-Cut:
$$\frac{(\alpha \wedge \beta) \rhd \gamma \; ; \; \exists \delta \, [\alpha \rhd \delta \;\&\; \delta \vdash_{\Sigma} \beta]}{\alpha \rhd \gamma}$$

We will see later that E-R-Cut implies that preferred explanations (*i.e.* those formulas γ such that $\alpha \rhd \gamma$ for some α) are linearly pre-order. Moreover, when the underlying language is finite, E-R-Cut turns out to be equivalent to assigning a natural number to each formula and thus the preferred explanation of α are those explanations of α with minimal value.

To get a little intuition about this postulate we can paraphrase one of its equivalent forms as follows: if γ is a good explanation of $\alpha \wedge \beta$ but it is not a good explanation of α then any good explanation of α is consistent with $\neg\beta$.

Theorem 5.13 *Let $\rhd$ be an E-preferential explanatory relation that satisfies E-R-Cut. Then $\mid\!\sim_{ab}$ is rational.*

We will see next that E-R-Cut gives a fine structure to the set $\{\gamma : (\alpha \vee \beta) \rhd \gamma\}$.

Proposition 5.14 *Suppose $\rhd$ is an E-preferential explanatory relation that satisfies E-R-Cut. Then for every α and β one of the following holds:*

(a) $\{\gamma : (\alpha \vee \beta) \rhd \gamma\} = \{\gamma : \alpha \rhd \gamma\}$

(b) $\{\gamma : (\alpha \vee \beta) \rhd \gamma\} = \{\gamma : \beta \rhd \gamma\}$

(c) $\{\gamma : \alpha \rhd \gamma\} \cup \{\gamma : \beta \rhd \gamma\} \subseteq \{\gamma : (\alpha \vee \beta) \rhd \gamma\} \subseteq$
$\{\gamma : \alpha \rhd \gamma\} \cup \{\gamma : \beta \rhd \gamma\} \cup \{\gamma : \vdash_\Sigma \gamma \leftrightarrow (\delta \vee \rho) \;\&\; \alpha \rhd \delta \;\&\; \beta \rhd \rho\}$

Remark: The second $\subseteq$ in (c) above could be an equality if $\rhd$ satisfies E-RW . In fact, in case 3 in the proof of 5.14 we show that if $\alpha \rhd \delta$ and $\beta \rhd \rho$ then $(\alpha \vee \beta) \rhd \delta$ and $(\alpha \vee \beta) \rhd \rho$, therefore by E-RW we get that $(\alpha \vee \beta) \rhd (\delta \vee \rho)$. In this case, 5.14 is the analogous of the following well known fact about rational relations (which was found first in the context of belief revision [36, 70]): If $\mid\!\sim$ is rational then for every α and β one of the following holds: (a) $C(\alpha \vee \beta) = C(\alpha)$, (b) $C(\alpha \vee \beta) = C(\beta)$, (c) $C(\alpha \vee \beta) = C(\alpha) \cap C(\beta)$. The proof of 5.14 follows closely the proof of this fact about $\mid\!\sim_{ab}$. ∎

It is well known that any rational relation satisfies DR [69]. We will show next the corresponding result for E-DR (it will be used later in the paper).

Proposition 5.15 *Suppose $\rhd$ is E-preferential and satisfies E-R-Cut. Then it also satisfies E-DR.*

We consider next the role of full Monotony. It turns out that full Monotony is implied the strongest cut rule which we call E-Cut.

E-Cut:
$$\frac{(\alpha \wedge \beta) \rhd \gamma}{\beta \rhd \gamma}$$

As we have noticed before, E-Cut is equivalent, under the presence of E-CM, to the following:

$$\frac{\alpha \rhd \gamma \;\; ; \;\; \alpha \vdash_\Sigma \beta}{\beta \rhd \gamma} \tag{5.5}$$

Theorem 5.16 *Suppose $\rhd$ satisfies E-Cut then $\mid\!\sim_{ab}$ is monotonic.*

On the light of the previous results we will complete the definition 5.6 as follows

Definition 5.17 *Let Σ be a background theory and $\rhd$ be an explanatory relation. We say that $\rhd$ is **E-cumulative** if it satisfies E-CM, E-C-Cut and LLE_Σ. $\rhd$ is **E-preferential** if it is E-cumulative and in addition satisfies RA. $\rhd$ is **E-rational** if it is E-preferential and in addition satisfies E-R-Cut.*

We are about to finish the presentation of the postulates for explanatory reasoning. There is however one natural question that we have not considered yet. When an observation has a preferred explanation? The following postulate, that we call *Explanatory Consistency Preservation*, says that α has a preferred explanation iff it is consistent with Σ. Our last results are somewhat technical but they will be needed in the sequel.

E-Con$_\Sigma$: $\not\vdash_\Sigma \neg\alpha$ iff there is γ such that $\alpha \rhd \gamma$

The corresponding postulate for consequence relations will be called *Consistency Preservation* (with respect to Σ).

Con$_\Sigma$: For every formula α, (i) $\alpha \mathrel{|\!\sim} \bot$ iff $\vdash_\Sigma \neg\alpha$ and (ii) for every $\sigma \in \Sigma$, $\alpha \mathrel{|\!\sim} \sigma$.

Part (ii) in Con$_\Sigma$ was included since it necessarily holds for $\mathrel{|\!\sim}_{ab}$. The following observation is obvious.

Proposition 5.18 *Let* $\rhd$ *be an explanatory relation satisfying E-Con*$_\Sigma$ *, then* $\mathrel{|\!\sim}_{ab}$ *satisfies Con*$_\Sigma$*.* □

Under E-Con$_\Sigma$ E-R-Cut is stronger than E-C-Cut. More precisely we have the following

Proposition 5.19 *Any explanatory relation satisfying E-Con*$_\Sigma$ *and E-R-Cut satisfies E-C-Cut.*

As a corollary of 5.19 and 5.15 we have the following result:

Proposition 5.20 *Suppose that* $\rhd$ *satisfies LLE*$_\Sigma$*, E-CM, RA, E-R-Cut and E-Con*$_\Sigma$*. Then it also satisfies E-DR.*

As a corollary of 5.13, 5.19 and 5.18 we have the following result:

Proposition 5.21 *Let* $\rhd$ *be an explanatory relation that satisfies LLE*$_\Sigma$*, E-CM, E-R-Cut, E-Con*$_\Sigma$*, and RA.Then* $\mathrel{|\!\sim}_{ab}$ *is rational and satisfies Con*$_\Sigma$*.* ∎

Proposition 5.22 *Suppose* $\rhd$ *satisfies E-Cut and E-Con*$_\Sigma$*, then* $\mathrel{|\!\sim}_{ab} = \vdash_\Sigma$*.*

5.2.1 Two examples

We will present examples of E-preferential and E-rational explanatory relations. Both examples are based on preferential models. Preferential models are the main tool for representing and studying non-monotonic consequence relations (see [49] and the references therein). Given an order of the models of Σ we define a notion of preferred explanation. The intuition is that to explain an observation we only look at the closest worlds where the observation hold. It is not by accident that we use preferential models. In fact, explanatory relations defined in this way are quite universal in the sense that many explanatory relations are of that form. This will be addressed in §5.3.

We could have presented the examples just as a formal manipulation of symbols, but instead we choose to provide a context where to interpret the symbols. This kind of interpretations (that makes the reading more enjoyable) have a drawback: important aspects of the context are not included into the formalism used to model it; so one get the impression that the formalism is an over simplification of the problems under consideration. Our examples mainly pretend to illustrate some of the concepts we have introduced.

Example 1: Consider the following scenario. A message consisting of a finite sequence of 0 and 1 is sent by either one of two independent senders A and B. Messages sent by A always start with 0 and messages sent by B always start with 1. Sometimes only a portion of the message is received and thus it is necessary to recover the lost part. The person in charge of recovering messages, after many years of persistent work, has developed a quite simple preference criterion for guessing the correct message. He has observed that normally both A and B send messages starting with a constant sequence and moreover the sequence has even length. Since the senders are independent of each other he has not preference about who send the message. To make the example manageable we will assume that all messages have length 4. We will analyze later in the paper a similar example allowing messages of any length.

The preference criterion can then be represented as follows:

$$\begin{array}{ccc} \{0100, 0101, 0110, 0111, 0001\} & & \{1000, 1001, 1010, 1011, 1111\} \\ | & & | \\ | & & | \\ \{0000, 0010, 0011\} & & \{1111, 1100, 1101\} \end{array}$$

Where the messages at the bottom are more preferred than those at the top, but there is no relation between a message starting with 0 and a message starting with 1.

Let the letters a, b, c and d represent, in that order, the four digits of a message. The language $\mathcal{L}$ is the propositional language in the variables a, b, c and d and Σ is the empty set (any message can be either sent or received and there is no logical connection between the digits of a given message). Every message is a valuation of $\mathcal{L}$ and therefore the preference relation described is a partial order over the collection of all interpretations of the language. This partial order will be denoted by $\prec$. Notice that all valuations at the bottom (or top) are mutually incomparable. Given a formula α we define its minimal models as usual:

$$min(\alpha) = \{N : \ N \models \alpha \ \& \ M \not\models \alpha \text{ for all } M \prec N\}$$

Now we define an explanatory relation $\rhd$ as follows:

$$\alpha \rhd \gamma \stackrel{def}{\Leftrightarrow} Mod(\gamma) \subseteq min(\alpha)$$

for any pair of consistent formulas α and γ. We interpret $min(\alpha)$ as containing those messages encoded by α that have the most preferred features. Thus our definition says that γ is a preferred explanation for α if every message encoded by γ is one of the preferred messages encoded by α. This is not quite the same as saying that every *preferred* message encoded by γ is also one of the preferred messages encoded by α. The last statement holds if we ask that $min(\gamma) \subseteq min(\alpha)$. This alternative will be considered later.

It is easy to show that $\alpha \mid\!\sim_{ab} \beta$ iff $N \models \beta$ for all $N \in min(\alpha)$. This can be stated equivalently as $Mod(C_{ab}(\alpha)) = min(\alpha)$. Readers familiar with the theory of non-monotonic consequence relations will realize the motivation for our definition. We will make this connection clear in the forthcoming sections.

It is not difficult to show that $\rhd$ is a preferential explanatory relation. We will not proved this now since it is a consequence of a general result that will be shown later. We will compute some preferred explanations.

Suppose that the portion of the message we were able to get is expressed by the formula d (*i.e.* we only know that the fourth digit is 1). Then it is easy to check that the most likely sent messages are 0011, 1101 or 1111. Thus the preferred explanation of d are $\neg a \wedge \neg b \wedge c \wedge d$, $a \wedge b \wedge \neg c \wedge d$, $a \wedge b \wedge c \wedge d$ and the disjunction of them. In particular, $\rhd$ is not reflexive, for instance $d \not\rhd d$. Notice that $d \mid\!\sim_{ab} (\neg a \wedge \neg b \ \vee \ a \wedge b)$, which reflects the agent's preferences.

Let us suppose that in addition we know that the second digit was 0. Now the observation is encoded by $\neg b \wedge d$. In this case the most likely sent messages are 0011, 1001 and 1011. The formulas encoding these messages together with their disjunction are all the preferred explanation of $\neg b \wedge d$. Notice that E-R-Cut fails. In fact, 1001 is a preferred explanation of $d \wedge \neg b$ which is not a preferred explanation of d but there is a preferred explanation of d (namely 0011) that implies $\neg b$.

We have already suggested that there are others natural alternatives to define $\rhd$ based on a preferential model. For example, requiring that $min(\gamma) \subseteq min(\alpha)$ instead of $Mod(\gamma) \subseteq min(\alpha)$. The main difference of this alternative definition with respect to the one given above is that the former is reflexive and fails to satisfy RA but the later is not. This will be treated in section §5.4. ∎

Example 2: Leonidas, an old taxi driver, retired two month ago after 50 years of work. He lent his car to Julio, a cousin of him.Every time Leonidas has an opportunity he enjoyed himself by guessing which streets his cousin have driven his car by. Leonidas just needs to ask a couple of questions and then he is able to tell very precisely the exact route Julio took. He uses to say, making fun of Julio, *"my car is more like a metro train that needs no driver and you are in the car not really to drive it but only to collect the fare".* Once he got into a big trouble by trying to impress his cousin with his divining skills. He could not help himself and approached a young couple that just got off the car. Very politely he addressed them with his usual questions. "*Did you pass by* Cafe Kawi?, "*Did you pass by* Cine Paraíso?" The young couple got into a awful argument. The outburst, Leonidas and Julio thought, had nothing to do with the question they were asked. He said *"we did pass by Cafe Kawi but not by the movie theater"* and she replied, *"as usual, you were absent mind, thinking about god knows what! We did not pass by the Cafe but we did pass by the theater".* That day Julio made his uncle swear that he will never again bother his customers with such nagging questions. The old taxi driver slowly walked away, then turned his head and smiling said to Julio *"You did pass by the movie theater, anyway".* The reason for Leonidas's success in guessing the routes was that he has given Julio very precise indications about which were the best routes for avoiding traffic and finding good customers. He said to Julio: "*Always try to pass by either one of the two metro station* Chacaito *or* La Hoyada. *In case this is not possible, then try to pass by either* Cafe Kawi *or* Cine Paraíso. *If neither of these two alternatives are possible, do whatever you feel like*". Julio always follows Leonidas's advice to the letter.

The street map of the area covered by Leonidas's car is indicated below.

P	←	(2,4)	←	(3,4)	←	H		4th St.
↓		↑		↓		↑		
(1,3)	→	(2,3)	→	(3,3)	→	(4,3)		3rd St.
↓		↑		↓		↑		
(1,2)	←	(2,2)	←	K	←	(4,2)		2nd St.
↓		↑		↓		↑		
C	→	(2,1)	→	(3,1)	→	(4,1)		1th St.
1th Ave.		2nd Av.		3rd Av.		4th Av.		

Chacaito station is at C, *La Hoyada* station is at H, *Cafe Kawi* is at K and *Cine Paraíso* is at P.

To model this example we introduce one propositional variable $z_{i,j}$ for each one of the 16 corners in the map. The models of Σ will be paths without cycles through this map. Σ should contains formulas ensuring that each of its model consists of only one connected path.It is also convenient to add another 32 new variables to denote the starting and ending points. Let $s_{i,j}$ denote that the starting point was at (i,j) and similarly $e_{i,j}$ for the ending point. Some constrains about the starting and ending points must be added to Σ. [4]

Leonidas's preferences are given by a three level preferential model.

$$\begin{aligned} L_0 &= Mod(\Sigma \cup \{z_{1,1} \vee z_{4,4}\}) \\ L_1 &= Mod(\Sigma \cup \{\neg z_{1,1} \wedge \neg z_{4,4}, z_{3,2} \vee z_{1,4}\}) \\ L_2 &= Mod(\Sigma) \setminus (L_0 \cup L_1) \end{aligned}$$

This gives a total pre-order (*i.e.* a transitive and reflexive relation) of $Mod(\Sigma)$. The explanatory relation $\rhd$ is defined as in example 1. A general result, which will be proved later, guarantees that $\rhd$ is E-rational, satisfies E-RW and, since $\Sigma = L_0 \cup L_1 \cup L_2$ then $\rhd$ also satisfies E-Con$_\Sigma$.

Let us suppose that the couple got in the car at (3,4) and off the car at (2,2). Let α be $(z_{3,2} \wedge \neg z_{1,4} \vee \neg z_{3,2} \wedge z_{1,4}) \wedge s_{3,4} \wedge e_{2,2}$. Since α has models in L_0, then a preferred explanation of α must be a formula γ such that $mod(\gamma) \subseteq L_0$ and $\gamma \vdash_\Sigma \alpha$. It is clear that any path starting at (3,4) and ending at (2,2) can not pass by H. Hence any preferred explanation of α necessarily is a path passing by C. From this it is easy to check that there is only one solution and it includes P. Notice that there are several formulas describing this unique solution. For instance, $s_{3,4} \wedge z_{1,3} \wedge z_{2,1} \wedge e_{2,2}$. We do not need to mention all corners in this path. Some of them will be forced to be in the path by the rules of Σ. Observe that the preferred explanations of α are exactly the preferred explanations of $\neg k \wedge p \wedge s_{3,4} \wedge e_{2,2}$ (here recall 5.14).

Let β be the following "observation" $s_{2,1} \wedge z_{2,2} \wedge z_{2,3} \wedge e_{2,4} \wedge \neg k \wedge \neg z_{3,4}$. Any path satisfying β starts at (2,1), then it can not pass by C and since it does not pass by (3,4) then it can not pass by H. In fact, we have that $\beta \vdash_\Sigma \neg z_{1,4} \wedge \neg z_{3,2} \wedge \neg z_{1,1} \wedge \neg z_{4,4}$. This says that all models of β belong to L_2. Therefore the preferred explanations of β are formulas all whose models must be in L_2. What if we do not know the starting point? For instance, let α be

[4]For instance, $z_{1,1} \to (z_{2,1} \vee e_{1,1})$ and $z_{3,2} \to (z_{2,2}, \vee z_{3,1} \vee e_{3,2})$ must be in Σ. The first formula says that if a path pass through $(1,1)$ then either it ends at $(1,1)$ or goes to $(2,1)$. An analogous interpretation for the second formula. Also formulas like $e_{1,1} \to (\neg z_{2,1} \vee s_{1,1})$ and $e_{3,2} \to (\neg z_{2,2} \wedge \neg z_{3,1}) \vee s_{3,2}$ must be in Σ. These formula guarantee that the path ends at $(1,1)$ and $(3,2)$ respectively. Finally $\bigvee_{i,j} \bigvee_{i',j'} s_{i,j} \wedge e_{i',j'}$ ensures that there is a starting and ending point in each path. To ensure uniqueness of the starting and ending point we add $s_{i,j} \to z_{i,j} \wedge \neg s_{i',j'}$ and $e_{i,j} \to z_{i,j} \wedge \neg e_{i',j'}$ for all i, i', j, j' with $i \neq i'$ and $j \neq j'$.

$z_{2,2} \wedge z_{2,3} \wedge e_{2,4} \wedge \neg k \wedge \neg z_{3,4}$. This is a weaker observation and moreover α has models in L_0 (for instance a path starting at C, then it goes to (2,1), then goes through 2nd Ave. and finally stops at (2,4)). Hence none of the preferred explanations of β is a preferred explanation of α. This example shows that some parts of an observation are more important (because they are more preferred) than others and therefore cut rules must be constrained.

Let now β' be the following formula: $s_{2,1} \wedge z_{2,3} \wedge e_{2,4} \wedge \neg k \wedge \neg z_{3,4}$. We claim that the preferred explanations of β' are exactly the preferred explanations of β. In fact, it is easy to check that there are preferred explanations of β' that implies $z_{2,2}$. Then by E-R-Cut we conclude that any preferred explanation of β is also a preferred explanation of β'. This says that in this case $z_{2,2}$ is irrelevant and therefore can be ignored.

To relate the meaning of E-R-Cut with the ranked model that defines the explanatory relation, let us suppose that $(\alpha \wedge \beta) \rhd \gamma$. The constrain in E-R-Cut says that there must exist δ such that $\alpha \rhd \delta$ and $\delta \vdash_\Sigma \beta$. This implies that $min(\alpha)$ are at the same level as $min(\alpha \wedge \beta)$, therefore γ remains a preferred explanation for α. ∎

5.3 Explaining our reasoning

In the previous section we have shown that each explanatory relation has associated a consequence relation which reflects many properties of the explanatory relation. The intuition was: *if you tell me how to explain an observation, then I will tell you which are its usual or normal consequences.* In this section we will address the converse of the previous statement: *If you know which are the normal consequences of an observation, can you explain it?* In this setting there are two obvious thing one has to remark. The first one is that we are viewing the process of getting conclusions out of an observation and the process of explaining it as dual processes. But then it is natural to ask: are these two processes one the inverse of the other? In this section we will address also this question. We will introduce a notion of *causal explanatory relations* which corresponds to explanatory mechanisms that can be formally regarded as performing reversed non-monotonic deduction.

The normal consequences of an observation will be given by a consequence relation $\mid\!\sim$. We will assume that every such $\mid\!\sim$ is reflexive, *i.e.* $\alpha \mid\!\sim \alpha$ for all α. The first thing we must answer is under which conditions $\mid\!\sim$ is of the form $\mid\!\sim_{ab}$. It is obvious from the definition of $\mid\!\sim_{ab}$ that the question is then when the following holds:

$$C(\alpha) = \bigcap\{Cn\ (\Sigma \cup \{\gamma\}) : C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})\} \tag{5.6}$$

We formally introduce this condition in the following definition.

Definition 5.23 *A consequence relation $\mid\!\sim$ is said to be* adequate with respect to Σ *if (5.6) holds for every formula α.*

If $\rhd$ is an explanatory relation then, from the definition of $\mid\!\sim_{ab}$, it is clear that $\mid\!\sim_{ab}$ is adequate with respect to Σ. The classical entailment relation $\vdash$ is adequate with respect to $\{\top\}$ and $\vdash_\Sigma$ is adequate with respect to Σ. If there is no danger of confusion we will just say *adequate* instead of *adequate with respect to* Σ.

Given an adequate w.r.t. Σ consequence relation $\mid\!\sim$ it is clear that $\alpha \mid\!\sim \sigma$ for all $\sigma \in \Sigma$. Moreover, if $\alpha \not\mid\!\sim \bot$, then there must exist γ consistent with Σ such that $\gamma \vdash_\Sigma \alpha$. In particular, if $\alpha \not\mid\!\sim \bot$ then α is consistent with Σ. Hence $\mid\!\sim$ almost satisfies Con_Σ except that it might happen that $\alpha \mid\!\sim \bot$ for some α consistent with Σ. Also observe that an adequate consequence relation satisfies the following form of supraclassicality: if $\alpha \vdash_\Sigma \beta$, then $\alpha \mid\!\sim \beta$.

The notion of an adequate consequence relation is relevant only if the language is infinite. In fact, for a finite language, it is not hard to show that every consequence relation satisfying the following mild conditions is adequate: (i) $C(\alpha) = Cn(C(\alpha))$ and (ii) $\alpha \mid\!\sim \sigma$ for all α and all $\sigma \in \Sigma$. However, for infinite languages there are even rational relations satisfying Con_Σ which are not adequate (see example 5 in §5.3.2).

It is clear from (5.6) what should be the definition of the explanatory relation associated with a consequence relation.

Definition 5.24 *Let $\mid\!\sim$ be a consequence relation $\mid\!\sim$. We associate with $\mid\!\sim$ a binary relation $\tilde{\rhd}$ as follows:*

$$\alpha \tilde{\rhd} \gamma \overset{def}{\Leftrightarrow} \gamma \not\vdash_\Sigma \bot \ \& \ C(\alpha) \subseteq Cn\,(\Sigma \cup \{\gamma\}) \tag{5.7}$$

Notice that $\tilde{\rhd}$ is indeed an explanatory relation (using that $\mid\!\sim$ is reflexive). We have put a tilde above the symbol $\rhd$ to remind the reader that this explanatory relation is defined using a consequence relation $\mid\!\sim$. Suppose that $\mid\!\sim$ satisfies the following form of supraclassicality: if $\alpha \vdash_\Sigma \beta$, then $\alpha \mid\!\sim \beta$.

Then it is clear that if $\alpha \tilde{\rhd} \gamma$ then $\gamma \mid\!\sim \alpha$. However, in general $\gamma \mid\!\sim \alpha$ does not imply $\alpha \tilde{\rhd} \gamma$ as we will see in the examples. This suggests an alternative definition which will be treated in §5.4.

The following result is easy to show.

Proposition 5.25 *Every adequate consequence relation is of the form $\mid\!\sim_{ab}$.*

The next theorem shows the correspondence between the postulates satisfied by $\mid\!\sim$ and those satisfied by $\tilde{\rhd}$.

Theorem 5.26 *Let $\mid\!\sim$ be an adequate consequence relation, then*

1. *$\tilde{\rhd}$ satisfies RA and RLE_Σ.*
2. *If $\mid\!\sim$ satisfies LLE, then $\tilde{\rhd}$ satisfies LLE_Σ.*
3. *If $\mid\!\sim$ satisfies Con_Σ, then $\tilde{\rhd}$ satisfies E-Con_Σ.*
4. *If $\mid\!\sim$ satisfies CM, then $\tilde{\rhd}$ satisfies E-C-Cut.*
5. *If $\mid\!\sim$ satisfies the S-rule,* i.e. *$\alpha \wedge \beta \mid\!\sim \rho$ implies $\alpha \mid\!\sim \beta \to \rho$, then $\tilde{\rhd}$ satisfies E-CM.*
6. *If $\mid\!\sim$ satisfies W-DR, then $\tilde{\rhd}$ satisfies LOR.*
7. *If $\mid\!\sim$ is preferential and satisfies DR, then $\tilde{\rhd}$ satisfies E-DR.*

8. If $\mid\!\sim$ satisfies RM, then $\tilde{\rhd}$ satisfies E-R-Cut.

9. If $\mid\!\sim$ is monotone, then $\tilde{\rhd}$ satisfies E-Cut.

Remark: The hypothesis that $\mid\!\sim$ is adequate was used only to show that $\mid\!\sim = \mid\!\sim_{ab}$, E-Con$_\Sigma$ and E-C-Cut.

It is interesting to notice that we needed the S-rule, which is part of the preferential system, to get that $\tilde{\rhd}$ satisfies E-CM which is part of the cumulative system for explanatory relations. Also notice that we did not include the corresponding result for $\mid\!\sim$ cumulative. We will handle this case only for finite languages.

Proposition 5.27 *Suppose the language is finite. Let $\mid\!\sim$ be a cumulative relation such that $\alpha \mid\!\sim \sigma$ for all α and all $\sigma \in \Sigma$. Then there is an explanatory relation $\rhd$ satisfying, LLE$_\Sigma$, RLE$_\Sigma$, E-CM and E-C-Cut such that $\mid\!\sim = \mid\!\sim_{ab}$.*

5.3.1 Causal explanatory relations and reversed deduction

In the previous section we have shown that many consequence relations are of the form $\mid\!\sim_{ab}$. In this section we will address the dual question for explanatory relations. Namely, which explanatory relations are of the form $\tilde{\rhd}$? Let $\rhd$ be an explanatory relation and $\mid\!\sim_{ab}$ its associated consequence relation. Let $\tilde{\rhd}$ be the explanatory relation associated to $\mid\!\sim_{ab}$. Then the question is whether $\tilde{\rhd}$ is equal to $\rhd$. Consider the following condition on $\rhd$.

$$\alpha \rhd \gamma \text{ iff } C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \gamma) \tag{5.8}$$

Then our question can be equivalently stated as: Which explanatory relations satisfy (5.8)? First let us notice that in (5.8) the direction from left to right always holds. So the direction from right to left is what (5.8) really ask for.

Condition (5.8) says that $\rhd$ can be recuperated from $\mid\!\sim_{ab}$ and thus explanatory reasoning based on $\rhd$ can be viewed as performing a sort of reversed deduction with respect to $\mid\!\sim_{ab}$. We will give more evidence about the last claim later in this section. The failure of (5.8) means that even if we know that an agent is reasoning abductively, we might not be sure which explanatory relation the agent is using. In other words, looking only at $\mid\!\sim_{ab}$ we can not tell what are the agent's preferred explanations. We will isolated (5.8) in the following definition.

Definition 5.28 *An explanatory relations is said to be* ***causal*** *if it satisfies (5.8).*

In the following sections we will show some examples of explanatory relations which are far from being causal relations. These relations use a notion of explanation based on belief revision. Notice that $\mid\!\sim_{ab} = \vdash_\Sigma$ for any explanatory relation satisfying full reflexivity (*i.e.* $\alpha \rhd \alpha$ for every α consistent with Σ), thus such relations can not be causal unless they are trivial.

So far we have not presented any semantic characterization of explanatory relations. It is not difficult to see that some causal explanatory relations can be easily characterized in terms of preferential models. Cumulative, preferential and rational relations are represented by cumulative, preferential and ranked models respectively (see [49, 55, 30, 96]). Those models can also be used to represent causal explanatory relations. In fact, from (5.8) it follows that one can check whether $\alpha \rhd \gamma$ holds by looking at the model that represents $\mid\!\sim_{ab}$. To give an example we state the theorem corresponding to E-rational causal relations.

Theorem 5.29 *Let $\rhd$ be a E-rational causal relations satisfying E-Con$_\Sigma$. Then there is a ranked model $(Mod(\Sigma), \preceq)$ such that the following holds for every γ consistent with Σ*

$$\alpha \rhd \gamma \textit{ iff } Mod(\Sigma \cup \{\gamma\}) \subseteq min(\alpha)$$

∎

Now we will address the question of when a relation is causal. The first observation is that any relation of the form $\tilde{\rhd}$ trivially satisfies E-RW and RA. We will need a bit more than these two postulates to get a characterization of causal relations.

The following postulate is called *causality axiom* .

C Let α and γ be formulas consistent with Σ. If for all δ such that $\delta \not\vdash_\Sigma \bot$ and $\delta \vdash_\Sigma \gamma$ there is ρ such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \delta$, then $\alpha \rhd \gamma$

This postulates says that if any consistent extension of γ can also be extended to a preferred explanation of α, then γ itself is a preferred explanation of α. [5].

Theorem 5.30 *Let $\rhd$ be an explanatory relation. The following are equivalent.*

(i) $\rhd$ is causal.

(ii) $\rhd$ satisfies RA and C.

If the language is finite, causal explanatory relations are characterized by RA and E-RW. We will present a more general result that also applies to infinite languages. For that end we will require that every observation has at most finitely many preferred explanations. First, we introduce an auxiliary notion.

Definition 5.31 *A set of formulas A is said to have an upper bound (in A w.r.t Σ) if there are finitely many formulas $\alpha_1, \ldots, \alpha_n \in A$ such that for all $\alpha \in A$, $\alpha \vdash_\Sigma (\alpha_1 \vee \cdots \vee \alpha_n)$ (i.e, $\alpha_1 \vee \cdots \vee \alpha_n$ is an upper bound of A in the lattice of formulas modulo Σ).*

Definition 5.32 *An explanatory relation $\rhd$ is said to be* ***logically finite on the right*** *and denoted by RLF, if for every formula α the set $\{\gamma : \alpha \rhd \gamma\}$ has an upper bound.*

The previous notion can be stated equivalently in terms of models, as we show next.

Lemma 5.33 *Let A be a set of formulas. The following statements are equivalent:*

(i) A has an upper bound in A w.r.t. Σ.

(ii) $\mathrm{mod}(\bigcap_{\alpha \in A} Cn(\Sigma \cup \{\alpha\})) = \bigcup_{\alpha \in A} \mathrm{mod}(\Sigma \cup \{\alpha\})$.

Notice that if the language is finite then every explanatory relation obviously satisfies RLF.

[5] In topological terms, this postulates says that if the collection of preferred explanations of α is dense below γ (in the lattice of formulas), then γ is a preferred explanation of α.

Proposition 5.34 *Let $\triangleright$ be an explanatory relation satisfying RA, E-RW and RLF. Then $\triangleright$ is causal.*

We will show in §5.3.2 an example of a causal explanatory relations which does not satisfy RLF.

Corollary 5.35 *Suppose the language is finite. Let $\triangleright$ be explanatory relation. Then $\triangleright$ is causal iff it satisfies E-RW and RA.* ∎

What kind of relations are not causal? The examples that we will present in §5.4 use a notion of explanation based on belief revision which is the typical notion that does not satisfy RA; so by 5.30 thay are not causal.

5.3.2 More examples

It is easy to verified that the explanatory relations given in §5.2.1 are both causal. Example 1 is of the form $\tilde{\triangleright}$ for $\mid\!\sim$ a preferential consequence relation since we used a partial order to define the preferential model. In example 2 the preference relation is a total pre-order and hence the consequence relations is rational and the associate explanatory relation is E-rational.

Example 3: This example is a minor modification of one given in [65]. Consider the following scenario: Lisa lives in a high-rise and parks her car in the 16-floor parking garage of her building. One morning, Lisa was looking for her car and did not find it where she thought she left it the night before. She considered the possibility that she was in the wrong floor and went to the next floor. There was also the possibility that the car was stolen and she must had called the police, but Lisa looked for the elevator and went to the next floor instead before taking the extreme decision of calling the police. We could model part of her background theory as follows: Let the language consist of the propositional variables $\{c, r, s, f, p\}$, where r stands for *right_floor*, c for *car*, s for *stolen_car*, f for *go_to_next_floor* and p for *call_police*. The background theory Σ will be the following:

$$\Sigma = \begin{cases} \neg r & \rightarrow & \neg c \\ s & \rightarrow & \neg c \\ \neg r & \rightarrow & f \\ s & \rightarrow & p \end{cases}$$

Lisa's preference are linearly pre-ordered. She prefers "worlds" where her car has not been stolen. In case the car is not found, she would think that she is not at the right place. So she has a three level preferential model:

$$\begin{array}{rcl} L_0 & = & \{\{r, c\}\} \\ L_1 & = & \{\{f\}, \{f, p\}\} \\ L_2 & = & \{\{r\}, \{r, p\}, \{r, f\}, \{r, c, f\}, \{r, c, p\}, \{r, s, p\}, \{r, f, p\}, \\ & & \{r, s, p, f\}, \{r, c, p, f\}, \{s, p, f\}\} \end{array}$$

Notice that $mod(\Sigma) = L_0 \cup L_1 \cup L_2$. L_0 contains the initial states, in this case $\{r, c\}$. This is what Lisa expected before arriving to the parking place: the car will be there and she will not need to do anything else.

Let $\mid\!\sim$ be the rational consequence relation associated to this ranked model. That is to say

$$\alpha \mid\!\sim \beta \text{ iff } min(\alpha) \subseteq Mod(\beta).$$

Let $\tilde{\triangleright}$ be the explanatory relation associated to $\mid\!\sim$. Notice that $mod(\Sigma) = L_0 \cup L_1 \cup L_2$. We have that $mod(C(\neg c)) = \{\{f\}, \{f, p\}\}$. It is easy to check that $Mod(\Sigma \cup \{\neg r\}) = \{\{f\}, \{f, p\}, \{s, p, f\}\}$. Thus $\neg c \not\tilde{\triangleright} \neg r$, but it is clear that $\neg c \tilde{\triangleright} (\neg r \wedge \neg s)$ (notice that in this example Σ is playing a role). So $\neg r$ is not enough to explain why the car was not found. Since $\neg r \wedge \neg s \mid\!\sim f$, then Lisa will go to the next floor. Notice also that $s \in Expla(\neg c)$, however $\neg c \not\tilde{\triangleright} s$ because $mod(\Sigma \cup \{s\}) \not\subseteq mod(C(\neg c))$ (Lisa does not wish to think that the car was stolen as an explanation for not finding it). Observe also that $s \mid\!\sim \neg c$, so it is not sufficient that $\gamma \mid\!\sim \alpha$ in order that $\alpha \tilde{\triangleright} \gamma$. Finally, to illustrate how $\tilde{\triangleright}$ treats a disjunction, let us observe that $C(\neg c \vee s) = C(\neg c)$ and thus $(\neg c \vee s) \tilde{\triangleright} (\neg r \wedge \neg s)$ but notice that $s \not\tilde{\triangleright} (\neg r \wedge \neg s)$. ∎

Example 4: This example is similar to example 1 given in §5.2.1 but now we will allow messages of any length but only one sender. Again the preference criterion is simple: messages starting with an even number of 0 are the most preferred ones. To make easier the presentation let γ_n be, for each $n \geq 1$, the formula encoding the message of $2n + 1$ digits such that the first $2n$ digits are equal to 0 and the $2n + 1$ digit is equal to 1. Our language will be propositional on the countable set of variables $\{p_1, p_2, p_3, \cdots,\}$ and Σ will be the empty set[6]. Let

$$L_0 = \bigcup_{n \geq 1} Mod(\gamma_n)$$

and L_1 consists of the rest valuations. We have then a two level ranked model. Let $\mid\!\sim$ be the rational consequence relation defined by this model and let $\tilde{\triangleright}$ be explanatory relation associated with $\mid\!\sim$. We will show that $\tilde{\triangleright}$ is not logically finite on the right. In fact, suppose that the only portion of the message we were able to get only consists of ceros. Let us say $\alpha = \neg p_3 \wedge \neg p_5$. Then it is easy to check that $mod(\gamma_n) \subseteq min(\alpha)$ for all $n \geq 6$. Thus $\alpha \tilde{\triangleright} \gamma_n$ for all $n \geq 6$ and therefore no preferred explanation of α is an upper bound for all preferred explanations of α. This shows that $\tilde{\triangleright}$ is not logically finite on the right, but it is a causal explanatory relation by definition.

On the other hand, if the portion of the message contains at least one 1, then there is an upper bound for the set of preferred explanation for that message. For instance, let β be $\neg p_2 \wedge p_5$ be the incomplete message we received. Then $\gamma_1 \wedge p_5$ and γ_2 are preferred explanations for β. In other words, the beginning ofthe most likely messages sent were 00101, 00111 and 00001. In this case the upper bound is $(\gamma_1 \wedge p_5) \vee \gamma_2$. ∎

Example 5: We will present examples of an adequate and non adequate relation for an infinite language.

(i) Let $\{p_i : i \geq 1\}$ be the variables of the language and $\Sigma = \{p_1\}$. Consider the following two-level ranked preferential model: at the lowest level there will be only one model, M, defined by $M \models p_i$ for all $i \geq 1$ and at the second level we put all the other models of Σ (but not M). Let $\mid\!\sim$ be the relation associated with this ranked preferential model. Clearly $\mid\!\sim$ satisfies Con_Σ. Let $\alpha = p_1$. It is clear that $C(\alpha) = Th(M)$, thus there is no γ (consistent with Σ) such that $C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})$. Therefore (5.6) does not hold because its right hand side is empty and its left hand side is not empty.

(ii) Let Σ be the empty background theory and as in (i) we define a two-level ranked model: at the lowest level we put all models of p_1 and at the second level we put the other valuations

[6]We could have put $\Sigma = \{\neg p_1\}$ to make this example closer to example 1. But this is not important.

of the language (*i.e.* those which do not satisfy p_1). Let $\mid\!\sim$ be the rational relation associated with this ranked preferential model. We claim that $\mid\!\sim$ is adequate. In fact, let α be any consistent formula. We consider two cases: (a) Suppose $\alpha \vdash \neg p_1$, then it is easy to check that $C(\alpha) = Cn(\alpha)$. From this it follows that (5.6) holds. (b) Suppose $\alpha \not\vdash \neg p_1$, then it is easy to check that $C(\alpha) = Cn(\alpha \wedge p_1)$ and as before this implies that (5.6) holds. ∎

5.4 Connection with belief revision

We will show in this section the connection of our approach with the theory of belief revision. In particular, we will see the peculiar place that causal explanatory relations occupy when they are viewed from the perspective of belief revision.

Belief revision is the process of changing the beliefs an agent has in order to incorporate incoming information (which might contradict the old one). The best known formalism for belief revision is the so called AGM postulates [1]. Let K be the belief set of an agent (which we assume to be a propositional theory) and suppose that the new incoming information is represented by a formula α. The revision of K with α is denoted by $K * \alpha$. It is natural to assume that $K * \alpha$ is also a belief set (i.e. closed under logical consequences) and obviously that $\alpha \in K * \alpha$. The AGM postulates impose other non trivial conditions on $*$ in order to make minimal the changes it performs in K. For instance, if α is consistent with K then $K * \alpha = Cn(K \cup \{\alpha\})$. Gärdenfors and Makinson [37] have shown a tight connection of belief revision with the theory of non-monotonic consequence relations. Given an AGM revision operator $*$ they define a consequence relation by letting $\alpha \mid\!\sim_K \beta$ if $\beta \in K * \alpha$. In words, its says that the agent is willing to conclude β from α in the case that β belongs to the revised belief set obtained after α is incorporated into K (using the revision operator $*$). In [70] it is shown that $\mid\!\sim_K$ is a rational consequence relation in the sense of Kraus-Lehmann and Magidor [49]. On the other hand, they also have shown that every rational consequence relation $\mid\!\sim$ can be represented as a consequence relation of the form $\mid\!\sim_K$. In fact, let $\mid\!\sim$ be a rational consequence relation and let $K = \{\alpha : \top \mid\!\sim \alpha\}$. Define $*$ by $K * \alpha = C(\alpha)$. Then $*$ is a revision operator for K such that $\mid\!\sim$ is equal to $\mid\!\sim_K$ [7].

The connection between abduction and belief revision was already observed by Gärdenfors [36]. Boutilier and Becher [12] proposed a model of abduction based on the revision of the epistemic state of an agent. Aliseda [2] and also in [64] consider modeling belief revision with abduction. The main idea in all these papers is the same. We will follow the terminology of [12]. They consider various forms of explaining α relative to K and to an arbitrary (but fix) AGM revision operator. These type of explanations were called *epistemic explanations.* Epistemic explanations capture the intuition that *if the explanation were believed, so too would be the observation.* More precisely, they introduced the following.

Definition 5.36 [8] *Let $*$ be a AGM revision operator and K be a consistent set of formulas. An* epistemic explanation *for α relative to K and $*$ is any consistent formula γ such that*

[7] Formally $*$ can not be considered a revision operator because we have given only a description of how to revise a single knowledge base, namely $C(\top)$, and $*$ must be applicable to any knowledge base. Also $*$ might not satisfy one of the defining condition of an AGM operator. Namely, $*$ might not preserve consistency, it can happen that α is consistent but $K * \alpha$ is inconsistent. To avoid this problem one has to restrict to rational consequence relations that preserve consistency: $\alpha \not\vdash \bot$ iff $\alpha \not\mid\!\sim \bot$.

[8] This definition corresponds to what Boutilier and Becher called *predictive explanations.* This notion is the closer to our approach. We will not analyze other alternatives.

$\alpha \in K * \gamma$.

It is not difficult to see that the notion of epistemic explanations will not satisfy the postulate RA. Because if γ is an epistemic explanation of α, then $\gamma \wedge \delta$ is not in general an epistemic explanation of α. The reason is that $K * (\gamma \wedge \delta)$ is in general very different from $K * \gamma$. These notions of epistemic explanations " cannot be given a truly causal interpretation because they are simple beliefs that induce belief in the fact to be explained" [12]. The lack of a causal relationship between an observation and its epistemic explanations is precisely where our notion of explanation differs from theirs. There is also another very important difference. The relation "γ is an epistemic explanation of α" is not an explanatory relation in our sense. This is simply because an epistemic explanation might not have any deductive relationship with the explanandum. However, as revision operator preserves consistency, it is easy to see that an epistemic explanation has to be at least consistent with the explanandum. [9]. We will make a little detour in order to introduce a new concept that covers the notion of epistemic explanations.

Definition 5.37 *A binary relation* $\mid\!<$ *is called a* weak explanatory relation *if for all* α *and* γ

$$\alpha \mid\!< \gamma \quad \Rightarrow \quad \gamma \wedge \alpha \not\vdash_{\Sigma} \bot$$

Remark: Observe that for a weak explanatory relation the consequence relation $\mid\!\sim_{ab}$ is not necessarily reflexive. So $\mid\!\sim_{ab}$ might lose one of its more basic features and thus it does not make too much sense to study $\mid\!\sim_{ab}$ in this case. In the examples $\mid\!\sim_{ab}$ will be equal to $\vdash_{\Sigma}$, so it is trivial and does not give any information about the weak explanatory relation. All postulates we have introduced in §5.2 also apply to weak explanatory relations. Some of the results proved for explanatory relations are valid for weak explanatory relations. For instance 5.19 is valid.

The proof of 5.11 works for weak explanatory relation, so E-DR implies LOR in this case too.

It is easy to check that any weak explanatory relation satisfying RA is necessary an explanatory relation.

Let's go back the main theme of this section. Recall the rational consequence relation $\mid\!\sim_K$ associated to an AGM revision operator. The notion of epistemic explanation can the be restated as follows: γ is an epistemic explanation for α iff $\gamma \mid\!\sim_K \alpha$. From this it is obvious what are the logical properties satisfied by epistemic explanation. However, it is convenient to see which of our postulates for explanatory reasoning are satisfied by epistemic explanations.

Proposition 5.38 *Assume that* Σ *is the empty set. Let* $*$ *be an AGM revision operator and* K *be a consistent set of formulas. Let* $\mid\!<$ *be defined by* $\alpha \mid\!< \gamma$ *if* γ *is consistent and* $\alpha \in K * \gamma$. *Then* $\mid\!<$ *is a weak explanatory relation that satisfies LLE,* RLE_{Σ}*, E-CM, E-RW, ROR, LOR, E-Cut and full reflexivity (*i.e. $\alpha \mid\!< \alpha$ *for all consistent* α*).*

Epistemic explanation are very far from being causal in our sense, since neither the causal axiom C nor RA hold. Also let us remark that since transitivity of $\mid\!\sim$ implies monotonicity,

[9] We are assuming here that Σ is the empty set. This is not a crucial assumption. Our claims can easily extended to cover the case where Σ is not trivial

then notion of epistemic explanation is not transitive. [10]

The notion of epistemic explanation is too permissive. We can restrict it by asking a bit more from the explanations. Namely, we could say that γ is a *strong epistemic explanation* of α if

$$K * \alpha \subseteq K * \gamma \tag{5.9}$$

In other words, after revising K with the explanation we obtain all beliefs corresponding to the revision of K with the observation. If we state this new definition in terms of $\mid\!\sim_K$ we get the following condition: $C_K(\alpha) \subseteq C_K(\gamma)$. Where, as usual, $C_K(\alpha) = \{\beta : \alpha \mid\!\sim_K \beta\}$. It is convenient to see this condition as defining a notion of an explanation with respect to an arbitrary consequence relation $\mid\!\sim$. More precisely, consider the following condition for any γ such that $\gamma \not\mid\!\sim \bot$

$$C(\alpha) \subseteq C(\gamma) \tag{5.10}$$

This condition was suggested by Flach (Lehmman [54] has some preliminaries results about it[11]). In our setting it is quite natural to require that $\mid\!\sim$ satisfies Con_Σ. The next theorem shows which postulates are satisfied by this weak explanatory relation.

Proposition 5.39 *Let $\mid\!\sim$ be a preferential consequence relation satisfying Con_Σ. Define $\alpha \mid\!< \gamma$ if (5.10) holds for γ consistent with Σ. Then $\mid\!<$ is a weak explanatory relation and moreover*

(i) $\mid\!<$ is transitive, full reflexive for Σ-consistent formulas and satisfies LLE_Σ, RLE_Σ E-CM, E-RW and E-C-Cut.

(ii) If in addition $\mid\!\sim$ satisfies DR, then $\mid\!<$ satisfies LOR.

(iii) If in addition $\mid\!\sim$ satisfies RM, then $\mid\!<$ satisfies E-DR, ROR and E-R-Cut.

The relation $\mid\!<$ given in 5.39 does not satisfy E-Cut. This relation is also far from being causal, since neither satisfies C nor RA.

Since $K*\gamma$ is supposed to be closed under logical consequences and in our setting $\Sigma \subseteq K*\gamma$, then we have that $Cn(\Sigma \cup \{\gamma\}) \subseteq K * \gamma$. This suggests another way of strengthening (5.9). Consider the following notion of explanation

$$K * \alpha \subseteq Cn(\Sigma \cup \{\gamma\}) \tag{5.11}$$

This is exactly the defining condition of a causal explanatory relation. Let us see this in detail.

Let $\rhd$ be a causal explanatory relation. This means that the following holds

$$\alpha \rhd \gamma \text{ iff } C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \gamma) \tag{5.12}$$

Suppose also that $\rhd$ is E-rational and satisfies $\mathsf{E\text{-}Con}_\Sigma$. Then by 5.13 we know that $\mid\!\sim_{ab}$ is a rational consequence relation satisfying Con_Σ. As before, let $*$ be the revision operator

[10] We should mention that the original definition of predictive explanation given by Boutilier and Becher requieres an additional condition. When the observation α is entailed by K then they ask also that $\neg\gamma \in K * \neg\alpha$ which captures the intuition that if the observation had been absent, so too would be the explanation. With this new restriction we have that E-C-Cut holds but we do not have neither E-Cut nor E-R-Cut.

[11] We thank him for letting us have a copy of his manuscript.

associated with $\mid\!\sim_{ab}$ [12]. By definition $C_{ab}(\alpha)$ is equal to $K * \alpha$ and thus from (5.12) we have the following

$$\alpha \rhd \gamma \text{ iff } K * \alpha \subseteq Cn(\Sigma \cup \gamma)$$

which is exactly (5.11).

The initial knowledge base K is the collection $\{\alpha : \top \mid\!\sim_{ab} \alpha\}$. That is to say

$$K = \bigcap \{Cn(\Sigma \cup \{\gamma\} : \top \rhd \gamma\}$$

K represents the agent's belief before any observation is made. It is clear that $\Sigma \subseteq K$ and moreover, by E-Con$_\Sigma$, we have also that $\Sigma \subseteq K * \alpha$ for all α. It is not hard to check (using RA) that an observation α is consistent with K iff there is γ such that $\top \rhd \gamma$ and $\alpha \rhd \gamma$. Thus the intuition behind K is that it contains in addition to the background theory those beliefs initially considered by the agent to be the usual "causes" for explaining normal observations.

To give a clear picture of the meaning of (5.11) we must consider the following condition

$$Cn(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\}) \tag{5.13}$$

This corresponds to the notion of explanation given by Flach's postulates [29] and it is clear that (5.13) can be viewed as performing an expansion of the knowledge base instead of a revision.

Notice that (5.11) is stronger than (5.9). Thus any preferred explanation is a strong epistemic explanation. However, rather than saying that γ normally implies α (as Boutilier and Becher did) we say that γ implies everything that α normally implies. Condition (5.11) keeps some of the "epistemic" flavor of the belief revision approach and at the same time retains a strong causal relationship between an observation and its preferred explanations. Thus (5.11) is a sort of a middle ground between (5.13) and (5.9). Causal explanatory relations treats differently observations and explanation. Observations are viewed as beliefs but explanations are not. This epistemological distinction seems to capture the following idea. *We might be wrong about which is the "real" world (i.e. the preference relation might be incorrect), but we would like to be right about the causality relation used to explain the features of whichever world we happen to prefer.*

The selection mechanism behind causal explanatory relations depends basically on the observations in the following sense. After we observe α we collect the normal consequences of α (*i.e.* we compute $C(\alpha)$ based on either a preferential model or some sort of preferential order). A preferred causal explanation for α is *any* formula that together with the background theory accounts for α and its usual consequences. Thus the preference criteria is used mainly to compute those other facts ("symptoms") that usually occur simultaneously with α and thus should also be present. After doing that, we proceed to explain in the most straightforward way.

Example: To illustrate the differences between epistemic, strong epistemic and causal explanations let's go back to Lisa's example in §5.3.2. In this example K is the theory of $\{r, c\}$ which correspond to what Lisa expected before arriving to the parking place. The total pre-order defines a AGM revision operator in the usual way: $K * \alpha$ corresponds to the theory of the minimal models of α.

[12] As we said before, $*$ is not formally an AGM revision operator. However, it still capture the key idea of belief revision, that is to say, to minimize the changes of K.

It is easy to verify that f is a strong epistemic explanation of $\neg c$ (but notice that $f \wedge r$ is not). However, for us f is not even an explanation of $\neg c$ since $\Sigma \cup \{f\} \not\vdash \neg c$. Another instance, $\neg r$ is a strong epistemic explanation of $\neg c$, it entails $\neg c$ but it is not a preferred explanation in our sense. On the other hand, $\neg r \wedge \neg s$ is both a preferred explanation and a strong epistemic explanation of $\neg c$. Finally, $r \wedge p \wedge \neg c$ is an epistemic explanation of $\neg c$ but it is not a strong epistemic explanation of $\neg c$. ∎

5.5 Ordering the explanations.

As we have said in the introduction the most distinct feature of abduction is the emphasis it makes on preferred explanations rather than plain explanations. In this section we will focus on preference criteria for defining explanatory relations. We will show that these preference criteria are implicitly built in the structural properties of explanatory relations studied in §5.2. Even though these results seems new in the field of abductive reasoning (all formalism mentioned include preference as an external requirement), they are quite natural on the light of the well known facts about non-monotonic reasoning. In fact, it is well known that inference processes based on orders over formulas are one the "faces" of non monotonic reasoning [68]. For instance possibilistic orders [22] and expectations orders [37] characterize inference rational relations. Preferential orders [30] characterize preferential relations. We will comment about their connection with our results.

Perhaps the most natural way of defining an explanatory relation $\rhd$ is through a preference relation $\prec$ over formulas. The relation $\prec$ will tell which formulas in $Expla(\alpha)$ are the preferred ones: those formulas in $Expla(\alpha)$ that are $\prec$-minimal. Thus a natural question is what type of orders between formulas correspond to good explanatory relations. One of the purposes of this section is to answer this question. More precisely, we will address the problem of how to define explanatory relations based on preference relations over formulas. This will provide a formalization of the set $PE(\alpha)$ mentioned in the introduction. On the other hand, we are also interested in representing an arbitrary explanatory relation by means of an order among formulas.

We will start by making precise some basic notions. If $\prec$ is an irreflexive binary relation over a set S and $A \subseteq S$, then $a \in A$ is a *$\prec$-minimal* element of A if there is no $b \in A$ with $b \prec a$. The minimal elements of a set A will be denoted by $min(A, \prec)$ and when there is no confusion about which preference relation $\prec$ is used we will just write $min(A)$.

The most obvious explanatory relation $\rhd$ associated with $\prec$ is define as follows:

Definition 5.40 *Let $\prec$ be an irreflexive relation on formulas. The explanatory relation $\rhd$ associated with $\prec$ is defined by:*

$$\alpha \rhd \gamma \stackrel{def}{\Leftrightarrow} \gamma \in min(Expla(\alpha), \prec) \tag{5.14}$$

i.e. $\alpha \rhd \gamma$ *iff* $\not\vdash_\Sigma \neg\gamma$, $\gamma \vdash_\Sigma \alpha$ *and* $\delta \not\vdash_\Sigma \alpha$ *for all* δ *such that* $\delta \prec \gamma$.

Definition 5.40 is the same one given in [56, 66] (but notice that they worked with reflexive relations). Notice that $\prec$ is not necessarily transitive, thus the notion of $\prec$-minimal element might not be intuitive. We will be interested mainly in the case where $\prec$ is at least smooth (see definition below). We are interested in finding under which conditions there is, for a given an explanatory relation $\rhd$, a relation $\prec$ such that $\rhd$ satisfies (5.14). In this case we will say that $\prec$ *represents* $\rhd$.

First we point out a simple fact.

Proposition 5.41 *Let $\prec$ be a binary relation as in 5.40 and $\rhd$ be the corresponding explanatory relation. Then $\rhd$ satisfies E-Reflexivity and E-CM.*

To obtain other postulates we will impose some constrains over $\prec$. We formally define the notion of a preference relation.

Definition 5.42 *A **preference relation** $\prec$ will be any binary irreflexive relation $\prec$ over $\mathcal{L}$ which is invariant under logical equivalence w.r.t. Σ,* i.e. *if $\alpha \prec \beta$ and $\vdash_\Sigma \alpha \leftrightarrow \alpha'$ and $\vdash_\Sigma \beta \leftrightarrow \beta'$, then $\alpha' \prec \beta'$.*

The following notion of a smooth relation is inspired by the notion of smoothness used in the study of consequence relations ([49]).

Definition 5.43 *Let $\prec$ be a reflexive binary relation over a set S. We say that a subset $A \subseteq S$ is **smooth** if for every $a \in A$ either a is minimal in A or there is $b \in A$ with $b \prec a$ and b minimal in A. A preference relation $\prec$ as in 5.42 is called **smooth**, if for every formula α the set $Expla(\alpha)$ is smooth.*

To understand better the meaning of smoothness we remark the following: Let $A \subseteq B \subseteq S$, then it is clear that $min(B) \cap A \subseteq min(A)$. Suppose now that $min(B) \subseteq A$, hence $min(B) \subseteq min(A)$. It is reasonable then to expect that $min(A) = min(B)$. This is true when B is smooth since, in this case, $min(A) \subseteq min(B)$.

Theorem 5.44 *If $\prec$ is a smooth preference relation and $\rhd$ is defined as in (5.40), then $\rhd$ is an explanatory relation that satisfies LLE$_\Sigma$, RLE$_\Sigma$, E-CM, E-C-Cut and E-Con$_\Sigma$.*

It seems natural to expect that under the conditions in the conclusion of 5.44 the relation $\rhd$ is represented by a smooth preference relation as in 5.40. However, in order to get such representation we will need more than just E-CM and E-C-Cut. First we introduce an auxiliary notion.

Definition 5.45 *Let $\rhd$ be an explanatory relation. We will say that a formula γ is* admissible for $\rhd$ *if $\alpha \rhd \gamma$ for some formula α.*

The following definition is motivated by the results in [96].

Definition 5.46 *Let $\rhd$ be an explanatory relation that satisfies RLE$_\Sigma$. The* essential preference relation *associated with $\rhd$ is denoted by $\prec_e$ and defined by:*

(a) For δ not admissible: $\gamma \prec_e \delta$ for every admissible γ.

(b) For γ and δ admissible: $\gamma \prec_e \delta$ if $Cn(\Sigma \cup \{\gamma\}) \cap \{\beta : \beta \rhd \delta\} = \emptyset$.

The only relevant formulas for the definition of $\prec_e$ are admissible formulas. Since $\rhd$ satisfies RLE_Σ then $\prec_e$ is invariant under logically equivalence and thus it is indeed a preference relation. Notice that admissible formulas are consistent with Σ. Admissible formulas play in our paper the same role as normal models in [49, 65]. A concept similar to that of an admissible formula was defined in [29].

Remark: Suppose $\prec$ is a preference relation and $\rhd$ is the associated explanatory relation (5.40). It is easy to verify that if $\gamma \prec \delta$, then $\gamma \prec_e \delta$. In other words, $\prec_e$ is larger than $\prec$.

The proof that $\prec_e$ represents $\rhd$ works for explanatory relations which are logically finite either on the right (see 5.32) or on the left (see definition below).

Definition 5.47 *An explanatory relation $\rhd$ is said to be* ***logically finite on the left*** *and denoted by LLF, if for every admissible formula γ the set $\{\alpha : \alpha \rhd \gamma\}$ has an upper bound.*

Definition 5.48 *An explanatory relation $\rhd$ is said to be* ***logically finite*** *if it satisfies either RLF or LLF.*

Notice that if the language is finite then every explanatory relation is logically finite. We will give two example of logically finite relations:

Example 7: Let $\mid\!\sim$ be an adequate consequence relation and $\tilde{\rhd}$ be the explanatory relation defined in 5.24. If $\tilde{\rhd}$ is logically finite on the right, then there is a map F from formulas into formulas such that $C(\alpha) = Cn(\Sigma \cup \{F(\alpha)\})$. In fact, for every α let $F(\alpha)$ be $\gamma_1 \vee \cdots \vee \gamma_n$ the upper bound for $\{\gamma : \alpha \rhd \gamma\}$ given by 5.32 (if α is inconsistent with Σ, then we let $F(\alpha)$ be $\bot$). Conversely, it is clear that if such function F exists then $\tilde{\rhd}$ satisfies RLF.

Example 8: We will present an example of a LLF explanatory relation $\tilde{\rhd}$. We define first an adequate rational relation $\mid\!\sim$ as follows: Consider an infinite language $\mathcal{L} = \{p_1, p_2, \ldots\}$. Let $\mathcal{L}_n = \{p_1, p_2, \ldots, p_n\}$ and fix m models $M_1, \ldots, M_m$ for the language $\mathcal{L}_n$. Let $L'_1 \cdots, L'_k$ be a partition of $\{M_1, \ldots, M_m\}$ in k levels and let $L'_i = \{M_i^1, \ldots, M_i^{n_i}\}$ for $i = 1, \ldots, k$. Now consider the ranked model in the language $\mathcal{L}$ given by k levels $L_1, \ldots, L_k$, where $M \in L_i$ iff the restriction of M to $\mathcal{L}_n$ is in L'_i. Let γ_i^j be formulas in the language $\mathcal{L}_n$ such that $mod(\gamma_i^j) = M_i^j$. For $r = 1, \ldots, k$ we let β_r be the following formula:

$$\beta_r = \bigvee_{i=r}^{k} (\bigvee_{j=1}^{n_i} \gamma_i^j)$$

Let $\Sigma = \{\beta_1\}$. It is not hard to see that the rational relation $\mid\!\sim$ generated by this model is adequate (with respect to Σ). Moreover, this ranked model is standard, *i.e.* for every formula α, $mod(C(\alpha)) = mod(\alpha) \cap L_i$, where i is the first integer j such that $mod(\alpha) \cap L_j \neq \emptyset$. It is easy to check that a formula γ is admissible iff $C(\gamma) = Cn(\Sigma \cup \{\gamma\})$. Let γ be an admissible formula, then there is r such that $mod(\Sigma \cup \{\gamma\}) \subseteq L_r$. We claim that β_r is an upper bound for $\{\alpha : \alpha \tilde{\rhd} \gamma\}$. In fact, it is easy to see that $mod(\beta_r) = \bigcup_{i=r}^{k} L_i$ and $mod(C(\beta_r)) = L_r$. Hence $\beta_r \tilde{\rhd} \gamma$. Now, if $\alpha \tilde{\rhd} \gamma$, then $mod(\Sigma \cup \{\alpha\}) \subseteq \bigcup_{i=r}^{k} L_i$. Thus $\alpha \vdash_\Sigma \beta_r$.

Remark: It seems that the idea used in the example 8 above can be generalized to the case of a finite Σ. In fact, we do not know any example of an adequate consequence relation $\mid\!\sim$ for which $\tilde{\rhd}$ is not logically finite. We conjecture that there are none. ∎

The next theorem gives a characterization of those logically finite explanatory relations representable by preference relations.

Theorem 5.49 *Let $\rhd$ be a logically finite explanatory relation. The following are equivalent:*

(i) The relation $\rhd$ satisfies E-CM, LLE_Σ, RLE_Σ, E-C-Cut, E-Con_Σ and LOR.

(ii) There is a smooth preference relation $\prec$ such that

$$min(Expla(\alpha)) \cap min(Expla(\beta)) \subseteq min(Expla(\alpha \vee \beta)) \tag{5.15}$$

and for every formula α the following holds

$$\alpha \rhd \gamma \quad \textit{iff} \quad \gamma \in min(Expla(\alpha), \prec) \tag{5.16}$$

We will continue now analyzing the properties that $\prec_e$ has when $\rhd$ satisfies extra axioms. We will leave the analysis of the effect that RA has on $\prec_e$ to section §5.5.1. This postulate is related, as we already know from §5.3, with causal explanatory relations.

When $\rhd$ satisfies E-DR, then $\prec_e$ can be described in a different way (a similar idea was used in [65, 96]). Recall that from 5.11 we know that E-DR implies LOR. We introduce the following definition

Definition 5.50 *Let $\rhd$ be an explanatory relation that satisfies RLE_Σ. Define a binary relation $\prec_u$ by:*

(a) For δ not admissible: $\gamma \prec_u \delta$ for every admissible γ.

(b) For γ and δ admissible:

$$\gamma \prec_u \delta \overset{def}{\Leftrightarrow} \forall\alpha\forall\beta\,[\alpha \rhd \gamma \;\&\; \beta \rhd \delta \;\Rightarrow\; (\alpha \vee \beta) \rhd \gamma \;\&\; (\alpha \vee \beta) \not\rhd \delta]$$

Proposition 5.51 *Let $\rhd$ be an explanatory relation that satisfies LLE_Σ, RLE_Σ, E-CM, E-C-Cut, and E-DR. Then $\prec_e = \prec_u$. Moreover, $\prec_u$ (and therefore $\prec_e$) is transitive.*

Remarks: In [30] it was used a notion of *filtered* relation. We adapt next this notion to our context. We say that a preference relation $\prec$ is filtered if for every α and every $\gamma, \gamma' \in Expla(\alpha)$ such that $\gamma \notin min(Expla(\alpha))$ and $\gamma' \notin min(Expla(\alpha))$, there is $\delta \in min(Expla(\alpha))$, such that $\delta \prec \gamma$ and $\delta \prec \gamma'$. Using an argument similar to that in the proof of 5.49 it can be proved that if $\rhd$ is a logically finite explanatory relation that satisfies RLE_Σ, E-CM, E-C-Cut, E-Con_Σ and E-DR then $\prec_u$ (therefore $\prec_e$) is filtered. We do not know if the converse is true. ∎

5.5.1 Causal orders

Now we will look more closely at the relation $\prec_e$ when $\rhd$ satisfies RA and C.

The following proposition says that if RA holds, then $\prec_e$ satisfies almost all the properties of a *preferential preordering* as defined by Freund in [30]. In §5.6 we will compare in more detail the properties of preferential orders and the essential relation $\prec_e$.

Proposition 5.52 *Let $\rhd$ be an explanatory relation that satisfies RLE_Σ. Then the following holds:*

(i) Let γ, γ' and δ be admissible formulas such that $\gamma \vdash \gamma'$. If $\gamma \prec_e \delta$, then $\gamma' \prec_e \delta$.

(ii) Let γ and δ be admissible formulas. If $(\delta \vee \gamma) \prec_e \gamma$, then $\delta \prec_e \gamma$.

(iii) Suppose that $\rhd$ also satisfies RA. Then we have

(a) Let γ, γ' and δ be formulas such that $\gamma \nvdash_\Sigma \bot$ and $\gamma \vdash \gamma'$. If $\delta \prec_e \gamma$, then $\delta \prec_e \gamma'$.

(b) For any formulas γ, γ' and δ, if $\delta \prec_e \gamma$ and $\delta \prec_e \rho$, then $\delta \prec_e (\gamma \vee \rho)$.

Remark: A way of understanding iii(a) is as follows: Assume that $\rhd$ satisfies RLE_Σ and RA. Let γ_1 and γ_2 be two admissible formulas such that $\gamma_1 \vee \gamma_2$ is also admissible. Let $\gamma' = \gamma_1 \vee \gamma_2$. It is easy to check using 5.52 that $\gamma_i \not\prec_e \gamma'$ and $\gamma' \not\prec_e \gamma_i$, *i.e.* for each i, γ_i and γ' are $\prec_e$-incomparable. But in fact, 5.52 (i) (resp. iii(a)) says more, namely that every formula above (resp. below) γ_i is also above (resp. below) $\gamma_1 \vee \gamma_2$. So in some sense $\gamma_1 \vee \gamma_2$ contains the information "coded" by γ_1 and γ_2. Since explanatory relations are defined using $\prec$-minimal explanations it is clear the relevance of iii(a). ∎

The property iii(a) correspond to RA and we will denote this property by C-U (Continuing Up).To define it formally, we say that a binary relation $<$ over formulas satisfies C-U if the following holds:

C-U $$\forall \gamma, \gamma', \delta \ \ (\gamma \nvdash_\Sigma \bot \ \& \ \gamma \vdash \gamma' \ \& \ \delta < \gamma \ \Rightarrow \delta < \gamma')$$

Proposition 5.53 *If $\prec$ is a preference relation satisfying C-U then the explanatory relation associated with $\prec$ (defined in 5.40) satisfies RA.*

In the result that follows, it is interesting to notice that the hypothesis of logically finiteness is not needed. We will use this result in the sequel.

Proposition 5.54 *Let $\rhd$ be an E-rational explanatory relation satisfying E-Con$_\Sigma$. Then the following holds: for all admissible formulas γ and δ,*

$$\gamma \prec_u \delta \Leftrightarrow \exists\alpha\, \exists\beta \ [\, \alpha \rhd \gamma \ \& \ \beta \rhd \delta \ \& \ (\alpha \vee \beta) \rhd \gamma \ \& \ (\alpha \vee \beta) \not\rhd \delta \,] \tag{5.17}$$

Moreover, $\prec_u$ (and therefore $\prec_e$) is smooth and represents $\rhd$.

We will show next that when $\rhd$ satisfies E-R-Cut then $\prec_u$ is *modular*. We recall the definition of a modular relation (see [55]):

Definition 5.55 *A relation $\prec$ on a set E is said to be* modular *iff there exists a linear order $<$ on some set Ω and a function $r : E \to \Omega$ such that $a \prec b$ iff $r(a) < r(b)$. If $\prec$ is transitive, modularity is equivalent to the following property: for all a, b and c in E if a and b are $\prec$-incomparable and $a \prec c$ then $b \prec c$.*

Proposition 5.56 *Let $\rhd$ be an E-rational explanatory relation satisfying E-Con$_\Sigma$, E-R-Cut. Then $\prec_u$ is modular and satisfies C-U.*

There is a converse of the previous result:

Proposition 5.57 *Let $\prec$ be a smooth and modular preference relation satisfying C-U, then the associated explanatory relation $\rhd$ is E-rational and satisfies E-Con$_\Sigma$.*

From 5.56 and 5.57 we obtain the following

Theorem 5.58 *Let $\rhd$ be an explanatory relation, the following are equivalent:*

(i) The relation $\rhd$ is E-rational and satisfies E-Con$_\Sigma$.

(ii) There is a smooth and modular preference relation $\prec$ satisfying C-U such that for every α we have

$$\alpha \rhd \gamma \quad \textit{iff} \quad \gamma \in min(Expla(\alpha), \prec)$$

∎

5.5.2 Too many orders, too much disorder

In this section we will show some examples and see the connection of our results and the literature.

Since there are various explanatory, preference and consequence relations that have been used so far, it might be proper to see the relationship among them.

(i) Consider an injective preferential relation $\mid\!\sim$ satisfying Con$_\Sigma$. Suppose that $\mathcal{M} = \langle W, < \rangle$ is an injective standard preferential model defining $\mid\!\sim$ (standard means that for every α, $mod(C(\alpha)) = min(mod(\alpha), <)$). There are two natural explanatory relations associated with $\mid\!\sim$. One is $\tilde{\rhd}$ as defined in 5.24. The other one is given by a preference relation $<_\sim$ associated with $\mid\!\sim$ which is given by $\alpha <_\sim \beta$ if $\alpha \vee \beta \mid\!\sim \neg\beta$. It is known that $\alpha \mid\!\sim \beta$ iff $\alpha <_\sim \alpha \wedge \neg\beta$ (see [30]). Let $\rhd$ be the explanatory relation associated with $<_\sim$ as in 5.40. It is natural to ask whether $\rhd$ and $\tilde{\rhd}$ are equal. It is not hard to check that $\tilde{\rhd} \subseteq \rhd$. In fact, suppose $\alpha\tilde{\rhd}\gamma$ then $mod(\Sigma \cup \{\gamma\}) \subseteq mod(C(\alpha))$. We want to show that $\gamma \in min(Expla(\alpha), <_\sim)$. If $\delta <_\sim \gamma$ and $\delta \not\vdash_\Sigma \bot$ then there are models $M, M' \in W$ such that $M \in min(mod(\delta) \cap W, <)$, $M' \in min(mod(\gamma) \cap W, <)$ and $M < M'$. Since $mod(\Sigma \cup \{\gamma\}) \subseteq mod(C(\alpha))$ then $M' \in mod(C(\alpha))$ so $M' \in min(mod(\alpha), <)$ and therefore $M \models \neg\alpha$. Thus $\delta \notin Expla(\alpha)$. It follows that $\gamma \in min(Expla(\alpha), <_\sim)$. The other inclusion is not true in general (see the example that follows). However, for a finite language, if $\alpha \rhd \gamma$, then there is γ' such that $\alpha\tilde{\rhd}\gamma'$ and $C(\gamma) = C(\gamma')$: Namely, take γ' such that $mod(\gamma') = min(mod(\gamma))$.

(ii) On the other hand, we have two preference relations $<_\sim$ and $\prec_e$ (where this last one is the relation associated to $\tilde{\rhd}$). Since $\rhd$ and $\tilde{\rhd}$ are not necessarily the same, neither are $<_\sim$ and $\prec_e$.

Example: Suppose that the language is finite. Let $\mid\!\sim$ be defined by the following injective preferential model $\mathcal{M} = \langle \{M_1, M_2\}, \prec \rangle$ where $M_1 \prec M_2$. Let Σ be such that $mod(\Sigma) = \{M_1, M_2\}$, so $\mid\!\sim$ satisfies Con$_\Sigma$. Let γ be any formula such that $mod(\gamma) = \{M_1, M_2\}$. Then $\gamma \rhd \gamma$ because γ is minimal for $<_\sim$. But $\gamma \not\tilde{\rhd} \gamma$ because $C(\gamma) \not\subseteq Cn(\Sigma \cup \{\gamma\})$. However note that for γ' such that $mod(\gamma') = \{M_1\}$ we have $\gamma\tilde{\rhd}\gamma'$ and $C(\gamma) = C(\gamma')$. ∎

5.6 Related works

We will comment in this section about the connection of our results and the work of Flach [29], Cialdea-Pirri [66], Aliseda [2], Zadrozny [114], [65] and Freund [30].

1. Flach's work [29] is the one which is closer to ours. He presented some postulates for explanatory and inductive reasoning. Some of our postulates are similar to his. He studied the relations "γ is a possible inductive hypothesis given evidence α" and " γ is a possible explanation of α" (denoted by $\alpha \mid\!< \gamma$). Below we will compare his postulates with ours (we state them using our notation).

   I1: If $\alpha \rhd \gamma$ and $\alpha \wedge \gamma \vdash_\Sigma \beta$, then $(\alpha \wedge \beta) \rhd \gamma$. Since $\rhd$ is an explanatory relation then by hypothesis $\gamma \vdash_\Sigma \alpha$, hence it is not difficult to see that **I1** is, in our context, equivalent to E-CM.

   I2: If $\alpha \rhd \gamma$ and $\alpha \wedge \gamma \vdash_\Sigma \beta$, then $(\alpha \wedge \neg\beta) \not\rhd \gamma$. This holds trivially in our case because our explanations have to be consistent with Σ.

   I3: If $\alpha \rhd \gamma$ and $\alpha \wedge \gamma \vdash_\Sigma \beta$, then $\alpha \rhd \gamma \wedge \beta$. This is a consequence of RLE_Σ and 5.2.
   He considered two versions of Reflexivity,

   I4: If $\alpha \rhd \gamma$, then $\alpha \rhd \alpha$. This will not be valid in our case, because α might not be a *preferred* explanation of itself.

   I5: If $\alpha \rhd \gamma$, then $\gamma \rhd \gamma$. We already have mentioned that this holds for $\rhd$, when it is defined as in 5.40 and also if $\rhd$ satisfies E-CM and E-C-Cut.

   The other two postulates for induction **I6** and **I7** correspond to RLE_Σ and LLE_Σ respectively. There are other postulates studied by Flach that he considers more specific of explanatory relations.

   E1: If $\alpha \rhd \delta$, $\gamma \rhd \gamma$ and $\vdash \gamma \rightarrow \delta$, then $\alpha \rhd \gamma$. This is essentially our RA,except that we do not require that $\gamma \rhd \gamma$.

   E2: If $\gamma \rhd \gamma$ and $\neg\alpha \not\rhd \gamma$, then $\alpha \rhd \alpha$. This does not necessarily hold in our case. In our context, this rule is quite strange because it says that when a formula α is not a good explanation for itself then any good explanation is a good explanation for the negation of α. This seems to be true only when $\mid\!\sim_{ab}$ is monotone.

   E3: If $\alpha \rhd (\beta \wedge \gamma)$, then $(\beta \rightarrow \alpha) \rhd \gamma$. This is not necessarily true for $\tilde{\rhd}$ even when $\mid\!\sim$ is rational.

   E4: If $\alpha \rhd \gamma$ and $\beta \rhd \gamma$, then $(\alpha \wedge \beta) \rhd \gamma$. This is a consequence of E-CM.

   E5: If $\alpha \rhd \gamma$ and $\alpha \vdash_\Sigma \beta$, then $\beta \rhd \gamma$. This implies E-Cut and in fact, it is equivalent to E-Cut under the presence of E-CM.

   He then presented five postulates for "confirmatory induction" which does not seem applicable for explanatory reasoning, except for his postulate C4 which corresponds to our LOR. For Flach "intuition constitutes the primary source of justification for his rationality postulates". Our results confirm that his intuition also has a formal justification. The more important difference is that he did not consider weaker cut rules than E-Cut thus his postulates force $\mid\!\sim_{ab}$ to be monotonic. But moreover, his explanatory relation is restricted to reversed deduction: $\alpha \mid\!< \gamma$ iff $\gamma \vdash_\Sigma \alpha$. This is his main representation theorem for explanatory relations.

2. Cialcea and Pirri [66] defined a relation $\Sigma \vdash \gamma \rightsquigarrow \alpha$ to capture the notion that "in the theory Σ, γ is a good reason for α". The definition of $\rightsquigarrow$ is based on a preference relation over formulas like 5.40. They presented some basic postulates and some conditions where they hold. Our Postulate E-CM is stronger than their **And-Right**. Their Left Logical Equivalence is our RLE_Σ. Our Cut rules (E-C-Cut, E-R-Cut and E-Cut) has nothing to do with their E-Cut. Here there is an important difference between our approach and theirs. As we said in the introduction, our postulates are concerned with a fixed background theory Σ, but they considered postulates concerning properties of abduction when the background theory changes. For instance, their E-Cut rule says

$$\text{If } \Sigma \vdash \alpha \text{ and } \Sigma \cup \{\alpha\} \vdash \gamma \rightsquigarrow \beta, \text{ then } \Sigma \vdash \gamma \rightsquigarrow \beta$$

and their E-Monotonicity rule says

$$\text{If } \Sigma \vdash \alpha \text{ and } \Sigma \vdash \gamma \rightsquigarrow \beta, \text{ then } \Sigma \cup \{\alpha\} \vdash \gamma \rightsquigarrow \beta$$

These last two postulates are very weak, since they are valid for $\rhd$ regardless of which preference relation $\preceq$ is used (when $\rhd$ is defined as in 5.40). Another difference with our paper is that they did not address the problem of whether their postulates will guarantee that $\rightsquigarrow$ is given by a preference relation.

3. Aliseda's Ph.D thesis is a comprehensive presentation of abduction from several points of view. It is a very good source for the vast literature on abduction. We will make some comments only about the part of her work which is close related to our paper. Similar to Cialcea and Pirri's approach, Aliseda regards abduction as a relation with three parameters: a background theory, an observation and an explanation. Her notation is $\Sigma \mid \gamma \Rightarrow \alpha$ to express that γ is an explanation for α w.r.t Σ. She presented sets of rules for formalizing the various versions of abduction she introduced: Plain, Consistent, Explanatory, Minimal and Preferential abduction. Some of her postulates are not valid in our context, for instance her Weak Explanatory Reflexivity says

$$\text{If } \Sigma \mid \gamma \Rightarrow \alpha, \text{ then } \Sigma \mid \alpha \Rightarrow \alpha$$

Which is Flach **I4** and as we already said it is not valid in our context because in most cases an observation is not a preferred explanation of itself. She also consider cut and monotonicity rules, similar to those used by Cialcea and Pirri. However, no cut rule for observations (as ours) was studied, except the rule of transitivity (which follows from RA.) This type of rules are missing in all papers we have cited. Preferential abduction is among all versions she considered the closest to our approach. It naturally requires that γ has to be minimal with respect to a preference relation among formula. The crucial rule for formalizing Preferential abduction is not viewed as structural rule since it adds a preference relation that she thought cannot be expressed in terms of the inference relation itself. But we have shown that preference criteria can be coded by the structural rules without explicitly mention them.

4. We will comment next about [65]. In logic-based abduction usually together with the background theory Σ there is also a distinguished set of atoms Ab called *abducibles* and explanations have to be built using only atoms in Ab. The pair (Σ, Ab) is referred to as the *Abductive framework*. In [65] were studied consequence relations based on an abductive framework. They defined a notion of abducible explanation and used an order among them to select the preferred ones. They defined abductive rational consequence relations, denoted also by $\mid\!\sim_{ab}$, based on this notion of preferred abducible explanation.

These relations are very closed to the rational relations studied by Lehmann and Magidor in [55]. The definition of our $\mid\!\sim_{ab}$ was motivated by their definition. Abductive consequence relations can also be represented by explanatory relations, but for this case the explanatory relation is between a formula and an abducible formula (*i.e.* those formulas built using only atoms in *Ab*) and the corresponding preference relation will be over the set of abducible formulas. There are some extra difficulties that one has to deal with in this more general setting of abduction. For instance, even for the Ab-rational case studied in [65], the rule E-Disj fails. This is due to the fact that explanations have to be built using only abducible atoms, thus a preferred explanation of $\alpha \vee \beta$ is not necessarily a preferred explanation of neither α nor β. For certain abductive consequence relations $\mid\!\sim$, it can be shown that if γ is an abducible formula and $\alpha \tilde{\rhd} \gamma$, then γ is also a preferred explanation of α in the sense of [65]. The role of abducibles formulas in this paper is quite closed to our admissible formulas. It would be interesting to clarify this analogy.

5. Zadrozny [114] approached abduction from a quite abstract point of view based on the concept of invariant of reasoning. Abduction is viewed as an inference process that preserves sets of explanations. It is not clear the relation with our results, but it seems an interesting topic of research. He has some rules similar to ours but his presentations is quite complex. His explanation systems are formulated using higher-order logic as a metalanguage.

6. Possibility and Expectation orders. We will comment next about [30]. In that article Freund characterize preferential consequence relation in terms of 'preferential orders'. He called *preferential order* any relation $<$ on formulae satisfying the following four properties:
   P_0: $\alpha < \bot$
   P_1: If $\alpha \vdash \beta$, then (a) $\alpha < \gamma \Rightarrow \beta < \gamma$
   (b) $\delta < \beta \Rightarrow \delta < \alpha$
   P_2: $\alpha < \gamma$ and $\alpha < \delta$ implies $\alpha < \gamma \vee \delta$
   P_3: $\alpha \vee \beta < \beta$ implies $\alpha < \beta$

   Now observe that $\prec_e$ satisfies P_0 when the formula α is admissible and, except for P_1(b), the others properties are also satisfied by $\prec_e$ (this follows from 5.52). However from 5.52 and 5.53 it is clear that C-U and P_1(b) play dual roles.

5.7 Conclusions

We have analyzed three aspects of explanatory reasoning: Its logical properties, its relation with reversed deduction and the role of selection criteria. The logical properties have been isolated in a fairly complete list of postulates. Our list extends and confirm the intuition of previous approaches (Flach, Cialdea-Pirri and Aliseda). The key idea was to use $\mid\!\sim_{ab}$ as an heuristic device for isolating the logical properties of an explanatory relation $\rhd$. It is important to point out the special role that explanatory cut rules play in our presentation. We have not seen these rules in other formalism and they are the key feature for encoding preference criteria.

When we started this research we were focused on getting $\mid\!\sim_{ab}$ to be nice in the KLM sense. Moreover, we thought that an explanatory relation $\rhd$ and its associate consequence relation $\mid\!\sim_{ab}$ were somewhat interchangeable. But this turns out to be true only for those explanatory relations that we have called *causal*. Our results gave an answer to one of our initial question: Causal explanatory reasoning is non-monotonic reasoning in reverse.

Selection mechanisms are a fundamental part of abduction. However, most formalism have treated them as external devices which work on top of the logical part of abduction. We have shown that preference criteria are built in the structural properties of explanatory relations.

Causal explanatory relations have also a interpretation in terms of belief revision. The key feature that distinguish causal explanations from other notions of explanations is the fact that causal explanatory relations treat observations and explanations in a different way. Observations are viewed as beliefs and therefore as something defeasible. On the other hand, explanations are not treated as beliefs and thus the logical connection between an observation and its preferred explanations is retained in a very strong form. The underlying idea of causal explanatory relations is the following. After observing α, we collect the concomitant facts that are normally present (i.e. we compute $C(\alpha)$). Then a preferred causal explanation of α is *any* formula that entails α and its usual consequences. In other words, rather than saying that γ normally implies α we say that γ implies everything that α normally implies.

It turns out that there are explanatory relations for which $\mid\!\sim_{ab}$ does not give any information at all, however, they can be classified using our postulates. So the set of postulates for explanatory relations coming out of our analysis are more general than we thought at the beginning. It is well known that non-monotonic consequence relations, revision operators and preference orders (like expectation, posibilistic, preferential orders) are essentially the same formal objects. We can say, as a summary, that explanatory relations can be added to this list of tools for studying common sense reasoning.

Finally, we will mention two possible lines of research related to our results. The first one is to study more carefully the hierarchy we have presented for classifying the logical properties of abduction. Specially relevant is to determine up to which extend this hierarchy classifies (non causal) weak explanatory relations. The second one is related to the role of the background theory. Usually it is said there are three kinds of reasoning processes: deductive, abductive and inductive. We have shown that abduction is very tightly related to a "non-monotonic-deduction". On the other hand, inductive reasoning (when it is understood as the process of inferring general rules out of specific observations) did not play any role in our setting. This is probably due to the fact that we have fixed the background theory. There are many situations where Σ is the natural outcome of an inductive reasoning process. As we said in the introduction, Cialdea-Pirri and Aliseda presented a view of abduction as a relation with three parameters: an observation, an explanation and a background theory. We think that an extension of our results to allow the change of the background theory will provide some hints for a better understanding of inductive reasoning.

Acknowledgments: We would like to thank Jorge Lobo, M. Freund, D. Lehmann and the anonymous referees for their comments and criticism which have improved the quality of the paper.

Partial support for Carlos Uzcátegui was provided by a CDCHT-ULA (Venezuela) grant. This work was initiated in 1996 while he was visiting the Laboratoire d'Informatique Fondamentale de Lille (LIFL), France. He would like to thank LIFL for the financial assistance and facilities they provided.

5.8 The proofs of results of this chapter

Proof of theorem 5.4: Suppose $\rhd$ is a relation as in the hypothesis. We will show that $\mid\!\sim_{ab}$ is cumulative. From $\mathsf{LLE}_\Sigma$for $\rhd$ we easily get that $\mid\!\sim_{ab}$ satisfies Left Logical Equivalence and from the definition of $\mid\!\sim_{ab}$ (5.2) it is obvious that Reflexivity and RW holds. It remains to be checked the rules Cut and Cautious Monotony.

Let's suppose that $\alpha\mid\!\sim_{ab}\beta$, then the second condition in the rule E-C-Cut is satisfied, *i.e.* $\forall\delta\,[\,\alpha\rhd\delta\;\Rightarrow\vdash_\Sigma\delta\rightarrow\beta\,]$. Therefore from E-C-Cut and E-CM we easily conclude $\{\gamma:\alpha\rhd\gamma\}=\{\gamma:(\alpha\wedge\beta)\rhd\gamma\}$ and hence $C(\alpha\wedge\beta)=C(\alpha)$ (where as usual for a fixed consequence relation $\mid\!\sim$ and any formula δ, $C(\delta)$ is the set $\{\theta:\delta\mid\!\sim\theta\}$). That is to say, $\mid\!\sim_{ab}$ satisfies Cut and Cautious Monotony. □

Proof of proposition 5.5: (i) That RLE_Σ holds is straightforward. To see that ROR holds, suppose that $\alpha\rhd(\gamma\vee\rho)$. First note that either $(\gamma\vee\rho)\not\vdash_\Sigma\neg\gamma$ or $(\gamma\vee\rho)\not\vdash_\Sigma\neg\rho$, otherwise $(\gamma\vee\rho)\vdash_\Sigma(\neg\gamma\wedge\neg\rho)$ so $(\gamma\vee\rho)\vdash_\Sigma\bot$, which is a contradiction since $\rhd$ is an explanatory relation. Therefore by RA either $\alpha\rhd(\gamma\vee\rho)\wedge\gamma$ or $\alpha\rhd(\gamma\vee\rho)\wedge\rho$ so by RLE_Σ either $\alpha\rhd\gamma$ or $\alpha\rhd\rho$.

(ii) and (iii) are straightforward.

(iv) The proof that RA implies E-Disj is as in (i) above.Conversely suppose that $\rhd$ satisfies E-Disj and RLE_Σ, we want to show that RA holds. Let α, γ and γ' be such that $\alpha\rhd\gamma$, $\gamma'\vdash_\Sigma\gamma$ and $\gamma'\not\vdash_\Sigma\bot$. Since $\gamma'\vdash_\Sigma\gamma$, we have $\vdash_\Sigma\gamma\leftrightarrow(\gamma'\vee\gamma)$ so by RLE_Σ $\alpha\rhd(\gamma'\vee\gamma)$. Since by hypothesis $\gamma'\not\vdash_\Sigma\bot$ then by E-Disj we have $\alpha\rhd\gamma'$. ∎

Proof of proposition 5.7: Suppose $(\alpha_1\vee\alpha_2)\rhd\gamma$ and $\alpha_i\not\rhd\gamma$ for $i=1,2$. We claim that $\gamma\not\vdash_\Sigma\alpha_i$ for $i=1,2$. Otherwise by E-CM we have for some $i\in\{1,2\}$, $(\alpha_1\vee\alpha_2)\wedge\alpha_i\rhd\gamma$ and therefore by LLE we conclude $\alpha_i\rhd\gamma$ which is a contradiction. Let $\gamma_i=\gamma\wedge\alpha_i$. Since $\gamma\vdash_\Sigma(\alpha_1\vee\alpha_2)$, then it is clear that γ is equivalent modulo Σ to $\gamma_1\vee\gamma_2$. On the other hand, $\gamma_i\vdash_\Sigma\alpha_i$ and $\gamma_i\not\vdash_\Sigma\bot$ for $i=1,2$ (otherwise $\gamma\vdash_\Sigma\alpha_i$ for some i). Finally by RA we have that $(\alpha_1\vee\alpha_2)\rhd\gamma_i$ and by E-CM and LLE_Σ we conclude $\alpha_i\rhd\gamma_i$ for $i=1,2$. □

Proofof theorem 5.9: We already know from 5.4 that $\mid\!\sim_{ab}$ is cumulative, so it remains to be shown that $\mid\!\sim_{ab}$ satisfies the rule Or. Let's suppose that $\alpha\mid\!\sim_{ab}\rho$ and $\beta\mid\!\sim_{ab}\rho$, we will show that $\alpha\vee\beta\mid\!\sim_{ab}\rho$. Let γ be such that $(\alpha\vee\beta)\rhd\gamma$, we have to show that $\gamma\vdash_\Sigma\rho$. By proposition 5.7 we have to consider three cases: (a) $\alpha\rhd\gamma$. Since $\alpha\mid\!\sim_{ab}\rho$ then we have $\gamma\vdash_\Sigma\rho$. (b) $\beta\rhd\gamma$. We conclude that $\gamma\vdash_\Sigma\rho$ as in the first case. (c) There are γ_1 and γ_2 such that $\vdash_\Sigma\gamma\leftrightarrow(\gamma_1\vee\gamma_2)$ with $\alpha\rhd\gamma_1$ $\beta\rhd\gamma_2$. Then, by hypotheses we have $\gamma_i\vdash_\Sigma\rho$ for $i=1,2$. Since $\vdash_\Sigma\gamma\leftrightarrow(\gamma_1\vee\gamma_2)$ we conclude $\gamma\vdash_\Sigma\rho$. □

Proof of theorem 5.10: By 5.9 we know that $\mid\!\sim_{ab}$ is preferential. So it remains to be shown that $\mid\!\sim_{ab}$ satisfies W-DR. We define an auxiliary function F that maps formulas into formulas as follows: $F(\alpha)=\bigvee\{\gamma:\alpha\rhd\gamma\}$ in case there is γ such that $\alpha\rhd\gamma$, otherwise we let $F(\alpha)=\bot$. Notice that $\alpha\mid\!\sim_{ab}\beta$ iff $F(\alpha)\vdash_\Sigma\beta$. To see that W-DR holds it clearly suffices to show that $F(\alpha)\wedge F(\beta)\vdash_\Sigma F(\alpha\vee\beta)$. Let $\alpha\rhd\gamma$ and $\beta\rhd\delta$, it is enough to verify that when $\gamma\wedge\delta$ is consistent with Σ, then $(\alpha\vee\beta)\rhd(\gamma\wedge\delta)$. Since $\not\vdash_\Sigma\gamma\rightarrow\neg\delta$, from RA we easily conclude $\alpha\rhd(\gamma\wedge\delta)$ and $\beta\rhd(\gamma\wedge\delta)$, therefore from LOR we obtain $(\alpha\vee\beta)\rhd(\gamma\wedge\delta)$. □

Proof of proposition 5.11: It is clear that E-DR implies LOR. To check that E-DR implies that $\mid\!\sim_{ab}$ satisfies DR, suppose that $\alpha\vee\beta\mid\!\sim_{ab}\rho$ and $\alpha\not\mid\!\sim_{ab}\rho$. We have to show that $\beta\mid\!\sim_{ab}\rho$. Let δ be such that $\beta\rhd\delta$, it suffices to check that $\delta\vdash_\Sigma\rho$. Since $\alpha\not\mid\!\sim_{ab}\rho$, then there is γ such that

$\alpha \rhd \gamma$ and $\gamma \not\vdash_\Sigma \rho$. By E-DR either $(\alpha \vee \beta) \rhd \gamma$ or $(\alpha \vee \beta) \rhd \delta$. Since $\alpha \vee \beta \mid\!\sim_{ab} \rho$ and $\gamma \not\vdash_\Sigma \rho$, we conclude $(\alpha \vee \beta) \rhd \delta$. Therefore $\delta \vdash_\Sigma \rho$. ∎

Proof of theorem 5.13: By 5.9 $\mid\!\sim_{ab}$ is preferential. Thus it suffices to show that $\mid\!\sim_{ab}$ satisfies Rational Monotony. Let α, β and ρ be formulas such that $\alpha \mid\!\sim_{ab} \rho$ and $\alpha \not\mid\!\sim_{ab} \neg\beta$. Let γ be such that $(\alpha \wedge \beta) \rhd \gamma$, we want to show that $\gamma \vdash_\Sigma \rho$. Since $\alpha \not\mid\!\sim_{ab} \neg\beta$, then by definition of $\mid\!\sim_{ab}$ there is δ such that $\alpha \rhd \delta$ and $\delta \not\vdash_\Sigma \neg\beta$. By RA (see 5.5.iii)there is $\delta' \vdash_\Sigma \delta$ such that $\alpha \rhd \delta'$ and $\delta' \vdash_\Sigma \beta$. Therefore by E-R-Cut we conclude that $\alpha \rhd \gamma$. Since $\alpha \mid\!\sim_{ab} \rho$, then $\gamma \vdash_\Sigma \rho$. □

Proof of proposition 5.14: We will consider three cases.

(Case 1) Suppose that $(\alpha \vee \beta) \mid\!\sim_{ab} \neg\alpha$. In particular, we have that for all $(\alpha \vee \beta) \rhd \gamma$, $\gamma \vdash_\Sigma \beta$. We will show that (b) holds. Let γ be such that $(\alpha \vee \beta) \rhd \gamma$. Then by our hypothesis $\gamma \vdash_\Sigma \beta$. By E-CM $(\alpha \vee \beta) \wedge \beta \rhd \gamma$ and by LLE $\beta \rhd \gamma$. On the other hand, let γ be such that $\beta \rhd \gamma$, then by E-CM $(\alpha \vee \beta) \wedge \beta \rhd \gamma$. Since $(\alpha \vee \beta) \mid\!\sim_{ab} \neg\alpha$, then it follows from E-C-Cut that $(\alpha \vee \beta) \rhd \gamma$.

(Case 2) Suppose that $(\alpha \vee \beta) \mid\!\sim_{ab} \neg\beta$. Then as in case 1 it follows that (a) holds.

(Case 3) Suppose that $(\alpha \vee \beta) \not\mid\!\sim_{ab} \neg\alpha$ and $(\alpha \vee \beta) \not\mid\!\sim_{ab} \neg\beta$. We will show that (c) holds. By 5.7 it suffices to show that $\{\gamma : \alpha \rhd \gamma\} \cup \{\gamma : \beta \rhd \gamma\} \subseteq \{\gamma : (\alpha \vee \beta) \rhd \gamma\}$ By hypothesis there is γ' such that $(\alpha \vee \beta) \rhd \gamma'$ and $\gamma' \not\vdash_\Sigma \neg\alpha$. By RA we can assume that $\gamma' \vdash_\Sigma \alpha$. Let γ be such that $\alpha \rhd \gamma$, then by E-CM $(\alpha \vee \beta) \wedge \alpha \rhd \gamma$. Using γ' and E-R-Cut we conclude that $(\alpha \vee \beta) \rhd \gamma$. It can be shown analogously that if $\beta \rhd \gamma$, then $(\alpha \vee \beta) \rhd \gamma$. □

Proof of proposition 5.15: Suppose $\alpha \rhd \gamma$ and $\beta \rhd \delta$ and $(\alpha \vee \beta) \not\rhd \delta$, we want to show that $(\alpha \vee \beta) \rhd \gamma$. Since $\vdash \beta \leftrightarrow (\alpha \vee \beta) \wedge \beta$ and $\beta \rhd \delta$ then it follows from E-R-Cut that for all γ' if $(\alpha \vee \beta) \rhd \gamma'$, then $\gamma' \not\vdash_\Sigma \beta$. Since $\alpha \rhd \gamma$ we have $(\alpha \vee \beta) \wedge \alpha \rhd \gamma$. Suppose, towards a contradiction, that $(\alpha \vee \beta) \not\rhd \gamma$. By E-C-Cut there is γ' such that $(\alpha \vee \beta) \rhd \gamma'$ and $\gamma' \not\vdash_\Sigma \alpha$. By RA (5.5) there is γ'' such that $(\alpha \vee \beta) \rhd \gamma''$ and $\gamma'' \vdash_\Sigma \alpha$. Finally, since $\vdash \alpha \leftrightarrow (\alpha \vee \beta) \wedge \alpha$ and $\alpha \rhd \gamma$, then by E-R-Cut we conclude that $(\alpha \vee \beta) \rhd \gamma$ a contradiction. □

Proof of theorem 5.16: Suppose $\alpha \mid\!\sim_{ab} \rho$, *i.e.* for all γ if $\alpha \rhd \gamma$ then $\gamma \vdash_\Sigma \rho$. Let δ be any formula such that $(\alpha \wedge \beta) \rhd \delta$. By E-Cut we have $\alpha \rhd \delta$ so $\delta \vdash_\Sigma \rho$. Thus $(\alpha \wedge \beta) \mid\!\sim_{ab} \rho$. □

Proof of proposition 5.19: Suppose that $(\alpha \wedge \beta) \rhd \gamma$ and also that $\delta \vdash_\Sigma \beta$ for all δ such that $\alpha \rhd \delta$. It suffices to show that there is δ such that $\alpha \rhd \delta$. Since $(\alpha \wedge \beta) \rhd \gamma$ then (by the definition of an explanatory relation) α is consistent with Σ, therefore by E-Con$_\Sigma$ there is δ such that $\alpha \rhd \delta$. ∎

Proof of proposition 5.22: It is obvious that if $\alpha \vdash_\Sigma \beta$ then $\alpha \mid\!\sim_{ab} \beta$. On the other hand, if $\alpha \not\vdash_\Sigma \beta$, then $\alpha \wedge \neg\beta \not\vdash_\Sigma \bot$. Thus by E-Con$_\Sigma$ there is γ such that $(\alpha \wedge \neg\beta) \rhd \gamma$. Therefore by E-Cut $\alpha \rhd \gamma$, hence $\alpha \not\mid\!\sim_{ab} \beta$. □

Proof of proposition 5.25: Let $\tilde{\rhd}$ be the explanatory relation associated with $\mid\!\sim$ and let $\mid\!\sim_{ab}$ be the consequence relation associated with $\tilde{\rhd}$. We will show that $\mid\!\sim$ is equal to $\mid\!\sim_{ab}$. By definition of $\mid\!\sim_{ab}$ and the hypothesis that $\mid\!\sim$ is adequate we have

$$\begin{aligned} C_{ab}(\alpha) &= \bigcap\{Cn(\Sigma \cup \{\gamma\} : \alpha \tilde{\rhd} \gamma\} \\ &= \bigcap\{Cn(\Sigma \cup \{\gamma\} : C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\} \;\&\; \gamma \not\vdash_\Sigma \bot\} \\ &= C(\alpha) \end{aligned}$$

Observe that these equalities are valid even in the case that there is no γ such that $\alpha \tilde{\rhd} \gamma$ (equivalently, when $C(\alpha)$ contains all formulas). ∎

Proof of theorem 5.26:

1. It is obvious from the definition of $\tilde{\rhd}$ that it satisfies RA and RLE_Σ.

2. It is obvious that if $\mid\!\sim$ satisfies LLE then $\tilde{\rhd}$ satisfies LLE_Σ.

3. Suppose that $\mid\!\sim$ satisfies Con_Σ, then $\tilde{\rhd}$ satisfies E-Con_Σ follows easily from the hypothesis that $\mid\!\sim$ is adequate.

4. Suppose that $\mid\!\sim$ satisfies CM. To see that $\tilde{\rhd}$ satisfies **E-C-Cut** let us suppose that $(\alpha \wedge \beta)\tilde{\rhd}\gamma$ and also that $\delta \vdash_\Sigma \beta$ for all δ such that $\alpha\tilde{\rhd}\delta$. We have to show that $\alpha\tilde{\rhd}\gamma$. Suppose $\alpha \mid\!\sim \rho$, it suffices to show that $\gamma \vdash_\Sigma \rho$. Since $\mid\!\sim$ is adequate, from the second part of the hypothesis of E-C-Cut we conclude that $\alpha \mid\!\sim \beta$. Therefore by CM we have $C(\alpha) \subseteq C(\alpha \wedge \beta)$, since $C(\alpha \wedge \beta) \subseteq Cn\ (\Sigma \cup \{\gamma\})$, then the result follows.

5. Suppose that $\mid\!\sim$ satisfies the S-rule. To see that $\tilde{\rhd}$ satisfies E-CM let $\alpha\tilde{\rhd}\gamma$ and $\gamma \vdash_\Sigma \beta$. We want to show that $(\alpha \wedge \beta)\tilde{\rhd}\gamma$. Since γ is consistent with Σ, it suffices to show that $C(\alpha \wedge \beta) \subseteq Cn\ (\Sigma \cup \{\gamma\})$. Let $\alpha \wedge \beta \mid\!\sim \rho$, then by the S-rule $\alpha \mid\!\sim \beta \rightarrow \rho$. Since $\alpha\tilde{\rhd}\gamma$, then $\gamma \vdash_\Sigma \beta \rightarrow \rho$. Hence $\gamma \vdash_\Sigma \rho$.

6. Suppose $\mid\!\sim$ satisfies W-DR. We will show that $\tilde{\rhd}$ satisfies LOR. Suppose $\alpha\tilde{\rhd}\gamma$ and $\beta\tilde{\rhd}\gamma$. By W-DR we have that $C(\alpha \vee \beta) \subseteq Cn\ (C(\alpha) \cup C(\beta))$. Then it is clear that $(\alpha \vee \beta)\tilde{\rhd}\gamma$.

7. Suppose $\mid\!\sim$ is preferential and satisfies DR. We will show that $\tilde{\rhd}$ satisfies E-DR. Suppose $\alpha\tilde{\rhd}\gamma$, $\beta\tilde{\rhd}\rho$ and $(\alpha \vee \beta)\ \not\tilde{\rhd}\gamma$. Then there is δ such that $\alpha \vee \beta \mid\!\sim \delta$ and $\gamma \not\vdash_\Sigma \delta$. Since $\alpha\tilde{\rhd}\gamma$ and $C(\alpha \vee \beta) \subseteq C(\alpha) \cup C(\beta)$ we have $\delta \in C(\beta)$. Now consider any δ' such that $\delta' \in C(\alpha \vee \beta)$. We want to show that $\rho \vdash_\Sigma \delta'$. By preferentiality $\delta \wedge \delta' \in C(\alpha \vee \beta)$. But $\delta \wedge \delta' \notin C(\alpha)$, otherwise $\gamma \vdash_\Sigma \delta \wedge \delta'$ and therefore $\gamma \vdash_\Sigma \delta$ which is a contradiction. Then by DR $\delta \wedge \delta' \in C(\beta)$. Hence $\rho \vdash_\Sigma \delta \wedge \delta'$ and thus $\rho \vdash_\Sigma \delta'$.

8. Suppose $\mid\!\sim$ satisfies RM. We will show that $\tilde{\rhd}$ satisfies **E-R-Cut**. Suppose $(\alpha \wedge \beta)\tilde{\rhd}\gamma$ and there is δ such that $\alpha\tilde{\rhd}\delta$ with $\delta \vdash_\Sigma \beta$. From the last assumption and the definition of $\tilde{\rhd}$ we conclude that $\alpha \not\mid\!\sim \neg\beta$. Therefore by RM we have $C(\alpha) \subseteq C(\alpha \wedge \beta)$, and the result follows.

9. Suppose that $\mid\!\sim$ is monotone. Since $\mid\!\sim$ is monotone, then $C(\alpha) \subseteq C(\alpha \wedge \beta)$. Therefore, if $(\alpha \wedge \beta)\tilde{\rhd}\gamma$ then $\alpha\tilde{\rhd}\gamma$. This says that $\tilde{\rhd}$ satisfies E-Cut.

∎

Proof pf proposition 5.27: For every α, let $F(\alpha)$ be a formula such that $C(\alpha) = Cn(F(\alpha))$. Let us define $\rhd$ as follows: $\alpha \rhd \gamma$ if $\gamma \not\vdash_\Sigma \bot$ and $\gamma \equiv F(\alpha)$. It is obvious that $\rhd$ is indeed an explanatory relation satisfying RLE_Σ. Let $\mid\!\sim_{ab}$ be the consequence relation associate with $\rhd$. It is easy to see that $\mid\!\sim$ is equal to $\mid\!\sim_{ab}$. Now we will check the other postulates. It follows that LLE_Σ (for $\rhd$) follows from LLE for $\mid\!\sim$. To see that E-CM holds, suppose $\alpha \rhd \gamma$ and $\gamma \vdash_\Sigma \beta$. We need to show that $(\alpha \wedge \beta) \rhd \gamma$. By hypothesis $F(\alpha) \vdash_\Sigma \beta$, then it follows that $\alpha \mid\!\sim \beta$. Since $\mid\!\sim$ is cumulative, then $C(\alpha) = C(\alpha \wedge \beta)$. From this it follows that $F(\alpha) \equiv F(\alpha \wedge \beta)$ and therefore $(\alpha \wedge \beta) \rhd \gamma$. The proof that **E-C-Cut** holds is similar. ∎

Proof of theorem 5.30: $((i) \Rightarrow (ii))$. It is obvious that any causal relation satisfies RA. To check the causal axiom let α and γ be two formulas consistent with Σ. Suppose that for all δ consistent with Σ such that $\delta \vdash_\Sigma \gamma$ there is ρ such that $\rho \vdash_\Sigma \delta$ and $\alpha \rhd \rho$. We want to show that $C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})$. Let $\alpha \mid\!\sim_{ab} \beta$ and suppose toward a contradiction that $\gamma \not\vdash_\Sigma \beta$. There is δ consistent with Σ such that $\delta \vdash_\Sigma \gamma$ and $\delta \vdash_\Sigma \neg\beta$. Let ρ be such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \delta$.

Therefore $\rho \vdash_\Sigma \neg\beta$. On the other hand, since $\alpha \mid\!\sim_{ab} \beta$, then $\rho \vdash_\Sigma \beta$. Thus $\rho \vdash_\Sigma \bot$ which is a contradiction.

($(ii) \Rightarrow (i)$). Suppose that $\rhd$ satisfies RA and C. It suffices to show that if $\gamma \not\vdash_\Sigma \bot$ and $C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})$, then $\alpha \rhd \gamma$. Let δ be any formula consistent with Σ such that $\delta \vdash_\Sigma \gamma$. Then there must exist ρ such that $\alpha \rhd \rho$ and $\rho \not\vdash_\Sigma \neg\delta$ (otherwise $\neg\delta \in C_{ab}(\alpha)$ which is not possible). Let ρ' be a formula consistent with Σ such that $\rho' \vdash_\Sigma \rho \wedge \delta$. By RA we conclude that $\alpha \rhd \rho'$. Therefore by C we get that $\alpha \rhd \gamma$. ∎

Proof of lemma 5.33: It is straightforward that (i) implies (ii). For the other direction, notice that (ii) implies that $\bigcap_{\alpha \in A} Cn(\Sigma \cup \{\alpha\}) \cup \{\neg\alpha : \alpha \in A\}$ is inconsistent. Therefore by compactness, there are $\alpha_1, \dots, \alpha_n \in A$, such that $(\alpha_1 \vee \cdots \vee \alpha_n) \in \bigcap_{\alpha \in A} Cn(\Sigma \cup \{\alpha\})$. From this (i) easily follows. □

Proof of proposition 5.34: It suffices to show that $\rhd$ satisfies **CA**. Let $\gamma_1, \dots, \gamma_k$ be an upper bound for $\{\gamma : \alpha \rhd \gamma\}$. Let $\theta = \gamma_1 \vee \cdots \vee \gamma_k$. By E-RW we have that $\alpha \rhd \theta$. Let α and γ be formulas consistent with Σ. Suppose that for all δ such that $\delta \not\vdash_\Sigma \bot$ and $\delta \vdash_\Sigma \gamma$ there is ρ such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \delta$. We want to show that $\alpha \rhd \gamma$. Suppose that $\alpha \not\rhd \gamma$ towards a contradiction. Then by RA we have that $\gamma \not\vdash_\Sigma \theta$. Therefore there is δ consistent with Σ such that $\delta \vdash_\Sigma \gamma \wedge \neg\theta$ and by hypothesis there is ρ such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \delta$. Thus $\rho \vdash_\Sigma \neg\theta$ which contradicts that θ is an upper bound. ∎

Proof of proposition 5.38: Since $*$ preserves consistency, then it is clear that $\mid\!<$ is a weak explanatory relation (as defined in 5.37). It is obvious that LLE_Σ, RLE_Σ, E-Cut and full reflexivity holds. Notice that LOR follows from E-Cut and LLE_Σ.To check E-CM, assume that $\alpha \mid\!< \gamma$ and also that $\gamma \vdash \beta$. Then $\gamma, \alpha \in K * \gamma$. Thus $\alpha \wedge \beta \in K * \gamma$. Finally, E-RW follows from the Or rule for $\mid\!\sim_K$ and ROR follows from DR for $\mid\!\sim_K$. ∎

Proof of proposition 5.39: From Con_Σ it follows that $\mid\!<$ is a weak explanatory relation. (i) It is clear that $\mid\!<$ is transitive, reflexive for Σ-consistent formulasand satisfies LLE_Σ, RLE_Σ. E-RW follows easily from the Or rule. To check E-CM, let us assume that $\alpha \mid\!< \gamma$ and $\gamma \vdash_\Sigma \beta$. Let $\alpha \wedge \beta \mid\!\sim \rho$, then by the S-rule we have that $\alpha \mid\!\sim \beta \to \rho$. By hypothesis $C(\alpha) \subseteq C(\gamma)$, thus $\gamma \mid\!\sim \beta \to \rho$. By preferentiality and Con_Σ from $\gamma \vdash_\Sigma \beta$ is easy to obtain $\gamma \mid\!\sim \beta$. Hence by **RW** $\gamma \mid\!\sim \rho$. Therefore $C(\alpha \wedge \beta) \subseteq C(\gamma)$. To check E-C-Cut, assume that $\alpha \wedge \beta \mid\!< \gamma$ and also that $\delta \vdash_\Sigma \beta$ for all δ such that $\alpha \mid\!< \delta$. In particular, since $\mid\!<$ is reflexive, we have that $\alpha \vdash_\Sigma \beta$. Thus $\alpha \mid\!\sim \beta$ and therefore $C(\alpha) = C(\alpha \wedge \beta)$.

(ii) DR says that $C(\alpha \vee \beta) \subseteq C(\alpha) \cup C(\beta)$ from which it is obvious that LOR holds.

(iii) Suppose that $\mid\!\sim$ is rational. We will use the following well known fact about rational relations. For every pair of formulas α and β one of the following holds: (i) $C(\alpha \vee \beta) = C(\alpha)$, (ii) $C(\alpha \vee \beta) = C(\beta)$ (iii) $C(\alpha \vee \beta) = C(\alpha) \cap C(\beta)$. From this is obvious that E-DR and the postulate stated in (3) hold. It remains to be checked that E-R-Cut holds. Suppose that $\alpha \wedge \beta \mid\!< \gamma$ and also that there is δ such that $\alpha \mid\!< \delta$ and $\delta \vdash_\Sigma \beta$. By RM it suffices to show that $\alpha \not\mid\!\sim \neg\beta$. Assume $\alpha \mid\!\sim \neg\beta$ towards a contradiction.Since $\alpha \mid\!< \delta$, then $\delta \mid\!\sim \neg\beta$. Since $\delta \vdash_\Sigma \beta$, then by preferentiality and Con_Σ $\delta \mid\!\sim \beta$. Which together with Con_Σ contradicts the fact that δ is Σ-consistent. ∎

Proof of proposition 5.41: To see E-Reflexivity just notice that if $\alpha \rhd \gamma$ then it is obvious that $\gamma \in min(Expla(\gamma)), \prec)$.To check that E-CM holds, suppose $\alpha \rhd \gamma$ and $\gamma \vdash_\Sigma \beta$, then $\gamma \in min(Expla(\alpha)) \cap Expla(\alpha \wedge \beta) \subseteq min(Expla(\alpha \wedge \beta))$. ∎

Proof of theorem 5.44: That LLE_Σ and RLE_Σ hold follows from the fact that $\prec$ is logically invariant. We already have shown in 5.41 that E-CM holds. To see that E-Con$_\Sigma$ holds, suppose thatγ is consistent with Σ then $Expla(\alpha)$ is not empty. By smoothness there is γ such that $\alpha \rhd \gamma$. To see that E-C-Cut, suppose that the premises in the rule E-C-Cut holds. Hence $min(Expla(\alpha)) \subseteq Expla(\beta)$ and since $Expla(\alpha \wedge \beta) \subseteq Expla(\alpha)$, then $min(Expla(\alpha)) \subseteq min(Expla(\alpha \wedge \beta))$. Since $Expla(\alpha)$ is smooth we conclude $min(Expla(\alpha)) = min(Expla(\alpha \wedge \beta))$ and this finishes the proof. □

Proof of theorem 5.49: $(ii) \Rightarrow (i)$. By 5.44 we only need to show that $\rhd$ satisfies LOR. But this follows immediately from (5.15).

$(i) \Rightarrow (ii)$. We will show that $\prec_e$ works. We already have observed that since $\rhd$ satisfies RLE$_\Sigma$ then $\prec_e$ is a preference relation. First, notice that (5.15) follows immediately from (5.16) and LOR.

We will show that (5.16) holds. Let us suppose that $\alpha \rhd \gamma$ and let $\delta \in Expla(\alpha)$, then $\alpha \in Cn(\Sigma \cup \{\delta\}) \cap \{\beta : \beta \rhd \gamma\}$. Therefore $\delta \not\prec_e \gamma$ and $\gamma \in min(Expla(\alpha))$. This shows that the *only if* in (5.16) holds.

Fix a formula α' consistent with Σ and let δ' be any formula in $Expla(\alpha')$. We will show that if $\alpha' \not\rhd \delta'$, then there is γ such that $\alpha' \rhd \gamma$ and $\gamma \prec_e \delta'$. In particular, this will prove that $\prec_e$ is smooth and also that the other direction in (5.16) holds. Suppose $\alpha' \not\rhd \delta'$. If δ' is not admissible, then there is nothing to show because of the definition of $\prec_e$ and E-Con$_\Sigma$. Hence we will assume that δ' is admissible. By E-Con$_\Sigma$ there is γ such that $\alpha' \rhd \gamma$, so let

$$C_{\alpha'} = \bigcap \{Cn\ (\Sigma \cup \{\gamma\}) : \alpha' \rhd \gamma\}$$

and

$$S = C_{\alpha'} \cup \{\neg\beta : \beta \rhd \delta'\}.$$

We claim that S is consistent. In fact, suppose, towards a contradiction, that S is inconsistent. By compactness there are β_i's for $i = 1, \ldots, n$ such that $\beta_i \rhd \delta'$ and $(\beta_1 \vee \cdots \vee \beta_n) \in C'_\alpha$. Let $\beta = \beta_1 \vee \cdots \vee \beta_n$. By LOR we know that $\beta \rhd \delta'$. By E-CM we have that $(\alpha' \wedge \beta) \rhd \delta'$. Since $\beta \in C_{\alpha'}$, then by E-C-Cut we conclude $\alpha' \rhd \delta'$, which is a contradiction. Therefore S is consistent.

Since $\rhd$ is logically finite there are two cases to be considered:

(a) $\rhd$ satisfies RLF, *i.e.* for every formula α the set $A = \{\gamma : \alpha \rhd \gamma\}$ has an upper bound. Let $\gamma_i \in A$, $i \leq n$ be an upper bound for A. It is easy to check that

$$C_{\alpha'} = \bigcap \{Cn\ (\Sigma \cup \{\gamma\}) : \gamma \in A\} = \bigcap \{Cn\ (\Sigma \cup \{\gamma_i\}) : i \leq n\} = Cn\ (\Sigma \cup \{(\gamma_1 \vee \cdots \gamma_n)\}.$$

Let N be a model of S, then there is i such that $N \models \Sigma \cup \{\gamma_i\}$. As N is also a model of $\{\neg\beta : \beta \rhd \delta'\}$, then it is clear that $\gamma_i \prec_e \delta'$.

(b) $\rhd$ satisfies LLF, *i.e.* for every admisible formula γ the set $\{\beta : \beta \rhd \gamma\}$ has a upper bound. Since δ' is admisible, let $\beta_1, \ldots, \beta_n$ be such that $\beta_i \rhd \delta'$ and $\beta \vdash_\Sigma \beta_1 \vee \cdots \vee \beta_n$ for every β such that $\beta \rhd \delta'$. Let $\beta' = \beta_1 \vee \cdots \vee \beta_n$, then by LOR $\neg\beta' \in S$. Since S is consistent then $\beta' \notin C_{\alpha'}$, hence there is γ such that $\alpha' \rhd \gamma$ and $\gamma \not\vdash_\Sigma \beta'$. Therefore $\gamma \not\vdash_\Sigma \beta$, for all β such that $\beta \rhd \delta'$, *i.e.* $\gamma \prec_e \delta'$. □

Proof of proposition 5.51: ($\prec_e \subseteq \prec_u$): Let γ, δ be admissible formulas with $\gamma \prec_e \delta$. Let α, β be such that $\alpha \rhd \gamma$ and $\beta \rhd \delta$. We want to show that $(\alpha \vee \beta) \rhd \gamma$ and $(\alpha \vee \beta) \not\rhd \delta$. By E-DR it suffices to show that $(\alpha \vee \beta) \not\rhd \delta$. But this follows directly from the definition of $\prec_e$ and the fact that $\gamma, \delta \in Expla(\alpha \vee \beta)$ and $\gamma \prec_e \delta$.

($\prec_u \subseteq \prec_e$): Let γ, δ be admissible formulas with $\gamma \prec_u \delta$. Suppose, towards a contradiction, that there is β such that $\beta \rhd \delta$ and $\gamma \vdash_\Sigma \beta$. Let α be any formula such that $\alpha \rhd \gamma$. Since $\gamma \vdash_\Sigma \beta$, then by E-CM we have $(\alpha \wedge \beta) \rhd \gamma$. Since $\vdash ((\alpha \wedge \beta) \vee \beta) \leftrightarrow \beta$ and $\gamma \prec_u \delta$, then (by LLE$_\Sigma$) we conclude that $\beta \not\rhd \delta$, which is a contradiction.

To see that $\prec_u$ is transitive, let γ_i be formulas such that $\gamma_1 \prec_u \gamma_2$ and $\gamma_2 \prec_u \gamma_3$. Without lost of generality we can assume that each γ_i is admissible. Let α_i be formulas such that $\alpha_i \rhd \gamma_i$. By E-DR it suffices to show that $(\alpha_1 \vee \alpha_3) \not\rhd \gamma_3$. Suppose, towards a contradiction, that $(\alpha_1 \vee \alpha_3) \rhd \gamma_3$. Since $\gamma_2 \prec_u \gamma_3$, then by definition of $\prec_u$ we have $(\alpha_1 \vee \alpha_2 \vee \alpha_3) \rhd \gamma_2$ and $(\alpha_1 \vee \alpha_2 \vee \alpha_3) \not\rhd \gamma_3$. Since $\gamma_1 \prec_u \gamma_2$, then analogously we have $(\alpha_1 \vee \alpha_2 \vee \alpha_3) \rhd \gamma_1$ and $(\alpha_1 \vee \alpha_2 \vee \alpha_3) \not\rhd \gamma_2$, which is a contradiction. □

Proof of proposition 5.52: To see (ii), suppose $\delta \not\prec_e \gamma$ and let β be such that $\delta \vdash_\Sigma \beta$ and $\beta \rhd \gamma$. Then clearly $\gamma \vee \delta \vdash_\Sigma \beta$ and thus $(\gamma \vee \delta) \not\prec_e \gamma$. The proof of (i) is similar. For (iiia), suppose that $\delta \prec_e \gamma$. Thus by definition δ is admissible. If γ' is not admissible then by definition $\delta \prec_e \gamma'$. Now suppose that γ' is admissible and also, towards a contradiction, that $\delta \not\prec_e \gamma'$. Let β be such that $\beta \rhd \gamma'$ and $\delta \vdash_\Sigma \beta$. Since $\gamma \vdash_\Sigma \gamma'$ and $\gamma \not\vdash_\Sigma \bot$, we have by RA that $\beta \rhd \gamma$ and therefore $\delta \not\prec_e \gamma$. To see (iiib), assume $\delta \prec_e \gamma$ and $\delta \prec_e \rho$. By this assumption and by definition of $\prec_e$ δ is admissible. If $(\gamma \vee \rho)$ is not admissible then there is nothing to show. Assume $(\gamma \vee \rho)$ is admissible and suppose, towards a contradiction, that $\delta \not\prec_e (\gamma \vee \rho)$. Let β be such that $\delta \vdash_\Sigma \beta$ and $\beta \rhd (\gamma \vee \rho)$, then by 5.5 either $\beta \rhd \gamma$ or $\beta \rhd \rho$. Since $\delta \vdash_\Sigma \beta$, then either $\delta \not\prec_e \gamma$ or $\delta \not\prec_e \rho$, a contradiction. ∎

Proof of proposition 5.53 Suppose that $\alpha \rhd \gamma$, $\gamma' \vdash_\Sigma \gamma$ and $\gamma' \not\vdash_\Sigma \bot$. We want to show that $\alpha \rhd \gamma'$, *i.e.* $\gamma' \in min(Expla(\alpha), \prec)$. Since $\gamma' \not\vdash_\Sigma \bot$ then it is clear that $\gamma' \in Expla(\alpha)$. For reductio, assume there is $\delta \in Expla(\alpha)$ such that $\delta \prec \gamma'$. Since $\gamma' \vdash_\Sigma \gamma$ we have, by C-U , $\delta \prec \gamma$ contradicting the minimality of γ in $Expla(\alpha)$. □

Proof of proposition 5.54: The $\Rightarrow$ direction comes directly from the definition of $\prec_u$. For the other direction, let α and β be as in the right hand side of (5.17) and α' and β' be formulas such that $\alpha' \rhd \gamma$ and $\beta' \rhd \delta$. We need to show that $(\alpha' \vee \beta') \rhd \gamma$ and $(\alpha' \vee \beta') \not\rhd \delta$. By 5.15 we know that $\rhd$ satisfies E-DR, hence it suffices to show that $(\alpha' \vee \beta') \not\rhd \delta$. Suppose, towards a contradiction, that $(\alpha' \vee \beta') \rhd \delta$. By E-CM we have $(\alpha' \vee \beta') \wedge (\alpha \vee \beta) \rhd \delta$. And by hypothesis $(\alpha \vee \beta) \rhd \gamma$ and clearly $\gamma \vdash_\Sigma (\alpha' \vee \beta')$, hence by E-R-Cut $(\alpha \vee \beta) \rhd \delta$, which contradicts the choice of α and β.

To see that $\prec_u$ is smooth, we first recall thatby 5.15 $\rhd$ satisfies E-DR and therefore by 5.51 $\prec_u = \prec_e$. As in the proof of 5.49 we have that if $\alpha \rhd \gamma$ then $\gamma \in min(Expla(\alpha), \prec_e)$. For the other direction, let $\delta \in Expla(\alpha)$ such that $\alpha \not\rhd \delta$. We will find γ such that $\alpha \rhd \gamma$ and $\gamma \prec_u \delta$. This will show that $\prec_u$ is smooth and also that it represents $\rhd$. We can assume without loss of generality that δ is admissible and thus let β be such that $\beta \rhd \delta$. Hence by E-CM $(\alpha \wedge \beta) \rhd \delta$. By E-Con$_\Sigma$ there is γ such that $\alpha \rhd \gamma$. Since α is logically equivalent to $(\alpha \wedge \beta) \vee \alpha$, then $((\alpha \wedge \beta) \vee \alpha) \not\rhd \delta$ and $((\alpha \wedge \beta) \vee \alpha) \rhd \gamma$. From (5.17) we conclude that $\gamma \prec_u \delta$. This finishes the proof. ∎

Proof of proposition 5.56: Let γ, δ and ρ be formulas such that $\gamma \not\prec_u \delta$, $\delta \not\prec_u \gamma$ and $\gamma \prec_u \rho$. We want to show that $\delta \prec_u \rho$. Without lost of generality we can assume that γ, δ and ρ are admissible. Let α, β, ω formulas such that $\alpha \rhd \gamma$, $\beta \rhd \delta$ and $\omega \rhd \rho$. Since γ and δ are $\prec_u$-incomparable then from 5.54 it follows that $(\alpha \vee \beta) \rhd \gamma$ and $(\alpha \vee \beta) \rhd \delta$. Again by 5.54 it suffices to show that $(\alpha \vee \beta \vee \omega) \rhd \delta$ and $(\alpha \vee \beta \vee \omega) \not\rhd \rho$. By E-DR, which is true by 5.15,it is enough to show that $(\alpha \vee \beta \vee \omega) \not\rhd \rho$. Since $\gamma \prec_u \rho$, then by definition of $\prec_u$ we have $(\alpha \vee \beta \vee \omega) \rhd \gamma$ and $(\alpha \vee \beta \vee \omega) \not\rhd \rho$. Finally that $\prec_u$ satisfies C-U follows from 5.53. □

Proof of proposition 5.57: From 5.44 we know that $\rhd$ satisfies LLE_Σ, RLE_Σ, E-CM, E-C-Cut and E-Con_Σ. From 5.53 we obtain RA. It remains to be shown that E-R-Cut holds. Let α, β, γ and δ formulas such that $(\alpha \wedge \beta) \rhd \gamma$, $\alpha \rhd \delta$ and $\delta \vdash_\Sigma \beta$. We need to show that $\alpha \rhd \gamma$. Suppose, towards a contradiction, that $\alpha \not\rhd \gamma$. Since $\gamma \vdash_\Sigma \alpha$, then by smoothness and the definition of $\rhd$, there is δ' such that $\alpha \rhd \delta'$ and $\delta' \prec \gamma$. Since $\alpha \rhd \delta$ then $\delta \not\prec \delta'$ and $\delta' \not\prec \delta$. By E-CM $(\alpha \wedge \beta) \rhd \delta$ and by modularity, $\delta \prec \gamma$, which contradicts the hypothesis that $(\alpha \wedge \beta) \rhd \gamma$. ∎

Chapitre 6

Analysing rational properties of change operators based on forward chaining

We propose an abstract framework to analyse the rationality of change operators defined in a syntactical way. More precisely we propose "syntactical" postulates of rationality stemming from AGM ones. Then we introduce five change operators based on forward chaining. Finally we apply our abstract framework to analyse the rationality of our operators.

Introduction

Revision is the process of according an old knowledge base with a new evidence. In order to have a good behaviour, a revision operator must obey a minimal set of rationality requirements. For example it must obey the principle of primacy of update that demands the new evidence to be true in the new knowledge base, and the principle of minimal change that imposes that the new knowledge base has to be as close as possible to the old one. These properties are intuitive requirements one can expect from revision operators. These operators and their properties have been formally studied in philosophy, artificial intelligence and databases [1, 36, 45] and several operators have already been proposed [14, 27, 112, 113, 101].

In general, revision is a complex process [24, 60] and is not efficiently computable. The problem is that revision operators usually handle theories closed under logical consequences. Then, the computation of (all the consequences of) the new theory according to the old one and to the new information is generally prohibitive. One solution is to work with theories that are not closed under logical consequences [34, 42, 41, 78, 80, 110] and to take their logical closure only when one needs them. Of course such an approach is syntax sensitive. Another solution is to work in a restrained framework, a "weaker" (tractable) logic, instead of the classical one. It is a combination of these two approaches that is proposed here.

In this paper, we investigate five change operators based on forward chaining. The use of forward chaining provides us with an efficient way to compute the revision of a knowledge base.

Furthermore these operators are readily suitable for expert systems based on the same kind of inference. Thus, we have an easy way to include non-monotonic reasoning in such systems. Such operators may have numerous other applications, in diagnosis systems for example.

For some operators, our approach is close to REVISE [16] and Revision Programming [71] but is simpler, since we use only forward chaining on propositional formulae; in particular, we don't assume negation by failure.

We propose five knowledge change operators. The first one, called *factual update*, updates a set of facts with another set of facts coding a new evidence, according to a set of rules which can be seen as integrity constraints of the system. The other operators revise programs by programs. More precisely the second one, called *ranked revision*, is based on a hierarchy over the rules which denotes how *exceptional* the rules are, and when a new evidence arrives, it finds the least exceptional rules consistent with this new information. The third one, called *hull revision*, extends the result of ranked revision to a set which remains consistent with the new information. The fourth operator, *extended hull*, combines the approaches of hull revision and factual update operators. The fifth operator, called *selection hull*, is actually a family of operators defined by selection functions.

One of the main contributions of this work is the study of the rationality properties of these operators. To do that we introduce syntactical relativizations of the main postulates proposed in the literature for theory change. We prove that factual update can be seen as an update operator, *i.e.* satisfying a syntactical version of Katsuno and Mendelzon postulates [46]. In the same way, ranked revision and selection hull (when the selection function used to define it has good properties) can be seen as revision operators, according to Alchourrón-Gärdenfors-Makinson (AGM) postulates [1, 36]. Concerning hull revision and extended hull revision, although they seem to be rational extensions of the ranked operator, we prove that only some basic postulates of change hold.

The paper is organized as follows: section 6.1 contains the basic definitions of our abstract framework; in section 6.2.1 we define factual revision (an algorithm to compute it is given in the appendix); section 6.2.2 is devoted to definitions of ranked revision, hull revision and extended hull revision; in section 6.3 we study the properties of these knowledge change operators. In section 6.4 we introduce the selection hull operators. Finally, we conclude with some remarks and some perspectives for future work.

6.1 Preliminaries

Our framework is finite propositional logic.

A literal (or fact) is an atom or a negation of an atom. The set of literals will be denoted *Lit*. A rule is a formula of the shape $l_1 \wedge l_2 \wedge \cdots \wedge l_n \rightarrow l_{n+1}$ where l_i is a literal for $i = 1, \ldots, n+1$. A rule as before will be denoted $l_1, l_2, \ldots, l_n \rightarrow l_{n+1}$. We admit rules of the form $\rightarrow l$ which actually code facts.

Let R and L be a finite set of rules and a finite set of literals (both possibly empty) respectively. A program P is a set of the form $R \cup L$ and we will say that the elements of R are the rules of P and the elements of L are the facts of P. The set of programs will be denoted by *Prog*

Let $P = R \cup L$ be a program. We define the set of consequences by forward chaining of P, denoted $C_{fc}(P)$, as the smallest set of literals L' such that:(i) $L \subseteq L'$. ii) If $l_1, l_2, \ldots, l_n \rightarrow l$

is in R and $l_i \in L'$ for $i = 1, \ldots, n$ then $l \in L'$.(iii) If L' contains two opposite literals then $L' = Lit$

A program P is said to be consistent iff $C_{fc}(P) \neq Lit$.

Let P and L be a program and a finite set of literals respectively. L is said to be P-consistent iff $P \cup L$ is consistent (with respect to forward chaining).

In the sequel ω will denote the set of positive integers.

6.1.1 Revision and update postulates

We begin this section by recalling the rationality postulates that have been proposed [1, 36, 45, 46] in the area of revision theory, i.e. properties that an operator has to satisfy in order to have a "good" behaviour as a change operator. Then we will give a relativization of these notions to a syntactical abstract framework.

First let's consider the postulates for revision operators. Let φ be a formula representing a knowledge base and let μ be a formula representing a new piece of information. $\varphi \circ \mu$ will denote a formula representing the changes that μ produces on φ. The operator $\circ$ is a revision operator [1, 45] if it satisfies the following postulates:

(R1) $\vdash (\varphi \circ \mu) \rightarrow \mu$.

(R2) If $\varphi \wedge \mu$ is consistent then $\vdash (\varphi \circ \mu) \leftrightarrow (\varphi \wedge \mu)$.

(R3) If μ is consistent then $\varphi \circ \mu$ is consistent.

(R4) If $\vdash (\varphi_1 \leftrightarrow \varphi_2)$ and $\vdash (\mu_1 \leftrightarrow \mu_2)$ then $\vdash (\varphi_1 \circ \mu_1) \leftrightarrow (\varphi_2 \circ \mu_2)$.

(R5) $\vdash ((\varphi \circ \mu) \wedge \phi) \rightarrow (\varphi \circ (\mu \wedge \phi))$.

(R6) If $(\varphi \circ \mu) \wedge \phi$ is consistent then $\vdash (\varphi \circ (\mu \wedge \phi)) \rightarrow ((\varphi \circ \mu) \wedge \phi)$.

The intuitive meaning of these postulates is the following: the new piece of information must be true in the new knowledge base, which is required by (R1). (R2) states that if the new piece of information is consistent with the old knowledge base then the revision is reduced to the addition of the new piece of information to the old knowledge base. (R3) assures that if the new piece of information is consistent then the new knowledge base must be consistent. (R4) is the so called *Dalal's principle of irrelevance of syntax* and says that the result of the revision depends neither on the syntax of the new piece of information nor on the one of the knowledge base. (R5) and (R6) assure that the result of the revision is "closest" to the old base and that this notion of closeness behaves well. For more explanations on the meaning of these postulates see [36, 45].

Revision is adequate to model change of belief about a static world but, as shown in [46], is not able to cope with change in a dynamic world. Katsuno and Mendelzon have defined update operators for this case. The rationality postulates for update operators they propose are given next.

The operator $\circ$ is an update operator [46] if it satisfies the following postulates:

(U1) $\vdash (\varphi \circ \mu) \rightarrow \mu$.

(U2) If $\vdash \varphi \rightarrow \mu$ then $\vdash (\varphi \circ \mu) \leftrightarrow \varphi$.

(U3) If both φ and μ are consistent then $\varphi \circ \mu$ is also consistent.

(U4) If $\vdash (\varphi_1 \leftrightarrow \varphi_2)$ and $\vdash (\mu_1 \leftrightarrow \mu_2)$ then $\vdash (\varphi_1 \circ \mu_1) \leftrightarrow (\varphi_2 \circ \mu_2)$.

(U5) $\vdash ((\varphi \circ \mu) \wedge \phi) \rightarrow (\varphi \circ (\mu \wedge \phi))$.

(U6) If $\vdash (\varphi \circ \mu_1) \rightarrow \mu_2$ and $\vdash (\varphi \circ \mu_2) \rightarrow \mu_1$ then $\vdash (\varphi \circ \mu_1) \leftrightarrow (\varphi \circ \mu_2)$.

(U7) If φ is complete then $\vdash ((\varphi \circ \mu_1) \wedge (\varphi \circ \mu_2)) \rightarrow (\varphi \circ (\mu_1 \vee \mu_2))$.

(U8) $\vdash ((\varphi_1 \vee \varphi_2) \circ \mu) \leftrightarrow ((\varphi_1 \circ \mu) \vee (\varphi_2 \circ \mu))$.

These postulates are close to those for revision. Postulates (U1)-(U5) correspond to postulates (R1-R5) and the intuitive meaning of these postulates is: (U1) is exactly (R1),*i.e.* the new piece of information must be true in the new knowledge base, (U2) states that if the new piece of information is weaker than the knowledge base then updating by this new piece of information has no effect on the knowledge base. Notice that if the knowledge base is consistent then (U2) is weaker than (R2). (U3) assures that if the new piece of information and the old knowledge base are consistent then the new knowledge base is consistent. (U4) is exactly (R4), the principle of irrelevance of syntax. (U5) is exactly (R5). It assures that the notion of "minimal change" behaves well. (U6) says that if μ_1 is true when we update the knowledge base by μ_2 and if μ_2 is true when we update the knowledge base by μ_1, then the two updates are equivalent. (U7) states that for a complete knowledge base the conjunction of two updates contains the information of the update by the disjunction of the two pieces of information. (U8) is the disjunction rule: from a semantical point of view a knowledge base can be considered as the sum of all its possible worlds, so (U8) states that updating this sum is the sum of updating. This assures that each possible world of the knowledge base is given independent consideration.

Note that in revision and update postulates the notions of consequence, consistency and equivalence are the classical ones. We will investigate the instantiation of these postulates to different "logics". This point is essential in this paper because when manipulating syntactical objects (such as databases) we have to define some abstract notions of *consequence, conjunction, disjunction* in order to be able to analyse the properties of the operators. Particularly, we will focus in this paper on a "forward chaining logic".

More precisely, a first case of this sort of instantiation concerns the postulates for revision. This is done in next definition.

Definition 6.1 *Suppose that we are manipulating objects of a set Ω (a set of "formulas" or "knowledge bases"), a set $\Gamma \subseteq \Omega$ and a set $\mathcal{L}$ (the "deducible atoms") such that $\mathcal{P}(\mathcal{L}) \subseteq \Omega$. Consider we have a map $\mathcal{C} : \Omega \longrightarrow \mathcal{P}(\mathcal{L})$ ($\mathcal{C}$ is a consequence operator for the chosen logic). Finally suppose we have a function (a change operator) $\vartriangle : \Omega{\times}\Gamma \longrightarrow \Omega$ and a function (the "conjunction") $\otimes : \Omega{\times}\Gamma \longrightarrow \Omega$, such that $\otimes : \Gamma{\times}\Gamma \longrightarrow \Gamma$,* i.e. *the restriction of $\otimes$ to couples in Γ takes its values in Γ. Then $\vartriangle$ is said to be a **syntactical revision operator** (with respect to $\mathcal{C}$ and $\otimes$) if the following postulates hold:*

(SR1) $\mathcal{C}(Y) \subseteq \mathcal{C}(X \vartriangle Y)$.

(SR2) *If* $\mathcal{C}(X \otimes Y) \neq \mathcal{L}$ *then* $\mathcal{C}(X \vartriangle Y) = \mathcal{C}(X \otimes Y)$.

(SR3) *If* $\mathcal{C}(Y) \neq \mathcal{L}$ *then* $\mathcal{C}(X \triangle Y) \neq \mathcal{L}$.

(SR5) $\mathcal{C}(X \triangle (Y \otimes Z)) \subseteq \mathcal{C}((X \triangle Y) \otimes Z)$.

(SR6) *If* $\mathcal{C}((X \triangle Y) \otimes Z) \neq \mathcal{L}$ *then* $\mathcal{C}((X \triangle Y) \otimes Z) \subseteq \mathcal{C}(X \triangle (Y \otimes Z))$.

Let us remark that the postulates **(SRi)** are the natural counterparts of postulates **(Ri)** when we interpret $\otimes$ as the conjunction of formulas and thinking X consistent iff $\mathcal{C}(X) \neq \mathcal{L}$.

There is no postulate corresponding to **(R4)** (alias **U4**), the postulate of irrelevance of syntax, because the operators we will define are in general syntax-sensitive. Nevertheless in our abstract setting we could define the counterpart of **(R4)** in the following way:

(IS) If $\mathcal{C}(X) = \mathcal{C}(Y)$ and $\mathcal{C}(Z) = \mathcal{C}(W)$ then $\mathcal{C}(X \triangle Z) = \mathcal{C}(Y \triangle W)$.

Unfortunately this does not hold for our operators as we will see in observation 6.23.

The second case of instantiation we consider concerns the postulates for update. This is the object of next definition.

Definition 6.2 *Consider two sets* Ω *and* $\mathcal{L}$ *such that there is a set* $\Gamma \subseteq \mathcal{P}(\mathcal{L})$ *with* $\Gamma \subseteq \Omega$. *Let* $\mathcal{C}$ *be a function* $\mathcal{C} : \Omega \longrightarrow \mathcal{P}(\mathcal{L})$. *Suppose we have a function (a change operator)* $\triangle : \Omega \times \Gamma \longrightarrow \Omega$. *Now suppose that we have an associative "connector" (our "disjunction")* $\oplus : \Gamma \times \Gamma \longrightarrow \Gamma$ *such that* $\mathcal{C}(X \oplus Y) = \mathcal{C}(X) \cap \mathcal{C}(Y)$, i.e. *the behaviour of* $\oplus$ *with respect to* $\mathcal{C}$ *is like a disjunction. We also suppose we have a function* $\otimes : \Omega \times \Gamma \longrightarrow \Omega$, *such that* $\otimes : \Gamma \times \Gamma \longrightarrow \Gamma$, i.e. *the restriction of* $\otimes$ *to couples in* Γ *takes its values in* Γ. *The operator* $\triangle$ *is said to be a* ***syntactical update operator*** *with respect to* $\mathcal{C}$, $\otimes$ *and* $\oplus$ *if the following postulates hold:*

(SU1) $\mathcal{C}(Y) \subseteq \mathcal{C}(X \triangle Y)$.

(SU2) *If* $\mathcal{C}(Y) \subseteq \mathcal{C}(X)$ *then* $\mathcal{C}(X \triangle Y) = \mathcal{C}(X)$.

(SU3) *If* $\mathcal{C}(X) \neq \mathcal{L}$ *and* $\mathcal{C}(Y) \neq \mathcal{L}$ *then* $\mathcal{C}(X \triangle Y) \neq \mathcal{L}$.

(SU5) $\mathcal{C}(X \triangle (Y \otimes Z)) \subseteq \mathcal{C}((X \triangle Y) \otimes Z)$.

(SU6) *If* $\mathcal{C}(Y_1) \subseteq \mathcal{C}(X \triangle Y_2)$ *and* $\mathcal{C}(Y_2) \subseteq \mathcal{C}(X \triangle Y_1)$ *then* $\mathcal{C}(X \triangle Y_1) = \mathcal{C}(X \triangle Y_2)$.

(SU8) $\mathcal{C}((X \oplus Y) \triangle Z) = \mathcal{C}((X \triangle Z) \oplus (Y \triangle Z))$.

The postulates **(SUi)** are the natural counterparts of postulates **(Ui)** (notice that we have asked the "connector" $\oplus$ to have the behaviour of a disjunction with respect to the notion of consequence $\mathcal{C}$). Remark also that there is no postulate corresponding to **(U7)** because in general the image of a couple of elements of Γ under $\oplus$ will not belong to Γ.

As for syntactical revision operators, there is no postulate corresponding to **U4**.

6.2 Some syntactical change operators

The purpose of this section is to define some change operators essentially based on forward chaining. The first one, factual update, updates a set of facts by a set of facts coding a change in the world according to a set of integrity constraints. The following three ones, namely ranked revision, hull revision and extended hull revision are based on a ranking of sentences of the programs. We will analyse the rational properties they satisfy in section 6.3.

6.2.1 Factual update

Let P be a fixed program which in this context can be seen as our background theory or our integrity constraints. Let L be a set of facts which can be seen as our beliefs about the world. We would like to define the change produced by a set of facts L' coding a new piece of information about the world. The following definition describes the result of this change:

$$L \diamond_P L' = \begin{cases} Lit & \text{if } L \text{ or } L' \text{ is not } P\text{-consistent} \\ \langle L_1 \cup L', \ldots, L_n \cup L' \rangle & \text{otherwise} \end{cases}$$

where $\{L_1, \ldots, L_n\}$ is the set of subsets of L which are maximal and $P \cup L'$-consistent.

So more generally than a set of facts L we are considering unordered tuples of sets of facts $\langle L_1, \ldots, L_n \rangle$ called *flocks* in the literature [27]. Such flocks can be also seen as multisets.

We define the concatenation of flocks '$\cdot$' in the obvious way:

$$\langle L_1, \ldots, L_n \rangle \cdot \langle L'_1, \ldots, L'_m \rangle \stackrel{def}{=} \langle L_1, \ldots, L_n, L'_1, \ldots, L'_m \rangle$$

and we define the change produced in a flock by a new piece of information by the following

$$\langle L_1, \ldots, L_n \rangle \diamond_P L' \stackrel{def}{=} \begin{cases} Lit & \text{if } L' \text{ or all the } L_i \text{ are not } P\text{-consistent} \\ (L_{i_1} \diamond_P L') \cdot (L_{i_2} \diamond_P L') \cdots (L_{i_k} \diamond_P L') & \text{otherwise} \end{cases}$$

where $\{L_{i_1}, \ldots, L_{i_k}\}$ is the set of all sets in $\{L_1, \ldots, L_n\}$ which are consistent with P.

In order to investigate the relation between $\diamond_P$ and the postulates of change we need to define the intensional content (the consequences) of a flock $\mathcal{F} = \langle L_1, \ldots, L_n \rangle$. So we define the consequences by forward chaining (with respect to P) of such a flock, denoted $C^P_{fc}(\mathcal{F})$, by the following:

$$C^P_{fc}(\mathcal{F}) = \bigcap_{i=1}^{n} C_{fc}(L_i \cup P)$$

So, we adopt here a sceptical point of view, since the consequences of a flock are the facts that are true in every elements of the flock.

In section 6.3 we will show that $\diamond_P$ is a syntactical update operator.

Example 6.3 *We consider the program P and the set of literals L, defined by $P = \{a, b \rightarrow c \; ; \; a, d \rightarrow c\}$ and $L = \{a, b, d\}$. Put $L' = \{\neg c\}$. L' is not $P \cup L$ consistent and then some facts must be "retracted" from the old base L. Then it is easy to see that*

$$L \diamond_P L' = \langle \{a\} \cup \{\neg c\}, \{b, d\} \cup \{\neg c\} \rangle$$

For the sake of completeness we give in the appendix an algorithm to compute $L \diamond_P L'$.

6.2.2 Ranked revision, hull revision and extended hull revision

In the case of factual update the program is fixed and we restrain the new piece of information to be a set of facts. When it is not the case a natural question that one may ask is how to change a program when a new piece of information arrives. The aim of this section is to give an answer to this question even when the new piece of information is a program.

The change operators introduced in this section are inspired by the duality existing between revision and rational inference relations [31, 70]. So the first operator can be seen as the 'relativization' of the rational closure [55] to the forward chaining logic. The second operator is an extension of the first one and it is aimed to satisfy a little bit more of transitivity [9, 7].

Definition 6.4 (Exceptional sets of literals and rules) *Let P be a program. A set of literals L is said to be exceptional with respect to P iff L is not P-consistent and a rule $L \longrightarrow l$ of P is said to be exceptional in P iff L is exceptional in P.*

Notice that, when the body of a rule is empty, this rule will be exceptional iff P is not consistent. In this case all the rules are exceptional.

A similar definition of exceptionality for a formula can be found in [55].

Definition 6.5 (Base) *Let P be a program. Let $(P_i)_{i\in\omega}$ be the decreasing sequence defined by: P_0 is P and P_{i+1} is the set of all exceptional rules of P_i. Since P is finite there is a smallest integer n_0 such that for all $m \geq n_0$ we have $P_m = P_{n_0}$. If $P_{n_0} \neq \emptyset$ we say that $\langle P_0, \ldots, P_{n_0}, \emptyset\rangle$ is the base of P. If $P_{n_0} = \emptyset$ we say that $\langle P_0, \ldots, P_{n_0}\rangle$ is the base of P.*

Thus a program P has intrinsically a hierarchy, its base, in which the greater n is, the more exceptional the information in P_n is.

Definition 6.6 (Rank function) *Fix a program P and let $\langle P_0, \ldots, P_n\rangle$ be the base of P. Let $\rho : Prog \longrightarrow \omega$ be the rank function defined as follows: $\rho(P') = min\{i \in \omega : P'$ is P_i-consistent$\}$ if P is consistent, otherwise $\rho(P') = n$. It is a fact that if $P' \subseteq P''$ then $\rho(P') \leq \rho(P'')$.*

Notice that actually the rank function has two parameters. Thus in the notation $P_{\rho(P')}$, $\rho(P')$ denotes the rank of P' with respect to P.

The rank of a new program P' denotes how this program is exceptional according to the old program P.

Definition 6.7 (Ranked revision) *Let P and P' be two programs. We define the ranked revision of P by P', denoted $P \circ_{rk} P'$, as follows:*

$$P \circ_{rk} P' = P_{\rho(P')} \cup P'$$

In other words we take from the base of P the first program (the least exceptional) that agrees with the new piece of information.

We will now slightly generalize the ranked revision operator, and define the hull revision operator from the definition of the "hull of P" $((h(P))$.

Let $I_P(P')$ be the set of maximal subsets of P which are consistent with P' and which contain $P_{\rho(P')}$. Define $h : Prog \longrightarrow \mathcal{P}(P)$ by $h(P') = \bigcap I_P(P')$

Definition 6.8 (Hull revision) *The hull revision of a program P by a program P' denoted $P \circ_h P'$ is defined as follows*

$$P \circ_h P' = h(P') \cup P'$$

Remark 6.9 *By the definitions it is easy to see that*

$$C_{fc}(P \circ_{rk} P') \subseteq C_{fc}(P \circ_h P')$$

Thus one can say that $\circ_h$ is a conservative extension of $\circ_{rk}$. It keeps information that does not come into account in the contradiction.

Remember that in the definition of hull revision of P by P' we first calculate the set $I_P(P')$ of subsets of P maxiconsistent with P' and containing $P_{\rho(P')}$; then we take a very sceptical approach putting $P \circ_h P' = (\bigcap I_P(P')) \cup P'$. What we want now is to be more permissive and we are going to manipulate $I_P(P')$ as a flock. The ideas here are very close to the ones of factual update (cf section 6.2.1).

First given two programs P and P' we are going to define $P \circ_{eh} P'$. Let $I_P(P')$ be as before. Then we put:

$$P \circ_{eh} P' = \begin{cases} \langle H_1 \cup P', \ldots, H_n \cup P' \rangle & \text{if } I_P(P') = \{H_1, \ldots, H_n\} \\ P' & \text{if } I_P(P') = \emptyset \end{cases}$$

Remark that the result is a flock of programs.

Now suppose that $\mathcal{F}$ is a flock of programs, say $\mathcal{F} = \langle Q_1, \ldots, Q_n \rangle$. Then we define $\mathcal{F} \circ_{eh} P$ by putting

$$\mathcal{F} \circ_{eh} P = (Q_1 \circ_{eh} P) \cdot (Q_2 \circ_{eh} P) \cdots (Q_n \circ_{eh} P)$$

where as in section 6.2.1 '$\cdot$' is the concatenation of flocks.

If $\mathcal{F} = \langle Q_1, \ldots, Q_n \rangle$ is a flock of programs we define $C_{fc}(\mathcal{F})$ by putting $C_{fc}(\mathcal{F}) = \bigcap_{i=1}^{n} C_{fc}(Q_i)$.

Remark 6.10 *Notice that with this definition $\circ_{eh}$ is a conservative extension of $\circ_h$,* i.e. $C_{fc}(P \circ_h P') \subseteq C_{fc}(P \circ_{eh} P')$. *For this reason the operator $\circ_{eh}$ is called* extended hull revision .

We will identify a program P with the flock $\langle P \rangle$.

In section 6.3 we will show that $\circ_{rk}$ is a syntactical revision operator and that the operators $\circ_h$ and $\circ_{eh}$ enjoy some of the properties of syntactical revision operators.

6.2.3 Examples of ranked revision, hull revision and extended hull revision.

In this subsection we give examples that illustrate the behaviour of ranked revision, hull, and extended hull operators and at the same time the differences between them.

In the following examples we are interested in facts one can infer from $P \circ P'$, *i.e.* the facts $l \in C_{fc}(P \circ P')$ when $\circ$ is $\circ_{rk}$, $\circ_h$ or $\circ_{eh}$. For simplicity we will take for P a set of rules with non empty body and P' a set of facts.

Example 6.11 *Consider* $P = \{b \to f \ ; \ b \to w \ ; \ o \to b \ ; \ o \to \neg f\}$ *where* b, o, f, w *stand respectively for birds, ostriches, fly and have wings. It is easy to see that the base is* $< P_0, P_1, P_2 >$ *where*

$$\begin{aligned} P_0 &= \{b \to f \ ; \ b \to w \ ; \ o \to b \ ; \ o \to \neg f\} \\ P_1 &= \{o \to b \ ; \ o \to \neg f\} \\ P_2 &= \emptyset \end{aligned}$$

Notice that $\rho(b) = 0$ *so* $I_P(b) = P_0 = P$ *and therefore*

$$P \circ_{rk} \{b\} = P \circ_h \{b\} = P \circ_{eh} \{b\} = P \cup \{b\}$$

$$C_{fc}(P \circ_{rk} \{b\}) = \{b, f, w\}$$

Thus on this example the three operators have exactly the same behaviour.

For the same P*, an easy computation shows that* $\rho(o) = 1$*. Since the set*

$$\{o \to b \ ; \ o \to \neg f \ ; \ b \to w\}$$

is the unique extension of P_1 *consistent with* $\{o\}$ *we have* $h(o) = \{o \to b \ ; \ o \to \neg f \ ; \ b \to w\}$ *so*

$$P \circ_h \{o\} = P \circ_{eh} \{o\} = \{o \to b \ ; \ o \to \neg f \ ; \ b \to w\} \cup \{o\}$$

Since $\rho(o) = 1$ *we have* $P \circ_{rk} \{o\} = \{o \to b \ ; \ o \to \neg f\} \cup \{o\}$*. Therefore*

$$C_{fc}(P \circ_h \{o\}) = C_{fc}(P \circ_{eh} \{o\}) = \{b, o, \neg f, w\}$$

but $C_{fc}(P \circ_{rk} \{o\}) = \{b, o, \neg f\}$*. Thus this example shows that hull revision and extended hull revision keep more information from the old program than ranked revision.*

Another classic taxonomic example (the calculations are left to the reader) is given by

Example 6.12 $P = \{m \to s \ ; \ c \to m \ ; \ c \to \neg s \ ; \ n \to c \ ; \ n \to s\}$ *where* m, s, c, n *stand respectively for mollusc, shell, cephalopod and nautili. The base is* $< P_0, P_1, P_2, P_3 >$ *where*

$$\begin{aligned} P_0 &= \{m \to s \ ; \ c \to m \ ; \ c \to \neg s \ ; \ n \to c \ ; \ n \to s\} \\ P_1 &= \{c \to m \ ; \ c \to \neg s \ ; \ n \to c \ ; \ n \to s\} \\ P_2 &= \{n \to c \ ; \ n \to s\} \\ P_3 &= \emptyset \end{aligned}$$

We have

$$C_{fc}(P \circ_h \{n\}) = \{n, c, s, m\} = C_{fc}(P \circ_{eh} \{n\})$$

and

$$C_{fc}(P \circ_{rk} \{n\}) = \{n, c, s\}$$

this shows that the hull (and extended hull) revision allows more inferences than ranked revision. In some other cases the revisions coincide, for instance

$$C_{fc}(P \circ_h \{c, \neg n\}) = C_{fc}(P \circ_{eh} \{c, \neg n\}) = \{c, \neg n, m, \neg s\} = C_{fc}(P \circ_{rk} \{c, \neg n\})$$

Now we consider an example showing that in general the extended hull revision allows more inferences than the hull revision

Example 6.13 *Take the following program*

$$P = \{a, b \to c \; ; \; a, \neg c \to d \; ; \; b, \neg c \to d \; ; \; a \; ; \; b\}$$

Put $P' = \{\neg c\}$. *The base of* P *is* $\langle P_0, P_1, P_2 \rangle$ *with*

$$\begin{aligned} P_0 &= \{a, b \to c \; ; \; a, \neg c \to d \; ; \; b, \neg c \to d \; ; \; a \; ; \; b\} \\ P_1 &= \{a, \neg c \to d \; ; \; b, \neg c \to d\} \\ P_2 &= \emptyset \end{aligned}$$

Clearly $\rho(P') = 1$ *and* $I_P(P') = \{P_1 \cup \{a, b \to c, a\}, P_1 \cup \{a, b \to c, b\}, P_1 \cup \{a, b\}\}$. *Thus* $h(P') = P_1$ *and therefore* $d \notin C_{fc}(P \circ_h P') = C_{fc}(P_1 \cup \neg c)$. *Whereas* $d \in C_{fc}(P \circ_{eh} P')$ *because for any* $Q \in I_P(P')$ *we have* $d \in C_{fc}(Q \cup \{\neg c\})$.

6.2.4 Computing hull revision and extended hull revision

In this subsection we show how, via a simple coding, we can compute the hull revision by using the factual revision defined in section 6.2.1.

The base $< P_0, \ldots, P_n >$ of P is easily computed.

To compute the class of maximal subsets of P which are consistent with L and which contain $P_{\rho(P')}$ we use the update algorithm given in the appendix in the following way.

Let $\ell' : P \longrightarrow \{r_1, \ldots, r_m\}$ and $\ell'' : P' \longrightarrow \{q_1, \ldots, q_k\}$ be two bijections where the r_i and the q_j are new atoms. Define $\ell : P \cup P' \longrightarrow \{r_1, \ldots, r_m\} \cup \{q_1, \ldots, q_k\}$ by $\ell(r) = \ell'(r)$ if $r \in R$ and $\ell(r') = \ell''(r')$ if $r' \in P'$. Let $M(P)$ be the modification of P in the following way: each rule $L \to l$ of P is replaced by the rule $r, L \to l$ where $r = \ell(L \to l)$. Analogously, let $M(P')$ be the modification of P' in the following way: each rule $L \to l$ of P' is replaced by the rule $r, L \to l$ where $r = \ell(L \to l)$. Note that the maximal subsets of P which are consistent with P' and which contain $P_{\rho(P')}$ are then those corresponding to the maximal subsets of the base of facts $\ell(P)$ computed as being its possible updatings with respect to $M(P) \cup M(P')$ by $\ell(P') \cup \ell(P_{\rho(P')})$. More precisely we have:

$$P \circ_h P' = \ell^{-1}\{ \; \bigcap[\ell(P) \diamond_{M(P) \cup M(P')} (\ell(P') \cup \ell(P_{\rho(P')})) \;] \; \}$$

In an analogous way

$$P \circ_{eh} P' = \ell^{-1}[\ell(P) \diamond_{M(P) \cup M(P')} (\ell(P') \cup \ell(P_{\rho(P')})) \;]$$

In order to illustrate this method take the following example:

Example 6.14 $P = \{b \to f; b \to w; o \to b; o \to \neg f\}$. *Let* $P' = \{o\}$. *Define* $\ell : P \cup P' \longrightarrow \{1, 2, 3, 4, 5\}$ *such that* $M(P) = \{1, b \to f \; ; \; 2, b \to w \; ; \; 3, o \to b \; ; \; 4, o \to \neg f\}$ *and* $M(P') = \{5 \to o\}$. *We have seen above that* $P_{\rho(P')} = \{o \to b; o \to \neg f\}$ *so* $\ell(P_{\rho(P')}) = \{3, 4\}$. *Therefore* ,

$$\begin{aligned} \ell(P) \diamond_{M(P) \cup M(P')} (\ell(P') \cup \ell(P_{\rho(P')})) &= \{1, 2, 3, 4\} \diamond_{M(P) \cup M(P')} \{5\} \cup \{3, 4\} \\ &= \langle \{2, 3, 4, 5\} \rangle \end{aligned}$$

and so $C_{fc}(P \circ_h P') = C_{fc}(\ell^{-1}(\{2,3,4,5\})) = C_{fc}(\{o \to b; o \to \neg f; b \to w; o\}) = \{o, b, \neg f, w\}$.

6.3 Change properties for $\diamond_P$, $\circ_{rk}$, $\circ_h$ and $\circ_{eh}$

In this section we analyse the rationality of our operators.

More exactly we will show that factual update can be seen as an update operator in our relativized version of the Katsuno-Mendelzon postulates and that ranked revision can be considered as a revision operator in in our relativized version of the Alchourrón-Gärdenfors-Makinson postulates. And we give some properties satisfied by hull revision and extended hull revision.

We begin with an observation the proof of which is straightforward.

Observation 6.15 *The functions C_{fc} and C^P_{fc} are idempotent and monotonic,* i.e. $\mathcal{C}(\mathcal{C}(X)) = \mathcal{C}(X)$ *and* $\mathcal{C}(X) \subseteq \mathcal{C}(X \cup Y)$. *Thus if* $X \subseteq \mathcal{C}(Y)$ *then* $\mathcal{C}(X) \subseteq \mathcal{C}(Y)$ *for* $\mathcal{C} = C_{fc}$ *or* $\mathcal{C} = C^P_{fc}$ *and X and Y in the appropriate domains.* ∎

We will show that $\diamond_P$ is a syntactical update operator. In order to do that we must give the instantiations for $\mathcal{L}$, Ω, $\mathcal{C}$ $\oplus$ and $\otimes$ used in definition 6.2. We do that in next definition.

Definition 6.16 $\mathcal{L} = Lit$; $\Gamma = \mathcal{P}(\mathcal{L})$; *$\Omega$ is the set of flocks in which each element is in Γ. We identify $L \in \Gamma$ with the flock $\langle L \rangle$. With this identification we have $\Gamma \subseteq \Omega$. The function* $\mathcal{C} : \Omega \longrightarrow \mathcal{P}(\mathcal{L})$ *is defined by* $\mathcal{C} = C^P_{fc}$. *The function* $\oplus : \Omega \times \Omega \longrightarrow \Omega$ *is defined by* $\mathcal{F}_1 \oplus \mathcal{F}_2 = \mathcal{F}_1 \cdot \mathcal{F}_2$ *(notice that this definition satisfies the requirement* $\mathcal{C}(\mathcal{F}_1 \oplus \mathcal{F}_2) = \mathcal{C}(\mathcal{F}_1) \cap \mathcal{C}(\mathcal{F}_2)$. *The function* $\otimes : \Omega \times \Gamma \longrightarrow \Omega$ *is defined in the following way:*

$$\langle L_1, \ldots, L_n \rangle \otimes L = \langle L_1 \cup L, \ldots, L_n \cup L \rangle$$

Notice that with the previous identification we have $L \otimes L' = L \cup L'$, *that is $\otimes$ satisfies the requirement that its restriction to couples of Γ takes its values in Γ.*

With this definition we can state the following theorem:

Theorem 6.17 *The operator $\diamond_P$ is a syntactical update operator. More precisely, taking $\mathcal{L}$, Γ, Ω, $\mathcal{C}$ and $\otimes$ as in definition 6.16 the postulates* ***SU1, SU2, SU3, SU5, SU6*** *and* ***SU8*** *hold.*

Proof: **SU1**: We want to show that $\mathcal{C}(L') \subseteq \mathcal{C}(\mathcal{F} \diamond_P L')$, *i.e.* that $C^P_{fc}(L')$ is a subset of $C^P_{fc}(\mathcal{F} \diamond_P L')$. If L' is P-inconsistent or all the elements of $\mathcal{F}$ are P-inconsistent, the result follows trivially.

Now suppose that $\mathcal{F} \diamond_P L' = \langle L_1, \ldots, L_n \rangle$. By definition of $\diamond_P$, we have $L' \subseteq L_i$ for $i = 1, \ldots, n$. Therefore $L' \subseteq \bigcap_{i=1}^n C^P_{fc}(L_i) = C^P_{fc}(\mathcal{F} \diamond_P L')$. We conclude using observation 6.15

SU2: Suppose that $C^P_{fc}(L') \subseteq C^P_{fc}(\mathcal{F})$. We want to show that $C^P_{fc}(\mathcal{F}) = C^P_{fc}(\mathcal{F} \diamond_P L')$. If $C^P_{fc}(\mathcal{F}) = Lit$ or $C^P_{fc}(L') = Lit$ then the result follows trivially from definitions. Thus, assume that $C^P_{fc}(\mathcal{F}) \neq Lit$ and $C^P_{fc}(L') \neq Lit$ By the assumption, we can suppose that $\mathcal{F} = \langle L_1, \ldots, L_n \rangle$

and $\mathcal{F} \diamond_P L' = (L_{i_1} \diamond_P L') \cdots (L_{i_K} \diamond_P L')$ where the set $\{L_{i_1} \ldots L_{i_K}\}$ is the maximal subset of $\{L_1 \ldots L_n\}$ such that each L_{i_j} is P-consistent. Since $C^P_{fc}(L') \subseteq C^P_{fc}(L_{i_j})$ we have that $L_{i_j} \diamond_P L' = L_{i_j} \cup L'$ and by observation 6.15 that $C^P_{fc}(L_{i_j} \cup L') = C^P_{fc}(L_{i_j})$. Thus

$$C^P_{fc}(\mathcal{F} \diamond_P L') = C^P_{fc}(L_{i_1} \cup L', \ldots, L_{i_k} \cup L') = \bigcap_{j=1}^{k} C^P_{fc}(L_{i_j}) = \bigcap_{i=1}^{n} C^P_{fc}(L_i) = C^P_{fc}(\mathcal{F})$$

where the next to last equality is due to the fact that if L_i is different of all L_{i_j} then $C^P_{fc}(L_i) = Lit$.

SU3: It is straightforward by definition of $\diamond_P$.

SU5: We want to show that $C^P_{fc}(\mathcal{F} \diamond_P (L \otimes L')) \subseteq C^P_{fc}((\mathcal{F} \diamond_P L) \otimes L')$. First we prove the result when $\mathcal{F}$ is a flock with one element, say $\mathcal{F} = \langle H \rangle$. If H is P-inconsistent or $L \cup L'$ is P-inconsistent the result is quite straightforward. Now suppose that both H and $P \cup P'$ are P-consistent. By definition we have

$$\begin{aligned} H \diamond_P (L \otimes L') &= \langle L_1 \cup L \cup L', \ldots, L_n \cup L \cup L' \rangle \\ (H \diamond_P L) \otimes L' &= \langle K_1 \cup L, \ldots, K_p \cup L \rangle \otimes L' \\ &= \langle K_1 \cup L \cup L', \ldots, K_p \cup L \cup L' \rangle \end{aligned}$$

where L_i is a maximal subset of H such that $L_i \cup L \cup L'$ is P-consistent, for $i = 1, \ldots, n$, K_j is a maximal subset of H such that $K_j \cup L$ is P-consistent, for $j = 1, \ldots, p$. Notice that either $C^P_{fc}(K_j \cup L \cup L') = Lit$ or there exists an m such that $K_j = L_m$. Therefore if $l \in \bigcap_{i=1}^{n} C_{fc}(P \cup L_i \cup L' \cup L'')$ then $l \in \bigcap_{j=1}^{r} C_{fc}(P \cup K_{i_j} \cup L' \cup L'')$.

Now we prove the general case. Suppose $\mathcal{F} = \langle L_1, \ldots, L_n \rangle$. If $\mathcal{F}$ is P-inconsistent or $L \cup L'$ is P-inconsistent the result is straightforward. Thus assume that $\mathcal{F}$ and $P \cup P'$ are P-consistent. Then

$$\begin{aligned} \mathcal{F} \diamond_P (L \otimes L') &= (L_{i_1} \diamond_P (L \cup L')) \cdots (L_{i_k} \diamond_P (L \cup L')) \\ (\mathcal{F} \diamond_P L') \otimes L' &= ((L_{i_1} \diamond_P L \cdots L_{i_k} \diamond_P L)) \otimes L' \\ &= ((L_{i_1} \diamond_P L) \otimes L') \cdots ((L_{i_k} \diamond_P L) \otimes L') \end{aligned}$$

where $\{L_{i_1}, \ldots, L_{i_k}\}$ is the maximal subset of $\{L_1, \ldots, L_n\}$ such that each element is P-consistent. By the first case we have

$$C^P_{fc}((L_{i_j} \diamond_P (L \otimes L')) \subseteq C^P_{fc}((L_{i_j} \diamond_P L) \otimes L') \quad \text{for } j = 1, \ldots, k$$

Therefore $C^P_{fc}((\mathcal{F} \diamond_P (L \otimes L')) \subseteq C^P_{fc}((\mathcal{F} \diamond_P L) \otimes L')$.

SU6: Suppose $C^P_{fc}(L_2) \subseteq C^P_{fc}(\mathcal{F} \diamond_P L_1)$ and $C^P_{fc}(L_1) \subseteq C^P_{fc}(\mathcal{F} \diamond_P L_2)$. We want to show that $C^P_{fc}(\mathcal{F} \diamond_P L_1) = C^P_{fc}(\mathcal{F} \diamond_P L_2)$. When one of $\mathcal{F}$, L_1, L_2 is P-inconsistent the result is trivial. So suppose all three of them P-consistent.

First we suppose that $\mathcal{F} = L$. By the assumption it is clear that $L_2 \subseteq C^P_{fc}(L \diamond_P L_1)$ and $L_1 \subseteq C^P_{fc}(L \diamond_P L_2)$ Put

$$\begin{aligned} L \diamond_P L_1 &= \langle L^1_1, \ldots, L^1_{n_1} \rangle \otimes L_1 \\ L \diamond_P L_2 &= \langle L^2_1, \ldots, L^2_{n_2} \rangle \otimes L_2 \end{aligned}$$

where L^1_i is a maximal subset of L such that $L^1_i \cup L_1$ is P-consistent for $i = 1, \ldots, n_1$ and L^2_j is a maximal subset of L such that $L^2_j \cup L_2$ is P-consistent for $j = 1, \ldots, n_2$. From the hypothesis it is easy to see that:(a) $L_1 \cup L^2_j$ is P-consistent for $j = 1, \ldots n_2$ and(b) $L_2 \cup L^1_i$ is P-consistent for $i = 1, \ldots n_1$. From (b) we have $\forall\, i \in \{1, \ldots, n_1\}\ \exists\, j \in \{1, \ldots, n_2\}$ such that $L^1_i \subseteq L^2_j$ and from (a) we have $\forall\, j \in \{1, \ldots, n_2\}\ \exists\, i \in \{1, \ldots, n_1\}$ such that $L^2_j \subseteq L^1_i$. But this implies, by maximality of sets L^1_i and L^2_j, that $n_1 = n_2$ and there is a permutation σ of $\{1, \ldots, n_1\}$ such that $L^1_i = L^2_{\sigma(i)}$. Without loss of generality we can suppose that σ is the identity. Finally note that

$$\begin{aligned} C^P_{fc}(\langle L^1_1, \ldots, L^1_{n_1}\rangle \otimes L_1) &= C^P_{fc}(\langle L^1_1, \ldots, L^1_{n_1}\rangle \otimes (L_1 \cup L_2)) \\ &= C^P_{fc}(\langle L^2_1, \ldots, L^2_{n_1}\rangle \otimes (L_2 \cup L_1)) \\ &= C^P_{fc}(\langle L^2_1, \ldots, L^2_{n_1}\rangle \otimes L_2) \end{aligned}$$

The first and third equalities follow from the hypothesis and observation 6.15.

Now we prove the general case. Put $\mathcal{F} = \langle H_1, \ldots, H_n\rangle$ and suppose $C^P_{fc}(L_2) \subseteq C^P_{fc}(\mathcal{F} \diamond_P L_1)$ and $C^P_{fc}(L_1) \subseteq C^P_{fc}(\mathcal{F} \diamond_P L_2)$. Then

$$\mathcal{F} \diamond_P L_i \;=\; (H_{j_1} \diamond_P L_i) \cdots (H_{j_k} \diamond_P L_i) \quad i = 1, 2$$

But it is easy to see that $L_1 \subseteq C^P_{fc}(H_{j_m} \diamond_P L_2)$ and $L_2 \subseteq C^P_{fc}(H_{j_m} \diamond_P L_1)$ for $m = 1, \ldots, k$. Then, because of the first case, $C^P_{fc}(H_{j_m} \diamond_P L_1) = C^P_{fc}(H_{j_m} \diamond_P L_2)$ and therefore $C^P_{fc}(\mathcal{F} \diamond_P L_1) = C^P_{fc}(\mathcal{F} \diamond_P L_2)$.

SU8: It is trivially verified because of definitions. ∎

We will show now that $\circ_{rk}$ is a syntactical revision operator and that the operator $\circ_h$ has some of the properties of syntactical revision operators. In order to do that we give the instantiations of the sets $\mathcal{L}$, Ω and the functions C and $\otimes$ used in 6.1.

Definition 6.18 *$\mathcal{L} = Lit$. $\Omega = \Gamma = Prog$. Clearly $\mathcal{P}(\mathcal{L}) \subseteq \Omega$. The consequence operator $C : \Omega \longrightarrow \mathcal{P}(\mathcal{L})$ is defined by $C = C_{fc}$. The function $\otimes : \Omega \times \Omega \longrightarrow \Omega$ is defined by $P \otimes P' = P \cup P'$.*

Theorem 6.19 *The operator $\circ_{rk}$ is a syntactical revision operator. More precisely, taking $\mathcal{L}$, Ω, C and $\otimes$ like in definition 6.18, the postulates **SR1**, **SR2**, **SR3**, **SR5** and **SR6** hold.*

*The operator $\circ_h$ satisfies the postulates **SR1**, **SR2** and **SR3** but it does not satisfy **SR5** nor **SR6**.*

Proof: First we will do the verifications for $\circ_h$ concerning the postulates **SR1**, **SR2**, **SR3** (the postulates for $\circ_{rk}$ are verified in an analogous way). Then we will verify **SR5** and **SR6** for $\circ_{rk}$. Finally we will show counterexamples to **SR5** and **SR6** for $\circ_h$.

SR1: We want to show that $C_{fc}(P') \subseteq C_{fc}(P \circ_h P')$. This is clearly true because $P \circ_h P' = h(P') \cup P'$.

SR2: Suppose that $C(P \otimes P') \neq \mathcal{L}$, *i.e.* $C_{fc}(P \cup P') \neq Lit$. We want to show that $P \circ_h P' = P \cup P'$. This is straightforward because $C_{fc}(P \cup P') \neq Lit$ implies $\rho(P') = 0$.

SR3: Suppose $C(P') \neq \mathcal{L}$, *i.e.* P' is consistent. We want to show that $P \circ_h P'$ is also consistent, *i.e.* $C(P \circ_h P') \neq \mathcal{L}$. This is true because $P \circ_h P' = h(P') \cup P'$ and $h(P')$ is by definition contained in a set consistent with P'.

SR5 and **SR6** for $\circ_{rk}$: We suppose that $C((P \circ_{rk} P') \otimes P'') \neq \mathcal{L}$, *i.e.* $(P \circ_{rk} P') \cup P''$ is consistent (otherwise **SR5** is trivial). We want to prove that $C((P \circ_{rk} P') \otimes P'') = C(P \circ_{rk} (P' \otimes P''))$. In order to do that it is enough to show that $P \circ_{rk} (P' \cup P'') = (P \circ_{rk} P') \cup P''$. By hypothesis $(P_{\rho(P')}\cup P')\cup P''$ is consistent. Thus $\rho(P'\cup P'') \leq \rho(P')$ and then $\rho(P'\cup P'') = \rho(P')$. From this we conclude easily.

In order to show that **SR5** does not hold for $\circ_h$ it is enough to consider the following example: the program P is defined by $P = \{b \to w, w \to w', w' \to f, o \to b, o \to \neg f\}$. The base is in this case $\langle P_0, P_1, P_2\rangle$ with $P_0 = P$, $P_1 = \{o \to b, o \to \neg f\}$ and $P_2 = \emptyset$. Put $P' = \{o\}$ and $P'' = \{w'\}$. It is not hard to establish that $h(P') = P_1$ and $h(P' \cup P'') = P_1 \cup \{b \to w, w \to w'\}$. Thus $C_{fc}((P \circ_h P') \cup P'') = \{o, w', b, \neg f\}$ and $C_{fc}(P \circ_h (P' \cup P'')) = \{o, w', b, \neg f, w\}$. Therefore $C_{fc}(P \circ_h (P' \cup P'')) \not\subseteq C_{fc}((P \circ_h P') \cup P'')$, that is **SR5** does not hold.

To prove that **SR6** does not hold for $\circ_h$ we consider the following example: Put $P = \{r_0, r_1, r_2\}$ where $r_0 = a \to c$, $r_1 = e \to \neg c$, $r_2 = b \to \neg c$. Put $P' = \{a, e\}$ and $P'' = \{b\}$. The base for P is $\langle P, \emptyset\rangle$. Then it is easy to see that $P_{\rho(P')} = P_{\rho(P'\cup P'')} = \emptyset$ and

$$\begin{array}{lcl} I_P(P') & = & \{\{r_0, r_2\}, \{r_1, r_2\}\} \text{ and} \\ I_P(P' \cup P'') & = & \{\{r_0\}, \{r_1, r_2\}\} \end{array}$$

Thus $h(P') = \{r_2\}$ and $h(P' \cup P'') = \emptyset$. Therefore $\neg c \in C_{fc}((P \circ_h P') \cup P'')$ and $\neg c \notin C_{fc}(P \circ_h (P' \cup P''))$, so **R6** fails. ∎

Now, in order to analyse the postulates of syntactical revision satisfied by $\circ_{eh}$, we need to state in a very precise way what are the sets and functions of definition 6.1. This is the subject of the next definition.

Definition 6.20 *We put $\mathcal{L} = Lit$; Ω is the set of flocks of programs; $\Gamma = Prog$; notice that with the above identification we have $\Gamma \subseteq \Omega$; $\circ_{eh} : \Omega \times \Gamma \longrightarrow \Omega$ as defined above; the function $\otimes : \Omega \times \Gamma \longrightarrow \Omega$ is defined by:*

$$\langle Q_1, \ldots, Q_n\rangle \otimes P = \langle Q_1 \cup P, \ldots, Q_n \cup P\rangle$$

Notice that the restriction of $\otimes$ to couples of elements in Γ takes its values in Γ; and finally we define $\mathcal{C} : \Omega \longrightarrow \mathcal{P}(\mathcal{L})$ by putting $\mathcal{C} = C_{fc}$.

The extended hull revision operator satisfies some of our syntactical postulates. More precisely we have the following theorem:

Theorem 6.21 *The operator $\circ_{eh}$ satisfy **SR1**, **SR3** and **SR5**, when we take the definitions of 6.20. It satisfies a weak version of **SR2**: if P is consistent with all the elements of $\mathcal{F}$ then $\mathcal{F} \circ_{eh} P = \mathcal{F} \otimes P$.*

Proof: **SR1** is proved in an analogous way than the same postulate (**SU1**) for the operator $\diamond_P$ (see the proof of theorem 6.17).1.5mm] **SR3** and the weak form of **SR2** are straightforward from definitions.1.5mm] Now we prove **SR5**. First, we consider the case $\mathcal{F} = P$. Then we want to prove that $\mathcal{C}(P \circ_{eh} (P' \otimes P'')) \subseteq \mathcal{C}((P \circ_{eh} P') \otimes P'')$. Suppose that

$$\begin{array}{l} P \circ_{eh} P' = \langle Q_1 \cup P', \ldots, Q_n \cup P'\rangle \\ P \circ_{eh} (P' \otimes P'') = P \circ_{eh} (P' \cup P'') = \langle H_1 \cup P' \cup P'', \ldots, H_k \cup P' \cup P''\rangle \end{array}$$

Then $(P \circ_{eh} P') \otimes P'' = \langle Q_1 \cup P' \cup P'', \ldots, Q_n \cup P' \cup P''\rangle$. If $C_{fc}((P \circ_{eh} P') \otimes P'') = Lit$ we are done. Otherwise there is a Q_i such that $Q_i \cup P' \cup P''$ is consistent. But since Q_i is a subset of P

maxiconsistent with P' and containing $P_{\rho(P')}$ necessarily $\rho(P') = \rho(P' \cup P'')$. Now we claim that for all $i = 1, \ldots, n$ either $Q_i \cup P' \cup P''$ is inconsistent or there is a $j \leq k$ such that $Q_i = H_j$. To see that suppose that $Q_i \cup P' \cup P''$ is consistent then Q_i is a subset maximal consistent with $P' \cup P''$ containing $P_{\rho(P')} = P_{\rho(P' \cup P'')}$, *i.e.* $Q_i = H_j$ for a j, by definition of sets H_q for $q = 1, \ldots, k$. Finally from the claim we get easily $\bigcap_{j=1}^{k} C_{fc}(H_j \cup P' \cup P'') \subseteq \bigcap_{i=1}^{n} C_{fc}(Q_i \cup P' \cup P'')$, *i.e.* $C(P \circ_{eh} (P' \otimes P'')) \subseteq C((P \circ_{eh} P') \otimes P'')$.

The general case, when $\mathcal{F} = \langle P_1, \ldots, P_n \rangle$, follows from the first case by definition of $\circ_{eh}$ and using the same trick that we used in the proof of **SU5** in theorem 6.17. ∎

Observation 6.22 *The postulates **SR2**, **SU2**, **SU6** and **SR6** don't hold for* $\circ_{eh}$.

Proof: We build counterexamples for each of those postulates.

For **SR2**: Take $\mathcal{F} = \langle \{a\}, \{\neg b\} \rangle$ and $P = \{b\}$. Then

$$C_{fc}(\mathcal{F} \otimes P) = C_{fc}(\langle \{a, b\}, \{\neg b, b\} \rangle) = \{a, b\}$$

So $\mathcal{F} \otimes P$ is consistent. But $C_{fc}(\mathcal{F} \circ_{eh} P) = C_{fc}(\langle \{a, b\}, \{b\} \rangle) = \{b\}$; so **SR2** fails.

For **SU2**: Take $P = \{a\}$ and $P' = \{a \rightarrow b\}$. Then $\emptyset = C_{fc}(P') \subseteq C_{fc}(P) = \{a\}$. But $C_{fc}(P \circ_{eh} P') = C_{fc}(P \cup P') = \{a, b\}$; so **SU2** fails.

For **SU6**: Take $P_1 = \{a \rightarrow b\}$, $P_2 = \{c \rightarrow d\}$ and $Q = \{a\}$. Clearly $C_{fc}(P_1) = C_{fc}(P_2) = \emptyset$, so the hypothesis of **SU6** is true. But

$$C_{fc}(Q \circ_{eh} P_1) = C_{fc}(Q \cup P_1) = \{a, b\} \neq \{a\} = C_{fc}(Q \cup P_2) = C_{fc}(Q \circ_{eh} P_2)$$

For **SR6**: The same counterexample given in the proof of theorem 6.19 to show that **SR6** fails for $\circ_h$, works in this case. ∎

Observation 6.23 *All the operators previously defined are syntax-sensitive,* i.e. *(IS) fails for the operators* $\diamond_P$, $\circ_{rk}$, $\circ_h$ *and* $\circ_{eh}$.

Proof: First we give a counterexample for $\diamond_P$. Put $P = \{a \rightarrow b\}$. Define $L_1 = \{a, b\}$ and $L_2 = \{a\}$. Put $L' = \{\neg a\}$. Clearly $C^P_{fc}(L_1) = C^P_{fc}(L_2) = \{a, b\}$. It is easy to see that $L_1 \diamond_P L' = \{b, \neg a\}$. Thus $C^P_{fc}(L_1 \diamond_P L') = \{b, \neg a\}$. But it is easy to see that $L_2 \diamond_P L' = \{\neg a\}$ so $C^P_{fc}(L_2 \diamond_P L') = \{\neg a\}$. Therefore $C^P_{fc}(L_1 \diamond_P L') \neq C^P_{fc}(L_2 \diamond_P L')$.

Now we give a counterexample for $\circ_{rk}$, $\circ_h$ and $\circ_{eh}$. Put $P_1 = \{a \rightarrow b\}$, $P_2 = \{a \rightarrow c\}$ and $P' = \{a\}$. Then $C_{fc}(P_1) = \emptyset = C_{fc}(P_2)$. But $C_{fc}(P_1 \circ P') = C_{fc}(P_1 \cup P') = \{a, b\} \neq \{a\} = C_{fc}(P_2 \cup P') = C_{fc}(L_2 \circ P')$ for any $\circ \in \{\circ_{rk}, \circ_h, \circ_{eh}\}$. ∎

6.4 Still another operator: selection hull

In the previous section we have seen that the two sceptical approaches to extend the ranked revision fail to be syntactical revision operators. In this section we give another extension of ranked revision based on the idea of using a selection function.

Let S be a function mapping sets of programs into programs, *i.e.* $S : \mathcal{P}(Prog) \longrightarrow Prog$. We will say that S is a selection function iff the following properties hold:

(i) $S(\emptyset) = \emptyset$

(ii) If $D \neq \emptyset$ then $S(D) \in D$

Let us remark that this kind of selection functions are known as maxichoice functions in the literature [1, 36].

Let P and P' be two programs. Let $I_P(P')$ be as defined in section 6.2.2, *i.e.* the set of subsets of P which are maxiconsistent with P' and which contain $P_{\rho(P')}$. Let S be a selection function. We define the operator $\circ_{sh}$ by the following:

$$P \circ_{sh} P' = S(I_P(P')) \cup P'$$

We will take the same instantiations as in definition 6.18 in order to analyse the postulates satisfied by $\circ_{sh}$.

The following definition gives us a class of selection functions for which the operator $\circ_{sh}$ is a syntactical revision operator.

Definition 6.24 *A selection function S is said to be sensible iff the following property holds: for any programs P, P' and P''*

$$I_P(P') \cap I_P(P' \cup P'') \neq \emptyset \;\Rightarrow\; S(I_P(P')) = S(I_P(P' \cup P''))$$

Theorem 6.25 *If S is an sensible selection function then $\circ_{sh}$ is a syntactical revision operator when we consider $\mathcal{L}$, Ω, Γ, $\mathcal{C}$ and $\otimes$ as in definition 6.18.*

Proof: **SR1** and **SR3** follow easily from definition of $\circ_{sh}$. The verification of **SR3** is also easy using the property (ii) of selection function. Now we prove **SR5** and **SR6**. Actually with the hypothesis on S we are going to prove that if $(P \circ_{sh} P') \cup P''$ is consistent then $(P \circ_{sh} P') \cup P'' = P \circ_{sh} (P' \cup P'')$ which is obviously stronger than **SR5** and **SR6**.

Suppose that $I(P') = \{Q\} \cup D_1$ and $S(I(P')) = Q$. By hypothesis $Q \cup P' \cup P''$ is consistent so $\rho(P') = \rho(P' \cup P'')$ and $Q \in I(P' \cup P'')$. Therefore $I(P' \cup P'') = \{Q\} \cup D_2$. Since $\rho(P') = \rho(P' \cup P'')$ and $P' \subseteq P' \cup P''$ we have that for any R in D_2 there exists R' in D_1 such that $R \subseteq R'$. Then by the property assumed for S we have $S(I(P' \cup P'')) = Q$; and from this we conclude. ∎

We remark that the property required for S in definition 6.24 can be explained in an intuitive way: if you are choosing among elements of $\{Q\} \cup D_1$ and your preference is Q, it means that you think Q is the set which best fits P, then in the situation where you are choosing among elements of $\{Q\} \cup D_2$ with elements of D_2 contained in elements of D_1, you must reasonably choose Q.

Notice also that with this definition $\circ_{sh}$ is a conservative extension of $\circ_{eh}$, *i.e.* $C_{fc}(P \circ_{eh} P') \subseteq C_{fc}(P \circ_{sh} P')$. Thus we have

$$C_{fc}(P \circ_{rk} P') \subseteq C_{fc}(P \circ_{h} P') \subseteq C_{fc}(P \circ_{eh} P') \subseteq C_{fc}(P \circ_{sh} P')$$

A natural question one can ask is if there are selection functions with the property required in theorem 6.25. We show next a method for building such selection functions.

6.4.1 Building sensible selection functions

An easy way to build a sensible selection function is to use a linear ordering among propositions, that codes agent preferences in the program.

Let $|Q|$ be the cardinality of a set Q. Let $\{r_1, \ldots, r_n\}$ be an enumeration without repetition of rules and facts. Let π be a function (weighting) $\pi : \{r_1, \ldots, r_n\} \longrightarrow \omega$ such that $\pi(r_i) = 2^i$. We extend in a natural way π to $\mathcal{P}(Prog)$ by putting

$$\pi(\{r_{i_1}, \ldots, r_{i_k}\}) = \sum_{j=1}^{k} 2^{i_j}$$

Notice that if $P, P' \in Prog$ and $P \neq P'$ then $\pi(P) \neq \pi(P')$.

Let $<_\ell$ the lexicographical order on ω^2. Then define $S : \mathcal{P}(Prog) \longrightarrow Prog$ by $S(\emptyset) = \emptyset$ and if D is nonempty

$$S(D) = Q \text{ iff } Q \in D \text{ and } \forall R \in D\ (R \neq Q \ \Rightarrow\ (|R|, \pi(R)) <_\ell (|Q|, \pi(Q))\,)$$

That is S chooses among the sets of greatest cardinality the set with higher weighting.

Because of definition of π it is quite easy to see that S is a sensible selection function.

Conclusion

We have proposed in this paper a methodological framework in order to analyse rational properties for syntactical change operators. We have also introduced five knowledge change operators based on forward chaining. The ideas behind the definition of these operators are very natural and simple. In most of the cases, they are connected with well known methods [27, 5, 71]. The originality of our work relies on the definition of the rank function for the revision operators and concerning the factual update on the fact that we consider the flocks with a logical content: their consequences by forward chaining. This point -endowing the result of an operator with a logical content- is particularly important because it makes possible the analysis of the operators in our abstract framework.

The following table summarizes the main results:

Operator	Satisfied postulates	Unsatisfied postulates
$\diamond_P$	**SU1+SU2+SU3+SU5+SU6+SU7+SU8**	
$\circ_{rk}$	**SR1+SR2+SR3+SR5+SR6**	
$\circ_h$	**SR1+SR2+SR3**	**SR5+SR6**
$\circ_{eh}$	**SR1+SR3+SR5**	**SR2+SU2+SR6+SU6**
$\circ_{sh}$	**SR1+SR2+SR3+SR5+SR6**	

Thus three of our operators have desirable properties. In particular factual update is a syntactical update operator according to our version of Katsuno and Mendelzon postulates. Ranked revision and selection hull are, syntactical revision operators according to our version of Alchourrón-Gärdenfors-Makinson postulates. The other operators, hull and extended hull revision, don't have good rational properties. Nevertheless they extend ranked revision in order to keep more information from the old program and thus seem to be less drastic than ranked revision.

These operators based on forward chaining are easily computable. Notice that, in particular, the operator of ranked revision is polynomial. The factual update operator, from a complexity point of view, is exponential (actually it is NP-complete since it requires the computation of all minimal hitting sets).

The operators based on hull revision are more complicated but can be computed with the help of the two others.

An interesting further work is to investigate the properties of our operators with respect to some relativisation of iteration postulates [18, 53]. Another interesting point to develop is the extension of these operators to the first order. By the way, our results can be translated in an obvious way to Datalog without negation by failure.

Chapitre 7

On the logic of merging

This work proposes an axiomatic characterization of merging operators. It underlines the differences between arbitration operators and majority operators. A representation theorem is stated showing that each merging operator corresponds to a family of partial pre-orders on interpretations. Examples of operators are given. They show the consistency of the axiomatic characterization. A new merging operator Δ_{GMax} is provided. It is proved that it is actually an arbitration operator.

7.1 Introduction

In a growing number of applications, we face conflicting information coming from several sources. The problem is to reach a coherent piece of information from these contradicting ones. A lot of different merging methods have already been given [10, 62, 4, 5, 108]. Instead of giving one particular merging method we propose, in this paper, a characterization of such methods following the rationality of the postulates they satisfy. We shall call merging operators those methods that obey a minimal set of rational merging postulates. Then we shall investigate two subclasses of merging operators: arbitration operators and majority operators.

Merging operators are useful in a lot of applications: to find a coherent information in a distributed database system, to solve a conflict between several people or several agents, to find an answer in a decision-making committee, to take a decision when information given by some captors is contradictory, etc.

This work is related to the AGM (Alchourrón, Gärdenfors, Makinson) framework of revision theory [1, 36, 45]. Revision is the process of according a knowledge base in the view of a new evidence. One basic assumption of revision is that the new information is more reliable that the knowledge base, but it is not always the case. We can distinguish 3 cases:

- *The new piece of information is more reliable than the knowledge base:* it is the assumption made in the revision theory so we can revise our knowledge base by the new piece of information.
- *The new piece of information is less reliable than the knowledge base:* a drastic point of view could be to ignore this unreliable piece of information but if we want to be more constructive we can take this piece of information into account if it is consistent with the knowledge base and ignore it only if it is inconsistent with our belief. Another interesting

way would be to reverse the revision, i.e. to revise the new piece of information by the knowledge base.

- *The new piece of information is as reliable as the knowledge base:* here we can't give the preference to one of the two items of knowledge, so we have to find something else. This is the aim of merging operators.

The intuitive difference between arbitration and majority operators is that arbitration operators reach a consensus between the protagonists' views by trying to satisfy as much as possible all the protagonists, whereas majority operators elect, in a sense, the result of the merging by taking the majority into account. In other words arbitration operators try to minimize individual dissatisfaction, whereas majority operators try to minimize global dissatisfaction. One of our main concerns in this work is to state these intuitions in a formal way.

Some operators quite close to merging operators have already been formally studied. Revesz defined in [99, 100] model-fitting operators which can be considered as a generalization of revision for multiple knowledge bases. Revesz also defined arbitration operators from model-fitting operators. We make a criticism about Revesz's postulates: they do not distinguish between majority and arbitration.

Liberatore and Schaerf have proposed postulates to characterize arbitration [58, 59]. Their definition has a strong connection with revision operators, but the major drawback, in our opinion, is that those operators arbitrate only two knowledge bases. Furthermore they select some interpretations in the two knowledge bases as the result of the arbitration. We consider that we can't ignore interpretations which do not belong to these knowledge bases, consider the following example:

Example 7.1 *Suppose that we want to speculate on the stock exchange. We ask two financial experts about four shares A,B,C,D. We denote 1 if the share rises and 0 if it falls (we suppose that its value can't be stable). These agents have the same expert level and so they are both equally reliable. The first one says that all the shares will rise:* $\varphi_1 = \{(1,1,1,1)\}$*, the second one thinks that all the shares will fall:* $\varphi_2 = \{(0,0,0,0)\}$*. The Liberatore and Schaerf operators will arbitrate these opinions and give the following result:* $R = \{(0,0,0,0),(1,1,1,1)\}$*. So it means that either* φ_1 *is totally wrong or it's* φ_2 *who is completely mistaken. But intuitively, if the two experts are equally reliable, there is no reason to think that one of them has failed more than the other: they both have to be at the same "distance" of the truth. So they are certainly both wrong on two shares and the result has to be:* $R' = \{(0,0,1,1),(0,1,0,1),(0,1,1,0),(1,0,0,1),(1,0,1,0),\ (1,1,0,0)\}$*. So two of the shares will rise and two will fall but we don't know which ones.*

In our opinion Liberatore and Schaerf's operators have to be seen as selection operators and have to be used in applications which require the result be one of the possibilities given by the protagonists. For example, if the result of the arbitration is a medical treatment, we can't "merge" several therapies and so we have to use Liberatore and Schaerf operators. Liberatore and Schaerf's operators take, in a sense, the interpretation as unit of change, we propose to take the propositional variable as such a unit, as Dalal says in [15]: "*Change in truth value of a single symbol can be considered as the smallest unit of change*", we want to apply this to arbitration.

Lin and Mendelzon proposed a *theory merging by majority* operator [62, 61] which solves conflicts between knowledge bases by taking the majority into account. Their *theory merging operators* are what we call majority operators.

The paper is organized as follows: in section 2 we give some definitions and state some notations. In section 3 we propose postulates for merging operators, majority operators and arbitration operators and we study the relationships between some of the postulates. In section 4 we give a model-theoretic characterization of those operators. In section 5 we give some examples of merging operators, especially we show that an operator, called Δ_{GMax}, is an arbitration operator. Finally, in section 6 we give some conclusions and discuss open problems.

7.2 Preliminaries

We consider a propositional language $\mathcal{L}$ over a finite alphabet $\mathcal{P}$ of propositional letters. An interpretation is a function from $\mathcal{P}$ to $\{0, 1\}$. The set of all the interpretations is denoted $\mathcal{W}$. An interpretation I is a model of a formula if and only if it makes it true in the usual classical truth functional way. Let φ be a formula, $Mod(\varphi)$ denote the set of models of φ. And let M be a set of interpretations, $form(M)$ denote a formula which set of models is M. When $M = \{I\}$ we will use the notation $form(I)$ for reading convenience.

A *knowledge base* K is a finite set of propositional formulae which can be seen as the formula φ which is the conjunction of the formulae of K. By abuse, we will use K to denote the formula φ. We will note K_I a knowledge base the sole model is I.

Let $K_1, \ldots, K_n$ be n knowledge bases (not necessarily different). We call *knowledge set* the multi-set E consisting of those n knowledge bases: $E = \{K_1, \ldots, K_n\}$. We note $\bigwedge E$ the conjunction of the knowledge bases of E, i.e. $\bigwedge E = K_1 \wedge \cdots \wedge K_n$. The union of multi-sets will be noted $\sqcup$.

Remark 7.2 *Since an inconsistent knowledge base gives no information for the merging process, we'll suppose in the rest of the paper that the knowledge bases are consistent.*

$\mathcal{K}$ will denote the set of consistent knowledge bases and $\mathcal{E}$ will denote the set of non empty finite multi-sets with elements in $\mathcal{K}$.

Let's denote $\mathcal{S}$ the set of sets of interpretations without the empty set, *i.e.* $\mathcal{S} = \mathcal{P}(\mathcal{W}) \setminus \{\emptyset\}$; and let's denote $\mathcal{M}$ the set of finite non empty multi-sets with elements in $\mathcal{S}$. Elements of $\mathcal{S}$ and $\mathcal{M}$ will be denoted by the letters S and M respectively with possibly subscripts. So a typical element $M \in \mathcal{M}$ will be of the shape $\{S_1, \ldots, S_n\}$. Let $M = \{S_1, \ldots, S_n\}$, we define $\bigcap M$ in the usual way: $I \in \bigcap M$ iff $\forall S_i \in M \quad I \in S_i$.

Definition 7.3 *A knowledge set E is consistent if and only if $\bigwedge E$ is consistent. We will use $Mod(E)$ to denote $Mod(\bigwedge E)$.*

Definition 7.4 *Let E_1, E_2 be two knowledge sets. E_1 and E_2 are equivalent, noted $E_1 \leftrightarrow E_2$, iff there exists a bijection f from $E_1 = \{K_1^1, \ldots, K_n^1\}$ to $E_2 = \{K_1^2, \ldots, K_n^2\}$ such that $\vdash f(K) \leftrightarrow K$.*

Note that the relation $\leftrightarrow$ is an equivalence relation on knowledge sets. As usual, we denote by $\mathcal{E}/\leftrightarrow$ the quotient of $\mathcal{E}$ by the relation $\leftrightarrow$. Thus the function $\imath : \mathcal{E}/\leftrightarrow \longrightarrow \mathcal{M}$, defined by $\imath([\{K_1, \ldots, K_n\}]_\leftrightarrow) = \{Mod(K_1), \ldots, Mod(K_n)\}$ is a bijection. By abuse we will write $\imath(E)$ instead of $\imath([E]_\leftrightarrow)$.

A pre-order over $\mathcal{W}$ is a reflexive and transitive relation on $\mathcal{W}$. Let $\leq$ be a pre-order over $\mathcal{W}$, we define $<$ as follows: $I < J$ iff $I \leq J$ and $J \not\leq I$. And $\simeq$ as $I \simeq J$ iff $I \leq J$ and $J \leq I$. Let I be an interpretation, we wrote $I \in min(\leq)$ iff $\nexists J \in \mathcal{W}$ s.t. $J < I$.

By abuse if R is in $\mathcal{K}$ (respectively in $\mathcal{S}$) then R will denote also the multi-set $\{R\}$ which is in $\mathcal{E}$ (resp. in $\mathcal{M}$). For a positive integer n we will denote R^n the multi-set $\{\underbrace{R, \ldots, R}_{n}\}$. Thus $R^n = \underbrace{R \sqcup \ldots \sqcup R}_{n}$.

An operator Δ will be a function mapping knowledge sets into knowledge bases. In the rest of the paper we will distinguish between *operator* and *merging operator*: the former when no special properties are satisfied the later to indicate that the operator satisfies the postulates of definition 7.5. Let K, E and Δ be a knowledge base, a knowledge set and an operator respectively. We define the sequence $\langle \Delta^n(E,K) \rangle_{n \geq 1}$ by the following:

$$\Delta^1(E,K) = \Delta(E \sqcup K)$$
$$\text{and} \quad \Delta^{n+1} = \Delta(\Delta^n(E,K) \sqcup K)$$

7.3 Postulates

In this section, we are going to propose a characterization of merging operators, i.e. we give a minimal set of properties an operator has to satisfy in order to have a rational behaviour concerning the merging. Let E be a knowledge set, and let Δ be an operator which assigns to each knowledge set E a knowledge base $\Delta(E)$.

Definition 7.5 *Δ is a merging operator if and only if it satisfies the following postulates:*

(A1) *$\Delta(E)$ is consistent*

(A2) *If E is consistent, then $\Delta(E) = \bigwedge E$*

(A3) *If $E_1 \leftrightarrow E_2$, then $\vdash \Delta(E_1) \leftrightarrow \Delta(E_2)$*

(A4) *If $K \wedge K'$ is not consistent, then $\Delta(K \sqcup K') \not\vdash K$*

(A5) *$\Delta(E_1) \wedge \Delta(E_2) \vdash \Delta(E_1 \sqcup E_2)$*

(A6) *If $\Delta(E_1) \wedge \Delta(E_2)$ is consistent, then $\Delta(E_1 \sqcup E_2) \vdash \Delta(E_1) \wedge \Delta(E_2)$*

These six postulates are the basic properties a merging operator has to satisfy, the intuitive meaning of the postulates is easy to understand: we always want to extract a piece of information from the knowledge set, what is forced by $(A1)$ (Notice that, as assumed in remark 7.2, all the knowledge bases of the knowledge set are consistent). If all the knowledge bases agree on some alternatives, $(A2)$ assures that the result of the merging will be the conjunction of the knowledge bases. $(A3)$ states that the operator Δ obeys a principle of irrelevance of syntax, i.e. if two knowledge sets are equivalent in the sense of definition 7.4, then the two knowledge bases resulting from the merging will be logically equivalent. $(A4)$ is the fairness postulates, the point is that when we merge two knowledge bases, merging operators must not give preference to one of them. We will see (theorem 7.11) that (A4) is the clue for distinguishing arbitration operators from majority operators. $(A5)$ expresses the following idea: if a group E_1 compromises on a set of alternatives which I belongs to, and another group E_2 compromises on an another set of alternatives which contains I, so I has to be in the chosen alternatives if we join the two groups. ($A5$) and ($A6$) together state that if you could find two subgroups which agree on at least one alternative, then the result of the global arbitration will be exactly those alternatives the two groups agree on. The postulates ($A5$) and ($A6$) have been given in [100] by Revesz for weighted model fitting operators.

Observation 7.6 *By definition, merging operators are commutative, i.e. the result of a merging does not depend on any order of elements of the knowledge set.*

Let's now turn our attention to the difference between majority and arbitration operators. We give here a postulate that renders the behaviour of majority operators, that is to say that if an opinion has a large audience, then it will be the opinion of the group:

$$(M7) \quad \forall K \; \exists n \; \Delta(E \sqcup K^n) \vdash K$$

Thus we define majority operators by the following:

Definition 7.7 *A merging operator is a majority operator if it satisfies* ($M7$).

Besides, arbitration operators are those operators which are, in a large extent, majority insensitive. We first give a postulate which seems to be a good characterization of arbitration operator:

$$(A7') \quad \forall K \; \forall n \; \Delta(E \sqcup K^n) = \Delta(E \sqcup K)$$

This postulate states that the result of an arbitration is fully independent from the frequency of different views. Unfortunately the set of postulates $\{A1, \ldots, A6, A7'\}$ is not consistent. The proof of this result has been pointed out by P. Liberatore (personal communication):

Theorem 7.8 *There is no merging operator satisfying* ($A7'$).

Proof: Let $E_1 = \{K, \neg K\}$ and $E_2 = \{K\}$ be two knowledge sets. By ($A7'$) we have that $\Delta(E_1 \sqcup E_2) = \Delta(E_1)$. By ($A4$) we have also that $\Delta(E_1) \not\vdash K$ and $\Delta(E_1) \not\vdash \neg K$. Furthermore by ($A2$) we deduce $\Delta(E_2) = K$. So $\Delta(E_1) \wedge \Delta(E_2)$ is consistent and by ($A6$) we have $\Delta(E_1 \sqcup E_2) \vdash \Delta(E_1) \wedge \Delta(E_2)$, it can be rewritten as $\Delta(E_1) \vdash \Delta(E_1) \wedge K$. Then $\Delta(E_1) \vdash K$, which contradicts ($A4$). ∎

Thus if we want to have a postulate expressing majority insensitivity while being consistent with ($A1 - A6$) we must weaken ($A7'$). We propose the following alternative:

$$(A7) \quad \forall K' \; \exists K \; K' \not\vdash K \; \forall n \; \Delta(K' \sqcup K^n) = \Delta(K' \sqcup K)$$

($A7$) states that, to a large extent, the result of the arbitration is independent from the frequency of the different views.

And we define arbitration operator in the following way:

Definition 7.9 *A merging operator is an arbitration operator if it satisfies* ($A7$).

Now we investigate some relations between the postulates.

Theorem 7.10 *If an operator satisfies* $(A1)$, *then it can't satisfy both* $(A7')$ *and* $(M7)$.

Proof: From $(A7')$ and $(M7)$ we deduce that for any arbitrary E

$$\forall\ K\ \Delta(E \sqcup K) \vdash K \quad (*)$$

Take K' such that $K \wedge K' \vdash \bot$. Now putting $E = K'$, by $(*)$, we have $\Delta(K' \sqcup K) \vdash K$. In a symmetrical way we have $\Delta(K \sqcup K') \vdash K'$ so $\Delta(K \sqcup K') \vdash K \wedge K'$ and then $\Delta(K \sqcup K') \vdash \bot$ which contradicts $(A1)$. ∎

A merging operator can't be an arbitration operator and a majority operator, more precisely we have the following:

Theorem 7.11 *If an operator satisfies* $(A4)$, *then it can't satisfy both* $(A7)$ *and* $(M7)$.

Proof: From $(A7)$ and $(M7)$ we deduce easily $\forall K'\ \exists K\ K' \not\vdash K\ \Delta(K' \sqcup K) \vdash K$. Let's choose $K' = K_I = form(I)$, then $\exists K\ K_I \not\vdash K\ \Delta(K_I \sqcup K) \vdash K$. But $K_I \not\vdash K$ is equivalent to $K_I \wedge K \vdash \bot$ and so by $(A4)$ we have that $\Delta(K_I \sqcup K) \not\vdash K$. Contradiction. ∎

So, although it seems very weak, the fairness postulate $(A4)$ play a very important role, since it allows us to differentiate arbitration operators and majority operators.

In addition to these basic postulates we can find various other properties, we investigate some of them below.

An interesting property for a merging operator is the following which we call the *iteration* property:

(A_{it}) $\quad \exists n \Delta^n(E, K) \vdash K$

The intuitive idea is that, since the merging operators give, in a sense, the average knowledge of a knowledge set, if we always take the result of a merging and iterate with the same knowledge base, we have to reach this knowledge base after enough iterations. But, even if it seems to be a reasonable requirement, we don't know if all merging operators obey (A_{it}), more exactly we suspect that those operators satisfying (A_{it}) are topological operators, *i.e.* operators defined from a distance.

Now let's turn our attention to the two properties of associativity and monotony. We claim that they are not desirable for merging operators and we show that merging operators do not satisfy any of them. First let's give a formal definition of associativity and monotony:

(Ass) $\quad \Delta(E_1 \sqcup \Delta(E_2)) = \Delta(E_1 \sqcup E_2)$

Associativity seems to be an interesting property since it would allow sub-merging within the knowledge set. So merging could be implemented more easily and more efficiently.

(Mon) $\quad$ If $K_1 \vdash K_1', \ldots, K_n \vdash K_n'$ then $\Delta(K_1 \sqcup \ldots \sqcup K_n) \vdash \Delta(K_1' \sqcup \ldots \sqcup K_n')$

The monotony property expresses that if a knowledge set E_1 is "stronger" than a knowledge set E_2, then the merging of E_1 has to be logically stronger than the merging of E_2.

Theorem 7.12 *If an operator satisfies (A2) and (A4), then it doesn't satisfy (Mon).*

Proof: Let I, J be two different interpretations. Let $K_1 = K_1' = form(I)$, $K_2 = form(J)$, and $K_2' = form(I, J)$, so we have $K_1 \vdash K_1'$ and $K_2 \vdash K_2'$. From $(A2)$ $\Delta(K_1' \sqcup K_2') = form(I)$ and from $(A4)$ $\Delta(K_1 \sqcup K_2) \not\vdash form(I)$. So we have $\Delta(K_1 \sqcup K_2) \not\vdash \Delta(K_1' \sqcup K_2')$. ∎

So it is clear that monotony is not satisfied by merging operators, it is not exactly the same with associativity, we show that it is not satisfied by majority operators and that it is not compatible with the iteration property:

Theorem 7.13 *If an operator satisfies (A2) (A4) and (M7), then it can't satisfy (Ass).*

Proof: Let's take K_I and K_J two different complete formulae, by $(M7)$ we have that $\exists n \Delta(K_I \sqcup K_J^n) \vdash K_J$. By (Ass) we have that $\Delta(K_I \sqcup K_J^n) = \Delta(K_I \sqcup \Delta(K_J^n))$. But by $(A2)$ we have $\Delta(K_J^n) = K_J$. So we obtain that $\Delta(K_I \sqcup K_J) \vdash K_J$. What contradicts $(A4)$. ∎

Theorem 7.14 *If an operator satisfies (A2) and (A4), then it can't satisfy both (A_{it}) and (Ass).*

Proof: (A_{it}) $\exists n\ \Delta^n(E, K) \vdash K$, but by (Ass) we find that $\Delta^n(E, K) = \Delta(E \sqcup K^n) = \Delta(E \sqcup \Delta(K^n))$ and by (A2) we have that $\Delta(E \sqcup \Delta(K^n)) = \Delta(E \sqcup K)$. So we have that $\Delta(E \sqcup K) \vdash K$, what, taking $E = K'$ with $K' \wedge K \vdash \bot$, contradicts (A4). ∎

So, if we want some additional property for a merging operator, we have to choose between iteration and associativity. We claim that iteration is a desirable property for merging operators, so associativity is not.

7.4 Semantical characterizations

In this section we give a model-theoretic characterization of merging operators first in terms of functions on sets of interpretations and then in terms of family of orders. More exactly we show that each merging operator corresponds to a function from multi-sets of sets of interpretations to sets of interpretations and then we show that each merging operator corresponds to a family of partial pre-orders on interpretations. The semantical characterization of the merging operators in terms of pre-orders is very close to the axiomatic characterization. This is due to the fact that we can't have a definition of the pre-order as subtle as in the case of belief revision. But this semantical characterization is very useful in the proofs and is a starting point for generalizing merging operators (*e.g.* when one considers the set of alternatives as a parameter).

First we define what is a merging function:

Definition 7.15 *A function $\delta : \mathcal{M} \longrightarrow \mathcal{S}$ is said to be a merging function if the following properties hold for any $M, M_1, M_2 \in \mathcal{M}$ and $S, S' \in \mathcal{S}$:*

1. *If $I \in \bigcap M$, then $I \in \delta(M)$*
2. *If $\bigcap M \neq \emptyset$ and $I \notin \bigcap M$, then $I \notin \delta(M)$*

3. If $S \cap S' = \emptyset$, *then* $\delta(S \sqcup S') \not\subseteq S$

4. If $I \in \delta(M_1)$ *and* $I \in \delta(M_2)$, *then* $I \in \delta(M_1 \sqcup M_2)$

5. If $\delta(M_1) \cap \delta(M_2) \neq \emptyset$ *and* $I \notin \delta(M_1)$, *then* $I \notin \delta(M_1 \sqcup M_2)$

A majority merging function is a merging function that satisfies the following:

6. $\forall M \in \mathcal{M}\ \forall S \in \mathcal{S}\ \exists n\ \delta(M \sqcup S^n) \subseteq S$

A fair merging function is a merging function that satisfies the following:

7. $\forall S' \in \mathcal{S}\ \exists S \in \mathcal{S}\ S' \not\subseteq S\ \forall n\ \delta(S' \sqcup S^n) = \delta(S' \sqcup S)$

It is easy to see, via the bijection $\imath$ of section 7.2 that the properties $1-5$ are the semantical counterparts of postulates $(A1-A6)$ (notice that postulate (A_1) corresponds to the fact $\emptyset \notin \mathcal{S}$), property 6 corresponds to postulate $(M7)$ and property 7 corresponds to postulate $(A7)$. More precisely we have the following representation theorem which proof is straightforward:

Theorem 7.16 *An operator* Δ *is a merging operator (it satisfies* $(A1-A6)$*) if and only if there exists a merging function* $\delta : \mathcal{M} \longrightarrow \mathcal{S}$ *such that*

$$Mod(\Delta(E)) = \delta(\imath(E)).$$

Furthermore Δ *is a majority merging operator iff* δ *is a majority merging function; and* Δ *is an arbitration operator iff* δ *is a fair merging function.*

As in the AGM framework for revision, we can suppose the existence of some relation which intuitively represents how credible each interpretation is for some given knowledge set. We will see that there is a close relationship between merging function and these relations on knowledge sets. First we define what a syncretic assignment is:

Definition 7.17 *A syncretic assignment is an assignment which maps each knowledge set* E *to a pre-order* $\leq_E$ *over interpretations such that for any* $E, E_1, E_2 \in \mathcal{E}$ *and for any* $K, K' \in \mathcal{K}$:

1. If $I \in Mod(E)$ *and* $J \in Mod(E)$, *then* $I \simeq_E J$

2. If $I \in Mod(E)$ *and* $J \notin Mod(E)$, *then* $I <_E J$

3. If $E_1 \leftrightarrow E_2$, *then* $\leq_{E_1} = \leq_{E_2}$

4. If $Mod(K) \cap Mod(K') = \emptyset$, *then* $min(\leq_{K \sqcup K'}) \not\subseteq Mod(K)$

5. If $I \in min(\leq_{E_1})$ *and* $I \in min(\leq_{E_2})$, *then* $I \in min(\leq_{E_1 \sqcup E_2})$

6. If $min(\leq_{E_1}) \cap min(\leq_{E_2}) \neq \emptyset$ *and* $I \notin min(\leq_{E_1})$, *then* $I \notin min(\leq_{E_1 \sqcup E_2})$

A majority syncretic assignment is a syncretic assignment which satisfies the following:

7. $\forall E \in \mathcal{E}\ \forall K \in \mathcal{K}\ \exists n\ min(\leq_{E \sqcup K^n}) \subseteq Mod(K)$

A fair syncretic assignment is a syncretic assignment which satisfies the following:

8. $\forall K' \, \exists K$ *if* $Mod(K') \not\subseteq Mod(K)$, *then* $\forall n \; min(\leq_{K' \sqcup K^n}) = min(\leq_{K' \sqcup K})$

If we have an assignment that maps each knowledge set E to a pre-order $\leq_E$ on $\mathcal{W}$, then we can define a function $\delta : \mathcal{M} \longrightarrow \mathcal{S}$ by the following: let $M \in \mathcal{M}$ and let $E \in \mathcal{E}$ be such that $\imath(E) = M$, put

$$\delta(M) = \min(\leq_E) \tag{7.1}$$

If the assignment satisfies property 3 above then δ is well defined.

Conversely, if we have a function $\delta : \mathcal{M} \longrightarrow \mathcal{S}$ we can define a corresponding family of relations on interpretations as $\forall E \in \mathcal{E}$:

$$\leq_E = [\delta(\imath(E)) \times (\mathcal{W} \setminus \delta(\imath(E)))] \cup \{\{I, I\} \setminus I \in \mathcal{W}\} \tag{7.2}$$

It is easy to show that if we have a (majority, fair) syncretic assignment, then the merging function obtained by equation 7.1 is a (majority, fair) merging function. Conversely, if we have a (majority, fair) merging function, then the family of relations obtained by equation 7.2 is a (majority, fair) syncretic assignment. This observation together with theorem 7.16 gives us straightforwardly the following:

Theorem 7.18 *An operator is a merging operator (respectively majority merging operator or arbitration operator) if and only if there exists a syncretic assignment (respectively majority syncretic assignment or fair syncretic assignment) that maps each knowledge set E to a pre-order $\leq_E$ such that*

$$Mod(\Delta(E)) = min(\leq_E).$$

As pointed out by D. Makinson (personal communication), this definition of merging operators from such assignments can be compared to the framework of social choice theory [47, 3]. The aim of social choice theory is to aggregate individual choices into a social choice, *i.e.* to find, for a given set of agents (corresponding to our knowledge sets) with individual preference relations, a social preference relation which reflects the preferences of the set of agents. This allows the definition of a welfare function selecting from a set of alternatives those that best fit the social preference relation.

7.5 Some merging operators

In this section we show the consistency of our merging postulates by giving three examples of operators. The first one is not a merging operator but it illustrates an approach to arbitration operators. The second one is a majority merging operator and the last one is a true arbitration operator.

For the following operators we will use the Dalal's distance [15] to calculate the distance between two interpretations: let I, J be interpretations, $dist(I, J)$ is the number of propositional letters the two interpretations differ.

We also define the distance between an interpretation and a knowledge base as the minimum distance between this interpretation and the models of the knowledge base, that is:

$$dist(I,\varphi) = \min_{J\in Mod(\varphi)} dist(I,J)$$

Finally we define the distance between two knowledge bases by the following:

$$dist(\varphi,\varphi') = \min_{I\in Mod(\varphi)\ J\in Mod(\varphi')} dist(I,J)$$

The first operator we consider is the Δ_{Max} operator. It comes from an example of model fitting operator given by Revesz in [100]. It is close to the minimax rule used in decision theory [103]. The idea is to find the closest information to the overall knowledge set. Therefore it seems to be a good arbitration operator. But, as we will see, it doesn't satisfy all the postulates.

Definition 7.19 *Let φ be a knowledge base and E be a knowledge set:*

$$dist_{Max}(I,E) = \max_{\varphi\in E} dist(I,\varphi)$$

So, we define the following order:

$$I \leq_E^{Max} J \quad \textit{iff } dist_{Max}(I,E) \leq dist_{Max}(J,E)$$
$$\textit{and} \qquad Mod(\Delta_{Max}(E)) = min(\leq_E^{Max})$$

The second operator we consider is the Δ_Σ operator. This is a majority merging operator as we will see below. Lin and Mendelzon give it as an example of what they called operators of *theory merging by majority* in [62]. Independently Revesz gives it as an example of weighted model fitting in [99]. The Σ operator comes from a natural idea: the distance between an interpretation and a knowledge set is the sum of the distances between this interpretation and the knowledge bases of the knowledge set.

Definition 7.20 *Let E be a knowledge set and let I be an interpretation we put:*

$$dist_\Sigma(I,E) = \sum_{\varphi\in E} dist(I,\varphi)$$

$$I \leq_E^{\Sigma} J \quad \textit{iff } dist_\Sigma(I,E) \leq dist_\Sigma(J,E)$$
$$\textit{and} \qquad Mod(\Delta_\Sigma(E)) = min(\leq_E^{\Sigma})$$

Next we present a new merging operator: Δ_{GMax} (stands for *Generalized Max*). The operator Δ_{GMax} is an arbitration operator and is a refinement of the Δ_{Max} operator.

Definition 7.21 *Let E be a knowledge set. Suppose $E = \{\varphi_1,\ldots,\varphi_n\}$. For each interpretation I we build the list $(d_1^I \ldots d_n^I)$ of distances between this interpretation and the n knowledge bases in E,* i.e. *$d_j^I = dist(I,\varphi_j)$. Let L_I be the list obtained from $(d_1^I \ldots d_n^I)$ by sorting it in descending order. Define $dist_{GMax}(I,E) = L_I$. Let $\leq_{lex}$ be the lexicographical order between sequences of integers. Now we put:*

$$I \leq_E^{GMax} J \quad \textit{iff} \quad dist_{GMax}(I,E) \leq_{lex} dist_{GMax}(J,E)$$
$$\textit{and} \qquad Mod(\Delta_{GMax}(E)) = min(\leq_E^{GMax})$$

We will illustrate the behaviour of these three operators on the database class example given by Revesz in [99]:

Example 7.22 *Consider a database class with three students: $E = \{\varphi_1, \varphi_2, \varphi_3\}$. The teacher can teach SQL, Datalog and O_2. He asks his students in turn to choose what to teach to satisfy the class best. The first student wants to learn SQL or O_2: $\varphi_1 = (S \vee O) \wedge \neg D$. The second wants to learn Datalog or O_2 but not both: $\varphi_2 = (\neg S \wedge D \wedge \neg O) \vee (\neg S \wedge \neg D \wedge O)$. The third wants to learn the three languages: $\varphi_3 = (S \wedge D \wedge O)$. Considering the propositional letters S, D and O in that order we have: $Mod(\varphi_1) = \{(1,0,0),(0,0,1),(1,0,1)\}$, $Mod(\varphi_2) = \{(0,1,0),(0,0,1)\}$, $Mod(\varphi_3) = \{(1,1,1)\}$.*

The following table contains all distances relevant to computations in order to calculate $\Delta_{Max}(E)$, $\Delta_{\Sigma}(E)$ and $\Delta_{GMax}(E)$.

	φ_1	φ_2	φ_3	$\mathbf{dist_{Max}}$	$\mathbf{dist_{\Sigma}}$	$\mathbf{dist_{GMax}}$
$(0,0,0)$	1	1	3	3	5	(3,1,1)
$(0,0,1)$	0	0	2	2	2	(2,0,0)
$(0,1,0)$	2	0	2	2	4	(2,2,2)
$(0,1,1)$	1	1	1	1	3	(1,1,1)
$(1,0,0)$	0	2	2	2	4	(2,2,0)
$(1,0,1)$	0	1	1	1	2	(1,1,0)
$(1,1,0)$	1	1	1	1	3	(1,1,1)
$(1,1,1)$	1	2	0	2	3	(2,1,0)

As the min in the column of $dist_{Max}$ is 1 we have $Mod(\Delta_{Max}(E)) = \{(0,1,1),(1,0,1),(1,1,0)\}$, thus the teacher has to teach two of the three languages to best satisfy the class when the criterion to solve conflicts is Δ_{Max}. Similarly as the min in the column of $dist_{\Sigma}$ is 2 we have $Mod(\Delta_{\Sigma}(E)) = \{(0,0,1),(1,0,1)\}$, thus the teacher has to teach both SQL and O_2 or O_2 alone to best fit the class when the criterion to solve conflicts is Δ_{Σ}. Finally as the min in the column of $dist_{\Delta_{GMax}}$ is $(1,1,0)$ we have $Mod(\Delta_{GMax}(E)) = \{(1,0,1)\}$, thus the teacher has to teach SQL and O_2 to best satisfy the class when the criterion to solve conflicts is Δ_{GMax}.

As we can expect the result of the merging highly depends on the operator we choose. Note in particular that the Δ_{Max} operator has selected interpretations that satisfy as much as possible each student, whereas the Δ_{Σ} operator has selected interpretations that satisfy the majority of students. Notice also that in this example the Δ_{GMax} operator selects the interpretation chosen by both Δ_{Max} and Δ_{Σ} operators, showing its good behaviour.

We will see now the logical properties of these three operators.

We first show that Δ_{Max} is not a merging operator.

Theorem 7.23 *Δ_{Max} satisfies postulates $(A1 - A5)$, $(A7')$ and (A_{it}) but it doesn't satisfy $(A6)$.*

Proof: The proof of $(A1 - A3)$ and $(A5)$ is straightforward. To prove that $(A4)$ is satisfied suppose $K \wedge K' \vdash \bot$. We consider two cases: $dist(K,K') = 1$ or $dist(K,K') > 1$ If $dist(K,K') = 1$ then $\exists I \in Mod(K), \exists J \in Mod(K')$ such that $dist(I,J) = 1$, so as $dist(I,J)$

is minimum $I \in Mod(\Delta(K \sqcup K'))$ and $J \in Mod(\Delta(K \sqcup K'))$, so $\Delta(K \sqcup K') \not\vdash K$. Otherwise $dist(K, K') > 1$, and then $\exists I \in Mod(K), \exists J \in Mod(K')\ \forall I' \in Mod(K), \forall J' \in Mod(K')$ $dist(I, J) \leq dist(I', J')$ and $dist(I, J) > 1$. But it is easy to see that if $dist(I, J) = a > 1$ then there exists $L \in \mathcal{W}$ such that $dist(L, I) < a$ and $dist(L, J) < a$, so $dist_{Max}(L, K \sqcup K') < a$. Therefore $L <^{Max}_{K \sqcup K'} I$ so $I \notin Mod(\Delta(K \sqcup K'))$, so $\Delta(K \sqcup K') \not\vdash K$. $(A7')$ is satisfied because $\max_{\varphi \in E \sqcup K^n} dist(I, \varphi) = \max_{\varphi \in E \sqcup K} dist(I, \varphi)$. So $\Delta(E \sqcup K^n) = \Delta(E, K)$. As $(A7')$ is satisfied, $(A7)$ is satisfied. In order to show that $(A6)$ is not satisfied consider the example 7.22 and observe that if we take $E_1 = \{\varphi_1\}$ and $E_2 = \{\varphi_2, \varphi_3\}$, then $\Delta(E_1) \wedge \Delta(E_2) = form(\{(1, 0, 1)\})$ is consistent, and $\Delta(E_1 \sqcup E_2) = form(\{(0, 1, 1), (1, 0, 1), (1, 1, 0)\})$, so $\Delta(E_1 \sqcup E_2) \not\vdash \Delta(E_1) \wedge \Delta(E_2)$.

It remains to show that (A_{it}) holds. First, by induction on $dist(K, K')$ we prove that

$$\exists n \text{ such that } \Delta^n_{Max}(K', K) \vdash K \quad (*)$$

If $dist(K, K') = 0$ the proof is straightforward. Suppose $dist(K, K') = 1$. Then $\exists I \in Mod(K)\ \exists J \in Mod(K')\ dist(I, J) = 1$. So $I \in \Delta_{Max}(K, K')$ and then, by (A2), $\Delta^2_{Max}(K', K) = \Delta_{Max}(\Delta_{Max}(K', K), K) = \Delta_{Max}(K', K) \wedge K$. So $\Delta^2_{Max}(K', K) \vdash K$. Suppose that $dist(K, K') > 1$. Put $a = dist(K', K)$, i.e. $\exists I \in Mod(K)\ \exists J \in Mod(K')\ dist(I, J) = a$. Let $a/2$ be the integer part of the quotient of a by 2. Since I and J disagree on a letters, we can find an interpretation I' such that I' agrees with I on the letters on which I and J agree, and I' agrees with J on $a/2$ letters on which I and J disagree and I' agrees with I for the $a/2$ remaining letters if a is even and for the $a/2 + 1$ remaining letters if a is odd. So we have $dist(I', K) \leq a/2$ and $dist(I', K') \leq a/2$ if a is even or $dist(I', K') \leq a/2 + 1$ if a is odd. If a is even then $dist_{Max}(I', \{K, K'\}) \leq a/2$, so if $J' \in Mod(\Delta_{Max}(K, K'))$ then $dist_{Max}(J', \{K, K'\}) \leq a/2$. So we have that if $dist(K, K') = a$ with $a > 1$ then $dist(K, \Delta_{Max}(K, K')) \leq a/2$. By induction hypothesis there exists n such that $\Delta^n_{Max}(\Delta_{Max}(K, K'), K) \vdash K$ that is $\Delta^{n+1}_{Max}(K', K) \vdash K$. The case where a is odd is similar. Now (A_{it}) follows from $(*)$ by putting $K' = \Delta(E \sqcup K)$. ∎

The operator Δ_Σ is a majority merging operator as stated in the following theorem.

Theorem 7.24 *Δ_Σ satisfies postulates $(A1 - A6)$, $(M7)$ and (A_{it}).*

Proof: We will prove that the assignment $E \mapsto \leq^\Sigma_E$ is a majority syncretic assignment. Then by theorem 7.18 we conclude that Δ_Σ satisfies $(A1 - A6)$ and $(M7)$. Let's verify the conditions of a majority syncretic assignment:

1. If $I \in Mod(E)$ and $J \in Mod(E)$, then $dist_\Sigma(I, E) = 0$ and $dist_\Sigma(J, E) = 0$, so $I \simeq_E J$.

2. If $I \in Mod(E)$ and $J \notin Mod(E)$, then $dist_\Sigma(I, E) = 0$ and $dist_\Sigma(J, E) > 0$, so $I <_E J$.

3. Straightforward.

4. Suppose $K \wedge K' \vdash \bot$, so $dist(K, K') > 0$. So $\exists I \in Mod(K), \exists J \in Mod(K')\ \forall I' \in Mod(K), \forall J' \in Mod(K')\ dist(I, J) \leq dist(I', J')$ and $dist(I, J) = a > 1$. It is easy to see that $a = \min\{dist_\Sigma(L, K \sqcup K') : L \in \mathcal{W}\}$ thus $I \in Mod(\Delta(K \sqcup K'))$ and $J \in Mod(\Delta(K \sqcup K'))$, so $\Delta(K \sqcup K') \not\vdash K$.

5. If $I \in \min(\leq_{E_1})$ and $I \in \min(\leq_{E_2})$, then $\forall J\ dist_\Sigma(I, E_1) \leq dist_\Sigma(J, E_1)$ and $dist_\Sigma(I, E_2) \leq dist_\Sigma(J, E_2)$. So $\forall J\ dist_\Sigma(I, E_1) + dist_\Sigma(I, E_2) \leq dist_\Sigma(J, E_1) + dist_\Sigma(J, E_2)$. By definition of $dist_\Sigma$ is easy to see that for any L, E, E', $dist_\Sigma(L, E \sqcup E') = dist_\Sigma(L, E) + dist_\Sigma(L, E')$. Then $\forall J\ dist_\Sigma(I, E_1 \sqcup E_2) \leq dist_\Sigma(J, E_1 \sqcup E_2)$. So $I \in \min(\leq_{E_1 \sqcup E_2})$.

6. If $\min(\leq_{E_1}) \cap \min(\leq_{E_2}) \neq \emptyset$, then $\exists J$ s.t. $J \in \min(\leq_{E_1})$ and $J \in \min(\leq_{E_2})$. Suppose $I \notin \min(\leq_{E_1})$, then $dist_\Sigma(J, E_1) < dist_\Sigma(I, E_1)$ and $dist_\Sigma(J, E_2) \leq dist_\Sigma(I, E_2)$. So $dist_\Sigma(J, E_1) + dist_\Sigma(J, E_2) < dist_\Sigma(I, E_1) + dist_\Sigma(I, E_2)$. Then $dist_\Sigma(J, E_1 \sqcup E_2) < dist_\Sigma(I, E_1 \sqcup E_2)$. Then $I \notin \min(\leq_{E_1 \sqcup E_2})$.

7. We have to find a n such that $\min(\leq_{E \sqcup K^n}) \subseteq Mod(K)$. Consider $x = \max_{I \in \mathcal{W}} dist_\Sigma(I, E)$, i.e. x is the distance of the furthest interpretation from E. We choose $n = x + 1$, it is easy to see that if $I \in Mod(K)$ then $dist_\Sigma(I, E \sqcup K^n) < n$. And if $I \notin Mod(K)$ then $dist_\Sigma(I, E \sqcup K^n) \geq n$. So if $I \in \min(\leq_{E \sqcup K^n})$ then $I \in Mod(K)$.

Now we prove that (A_{it}) holds. We want to show that $\exists n\ \Delta^n_\Sigma(K', K) \vdash K$. Let a be the distance between K and K'. Take $I \in Mod(K)$ and $J \in Mod(K')$ such that $dist(I, J) = a$. It is easy to see that $a = \min\{dist_\Sigma(L, K \sqcup K') : L \in \mathcal{W}\}$ thus $I \in Mod(\Delta(K \sqcup K'))$ and then $\Delta_\Sigma(\Delta_\Sigma(K' \sqcup K), K) \vdash K$. Therefore $\exists n\ \Delta^n_\Sigma(K', K) \vdash K$. And with $K' = \Delta_\Sigma(E \sqcup K)$ we have $\exists n\ \Delta^n_\Sigma(E, K) \vdash K$. ∎

Now, we will state some lemmas in order to prove that Δ_{GMax} has desirable properties.

Definition 7.25 *Let L_1 and L_2 be two lists of n numbers sorted in descending order. We define $L_1 \odot L_2$ the list obtained by sorting in descending order the concatenation of L_1 with L_2.*

Lemma 7.26 *Let L_1, L_1', L_2, L_2' be 4 lists of integers sorted in descending order. If $L_1 \leq_{lex} L_1'$ and $L_2 \leq_{lex} L_2'$ then $L_1 \odot L_2 \leq_{lex} L_1' \odot L_2'$.*

Proof: Suppose that $L_1 \leq L_1'$ and $L_2 \leq L_2'$. It is easy to see that the two following inequalities hold: $L_1 \odot L_2 \leq_{lex} L_1' \odot L_2$ and $L_1' \odot L_2 \leq_{lex} L_1' \odot L_2'$. So by transitivity $L_1 \odot L_2 \leq_{lex} L_1' \odot L_2'$. ∎

Lemma 7.27 *Let L_1, L_1', L_2, L_2' be 4 lists of integers sorted in descending order. If $L_1 \leq_{lex} L_1'$ and $L_2 <_{lex} L_2'$ then $L_1 \odot L_2 <_{lex} L_1' \odot L_2'$.*

Proof: With the assumptions it is easy to see that $L_1 \odot L_2 \leq_{lex} L_1' \odot L_2$ and $L_1' \odot L_2 <_{lex} L_1' \odot L_2'$. We conclude by transitivity of $\leq_{lex}$. ∎

The operator Δ_{GMax} is a true arbitration operator as showed in the following theorem.

Theorem 7.28 *The operator Δ_{GMax} satisfies postulates $(A1 - A6)$ and (A_{it}). Furthermore Δ_{GMax} satisfies $(A7)$ iff $card(\mathcal{P}) > 1$. But it doesn't satisfy $(A7')$.*

Proof: In order to show that $GMax$ satisfies $(A1 - A7)$ we use the representation theorem and we show that the assignment $E \mapsto \leq^{Gmax}_E$ is a fair syncretic assignment.

1. If $I \in Mod(E)$ and $J \in Mod(E)$, then $\forall K_i \in E\ I \in Mod(K_i)$ and $J \in Mod(K_i)$, then $L_I = (0, \ldots, 0)$ and $L_J = (0, \ldots, 0)$, so $I \simeq_E J$.

2. If $I \in Mod(E)$ and $J \notin Mod(E)$, then $L_I = (0, \ldots, 0)$ and $L_J \neq (0, \ldots, 0)$, so $I <_E J$.

3. If $E_1 \leftrightarrow E_2$, then is obvious that $\leq_{E_1} = \leq_{E_2}$.

4. This property is proved in a similar way as (A_4) for Δ_{Max} (theorem 7.23).

5. If $I \in \min(\leq_{E_1})$ and $I \in \min(\leq_{E_2})$, then $\forall J \in \mathcal{W}$ $L_I^{E_1} \leq_{lex} L_J^{E_1}$ and $L_I^{E_2} \leq_{lex} L_J^{E_2}$. So, by lemma 7.26, we have $\forall J$ $L_I^{E_1 \sqcup E_2} \leq_{lex} L_J^{E_1 \sqcup E_2}$. Then $I \in \min(\leq_{E_1 \sqcup E_2})$.

6. If $\min(\leq_{E_1}) \cap \min(\leq_{E_2}) \neq \emptyset$ and $I \notin \min(\leq_{E_1})$, let $J \in \min(\leq_{E_1}) \cap \min(\leq_{E_2})$, so $L_J^{E_1} <_{lex} L_I^{E_1}$ and $L_J^{E_2} \leq_{lex} L_I^{E_2}$, and by lemma 7.27 follows $L_J^{E_1 \sqcup E_2} <_{lex} L_I^{E_1 \sqcup E_2}$. Then $I \notin \min(\leq_{E_1 \sqcup E_2})$.

7. Consider a knowledge base K'. We will show that if there are 2 or more propositional variables then there exists a K s.t. $K' \not\vdash K$ and $\forall n$ $\min(\leq_{K' \sqcup K^n}) = \min(\leq_{K' \sqcup K})$. We consider 2 cases, first if $card(Mod(K')) > 1$ then let $I \in Mod(K')$, we choose $K = form(I)$. So, by condition 1 and 2, $\min(\leq_{K' \sqcup K}) = \{I\}$ and $\forall n$ $\min(\leq_{K' \sqcup K^n}) = \{I\}$. Hence $\forall n$ $\min(\leq_{K' \sqcup K^n}) = \min(\leq_{K' \sqcup K})$. Second, if $card(Mod(K')) = 1$, let $Mod(K') = \{J\}$, we choose $K = form(I)$ s.t. $dist(I, J) = 2$, this is possible because there are at least two propositional variables. So there exists I' s.t. $dist(I', I) = 1$ and $dist(I', J) = 1$. So $\min(\leq_{K' \sqcup K^n}) = \{I' : dist(I', I) = 1 \text{ and } dist(I', J) = 1\}$ otherwise if $\exists J'$ such that $dist(J', K) = 0$ then $dist(J', K') \geq 2$ or if $dist(J', K') = 0$ then $dist(J', K) \geq 2$, and so $L_{I'} < L_{J'}$. So $\forall n$ $\min(\leq_{K' \sqcup K^n}) = \{I' : dist(I', I) = 1 \text{ and } dist(I', J) = 1\}$. Then $\forall n$ $\min(\leq_{K' \sqcup K^n}) = \min(\leq_{K' \sqcup K})$. Conversely suppose that $\mathcal{P} = \{p\}$. Put $K' = p$. Then the only consistent K (up to logical equivalence) such that $K' \not\vdash K$ is $K = \neg p$ but $\Delta_{GMax}(K' \sqcup K^n) = \neg p$ for any $n \geq 2$ whereas $\Delta_{GMax}(K' \sqcup K) = \neg p \vee p$.

To show that Δ_{GMax} doesn't satisfy $(A7')$ consider the following example: Suppose that $\mathcal{P} = \{p, q\}$ and that $K' = \neg p \wedge \neg q$ and $K = \neg p \wedge q$. It is easy to see that $\Delta_{GMax}(K \sqcup K') = \neg p$ whereas $\Delta_{GMax}(K' \sqcup K^n) = \neg p \wedge q$ for any $n \geq 2$.

Finally the proof that the postulate (A_{it}) holds for Δ_{GMax} goes exactly the same way that for Δ_{Max} (theorem 7.23). ∎

Actually $GMax$ operator is a refinement of the Max operator. More precisely we have the following observation the proof of which is straightforward:

Observation 7.29 $\Delta_{GMax}(E) \vdash \Delta_{Max}(E)$.

We end this section with the following table which sums up the properties of operators defined above. It is filled using the results of this section together with some results of section 3. The symbol ✓ (respectively –) in a square means that the corresponding operator satisfies (resp. does not satisfy) the corresponding postulate.

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A7′**	**M7**	**A$_{it}$**
Max	✓	✓	✓	✓	✓	–	✓	✓	–	✓
Σ	✓	✓	✓	✓	✓	✓	–	–	✓	✓
$GMax$	✓	✓	✓	✓	✓	✓	✓	–	–	✓

7.6 Conclusion and future work

We have proposed in this paper a set of postulates that a rational merging operator has to satisfy. We have made a distinction between arbitration operators striving to minimize individual dissatisfaction and majority operators striving to minimize global dissatisfaction. The fairness postulate is the key postulate in this distinction. We have shown that our characterization is equivalent to a family of pre-orders on interpretations. We show the consistency of the axiomatic characterization by giving examples of operators. In particular, we have proposed a new rational merging operator called Δ_{GMax} and shown that it is an arbitration operator.

Actually, in a committee, all the protagonists do not have the same weight on the final decision and so one needs to weight each knowledge base to reflect this. The idea behind weights is that the higher weight a knowledge base has, the more important it is. If the knowledge bases reflect the view of several people, weights could represent, for example, the cardinality of each group. We want to characterize logically the use of this weights. Majority operators are close to this idea of weighted operators since they allow to take cardinalities into account. But a more subtle treatment of weights in merging is still to do, in particular the notion of weighted arbitration operators is missing.

In this work the result of a merging is a subset of the set of all interpretations but a lot of systems have to conform to a set of integrity constraints, for that reason it is interesting to be able to merge some knowledge sets in the presence of these constraints [63]. And so one has to restrain the result of the merging to be a subset of the set of allowed interpretations. Suppose that these integrity constraints are denoted by the knowledge base IC. If we consider a weighted rational merging, a way to incorporate integrity constraints is to add IC to E with a weight "infinity". Thus we would ensure that the interpretations selected were models of IC. Intuitively, it amounts to consider a person in the committee whose view is unquestionable and therefore one has to choose among the alternatives given by that person.

But the best way to include integrity constraints seems to be to select the minimal models in the models of the IC base rather than in $\mathcal{W}$. Intuitively, we restrict the choices of interpretations to those which satisfy IC. It is in a sense what Revesz called model fitting operators [100].

In that paper we use only the Dalal's distance to define the distance between two interpretations, it would be interesting to study operators defined with other distances, in particular distances which give partial orders.

Notice also that the three merging operators defined in the paper are based on the Dalal's distance. But if one chooses an other distance between interpretations and keeps the same definitions, then one obtains other merging operators. So, more exactly, we have defined in this paper three families of merging operators, function of the definition of the distance between interpretations. It would be interesting to find what the minimum conditions on that distance are to ensure that the operators satisfy the axiomatic characterization.

Two other points of interest are to study merging operators which are not defined from a distance and to study syntactic definition of merging operators.

Acknowledgements

We would like to thank D. Makinson, P. Liberatore, P. Revesz and two anonymous referees for their comments which have improved the quality of this paper.

Chapitre 8

Remarques finales et perspectives

Ci-dessous nous donnons quelques remarques et commentaires sur nos résultats et ensuite nous énoncerons quelques points qui nous semblent importants de développer dans des travaux futurs.

8.1 Remarques finales

Nous voulons commencer en insistant sur le fait qu'au centre de ces travaux se situe la logique préférentielle et quelques-unes de ses extensions notamment la logique rationnelle avec son "dual" le cadre AGM pour la révision de la connaissance.

Nous avons étudié dans le chapitre 3 différentes extensions de la logique préférentielle. Certaines avec des règles non-Horn plus fortes que la monotonie rationnelle d'autres avec des règles Horn qui sont incomparables à la monotonie rationnelle. Pour cela nous nous sommes servis des outils syntaxiques et sémantiques. En particulier des théorèmes de représentation, certains bien connus d'autres que nous avons établis.

Une caractéristique importante des théorèmes de représentation, en dehors de leur intérêt intrinsèque comme puissants outils de raisonnement, est que leur nature "géometrique" permet d'introduire des "déformations" qui sont une excellente heuristique pour la recherche des nouveaux postulats pour les relations de conséquence non monotones.

Dans le chapitre 4 la notion de relation essentielle $<_e$ entre valuations, associée à une relation de conséquence non monotone, s'est avérée fort utile pour donner des preuves uniformes des théorèmes de représentation. Cette même idée a été exploitée au chapitre 5 dans le cadre des relations d'abduction pour construire des ordres sur les formules qui nous donnent les meilleures explications.

Toujours dans le chapitre 4 nous avons cherché à caractériser l'existence des modèles injectifs d'une relation préférentielle. Nous montrons que la caractérisation dans le cas fini par la règle W-DR due à Freund marche partiellement dans le cas infini. Plus précisément nous prouvons, qu'en présence de W-DR, la relation essentielle définit un modèle MAK qui représente la relation de conséquence non monotone de départ.

Or les problèmes que l'on rencontre dans le cas infini (notamment que la relation $<_e$ ne soit pas, en général, transitive) sont dus à des comportements des "approximations" d'une valuation. Ceci mène à considérer naturellement des notions topologiques sur les valuations qui donnent une lumière nouvelle sur les représentations. Ces notions sont très utiles pour caractériser l'unicité de certaines représentations et aussi dans la recherche de contre-exemples.

Dans le chapitre 5 nous avons aussi utilisé la logique préférentielle et ses extensions pour faire une étude systématique de l'abduction. En effet, nous proposons des postulats pour les relations abductives qui s'avéreront être duaux d'une façon assez précise de ceux de familles de relations non monotones bien connues. On peut voir cette dualité comme une justification de nos postulats pour l'abduction. Or, on peut aussi donner des justifications intuitives des postulats pour l'abduction et voir les liens étroits entre l'abduction et les relations de conséquence non monotones comme un argument renforçant la thèse que la logique préférentielle (et quelques-unes de ses extensions) joue un rôle central dans le cadre du raisonnement dynamique.

Nous montrons aussi dans le chapitre 5 que le raisonnement abductif peut être vu comme du raisonnement non monotone à l'envers.

Une remarque importante concernant nos postulats pour l'abduction est qu'ils sont assez puissants pour définir un ordre $\prec_e$ sur les formules qui permet de récupérer la relation d'abduction : les bonnes explications d'une observation α seront les éléments minimaux, par rapport à $\prec_e$, parmi l'ensemble des candidats à être de bonnes explications.

Dans le chapitre 6 nous avons utilisé la dualité existante entre les relations rationnelles et les opérateurs de révision pour définir des opérateurs syntaxiques avec le but de diminuer la complexité du traitement et de garder de bonnes propriétés logiques. Pour ce faire nous avons d'une part affaibli la logique et d'autre part donné une méthodologie pour relativiser les postulats AGM de la révision et ceux de KM de la mise-à-jour au cadre syntaxique dans lequel on travaille. Plusieurs opérateurs ont été définis. En particulier les opérateurs $\diamond_P$, $\circ_{rk}$ et $\circ_{sh}$ (actualisation factuelle, révision rangée et révision avec sélection respectivement). Ces trois opérateurs ont de bonnes propriétés logiques. Notamment le premier est un opérateur de mise-à-jour syntaxique et les deux derniers sont des opérateurs de révision syntaxique. Mais du point de vue de la complexité le seul opérateur dont les calculs restent polynomiaux est celui de la révision rangée, opérateur qui s'inspire du calcul de la clôture rationnel d'une base conditionnelle définit par Lehmann et Magidor [55].

Dans le chapitre 7 nous abordons le problème de la fusion de bases de connaissances selon un point de vue qualitatif, plus précisément avec des représentations logiques. Une caractérisation des opérateurs de fusion en termes de postulats de rationalité est proposée. Pour ce faire nous introduisons deux notions importantes :

- La notion de groupe (*knowledge set*) qui n'est autre chose qu'un multi-ensemble de formules
- La notion de groupes équivalents. Deux groupes sont équivalents s'il y a une bijection entre ces groupes telle que les formules en correspondance sont logiquement équivalentes.

A l'aide de ces notions nous formulons les postulats qu'un opérateur doit posséder pour avoir un comportement rationnel par rapport à la fusion. Notamment nous proposons un postulat d'équité qui dit intuitivement que dans un groupe de deux individus le résultat de la fusion ne priviligiera aucun des deux individus. Ce postulat sera essentiel pour distinguer entre opérateurs de fusion majoritaires et opérateurs de fusion consensuels (*arbitration operators*). Il est important de remarquer que l'on ne peut pas réduire la notion de groupe à la notion d'ensemble de formules car alors on perd la cohérence des postulats. Des exemples d'opérateurs montrant la cohérence des postulats sont donnés. En particulier nous introduisons l'opérateur Δ_{GMax} et montrons que c'est un exemple d'opérateur d'arbitrage. Didier Dubois nous a communiqué que c'est un opérateur déjà introduit en théorie de la décision il y a quelques années et qui a aussi récemment été redécouvert par Hélène Fargier [20].

La simple traduction des postulats en termes sémantiques donne un théorème de représentation trivial mais quand même utile pour la construction d'opérateurs de fusion. Un autre intérêt de cette traduction est qu'elle laisse entrevoir des rapports étroits entre la théorie de la décision qualitative et la logique de la fusion.

8.2 Perspectives

1. Nous avons introduit plusieurs règles dans le chapitre 3 pour lesquelles la sémantique n'as pas été établie ou pas complètement. En particulier les règles CI et n-M pour $n \geq 2$. Dans un premier temps nous pensons que démarrer l'étude dans le cas fini avec en plus la règle W-DR (*i.e.* l'injectivité) simplifie énormement la tâche et doit donner des résultats immédiats. Par exemple pour CI les modèles sont plats (monotonie) ou sinon ils se reduisent à deux points l'un plus petit que l'autre.

   Un autre point concernant ce chapitre est de savoir s'il y a une notion naturelle de clôture rationnelle transitive. Par exemple à partir de $B = \{\alpha_i \mid\!\sim \beta_i : i \in \mathcal{I}\}$ on peut construire, "en orientant" la règle RT, une relation $\mid\!\sim$ de la façon suivante. D'abord on définit $\mid\!\sim_0$ comme la clôture rationnelle de B si elle existe, sinon c'est la clôture préferentielle. Ensuite on pose
   $$\mid\!\sim_{n+1} = (\mid\!\sim_n \cup \{(\alpha,\gamma) : \alpha \mid\!\sim_n \beta \mid\!\sim_n \gamma \ \&\ \alpha \not\mid\!\sim \neg\gamma\})^P$$
   où K^P dénote la clôture préférentielle de K. Finalement on pose
   $$\mid\!\sim = \bigcup_{n\in\omega} \mid\!\sim_n$$

   Est-ce que cette relation a de bonnes propriétés? Il est facile de voir quelle est rationnelle transitive. Mais la question est de savoir s'il y a une notion simple pour laquelle elle soit la meilleure.

2. Concernant la caractérisation des relations préférentielles admettant des modèles injectifs (KLM) la question reste encore ouverte. Néanmoins nous savons que s'il y a un ordre $<$ représentant injectivement la relation de conséquence $\mid\!\sim$ il doit être contenu dans $<_e$. Ainsi nous avons essayé "d'élaguer" la relation $<_e$ afin d'obtenir un ordre $<_e^*$ qui représente $\mid\!\sim$. Plus précisément on définit $<_e^*$ de la façon suivante :
   $$<_e^* = <_e \setminus \{(N,P) : \exists M\ M <_e N <_e P \ \&\ M \not<_e P\}$$
   C'est-à-dire on enlève de $<_e$ certains des couples où la prémisse de la transivité est vraie et la conclusion est fausse. Curieusement la relation obtenue est transitive. Le problème est que nous ne savons pas si elle est encore smooth ni si elle représente la relation $\mid\!\sim$ de départ.

3. Concernant nos relations abductives plusieurs problèmes non encore résolus nous semblent importants. Le premier est de nature calculatoire et en paraphrasant le titre de [55] (*What does a conditional base entail?*) il peut s'énoncer par la question suivante

'Que doit impliquer une base abductive?' Plus précisément si l'on a une base abductive $B_A = \{\alpha_i \rhd \gamma_i : i \in \mathcal{I}\}$ quelle est la meilleure relation explicatoire qui la contient et comment la calculer? Notons que la règle E-C-Cut ne passe pas forcément aux intersections et pour cette raison on ne peut même pas parler de plus petite relation E-préférentielle contenant B_A.

Un autre problème intéressant concerne le rôle de la théorie Σ dans la notion de relation d'explication. Nous avons assumé que la théorie Σ était fixée. Or il se pourrait qu'il y ait des situations où l'on doit changer Σ. Quels seraient les postulats lorsque Σ est un paramètre? Serait-ce encore de l'abduction ou plutôt de l'induction?

Nous avons esquissé des rapports entre nos relations d'abduction et des relations d'abduction épistémiques. Mais nous n'avons pas étudié en détail ces dernières. C'est une étude qui devrait être faite du point de vue de la dualité entre la révision et ces relations et aussi de la représentation de ces relations en termes d'ordres de préférence entre les formules.

4. Pour les opérateurs définis dans le chapitre 6 un calcul exact de la complexité reste à faire, ainsi qu'une analyse expérimentale des opérateurs moins complexes doit être réalisée. Un prototypage en Prolog de ces opérateurs a déjà été fait par S. Janot mais des essais sur des bases de grande taille restent encore à être effectués.

   Nous pensons que les idées du traitement syntaxique des opérateurs de révision devraient pouvoir s'adapter à l'étude des opérateurs de fusion du chapitre 7.

5. Une extension naturelle et puissante des opérateurs de fusion est de considérer que le résultat doit obéir à certaines contraintes. Nous avons commencé à étudier ce type d'opérateurs de fusion. Ils sont en fait une extension du cadre AGM pour la revision et admettent des théorèmes de représentation non triviaux. Avec le paramètre additionnel des contraintes (d'intégrité) on peux exprimer de façon plus positive l'idée d'un comportement consensuel.

   Un des problèmes intéressants est celui de la construction de tels opérateurs. Les méthodes qui semblent les plus naturelles pointent sur le rôle essentiel de la notion de distance afin d'obtenir des opérateurs possédant de bonnes propriétés.

   Nous pensons que les rapports étroits entre la fusion et la théorie qualitative de la décision, que le théorème de représentation laisse entrevoir, pourront être établis d'une façon assez précise.

Appendice A

Postulates for consequence and explanatory relations

A.1 Rationality Postulates for Consequence Relations

To make easier the reading of the paper we will include a list of all rationality postulates for consequence relations used in the paper.

REF (reflexivity)	$\alpha \mid\!\sim \alpha$		
LLE (left logical equivalence)	$\alpha \mid\!\sim \beta \;\&\; \vdash \alpha \leftrightarrow \gamma$	$\Rightarrow$	$\gamma \mid\!\sim \beta$
RW (right weakening)	$\alpha \mid\!\sim \beta \;\&\; \vdash \beta \rightarrow \gamma$	$\Rightarrow$	$\alpha \mid\!\sim \gamma$
CUT	$\alpha \wedge \beta \mid\!\sim \gamma \;\&\; \alpha \mid\!\sim \beta$	$\Rightarrow$	$\alpha \mid\!\sim \gamma$
CM (cautious monotony)	$\alpha \mid\!\sim \beta \;\&\; \alpha \mid\!\sim \gamma$	$\Rightarrow$	$\alpha \wedge \gamma \mid\!\sim \beta$
OR	$\alpha \mid\!\sim \gamma \;\&\; \beta \mid\!\sim \gamma$	$\Rightarrow$	$\alpha \vee \beta \mid\!\sim \gamma$
S	$\alpha \wedge \beta \mid\!\sim \gamma$	$\Rightarrow$	$\alpha \mid\!\sim \beta \rightarrow \gamma$
DR (disjunctive rationality)	$\alpha \vee \beta \mid\!\sim \rho$	$\Rightarrow$	$\alpha \mid\!\sim \rho$ or $\beta \mid\!\sim \rho$
RM (rational monotony)	$\alpha \mid\!\sim \rho \;\&\; \alpha \not\mid\!\sim \neg\beta$	$\Rightarrow$	$\alpha \wedge \beta \mid\!\sim \rho$
M (monotony)	$\alpha \mid\!\sim \gamma$	$\Rightarrow$	$\alpha \wedge \beta \mid\!\sim \gamma$

An inference relation $\mid\!\sim$ is said to be *cumulative* if it satisfies the rules REF, LLE, RW, CUT and CM. A consequence relation is called *preferential* if it satisfies, in addition to cumulative rules, the rule **OR** and it is called *rational* if it is preferential and satisfies RM. $\mid\!\sim$ is *monotone* if it satisfies **Mono.** A consequence relation satisfies W-DR if $C(\alpha \vee \beta) \subseteq Cn(C(\alpha) \cup C(\beta))$, for every formulas α and β. We used also Con_Σ (Σ-consistency preservation) which is a variant of a postulate introduced in [37]: for all α, $\alpha \mid\!\sim \bot$ iff $\vdash_\Sigma \neg\alpha$ and if $\sigma \in \Sigma$, then $\alpha \mid\!\sim \sigma$.

A.2 Rationality Postulates for Explanatory Relations

We list below all postulates for explanatory relations that we have introduced in this paper.

E-RW	$\alpha \rhd \gamma \;\&\; \alpha \rhd \delta$	$\Rightarrow$	$\alpha \rhd (\gamma \vee \delta)$
E-CM	$\gamma \rhd \alpha \;\&\; \gamma \vdash_\Sigma \beta$	$\Rightarrow$	$(\alpha \wedge \beta) \rhd \gamma$
E-C-Cut	$(\alpha \wedge \beta) \rhd \gamma \;\&\; \forall\delta\, (\alpha \rhd \delta \Rightarrow \delta \vdash_\Sigma \beta)$	$\Rightarrow$	$\alpha \rhd \gamma$
E-Reflexivity	$\alpha \rhd \gamma$	$\Rightarrow$	$\gamma \rhd \gamma$
LLE_Σ	$(\vdash_\Sigma \alpha \leftrightarrow \alpha') \;\&\; \alpha \rhd \gamma$	$\Rightarrow$	$\alpha \rhd \gamma'$
E-Con_Σ	$\nvdash_\Sigma \neg\alpha$	$\Leftrightarrow$	$\exists\gamma\; \alpha \rhd \gamma$
RA	$\alpha \rhd \gamma \;\&\; \gamma' \vdash_\Sigma \gamma \;\&\; \gamma' \nvdash_\Sigma \bot$	$\Rightarrow$	$\alpha \rhd \gamma'$
RLE_Σ	$(\vdash_\Sigma \gamma \leftrightarrow \gamma') \;\&\; \alpha \rhd \gamma$	$\Rightarrow$	$\alpha \rhd \gamma'$
E-Disj	$\gamma \nvdash_\Sigma \bot \;\&\; \rho \nvdash_\Sigma \bot \;\&\; \alpha \rhd (\gamma \vee \rho)$	$\Rightarrow$	$\alpha \rhd \gamma \;\&\; \alpha \rhd \rho$
LOR	$\alpha \rhd \gamma \;\&\; \beta \rhd \gamma$	$\Rightarrow$	$(\alpha \vee \beta) \rhd \gamma$
E-DR	$\alpha \rhd \gamma \;\&\; \beta \rhd \delta$	$\Rightarrow$	$(\alpha \vee \beta) \rhd \gamma$ or $(\alpha \vee \beta) \rhd \delta$
E-R-Cut	$(\alpha \wedge \beta) \rhd \gamma \;\&\; \exists\delta\, [\, \alpha \rhd \delta \;\&\; \delta \vdash_\Sigma \beta \,]$	$\Rightarrow$	$\alpha \rhd \gamma$
E-Cut	$(\alpha \wedge \beta) \rhd \gamma$	$\Rightarrow$	$\beta \rhd \gamma$

Appendice B

Relationships between explanatory and consequence relation

$\triangleright$		$\mid\!\sim_{ab}$
	$\Rightarrow$	Adequate
E-Con$_\Sigma$	$\Rightarrow$	Con$_\Sigma$
LLE$_\Sigma$+ E-CM+ E-C-Cut	$\Rightarrow$	Cumulative
LLE$_\Sigma$+ E-CM+ E-C-Cut+ RA	$\Rightarrow$	Preferential
LLE$_\Sigma$+ E-CM+ E-C-Cut+RA+ LOR+ finite language	$\Rightarrow$	Preferential + W-DR
LLE$_\Sigma$+ E-CM+ E-C-Cut+RA+ E-DR	$\Rightarrow$	Preferential + DR
LLE$_\Sigma$+ E-CM+ E-C-Cut+RA+ E-R-Cut	$\Rightarrow$	Rational
E-Cut	$\Rightarrow$	Monotonic

From explanatory relations to consequence relations

$\tilde{\triangleright}$		$\mid\!\sim$ adequate
E-Con$_\Sigma$	$\Leftarrow$	Con$_\Sigma$
LLE$_\Sigma$+ E-CM+ E-C-Cut+ RA	$\Leftarrow$	Preferential
LLE$_\Sigma$+ E-CM+ E-C-Cut+RA+ LOR	$\Leftarrow$	Preferential + W-DR
LLE$_\Sigma$+ E-CM+ E-C-Cut+RA+ E-DR	$\Leftarrow$	Preferential + DR
LLE$_\Sigma$+ E-CM+ E-C-Cut+RA+ E-R-Cut	$\Leftarrow$	Rational
E-Cut	$\Leftarrow$	Monotonic

From consequence relations to explanatory relations

Appendice C

Update algorithm

Let P be a fixed program which in this context can be seen as our background theory or our integrity constraints. Let L be a set of facts which can be seen as our beliefs about the world. We would like to define the change produced by a set of facts L' coding a new piece of information about the world. The following definition describes the result of this change:

$$L \diamond_P L' = \begin{cases} Lit & \text{if } L \text{ or } L' \text{ is not } P\text{-consistent} \\ \langle L_1 \cup L', \ldots, L_n \cup L' \rangle & \text{otherwise} \end{cases}$$

where $\{L_1, \ldots, L_n\}$ is the set of subsets of L which are maximal and $P \cup L'$-consistent.

This is computed in two steps: first, we compute sets of facts that lead to inconsistency, called *contradictory sets*. Then, given the set SC of contradictory sets, we compute the minimal hitting sets of SC. The maximal subsets of L such that $L \cup L'$ is P-consistent are the sets $L \setminus H$, where H is a minimal hitting set of SC.

First step. A contradictory set C is a subset of L corresponding to a way of proving a pair of opposite literals from $P \cup L \cup L'$: C is a contradictory set if C is a subset of L and there exists a minimal subset P' of P such that $P' \cup C \cup L'$ is not consistent and, for each l in C, l appears in the body of a rule of P'.

We assume that, for every atom a that appears in the knowledge base, we have an implicit rule $a, \neg a \rightarrow \perp$.

To compute the contradictory sets when updating L with L', we build a contradiction tree $T_{L,L'}$, starting from $\perp$: a node is a pair$(L; C)$, where L is a list of literals to prove to obtain contradiction and C is a partial contradictory set. We start with the node $(\perp; \{\}\ )$. Let $N = (l_1, l_2, \ldots, l_n; C)$ be a node of $T_{L,L'}$. The successors of N are computed as follows:

- if $l_1 \in L'$ or $l_1 \in P$ or if l_1 is already in C, then $(l_2, \ldots, l_n; C)$ is the only successor of N
- else for each rule $g_1, g_2, \ldots, g_p \rightarrow l_1$, $(g_1, g_2, \ldots, g_p, l_2, \ldots, l_n$; C is a successor of N and if $l_1 \in L$, then $(l_2, \ldots, l_n; C \cup \{l_1\}))$ is a successor of N.

A branch terminates with an empty list of literals or with a node that cannot be developed. If a branch ends with $(\emptyset; C)$, then C is a contradictory set of facts. Note that if we suppose that

L' is consistent with P, we can't obtain $(\emptyset; \emptyset)$. $T_{L,L'}$ doesn't give only the minimal contradictory sets, but all the ways to entail a pair of opposite literals from P and $L \cup L'$.

Example. We consider the program P and the set of literals L, with $P = \{a, b \rightarrow c \; ; \; a, d \rightarrow c\}$ and $L = \{a, b, d\}$. When updating L with $\{\neg c\}$, we obtain two contradictory sets $\{a, d\}$ and $\{a, b\}$. Fig 1 shows the contradiction tree (to simplify, we consider only the rule $\neg c, c \rightarrow \perp$ at the first step, since the only pair of contradictory literals that actually appears in this case is $c, \neg c$).

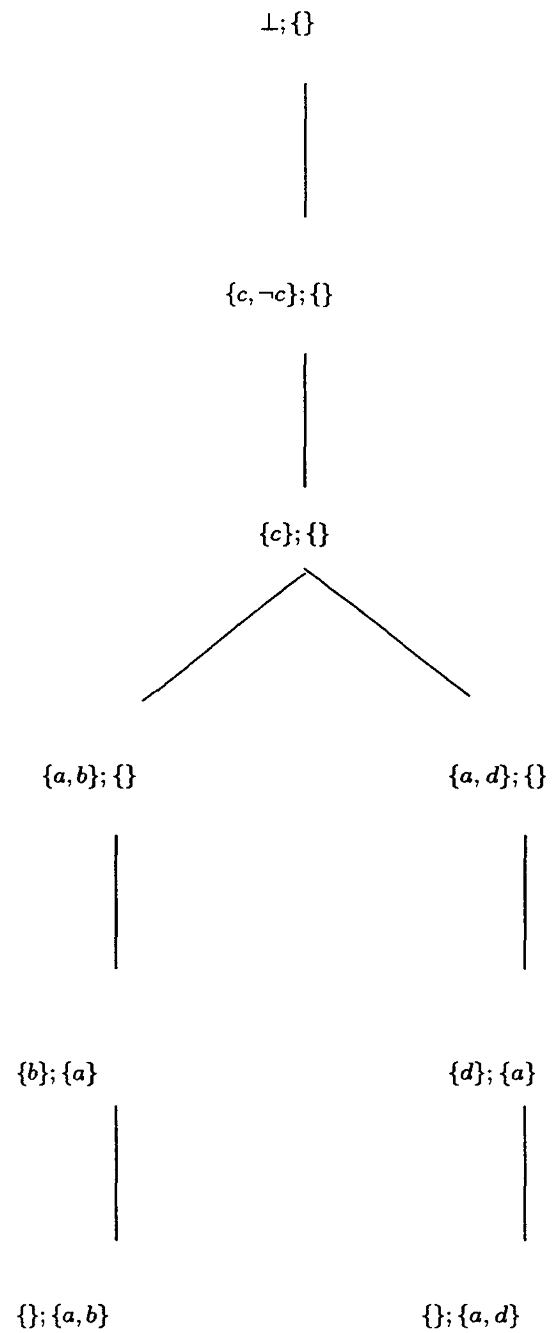
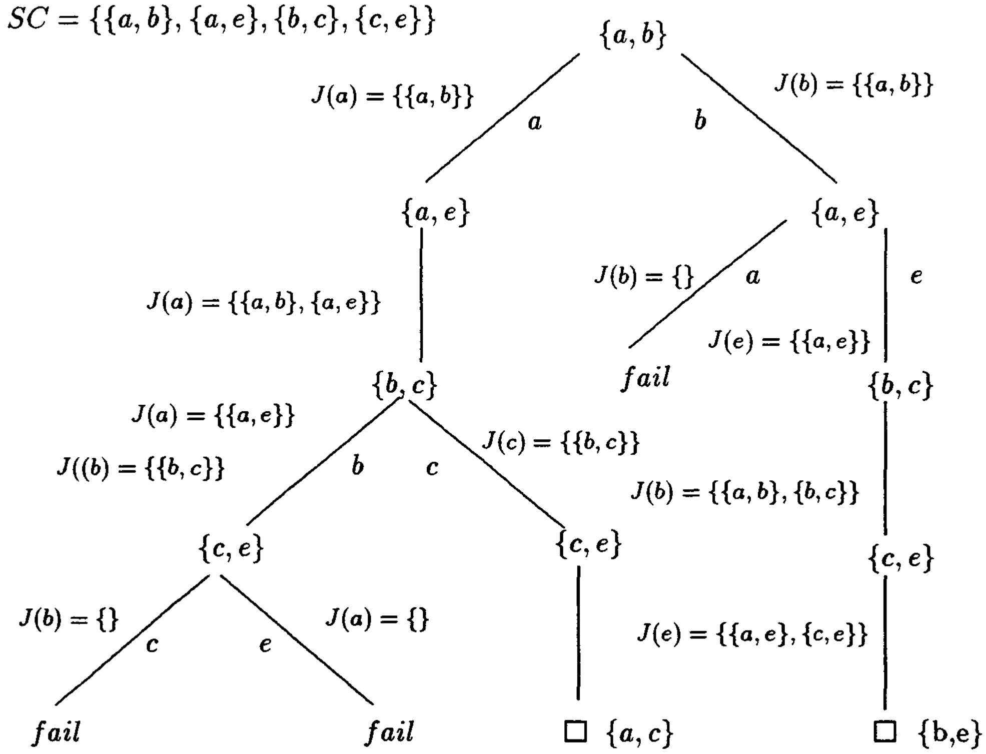


Fig 1. Contradiction tree

Second step. The contradiction tree produces a set of contradictory sets $SC = \{C_1, ..., C_n\}$. To update L with L' we compute all the maximal subsets S of L such that $S \cup L'$ is P-consistent. The subsets S of L are obtained by removing from L at least one element of each contradictory set: if H is a set of facts such that, for each element C_i of SC, $S \cap C_i \neq \emptyset$, then $(L \setminus H) \cup L'$ is consistent. H is usually called a hitting set of SC. To find the maximal consistent subsets of L, we need all the minimal hitting sets (by set inclusion) of SC.

The figure 2 illustrates the algorithm we implemented in Prolog to compute the minimal hitting sets.



Minimal Hitting sets: $\{a, c\}$, $\{b, e\}$

This algorithm is very close to the one given by Reiter in [98]. Let $SC = \{C_1, C_2, ..., C_n\}$. We try to construct a hitting set of SC by examining the elements of SC one by one: if the current set C_i is not already hit by the partial hitting set HS, we add one of the literals of C_i in HS. To know if a hitting set is minimal, we maintain a set of justifications $J(l)$ for each literal l of the hitting set: $J(l)$ contains all the sets of literals that are hit only by l. When a new literal l must be added to the hitting set, the justification sets are updated by removing all the sets containing l. If one of the justification sets becomes empty, the current hitting set is not minimal anymore and so the construction fails.

Concerning the relationships between our algorithm and Reiter's notice that we construct the same kind of HS-tree, where nodes are labeled with elements of SC and edges are labeled with elements of the hitting sets. The main difference is the use of justification sets instead of tree pruning. Tree pruning is used in Reiter's algorithm in order to compute not all the hitting sets but only the minimal ones. In our algorithm, this is done with justification sets and each minimal hitting set is computed in a unique branch.

Appendice D

Historique des publications des travaux qui constituent ce mémoire

CHAPITRE 3

Une version préliminaire est:

[1] H. BEZZAZI, R. PINO PÉREZ (1996), Rational Transitivity and its models. *Proceedings of the 26th International Symposium on Multiple-Valued Logics*, Santiago de Compostela, Spain. May 29-31. IEEE Computer Society Press, 1996. pp 160-165.

La version du memoire est

[2] H. BEZZAZI,D. MAKINSON, R. PINO PÉREZ (1996), Beyond Rational Monotony: Some strong non-Horn rules for nonmonotonic inference relations. **Journal of Logic and Computation**, vol. 7, pp 605-631. 1997.

CHAPITRE 4

Une version préliminaire est

[3] R. PINO PÉREZ, C. UZCÁTEGUI. On representation theorems for nonmonotonic inference relations. Rapport de recherche LIFL, No IT-300, Université de Lille I, 1997. 12pages.

La version du mémoire est une version enrichie de ce rapport à paraître dans le **Journal of Symbolic Logic.**

CHAPITRE 5

Une version préliminaire est

[4] R. PINO PÉREZ, C. UZCÁTEGUI. Jumping to explanations vs jumping to conclusions. Rapport de recherche LIFL, No IT-301, Université de Lille I, 1997. 29 pages.

Une première version a été présenté à WOLLIC'97 Workshop on Logic, Language, Information and Computation. Fortaleza, Brasil, August 1997.

Une version partielle est

[5] C. UZCÁTEGUI AND R. PINO PÉREZ. Abduction vs. Deduction in Nonmonotonic Reasoning. In *Proceedings of the Seventh International Workshop on Nonmonotonic Reasoning, NR'98.* Trento, Italy. May 30-June 1, 1998, pp 42-54.

Ce chapitre a été soumis à la revue **AI Journal**. Ils demandent une version corrigée qui est en révision actuellement. C'est la version du mémoire à peu de choses près.

CHAPITRE 6

Une première version est

[6] H. BEZZAZI, S. JANOT, S. KONIECZNY AND R. PINO PÉREZ. Forward chaining and change operators. Proceedings of DYNAMICS'97 Workshop on (Trans)Actions and Change in Logic Programming and Deductive Databases held in conjonction with the International Logic Programming Symposium ILPS'97. Port Jefferson, New York, USA, October 1997. pp. 135-146.

La version du mémoire est

[7] H. BEZZAZI, S. JANOT, S. KONIECZNY AND R. PINO PÉREZ. Analysing rational properties of change operators based on forward chaining. In *Transactions and Change in Logic Databases.* H. Decker, B. Freitag, M. Kifer and A. Voronkov, Eds. **Lecture Notes in Computer Science**, vol. 1472. 1998.

CHAPITRE 7

Une première version est

[8] S. KONIECZNY AND R. PINO PÉREZ. On the difference between arbitration and majority merging. Proceedings of WOLLIC'97 Workshop on Logic, Language, Information and Computation. Fortaleza, Brasil, August 1997.

La version du mémoire est

[9] S. KONIECZNY AND R. PINO PÉREZ. On the logic of Merging. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation And Reasoning, KR'98.* Trento, Italy. June 2-5, 1998, pp 488-498.

Appendice E

Curriculum Vitae

Nom: **Pino Pérez**
Prénom: **Ramón Augusto**
Date et lieu de naissance: **12 juin 1956 à La Paz (Bolivie)**
Situation de famille: **Marié, 2 enfants**
Adresse: **21, rue Royale 59800 Lille**
Téléphone : **03.20.31.67.87**

Fonction actuelle : **Maître de Conférences à L'Ecole Universitaire d'Ingénieurs de Lille (EUDIL) à L'Université de Lille I**
Adresse professionnelle : **Université de Lille I, LIFL Bât. M3, Cité Scientifique, 59655 Villeneuve d'Ascq.**
Tél. : **03.20.43.67.25** Fax : **03.20.43.65.66** E-mail : **pino@lifl.fr**

Etudes réalisées

1973 Baccalauréat Scientifique. Mention Très Bien. Mérida, Vénézuéla.

1978 Licence de Mathématiques. Mention Très Bien. Université Simón Bolívar. Caracas, Vénézuéla.

1979 Diplôme de Français et Civilisation Française. Mention Très Honorable. Alliance Française. Paris.

1980 DEA de Logique Mathématique. Mention Assez Bien. Sous la direction de Gabriel Sabbagh. Université Paris 7.

1983 Doctorat de Troisième Cycle en Logique Mathématique. Mention Très Honorable. Sous la direction de Jean-Pierre Ressayre. Université Paris 7. Juin 1983.

1986 DEA d'Informatique Fondamentale. Mention Très Bien. Sous la direction de Pierre-Louis Curien. Université Paris 7.

1992 Doctorat d'Informatique. Mention Très Honorable. Sous la direction de Pierre-Louis Curien. Université Paris 7. Novembre 1992.

Postes occupés

75-78 Assistant de Travaux Dirigés de Mathématiques. Université Simón Bolívar. Caracas, Vénézuéla.

83-85 Professeur de Mathématiques. Université Simón Bolívar. Caracas, Vénézuéla.

86-87 Maître de Conférences Associé en Informatique. Ecole Normale Supérieure (Ulm). Paris.

87-88 Maître de Conférences Associé en Informatique. Université Paris 13. Paris.

88- Maître de Conférences en Informatique. Université Lille I (EUDIL). Lille.

Activités d'enseignement

1. Pendant trois ans (75-78) en tant qu'Assistant de Travaux Dirigés de Mathématiques à l'Université Simón Bolívar j'ai eu à assurer des séances des TD d'Analyse et d'Algèbre dans les deux premières anneés de licence (l'équivalent d'un DEUG).

2. Pendant deux ans (83-85) en tant que Professeur de Mathématiques à l'Université Simón Bolívar j'ai été chargé de :

   - Cours de Logique et d'Analyse Complexe aux étudiants en Licence de Mathématiques.
   - Le cours de Logique pour l'Informatique aux élèves ingénieurs en Informatique.
   - Le cours d'informatique Théorique aux étudiants du "Master en Ciencias de la Computación". Notamment un cours intitulé *Elements théoriques pour l'Intelligence Artificielle.* Le contenu de ce cours portait sur le Principe de Résolution, le λ-calcul et la Logique Modale.

3. Pendant un an (86-87) en tant que Maître de Conférences Associé en Informatique à l'Ecole Normale Supérieure (Ulm) j'ai eu à assurer les séances des TD et des TP du cours d'Informatique 1 du M.M.F.A.I. (Magistère de Mathématiques Fondamentales et Appliquées et d'Informatique). Le contenu du cours portait sur l'environnement informatique, Lisp, Prolog, Algorithmique et structures des données et Mécanisation du raisonnement.

4. Pendant un an (87-88) en tant Maître de Conférences Associé en Informatique à l'Université Paris 13 j'ai été chargé de :

   - Le cours de Logique et Mécanisation du Raisonnement de la Licence en Informatique.
   - Les TD des cours de Langages Formels et de Compilation de la Maîtrise en Informatique.

5. Depuis 1988 année où j'ai commencé à travailler à l'Université Lille 1 en tant que Maître de Conférences j'ai été chargé des cours suivants :

   - A l'EUDIL (Ecole Universitaire d'Ingénieurs de Lille) :

(a) Algorithmique et Structures de Donnés (année 88-89). Cours de la première anné IMA.

(b) Techniques de Base en Intelligence Artificielle (année 88-89). Cours de la deuxième anné IMA.

(c) Mathématiques pour l'Informatique (années 89-91). Cours de la première année IMA. Ce cours dont j'ai fait un polycopié a été conçu par moi même et j'en ai la responsabilité depuis sa création en 1989.

(d) Informatique Théorique (années 89-91). Cours de la deuxième année IMA.

(e) TP de Compilation et de Techniques de base en Intelligence Artificielle (depuis 93).

(f) TP d'algorithmique en langage C (1998).

(g) TP de Bases de données en SQL (1998).

- Cours sur le λ-calcul dans le cadre du DEA d'Informatique (89-93).
- Conférences de formation dans le cadre du DEA d'Informatique sur la théorie de la révision et la logique non monotone (Depuis 94)
- Cours de DEA sur les relations de conséquence non monotones et le raisonnement révisable (1998).

Activités de recherche

Mots clés : *Logiques pour l'intelligence artificielle, raisonnement non-monotone, théorie de la révision de la connaissance, mise-à-jour, abduction, fusion, arbitrage, lambda-calcul, partialité*

I. Je fais ici un résumé historique succinct concernant mes recherches jusqu'à l'année 1993. Pour les dernières années voir l'appendice F.

De 1980 à 1983 des recherches en logique : théorie généralisée de la récursivité, problèmes d'indépendence et techniques de théorie des modèles. Ces recherches ont été faites au sein de l'équipe de logique de Paris 7 (thèse de 3ème cycle).

De 1983 à 1985 suite des recherches sur les travaux de ma thèse et début de recherches sur les fondements logiques de la programmation (principe de résolution de Robinson d'une part et le λ-calcul d'autre part). Ces recherches ont été faites à l'Université Simón Bolívar à Caracas.

De 1985 à 1993 des recherches sur la logique des langages fonctionnels en particulier sur la sémantique opérationnelle et dénotationnelle du λ-calcul et de quelques variantes de celui-ci. D'abord au Laboratoire d'Informatique de l'Ecole Normale Supérieure (LIENS) URA CNRS 1327 et à partir de 1988 au Laboratoire d'Informatique Fondamentale de Lille (LIFL) URA CNRS 369.

Pendant la période de 88 à 93 j'animai une petite équipe au LIFL sur cette thématique. Le thème central de cette équipe était plus particulièrement le λ-calcul partiel. Cet outil offre un cadre théorique pour les preuves d'équivalence de programmes fonctionnels éxécutés par le mécanisme d'appel par valeur. Parmi les résultats obtenus de cette époque nous pouvons citer

- Equivalence entre le λ-calcul partiel et les catégories cartésiennes fermées partielles (**pccc** en abregé). Ce résultat est basé sur une présentation équationnelle des **pccc** due à Curien et Obtułowicz.

- Décidabilité d'une sous-théorie du λ-calcul partiel. Ce travail joint à des résultats de E. Moggi permet de prouver la cohérence du calcul.
- Construction d'un modèle standard du λ-calcul partiel.
- Liens entre le λ-calcul partiel et la logique partielle.
- Construction d'algèbres combinatoires partielles extensionnelles.

Une partie de ces travaux a constitué mon mémoire de thèse en informatique, soutenue en 1992 à l'Université de Paris 7.

J'ai aussi pendant cette période dirigé la thèse de C. Even. Il l'a soutenue en février 1993 à l'Université de Lille I.

II. Depuis 1993 je dirige un groupe de travail sur la théorie de la révision des croyances , le raisonnement non monotone et la représentation des connaissances au sein de l'équipe Méthéol du LIFL (voir l'appendice F pour plus de détails concernant la recherche des 5 dernières années).

Travaux et publications

Chapitre de livre

1. H. Bezzazi, S. Janot, S. Konieczny and **R. Pino Pérez**. Analysing rational properties of change operators based on forward chaining. In *Transactions and Change in Logic Databases*. H. Decker, B. Freitag, M. Kifer and A. Voronkov, Eds. **Lecture Notes in Computer Science**, vol. 1472 (in press). 1998.

Revues internationales

2. **R. Pino Pérez**. Decidability of the Equational Restriction Theory in the Partial λ-calculus. **Theoretical Computer Science.** Vol. 67 pp 129-139. 1989.
3. C. Even and **R. Pino Pérez**. Extension of Scott-Koymans theorem to partial framework. **The Bulletin of Symbolic Logic**, vol **1** n° **4**, pp. 489-490, 1995.
4. H. Bezzazi, D. Makinson and **R. Pino Pérez**. Beyond Rational Monotony: Some strong non-Horn rules for nonmonotonic inference relations. **Journal of Logic and Computation**, vol. 7, pp 605-631. 1997.
5. **R. Pino Pérez** and C. Uzcátegui. On representation theorems for nonmonotonic inference relations. **Journal of Symbolic Logic.** A paraître.

Congrès, Colloques et Ateliers internationaux avec comité de lecture et actes

6. **R. Pino Pérez** and J.P. Ressayre. Definable Ultrafilters and Elementary End Extensions. In *Methods in Mathematical Logic: proceedings of the 6th Latin American Symposium of Mathematical Logic.* Caracas, Venezuela, August 1-6 1983. C. di Prisco Ed. **Lecture Notes in Mathematics** Vol. 1130, 1985. pp. 341-350.

7. **R. Pino Pérez**. An Extensional Partial Combinatory Algebra based on λ-terms. In *Proceedings of the 16th International Symposium on Mathematical Foundations of Computer Science 1991, MFCS'91.* Kazimierz Dolny, Poland, September 9-13, 1991. A.Tarlecki Ed. **Lecture Notes in Computer Science**, vol. 520. 1991. pp. 387-396.

8. C. Even and **R. Pino Pérez.** λ_p-calculus and algebras with partial elements. In *Proceedings of the XII International Conference of the Chilean Computer Science Society.* Santiago, Chile, October 14-16, 1992. S. Mujica and A. Viña Eds. Sociedad Chilena de Ciencias de la Computación, 1992. pp. 71-84.

9. **R. Pino Pérez** and C. Even. An abstract property of confluence applied to the study of the Lazy Partial Lambda Calculus. In *Proceedings of the 3th International Symposium on Logical Foundations in Computer Science*, Saint Petersburg, Russia. July 11-14, 1994. A. Nerode and Yu. Matiyasevich Eds. **Lecture Notes in Computer Science 813**, 1994. pp. 278-290.

10. H. Bezzazi and **R. Pino Pérez**. Rational Transitivity and its models. In *Proceedings of the 26th International Symposium on Multiple-Valued Logics*, Santiago de Compostela, Spain. May 29-31, 1996. IEEE Computer Society Press, 1996. pp. 160-165.

11. H. Bezzazi, S. Janot, S. Konieczny and **R. Pino Pérez**. Forward chaining and change operators. In *Proceedings of DYNAMICS'97 Workshop on (Trans)Actions and Change in Logic Programming and Deductive Databases held in conjonction with the International Logic Programming Symposium, ILPS'97.* Port Jefferson, New York, USA, October 1997. pp. 135-146.

12. S. Konieczny and **R. Pino Pérez.** On the logic of Merging. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation And Reasoning, KR'98.* Trento, Italy. June 2-5, 1998, pp 488-498.

13. C. Uzcátegui and **R. Pino Pérez**. Abduction vs. Deduction in Nonmonotonic Reasoning. In *Proceedings of the Seventh International Workshop on Nonmonotonic Reasoning, NR'98.* Trento, Italy. May 30-June 1, 1998, pp 42-54.

Congrès, Colloques et Ateliers internationaux avec comité de lecture sans actes

14. **R. Pino Pérez**. Non standard Methods in Set Theory. Travail présenté au Logic Colloquium'84. Manchester. 1984. 8 pages.

15. **R. Pino Pérez.** A Strict Partial Combinatory Algebra which modelizes Partial Lambda Calculus. Présenté à l'atelier de travail du Projet Stimulation sur le λ-calcul typé. Paris 1-6 février 1991. 12 pages.

16. S. Konieczny and **R. Pino Pérez.** On the difference between arbitration and majority merging. Proceedings of WOLLIC'97 Workshop on Logic, Language, Information and Computation. Fortaleza, Brasil, August 1997. 15 pages.

17. **R. Pino Pérez** and C. Uzcátegui. Jumping to explanations vs jumping to conclusions. Le résumé est dans Proceedings of WOLLIC'97, 4th Workshop on Logic, Language, Information and Computation. Fortaleza, Brasil, August 1997.

18. S. Konieczny and **R. Pino Pérez**. An analysis of merging from a logical point of view. Travail présenté dans le XI Simposio Latinoamericano de Lógica Matemática. Mérida, Venezuela, julio 1998.

Rapports de Recherche soumis à publication

19. **R. Pino Pérez** and C. Uzcátegui. Jumping to explanations vs jumping to conclusions. Rapport de recherche LIFL, No IT-301, Université de Lille I, 1997. 29 pages.*En révision dans le AI Journal.* (La version soumise a 45 pages)
20. S. Konieczny and **R. Pino Pérez**. Merging with integrity constraints. Rapport de recherche LIFL No 99-01, Université de Lille I, 1999. Soumis à **Ecsqaru'99**.

Monographies de recherche

21. **R. Pino Pérez**. Π_n-collection, indicatrices et ultrafiltres définissables. **Thèse de Troisième Cycle en Logique Mathématique.** Université Paris VII. 1983.
22. **R. Pino Pérez**. Le modèle de projections finitaires et l'égalité de Types dans une classe de Modèles du λ-calcul polymorphe. Mémoire de DEA en Informatique. Université Paris 7. 1986.
23. **R. Pino Pérez**. Contribution à l'étude du Lambda Calcul Partiel. **Thèse de Doctorat en Informatique.** Université Paris VII. 12 novembre 1992.

Rapports de Recherche

24. **R. Pino Pérez**. Le lambda calcul partiel. Rapport de recherche LIENS No 87-11. Ecole Normale Supérieure. Paris. 1987. 24 pages.
25. **R. Pino Pérez**. A standard model for the partial lambda calculus. Rapport de recherche du LIFL, No IT-170. Université de Lille I. 1989. 11 pages.
26. **R. Pino Pérez**. Semantics of Partial Lambda Calculus. Rapport interne LIFL, No IT-212. Université de Lille I. 1991. 14 pages.

Monographies de type pédagogique

27. **R. Pino Pérez**. *Introducción a la Lógica de primer orden.* Université Simón Bolívar. Caracas. 1985.
28. **R. Pino Pérez**. *Cours de Logique.* Trois volumes. Vol. 1: *Calcul des Propositions.* Vol. 2: *Calcul des Prédicats.* Vol.3: *La méthode de Résolution. Vers une mécanisation du raisonnement.* . Université Paris 13. 1987.
29. **R. Pino Pérez**. *Mathématiques pour l'Informatique.* EUDIL. Université de Lille 1. 1990.
30. **R. Pino Pérez**. *Lógica parcialidad y computabilidad.* Notes d'un cours donné comme conférencié invié à un symposium sur les fondements logiques de l'informatique organisé par l'Institut Vénézuélien de la Recherche Scientifique (IVIC) en février 1992.

31. **R. Pino Pérez.** *Lambda Calcul.* Notes de cours de DEA. Université de Lille I. 1992.

Participation aux séminaires

80-83 Séminaire de Logique de l'Université Paris 7.

83-85 Séminaire de Logique de Caracas (Université Central, Université Simón Bolívar et IVIC)

85-86 Séminaire de Logique de l'Université Paris 7.

86-88 Séminaire de Logique et sémantique de la programmation. Ecole Normale Supérieure-Université de Paris 7. Séminaire sur le λ-calcul typé Université de Paris 7.

88-93 Groupe de travail sur le λ-calcul. Université de Lille 1.

94-97 Groupe GNOM et CALC (anciennement ALP). Université de Lille 1.

Organisation des séminaires

- Organisation du séminaire de Logique de Caracas pendant deux ans (83-85) en collaboration avec le Dr. Carlos Di Prisco.
- De 88 à 93, organisation du Groupe de travail sur le λ-calcul à l'Université de Lille 1.
- Depuis 94, organisation du Groupe de travail GNOM sur la théorie de la révision, les relations de conséquence non classique et la représentation des connaissances (voir programmes annexes).

Direction des recherches

- Direction du mémoire de DEA d'Informatique de Gabriel Desmet. Université de Lille 1. 1989.
- Direction du mémoire de DEA d'Informatique de Olivier Dubuisson. Université de Lille 1. 1991.
- Direction du mémoire de DEA d'Informatique de Thierry Peltier. Université de Lille 1. 1991.
- Direction de la **thèse** de Christian Even. Université de Lille 1. **1993**.
- Direction du mémoire de DEA d'Informatique de Omar Malah. Université de Lille 1. 1994.
- Direction du mémoire de DEA d'Informatique de Sébastien Konieczny. Université de Lille 1. 1996.
- Direction du mémoire de DEA d'Informatique de Christophe Parent. Université de Lille 1. 1997.

- Direction de la **thèse** de Sébastien Konieczny. Université de Lille 1. Soutenance prevue en 99.
- Direction du groupe GNOM au sein de Méthéol au Laboratoire d'Informatique Fondamentale de Lille, URA CNRS 369. Voir l'appendice F.

Participation aux jurys de thèses

1. Thèse de Nathalie Devesa. *Proposition d'un schéma d'évaluation parallèle du langage fonctionnel FP sur un réseau de processeurs.* Université de Lille 1. Le 16 janvier 1990.
2. Thèse de Hassan Bezazzi. *Types de données et recurrence bien fondée dans un système de programmation par preuves.* Université de Lille 1. Le 4 juillet 1990.
3. Thèse de Christian Even. *Autour du Lambda Calcul Partiel.* Université de Lille 1. Février 1993.
4. Thèse de Nadia Benani. *Proposition d'un modèle d'évaluation parallèle pour les langages fonctionnels sans variables.* Université de Lille 1. Novembre 1994.

Responsabilités

1. Evaluation des recherches :
   - 'Referee' d'articles pour le revue *Journal of IGPL* dont le titre est depuis peu Logic Journal of IGPL et est publié par Oxford University Press.
   - 'Referee' d'articles pour des congrès internationaux notamment Caap (97, 96), STACS (99).
   - 'Referee' d'articles pour des congrès nationaux notamment RFIA (Reconnaissance de Formes et Intelligence Artificielle 98).
   - 'Reviewer' de Mathematical Reviews.
2. Membre élu de la CSE de la section 27 de l'Université de Lille I (91-97) .
3. Responsable pédagogique de la deuxième anné de l'option informatique du département IMA de l'Ecole Universitaire d'Ingénieurs de Lille de 1995 a 1997.

Collaborations Internationales

1. Nous collaborons avec le Professeur Carlos Uzcátegui, de l'Université des Andes (Mérida, Venezuela). Dans le cadre de cette collaboration le Professeur Uzcátegui a été invité au LIFL à travailler dans notre groupe pendant 4 mois de l'année 96.
2. Nos avons entamé une collaboration avec le Professeur Jorge Lobo de Bell Labs (New York)[1]. Dans le cadre de cette collaboration le Professeur Lobo a été invité au LIFL à travailler dans notre groupe pendant 2 mois à la fin du printemps 98.

[1] Anciennement Professeur à l'Université d'Illinois (Chicago).

Collaborations Nationales

1. En tant que composante du projet LOGIDIS (Logique distribuée et coopération) nous participons au programme interlaboratoires de la région, Ganymède, dont le thème central est le travail coopératif.
2. Nous avons des rapports privilégiés avec le Centre de Recherches en Informatique de Lens (CRIL) qui participe également au programme Ganymède. Les thèmes sur lesquels nous collaborons se situent autour des logiques non monotones, lesquelles sont en relation étroite avec la théorie de la révision de la connaissance.
3. Nous travaillons avec le Dr. David Makinson (UNESCO), un des pionniers dans la théorie de la révision. Dans ce cadre nous avons fait un article en collaboration.
4. Participation au PRC-I3 (créé en 1997), groupe modèles du raisonnement.

Invitations

- Invité par l'Institut Venezuelien des Recherches Scientifiques (IVIC) et par l'Université Centrale de Venezuela à donner une série de conférences sur le λ-calcul pendant la première quinzaine de février 92.
- Invité par L'Universié d'Asunción au Paraguay à donner un cours d'informatique théorique pendant le mois de septembre de 1993.
- Invité une semaine en décembre 1997 par l'Université des Andes, Mérida, Venezuela pour des collaboration scientifiques avec le Professeur C. Uzcátegui.
- Invité à donner une conférence dans le *Eleventh Latin American Symposium on Mathematical Logic*. 6-12 Juillet 1998, Mérida, Venezuela.
- Invité à donner une conférence dans le meeting FUSION, Condom, France, 28 novembre au 1er décembre 1998.

Distinctions

- Boursier Gran Mariscal de Ayacucho: 1974-1978 (Universidad Simón Bolívar).
- Boursier Foninves: 1980-1983 (Universidad Paris 7).
- Allocataire de la prime d'encadrement doctoral et de recherche de 1993 à 1997.
- Participation dans le programme TALVEN de l'UNESCO.

Appendice F

Rapport d'activités de l'équipe GNOM (Groupe NOn Monotone), Années 93-98

Composition

- Ramón Pino Pérez, Maître de Conférences, Université Lille I (responsable).
- Hassan Bezzazi, Maître de Conférences, Université Lille II.
- Stéphane Janot, Maître de Conférences, Université Lille I.
- Sébastien Konieczny, Thésard (bourse BDI CNRS-Région), Université LilleI.
- Omar Malah[1], Thésard, Université LilleI.

Mots clés : *révision, fusion, arbitrage, abduction, mise à jour de la connaissance, coopération, raisonnement non monotone.*

Thématique

Le thème de recherche de cette équipe est l'étude de la dynamique de l'information. En particulier nous nous intéressons à la théorie de la révision de la connaissance et au raisonnement non monotone. Ce sujet se situe dans le domaine de l'Intelligence Artificielle et a des connexions étroites avec les sciences cognitives, la logique et les bases de données.

La question à laquelle la théorie de la révision de la connaissance veut répondre peut se résumer très succinctement de la façon suivante: que faut-il faire pour ajouter une nouvelle information à une base de connaissances de telle sorte que l'on puisse garder le maximum de l'ancienne information tout en restant cohérent?

Dans des situations diverses, les données (les informations) sont dans un flux continu, elles sont dynamiques : de nouvelles données apparaissent et l'ancienne connaissance devient, dans beaucoup de cas, obsolète ou pour le moins révisable. Une théorie de la connaissance dynamique est donc importante si l'on veut rendre compte du changement dans les systèmes

[1]Suite à des problèmes de financement Malah n'a pas pu continuer en thèse à la rentrée 96. Il a créé une PME.

à bases de connaissances.

Depuis les travaux de Gärdenfors et Makinson[2] il est bien connu qu'il existe une dualité entre certains opérateurs de révision et certaines logiques non monotones. Ainsi l'étude des logiques non monotones et de leurs propriétés donne des informations importantes sur la manière de réviser et sur les algorithmes sous-jacents pour calculer les révisions. Un de nos buts est donc de comprendre mieux les opérateurs de révision via l'analyse des propriétés de logiques telles que la logique rationnelle introduite par Lehmann et Magidor[3].

Un cadre où les informations sont intrinsèquement dynamiques est celui où plusieurs agents rationnels coopèrent (échangent de l'information). Dans ce contexte il faut avoir des mécanismes précis qui décrivent le résultat de ce processus. La théorie de la révision de la connaissance s'avère un outil conceptuel très riche pour modéliser ces mécanismes. Ainsi un autre de nos buts est l'étude de ces outils issus de la théorie de la connaissance révisable. Parmi eux on trouve : la révision (proprement dite), l'arbitrage, l'abduction et la mise à jour. Très sommairement la révision correspond à privilégier la nouvelle information et à *minimiser* les changements que l'on doit opérer dans l'ancienne information. L'arbitrage doit trouver l'information qui soit le plus en *accord* avec un ensemble d'informations. L'abduction (et d'une certaine manière la mise à jour) consiste à trouver les informations les plus *pertinentes* à ajouter en vue de produire une information donnée.

Il s'agit pour nous de dégager les situations où ces outils sont pertinents, sa possibilité de réelle utilisation informatique (problèmes de faible complexité algorithmique) et d'introduire des nouvelles notions en théorie de la connaissance pour traiter des situations qui sortent du cadre des opérateurs de révision, d'abduction, d'arbitrage et de mise à jour.

Un dernier point auquel on s'intéresse ce sont les différentes façons d'agencer ces outils pour avoir des systèmes cohérents dans un cadre où plusieurs agents doivent coopérer (on peut imaginer un groupe d'experts qui collaborent dans la résolution d'un problème). On essaie de dégager les meilleures techniques, ainsi que le cadre précis où ces techniques sont applicables. Des plate-formes mises en œuvre en Bivouac et en Prolog permettront d'évaluer ces différentes techniques.

Résultats

Dans le volet théorique de notre travail nous citons :

- Les travaux de Bezzazi et Pino Pérez [2] sur la modélisation de la transitivité dans un cadre non monotone.
- Les travaux de Bezzazi, Makinson et Pino Pérez [3] sur la modélisation des relations de conséquence non monotones qui satisfont des règles non-Horn plus fortes que la monotonie rationnelle.
- Les travaux de Pino Pérez et Uzcátegui [6] sur des nouvelles techniques de représentation des relations non monotones.
- Les travaux de Pino Pérez et Uzcátegui [7] sur les rapports existant entre le raisonnement abductif et les relations non monotones. Ces résultats sont très encourageants car cela

[2] P. Gärdenfors, D. Makinson Relations between the logic of theory change and nonmonotonic logic. In *The Logic of Theory Change, Workshop, Konstanz, FRG, Octuber 1989.* **Lecture Notes in Artificial Intelligence** 465, pp. 185-205.

[3] D. Lehmann, M. Magidor, What does a conditional knowledge base entail? **Artificial Intelligence** 55 pp. 1-60, 1992

va permettre d'appliquer les algorithmes issus des relations non monotones (v.g. des relations rationnelles) au diagnostic.

- Les travaux de Konieczny et Pino Pérez [10] sur la logique de la fusion.

Dans le volet pratique, nous pouvons citer :

- Les travaux de Malah [1] sur des implantations d'un des algorithmes de calcul des faits déductibles dans des systèmes d'inférence non monotones.
- Les travaux de Konieczny [4] et [5] sur une architecture pour la coopération basée sur des opérateurs de révision.
- Les travaux de toute l'équipe [9] concernant des operateurs de révision basés sur le chainage avant.

Perspectives

Le futur de ce travail nous le situons dans une double perspective : d'une part les contributions théoriques à la théorie de la révision vont élargir le champ d'application de ces techniques. D'autre part, un des buts de notre groupe étant la réalisation d'une plate-forme modulaire rendant possible le choix du type de révision selon le contexte du problème, ce prototype permettra de développer des applications dans les domaines de la prise de décision lorsque plusieurs experts communiquent; en particulier dans le domaine médical et juridique.

Plus précisément nous envisageons :

A court terme :

- l'utilisation des résultats de Bezzazi et Pino Pérez dans le but de définir un algorithme pour calculer la "meilleure" relation rationnelle transitive contenant une base de connaissance finie;
- continuer des investigations sur l'abduction. Nous pensons que l'on peut extraire des travaux de Pino Pérez et Uzcátegui des algorithmes destinés à la recherche des meilleures explications. Ceci a des retombées importantes dans le domaine de la planification et du diagnostic.

A moyen terme:

- la réalisation d'un système expert modulaire où l'on puisse faire de l'inférence rationnelle, réviser une base de connaissance, mettre à jour d'une base de connaissance et finalement synthétiser l'information de plusieurs bases de connaissance.
- la réalisation d'un système expert qui modélise la coopération entre agents rationnels.

A plus long terme :

- dans le domaine de la dualité révision vs. logique non monotone, nous voulons étudier les opérateurs de révision qui se dégageraient de la programmation logique.
- de trouver des liens (équivalences) qui existeraient entre certaines techniques numériques pour faire de la synthèse de l'information et des techniques symboliques (logiques, comme l'arbitrage). Remarquons que des résultats qui mettent en rapport la logique rationnelle et la logique des possibilités ont été trouvés par Dubois et Prade [4].

Collaborations Internationales

1. Nous collaborons avec le Professeur Carlos Uzcátegui, de l'Université des Andes (Mérida, Venezuela). Dans le cadre de cette collaboration le Professeur Uzcátegui a été invité au

[4] D. Dubois, H. Prade, Possibilistic Logic, preferential models, non-monotonicity and related issues. In *Proceedings IJCAI'91* , 1991.

LIFL a travailler dans notre groupe pendant 4 mois de l'année 96. Deux travaux ont été rédigés suite à cette visite. Un troisième est en cours de rédaction.

2. Nos avons entamé une collaboration avec le Professeur Jorge Lobo de Bell Labs (New York)[5]. Dans le cadre de cette collaboration le Professeur Lobo a été invité au LIFL a travailler dans notre groupe pendant 2 mois à la fin du printemps 98.

Collaborations Nationales

1. En tant que composante du projet LOGIDIS (Logique distribuée et coopération) nous participons au programme interlaboratoires de la région, Ganymède, dont le thème central est le travail coopératif. Le sujet de la thèse de S. Konieczny est dans le cadre de ce projet. Cette thèse est cofinancée par la région.

2. Nous avons de rapports privilégiés avec le Centre de Recherches en Informatique de Lens (CRIL) qui participe également au programme Ganymède. Les thèmes sur lesquels nous collaborons se situent autour des logiques non monotones, lesquelles sont en relation étroite avec la théorie de la révision de la connaissance.

3. Nous travaillons avec le Dr. David Makinson (UNESCO), un des pionniers dans la théorie de la révision. Dans ce cadre nous avons fait un article en collaboration.

4. Participation au PRC-I3 (créé en 1997), groupe modèles du raisonnement.

Publications

[1] O. MALAH (1994) Un algorithme pour le calcul de la clôture rationnelle. Mémoire de DEA d'informatique. Université de Lille I. Septembre 1994.

[2] H. BEZZAZI, R. PINO PÉREZ (1996), Rational Transitivity and its models. *Proceedings of the 26th International Symposium on Multiple-Valued Logics*, Santiago de Compostela, Spain. May 29-31. IEEE Computer Society Press, 1996.

[3] H. BEZZAZI,D. MAKINSON, R. PINO PÉREZ (1996), Beyond Rational Monotony: Some strong non-Horn rules for nonmonotonic inference relations. **Journal of Logic and Computation**, vol. 7, pp 605-631. 1997.

[4] S. KONIECZNY (1996) Révision de la connaissance et coopération. Mémoire de DEA d'informatique. Université de Lille I. Juillet 1996.

[5] S. KONIECZNY (1996) Vers un modèle de la coopération. In Actes du 6ème colloque de l'Association pour la Recherche Cognitive. Décembre 1996.

[6] R. PINO PÉREZ, C. UZCÁTEGUI. On representation theorems for nonmonotonic inference relations. **Journal of Symbolic Logic**. A paraître. (Une version courte est parue comme Rapport de recherche LIFL, No IT-300, Université de Lille I, 1997).

[7] R. PINO PÉREZ, C. UZCÁTEGUI. Jumping to explanations vs jumping to conclusions. Rapport de recherche LIFL, No IT-301, Université de Lille I, 1997. Soumis à **AI Journal**.

[5] Anciennement Professeur à l'Université d'Illinois (Chicago).

[8] S. Konieczny and R. Pino Pérez. On the difference between arbitration and majority merging. Proceedings of WOLLIC'97 Workshop on Logic, Language, Information and Computation. Fortaleza, Brasil, August 1997.

[9] H. Bezzazi, S. Janot, S. Konieczny and R. Pino Pérez. Forward chaining and change operators. Proceedings of DYNAMICS'97 Workshop on (Trans)Actions and Change in Logic Programming and Deductive Databases held in conjonction with the International Logic Programming Symposium ILPS'97. Port Jefferson, New York, USA, October 1997. pp. 135-146.

[10] S. Konieczny and R. Pino Pérez. On the logic of Merging. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation And Reasoning, KR'98.* Trento, Italy. June 2-5, 1998, pp 488-498.

[11] R. Pino Pérez and C. Uzcátegui. Abduction vs. Deduction in Nonmonotonic Reasoning. In *Proceedings of the Seventh International Workshop on Nonmonotonic Reasoning, NR'98.* Trento, Italy. May 30-June 1, 1998, pp 42-54.

[12] H. Bezzazi. Stratified forward chaining. In *Proceedings of the Seventh International Workshop on Nonmonotonic Reasoning, NR'98.* Trento, Italy. May 30-June 1, 1998.

[13] H. Bezzazi, S. Janot, S. Konieczny and R. Pino Pérez. Analysing rational properties of change operators based on forward chaining. In *Transactions and Change in Logic Databases.* H. Decker, B. Freitag, M. Kifer and A. Voronkov, Eds. **Lecture Notes in Computer Science**, vol. 1472, pp 317-339. 1998.

[14] S. Konieczny Operators with memory for iteration. Rapport de recherche LIFL No IT-314, Université de Lille I, 1998.

[15] S. Konieczny and R. Pino Pérez. Merging with integrity constraints. Rapport de recherche LIFL No 99-01, Université de Lille I, 1999. Soumis à **Ecsqaru'99**.

[16] S. Benferhat, S. Konieczny, O. Papini et R. Pino Pérez. Révision itérée basée sur la primauté forte des observations. Soumis aux **Journées Modèles de Raisonnement du PRC-I3, 22-23 Mars 1999.**

Bibliographie

[1] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] Atocha Aliseda-Llera. *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence.* PhD thesis, Stanford University, Department of Computer Science, 1998.

[3] K. J. Arrow. *Social choice and individual values.* Wiley, New York, second edition, 1963.

[4] C. Baral, S. Kraus, and J. Minker. Combining multiple knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, 3(2):208–220, 1991.

[5] C. Baral, S. Kraus, J. Minker, and V.S. Subrahmanian. Combining knowledge bases consisting of first order theories. *Computational Intelligence*, 80:45–71, 1992.

[6] P. Besnard. *An Introduction to Default Logic.* Springer-Verlag, Berlin, 1989.

[7] H. Bezzazi, D. Makinson, and R. Pino Pérez. Beyond rational monotony: some strong non-horn rules for nonmonotonic inference relations. *Journal of Logic and Computation*, 7:605–631, 1997.

[8] H. Bezzazi, D. Makinson, and R. Pino Pérez. Beyond rational monotony: some strong non-horn rules for nonmonotonic inference relations. *Journal of Logic and Computation*, 7:605–631, 1997.

[9] H. Bezzazi and R. Pino Pérez. Rational transitivity and its models. In *Proc. of the Twenty-Sixth International Symposium on Multiple-Valued Logic*, pages 160–165, Santiago de Compostela, Spain, May, 1996. IEEE Computer Socieety Press.

[10] A. Borgida and T. Imielinski. Decision making in committees: A framework for dealing with inconsistency and non-monotonicity. In *Proceedings Workshop on Nonmonotonic Reasoning*, pages 21–32, 1984.

[11] C. Boutilier. Conditional logics of normality as modal systems. In *Proc. of the Eighth National Conference on Artificial Intelligence*, pages 594–598, 1990.

[12] C. Boutilier and V. Becher. Abduction as belief revision. *Artificial Intelligence*, 77:43–94, 1995.

[13] G. Brewka. *Nonmonotonic Reasoning:Logical Foundations of commonsense.* Cambridge Tracts in Theoretical Computer Science 12. Cambridge University Press, Cambridge, 1991.

[14] M. Dalal. Investigations into a theory of knowledge base revision: Preliminary report. In *Proc. of the Seventh National Conference on Artificial Intelligence (AAAI'88)*, pages 475–479, 1988.

[15] M. Dalal. Updates in propositional databases. Technical report, Rutgers University, 1988.

[16] C. Damasio, W. Nedjdl, and L. M. Pereira. Revise: An extended logic programming system for revising knowledge bases. In *Proc. of the 4th International Conference on Principles of Knowledge Representation and Reasoning*, pages 607–618. Morgan Kaufmann, 1994.

[17] Darwiche and J. Pearl. On the logic of iterated belief revision. In *Proc. of the 1994 Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 5–23, 1994.

[18] Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89:1–29, 1997. A first version of this work appeared in [17].

[19] J. Dix and D. Makinson. A note on the relationship between KLM amd MAK models for nonmonotonic inference operations. *Journal of Logic, Language and Information*, 1:131–140, 1992.

[20] D. Dubois, H. Fargier, and H. Prade. Refinements of the max-min approach to decision-making in fuzzy environment. *Fuzzy Sets and Systems*, 81:103–122, 1996.

[21] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In D. M. Gabbay, editor, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3. Oxford University Press, 1994.

[22] D. Dubois and H. Prade. Epistemic entrenchment and possibilistic logic. *Artificial Intelligence*, 50:223–239, 1991.

[23] T. Eiter and G. Gottlob. On the complexity of propositional knowledge base revision, updates, and counterfactuals. *Artificial Intelligence*, 57(2-3):227–270, 1992.

[24] T. Eiter and G. Gottlob. On the complexity of propositional knowledge base revision, updates and counterfactuals. In *Proc. of 11th Symposium on Principles of Database Systems*, pages 261–273. ACM Press, 1992.

[25] C. Even and R. Pino Pérez. λ_p-calculus and algebras with partial elements. In S. Mujica and A. Vi na, editors, *Proceedings of the XII International Conference of the Chilean Computer Science Society*, pages 71–84. Sociedad Chilena de Ciencias de la Computación, 1992.

[26] C. Even and R. Pino Pérez. Extension of scott-koymans theorem to partial framework. *The Bulletin of Symbolic Logic*, 1:489–490, 1995.

[27] R. Fagin, G. Kuper, J.D. Ullman, and M.Y. Vardi. Updating logical databases. In *Advances in Computing Research*, volume 3, pages 1–18, 1986.

[28] R. Fagin, J.D. Ullman, and M.Y. Vardi. On the semantics of updates in databases. In *Proc. Seventh ACM SIGACT/SIGMOD Symposium on Principles of Database Systems*, pages 352–365, 1983.

[29] P. A. Flach. Rationality postulates for induction. In Yoav Shoham, editor, *Proc. of the Sixth Conference of Theoretical Aspects of Rationality and Knowledge (TARK96)*, pages 267–281, The Netherlands, March 17-20, 1996.

[30] M. Freund. Injective models and disjunctive relations. *Journal of Logic and Computation*, 3:231–247, 1993.

[31] M. Freund and D. Lehmann. Belief revision and rational inference. Technical Report 94-16, Institute of Computer Science, The Hebrew University of Jerusalem, 1994.

[32] M. Freund and D. Lehmann. On negation rationality. *J. Logic and Computation*, 6:1–7, 1996.

[33] M. Freund, D. Lehmann, and P. Morris. Rationality, transitivity and contraposition. *Artificial Intelligence*, 52:191–203, 1991.

[34] A. Fuhrmann. Theory contraction through base contraction. *Journal of Philosophical Logic*, 20:175–203, 1991.

[35] D.M. Gabbay. Theoretical foundations for nonmonotonic reasoning in experts systems. In K. Apt, editor, *Logic and Models of Concurrent Systems.* Springer-Verlag, 1985.

[36] P. Gärdenfors. *Knowledge in Flux: modeling the dynamics of epistemic states.* MIT press, Cambridge, MA, 1988.

[37] P. Gärdenfors and D. Makinson. Nonmonotonic inferences based on expectations. *Artificial Intelligence*, 65:197–245, 1994.

[38] H. Geffner. *Default Reasoning: Causal and Conditional Theories.* PhD thesis, Computer Science Department, University of California at Los Angeles, 1989.

[39] G. Gentzen. *The Collected Papers of Gerhard Gentzen.* M.E. Szabo editor. Noth-Holland, Amsterdam, 1969.

[40] M. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30:35–79, 1986.

[41] S.O. Hansson. Reversing the levi identity. *Journal of Philosophical Logic*, 22:637–669, 1993.

[42] S.O. Hansson. Theory contraction and base contraction unified. *Journal of Symbolic Logic*, 58:602–625, 1993.

[43] W.L. Harper, R. Stalnaker, and G. Pearce. *Ifs: Conditionals, Belief, Decision, Chance and Time.* Reidel, Dordrecht, 1980.

[44] H. Katsuno and A.O. Mendelzon. Propositional knowledge base revision and minimal change. Technical Report KRR-TR-90-3, Computer Science Department, University of Toronto, 1990.

[45] H. Katsuno and A.O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991. First appeared in [44]KRR-TR-90-3.

[46] H. Katsuno and A.O. Mendelzon. On the difference between updating a knowledge database and revising it. In P. Gärdenfors, editor, *Belief Revision.* Cambridge university press, Cambridge, MA, 1992. Cambrigde tracts in theoretical computer science, # 29.

[47] J. S. Kelly. *Arrow impossibility theorems.* Series in economic theory and mathematical economics. Academic Press, New York, 1978.

[48] S. Kraus, D. Lehmann, and M. Magidor. Preferential models and cumulative logics. Technical Report 88-15, Institute of Computer Science, The Hebrew University of Jerusalem, 1988.

[49] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[50] S. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24:1–14, 1959.

[51] S. Kripke. Semantical analysis of modal logic I. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9:67–96, 1963.

[52] D. Lehmann. What does a conditional knowledge base entail? In R. Brachmann and H.L. Levesque, editors, *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, 1989.

[53] D. Lehmann. Belief revision, revised. In *Proceedings IJCAI'95*, pages 1–23, 1995.

[54] D. Lehmann. Draft on the logic of induction. Technical report, Institute of Computer Science, Hebrew University, 1995.

[55] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992.

[56] H.J. Levesque. A knowledge level account of abduction. In *Proceedings of the eleventh International Joint Conference on Artificial Intelligence*, pages 1061–1067, Detroit, 1989.

[57] D. Lewis. *Counterfactuals.* Blackwell, Oxford, 1973.

[58] P. Liberatore and M. Schaerf. Arbitration: A commutative operator for belief revision. In *Proceedings of the Second World Conference on the Fundamentals of Artificial Intelligence*, pages 217–228, 1995.

[59] P. Liberatore and M. Schaerf. Arbitration (or how to merge knowledge bases). *IEEE Transactions on Knowledge and Data Engineering*, 10(1), 1998. To appear.

[60] Paolo Liberatore. The complexity of iterated belief revision. In *Proceedings of the Sixth International Conference on Database Theory -ICDT'97*, pages 276–290, Delphi, Greece, 1997. Lecture Notes in Computer Science, Vol. 1186.

[61] J. Lin. Integration of weighted knowledge bases. *Artificial Intelligence*, 83(2):363–378, 1996.

[62] J. Lin and A. O. Mendelzon. Knowledge base merging by majority. Manuscript.

[63] J. Lin and A. O. Mendelzon. Merging databases under constraints. To appear in International Journal of Cooperative Information System.

[64] J. Lobo and C. Uzcátegui. Abductive change operators. *Fundamenta Informaticae*, 27:385–411, 1996.

[65] J. Lobo and C. Uzcátegui. Abductive consequence relations. *Artificial Intelligence*, 89:149–171, 1997.

[66] M. Cialdea Mayer and F. Pirri. Abduction is not deduction-in-reverse. *Journal of the IGPL*, 4:1–14, 1996.

[67] D. Makinson. General theory of cumulative inference. In M. Reinfrank et al., editors, *Non-Monotonic Reasoning*, volume 346 of *Lecture Notes in Artificial Intelligence*, pages 1–18. Springer-Verlag, 1989.

[68] D. Makinson. Five faces of minimality. *Studia Logica*, 52:339–379, 1993.

[69] D. Makinson. General patterns in nonmonotonic reasoning. In C. Hogger D. Gabbay and J. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume III, Nonmonotonic Reasoning and Uncertain Reasoning, pages 35–110. Clarendon Press, Oxford, 1994.

[70] D. Makinson and P. Gärdenfors. The relations between the logic of theory change and nonmonotonic logic. In *The Logic of Theory Change, Workshop, Konstanz, FRG, October 1989*, pages 185–205. Springer-Verlag, 1991. Lecture Notes in Artificial Intelligence 465.

[71] V. Marek and M.Truszczynski. Revision programming, database updates and integrity constraints. In *Proc. of 5th International Conference of Database Theory*, pages 368–382, Prague, Czech Republic, 1995. Lecture Notes in Computer Science, Vol. 893.

[72] V.W. Marek and M. Truszczyński. *Nonmonotonic Logic.* Springer–Verlag, 1993.

[73] J. McCarthy. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.

[74] D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13:41–72, 1980.

[75] E. Moggi. *The Partial Lambda Calculus.* PhD thesis, Universita de Pisa, Department of Computer Science, 1986.

[76] R. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1):75–94, 1985.

[77] R.C. Moore. *Logic and representation.* CSLI Publications, 1995.

[78] A. C. Nayak. Foundational belief change. *Journal of Philosophical Logic*, 23:495–533, 1992.

[79] B. Nebel. Principles of knowledge representation and reasoning: Proceedings of the 1st international conference. In R. Reiter R.J. Brachman, H.J. Levesque, editor, *Belief Revision*, pages 301–311, Toronto, Ont., 1989.

[80] B. Nebel. Syntax-based approches to belief revision. In P. Gärdenfors, editor, *Belief Revision*, pages 52–88. Cambridge university press, Cambridge, MA, 1992. Cambrigde tracts in theoretical computer science, # 29.

[81] D. Nute. Conditional logic. In F. Günther D. Gabbay, editor, *Handbook of of Philosophical Logic*, volume II, pages 387–439. Kluwer Academic Publishers, 1984.

[82] G. Paul. Approaches to abductive reasoning: an overview. *Artificial Intelligence Review*, 7:109–152, 1993.

[83] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, Los Altos, California, 1988.

[84] C.S. Peirce. *Collected Papers of Charles Sanders Peirce,* edited by Charles Hartshorne and Paul Weiss. Harvard University Press, 1960-1966.

[85] P. Peppas. Well behaved and multiple belief revision. In W.Wahlster, editor, *Proceedings of the 12th European Conference on Artificial Intelligence*, pages 90–94, Budapest, 1996.

[86] R. Pino Pérez. π_n-collection, indicatrices et ultrafiltres définissables. Thèse de Troisième Cycle en Logique Mathématique. Université de Paris 7, 1983.

[87] R. Pino Pérez. Non standard methods in set theory. Travail présenté au Logic Colloquium'84. Manchester, 1984.

[88] R. Pino Pérez. Le modèle de projections finitaires et l'égalité de types dans une classe de modèles du λ-calcul polymorphe. Mémoire de DEA en Informatique. Université de Paris 7, 1986.

[89] R. Pino Pérez. Le lambda calcul partiel. Technical Report 87-11, Laboratoire d'Informatique de l'Ecole Normale Supérieure, Paris, 1987.

[90] R. Pino Pérez. Decidability of the equational restriction theory in the partial λ-calculus. *Theoretical Computer Science*, 67:129–139, 1989.

[91] R. Pino Pérez. An extensional partial combinatory algebra based on λ-terms. In A.Tarlecki, editor, *Proceedings of the 16th International Symposium on Mathematical Foundations of Computer Science*, pages 387–396. Lecture Notes in Computer Science, vol. 520, 1991.

[92] R. Pino Pérez. Contribution à l'étude du lambda calcul partiel. Thèse de Doctorat en Informatique. Université de Paris 7, 1992.

[93] R. Pino Pérez and C. Even. An abstract property of confluence applied to the study of the lazy partial lambda calculus. In A. Nerode and Yu. Matiyasevich, editors, *Proceedings of the 3th International Symposium on Logical Foundations in Computer Science*, pages 278–290. Lecture Notes in Computer Science, vol. 813, 1994.

[94] R. Pino Pérez and J.P. Ressayre. Definable ultrafilters and elementary end extensions. In C. di Prisco, editor, *Methods in Mathematical Logic*, pages 341–350. Lecture Notes in Mathematics, vol. 1130, 1985.

[95] R. Pino Pérez and C. Uzcátegui. Jumping to explanations vs. jumping to conclusions. Rapport de recherche LIFL No IT-301. Université de Lille I. 1997.

[96] R. Pino Pérez and C. Uzcátegui. On representation theorems for non-monotonic inference relations. *Journal of Symbolic Logic.* To appear.

[97] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1 and 2):81–132, 1980.

[98] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[99] P. Z. Revesz. On the semantics of theory change: arbitration between old and new information. In *Proceedings of the* 12^{th} *ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Databases*, pages 71–92, 1993.

[100] P. Z. Revesz. On the semantics of arbitration. *International Journal of Algebra and Computation*, 7(2):133–160, 1997.

[101] K. Satoh. Nonmonotonic reasoning by minimal belief revision. In *Proceedings International Conference on Fifth Generation Computer Systems*, pages 455–462, Tokio, 1988.

[102] K. Satoh. A probabilistic interpretation for lazy non monotonic reasoning. In *Proceedings of AAAI-90*, pages 659–664, Boston, August 1990.

[103] L. J. Savage. *The foundations of statistics.* Dover Publications, New York, 1971. Second revised edition.

[104] K. Schlechta. Some completeness results for stoppered and ranked classical preferential models. *J. Logic and Computation*, 6:599–622, 1996.

[105] Y. Shoham. A semantical approche to nonmonotonic logics. In *Logic in Computer Science*, pages 275–279, 1987.

[106] Y. Shoham. *Reasoning about change.* Mit Press, Cambridge, MA, 1988.

[107] R. Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in Logical Theory, American Philosophical Quarterly Monograph Series, No. 2.* Blackwell, 1968. Reprinted in [43].

[108] V. S. Subrahmanian. Amalgamating knowledge bases. *ACM Transactions on Database systems.*, 19(2):291–331, 1994.

[109] A. Tarski. *Logic, Semantics, Metamathematics, Papers from 1923-1938.* Clarendon Press, Oxford, 1956.

[110] N. Tennant. Changing the theory of theory change: Towards a computational approach. *British Journal for Philosophy of Science*, 45:865–897, 1994.

[111] D.S. Touretzky. *The Mathematics of Inheritance Systems.* Pitman Reasearch Notes in Artificial Intelligence, London, 1986.

[112] A. Weber. Updating propositional formulas. In *Proceedings First Conference on Database Systems*, pages 487–500, 1986.

[113] M. Winslett. Reasoning about action using a possible model approach. In *Proc. of the Seventh National Conference on Artificial Intelligence*, pages 89–93, 1988.

[114] W. Zadrozny. On the rules of abduction. *Annals of Mathematics and Artificial Intelligence*, 9:387–419, 1993.