



HAL
open science

Les opérations de changement basées sur le test de Ramsey

Gabriella Crocco, Andreas Herzig

► **To cite this version:**

Gabriella Crocco, Andreas Herzig. Les opérations de changement basées sur le test de Ramsey. Livet, Pierre. Révision des croyances (Traité des sciences cognitives), Chapitre 2, Hermes Science Publications, pp.21-41, 2002, 978-2746204690. hal-03534103

HAL Id: hal-03534103

<https://hal.science/hal-03534103>

Submitted on 19 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapitre 2

Les opérations de changement basées sur le test de Ramsey

Introduction

Ce chapitre a une double ambition. D'une part, nous souhaiterions donner un exemple concret d'une fonction de mise à jour d'une base de données à partir d'hypothèses simples, de manière constructive et dans un langage formel aux caractéristiques déductives attrayantes du point de vue de la complexité du calcul. D'autre part, nous voudrions discuter, à partir de cet exemple, certains des présupposés, et peut-être même des préjugés, les plus répandus dans la littérature de la théorie de la révision. Ces présupposés (ces préjugés) trouvent leur origine dans l'hétérogénéité des traditions de recherche qui en constituent l'arrière plan. L'absence d'une véritable analyse intensionnelle fine est, à notre avis, l'aspect fédérateur de ces traditions de recherche hétérogènes qui ont été de fait sollicitées dans la constitution de la théorie moderne de la révision. Nous donnons à la section 2.1 quelques éléments permettant de justifier une telle assertion, en esquissant les buts et les méthodes de ces trois traditions quant au problème de la croyance.

Notre exemple – la fonction de mise à jour que nous définissons en section 2.6 – a été ici introduit dans le but de montrer combien l'absence d'une telle analyse intensionnelle fine est regrettable. Cette fonction de mise à jour est définie grâce à la notion intensionnelle de dépendance. Elle n'est qu'une extension constructive de la fonction de sélection propre aux logiques des conditionnels contrefactuels de Chellas.

Les hypothèses de travail qui nous ont conduit à assumer une telle sémantique sont illustrées dans les sections 2.2, 2.3 et 2.4. Ces hypothèses ne sont certainement pas les seules à avoir droit d'existence, mais elles nous semblaient les plus simples et les plus naturelles pour mettre en évidence certains des problèmes non résolus de la théorie de la révision.

2.1. Traditions et problèmes d'hétérogénéité

Trois sources convergent dans ce que l'on a pris l'habitude d'appeler la théorie de la révision.

La *première source* peut être reliée aux logiques modales, logiques de représentation des contextes intensionnels : logiques épistémiques et doxastiques (Hintikka, 1962 ; von Wright, 1951), logiques de la causalité, des conditionnels contrefactuels (Lewis, 1973) qui appartiennent au vaste domaine de ce que l'on appelle usuellement la logique philosophique. Dans le sillon de la conception carnapienne de la logique, le but commun de ces travaux était de fournir une explication de ces notions vagues telles que croyance ou causalité, dans le cadre d'une théorie formelle. Une explication au sens de Carnap (1947 ; 1950) n'est pas une fondation ou une réduction « constructive ». L'explication consiste en la transformation de ces notions obscures (vérité, croyance, causalité), les *explicanda*, appartenant à un stade non formel de la science, par des notions exactes (les *explicata*) définies dans une syntaxe dont les ressources ont été au préalable précisément délimitées. L'explication n'est qu'une proposition qui doit permettre de remplacer, dans le discours scientifique, le concept vague par le concept exact qui, quant à lui, doit être le plus « similaire » au concept vague.

Conforme à cela, le but des recherches en logique épistémique ou doxastique n'était pas celui de décrire la formation ou l'évolution d'un ensemble de croyances d'un ou plusieurs agents, mais plutôt de représenter *statiquement* l'état cognitif d'un agent dans le cadre d'une sémantique relationnelle, afin d'expliquer les différences entre les concepts de *connaissance* et de *croyance*. L'intelligence artificielle a redécouvert ces formalismes à partir des années 1980, en négligeant d'ailleurs pratiquement l'opposition philosophique entre croyance et connaissance qui en avait été la source.

Ensuite, la théorie des contrefactuels de David Lewis se proposait à son tour de fournir une explication de la causalité en termes d'énoncés conditionnels contrefactuels, et d'expliquer ces derniers en termes de la relation de similarité comparative. Dans la littérature sur la révision et plus généralement sur les raisonnements non monotones, on a repris les résultats et les méthodes de Lewis sans trop se soucier des limites que ce dernier assignait à son entreprise. Lewis jugeait en effet impossible de

donner une description constructive de la similarité comparative, notion « trop dépendante du contexte » pour être décrite dans le langage logique¹.

Les traits communs des recherches en logique épistémique et contrefactuelle quant au traitement de l'intensionnalité sont bien connus. On distingue l'extension d'une expression désignative de son intension. Dans le cas des énoncés, l'*extension* d'un énoncé *A* est sa valeur de vérité dans le monde actuel. L'*intension* de *A* (notée d'habitude [*A*]), en d'autres termes la proposition que *A* exprime, n'est rien d'autre que l'ensemble des mondes dans lequel l'énoncé *A* est vrai. Connaître le contenu de *A*, connaître son sens, revient donc à savoir dans quelles circonstances *A* est vrai et dans quelles circonstances *A* est faux, l'ensemble des mondes possibles constituant l'ensemble des circonstances possibles. De ce critère de définition de la notion d'intension pour les énoncés découle un critère d'identité pour les propositions. Deux propositions sont alors identiques si et seulement si elles sont logiquement équivalentes.

Les entités intensionnelles, ces notions obscures telles que le sens d'un énoncé, le concept d'une classe ou d'un individu, sont alors extensionnalisées, réduites à des ensembles d'extensions possibles. Il a été observé qu'une telle extensionnalisation de l'intensionnalité était insatisfaisante pour au moins deux raisons :

- 1) elle donne une explication du sens qui autorise des définitions circulaires (voir Bealer, 1982) ;
- 2) elle est tout à fait inadaptée pour rendre compte de la notion de croyance.

En ce qui concerne 2), la question avait déjà été évoquée à l'époque où Carnap posait les fondements de l'analyse sémantique de la logique modale : les énoncés « Jean croit que *D* » et « Jean croit que *D'* » peuvent ne pas avoir la même valeur de vérité bien que *D* et *D'* soient logiquement équivalents (Carnap, 1947).

Malgré la clarté de cette explication sémantique, l'équivalence logique semble être un critère trop faible d'identification des intensions ou, si l'on veut, des intensions lorsque celles-ci sont évoquées pour rendre compte du sens, du contenu d'un énoncé. Elle est probablement une condition nécessaire mais nullement suffisante pour l'identité du contenu de deux énoncés. Il y a, à l'évidence, des informations structurelles que l'extensionnalisation de l'intension ne prend pas en considération.

1. Il prédit explicitement « *[the failure of] any humanly possible attempt at a precise definition of comparative similarity of worlds. Not only would we go wrong by giving a precise analysis of an imprecise concept; our precise concept would not fall within—or even near—the permissible range of variation of the ordinary concept.* » (Lewis, 1973, 4.2, p. 95). En français : « [la faillite de] toute tentative humaine de définir avec précision la notion de similarité comparative. Non seulement nous nous tromperions en essayant de donner une analyse précise d'un concept imprécis, mais notre concept précis ne tomberait pas dans le domaine de variation du concept ordinaire et il ne s'en approcherait même pas » (notre traduction).

En *deuxième lieu*, la théorie de la révision s'est développée à partir de l'étude des paradigmes scientifiques et de leurs changements (Kuhn, Lakatos et Quine, pour ne citer que les plus connus).

Il s'agit ici non pas d'expliquer, au sens technique que l'on vient d'évoquer, la notion de croyance d'un individu, mais de décrire les changements de ces objets complexes, les théories scientifiques, auxquels on prête un caractère absolu et extensionnel².

Les critiques quiniennes de toute notion intensionnelle sont bien connues, et toutefois Quine est l'auteur de référence dans le débat sur l'enracinement épistémique des croyances. Une théorie (scientifique) n'est qu'un ensemble infini de croyances acceptées dans le corpus de notre science. Tout énoncé asserté constitue une croyance pour l'individu qui l'asserte. De plus, certaines croyances font l'objet de l'assentiment collectif de la part de la communauté des scientifiques. L'ensemble des croyances d'une collectivité est ainsi constitué en une sorte de réseau, où toute croyance dépend de toute autre, où toute croyance est également réfutable en droit, mais où certaines croyances occupent une place privilégiée à cause de certaines de leurs caractéristiques. La notion d'enracinement épistémique utilisée en IA pour construire des opérations de révision dérive directement de ce principe quinienn de réseau de connaissances. Une théorie scientifique est mise à l'épreuve de l'observation : si les données de l'observation sont en conflit avec les prédictions de la théorie, alors ces données réfutent la théorie. Pour retrouver l'accord entre la théorie et l'observation, certaines parties de la théorie doivent être rejetées et de nouvelles hypothèses doivent être formulées. Le caractère holistique de la théorie et le caractère diffus de la signification empirique empêchent de limiter le choix à une seule possibilité. Certains critères généraux gouvernent ces changements de théories. Quine en cite six : conservativité, modestie, simplicité, généralité, réfutabilité et précision (Quine et Ullian, 1978).

A partir de 1985, Gärdenfors *et al.* ont introduit en l'intelligence artificielle la problématique des changements rationnels des théories, restreignant leur intérêt à ce que Kuhn considérait comme des ajustements dans le cadre de la « science normale ». Une telle distinction n'appartient pas, à proprement parler, à la description quinienne. Toutefois, Gärdenfors, Alchourrón et Makinson ont proposé un cadre formel d'analyse de ces changements minimaux par les biais des notions de révision, expansion et contraction. Ce que, depuis lors, on appelle le paradigme AGM a donné lieu à une littérature extrêmement abondante. Il faut tout de même souligner le caractère philosophiquement hybride de leur démarche. Ils utilisent les moyens de l'explication carnapienne (recherche des contraintes générales, exprimées dans un langage formel

2. La description s'oppose alors à l'explication car elle n'a pas lieu dans un langage formel aux ressources préalablement stipulées : on ne cherche pas à énumérer rigoureusement les éléments du langage de l'explication.

d'explication) pour les adapter au cadre quinién de description, extensionnel et absolu, de surcroît réduit au seul critère de conservativité de la connaissance.

La *troisième source* de la théorie de la révision concerne les travaux sur la mise à jour des bases de données dans le domaine de la gestion informatique de bases d'informations. On ne parle pas ici de croyances mais plutôt d'informations stockées dans une base. La constructivité des opérations de mise à jour est le critère essentiel des formalismes de ce domaine. S'il y a un traitement intensionnel de l'information, celui-ci se réduit à la prise en considération de contraintes exclusivement syntaxiques. Le sens d'une expression se réduit alors à sa forme syntaxique, son mode de notation dans un langage. On parvient même à distinguer, par rapport à leur comportement face à la mise à jour, des bases telles que $\{p, q\}$, $\{p \wedge q\}$, $\{q \wedge p\}$, $\{p, p \rightarrow q\}$, etc.

Il nous semble difficile de justifier une telle distinction entre $\{p, q\}$, $\{p \wedge q\}$ et $\{q \wedge p\}$. En revanche, il a souvent été argumenté en IA (Nebel, 1991, 1992) que les bases $\{p, q\}$ et $\{p, p \rightarrow q\}$ doivent être révisées différemment : dans $\{p, q\}$ il n'y a pas de lien entre p et q , tandis que dans $\{p, p \rightarrow q\}$ l'implication $p \rightarrow q$ nous dit qu'il y a un tel lien (il est laissé ouvert s'il est de type causal, explicatif ou autre). En conséquence, la révision de $\{p, q\}$ par $\neg q$ devrait résulter en $\{p, \neg q\}$, et celle de $\{p, p \rightarrow q\}$ en $\{p \rightarrow q, \neg q\}$. Cela a motivé certains auteurs pour étudier la révision dans des logiques de la pertinence où $p \wedge q$ et $\{p, p \rightarrow q\}$ ne sont pas équivalents (Lakemeyer et Lang, 1996 ; Restall et Slaney, 1995). Cependant, ces recherches se sont heurtées à la complexité de ces logiques substructurelles (ayant par exemple une notion de consistance triviale).

On pourra tout de même remarquer une certaine convergence entre cette tradition et la première que l'on a évoquée. En effet, après des solutions purement algorithmiques, vers la fin des années 1980, des chercheurs en IA se sont intéressés à des modèles logiques pour la mise à jour, à partir des travaux de Winslett (1988, 1990). Des modèles ont été proposés qui sont basés sur une fonction particulière de sélection à la Chellas. De telles fonctions de sélection sont construites (expliquées) à partir d'une notion de distance minimale particulière (voir section 2.6).

Etant donné les traditions hétérogènes sollicitées par la théorie de la révision, on pourrait se demander s'il y a eu, en vingt ans de recherches, une harmonisation de leurs objets, de leurs buts et de leurs méthodes. La réponse semble devoir être négative. Les distinctions courantes entre révision et mise à jour, les polémiques entre fondationnalistes et cohérentistes, la constante remise en discussion des postulats AGM, censés constituer les contraintes minimales auxquelles toute opération de révision devrait se soumettre, montrent bien l'instabilité des fondements de la théorie. Une telle instabilité nous semble justifier une remise en question de certains des présupposés d'AGM³.

3. Les travaux récents de Rott (2000) vont dans le même sens (voir également son chapitre).

Toutefois, au lieu de discuter explicitement chacun de ces postulats, nous proposons de renverser le parcours d'analyse. Partons des *desiderata* qui nous semblent souhaitables pour une théorie des changements des bases de croyances. Toute théorie de tels changements devrait :

- satisfaire au moins partiellement à un critère de constructivité du changement ;
- intégrer une analyse intensionnelle fine des énoncés, nécessaire lorsque nous prenons en considération une notion aussi délicate que celle de croyance.

Comment alors rendre le choix des objets et des opérations mises en jeu par les changements des bases de croyances compatible avec ces *desiderata* ? Les trois sections suivantes tentent de répondre à cette question, quant à la nature de la base des croyances et à la nature et aux propriétés de l'opération de changement.

2.2. Base de données *versus* ensemble de croyances

La première question, directement liée à l'hétérogénéité des traditions que nous venons d'évoquer, porte sur la nature de l'objet de la révision. Devons-nous parler de théories et donc d'un ensemble infini de formules clos sous l'opération de conséquence logique ? Devons-nous au contraire nous restreindre à des bases finies de croyances et comment devons-nous représenter de telles bases ?

Dans ce chapitre, nous supposons que les ensembles de croyances sont finis et nous identifierons de telles bases de croyances avec la conjonction de leurs éléments.

Il y a au moins deux arguments pour procéder ainsi. Le premier est pratique. Si nous voulons mettre en application la croyance sur un ordinateur, nous sommes obligés de le faire avec des représentations finies. Le deuxième argument est d'ordre philosophique. Nous considérons un ensemble de croyances comme une sorte de carte cognitive de la réalité, où une croyance peut être ajoutée ou enlevée tout comme une route peut être inscrite ou effacée d'une vraie carte (Sahlin, 1988). Une telle carte n'est pas une description complète d'une région, mais plutôt le résultat d'un équilibre raisonné entre la nécessité d'être informatif et la nécessité d'être lisible. En ce sens, il est naturel de penser un ensemble de croyances comme représenté par un nombre fini d'énoncés qui correspondent approximativement aux croyances explicites dans un état mental⁴. Les états de croyance, vus comme états mentaux, sont habituellement incomplets. Ainsi, ils ne décrivent pas complètement toutes les conséquences

4. Par conséquent, des ensembles de croyances logiquement équivalents pourraient être exprimés de différentes manières. Notons que ceci ne nous force pas *a priori* à prendre position dans le débat entre fondationalistes et cohérentistes : tout en travaillant avec des bases de croyance, nous pouvons toutefois stipuler que la mise à jour de bases logiquement équivalentes avec la même entrée mène à des résultats équivalents.

d'un ensemble de croyances, mais plutôt quelques-unes d'entre elles qui peuvent être obtenues en complétant la base actuelle de croyances.

La sémantique de la logique modale permet de donner une interprétation à une telle structure. Soit L le langage dans lequel la base B est construite. Soit P_1, \dots, P_k, \dots les atomes de ce langage. Soit A_1, \dots, A_m les formules de L contenues dans B à un moment donné et telles que seul un sous-ensemble des atomes du langage (soit P_1, \dots, P_k) est utilisé dans l'écriture de B ; nous pouvons associer à B un ensemble de mondes W sous-ensemble de l'ensemble des interprétations de L , tel qu'aucun monde de W satisfait P_{k+1}, \dots . Chaque description complète possible résultant de la base actuelle de croyances s'appellera un monde possible dans la suite. L'ensemble des mondes permet de donner le sens des informations contenues dans la base.

Un tel choix implique :

- d'abandonner la notion absolue⁵ de théorie ;
- d'assumer le cadre intensionnel des logiques modales, pour expliquer la notion de croyance et ensuite raffiner la notion d'intensionnalité qui est présentée à travers les mondes possibles par des moyens plus constructifs.

2.3. Le test de Ramsey

La deuxième question laissée ouverte concernant l'opération de changement est la suivante : comment peut-on parler d'un opérateur de révision ou de mise à jour, et quelles propriétés devrait-il avoir ? Dans un renvoi de bas de page dans Ramsey (1931, p. 247-248), qui est devenu un classique dans des approches de l'IA à la révision, Ramsey dit : « *If two people are arguing about "If p will q ?" and are both in doubt as to p , they are adding p hypothetically to their stock of knowledge and arguing on that basis about q .* »⁶.

Ce test de Ramsey (RT dorénavant) a été utilisé comme une manière de donner une signification aux conditionnels en les expliquant par une opération de changement de croyance. Nous appelons de tels changements des mises à jour. Etant donné un ensemble de croyances B et une nouvelle croyance A (que nous appelons l'entrée), nous parlons de la mise à jour de B par A . Nous notons C la conséquence d'une mise à jour.

5. Absolue au sens « indépendante du support de la théorie ».

6. En français : « Si deux personnes discutent sur l'énoncé « si p alors q » et que toutes deux doutent de p , elles sont, en réalité, en train d'ajouter hypothétiquement p à leurs connaissances et de discuter, sur cette base, de q . » (notre traduction).

Hansson (1995) dit que le RT était à la base des trois tentatives principales pour donner la signification aux conditionnels, à savoir la théorie métalinguistique de Goodman (basée sur la dérivabilité), l'analyse basée sur une fonction de sélection de mondes de Stalnaker et Lewis, et l'analyse basée sur la révision de Gärdenfors et ses collègues. Sans aucun doute, le RT possède beaucoup d'attraits intuitifs. Mais si nous l'acceptons, alors que signifie mettre à jour un ensemble de croyances, ou – en utilisant notre métaphore – quels sont les effets du fait d'ajouter une nouvelle route sur une carte?

Nous pouvons poser la question d'une autre manière: le RT explique-t-il une notion complexe (à savoir le conditionnel) par une notion plus simple (à savoir le changement de croyance)? La réponse semble être négative: apparemment, les deux notions sont pareillement vagues (dans le sens de Lewis, 1973, 4.2, p. 95) et pareillement complexes. Nous avons ainsi deux actions exprimées par des moyens linguistiques: ajouter une croyance à un ensemble de croyances et en vérifier les conséquences, et affirmer un conditionnel sur le fond d'un ensemble de croyances. Ces actions doivent avoir, en quelque sorte, les mêmes droits.

Le choix du RT comme règle d'inférence, dictant le sens d'une opération de mise à jour *via* un conditionnel, implique des contraintes quant à la nature de l'information déduite. Un conditionnel ayant dans la conclusion un opérateur modal (de connaissance ou de croyance), ou plus généralement un opérateur parlant du contenu de la base, doit être exclu sous peine de contradiction. Un conditionnel tel que « si ma femme me trompe, je ne le saurai jamais » ne peut évidemment pas être interprété *via* une opération de mise à jour. On ne peut pas dire dans ce cas que si on rajoute à la base de croyances d'un agent le fait que sa femme le trompe, on peut alors déduire que cette information ne sera pas dans sa base. Une telle contrainte ne semble pas infirmer le RT; elle semble simplement suggérer que des énoncés modaux se situent sur un niveau supérieur et ne peuvent être traités de la même manière que les autres éléments de la base. Cependant, une telle restriction ne peut être utilisée comme argument contre RT, car le même constat a été fait pour la révision AGM, bien que celle-ci ne respecte pas RT (Fuhrmann, 1989).

Le choix technique de la règle RT implique aussi le rejet d'une distinction devenue désormais classique en IA entre révision et mise à jour. Katsuno et Mendelzon (1992) affirment que quand une base de croyances est avisée d'un changement se produisant dans le « monde externe » ou la « réalité », alors les mondes possibles associés à l'état de croyance doivent être modifiés en conséquence :

« Since we are confined to our set of possibilities, we must make the change come true in all of our candidate worlds. Semantically, we change each of the possible worlds “as little as possible” in order to make the new state of affairs hold. Our new syntactic description of the worlds of interest should now

correctly reflect the outcome of this set of changes. The function that maps the old description to the new is called an update. »⁷ (Grahne, 1991).

Au contraire, si rien ne change dans le monde externe, et qu'une nouvelle croyance est censée seulement étendre ou corriger notre base actuelle de croyances, alors nous sommes dans une situation complètement différente, et l'on préconise l'utilisation de la révision (comprise comme révision AGM).

Formellement, des opérations de mise à jour comme celles définies dans Katsuno et Mendelzon (1992) (« les mises à jour KM ») sont non seulement différentes des opérations de révision, mais aussi incompatibles avec elles : il n'existe aucune opération de changement satisfaisant à la fois les postulats AGM et les postulats KM de Katsuno et Mendelzon. Pour que ces deux définitions soient opérationnelles il est alors nécessaire que l'on puisse décider, pour une situation et une nouvelle information donnée, quel type d'opération doit être appliqué. Mais comment pouvons-nous décider si une nouvelle information correspond à un changement dans le monde réel ou non ? Le choix semble évident seulement pour des contextes particuliers comme les enquêtes policières ou le diagnostic de pannes, où il est justifié de supposer que le monde réel n'évolue pas : quand un meurtre ou un défaut dans un circuit a été détecté, alors le détective ou l'ingénieur révisé ses croyances sur un état fixe des choses (les circonstances du crime ou de la panne) à la lumière des nouvelles informations recueillies au fil du temps sur celui-ci.

Cela dit, nous pensons qu'il y a beaucoup de situations où nous n'avons pas suffisamment d'informations pour faire le bon choix entre mise à jour KM et révision AGM, entre un changement dans le monde réel et un changement seulement dans nos croyances. Un agent est partie du monde, et ses croyances sont donc aussi des faits du monde. Toute autre différenciation ne peut surgir que dans des contextes artificiellement simplifiés.

2.4. Langage objet *versus* métalangage

Ayant imposé l'équivalence entre conditionnel et mise à jour, nous avons deux options pour chacune de ces opérations : nous pourrions opter pour une opération dans le métalangage, mais nous pourrions tout aussi bien opter pour un connecteur non

7. En français : « Etant donné que nous sommes restreints à notre ensemble de possibilités, nous devrions intégrer le changement dans tous nos mondes possibles. Du point de vue sémantique, nous changeons chaque monde « aussi peu que possible » de manière à réaliser ce nouvel état de choses [exprimé par la nouvelle information]. Notre nouvelle description syntaxique des mondes pertinents devrait alors refléter correctement le résultat de cet *ensemble* de changements. On appellera *mise à jour*, la fonction qui permet le passage de l'ancienne description à la nouvelle » (notre traduction).

classique dans le langage objet. Une opération située sur le niveau métalinguistique ne peut pas être analysée en termes d'axiomes et règles d'inférence (qui sont *a priori* constructives), mais en termes de postulats, qui sont non constructifs et en quelque sorte plus informels. A l'inverse, un opérateur de mise à jour dans le langage objet nous permet de l'identifier avec un connecteur logique, auquel une signification pragmatique peut être attribuée, c'est-à-dire une signification en termes de conséquences de l'action de mise à jour sur l'état de croyance.

Jusqu'ici, on a considéré, la plupart du temps, que la mise à jour est dans le métalangage et le conditionnel dans le langage objet, et que le test de Ramsey stipule un lien formel entre les deux. Sous cette vue, il y a un résultat formel que l'on appelle parfois le théorème d'impossibilité de Gärdenfors (1978, 1988) qui dit *grosso modo* qu'il n'y a pas d'opérateur conditionnel du langage objet qui correspond aux opérations métalinguistiques de révision AGM par l'intermédiaire du test de Ramsey⁸. Ce résultat négatif a conduit à une grande discussion dans la littérature sur la révision, et de nombreuses tentatives ont été entreprises pour le surmonter. Nous ne citons ici que les approches de Levi (1988) et Lindström et Rabinowicz (1995) qui évitent le théorème d'impossibilité en choisissant de placer le conditionnel hors du langage objet.

A l'inverse, et puisque nous travaillons avec des bases finies de croyances (que nous identifions avec des énoncés), nous pouvons placer des mises à jour et des conditionnels dans le langage objet. Grahne (1991) a montré que, sous cette hypothèse, le théorème d'impossibilité de Gärdenfors ne s'applique pas si nous abandonnons la révision à la AGM en faveur de la mise à jour à la Katsuno et Mendelzon (1992). Plus précisément, il a montré qu'il existe une logique des conditionnels (le système VCU de Lewis) et une opération de mise à jour correspondante *via* RT.

2.5. L'analyse formelle

Le RT explicite que conditionnels et mise à jour impliquent trois entités, à savoir une base de croyances initiale B (que nous supposons finie et que nous identifions avec la conjonction de ses éléments), une entrée A et une base de croyances résultante C , ou plutôt une conséquence de celle-ci. Il s'agit donc d'une relation ternaire $R(B, A, C)$ entre croyances, et nous pouvons dire que conditionnel et mise à jour sont deux manières différentes d'interpréter cette relation algébriquement. En fait, nous avons au moins deux possibilités pour exprimer R au moyen d'opérateurs binaires :

– nous pouvons grouper B et A et construire ainsi explicitement la nouvelle base de croyances résultant du changement. Dans ce cas, ce que nous choisissons est un

8. Ce résultat, quoique bien moins difficile à établir, peut être rapproché du théorème de Gödel, qui nous dit que la notion métalinguistique de consistance ne peut être captée dans le langage objet par un prédicat de consistance.

opérateur binaire de *mise à jour* $\leftarrow\Diamond$. $B\leftarrow\Diamond A$ peut être lu « B a été mis à jour par A ». Vérifier que $R(B, A, C)$ revient alors à montrer que $B\leftarrow\Diamond A \vdash C$;

– nous pouvons grouper A et C . Dans ce cas-ci, nous choisissons un *opérateur conditionnel* $\Box\rightarrow$ qui agit sur une entrée A et une conséquence C du changement. $A\Box\rightarrow C$ peut être lu « si une mise à jour par A est effectuée, alors C en résulte ». C’est ce que l’on pourrait appeler une mise à jour hypothétique. Vérifier que $R(B, A, C)$ revient alors à montrer que $B \vdash A\Box\rightarrow C$ ⁹.

Nous pouvons donner une interprétation temporelle à ces algébrisations :

– utiliser le conditionnel revient à se placer à l’instant temporel *avant* le changement (celui associé à B). Le conséquent C du conditionnel $A\Box\rightarrow C$ parle alors du futur ;

– utiliser une mise à jour revient à se placer à l’instant temporel *après* le changement (celui associé à C). La partie B de la mise à jour $B\leftarrow\Diamond A$ parle alors du passé.

Nous sommes donc proche des logiques temporelles avec futur et passé, ce qui se reflétera dans la sémantique (v.i.).

Il est étonnant que cette dualité entre les deux connecteurs n’ait pas été explorée. Ceci pourrait être dû à un préjugé général contre la mise des opérations métalinguistiques dans le langage objet par l’intermédiaire des connecteurs modaux ou à un préjugé particulier venant avec le résultat d’impossibilité de Gärdenfors.

Après avoir introduit ces deux opérateurs non classiques, nous pouvons maintenant exprimer le RT plus formellement :

$$B \vdash A\Box\rightarrow C \quad \text{si et seulement si} \quad B\leftarrow\Diamond A \vdash C$$

Si \vdash est monotonique, alors il découle immédiatement de RT que l’opérateur de mise à jour est monotonique dans l’argument gauche et que l’opérateur conditionnel est antitonique dans l’argument droit¹⁰.

9. Notons que nous pouvons aussi bien grouper B et C . Un tel opérateur correspondrait à la génération de plans comme étudié en IA, mais n’a pas été étudié jusqu’ici dans la littérature sur des mises à jour.

10. Supposons que $B \vdash B'$ et $B\leftarrow\Diamond A \vdash C$. Par le RT, $B \vdash A\Box\rightarrow C$. Par monotonie $B' \vdash A\Box\rightarrow C$, et encore une fois par RT nous obtenons $B'\leftarrow\Diamond A \vdash C$. La démonstration pour le conditionnel est semblable. Notons également que les deux opérateurs peuvent être monotoniques dans l’entrée seulement sous peine de trivialité.

Nous allons supposer dans la suite que la relation d'inférence est monotonique, et par conséquent nous acceptons les propriétés de monotonie qui en découlent¹¹. Nous supposons également que \vdash satisfait le théorème de déduction par rapport à \rightarrow , c'est-à-dire $B \vdash C$ si et seulement si $\vdash B \rightarrow C$.

2.5.1. Logiques normales de mise à jour

Notre expression formelle du RT peut être interprétée déductivement et nous pouvons en extraire deux schémas de règle d'inférence :

$$\begin{array}{l} \text{(R.Conv1)} \quad \frac{B \leftarrow \diamond A \rightarrow C}{B \rightarrow A \square \rightarrow C} \\ \text{(R.Conv2)} \quad \frac{B \rightarrow A \square \rightarrow C}{B \leftarrow \diamond A \rightarrow C} \end{array}$$

Nous les appelons les règles de conversion de Ramsey, car elles permettent de « convertir » un opérateur parlant du passé en un opérateur parlant du futur, tout comme le font les règles de conversion en logiques temporelles avec futur et passé¹².

Quelle est la logique minimale possédant ces deux règles d'inférence ? C'est-à-dire quelles sont les propriétés des opérateurs $\square \rightarrow$ et $\leftarrow \diamond$ dérivables exclusivement des règles de conversion ?

Supposez que \rightarrow est implication matérielle. Comme l'a montré Wansing (1994) pour les logiques temporelles, (R.Conv1) et (R.Conv2) permettent alors de dériver tous les principes de la logique conditionnelle semi-normale de base \mathcal{K} de Chellas. Nous baptisons cette logique semi-normale de base $\mathcal{U}\mathcal{C}\mathcal{k}$.

En détail, avec (R.Conv1) et (R.Conv2) ainsi que les principes de la logique classique propositionnelle, nous pouvons dériver les axiomes et règles d'inférence suivants :

$$\begin{array}{l} \text{(RC.EC)} \quad \frac{C_1 \equiv C_2}{A \square \rightarrow C_1 . \equiv . A \square \rightarrow C_2} \\ \text{(RC.M)} \quad \frac{C_1 \rightarrow C_2}{A \square \rightarrow C_1 . \rightarrow . A \square \rightarrow C_2} \end{array}$$

11. Notons néanmoins qu'en plaçant la mise à jour et l'opérateur conditionnel dans le langage objet, la possibilité d'une relation d'inférence non monotone au niveau métalinguistique reste ouverte.

12. $B \leftarrow \diamond A$ correspond à $\overline{\diamond} B$, où $\overline{\diamond}$ est l'opérateur « possible dans le passé », et $A \square \rightarrow C$ correspond à $\overline{\square} C$, où $\overline{\square}$ est l'opérateur « toujours dans le futur ». Les règles (R.Conv1) et (R.Conv2) deviennent alors les axiomes de conversion standard de la logique temporelle avec futur et passé : $\frac{\overline{\diamond} B \rightarrow C}{B \rightarrow \overline{\square} C} ; \frac{B \rightarrow \overline{\square} C}{\overline{\diamond} B \rightarrow C}$.

$$\begin{array}{ll}
\text{(C.C)} & (A \Box \rightarrow C_1) \wedge (A \Box \rightarrow C_2) . \rightarrow . A \Box \rightarrow (C_1 \wedge C_2) \\
\text{(C.N)} & A \Box \rightarrow \top
\end{array}$$

Nous pouvons également dériver les contreparties mises à jour :

$$\begin{array}{ll}
\text{(RU.EC)} & \frac{B_1 \equiv B_2}{B_1 \leftarrow \diamond A . \equiv . B_2 \leftarrow \diamond A} \\
\text{(RU.M)} & \frac{B_1 \rightarrow B_2}{B_1 \leftarrow \diamond A . \rightarrow . B_2 \leftarrow \diamond A} \\
\text{(U.C)} & (B_1 \vee B_2) \leftarrow \diamond A . \rightarrow . (B_1 \leftarrow \diamond A) \vee (B_2 \leftarrow \diamond A) \\
\text{(U.N)} & \neg(\perp \leftarrow \diamond A)
\end{array}$$

$\mathcal{U}ck$ nous dit donc que si l'opérateur de mise à jour doit être interprété (*via* les règles (R.Conv1) et (R.Conv2)) comme le dual de l'opérateur conditionnel de prévision, alors l'extensionnalité de la base s'ensuit, pourvu que l'on accepte les contraintes de transitivité, monotonie et réflexivité de la déduction. Ce résultat nous permet d'utiliser la sémantique de $\mathcal{U}ck$ afin de compléter la construction du modèle esquissé à la section 2.1.

Les démonstrations utilisent la technique de Ryan et Schobbens (1997) où les énoncés sont transformés en formules avec un seul opérateur non classique, auquel les règles de conversion sont alors appliquées afin de transformer cet opérateur.

La démonstration la plus simple est celle de (U.N) : par la logique classique, $\perp \rightarrow A \Box \rightarrow \perp$ est un théorème. Par la règle de conversion (R.Conv2), nous obtenons $\perp \leftarrow \diamond A \rightarrow \perp$. De nouveau par la logique classique, cela est équivalent à $\neg(\perp \leftarrow \diamond A)$, c'est-à-dire (U.N).

Afin d'établir (RU.EC), supposons $B_1 \equiv B_2$. Soit C un schéma atomique nouveau, c'est-à-dire n'apparaissant ni dans B_1 ni dans B_2 ; alors :

$$\frac{B_1 \rightarrow A \Box \rightarrow C}{B_2 \rightarrow A \Box \rightarrow C}$$

Avec (R.Conv1) et (R.Conv2), cette règle est déductivement équivalente¹³ à :

$$\frac{B_1 \leftarrow \diamond A \rightarrow C}{B_2 \leftarrow \diamond A \rightarrow C}$$

13. Deux schémas (d'axiome ou de règle) sont déductivement équivalents dans un système logique \mathcal{L} si, pour tout ensemble de formules S fermé sous la relation de déduction de \mathcal{L} , on a que S est fermé sous le premier schéma si et seulement si S est fermé sous le second schéma. En logique classique, un schéma de la forme $F \rightarrow G$ est déductivement équivalent à un schéma de la forme $\frac{B \rightarrow F}{B \rightarrow G}$, où B est un schéma atomique nouveau (*fresh*) n'apparaissant ni dans F ni dans G .

mais ce dernier est déductivement équivalent à $B_2 \leftarrow \diamond A \rightarrow B_1 \leftarrow \diamond A$. De la même manière, nous pouvons démontrer dans l'autre sens que $B_1 \leftarrow \diamond A \rightarrow B_2 \leftarrow \diamond A$ peut être inféré de $B_1 \equiv B_2$, ce qui établit (RU.EC).

Ainsi, le RT nous donne un principe d'extensionnalité pour des bases de croyances, mais pas pour des entrées. Tout comme la logique conditionnelle normale de base \mathcal{CK} est obtenue à partir de \mathcal{K} en ajoutant un principe d'extensionnalité pour des entrées, la logique normale de base des mises à jour et des conditionnels \mathcal{UCK} sera obtenue de la même manière en ajoutant la règle suivante :

$$(RC.EA) \quad \frac{A_1 \equiv A_2}{A_1 \Box \rightarrow C \equiv A_2 \Box \rightarrow C}$$

2.5.2. Modèles

Rappelons brièvement les modèles de Chellas pour les logiques des conditionnels, qui sont basés sur des fonctions de sélection. Nous pouvons les reprendre en tant que tels pour \mathcal{UCK} , en rajoutant une interprétation des mises à jour.

Un modèle consiste en :

- un ensemble W (les « mondes possibles »);
- une fonction de sélection associant à un monde w et un ensemble de mondes U un ensemble de mondes $w \cdot U$ (les « mondes de U les plus proches de w »);
- une fonction d'interprétation associant à chaque énoncé A l'ensemble de tous les mondes possibles où A est vrai (les « A -mondes »). L'ensemble de mondes $[A \Box \rightarrow C]$ est l'ensemble des mondes possibles w tels que C est vrai dans les A -mondes les plus proches de w :

$$[A \Box \rightarrow C] = \{w \in W : w \cdot [A] \subseteq [C]\}$$

Afin de valider les règles de conversion de Ramsey, nous avons besoin de la condition de vérité suivante pour l'opérateur de mise à jour :

$$[B \leftarrow \diamond A] = \{w \in W : \exists u \in B, w \in u \cdot [A]\}$$

La complétude de \mathcal{UCK} peut alors être établie à partir de la complétude de \mathcal{CK} en suivant la démonstration de complétude de la logique temporelle avec futur et passé (Herzig, 1998).

2.5.3. Extensions

Avant de nous arrêter plus longuement sur la nature de la fonction de sélection, nous pourrions nous demander, dans l'esprit d'AGM ou KM, quelles autres propriétés l'opérateur de mise à jour devrait-il satisfaire ?

Un principe qui ne découle pas de RT est :

$$(U.ID) \quad B \leftarrow \diamond A \rightarrow A$$

Il semble assez naturel de postuler que la mise à jour d'une base par A doit contenir A (cependant, des opérations de changement sans un tel principe de priorité de la nouvelle information ont été étudiées sous le nom révision non prioritaire en philosophie (Hansson, 1999) et sous le nom fusion des données en IA). Sémantiquement, un tel principe peut être garanti en stipulant que $w \in w \cdot U$.

D'autres principes sont :

$$(U.MP) \quad B \wedge A . \rightarrow B \leftarrow \diamond A$$

$$(RU.ASC) \quad \frac{B \leftarrow \diamond A \rightarrow A'}{B \leftarrow \diamond (A \wedge A') \rightarrow B \leftarrow \diamond A}$$

Tous ces axiomes sont identiques ou dérivables des postulats de mise à jour de Katsuno et Mendelzon (1992).

Des axiomes et les règles en termes de conditionnels équivalents à ceux-là peuvent être obtenus par l'intermédiaire des règles de Ramsey suivant la technique de Ryan et Schobbens (1997).

Le choix des contraintes devrait toutefois se faire en accord avec le choix des aspects intensionnels pris en considération pour spécifier contextuellement la nature de la fonction de sélection. \mathcal{UCk} constitue l'ensemble des contraintes minimales acceptables dans toute analyse qui se propose de raffiner la description intensionnelle propre au cadre modal. Toute contrainte ultérieure devrait s'harmoniser avec l'analyse intensionnelle fine que nous esquissons à la section suivante.

2.6. Retour sur la question de l'intensionnalité

Comment rendre la fonction de sélection plus constructive ? Quel type de contraintes intensionnelles pouvons-nous prendre en considération pour la rendre effective ?

Reprenons la question au début. A chaque base B nous associons un ensemble W de mondes possibles représentant les différentes manières de compléter la base, c'est-à-dire plus formellement les différentes valuations (ou interprétations) des atomes contenus dans la base et qui rendent vraies les croyances de la base.

Soit par exemple la base $B = \{\neg p, \neg p \rightarrow \neg q, q \rightarrow r\}$; l'ensemble des mondes associé à B serait alors $W = \{w_1, w_2\}$, où $w_1 = \{\neg p, \neg q, \neg r\}$ et $w_2 = \{\neg p, \neg q, r\}$. Une mise à jour de B par la croyance A , $B \leftarrow \diamond A$, est donc interprétée par une fonction

qui sélectionne à partir de w_1 et w_2 les mondes les plus similaires contenant A (étant donné que nous acceptons le principe (ID)). Nous avons donc que $w_1 \cdot [A]$ est un sous-ensemble de $[A]$. Nous voudrions avoir maintenant un moyen de choisir parmi tous les mondes possibles satisfaisant A ceux qui diffèrent le moins possible de w_1 et w_2 , ceux qui leur sont les plus similaires. La simple prise en considération de $[B]$ et de $[A]$ ne nous permet pas d'obtenir $[B \leftarrow \diamond A]$. Cela n'est pas étonnant étant donné la faiblesse propre à la représentation intensionnelle rendue possible par leur extensionnalisation. Plus haut (section 2.1), nous avons déjà remarqué qu'il était nécessaire d'ajouter à une telle représentation extensionnelle des intensions, de l'information structurelle, permettant de rendre compte plus finement de la notion de contenu.

Une première suggestion nous est livrée par l'analyse de l'intensionnalité de Church. Les informations structurelles recherchées peuvent être rattachées à la structure syntaxique de l'énoncé dont on se propose d'expliquer le contenu. Selon l'alternative 0 présentée dans Church (1951), deux énoncés expriment la même proposition s'ils sont isomorphes et si leurs constituants ultimes sont synonymes. Une telle prise en considération de la structure de l'énoncé permet d'éviter les définitions circulaires des concepts et de donner une explication acceptable des contextes de croyance. « Jean croit que D » et « Jean croit que D' » n'expriment pas la même proposition sauf si D et D' sont non seulement logiquement équivalents mais aussi intensionnellement isomorphes, c'est-à-dire construits de la même manière à partir d'atomes synonymes.

Dans le cas des changements de croyance, ce n'est évidemment pas l'identité des propositions qui est en jeu, mais plutôt la maximisation dans la conservation du sens des informations véhiculées par B , après mise à jour par A . Toutefois, on peut utiliser l'analyse intensionnelle de Church dans le cadre des changements de croyance de manière assez simple. Lorsque de $[B]$, nous voulons passer à la composition de $[B]$ et $[A]$, nous devons prendre en considération les composantes ultimes des énoncés B et A , c'est-à-dire leurs atomes. Si donc $[B] = \{w_1, w_2\}$, lorsque la mise à jour par A est opérée sur w_1 , (respectivement w_2), la fonction de sélection choisit les mondes possibles qui préservent la valeur de vérité des atomes de w_1 (respectivement w_2) qui n'appartiennent pas à A . Une telle fonction de mise à jour a été effectivement définie par Winslett à l'aide d'une mesure qualitative de la distance entre deux mondes (ou interprétations) d'un ensemble d'atomes. *Grosso modo* (voir paragraphe 2.6) la distance entre deux mondes w et v (notée $Dist(w, v)$) est l'ensemble d'atomes qui ont une valuation différente dans les deux mondes. Le critère d'identité entre propositions de Church prend en considération les constituants ultimes des énoncés exprimant ces propositions et leurs structures. Le critère de similarité de Winslett se limite à prendre en considération les constituants ultimes, les atomes, des propositions mises en jeu dans la révision. Ainsi :

$$w_1 \cdot [A] = \{u \in [A] : Dist(w_1, u) \subseteq Atm(A)\}$$

Peut-on pousser plus loin notre analyse? Peut-on prendre en considération des composantes plus fines encore du sens d'un énoncé? Peut-on trouver d'autres types

d'informations structurelles devant être prises en considération lors d'un changement minimal ?

La signification d'un énoncé dépend évidemment de sa valeur de vérité, de sa dénotation, de sa référence dans le monde actuel, mais aussi de la « place » que cet énoncé occupe dans le langage servant à décrire le monde. Les linguistes structuralistes ont souvent insisté sur la particularité des associations symboliques caractéristiques du langage. De Saussure (1978, p. 169-182) affirmait que le langage est une forme et non une substance car l'idée désignée par un mot ne correspond jamais à une classification naturelle préexistante au code linguistique particulier de nos idiomes, mais dépend toujours du système caractéristique des différences et des oppositions intralinguistiques que ce mot entretient avec les autres mots du langage. Le contenu d'un mot est fixé par ses relations négatives avec les autres termes du cadre linguistique auquel il appartient. Prenons l'exemple des antonymes¹⁴. Dans toutes les langues, on constate de nombreux cas de termes ou d'énoncés élémentaires considérés comme contraires (par exemple noir/blanc, lourd/léger, etc.). Plus généralement, dans la plupart des langues, une série (pas seulement un couple) de termes antonymes subdivisent les stimulations sensorielles d'un certain genre (visuel, tactile, etc.). Une telle subdivision engendre une forme de dépendance négative entre un mot et l'ensemble de ses antonymes. De telles dépendances doivent être prises en considération dans l'analyse du sens d'une expression. Prenons le cas des couleurs. Il se pourrait bien que dans deux langues les mêmes stimulations sensorielles soient liées à l'énoncé « *a* est rouge ». Toutefois, si rouge s'oppose antinomiquement dans la première langue à une liste de six éléments et dans la deuxième langue à une liste de trois éléments, le sens de « *a* est rouge » n'est évidemment pas le même dans les deux langues.

Le cas de l'antinomie ainsi que celui de la synonymie ne représentent que certaines des formes de dépendance intralinguistique. Les connexions causales ou fonctionnelles pourraient être rajoutées à cette liste.

Lorsque nous considérons un signe non pas seulement pour sa valeur désignative par rapport au monde actuel, mais aussi pour sa valeur oppositive, nous prenons en considération une des composantes de son sens liée à la structure non pas du signe lui-même, mais du langage auquel le signe appartient. Cette valeur oppositive d'un signe peut, dans certaines circonstances, être complètement extensionnalisée. C'est le cas lorsqu'un signe, un mot, fait partie d'une théorie scientifique axiomatisable explicitement. Les énoncés élémentaires « *a* est un nombre premier » et « *a* est un nombre pair différent de 2 » s'opposent par le biais d'une structure argumentative explicite, une preuve, dans la théorie, et selon la modalité propre à la négation interprétée comme complémentation entre des ensembles d'objets.

14. Voir Vuillemin (1986, p. 34-39).

Toutefois, ce cas idéalisé ne constitue sûrement pas la norme dans le cas de simples ensembles de croyances, lorsque croyance s'oppose à science. L'opposition des couples antinomiques lourd/léger, grand/petit, animé/inanimé peut évidemment être extensionnalisée si l'on peut, même arbitrairement, poser une frontière nette entre l'ensemble d'objets satisfaisant le premier terme de l'opposition et l'ensemble d'objets satisfaisant le deuxième terme. Il serait toutefois absurde de prétendre que l'on ne comprend pas le sens du mot inanimé, si l'on est incapable de décider de cas limites comme des individus dans le coma ou des organismes unicellulaires.

De plus, si nous nous plaçons dans un contexte évolutif des croyances, où il ne s'agit pas de décrire statiquement un ensemble de croyances mais de représenter leurs changements, il est bien possible que les relations de dépendance puissent varier au sens où d'autres termes indépendants peuvent devenir dépendants entre eux et *vice versa*.

Une opération de mise à jour basée sur la dépendance

Dans Herzig (1996) et Herzig et Rifi (1999), il a été proposé une logique basée sur une relation de dépendance métalinguistique entre les énoncés atomiques (atomes).

Formellement, il s'agit d'une application *dep* associant des ensembles d'atomes à des atomes, telle que $p \in dep(p)$ pour tout atome p . Cette fonction *dep* est étendue à des formules complexes en stipulant :

$$dep(A) = \bigcup_{p \text{ apparaît dans } A} dep(p)$$

Donc, des énoncés logiquement équivalents peuvent ne pas avoir les mêmes dépendances, comme par exemple p et $p \wedge (q \rightarrow q)$.

L'idée est alors que si l'atome p ne dépend pas de la proposition A , alors p doit être préservé à travers une mise à jour par A :

$$\text{si } p \notin dep(A) \text{ et } w \in [p] \text{ alors } w \cdot A \subseteq [p]$$

Inversement, si p dépend de A , il se peut que la mise à jour par A affecte également p : la relation de dépendance ne donne aucun renseignement dans ce cas¹⁵. L'opération de mise à jour est donc définie comme :

$$w \cdot A = \{u \in [A] : DIST(w, u) \subseteq dep(A)\}$$

15. Nous avons donc ici un modèle de \mathcal{UCk} et non de \mathcal{UCK} : le second argument de la fonction de sélection n'est plus un ensemble de mondes, mais un énoncé.

où la distance entre w et v est l'ensemble d'atomes dont l'interprétation diffère entre w et v :

$$\begin{aligned} DIST(w, v) &= (w \setminus v) \cup (v \setminus w) \\ &= \{p : w \in [p] \text{ et } v \notin [p]\} \cup \{p : w \notin [p] \text{ et } v \in [p]\} \end{aligned}$$

2.7. Conclusion

Comment interpréter l'ajout d'une information à une base de croyances ? En partant des deux *desiderata* de la constructivité et de la finesse de l'analyse intensionnelle, nous avons opté pour trois choix méthodologiques :

- partir d'une base de croyances finie, interprétée à l'aide de la sémantique des mondes possibles ;
- interpréter l'opérateur de mise à jour, *via* le RT pris comme une règle d'inférence si-et-seulement-si, comme dual par rapport à un conditionnel contrefactuel ;
- exprimer l'opérateur de mise à jour dans le langage objet.

Le système formel obtenu en ajoutant la logique classique au RT se révèle être une logique conditionnelle minimale. La fonction de sélection associée à l'interprétation sémantique de ce conditionnel dans une structure de mondes possibles a été ensuite explicitée grâce à la notion intensionnelle de dépendance.

Ce schéma d'analyse ne se prétend ni le seul ni le meilleur répondant au *desideratum* de la constructivité. Il suggère toutefois trois conclusions :

- 1) il est bien difficile de conjuguer constructivité et généralité sans l'aide d'une analyse intensionnelle fine (hyperintensionnelle pour ainsi dire) du contenu de la base. Le credo extensionaliste quinien a été, en ce sens, une grosse entrave au développement de la théorie de la révision ;
- 2) il ne faut pas espérer pouvoir appliquer directement les résultats obtenus dans le cadre des analyses hyperintensionnelles statiques (à la Church ou à la Bealer) à un cadre dynamique où l'incomplétude de nos croyances est la donnée essentielle ;
- 3) le critère de mutilation minimale utilisé dans la littérature de la révision pour mettre en œuvre la notion de changement minimal est, à l'évidence, trop simpliste. Les notions de cohérence et d'ordre d'enracinement épistémique des croyances sont trop pauvres pour rendre compte de la conservation du sens, du contenu des informations de la base après une mise à jour ¹⁶.

16. Une version préliminaire de ce chapitre est apparue dans les actes du vingtième *International Wittgenstein Symposium* (Kirchberg am Wechsel, Autriche, août 1997).

2.8. Bibliographie

- Bealer, G. (1982). *Quality and Concept*. Oxford : Clarendon Press.
- Carnap, R. (1947). *Meaning and Necessity*. Chicago : Chicago University Press.
- Carnap, R. (1950). *Logical Foundations of Probability*. Chicago : Chicago University Press.
Deuxième édition Routledge and Kegan Paul, 1951 ; troisième édition 1962.
- Church, A. (1951). A formulation of the logic of sense and denotation. In P. Henle, H. Kallen, & S. Langer (Eds.), *Structure, Method, and Meaning: Essays in Honor of H. Sheffer*. Art Press.
- Crocco, G., Fariñas del Cerro, L., & Herzig, A. (Eds.) (1995). *Conditionals: From Philosophy to Computer Science*. Oxford University Press (Studies in Logic and Computation n° 5).
- Fuhrmann, A. (1989). Reflective modalities and theory change. *Synthese*, 81, 115-134.
- Gärdenfors, P. (1978). Conditionals and changes of belief. In I. Niiniluoto, & R. Tuomela (Eds.), *The Logic and Epistemology of Scientific Change* (pp. 381-404). Acta Philosophica Fennica n° 30.
- Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press.
- Gärdenfors, P. (Ed.) (1992). *Belief Revision*. Cambridge : Cambridge University Press.
- Grahne, G. (1991). Updates and counterfactuals. In J. Allen, R. Fikes, & E. Sandewall (Eds.), *Proceedings of the Third International Conference on Knowledge Representation and Reasoning (KR'91)* (pp. 269-276). Morgan Kaufmann Publishers. Version étendue à paraître dans *Journal of Logic and Computation*.
- Hansson, S. O. (1995). The emperor's new clothes: Some recurring problems in the formal analysis of counterfactuals. In G. Crocco, L. Fariñas del Cerro, & A. Herzig (Eds.), *Conditionals: From Philosophy to Computer Science* (pp. 13-31). Oxford University Press (Studies in Logic and Computation n° 5).
- Hansson, S. O. (1999). Special issue on non-prioritized belief revision. *Theoria*, 63, 1-134.
- Herzig, A. (novembre 1996). The PMA revisited. In L. C. Aiello, & S. Shapiro (Eds.), *Proceedings of the International Conference on Knowledge Representation and Reasoning (KR'96)* (pp. 40-50). Morgan Kaufmann Publishers.
- Herzig, A. (1998). Logics for belief base updating. In D. Dubois, D. Gabbay, H. Prade, & P. Smets (Eds.), *Handbook of Defeasible Reasoning and Uncertainty Management*, Vol. 3—Belief Change (pp. 189-231). Kluwer Academic Publishers.
- Herzig, A., & Rifi, O. (1999). Propositional belief base update and minimal change. *Artificial Intelligence Journal*, 115(1), 107-138.
- Hintikka, J. K. K. (1962). *Knowledge and Belief*. Ithaca, New York : Cornell University Press.
- Katsuno, H., & Mendelzon, A. O. (1992). On the difference between updating a knowledge base and revising it. In P. Gärdenfors (Ed.), *Belief Revision* (pp. 183-203). Cambridge : Cambridge University Press. Version préliminaire dans J. A. Allen, R. Fikes, & E. Sandewall (Eds.) (1991), *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference* (pp. 387-394). Morgan Kaufmann Publishers.

- Lakemeyer, G., & Lang, W. (1996). Belief revision in a nonclassical logic. In G. Görz, & S. Hölldobler (Eds.), *KI-96: Advances in Artificial Intelligence* (pp. 199-211). Springer-Verlag (LNCS n° 1137).
- Levi, I. (1988). Iteration of conditionals and the Ramsey test. *Synthèse*, 76, 49-81.
- Lewis, D. (1973). *Counterfactuals*. Oxford : Basil Blackwell.
- Lindström, S., & Rabinowicz, W. (1995). The Ramsey test revisited. In G. Crocco, L. Fariñas del Cerro, & A. Herzig (Eds.), *Conditionals: From Philosophy to Computer Science* (pp. 147-192). Oxford University Press (Studies in Logic and Computation n° 5).
- Nebel, B. (1991). Belief revision and default reasoning: Syntax-based approaches. In J.A. Allen, R. Fikes, & E. Sandewall (Eds.), *Proceedings of the Third International Conference on Knowledge Representation and Reasoning (KR'91)* (pp. 417-428). Cambridge, Massachusetts : Morgan Kaufmann Publishers.
- Nebel, B. (1992). Syntax-based approaches to belief revision. In P. Gärdenfors (Ed.), *Belief Revision* (pp. 247-275). Cambridge : Cambridge University Press.
- Quine, W. V. O., & Ullian, J. S. (1978). *The Web of Belief*. New York : Random House.
- Ramsey, F. P. (1931). General propositions and causality. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays* (pp. 237-257). New York : Routledge and Kegan Paul.
- Restall, G., & Slaney, J. (1995). Realistic belief revision. In *Proceedings of the First World Congress in the Fundamentals of Artificial Intelligence (WOCFAI'95, Paris, France)* (pp. 367-378).
- Rott, H. (2000). Two dogmas of belief revision. *The Journal of Philosophy*, 97, 503-522.
- Ryan, M., & Schobbens, P. Y. (1997). Intertranslating counterfactuals and updates. *Journal of Logic, Language, and Information*, 6(2), 123-146. Version préliminaire dans W. Wahlster (Ed.), *Proceedings of ECAI'96*.
- Sahlin, N. E. (1988). *The Philosophy of F. P. Ramsey*. Cambridge University Press.
- Saussure, F. de (1978). *Cours de linguistique générale. Edition critique préparée par T. De Mauro*. Paris : Payot.
- Vuillemin, J. (1986). *What are Philosophical Systems?* Cambridge University Press.
- Wansing, H. (1994). Sequent calculi for normal modal propositional logics. *Journal of Logic and Computation*, 4, 125-142.
- Winslett, M. A. (1988). Reasoning about action using a possible models approach. In *Proceedings of the Seventh Conference on Artificial Intelligence (AAAI'88, St. Paul)* (pp. 89-93).
- Winslett, M. A. (1990). *Updating Logical Databases*. Cambridge University Press (Cambridge Tracts in Theoretical Computer Science).
- Wright, G. H. von (1951). *An Essay in Modal Logic*. Amsterdam : North Holland.