N° d'ordre : 03 / 2005 – M / IN

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE D'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE
« HOUARI BOUMEDIENNE »

FACULTE D'ELECTRONIQUE ET INFORMATIQUE

MEMOIRE

Présenté pour l'obtention du diplôme de MAGISTER

EN : INFORMATIQUE

Spécialité : Intelligence Artificielle et Base de Données Avancées

Par : BOUTOUHAMI SARA

Sujet

MODELISATION QUALITATIVE DE L'EXPLICATION CAUSALE

Soutenu le 03/07/2005, devant le jury composé de :

Mr- Mr N.BADACHE, Professeur, USTHB	Président
Mme- A.MOKHTARI, Professeur, USTHB	Dteur de Thèse
Mr- A.AISSANI, Professeur, USTHB	Examinateur
Mr- S.LARABI, Maître de conférence, USTHB	Examinateur
Mme- F.KHELLAF, Chargée de cours	Invitée

Remerciements

Je tiens à exprimer ma reconnaissance à madame ***Aïcha Mokhtari-Aissani*** *professeur à l'USTHB, ma directrice de thèse. Je la remercie pour son orientation, ses suggestions pertinentes, ses explications précieuses, sa rigueur scientifique et sa patience.*

Je tiens à remercier les membres du jury :

- *M.* ***A.BADACHE*** *: Professeur à l'USTHB, qui m'a fait l'honneur de bien vouloir présider ce jury.*
- *M.* ***A.AISSANI*** *: Professeur à l'USTHB, qui a accepté d'être un examinateur de ma thèse.*
- *M.* ***S.LARABI*** *: Maître de conférence à l'USTHB, qui a accepté d'être un examinateur de ma thèse.*
- *Mme* ***F.KHELLAF*** *: Chargée de cours à l'USTHB, qui a accepté d'être invitée du domaine.*

Je tiens à remercier du fond du cœur M. ***Nouioua Farid*** *pour son aide.*

Résumé

Le concept de la causalité est central dans notre raisonnement de tous les jours et intuitivement banal. Pourtant diverses disciplines s'y intéressent, certaines depuis des siècles. La logique a souvent été un formalisme adopté par les auteurs pour représenter les relations causales. Plusieurs approches ont été développées pour la modélisation de la causalité.

Nous présentons dans cette thèse trois principales modélisations de la relation causale parmi les modélisations existantes ; une modélisation graphique, via les réseaux causaux bayésiens ; une modélisation agentive ; via l'approche normative de la causalité et une approche contre-factuelle par une modélisation via un modèle d'équations structurelles.

Nous proposons ensuite une comparaison entre les deux dernières approches, en soulignant les points communs ainsi que les points de divergences.

Les chercheurs en Intelligence Artificielle (IA) font souvent appel à la notion de causalité de manière implicite ou explicite, ne serait-ce que quand ils parlent de "relations causales", de "processus causaux". C'est en général le cas dans les problèmes de diagnostic, de planification ou d'explication. La génération automatique d'explications s'avère une tâche essentielle dans la planification, le diagnostic et le langage naturel. Le raisonnement qualitatif constitue un sujet classique de l'IA. La raison de ceci est que les activités cognitives des êtres humains sont au niveau qualitatif.

Dans ce travail, nous proposons une approche qualitative de l'explication dans le cadre représentationel fourni par la logique possibiliste. Ceci par une modification de la définition de Halpern et Pearl pour l'explication en remplaçant les probabilités par des possibilités. Une réorganisation stratifiée des explications fournies est donnée, selon l'ordre des possibilités des explications.

Sommaire

Introduction……
I. Introduction……. 1
II. Proposition…… 2
III. Lecture de thèse…… 3
Etude de la causalité……..
I. Introduction…… 5
II. Etude sur la causalité……. 6
II.1 Causalité en philosophie…… 6
II.2 Etude logique…….…… 7
II.2.1 Quelques observations sur les relations causales……... 8
II.2.2 Frame problem…… 8
III. Raisonnement temporel en I.A……. 10
III.1 Propriétés des logiques non-monotones…… 11
III.2 Caractéristiques et problèmes des logiques non-monotones…… 12
III.3 Théorie des changements…… 12
III.4.1 Théorie des anomalies……... 12
III.4.2 Les modèles préférés de Shoham…… 14
IV. Conclusion……... 16
Les réseaux bayésiens……
I. Introduction…… 17
II. Fondement nécessaire pour les réseaux bayésiens…… 17
II.1 L'incertitude……... 17
II.2 Une représentation graphique de la causalité……. 18
II.2.1 circulation de l'information dans un graphe causal…… 18
II.2.2 Notion de D-séparation……. 19
III. Les réseaux bayésiens……. 20
III.1 Définition d'un réseau bayésien…….. 20
III.1.1 Exemple explicatif……. 20
III.2 D-séparation……. 22
III.3 Indépendance conditionnelle……... 23
IV. Inférence …….. 24
IV.1 Inférence dans les réseaux sans boucles……. 24
IV.1.1 Définitions(notions de graphe)…….. 24
IV.1.2 Propriétés de D-séparation…… 25
IV.1.3 Le problème de mis à jour des probabilités…….. 26
IV.2 Inférence dans les réseaux avec boucles……. 27
IV.2.1 Méthode de conditionnement……. 27
IV.2.2 Méthode de regroupement…… 28
IV.3 Construction de l'arbre de jonction …… 31
IV.3.1 Construction de l'arbre de jonction pour un arbre triangulé……... 32
IV.3.2 Triangulation d'un graphe moral…….. 32
IV.3.3 recherche de l'arbre de poids maximal……. 32
V. Complexité…… 32
V. 1 Le problème 3SAT…… 33
V.2 Réduction de l'inférence…….. 33

VI. Avantages et limites des réseaux bayésiens……. 34
VI.1 Avantages des réseaux bayésiens…… 34
VI.2 Les limites des réseaux bayésiens……. 35
VIII. Conclusion……. 36
Approche normative de la causalité…….……..……..
I. Introduction…… 37
II. Action et causalité……. 37
II.1 Le temps dans la causalité…… 37
II.2 Le Langage ……..…. 38
II.3 Non-monotonie dans la causalité…….. 40
II.4 Implication normale…….. 40
II.4.1 La persistance normale……... 42
II.4.2 Rôle de l'agent…… 42
III. La théorie causale……. 43
III.1 Caractérisation de la causation…… 43
III.2 Explication…….. 44
III.2.1 Algorithme …… 44
III.2.2 Complexité…….. 46
III.3 Prédiction…… 46
III.2.1 Algorithme …… 46
III.2.2 Complexité…….. 47
IV. Ramification……. 47
IV.1 Ramification dans la méthode normative de causalité…….. 48
IV.2 Génération automatique des rapports des effets indirects……... 49
IV.3 Algorithme de génération des rapports des effets indirects…… 50
V. Conclusion……. 51
Approche contre-factuelle……
I. Introduction……. 52
II. Modèle causal…… 52
II.1 Modèle causal……. 53
II.2 Syntaxe et sémantique…….. 55
II.3 Définition de la cause……... 56
II.3.1 Définition…… 56
II.3.2 Exemple explicatif……. 58
III. Exemples de modélisations ……...…….. 60
IV. Définition plus raffinée de la cause…….. 65
IV.1 Exemples……. 65
V. Comparaison entre l'approche normative et l'approche contre-factuelle……. 69
V. Points communs……. 69
V. Divergences……. 70
VI. Conclusion……... 71

Modélisation Contre-factuelle de l'explication…….
I. Introduction…… 73
II. Explication…… 73
II.1 Définition…….. 74
II.2 Exemple explicatif…….. 74
III. Explication partielle et la puissance explicative……. 76
III.1 Exemple1……. 77
III.2 Exemple1……. 77
III.3 Définition……. 78
IV. Généralisation de l'explication…….. 80
IV.1 Définition…….. 80
IV.2 Exemples…… 81
V. Complexité…… 81
V.1 Notions de base…… 81
V.2 Complexité de l'approche d'halpern et Pearl…… 85
V.2.1 Complexité de cause faible…… 85
V.2.2 Complexité d'explication……... 85
V.2.3 Complexité d'explication partielle et puissance d'explication…….. 87
VI. Conclusion……... 89
Explication Possibiliste…….
I. Introduction……. 91
II. Logique possibiliste……. 91
II.1 Propriétés des mesures de nécessité et de possibilité…….. 92
II.2 Base de connaissance possibilistes…….. 94
II.3 Possibilités/Probabilités…… 96
III. Explication Possibilistic…… 97
III.1 Définition…… 98
III.2 Explication partielle est la puissance d'une explication…… 98
III.3 Définition…… 99
III.4 Puissance d'une explication…… 100
IV. Complexité……... 101
IV.1 Complexité du calcule des strates……... 101
V. Conclusion……. 103
Conclusion générale et perspectives……...
I. Conclusion et Perspectives…….. 104
Bibliographie……

I. Introduction

La notion de causalité joue un rôle essentiel dans l'expression de la perception des phénomènes de notre environnement et de leur description. Elle s'avère difficile à être modélisée pleinement, malgré de nombreuses tentatives anciennes ou récentes. Une source de difficultés semble être le fait que la notion de causalité se trouve reliée à d'autres idées comme celles d'explication, ou de responsabilité notamment, ce qui rend encore plus malaisée sa compréhension [Hall03]. Les chercheurs en Intelligence Artificielle (IA) font souvent appel à la notion de causalité de manière implicite ou explicite, ne serait-ce que quand ils parlent de "relations causales ", de "processus causaux". C'est en général le cas dans les problèmes de diagnostic où, à partir d'observations éventuellement imprécises et incertaines, on recherche par abduction la ou les causes plausibles d'une situation.

La causalité est un concept épistémologique formalisant le fondement même de toute démarche scientifique qui consiste à organiser systématiquement les faits empiriques et de leur donner sens. Cette définition constructiviste de l'épistémologie s'oppose à l'idée de causalité conçue comme un postulat métaphysique, selon lequel la Raison humaine peut saisir l'être et la raison d'être des choses. Dans le champ scientifique, la causalité est donc le lien construit entre une cause (ou un ensemble de causes) et une conséquence.

La logique a souvent été un formalisme adopté par les auteurs pour représenter les relations causales. Nos raisonnements même les plus élémentaires, ne sont pas infaillibles, plutôt des raisonnements réversibles ce qui n'est pas possible à modéliser par les logiques classiques et nécessite des logiques non-monotones [Mokhtari97a].

Plusieurs approches ont été développées pour la modélisation de la causalité. La conception majeure de la causalité développée au XXe siècle : la conception qui assimile la causalité à une explication. L'analyse probabiliste réduit la relation causale à celle d'une augmentation de probabilité. L'approche interventionniste ou « agentive » analyse le rapport causal à partir du rapport entre l'agent et son action. Selon l'approche contre-factuelle, le rapport entre cause et effet se réduit au rapport de dépendance contre-factuelle [Kistler04].

Que signifie, ou quand peut-on dire qu'un événement A "cause" un événement B ? Cette question a été depuis longtemps l'objet de spéculations de la part de philosophes, mais aussi de la part de physiciens, de juristes (puisque responsabilité et causalité sont liées), et de psychologues cogniticiens notamment. C'est le cas en particulier au plan philosophique de David Hume [Hume1758] qui a souligné qu'on ne pouvait pas déductivement inférer les effets des causes seules, ni les causes des effets [Dupois03].

Ce n'est qu'assez récemment que des chercheurs en IA se sont à leur tour intéressés aux problèmes posés par la modélisation de la causalité [Shoham88][Mokhtari97a], et récemment Pearl [Pearl98][Pearl00], et Halpern et Pearl [Halpern01][Halpern02][Halpern05]. Cette préoccupation est effectivement naturelle pour l'IA, car il est important de toujours pouvoir munir les systèmes d'inférence ou d'aide à la décision de capacités d'explications, évidemment destinées à un opérateur ou utilisateur humain [Dupois03]. Or l'idée de causalité est étroitement liée à celle d'explication (et peut-être encore plus particulièrement à l'idée d'explication négative » [Safar90], répondant à des questions de type «pourquoi pas ?).

Les liens causaux sont parfois représentés graphiquement par des flèches reliant la cause à l'effet. Les réseaux causaux bayésiens par exemple sont des modèles graphiques, où l'identification des causes s'appuie sur des connaissances reliant «causes » et «effets ». Ces relations sont modélisées par des probabilités conditionnelles de la forme Prob(effets|causes), les effets n'étant pas toujours connus avec précision et certitude. Le cadre bayésien permet notamment de mettre en évidence des phénomènes de renversement d'explication [Pearl88][Becker99][Dubois03].

Ils correspondent à l'influence négative d'une observation sur la plausibilité d'une cause potentielle, suggérée par une première observation. C'est-à-dire qu'une observation complémentaire conduit à écarter une cause pressentie au profit d'une autre cause maintenant plus plausible (compte tenu de la nouvelle observation).

Nous présentons ensuite une approche normative de la causalité [Mokhtari94][Mokhtari97a] [Kayser98]. Elle est basée sur le concept interventionniste de causalité où un agent a le choix pour exécuter (ou pas) une action (volonté libre), c'est une approche permettant un raisonnement non-monotone basé sur l'utilisation des normes et propose une solution au frame problème [Khalfallah01]. Elle se base sur le principe qui stipule que : «une action peut en causer un ou plusieurs effets » et « aucun effet ne précède sa cause ».

Plus récemment, dans une approche différente, Halpern et Pearl [Halpern01][Halpern05] se sont efforcés de distinguer cause réelle ("cause in fact") et cause potentielle, en s'inspirant des idées de I. Good sur le problème de la détermination des responsabilités (qui doit s'appuyer sur les « causes réelles »). Pour modéliser ce problème, Halpern et Pearl proposent un cadre où sont distinguées a priori variables endogènes (dont les valeurs possibles sont régies par des équations structurelles, correspondant par exemple à des lois physiques) et variables exogènes (déterminées par des facteurs extérieurs au modèle). Ces dernières ne peuvent pas fournir de causes. La définition de la causalité dans ce cadre reste étroitement liée à l'idée de conditionnelle « contre-factuelle » (c'est-à-dire que « A cause B » dans la mesure où « il est vrai que si A n'avait pas eu lieu, B ne se serait pas produit »).

Plus précisément, ces auteurs considèrent que le fait qu'un sous-ensemble de variables endogènes A ait pris certaines valeurs est la cause réelle d'un événement B, si A et B sont vrais dans le monde réel, si ce sous-ensemble est minimal, et si une autre affectation de valeurs à ce sous-ensemble de variables rendrait B faux, les valeurs des autres variables endogènes ne participant pas directement à la réalisation de B étant fixées d'une certaine manière, et si A seul suffit à provoquer B dans ce contexte.

II. Proposition

L'approche contre-factuelle offre un modèle raisonnable de l'idée de causalité et permet de traiter des exemples qui posent des problèmes dans d'autres approches de la causalité. La cause réelle est importante dans les applications de l'IA. En effet, chaque fois qu'on explique un ensemble d'événements dans un scénario spécifique, l'explication produite doit reconnaître la cause réelle de ces événements [Halpern05].

La génération automatique d'explications qui s'avère une tâche essentielle dans la planification, le diagnostic et le langage naturel, exige donc une analyse formelle du concept de la cause réelle.

Lorsque nous devons fournir une explication, quelles informations allons-nous chercher et sélectionner ? Dans quel ordre, de quelle façon les produisons-nous ? Expliquer, c'est faire comprendre par un développement oral ou écrit ou par des gestes (Petit Larousse 1995), mais c'est aussi éclaircir, exposer, commenter, formuler une justification, rendre explicite, argumenter, solutionner. Une explication concerne toute opération impliquée dans la constitution du sens d'un phénomène.

Dans la formalisation de l'explication de Halpern et Pearl [Halpern05]; la notion de probabilité est utilisée pour cette modélisation, alors que les probabilités rendent cette modélisation quantitative ce qui s'éloigne du raisonnement humain, qui se dispose de notions plutôt qualitatives. Nous proposons une modification de la définition de Halpern et Pearl pour l'explication en remplaçant les probabilités par des possibilités. Nous proposons une réorganisation stratifiée des explications fournies, selon l'ordre des possibilités des explications [Boutouhami05] ainsi qu'une estimation de la complexité de la génération de l'ensemble des strates des explications partielles possibles pour un événement donné. Notre approche est une approche qualitative de l'explication dans le cadre représentationnel fourni par la logique possibiliste. La logique possibiliste est une extension de la logique classique permettant de traiter qualitativement des informations totalement ordonnées.

Les priorités sont explicitées à l'aide de formules pondérées dont les poids représentent les bornes inférieures de mesures de nécessité. La logique possibiliste possède une inférence syntaxique, correcte et complète par rapport à une inférence sémantique basée sur des distributions possibilistes. Enfin sa complexité est très proche de la logique classique [Benferhat02a].

La spécificité du problème nous a amené à considérer plusieurs aspects de la causalité, philosophique.

III. Lecture de la thèse

- Dans le premier chapitre nous présentons une étude philosophique de la causalité ainsi qu'une une étude des logiques classiques pour montrer les faiblesses de ces dernières à capturer certains aspects de la causalité. Nous présentons ensuite quelques logiques non-monotonnes bien connues en littérature tel que les modèles préférés de Shoham.

- Nous aborderons dans le deuxième chapitre un autre formalisme de la notion de la causalité qui est plutôt graphique : les réseaux bayésiens. Les réseaux bayésiens constituent une approche possible pour intégrer l'incertitude dans le raisonnement, 'prise en compte des faits *précis, mais incertains*' [Becker99]. Dans ce chapitre, nous faisons le point sur les principaux aspects de cette approche ; une introduction générale aux concepts participant à la formalisation des réseaux bayésiens sera suivit par une formulation complète des réseaux bayésiens pour leur utilisation comme modèles d'inférence. Nous terminons avec un résumé sur les avantages et limites de cette approche.

- Nous présentons dans le troisième chapitre l'approche Normative de la causalité proposée par Mokhtari ainsi que son extension pour traiter le problème bien connu de la ramification[Mokhtari94][Mokhtari97b][Kayser98][Khalfallah01], une approche symbolique non-monotone.

- Dans le quatrième chapitre nous allons voir une autre modélisation de la relation causale. Une nouvelle définition de la «cause réelle » en utilisant des équations structurelles modélisant la notion de contre-factuelle proposée par J.Y.Halpern et J. Pearl. Cette approche semble être une bonne représentation de la causalité [Halpern01] [Halpern05]**.** Les énoncés dits contre-factuels [Lewis1973] sont des énoncés conditionnels dont l'antécédent est faux, Si Jean avait un vélo, il ne marcherait pas. C'est une autre façon de modéliser la notion de causalité. Nous terminons ce chapitre par une étude comparative de cette approche et l'approche normative sur la causalité.

- Dans le cinquième chapitre nous présentons la modélisation de l'explication contre-factuelle en se basant sur la définition de la cause réelle [Halpern 02][Halpern 05], ainsi qu'un ensemble d'exemples. Qui sera suivit par une étude de la complexité de cette dernière [Eiter04].
- Le sixième chapitre est consacré à notre proposition : une explication possibiliste ainsi que le calcul de complexité. Cette approche basée sur le raisonnement qualitatif possibiliste utilise les notions de base de la logique possibiliste qui fera l'objet d'une brève présentation au début de ce chapitre.
- Nous terminons ce mémoire par une conclusion générale et des perspectives à ce travail.

I. Introduction

Le concept de causalité est central dans notre raisonnement de tous les jours et intuitivement banal. Pourtant diverses disciplines s'y intéressent, certaines depuis des siècles.
Pour essayer de comprendre le pourquoi de cet intérêt, nous présentons dans ce chapitre une étude dans deux principales disciplines la philosophie et la logique.
La logique a été toujours utilisée comme un outil pour modéliser les relations causales, mais les logiques classiques s'avèrent inadaptées au raisonnement causal qui est plutôt non-monotone et temporel [Mokhtari97a][Mokhtari97b]. Dans la troisième partie de ce chapitre nous abordons le raisonnement temporel en IA ; en outre nous présentons les caractéristiques des logiques non-monotones. Nous survolons enfin un certain nombre de développements bien connus dans le raisonnement non-monotone comme la théorie des anomalies proposée par McCarty [McCarty86] et les modèles préférés de Shoham [Shoham88].

I.1. Première définition

Le premier réflexe que nous pouvons avoir pour savoir ce qu'est une cause est de consulter un dictionnaire. Que dit le dictionnaire Larousse à ce sujet ?

- La causalité est le lien qui unit la cause à son effet,
- Causation est l'action de produire un effet, et
- La cause est ce par quoi une chose est ou arrive (cause physique, morale, occasionnelle, occulte, etc…) ou ce pour quoi l'on fait quelque chose (motif, raison, etc…).

Ces définitions ne disent pas grand chose, la circularité étant largement utilisée : définition de la causalité par cause, puis cause par «ce par quoi », «ce pour quoi» ou «par» et «pour» sont définies comme des prépositions introduisant un complément de cause, et enfin la définition de la causation par effet qui lui-même renvoie à cause !

I.2. Second définition

Selon M.Espinoza [Espinoza92] «Aristote a distingué quatre types de causes : matérielle, formelle, efficiente et finale ». Les deux premières sont internes ou intrinsèques, les deux autres, externes ou extrinsèques. La matière est le flux adéquat et passif sur lequel agissent les autres causes. La cause formelle est la qualité, l'essence de la chose, l'idée qui la définit. La cause efficiente est la force, le pouvoir qui produit quelque chose. La cause finale est le but vers lequel la chose est orientée ».
Bunge est de ceux qui pensent que «les causes formelles et finales font partie d'une métaphysique spirituelle et pseudo-scientifique dont les explications sont illusoires. La science moderne a montré que les causes formelles et finales ne sont ni empiriquement vérifiables, ni mathématiquement exprimable… ».
Espinosa se demande, s'il est vrai, comme le pense Bunge, que les causes efficientes sont les seules clairement concevables et de ce fait les seules capables d'expression mathématique ? En tant que forces inobservables, elles sont aussi spéculatives que les autres causes. Et il n'est pas vrai non plus que les causes formelles et finales soient incapables d'être exprimées mathématiquement. Cela peut paraître ambitieux de vouloir traiter de la causalité vue ces grands débats et controverses.

I.3 Définition pratique

Nous savons que selon A.Mokhtari A cause B n'est pas une vérité du monde, elle est liée à un état de connaissances ou de croyances. Nous avons besoin de «poser» des causalités parce que *l'inférence déductive* (prédiction) ou *abductive* (explication, diagnostic) est une nécessité vitale. Même en admettant que l'état de nos connaissances et de nos croyances justifie (dans un certain sens à élucider) l'affirmation A cause B, nous ne pouvons pas efficacement l'utiliser telle quelle, car A est souvent une conjonction infinie de facteurs [Mokhtari94].

Un schéma causal intéressant doit satisfaire un certain compromis entre simplicité et précision. Il s'agit donc plutôt d'une organisation des connaissances ; si on maximise les liens entre les causes possibles et les effets possibles, on perd tout le bénéfice de l'indexation, car le système de planification ou de diagnostic est submergé et ne fonctionne plus ; on doit se limiter à relier causalement peu d'éléments, d'où nécessité d'approximation, de pertinence et de choix.

II. Etude sur la causalité

II.1 Causalité en philosophie

Nous trouvons en philosophie, principalement, deux points de vue, totalement différents, sur la causalité :

- Le point de vue de D.Hume [Hume75] qui décrit une relation causale quelconque comme une succession temporelle invariable de deux événements.
  Bien qu'il soit admis que les causes précèdent leurs effets, ce point de vue est en vérité insuffisant.
  - il n'y a qu`à reprendre la vieille discussion portant sur la question de savoir si le jour est la cause de la nuit, parce que le jour est suivi de la nuit, ou
  - l'exemple de D.McDermott [McDermott82b] : en s'éloignant de la direction du soleil, l'arrivée de l'ombre de a est suivie de l'arrivée de a, signifie-t-il que l'arrivée de l'ombre de a est la cause de l'arrivée de a ?
- Le second point de vue, basé sur la notion d'action, associé généralement à Kant, semble être la voie actuellement suivie par les philosophes [Wright73][Puech90].

  Puech énumère extensivement toutes les questions auxquelles nous devrions répondre, selon lui, pour posséder une théorie de la causalité [Puech90] :

  1. Peut-on préciser ce que signifient cause et effet ?
  2. Qu'est-ce qu'une relation causale singulière ?
  3. Qu'est qu'une loi causale ?
  4. Qu'est-ce qu'une explication causale dans la science ?
  5. Comment formuler le principe de causalité ?
  6. Quels sont les rapports du déterminisme et de l'indéterminisme ?
  7. Comment démontre-t-on l'exactitude d'énoncés singuliers de causalité ?
  8. Comment démontre-t-on l'exactitude d'explication causale ?
  9. Existe-t-il des légalités causales ?
  10. Toutes lois sont-elles causales ?
  11. Le principe de la causalité, dans l'une ou l'autre de ses versions, est-il valable ?

IL faut bien reconnaître que les recherches dans cette direction tendent plutôt à montrer notre incapacité à répondre à ces questions. Nous pensons savoir ce que sont les relations causales, et nous nous en servons tous les jours. Mais lorsqu'il s'agit de le dire, nous ne le savons plus.

Selon l'approche de Mokhtari [Mokhtari97a]Pour un fait *B* donné, la détermination de la cause *A* de ce fait dépend de nos possibilités d'action dans une situation donnée, et est toujours particulière.
Dans un ensemble $A1,A2,\ldots A_n$, de faits tels que, s'ils sont vérifiés, alors *B* est vérifié, et si l'un d'entre eux ne l'est pas, alors *B* ne l'est pas, on a tendance à appeler cause de *B* le terme A_n tel que A_n est le plus facilement 'réalisable' ou 'évitable' par nous, le plus 'accessible' à notre action.

Pourquoi, à première vue, le jet de pierre est-il tenu sans hésitation pour la cause de bris de vitre ? Parce qu'il est *plus facile* de ne pas lancer de pierre sur les vitres que de placer partout des vitres incassables. L'usure extrême des pneus est la cause de leur éclatement parce qu'il est plus facile de changer ses pneus à temps que de construire des routes qui ne provoquent aucun frottement, aucune usure des pneus.

Nous retrouvons dans cette analyse plusieurs idées qui nous sont familières dans le domaine de l'intelligence artificielle, et plus précisément dans la théorie de l'action : la notion de point de vue suivant l'agent ou la notion de contexte, la notion de qualification de l'action et la notion de lois générales avec possibilité d'exception. Ainsi que la notion d'ordre (l'action la plus facile à réaliser) pour le choix d'une cause [Mokhtari97b].

II.2 Etudes logiques

La logique a souvent été un formalisme adopté par les auteurs pour représenter les relations causales. Une phrase causale «*p cause q*» peut être analysée en terme de condition «*si p alors q*» où *p et q* sont respectivement la cause et l'effet.

1. Formaliser cette condition par le connecteur $\supset$ (l'implication matérielle) de la logique classique pose un certain nombre de problèmes [Kleene71]. En effet l'implication matérielle permet de dire qu'une proposition vraie est impliquée par n'importe quelle proposition et de ce fait, de '2+2=4', notre formalisme nous permet d'accepter la phrase : « si la lune est un fromage vert alors 2+2=4 ».Ceci montre les faiblesses de l'implication matérielle pour rendre compte de la causalité. Elle ne capture ni le sens de «preuve» ni le sens de «à partir de». A cela s'ajoutent deux problèmes, deux aspects inhérents à la causalité et qui ne sont pas rendus par la logique classique:

   - **P1 :** Celui de l'ordonnancement temporel entre causes et effets, et
   - **P2 :** Celui de la non-monotonie dans le raisonnement causal.

2. La logique modale a tenté de clarifier la relation entre implication et déductibilité, rappelons que la logique modale est une extension de la logique classique qui s'est intéressée à la possibilité et à la nécessité. Les notions utilisées sont celles de la logique classique augmentée du carré $\Box$ qui se lit «il est nécessaire que». Un opérateur dual, le losange $\Diamond$ qui se lit «il est possible que», est défini comme $\neg\Box$.

   Cette logique s'est beaucoup développée suite à un article de Lewis de 1912 dans lequel il explique les problèmes de l'implication matérielle en faisant remarquer que : des propositions vraies sont toutes impliquées (matériellement, mais non nécessairement) par toute proposition, et n'importe qu'elle proposition fausse implique (matériellement, mais non nécessairement) toute proposition, quelle qu'elle soit.

   C'est pourquoi Lewis a introduit un connecteur spécial symbolisé $\rightarrow$ pour représenter la notion de nécessité, ou comme il l'appela lui-même, l'implication stricte. Gödel 1933 analyse la notion d'implication stricte en termes de nécessité et d'implication matérielle en définissant $A \rightarrow B$ comme étant $\Box(A \supset B)$.

Mais outre l'implication stricte, il y a d'autres types d'implications qui sont intéressantes de signaler, et dont la logique modale constitue une source d'idée, parmi elle nous citons les deux notions suivantes:

3. « L'implication pertinente » ou 'entailment' : l'implication matérielle peut être nécessairement vraie sans que les prémisses soient, de façon visible, en substance ou formellement, pertinentes à la conclusion. Par exemple, une contradiction explicite comme 'il y a un vol de Paris à Londres les dimanches à deux heures et il n'y a pas de vol de Paris à Londres les dimanches à deux heures', implique nécessairement n'importe quelle proposition, comme par exemple « il y' a des vols à moitie prix de paris à Sidney tous les vendredis ». Nous sommes même tentés de considérer cette implication nécessaire comme un cas limite, dégénéré ou impropre, elle ne correspond à aucune connexion intrinsèque entre les prémisses et la conclusion. Les logiques de la 'pertinence' cherchent à saisir un concept d'implication nécessairement pertinente.

4. L'autre ligne d'étude qui semble intéressante pour le lien causal est la logique du *contre-factuel*. Cette logique proposée par Lewis [Lewis73] utilise une variante de l'implication stricte pour représenter la dépendance entre événements. Un événement *e* est causalement dépendant d'un événement e' si «le fait que e a lieu ou pas dépend contre-factuellement du fait que e' a lieu ou pas''. Formellement e dépend causalement de e' si et seulement si :

$$(O(e)\boxed{\ }\rightarrow O(e'))\wedge(\neg O(e)\boxed{\ }\rightarrow\neg O(e'))$$

Où O(e) représente le fait que l'événement e apparaît (occurre) et le symbole$\boxed{\ }\rightarrow$ est le symbole de l'implication contre-factuelle.
Sémantiquement, Lewis considère une relation de similitude comparative entre les mondes possibles de Kripke [Kripke63]. Pour cela, il ajoute la notion de monde possible à la relation d'accessibilité entre ces mondes, la notion de «comment les mondes sont semblables les uns aux autres».
Vu sa clarté et sa précision, cette théorie a influencé un grand nombre d'auteurs (nous citons en particuliers les récents travaux de Halpern et Pearl [Halpern01][Halpern02] [Halpern 05].

II.2.1 Quelques observations sur les relations causales

1. Le concept de condition représenté par l'implication matérielle, ne peut à lui seul rendre compte de la causalité. *Une pluie torrentielle* peut être la cause *d'une inondation*, mais le fait «il *n'y a pas d'inondation »* peut être vu comme un effet, mais en aucun cas il ne peut être vu comme la cause de *l'absence de pluie* (alors que si l'on admet $A\supset B$, on admet aussi la contra-posée $\neg B\supset\neg A$). La relation causale est asymétrique.
2. Ajouter la précédence temporelle à la condition ne suffit pas non plus à rendre compte de la causalité. Dans ce cas, pouvons nous dire que la cause de la mort de monsieur X est le fait qu'il soit né quelques décennies auparavant ?
3. La perception de la causalité dépend des circonstances, du contexte (connaissances du monde, culture, croyances, etc…).
4. De manière générale, l'impossibilité matérielle de fournir à une machine toutes les connaissances possibles sur l'univers et même en réalité, sur un domaine restreint, pose un certain nombre de problèmes : si une allumette s'allume parce que nous l'avons frottée, la présence de matériaux inflammables, d'oxygène dans l'air, etc…, ont également leur rôle à jouer mais nous les omettons par esprit pratique et/ou par ignorance.

Ce problème n'est pas nouveau en IA. McCarthy est l'un des premiers à l'avoir soulevé [McCarthy69] sous le nom '*frame problem'* notons que ce dernier connaît un certain regain d'intérêt actuellement. Nous présentons dans le quatrième chapitre une solution à ce problème proposée par Mokhtari [Moktari97a][Khelfallah01].

II.2.2 Frame problem

Une exigence importante pour plusieurs systèmes intelligents est la capacité de raisonner sur les actions et leurs effets dans le monde et il y a plusieurs difficultés qui émergent :

1. Le frame problem, identifié en premier par Mc Carthy [McCarthy69] ; la difficulté est d'indiquer toutes *les choses qui ne doivent pas changer* quand des actions sont effectuées et que *le temps passe.*
2. Le problème de ramification, ainsi appelé par Finger [Finger87] ; la difficulté ici est qu'il n'est pas raisonnable de stocker explicitement *toutes les choses qui doivent changer* quand des actions sont effectuées et que *le temps passe.*
3. Le troisième problème est appelé le problème de qualification ; la difficulté est que le nombre de *préconditions* pour chacune des actions est immense. McCarthy est l'un des premiers à avoir identifié le problème de qualification dans [McCarty77] avec le problème de la pomme de terre dans le carburateur. Ce dernier problème se résume comme suit :
   Une des préconditions pour que le moteur puisse démarrer est d'avoir la clé de contact mais il y a aussi plusieurs autres conditions. Par exemple, que la batterie soit connectée, que le contact soit en bon état, et qu'il ne peut y avoir une pomme de terre dans le carburateur!...etc, par conséquent il serait difficile de vérifier en pratique toutes les qualifications à chaque fois que nous voulons utiliser la voiture.

Tous ces problèmes nécessitent un raisonnement non-monotone. Pour le problème de la qualification par exemple nous avons besoin de travailler avec des informations partielles et par conséquent, les conclusions de l'inférence peuvent être rétractées (inférence defeasible).

L'inférence defeasible est typiquement non-monotone : à partir du fait qu'une balle est en train de rouler dans une certaine direction, nous inférons qu'elle continuera à le faire (persistance) mais si nous ajoutons le fait qu'il y ait une autre balle qui se dirige vers cette balle nous changeons notre prédiction. C'est pourquoi les logiques classiques ne peuvent être utilisées, telle quelles, pour ce type de résonnement. Des logiques dites non-monotones ont été proposées. De plus, à partir des deux premiers points, nous remarquons que le temps joue un grand rôle dans ces problèmes d'où l'intérêt des logiques non-monotones temporelles.

Cette étude a montré les carences de la logique classique à rendre compte de la causalité, notamment les problèmes de la prise en compte du temps, et de la monotonie. Les philosophes ont donné de bonnes analyses sur la causalité en montrant qu'il peut y avoir plusieurs causes pour un effet donné suivant le contexte ou les circonstances et qu'il y a dans ce cas un choix à faire. Cependant aucune de ces analyses n'a offert un critère raisonnable pour ce choix.

Nous avons vu aussi, qu'en IA le frame problem pose des problèmes essentiels qu'il faut résoudre pour modéliser la causalité. Les logiques non-monotones constituent un bon outil pour cette modélisation [Mokhtari97a]. La partie suivante constitue un survol sur ces dernières.

III. Raisonnement temporel en I.A

Outre son intérêt pour la modélisation de la causalité, le traitement des informations temporelles est également utile en IA, notamment pour la représentation du changement continu. Quel que soit le domaine traité, la première étape consiste toujours à définir une représentation des informations temporelles, et c'est à ce niveau que nous retrouvons les préoccupations communes aux traitements de la causalité.

Historiquement, la prise en compte de l'évolution des phénomènes dans le temps a été proposée dans le cadre du «calcul de situation» introduit par McCarthy et Hayes [McCarthy69]. Dans ce formalisme, étant donné une description complète du monde, pour un moment choisi, un événement particulier opère une transition vers une autre description complète du monde. Il faut noter que les mondes considérés ici sont très restreints, l'exemple type, si cher aux informaticiens, étant *le monde des blocs* (ensemble de cubes posés sur une table ou empilés les uns sur les autres). Par exemple, supposons que la situation initiale décrive un monde dans lequel un bloc A est posé sur un bloc B. Nous pouvons définir une nouvelle situation sous la forme *résultat (situation, action).*

Ce genre d'approche a l'inconvénient de conduire au célèbre «frame problem» soulevé par McCarthy lui-même : comme une situation est une description complète de l'état du monde à un instant donné, il faut pour définir la fonction *résultat* non seulement décrire ce qui a été changé, mais également tout ce qui reste en état. Depuis le premier article de McCarthy, ce problème qui correspond en fait à la possibilité de décrire de façon «raisonnable » ce qui persiste et ce qui change quand des actions sont appliquées, a intéressé de nombreux chercheurs et a donné lieu à une abondante littérature. Si l'approche précédente s'applique plus ou moins bien au monde des blocs (changement discret), il est également souhaitable de pouvoir décrire le changement continu du monde.

Les entités élémentaires associées aux objets temporels sont très variées dans les représentations proposées dans la littérature.

- Les *faits,* qui ou bien ont une valeur de vérité donnée qui ne change pas dans le temps : « la terre tourne », ou bien dont la valeur de vérité change de façon discrète «la maison de nos voisins est rouge ». Si demain, on la peint en blanc, on aura un autre fait. Les faits sont donc des assertions sur l'état du monde à un moment donné.
- *Les événements* qui modifient l'état du monde, où le déroulement interne n'est pas pris en compte, seul le résultat compte «déplacer le bloc de A et B ».
- *Les processus*, qui modifient également l'état du monde, mais pour lesquels on veut analyser le déroulement même : « la cuve est en train de se remplir ».

Un système de représentation du temps doit choisir les objets temporels primitifs qu'il utilise (points ou intervalles).

Un deuxième choix concerne les *relations* entre objets. Dans tous les cas, pour la représentation du temps, une relation fondamentale est celle de la précédence. Lorsque nous avons affaire à des intervalles temporels, d'autres relations se présentent naturellement, par exemple celle de l'inclusion entre deux intervalles ou le chevauchement de deux intervalles.
Enfin, un troisième problème important concerne le choix des propriétés que doi(ven)t satisfaire la (ou les) relation(s). Le temps doit-il être acyclique ? De structure discrète ou continue ? Admet-on des branchements ou le temps est linéaire ?...etc.

Dans une première phase, nous décrivons les propriétés qui caractérisent les logiques non-monotones et rappelons les concepts et principes fondamentaux de la logique classique. Nous survolons ensuite un certain nombre de développements bien connus dans le raisonnement non-monotone comme les logiques modales non-monotones de McDermott [McDermott82a][McDermott80], nous présentons également l'approche proposée par Shoham sur les modèles préférés [Shoham88].

III.1 Propriétés des logiques non-monotones

Nos raisonnements même les plus élémentaires, ne sont pas des raisonnements infaillibles. Nous pouvons par exemple, tenir le raisonnement trivial suivant : 'sachant que la plupart des oiseaux peuvent voler et que Titi est un oiseau, conclure que Titi peut voler' bien qu'acceptable cette inférence ne peut cependant être qualifiée d'absolument correct car elle ne tient pas compte d'exceptions possibles. Elle est donc incertaine et peut être sujette à révision. S'il est précisé que Titi est un manchot et que les manchots sont des oiseaux qui ne volent pas, l'assertion 'Titi peut voler' doit être rétractée. Ce résonnement ne peut se faire en logique classique ; pour le voir, il suffit de rappeler les caractéristiques principales de cette logique.

1. Rappel sur le raisonnement en logique classique

Un système formel de déduction de la logique classique est composé d'un ensemble de schémas d'axiomes et de règles d'inférence. Un tel système permet d'inférer des conclusions à partir de prémisses et définit donc une relation d'inférabilité entre formules, notée $\vdash$.
Cette relation possède les propriétés suivantes [Gabbay84] :

- Réflexivité : si $x \in A$ alors $A \vdash x$;
- Coupure à élément : si $A \vdash y$ et $A \cup \{y\} \vdash x$ alors $A \vdash x$;
- Monotonie à un élément : si $A \vdash x$ alors $A \cup \{y\} \vdash x$.

Où A est un ensemble de formules du langage. Ces propriétés ne sont pas désirables pour un résonnement révisable. En particulier, la propriété de monotonie s'oppose à la formalisation directe de ce type de raisonnement.
D'un point de vu syntaxique, construire un système d'inférence non-monotone nécessite d'affaiblir les propriétés caractérisant les systèmes déductifs de la logique classique. De même la relation de conséquence sémantique notée $\models$ pose problème : en logique classique $A \models q$ si q est vrai dans tous les modèles de A. Comme tous les modèles de $A \wedge r$ sont aussi des modèles de A, il s'ensuit que $A \wedge r \models q$ et donc la logique est monotone.
Construire une logique non-monotone nécessite donc d'un point de vue sémantique de définir une relation de conséquence permettant d'inférer des conclusions qui ne sont pas vérifier dans tous les modèles des prémisses. Les logiques non-monotones se proposent de modéliser le raisonnement révisable qui n'est pas valides au sens classique du terme. Pour ce faire, certaines caractéristiques sont exigées, qui se résume dans la section suivante.

III.2 Caractéristiques et problèmes des logiques non-monotones

- Les conséquences ne sont plus valides par rapport aux prémisses, elles sont simplement consistantes vis-à-vis de celles-ci.
- L'ajout de nouveaux axiomes peut supprimer des théorèmes.
- Il peut y avoir plusieurs ensembles incompatibles de conclusions possibles à partir d'un même ensemble de prémisses initiales : c'est la caractéristique d'extensions multiples. Se pose alors un problème de choix. Accepte-t-on toutes ces extensions ? Ou sinon, laquelle ou lesquelles retient-t-on pour calculer la réponse ? Quels sont les critères de préférence ?

- Comme les formules peuvent être successivement inférées puis rétractées, on dit qu'on a un ensemble de conclusions possibles. Pour cela, on parle souvent d'évaluation des 'points fixes''.
- Un théorème va correspondre dans une logique non-monotone, à une formule qui est présente dans tous les ensembles stables d'assertions pouvant être inférées. On peut aussi s'intéresser à la construction d'un de ces ensembles (une image particulière qu'un agent imagine de façon consistance sur la base d'un ensemble de croyances de départ).
- Enfin un système non-monotone assurant dynamiquement la rétraction doit contenir des règles d'inférence révisable (des règles dont l'application peut être bloquée de manière dynamique). Pour cela on peut concevoir des règles d'inférence munies de conditions d'applications dont la vérification peut évoluer dynamiquement avec l'ensemble des prémisses. Les pré-conditions de ces règles permettent de vérifier, avant inférence, qu'une assertion est consistante avec ce qui a déjà été inféré par le système à partir de l'ensemble actuel de prémisses.

Différentes solutions ont été données pour satisfaire ces caractéristiques et résoudre les problèmes qu'elles engendrent. Ces solutions ont donné lieu à différentes logiques. Dans les sections suivantes nous passons en revue quelques-unes.

Nous allons dans ce qui suit considérer deux approches basées sur les logiques non-monotones appliquées aux théories des changements. La théorie d'anormalité de McCarthy [McCarty86] et de la théorie causale de Shoham [Shoham88]. La première théorie est basée sur la circonscription des anomalies, la seconde théorie est basée sur la préférence chronologique.

III.3 Théories de changements

III.3.1 Théorie sur les anomalies

Pour résoudre le «frame problem», McCarthy propose une théorie sur les anomalies [McCarty86], une extension du calcul situationnel dans laquelle il introduit deux prédicats :

- *t(f, s)* signifiant, le fait f est vrai dans la situation s, et
- ab(f, e, s) signifiant le fait f est anormal à l'égard de l'événement e apparaissant dans l'état s.

Comment est interprété l'exemple du Yale Shooting problem (Y.S.P)[Hanks87] dans cette théorie ? Soient les axiomes de (1) à (4) ci dessous, ayant la signification suivante :

À une situation connue s_0, une personne est vivante (axiome(1)) et le pistolet devient chargé à chaque fois que l'événement charger a lieu (axiome (2)). L'axiome(3) signifie qu'a n'importe quel instant où l'on tire sur une personne avec un pistolet chargé. Cette personne devient non vivante. L'état «être non vivant» par un tir de pistolet chargé et anormal avec le fait rester vivant. L'axiome (4) correspond à l'assertion : les faits normaux persistent à travers l'occurrence d'événements normaux.

Puis on circonscrit les axiomes (1)-(4) sur ab en faisant varier t, on obtient alors toutes les formules vraies dans tous les modèles minimaux dans ab.

Notons par ***A*** les axiomes (1)-(4) et par Circum(A, ab, t) les axiomes circonscrits.

$$t(\text{vivant}, s_0) \qquad (1)$$

$$\forall\ s\ t(\text{chargé}, \text{resulte}(\text{charger}, s)) \quad (2)$$

$$\forall\ s\ t(\text{chargé}, s) \supset ab(\text{vivant}, \text{tirer}, s) \wedge t(\neg\text{vivant}, \text{resulte}(\text{tirer}, s)) \quad (3)$$

$$\forall\ f, e, s\ t(f, s) \wedge \neg ab(f, e, s) \supset t(f, \text{resulte}(e, s)) \quad (4)$$

Considérons maintenant le problème de projection, i.e., quels faits seront vrais dans les situations suivantes :

S_0,

S_1=result(charger,s_0),

S_2=result (attendre,s_1),

S_3=result(tirer,s_2)=result(tirer, result(attendre, result(charger,s0))).

En d'autres termes, on sait initialement que notre individu est vivant, puis le pistolet est chargé, puis en attend un certain temps, puis on tire avec le pistolet.

La projection consiste à déterminer quels faits sont vrais dans la situation s_i,

Attendre a pour signification : pendant une période de temps il n' y a rien d'intéressant qui arrive i.e., chaque fait vrai avant attendre, devrais rester vrai après.

Une interprétation des axiomes (1)-(4) est donnée par la figure suivante cette première figure représente les faits vrais dans tous les modèles de A i.e., ce que nous pouvons déduire si nous ne circonscrivons pas. A partir des axiomes (1)-(2) nous pouvons faire les déductions suivantes : t(vivant, s_0), et t(chargé, s_1).

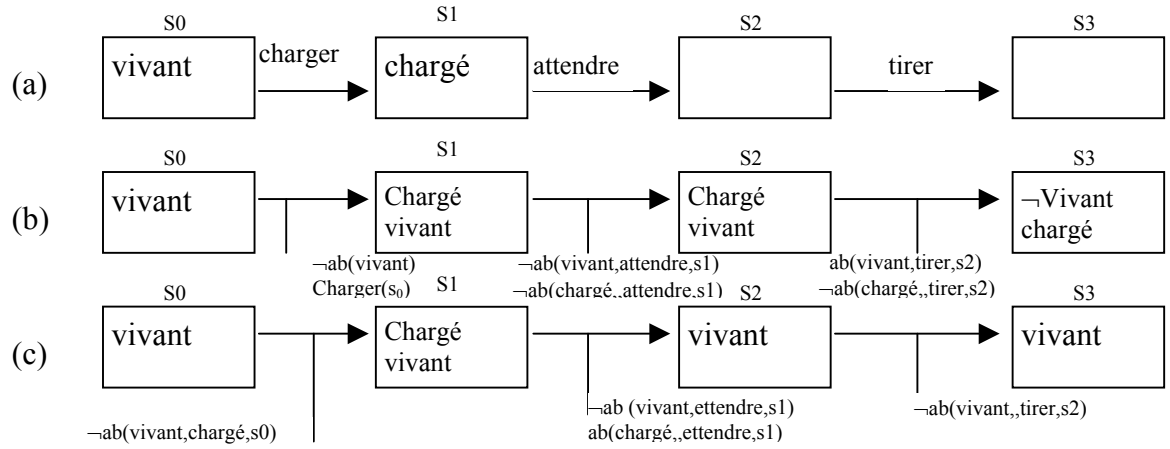


Figure 1: Trois modèles des axiomes (1)-(4)

Mais nous ne pouvons rien déduire sur ce qui est vrai dans s2 ou dans s3, de même, nous ne pouvons déduire ni des anomalies ni leurs négations : le fait vivant ne persiste pas parce que nous ne pouvons pas déduire qu'il est 'non ab' par rapport au chargement du pistolet, et le fait que le pistolet est chargé ne persiste pas parce que nous ne pouvons pas déduire qu'il est 'non ab' par rapport à l'attente. Intuitivement, McCarthy raisonne sur la «minimisation des anomalies» i.e., nous savons que vivant doit être vrai en s_0 et rien ne nous oblige à croire ab(vivant, charger, s_0), donc nous supposons sa négation.

A partir de l'axiome 3, nous déduisons t(vivant, attendre, s_1). Comme rien ne nous empêche de faire les suppositions : ¬ab(vivant, attendre, s_1) et ¬ab(chargé, attendre, s_1) nous déduisons donc t(vivant,s_2) et t(chargé,s_2).

A partir de l'axiome (2) nous pouvons déduire ab(vivant,attendre,s2) nous ne pouvons donc pas supposer sa négation. Par contre, nous pouvons supposer ab(chargé,tirer,s2) et dans ce cas, déduire t(vivant,s3).

Ce raisonnement est schématisé par la figure (b). Nous pouvons facilement vérifier que cette interprétation est un modèle de A i.e., les axiomes (1)-(4) sont satisfaits. Nous pouvons même dire que ce modèle est minimal dans ab (ce modèle devrait avoir une extension vide pour ab, ce qui n'est pas possible). Pour le voir, considérons t(vivant,s2) qui peut être vrai ou faux dans un modèle quelconque de A. S'il est vrai, nous pouvons immédiatement déduire une anomalie, à partir de l'axiome (3). S'il est faux, alors soit ab(vivant,charger,s0) soit ab(vivant, attendre,s1) devrait être vrai.

Dans l'un ou l'autre des cas, on doit avoir au moins une anomalie.

La question intéressante est de savoir si le modèle de la figure (b), le modèle attendu, est l'unique modèle minimal ou plus précisément, si t(¬vivant,s3) et t(chargé,s3) sont vrais dans tous les modèles minimaux.

Considérons la situation de la figure (c). Elle décrit un état de fait dans lequel le pistolet cesse d'être chargé comme un résultat de l'attente. L'individu n'est pas mort après le tir, puisque le pistolet n'est pas chargé. Bien sûr, cet état contredit directement l'intuition établie, i.e., toutes choses devrait être '*non ab'* si rien d'explicitement 'ab' n'a lieu pendant l'attente. Est-ce-que cette interprétation décrit un modèle minimal ?

Rappelons d'abord qu'il ne peut y avoir de modèles ayant des extensions de ab nulles.

Ainsi si cette interprétation est un modèle, mais en commençant par s_3 pour repartir vers l'arrière. Notons que t(vivant,s_3) est vrai, puis considérons ce que doit être vrai en s_2 et dans les situations précédentes.

La première décision d'anomalie à faire est si ab(vivant,tirer,s_2) est vrai. Comme nous n'avons pas pris de décision pour le contraire, nous supposons sa négation. Mais alors, à partir de la contraposée de l'axiome (3) nous pouvons déduire $\neg$t(chargé,s_2). Or si cela est le cas, et que t(chargé,s_1) doit aussi être vrai, nous pouvons déduire de l'axiome (4) que ab(chargé, attendre, s1) est vrai. La suite de la figure (c) est évidente, puisque nous pouvons supposer que vivant est 'non ab' selon charger et attendre et donc déduire que vivant est vrai dans s_1 et s_2.

Est ce qu'il y a d'autres modèles à partir de cette théorie de la circonscription des anormalités ? Il est facile de voir que les deux modèles présentés, sont les seuls modèles minimaux dans ab.

Nous pouvons déduire que vivant et chargé sont vrais dans s_1, que vivant est vrai dans s_2, mais ne nous pouvons rien dire sur ce qui est vrai en s_3 sauf des formules de type t(vivant,s_3)$\vee$ t($\neg$vivant,s_3). La circonscription est donc beaucoup plus faible qu'on ne l'espérait (voir aussi des résultats paradoxaux de la circonscription dans [Moinard88]).

III.3.2 Les modèles préférés de Shoham

Dans [Shoham88] Shoham suggère un mécanisme de raisonnement consistant à représenter la causalité dans un cadre temporel. L'interprétation de la représentation s'effectue selon les règles d'ignorance chronologique et la théorie est une théorie causale prédisant le futur. Shoham propose une structure de lignes de temps et définit la notion d'action comme la capacité de faire certains choix parmi l'ensemble des lignes de temps. Cette structure est assez générale, avec des branchements dans les deux sens (futur et passé).

Le formalisme utilisé est une logique non-monotone appelée logique d'ignorance chronologique (CI) qui consiste à associer à la logique modale standards une sémantique basée sur une relation de préférence entre les modèles.

Une formule dans le langage est de la forme □True(t1,t2,p) signifiant : on sait que la proposition p est vraie dans l'intervalle de temps(t1,t2) où t2 est par convention le dernier point dans le temps (l.t.p. ou last time point). Dans ce qui suit □(t1,t2,p) est une abréviation de □True(t1,t2,p).

Si A est une formule, une interprétation M satisfait préférentiellement A notée $M \models_{\sqsubset} A$ si $M \models A$, et s'il n' y a pas une autre interprétation M' telle que $M' \sqsubset M$ et $M' \not\models A$ ($\sqsubset$ est un ordre partiel strict sur les interprétations pour le langage). Dans ce cas nous disons que M est un modèle préféré de A.

L'ordre partiel sur les modèles établit simplement le fait que '*les modèles, dans lesquels quelqu'un sait le moins possible le plus longtemps possible*' sont préférés. Ces modèles sont appelés '*les modèles les plus ignorants chronologiquement*'' ou *c.m.i* models (chronologically maximally ignorant models). Dans CI, A implique B ssi B est vrai dans tous les cmi modèles de A.

Une théorie causale Ψ est une théorie dans laquelle toutes les formules ont la forme :

$$\Phi \wedge \Theta \supset \Box\varphi$$

1. φ est une formule atomique (positive ou négative) True(t1,t2, [¬] p) avec par convention t1<t2 et □est l'opérateur modal de nécessité,
2. Φ est une conjonction de formules $\Box\varphi_j$ où φ_j est une formule atomique de base (positive ou négative) dont le l.t.p t_i tel que $t_i<t1$,
3. Θ est une conjonction de formules $\Diamond\varphi_j$, où φ_j est une formule atomique de base (positive ou négative) dont le l.t.p t_j tel que $t_j<t1$ et ou ◊ est l'opérateur modal de possibilité.
4. Φ ou Θ ou les deux peuvent être vides. Une formule ou Φ est vide est appelée condition limite. Les autres formules sont appelées *règles causales,*
5. il y a un point de temps t0 tel que si $\Theta \supset \Box(t_1,t_2,[\neg]p)$ est une condition limite, alors $t_0<t_1$.
6. il ne doit pas exister deux formules dans telle que l'une contient ◊(t1,t2,p) à sa gauche et une autre contient ◊(t1,t2,¬p) à sa gauche pour tout p,t_1,t_2.
7. si $\Phi_1 \wedge \Theta_1 \supset \Box$ (t1,t2,p) et $\Phi_2 \wedge \Theta_2 \supset \Box$ (t1,t2,¬p) sont des formules de Ψ, alors $\Phi_1 \wedge \Theta_1 \wedge \Phi_2 \wedge \Theta_2$ est inconsistante.

Pour illustrer cette théorie, considérons le scénario de tir suivant (le 'Yale Shooting problem'' modifié)

1. □(1,1,pistolet chargé)
2. □ (5,5,tirer)
3. □ (t,t,chargé)∧◊(t,t, ¬tirer) ∧◊(t,t, déchargé manuellement)⊃□(t+1,t+1,chargé), pour tout t.
4. □ (t,t,chargé)∧□(t,t,tirer)∧◊(t,t,air)∧…◊…autres conditions⊃□(t+1,t+1,bruit), pour tout t.

Le troisième schéma d'axiome correspond au 'frame axiom' 'la persistance' dont on a besoin pour assurer que le pistolet resta chargé jusqu'au temps t=5. Shoham propose une autre solution à ce problème de persistance qu'il nomme 'extended prediction problem'. Le formalisme logique proposé est le même que celui de la qualification : la logique de l'ignorance chronologique.

L'idée clé est de capturer la propriété d'extension maximale des histoires potentielles, par la logique de l'ignorance chronologique qui *préfère* retarder les connaissances des formules de bases aussi tard que possible (le l.t.p.maximal). Pour cela une histoire potentielle est représentée par une formule atomique dont le second argument est quantifié.

L'histoire du chargement, par exemple, est représentée par :

$$\exists v(\ t_1 \leq v \leq t_2 \wedge \Box(t_1, v, \text{chargé}))$$

Cette formule signifie que le fait *chargé* doit rester vrai *aussi longtemps que possible* à partir du point de temps t_1, mais au-delà de t_2 il faut de nouvelles informations.

Avec la théorie d'inertie, Shoham propose la modification suivante de l'exemple précédent (dans ce qui suit, le préfixe p spécifie que l'on a affaire à une proposition dénotant une histoire potentielle).

1. $(\exists v\ (\Box(t_1, v, \text{p-chargé}) \wedge t \leq v)) \supset \Box(t, v, \text{chargé})))$, pour tout t_1 et t tel que $t_1 \leq t$.
2. $\exists v\ (1 \leq v \leq \infty \wedge(\neg\exists v'\ 1 \leq v' \leq v \wedge \Box(v', \text{tirer})) \wedge(\neg\exists\ v'\ 1 \leq v' \leq v \wedge \Box(v', \text{déchargé manuellement})) \wedge \Box(1, v, \text{p-chargé}))$.
3. $\Box$ (5, tirer).
4. $\Box(t, \text{chargé}) \wedge \Box(t, \text{tirer}) \wedge \Diamond(t, \text{air}) \wedge \ldots \Diamond \ldots$autres conditions $\supset \Box$ (t+1, bruit), pour tout t.

Puis Shoham démontre que cette nouvelle théorie possède également un unique cmi modèle. Baker montre dans [Baker89] que bien que l'approche de Shoham donne, intuitivement une réponse correcte au Y.S.P., son application semble se restreindre à ce que McDermott et Hanks [Hanks87] appellent « problèmes de projection temporelle » i.e., les problèmes dans lesquels étant données les conditions initiales, il faut prédire ce qui va arriver dans le futur. Notons néanmoins, que bien que l'approche de Shoham donne une réponse correcte au Y.S.P., si on veut considérer les explications temporelles, i.e., les problèmes nécessitant un retour dans le temps, l'approche de Shoham n'est plus adéquate.

IV. Conclusion

Nous avons présenté dans ce chapitre une étude sur la causalité : philosophique et logique. Rappelons que le raisonnement causal est un raisonnement temporel non-monotone et que les logiques classiques ne pouvaient pas en rendre compte. Nous avons présenté les caractéristiques principales des logiques non-monotones.

Plusieurs approches ont été proposées pour un raisonnement temporel non-monotone, un survol a été fait pour représenter quelques-unes telles que la théorie d'anomalie de McCarthy et la théorie causale de McDermott et de Shoham basée sur la préférence chronologique [Shoham88][Shoham89].

La logique n'est pas la seule représentation utilisée pour ce type de problème, les représentations graphiques telles que les réseaux sémantiques et… ont été utilisés sans grands sucées. Néanmoins, une représentation graphique proposée par Pearl en 1988 a eu ces derniers temps un grand regain d'intérêt. La représentation graphique la plus intuitive de l'influence d'un événement, d'un fait, ou d'une variable sur une autre, est probablement de relier la cause à l'effet par une flèche orientée. Dans le chapitre suivant, nous allons voir un formalisme, qui est graphique, pour la modélisation de la causalité, c'est le cadre des réseaux causaux bayésiens.

I. Introduction

L'information n'est pas la connaissance. A mesure que se développent les technologies permettant de stocker, d'échanger de l'information et d'y accéder, la question de l'analyse et de la synthèse de ces informations devient essentielle. Deux types d'approches connaissent donc tout naturellement un intérêt croissant. Les méthodes statistiques, parce qu'elles sont précisément conçues pour permettre le passage de l'observation à la loi, fut-elle loi de probabilité. Les technologies d'IA ensuite, parce que leur vocation est de permettre aux ordinateurs de traiter de la connaissance plutôt que de l'information.
Les réseaux bayésiens sont le résultat d'une convergence entre ces deux disciplines et constituent aujourd'hui l'un des formalismes le plus complet et le plus cohérent pour l'acquisition, la représentation et l'utilisation de connaissances par des ordinateurs [Balmisse02].
Dans le cadre de la représentation de connaissances, les réseaux bayésiens constituent une approche possible pour intégrer l'incertitude dans le raisonnement, 'prise en compte des faits *précis, mais incertains*'. Pour notre problème sur la causalité les liens causaux sont souvent représentés graphiquement par des flèches reliant la cause à l'effet, les réseaux causaux bayésiens sont des modèles graphique pour la représentation de la relation de causalité [Dubois03]. Dans ce chapitre, nous faisons le point sur les principaux aspects de cette approche; puis une formulation complète des réseaux bayésiens pour leur utilisation comme modèles d'inférence. Nous terminons par un résumé sur les avantages et les limites des réseaux bayésiens.

II Fondement nécessaire pour les Réseaux bayésiens

II.1 L'incertitude

Nous avons vu qu'il est difficile de déterminer de manière exacte la véracité (ou non) d'un fait. Par exemple, si l'on pose la question « Pleuvra-t-il demain ? », il sera certainement difficile de répondre par oui ou par non. En effet, selon la météo d'aujourd'hui, selon l'intensité de la douleur de vos rhumatismes..., on pourra en déduire qu'il pleuvra « peut-être », « sûrement », « certainement pas »... Selon Pearl inconsciemment, on associera une probabilité à l'événement « Il pleuvra demain » qui se traduira par un des adverbes cités ci-dessus. Mais on peut sans aucun doute annoncer qu'il est quasiment impossible d'affirmer cela à 0 ou à 100 %. Cette petite introduction montre bien qu'en fait nous vivons dans un monde incertain. Puis-je être sûr à 100 % de ne pas vivre jusqu'à 100 ans ? On le voit dans ces questions, il est impossible de donner l'exacte certitude d'un événement, en revanche, on peut la quantifier, notamment grâce aux probabilités. Ainsi, plutôt que de raisonner avec la véracité ou non d'une proposition, on peut raisonner avec la confiance que l'on accorde à une proposition ou à la réalisation d'un événement. Cette confiance se traduira pour certains chercheurs par l'attribution d'une probabilité.
Ainsi, pour une proposition A donnée, (exemple : « Il pleuvra demain »), on peut lui associer une probabilité P(A), telle que P(A) soit comprise entre 0 et 1. Si A est vraie, alors P(A)=1, et si A est fausse alors P(A)=0. La proposition A sera soit vraie, soit fausse, mais P(A) représentera notre degré de confiance dans le fait que A soit vraie ou fausse.

Exemples :

P(Temps = Ensoleillé) =0,7 signifie que l'on pense qu'il existe 70 % de chances que le temps soit ensoleillé. Dans ce cas, le temps est une variable aléatoire qui peut prendre des valeurs définies telles que « Ensoleillé », « Pluvieux », « Neigeux » ou « Nuageux ».

II.2 Une représentation graphique de la causalité

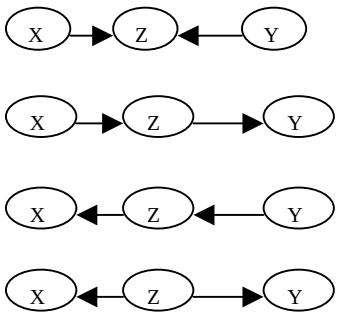
Les réseaux probabilistes peuvent être vus comme un langage de haut niveau utilisé pour décrire des relations de *dépendance* ou d'*indépendance* entre des *variables aléatoires* tout en ignorant les détails numériques ou fonctionnels spécifiques. Ils ont été développés et utilisés par de nombreuses communautés, dont la communauté des statisticiens, celle des chercheurs en IA. Aujourd'hui les réseaux probabilistes sont devenus un des outils incontournables pour qui s'intéresse aux problèmes posés par l'IA au sens large [Becker99].
Globalement on peut dire que les réseaux probabilistes ont pour caractéristique principale de spécifier une *distribution de probabilité*. Dotés de cette sémantique précise, ils peuvent être utilisés à des fins de diagnostic, d'apprentissage, d'explication, de contrôle et bien d'autres tâches nécessaires en IA [Balmisse02].

Les graphes causaux sont un mode intéressant de représentation d'une structure d'influence entre divers *faits*, *états* ou *hypothèses d'état* sur l'environnement. Lorsque *une durée et/ou une date de transition* est associée à chaque arc du graphe, un graphe *causal* permet de représenter une évolution de l'environnement : il s'agit de déduire de proche en proche quel est l'enchaînement des états successifs de l'environnement compte tenu du modèle d'évolution décrit par le graphe causal. Dans cette représentation, on peut réviser une représentation courante du monde pour l'adapter au cours du temps en prenant en compte au fur et à mesure des informations qui sont fournies en entrée ou qui sont issues de modèles de comportement ou d'évolution codés dans une base de connaissances temporelles [Fabiani96][Populaire00].

Les modèles graphiques sont le mariage entre la théorie des probabilités et celles des graphes. Ils fournissent des outils intuitifs et naturels pour traiter des problèmes dans lesquelles l'incertitude et la complexité des données joue un rôle important. L'idée fondamentale des modèles graphiques est la modularité: un système complexe est construit en combinant des parties plus simples. La théorie des probabilités combine alors ces parties tout en assurant une cohérence à l'ensemble du système. La théorie des graphes apporte une interface intuitive pour la modélisation des connaissances, permettant de structurer facilement les données[Becker99]. La représentation graphique la plus intuitive de l'influence d'un événement, d'un fait, ou d'une variable sur une autre, est probablement de relier la cause à l'effet par une flèche orientée.
Avant d'aller plus loin, il est important de comprendre que, bien que la flèche soit orientée, elle peut cependant fonctionner dans les deux sens, et ce même si la règle causale est stricte.

II.2.1 Circulation de l'information dans un graphe causal

Les chemins que peut prendre l'information à l'intérieur d'un graphe sont les cas suivants :



Connexion convergente (X et Y causent Z)

Connexion en série (X cause Z, Z cause Y).

Ou le cas symétrique.

Connexion divergente (Z cause X et Y).

Figure 1 : Circulation de l'information

Pour chacun de ces cas, nous allons essayer de déterminer si l'information peut circuler de X à Y.

Il existe trois cas qui décrivent l'ensemble des situations possibles faisant intervenir trois événements que nous résumons dans le tableau suivant.

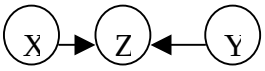
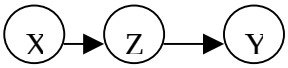
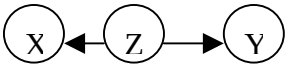
Graphe	propriété	Exemple
 *Connexion Convergente*	L'information ne peut circuler de X à Y que si Z est connu.	X : tremblement de terre Y : cambriolage Z : alarme.
 *Connexion en Série 'Linéaire'*	L'information ne peut circuler de X à Y que si Z n'est pas connu.	X : Il pleut Y : la chaussée est humide Z :la chaussée est glissante
 *Connexion Divergente*	L'information ne peut circuler de X à Y que si Z n'est pas connu	X : la pelouse de mon jardin est humide. Y : la pelouse de mon voisin est humide. Z : il a plu cette nuit.

Tableau 1 : Exemples de circulation de l'information.

Nous savons maintenant exactement dans quelles conditions une information peut circuler à l'intérieur d'un graphe. On voit qu'il ne s'agit pas de suivre le sens des flèches !
Supposons que nous disposions d'un graphe relativement complexe, pour lequel nous disposions déjà d'un certain nombre d'informations (i.e., certaines variables sont déjà connues). Si nous apprenons maintenant une autre information, devons nous réviser notre opinion sur l'ensemble des autres nœuds de ce graphe ?
Pour répondre à cette question, nous pouvons essayer de synthétiser l'étude des circuits d'informations ci-dessus en une règle appelée *d-séparation*, qui décrit dans quelles conditions l'information entre un nœud X et un nœud Y est bloquée.

II.2.2. Notion de D-Séparation

La notion de d-séparation est très importante dans l'étude des réseaux de causalités, elle permet de décrire dans quelles conditions l'information est traitée localement, sans perturber l'ensemble de graphe.
On dira que X et Y sont *d-séparés* par Z si pour tous les chemins entre X et Y, l'une au moins des deux conditions suivantes est vérifiée :

- Le chemin converge en un nœud W, tel que W≠Z, et W n'est pas une cause directe de Z.
- Le chemin passe par Z, et est soit divergent, soit en série au nœud Z.

X est **d-séparé** de Y par Z est notée : **<X|Z|Y>**.

Cette définition veut dire que si Z est la seule information connue dans le graphe et si X et Y qui sont d-séparés par Z. Une nouvelle information sur X ne modifie en rien mon opinion sur Y. Cette notion permet donc de préciser dans quelles conditions une information peut être traitée localement, sans percuter l'ensemble du graphe.

III. Les Réseaux Bayésiens

Un réseau bayésien est un modèle graphique dans lequel les connaissances sont représentées sous forme de variable. Chaque variable est un nœud du graphe et prend ses valeurs dans un ensemble discret ou continu. Le graphe est toujours dirigé et acyclique, il ne contient pas de boucle. Les arcs dirigés représentent un lien de dépendance directe (la plupart du temps il s'agit de causalité). Ainsi un arc allant de A à B exprimera le fait que B dépend directement de A. L'absence d'arc ne renseigne alors que sur la non-existence d'une dépendance directe. Les paramètres expriment le poids donné à ces relations et sont les probabilités conditionnelles des variables sachant leurs parents (exemple: p(B|A)) ou les probabilités a priori si la variable n'a pas de parents. Un réseau bayésien est donc un graphe causal auquel on a associé une représentation probabiliste sous-jacente ; Cette représentation, permet de rendre quantitatifs les raisonnements sur les causalités que l'on peut faire à l'intérieur du graphe [Becker99] [Geslin03][Balmisse02].

III.1 Définition

Un réseau bayésien est défini par :

- Un graphe acyclique orienté G, G=(V,E), où V est l'ensemble des nœuds de G, et E l'ensemble des arcs de G ;
- Une épreuve ε à laquelle est associée un espace probabilisé fini (Ω,Z,p), et n variables aléatoires $(Xi)_{1\leq i\leq n}$;

Ω : l'ensemble des événements (univers).

Z : une tribu de Ω.

P : mesure de probabilité, $\forall A \in Z$, $0\leq P(A)\leq 1$, et $P(\Omega)=1$.

G et ε définissent un réseau bayésien, que l'on note alors B=(G,P), si et seulement si :

- Il existe une bijection entre les nœuds du graphe G et les variables (Xi),
- La propriété suivante, appelée propriété de factorisation, est vérifiée : $\mathbf{P(X1, X2,..,Xn)=\Pi_{I=1,n}P(Xi|C(Xi))}$ où C(Xi) est l'ensemble des causes (parents) de Xi dans le graphe Ω.

III.1.1 Exemple explicatif

La meilleure façon de comprendre les réseaux bayésiens est d'essayer de modéliser une situation dans laquelle la causalité joue un rôle, mais où notre compréhension de ce qui se passe est incomplète, et telle que l'on ait besoin de décrire les faits de manière probabiliste.

Considérons la situation suivante. Un jardinier s'aperçoit que son plus beau pommier perd ses feuilles. On sait que ce phénomène peut-être dû à deux raisons : soit c'est l'automne, soit l'arbre est malade. Cette situation peut être décrite par la figure 1.

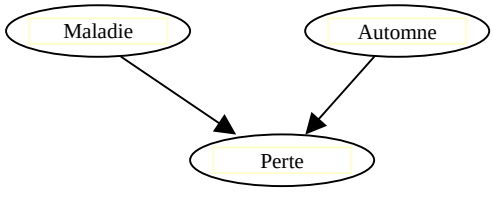


Figure 2 : Graphe du réseau bayésiens.

On est en présence de 3 nœuds : « Maladie » représente l'état de santé de l'arbre, « Automne» représente le fait que l'on soit ou non en automne, et « Perte des feuilles » représente le fait que l'arbre perde, ou non, ses feuilles. Ces 3 nœuds peuvent prendre 2 états différents : 'oui' ou 'non'. Les flèches représentent les relations causales « Si l'arbre est malade, alors il perd ses feuilles » et « Si on est en automne, alors l'arbre perd ses feuilles ».

Le modèle que nous avons défini précédemment est un modèle causal. Un réseau bayésien est en quelque sorte un réseau causal, auquel on ajoute des éléments de probabilités issus de la théorie bayésienne. Ainsi, pour un réseau bayésien, on ne dira pas « Si l'arbre est malade, alors il perd ses feuilles », mais « Si l'arbre est malade, alors il y a une probabilité $x \in [0:1]$ pour qu'il perde ses feuilles ». En effet, ce qui est important avec les réseaux bayésiens, c'est le fait que les relations causales ne sont pas absolues, mais sont associées à une probabilité, indiquant le degré de croyance que l'on a dans l'événement.

Chaque nœud du réseau représente une variable. Et chaque flèche représente les effets causaux des variables auxquelles elles sont rattachées. Ainsi, nous avons devant les yeux la structure du réseau bayésien. Ce qu'il nous faut maintenant, c'est de déterminer toutes les probabilités qui peuvent se rattacher aux nœuds. Pour cela, il faut donner les probabilités à priori de tous les nœuds «racines », qui sont des nœuds sans parents. Et il faut également donner les probabilités conditionnelles des nœuds « enfants » en fonction de tous les états possibles de leurs parents. Ce qui va nous donner, pour notre exemple, les probabilités suivantes.

Maladie= 'Oui'	Maladie= 'Non'
0.1	0.9

Tableau 2 : Probabilité de Maladie

Automne= 'Oui'	Automne= 'Non'
0.25	0.75

Tableau 3 : Probabilité de Automne

	Automne= 'Oui'		Automne = 'Non'	
	Maladie= 'Oui'	Maladie= 'Non'	Maladie= 'Oui'	Maladie= 'Non'
Perte= 'Oui'	1	1	0..9	0.05
Perte= 'Non'	0	0	0.1	0..95

Tableau 4 : Probabilité de Perte sachant Maladie et Automne.

Les réseaux bayésiens ont pour objectif d'acquérir, représenter et utiliser la connaissance. Ils sont constitués de deux composantes :

1. Graphe causal

Ce graphe est orienté et acyclique. Ses nœuds sont des variables d'intérêt du domaine et les arcs des relations de dépendance entre ces variables. L'ensemble des nœuds et des arcs forme ce que l'on appelle la structure du réseau bayésien. C'est la représentation qualitative de la connaissance.

2. Distributions locales de probabilité

L'ensemble des distributions de probabilité sont les paramètres du réseau. Pour chaque nœud on dispose d'une table de probabilité P(variable/parents(variable)) qui représente la distribution locale de probabilité. Il faut remarquer que chaque nœud ne dépend que de l'état de ses parents. Il s'agit de la représentation quantitative de la connaissance.

Par ailleurs, toute variable doit avoir au moins soit un parent soit un enfant. Les variables peuvent être discrètes (ce qui est le cas le plus fréquent) ou continues. A chaque variable discrète est associée une distribution de probabilité complète. Les variables continues obéissent à une distribution théorique dont les paramètres (moyenne, variance) sont donnés. Les distributions de probabilités nécessaires pour initialiser les modèles sont de deux types : marginales pour les variables qui n'ont pas de parent, conditionnelles aux valeurs de leurs parents pour les autres. Le modèle a pour fonction d'une part, de représenter les liens de causalité qui lient des phénomènes les uns aux autres et, d'autre part, d'utiliser l'information relative à l'état d'une variable quelconque A pour connaître celui d'une autre variable B du modèle, quel que soit par ailleurs le lien entre A et B, et cela en utilisant exclusivement les axiomes et théorèmes de la théorie des probabilités.

Pratiquement, un réseau bayésien est d'abord initialisé avec les valeurs *a priori* des différentes variables qui le composent. Ces valeurs proviennent soit d'observations historiques, soit d'une quelconque autre connaissance *a prior*i.

Une variable peut être dans l'un des trois états suivants :

- une distribution *a priori* provenant d'observations antérieures ou d'informations *a priori*
- une distribution *a posteriori* suite à l'introduction d'informations concernant les autres variables du modèle ;
- une certitude lorsque de nouvelles observations portant directement sur une variable qui permettent de connaître son état avec certitude. Celle-ci est introduite dans le modèle en remplaçant la distribution de probabilité *a priori* par la valeur 1 pour l'état observé et 0 pour les autres états.

En règle générale, les variables dont on cherche à connaître la distribution *a posteriori* sont des variables inobservables pour lesquelles il n'est jamais possible d'obtenir de certitude basée sur l'observation. Mais ces variables inobservables (dites latentes) sont reliées à d'autres pour lesquelles, en revanche, des observations sont possibles. Lorsqu'une variable inobservable est parente de l'une ou l'autre variable observable, celle-ci constitue un indicateur de celle-là. Connaissant la distribution de probabilité *a priori* de l'indicateur conditionnellement à la variable latente, toute observation sur l'indicateur peut être utilisée pour réviser les croyances relativement à la variable latente. On constate donc que le modèle du réseau bayésien offre une définition rigoureuse de la notion d'indicateur et une axiomatique qui permet de conceptualiser la relation entre la variable latente et ses indicateurs.

La contribution propre de la représentation sous forme de graphe au modèle par rapport à la théorie des probabilités est de fournir une représentation des relations d'indépendance conditionnelle entre les différentes variables (exprimée par la propriété dite de D-Séparation).

III.2 D-Séparation

A fin de modéliser cette notion nous présentons quelques notations et définitions de base.

- *Observation* : On appelle *observation* un événement du type : $f=(Xi=x^{ki}_{i})$, 'la variable Xi prend la valeur $x^{ki}_{i'}$.
- *Fait* : on appelle *fait* une conjonction d'observations, c'est à dire un événement de type $e=\cap_{i\in Ie}(X_i=x_i^{ki})$. Chacun des termes $Xi= x_i^{ki}$ est appelé *fait individuel* de *e*. 'Ie : l'ensemble des indices des variables aléatoires utilisées dans ce fait'.

- *Lieu d'un fait* : si $e=\cap_{i\in Ie}(X_i=x_i^{ki})$ est un fait, on appelle *lieu du fait e*, l'ensemble des variables aléatoires utilisées dans ce fait $E=\{Xi|(i\in Ie)\}$.
- *Variables observées* : si e est un fait de lieu E, les éléments de E s'appellent variables observées.
- **Chemin d-séparé par un nœud** : soient (X,Y,Z) des nœuds du graphe G=(V,A), et soit S un chemin entre les nœuds X et Y. on dit que le chemin S est d-séparé par Z, si au moins l'une des deux conditions est vérifiée :
  - Le chemin converge en un nœud W tel $W\neq Z$, et $W \notin C^+(Z)$, ($C^+(Z)$ c'est l'ensemble des ascendants du nœud Z).
  - Le chemin passe par Z, et il est soit divergent, soit en série au nœud Z.
- **Chemin d-séparé par un ensemble de nœuds** : soient (X,Y) deux nœuds du graphe G=(V,A), et soit **Z** un ensemble de nœuds de ce graphe. Soit S un chemin entre X et Y. on dit que S est d-séparé par **Z** si au moins l'une des conditions est vérifiée :
  - Le chemin converge en un nœud W tel $W\in$**Z**, et$\forall Z\in$**Z**, $W \notin C^+(Z)$
  - Le chemin passe par $Z\in$**Z**, et il est soit divergent, soit en série au nœud.
- **Nœuds d-séparé par un ensemble de nœuds** : soient (X,Y) deux nœuds du graphe G=(V,A), et soit **Z** un ensemble de nœuds de ce graphe. On dit que X et Y sont d-séparés par **Z** si tous les chemins entre X et Y sont d-séparés par **Z.**
- **D-séparation (ensembles) :**soient **(X,Y,Z)** trois ensembles de nœuds du graphe G=(V,A), on dira que **X** et **Y** sont d-séparé par **Z** si tous les éléments de **X** sont d-séparés par **Z** de tous les éléments de **Y**.

III.3 Indépendance conditionnelle

- *Indépendance conditionnelle*

Soit ε une épreuve et (Ω, Z,P) l'espace probabilisé associé. Soient X,Y, et Z trois vecteurs de variables aléatoires discrètes associées à ε.

On dit que X et Y sont *indépendants conditionnellement* à Z, et on note $\mathbf{X \perp Y|Z}$, si l'une des propriétés équivalentes suivantes sont vérifiées :

Ind1 $P(X|Z,Y)=P(X|Z)$

Ind2 $P(X,Y|Z)=P(X|Z)\cdot P(Y|Z)$.

- *Propriétés de l'indépendance conditionnelle*

Soit ε une épreuve et (Ω, Z,P) l'espace probabilisé associé. Soient X,Y,Z et W quatre vecteurs de variables aléatoires discrètes associées à ε. Les quatre propriétés suivantes sont vérifiées :

Symétrie	$X \perp Y\|Z \Leftrightarrow Y \perp X\|Z$.
Décomposition	$X \perp (Y\cup W)\|Z \Rightarrow X \perp Y\|Z$ et $X \perp W\|Z$.
Union faible	$X \perp (Y\cup W)\|Z \Rightarrow X \perp W\|(Z\cup Y)$.
Contraction	$X \perp Y\|Z \wedge X \perp W\|(Z\cup Y) \Rightarrow X \perp (Y\cup W)\|Z$.

Théorème fondamental [Pearl00]

Soit B=(G,P), un réseau bayésien. Soient X⊂V, Y⊂V et Z⊂V, trois sous-ensembles de nœuds.

Si X et Y sont **d-séparés** dans G par Z, **alors** X et Y sont indépendants conditionnellement à Z, i.e., **X ⊥ Y|Z**.

Ce résultat, démontré par Verma et Pearl en 1988, constitue le théorème fondamental des réseaux bayésiens, est très important, car il permet de limiter les calculs de probabilités grâce aux propriétés du graphe.
L'utilisation essentielle des réseaux bayésiens est donc de calculer des probabilités conditionnelles d'événements reliés les uns aux autres par des relations de cause à effet. Cette utilisation s'appelle Inférence.

IV. INFERENCE

Tout calcul portant sur la distribution de probabilité associée à un réseau bayésien relève de l'inférence. Certains types de calcul ont traditionnellement une plus grande importance, parce qu'ils peuvent correspondre à des utilisations pratiques. D'un point de vue intuitif, l'inférence dans un réseau de causalité consiste à propager une ou plusieurs informations certaines au sein de ce réseau, pour déduire comment sont modifiées les croyances concernant les autres nœuds. Le problème de l'inférence est uniquement un problème de calculs. Il n'y a aucun problème théorique ; en effet, la distribution de probabilité étant entièrement définie, on peut (en principe) tout calculer. Ces méthodes de calculs sont plus au moins complexes suivant la complexité du graphe, c'est à dire le niveau de factorisation de la distribution de probabilité [Populaire00].

En effet, une fois que sont définies les différentes probabilités du réseau bayésien, il faut savoir déterminer les nouvelles probabilités sachant les certitudes qui sont en notre possession. Il a été prouvé que c'est un problème qui est NP-complet (non résolvable en un temps polynomial)[Becker96]. L'inférence consiste à propager une ou plusieurs informations certaines au sein d'un réseau, pour en déduire comment sont modifiées les croyances concernant les autres nœuds. Ce qui revient au calcul de la probabilité d'une variable conditionnée à un ensemble 'observations P(X|e). Ce type d'inférence, appelée aussi *mise à jour des probabilités*, est essentiel dans des applications, où l'on doit reconsidérer son appréciation de la situation en fonction d'une ou plusieurs nouvelles observations. La complexité du calcul dépend de la complexité du graphe, de ce fait nous allons traiter le calcul pour deux types de graphe 'graphe sans boucle' et 'graphe avec boucle'.

IV.1. Inférence dans les réseaux sans boucles [Pearl00]

IV.1.1. Définitions (notions de graphe)

Soit G=(V,A), un CDAG(graphe connecté acyclique orienté) on à :

- G est une ***chaîne*** si chaque nœud a au plus un parent et au plus un fils.
- G est un ***arbre*** si et seulement si chaque nœud a au plus un parent.
- G est un ***polyarbre*** si et seulement si G ne comprend pas de cycle. On dira aussi que G est un CDAG simplement connecté.

Soit G=(V,A), un CDAG. Soit u un nœud, et v un parent de u.

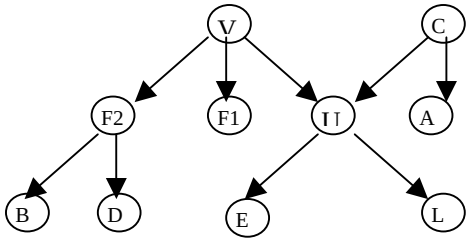


Figure 3 : Graphe G

- On appelle ***Frères*** de u relativement à v, et on note S(u,v), l'ensemble :

  S(u,v)=F(v)\{u}. Dans G, S(u,v)={F1,F2}.

- On appelle ***Coparents*** de u relativement à v, et on note Q(u,v), l'ensemble :

  Q(u,v)=C(u)\{v}. Dans G, Q(u,v)={c}.

- On appelle ***Clôture***, et on note Cl(G,u) ou $Cl_G(u)$ l'ensemble des nœuds connectés à u.

  $Cl_G(u)=\{v\in G|u\text{-}\ldots\text{-}v\}$.

- On appelle ***Clôture partielle de u*** dans G excluant v, et on note $V_{u\setminus v}$ la clôture de u dans le graphe $G_{V\setminus\{v\}}$. $V_{u\setminus v}=Cl(G_{V\setminus\{v\}},U)$. Dans G, $V_{u\setminus v}=\{C,A,U,E,L\}$.

- On appelle ***Clôture supérieure*** **de u** dans G, et on note Sup(u) l'ensemble défini par :

  $Sup(u)=\cup_{C\in C(u)}V_{C\setminus u}$. Dans G, Sup(u)={V,F1,F2,B,D,C,A}.

- On définit de même la ***clôture inférieure de u*** dans G , notée Inf(u) :

  $Inf(u)= \cup_{F\in F(u)}V_{F\setminus u}$.

- On appelle ***Clôture supérieure partielle de u*** dans G excluant v, et on note Sup(u\v) l'ensemble défini par : $Sup(u\setminus v)=\cup_{C\in C(u),C\neq u}V_{C\setminus u}$. Dans G, Sup(u\v)={C,A}. Réciproquement, on définit la ***Clôture inférieure partielle*** de u dans G excluant v :

  $Inf(u\setminus v)= \cup_{F\in F(u),\ F\neq u} V_{F\setminus u}$.

Propriété : si G est un arbre, on a la propriété : $sup(u)=sup(v)\cup Inf(v\setminus u)$.

IV.1.2. Propriétés de D-séparation

Soit G=(V,A), un CDAG.

- Soient U,V1,V2 et W quatre sous-ensembles de V. Si W d-sépare U et V1, et que W d-sépare U et V2 alors W d-sépare U et $V1\cup V2$.
- Soient trois sous-ensembles de V. Soient $W1\subset V1$ et $W2\subset V2$. Si W d-sépare V1 et V2, alors W d-sépare W1 et W2.
- Soit u un nœud. Soient V1, V2 deux sous ensembles disjoints de V, tels que $u\notin V1$, $u\notin V2$. Soit W un ensemble de nœuds vérifiant : $u\notin W$ et $W\cap Desc(u)=\emptyset$. Soit U un ensemble de nœuds tel que $u\in U$. On a les trois propriétés suivantes :
  - Si $V1\cap C(u)\neq\emptyset$ et $V2\cap F(u)\neq\emptyset$, alors U d-sépare V1 et V2.
  - Si $V1\cap F(u)\neq\emptyset$ et $V2\cap F(u)\neq\emptyset$, alors U d-sépare V1 et V2.
  - Si $V1\cap C(u)\neq\emptyset$ et $V2\cap C(u)\neq\emptyset$, alors U d-sépare V1 et V2.

- Soit G=(V,A), un CDAG simplement connecté. Soit u un nœud. Soit U un ensemble de nœuds tel que u∈U. Inf(u) et Sup(u) sont d-séparés par U.
- Soit G=(V,A), un CDAG simplement connecté. Soit u un nœud, et v∈C(u). {u}et Sup(u) sont d-séparés par v.
- G=(V,A), un CDAG simplement connecté. Soit u un nœud, et f1, f2 ∈F(u) deux enfants de u. u d-sépare $V_{f1\setminus u}$ et $V_{f2\setminus u}$.

Pour plus de détails sur les démonstration voir [Becker99]

IV.1.3 Le problème de mise à jour des probabilités

Probabilités

Soit B=(G,P) un réseau bayésien de graphe sous-jacent G=(V,A).
Si E1 et E2 sont deux ensembles de nœuds d-séparés par X dans G, on a la propriété suivante :

$$\mathbf{P(X|E1,E2) \propto P(X|E1).P(E2|X).}$$

Soit B=(G,P) un réseau bayésien de graphe sous-jacent G=(V,A)
Considérant $e=\cup_{i\in Ie}(Xi=x_i^K)$ un fait sur E, tel que $E=\{Xi|i\in I_e\}$. Comme hypothèse supplémentaire, nous supposerons que P(e)≠0.
Le problème de mis à jour des probabilités au nœud X consiste à calculer P(X|e).
Nous décomposons également le lieu du fait e en différents sous-ensembles :

$E_x^+ = E \cap Sup(X)$: Les observations se situant dans la partie du graphe au-dessus de X.

$E_x^- = E \cap Inf(X)$: Les observations se situant dans la partie du graphe au-dessous de X.

$E_{F\setminus X} = E \cap CL(F\setminus X)$: Les observations se situant dans la partie du graphe qui n'est pas connectée à F via X.

$E_{C\setminus X} = E \cap Inf(C\setminus X)$: Les observations se situant dans la partie du graphe qui est connectée à C via d'autres enfants que X.

1. La table de Transformation en un nœud

La table de probabilités P(X|C) est notée *Trans(X)* et est appelée *table de transformation* au nœud X. Cette table joue un rôle particulièrement important dans l'inférence dans les réseaux sans boucle. En effet, par définition des réseaux bayésiens, la distribution de probabilité définie sur un réseau bayésien est égale au produit des tables de transformation *Trans(X)* sur l'ensemble des nœuds de ce réseau. La distribution est donc entièrement définie par la donnée des tables de transformation. Toute inférence dans un réseau, c'est à dire le calcul d'une probabilité conditionnelle ou inconditionnelle dans ce réseau, devra donc être fondée sur l'utilisation des tables de transformation.

2. Mise à jour des probabilités dans une chaîne

Théorème

Dans un réseau bayésien dont le graphe est une chaîne, la mise à jour de la probabilité d'un nœud s'effectue par double récurrence à gauche et à droite du nœud, cette récurrence se terminant dés que l'on rencontre une observation, ou lorsqu'on arrive à un bout de la chaîne.
Notons F le fils de X et C le parent de X, on démontre facilement les relations suivantes[Becker99] :

$$P(X|e) \propto Amont(X) \,.\, Aval(X).$$

$$Aval(X) = Aval(F) \rightarrow Trans(F).$$

$$Amont(X) = Amont(C) \rightarrow Trans(X).$$

En notant : Amont (X)=$P(X|e_x^+)$ et Aval(X)=$P(e_x^-|X)$
Avec les cas particuliers suivants :

P(X|e)=Ind(X,e) si X est une observation ;
Aval(X)=Ind(X,e) si X est une observation ;
Amont(X)=Ind(X,e) si X est une observation ;
Amont (R)= Trans(R) pour le nœud à la racine de la chaîne ;
Aval(T)=1_T pour le nœud feuille de la chaîne.

Ind(X,e) est une table indicatrice.
On a Aval(X)= $P(e_x^-|X)=\sum_{fj\in F}P(e_F^-|fj).P(fj|X)$.
En terme de tables, cela s'écrit : $P(e_x^-|X)=P(e_F^-|F)\rightarrow P(X|F)$
Aval(X)=Aval(F)→Trans(F).

3. Mise à jour des probabilités dans un arbre

Théorème

Les relations suivantes sont démontrées [Becker99]

$$P(X|e) \propto Amont\ (X)\ .\ Aval(X).$$

$$Aval(X)=\Pi_{F\in Enfants(X)} Aval(F)\rightarrow Trans(F).$$

$$Amont(X)=((Amont(\ C).\ \Pi_{S\in Frères(X)} Aval(S))\ \rightarrow Trans(S))\ \rightarrow Trans(X).$$

Avec toujours les cas particuliers suivants :

P(X|e)=Ind(X,e) si X est une obseravtion ;
Aval(X)=Ind(X,e) si X est une observation ;
Amont(X)=ind(X,e) si X est une observation ;
Amont (R)= Trans(R) pour le nœud à la racine de la chaîne ;
Aval(T)=1_T pour le nœud feuille de la chaîne.

La méthode de propagation de probabilités appliquée dans un arbre peut se généraliser dans un poly-arbre.

IV.2 Inférence dans les réseaux avec boucles

L'inférence dans les réseaux avec boucles, c'est à dire pour lesquels il peut exister plusieurs chemins entre deux nœuds, ne peut pas se traiter facilement par des méthodes élémentaires. Deux approches principales ont été développées pour traiter ce problème : les méthodes de conditionnement et les méthodes de regroupement.

IV.2.1 Méthode de Conditionnement

Dans le cas général, il n'est pas possible d'effectuer une propagation locale des informations (dépendance et indépendance entre nœuds).
La méthode de conditionnement, consiste simplement à exécuter les étapes suivantes :

- Identifier un ensemble de nœuds tel que, si tous les arcs partant de ces nœuds étaient supprimés du réseau, le réseau n'aurait plus aucune boucle ;
- Considérer l'ensemble des hypothèses possibles sur les valeurs de chacun de ces nœuds ;

- Dans le cadre de ces hypothèses, effectuer les propagations 'locale' dans le réseau sans boucle correspondant, et en déduire la probabilité conditionnelle recherchée ;
- Sommer les probabilités obtenues dans chaque hypothèse, pondérées par la probabilité de chaque hypothèse.

Cette méthode repose sur la recherche de séparateurs, c'est à dire de nœuds dans le graphe permettant d'en éliminer les boucles. Ces méthodes ont été essentiellement développées par Pearl [Pearl00]. On voit que, dans ce type d'approche, il est important de bien choisir l'ensemble des N nœuds qui suppriment toutes les boucles. En effet, en supposant que chaque nœud a k états possibles, le nombre de propagations complètes à effectuer est égal à k^N.

IV.2.2 Méthode de Regroupement

Les méthodes de regroupement sont aujourd'hui les plus répondues, en grande partie grâce aux outils informatiques qui sont disponibles pour mettre en œuvre la méthode dite de l'arbre de jonction développée par Lauritzen, Spiegelhalter, et Jenson [Lauritzen88][Jenson96].

L'idée de base de la méthode de regroupement est simplement d'essayer de se ramener à un réseau sans boucle en créant des nœuds plus complexes, qui représentent plusieurs nœuds du graphe original. La méthode de l'arbre de jonction consiste à chercher à regrouper les nœuds d'une façon relativement naturelle. On forme ainsi un arbre reliant des variables plus complexes que les variables initiales.

- Les arbres de regroupement sont des arbres dont les nœuds sont des ensembles de nœuds du graphe original. Ces ensembles ne sont pas quelconques, mais doivent présenter une sorte de continuité le long d'un chemin.
- Les arbres de jonction sont des arbres de regroupement formés à partir du graphe initial auquel on a fait subir deux transformations (moralisation et triangulation). Cette méthode est aujourd'hui la meilleure connue en terme de complexité d'algorithmiques. Cependant, il a été démontré que le problème général de l'inférence dans un réseau bayésien est un problème NP-complet [Cooper90][Shimony03].

IV.2.2.1. Graphes

- ***Graphe de regroupement :*** Soit V un ensemble fini. On note **V** un ensemble de parties de V. on dite que H=(**V**,A) est un graphe de regroupement sur V si et seulement si :
  - H est un graphe sur **V.**
  - $\cup_{u \in V} \mathbf{u} = V$.
- ***Arbre de regroupement :*** Si H=(**V**,A) est un graphe de regroupement sur V, on dit sue H est un arbre de regroupement sur V si et seulement si :
  - H est arbre sur **V.**
  - Pour tout couple $\mathbf{(u,v) \in V*V}$ connectés par l'unique chemin $\mathbf{[u,v_0,v_1,\ldots,v_n]}$, on a la propriété : $\forall\ 0 \leq i \leq n\ \mathbf{u \cap v \subset v_i}$.
- ***Graphe moral :*** Soit G=(V,A) un graphe orienté. On dit que le graphe M=(V,E_M) est le graphe moral de G si et seulement si :
  - *M n'est pas orienté ;*
  - $A \subset E_M$;
  - $\forall\ (u,v) \in V*V,\ F(u) \cap F(v) \neq \varnothing \Rightarrow (u,v) \in E_M$.

L'opération de moralisation consiste à marier les parents.

- ***Graphe triangulé :*** Soit G=(V,E) un graphe non orienté. Un graphe T=(v,E_T) est un graphe triangulé de G si te seulement si :
  - T n'est pas orienté ;
  - E⊂E_T.
  - Pour tout cycle [v_0,v1,..,v_n,v_0] de longueur supérieure ou égale à 4 (n≥ 3), il existe i>j+1, tel que (v_i,v_j) ∈E_T. (v_i,v_j) est appelé une corde.
- ***Clique :*** Soit G=(V,E), un graphe. Soit W⊂V. W est une clique si et seulement si :

$$\forall(u,v) \in W*W, (u,v) \in E.$$

- ***Clique maximale :*** Soit G=(V,E) un graphe. Soit W une clique. W est une clique maximale si et seulement s'il n'existe aucun sur-ensemble U⊃W, tel que U soit une clique.
- ***Arbre de jonction :*** Soit G=(V,E) un CDAG. Soit M =(V,E_M) le graphe moral associé à G, et soit T =(V,E_T), un graphe triangulé associé à M. on dit que **J=(V,**A_j) est un arbre de jonction associé à G si et seulement si :
  - **J** est arbre de regroupement ;
  - Toute clique maximale dans T est un nœud de **V**.
- ***Voisins :*** soit G=(V,E)un graphe. Soit u un nœud dans G. On appelle voisins de u l'ensemble : Γ(u)={v∈V|(u,v)∈E}.
- ***Nœud simple :*** Soit G=(V,E) un graphe non orienté. Soit u un nœud dans G. On dit que u est nœud simple si Γ(u) est une clique.
- ***Séparateur :*** Soit G=(V,E) un graphe tel que V ne soit pas une clique. Soient u et v deux nœuds tels que (u,v)∉E. Si W⊂ V\{v} est un ensemble de nœuds tel que u et v ne sont pas connectés dans $G_{V\setminus W}$, alors W est appelé séparateur de u et v.
- ***Séparateur minimal :*** Si W est séparateur de u et v, mais qu'aucun sous-ensemble de W n'en est un, alors W est appelé un séparateur minimal de u et v.
- ***Arbre non orienté :*** un graphe non orienté G=(V,E) est appelé arbre non orienté s'il est connecté et n'a pas de cycles.
- ***Arbre recouvrant :*** Soit G=(V,E) un graphe. Soit H=(V,E_H) un arbre non orienté tel que E_H⊂E. On dit alors que H est un arbre recouvrant sur G.
- ***Arbre recouvrant de poids maximal :*** Soit G=(V,E). Soit H=(V,E_H) un arbre recouvrant sur G. Soit W : E→R^+ une fonction appelée poids. On dit que H=(V,E_H) est un arbre recouvrant de poids maximal sur G suivant w si et seulement si :

$$\Sigma_{e\in E_H} w(e)=\max_k (\Sigma_{e\in E_k} w(e))$$

IV.2.2.2. Les Tables

- ***Factorisation sur un arbre de regroupement :*** Soient V un ensemble fini, H=(**V**,A) un arbre de regroupement sur V, et T une table définie sur l'ensemble X=$X_{1,...,}X_n$ qui est le produit cartésien de n ensembles discrets $(X_i)_{1\le i \le n}$.

On dit que T se factorise sur H suivant Trans si et seulement si :

- Il existe une bijection entre les éléments de V et les variables (X_i) ;
- Pour chaque élément **v∈V,** on peut définir une table sur **v∪C(v)** notée Trans(**v**) telle que : T($X_{1,...,}X_n$)=$\Pi_{\mathbf{v}\in\mathbf{V}}$Trans(**v**). On appellera Trans une application de transformation.

- ***Conditionnement sur un arbre de regroupement :*** Si $T_{\mathbf{u}}$ est une table définie sur le produit cartésien des variables contenues dans **u**, et $T_{\mathbf{u}\cup\mathbf{v}}$ une table définie sur le produit cartésien des variables contenues dans **u∪v**, on définit :

$$T_{\mathbf{u}} \rightarrow T_{\mathbf{u}\cup \mathbf{v}} = \oplus_{\mathbf{u}\setminus\mathbf{v}}(T_{\mathbf{u}} \otimes_{\mathbf{u}} T_{\mathbf{u}\cup\mathbf{v}})$$

Cela revient à marginaliser le produit des deux tables par les variables se trouvant uniquement dans **u.**

Exemple : si **u** ={A,C}et **v**={B,C} on notera :

$T_{\{A,B\}} \rightarrow T_{\{A,B,C\}} = \oplus_{\{A\}}(T_{\{A,B\}} \otimes_{\{A,B\}} T_{\{A,B,C\}})$. La table résultante sera donc définie sur {B,C}.

IV.2.2.3. Propagation dans un arbre de regroupement

Soient V un ensemble fini, H=(**V**,A) un arbre de regroupement sur V, et *Trans*(**v**) une table définie sur **v∪c**, pour chaque **v∈V,** dont le parent dans H est **c.**

On définit alors la table $\mathbf{Q}_{Trans}(\mathbf{v})$ par la relation de récurrence suivante :

$$\mathbf{Q}_{Trans}(\mathbf{v}) = \text{Amont}(\mathbf{v}).\text{Aval}(\mathbf{v})$$

$$\text{Amont}(\mathbf{v}) = (\text{Amont}(\mathbf{c}) . \Pi_{\mathbf{s}\in \text{Frères}(\mathbf{v})} \text{Aval}(\mathbf{s}) \rightarrow \text{Trans}(\mathbf{s})) \rightarrow \text{Trans}(\mathbf{v})$$

$$\text{Aval}(\mathbf{v}) = \Pi_{\mathbf{f}\in \text{Enfants}(\mathbf{v})} \text{Aval}(\mathbf{f}) \rightarrow \text{Trans}(\mathbf{f})).$$

Avec les deux cas particuliers, où **r** est la racine de H, et **t** est une feuille de H :

$$\text{Amont}(\mathbf{r}) = \text{Trans}(\mathbf{r})$$
$$\text{Aval}(\mathbf{t})) = \mathbf{1}_{\mathbf{t}}.$$

Lemmes :

- **Marginalisation d'une table factorisée :** Si T est une table sur un ensemble de variables discrètes X_i qui se factorise sur H suivant Trans. On a :

$$\mathbf{Q}_{Trans}(\mathbf{u}) = \oplus_{\mathbf{V}\setminus\mathbf{u}} T(X_{1,..,} X_n).$$

- **Propagation des probabilités dans un arbre de regroupement :** Soient V un ensemble fini, H=(**V**,A) un arbre de regroupement sur V, et ε une épreuve à laquelle est associée un espace probabilisé fini, et n variables aléatoires $(X_i)_{1\le i\le n}$, tels que P se factorise sur H suivant Trans. Alors P(**v**)= $\mathbf{Q}_{Trans}(\mathbf{v})$.

$$\text{Donc } \mathbf{Q}_{Trans}(\mathbf{v}) = \oplus_{X(V\setminus \mathbf{v})} P(X_{1,..,}X_n).$$

- **Application de projection sur l'arbre de jonction :** Soit G=(V,A) un CDAG et soit **J=(V,**A_j) un arbre de jonction associé à G. Alors on peut définir une application $\Gamma_j : V \rightarrow \mathbf{V}$, telle que :

$$\forall\, v\in V : \{v\} \cup C(v) \subset \Gamma_j(v).$$

- **Factorisation d'un réseau bayésien sur son arbre de jonction :** Soit B=(G,P), un réseau bayésien de graphe sous-jacent G=(V,A). Soit **J=(V,**A_j) un arbre de jonction associé à G.

Si l'on définit, de plus :Trans (**v**) =$\Pi_{v\in V,\Gamma j(v)=\mathbf{v}}$ P(v|C(v)).

Alors P se factorise sur **J,** avec l'application Trans.

IV.2.2.4. Inférence dans le réseau original

La factorisation sur un arbre de regroupement ou sur un arbre de jonction nous a permis de montrer que $P(\mathbf{v})=Q_{Trans}(\mathbf{v})$, ce qui implique la possibilité de calculer $P(\mathbf{v})$ par propagation de l'application Trans, de façon similaire à ce qui est fait dans les réseaux sans boucles. Notre objectif est bien de calculer des probabilités dans le réseau bayésien initial, et plus précisément de calculer des probabilités conditionnelles à l'observation d'un fait.
D'une façon générale, si $W \subset V$ est un ensemble de nœuds du graphe initial, nous devons calculer P(W). Une approche possible est de calculer $P(\mathbf{v} \cup W)$ pour un nœud quelconque de l'arbre de regroupement **v** puis de marginaliser le résultat sur les variables contenues dans $\mathbf{v} \backslash W$. Dans la pratique, on choisira de préférence un nœud **v** proche de W, pour faciliter l'opération de marginalisation.

1. Calcul des probabilités conditionnelles

Pour calculer P(X|e) pour un nœud du graphe original, nous utiliserons simplement la formule de bayes : P(X|e)=P(X,E)/ P(e) .
P(e) est une constante. De plus comme $\oplus_X(P(X|e)) = 1$, nous pouvons écrire $P(W|e) \propto P(X,e)$. De même nous pouvant utiliser P(W,e) pour calculer P(W|e).

IV.3. Construction de l'arbre de jonction

La construction de l'arbre de jonction est particulièrement importante dans un réseau bayésien avec boucles. La construction de l'arbre de jonction d'un réseau bayésien s'effectue en trois étapes :

- La moralisation du graphe original ;
- La triangulation du graphe moral ;
- La construction de l'arbre de jonction à partir du graphe triangulé.

L'opération de moralisation est immédiate puisqu'il s'agit simplement de marier les parents du graphe initial. La construction de l'arbre de jonction étant possible de façon rigoureuse pour un graphe triangulé, nous abordons ce point en premier lieu.

IV.3.1. Construction de l'arbre de jonction pour un graphe triangulé

Lemmes : Pour plus de détails sur la démonstration des lemmes suivants se trouve dans [Becker99].

- **Séparateur minimal d'un graphe triangulé :** Soit T=(V,E) un graphe triangulé comprenant plus de trois nœuds, et que V ne soit pas une clique. Soient u et v deux nœuds tels que $(u,v) \notin E$. Alors tout séparateur minimal de u et v est une clique.
- **Existence de nœuds simple dans un graphe triangulé :** Soit T=(V,E) un graphe triangulé. Alors ou bien T est une clique, ou bien il existe deux nœuds simples u et v tels que $(u,v) \notin E$

Théorèmes :

1) **Existence de l'arbre de jonction pour un graphe triangulé :** Soit T=(V,E), un graphe triangulé. Alors il existe un arbre de jonction $\mathbf{J=(V,}A_j)$ sur T.

2) **Construction de l'arbre de jonction d'un graphe triangulé :** Soit T=(V,E), un graphe triangulé et H=(**V,**A) un graphe de regroupement sur T, tel que toute clique maximale dans T soit un nœud de **V**. Alors un arbre de jonction **J=(V,**A_j) sur T est un arbre de poids maximal sur H suivant l poids w(u,v)=|**u**∩**v**|.

IV.3.2. Triangulation d'un graphe moral

Il est possible de construire différentes triangulations d'un graphe moral. En particulier, on peut simplement construire une seule clique maximale connectant tous les nœuds du graphe. Intuitivement, il est clair qu'une telle triangulation n'est pas très intéressante par rapport à l'objectif initial. Donc, il faut définir des critères permettant de mesurer la qualité d'une triangulation. On peut utiliser la taille de la plus grande clique. De façon plus intéressante, comme la complexité d'une table Trans sera proportionnelle au produit du nombre de modalités de l'ensemble des variables contenues dans un nœud de l'arbre de regroupement, on peut utiliser une mesure de la taille totale des tables Trans nécessaires à la factorisation de la distribution sur l'arbre de regroupement .

Il a été montré que la recherche d'un arbre de jonction optimal est un problème NP-complet, pour la plupart des critères d'optimalité. Un algorithme exact existe pour le critère de la taille de la plus grande clique, mais sa complexité est n^{k+1} où n est le nombre variables et k la taille de la clique maximale dans le graphe optimal, ce qui le rend impraticable même pour des graphes de taille raisonnable.

IV.3.3. Recherche de l'arbre de poids maximal

Il existe plusieurs algorithmes simples pour la recherche d'un arbre de poids maximal. Celui de Kruskal présente l'avantage de la simplicité.

Soit G=(V,E) un graphe. Soit w : $E \rightarrow R^+$ un poids défini sur E. Notons E_T l'ensemble des arcs de l'arbre.

Recherche de l'arbre de poids maximal
(Algorithme de kruskal)
Ordonner tous les arcs possibles par poids décroissant
Initialiser E_T à ∅
Répéter pour chaque arc e
Si le graphe obtenu en ajoutant e à E_T n'a pas de cycle Alors Ajouter e à E_T.

Théorème :Arbre de pois maximal

L'arbre T=(V,E_T) obtenu par l'algorithme de Kruskal est un arbre de pois maximal sur G.

Voir Démonstration dans [Gondran95].

V. Complexité

L'inférence dans un réseau bayésien est NP-complet, par réduction à un problème de réalisabilité d'une clause logique. La démonstration de cette propriété est due à Cooper [Cooper92].

Les réseaux bayésiens simplement connectés sont un cas spécial incluant les poly-arbres et les réseaux à deux niveaux. L'inférence dans les réseaux bayésiens est difficile dans le cas général [Cooper90][Becker99]. L'analyse de la complexité d'une sous-classe est d'une extrême importance. Le problème de l'inférence dans les réseaux bayésien simplement connecté (polyarbre) est un problème difficile NP-difficile [Shimony03] pour plus de détails sur les démonstrations voir [Shimony03].

V.1 Le problème 3SAT

Le problème 3SAT est un problème NP-complet. Il est défini de la façon suivante. On considère un ensemble de variables booléennes $(X_i)_{i \leq i \leq n}$, c'est à dire des variables pouvant prendre les valeurs {V,F}.

Une clause logique d'ordre 3 est une expression de la forme :
$L(X_1,X_2,\ldots,X_n)=[\neg]X_{l1} \vee [\neg]X_{l2} \vee [\neg]X_{l3}$ où $1 \leq l_1 \leq l_2 \leq l_3 \leq n$, et où $[\neg]$ signifie que la négation est optionnelle (le signe $\neg$ peut être présent ou absent).

On dit que la clause L est satisfaite par $(x_1,x_2,\ldots,x_n)$ si et seulement si $L(x_1,x_2,\ldots,x_n)=V$.

On dit qu'une collection de clauses $L=\{L_1, L_2,\ldots, L_n\}$ est *réalisable* si et seulement s'il existe un jeu de variables booléennes $(x_1,x_2,\ldots,x_n)$ qui satisfont simultanément toutes les clauses.

Le problème 3SAT consiste à déterminer si une collection de clauses $L=\{L_1, L_2,\ldots, L_n\}$ est réalisable ou non.

V.2 Réduction de l'inférence

Il a été démontrer que l'inférence dans un réseau bayésien est réductible au problème 3SAT. Soit une collection de clauses $L=\{L_1, L_2,\ldots, L_n\}$ d'ordre 3 sur n variables booléennes $(X_i)_{1 \leq i \leq n.}$ Nous cherchons à décider s'il existe un jeu de valeurs $(x_1,x_2,\ldots,x_n)$qui satisfasse simultanément toutes les clause de L.

Considérons un graphe acyclique orienté G=(V,A) comme sur la figure. Pour chacune des n variables booléennes, nous créons un nœud correspondant $(X_i)_{1 \leq i \leq n.}$. Pour chacune des p clauses d'ordre 3, nous créons un nœud$(L_i)_{1 \leq i \leq p}$. Finalement nous ajoutons un nœud Y qui représente la collection, et dont les parents sont tous les $(L_i)_{1 \leq i \leq p.}$

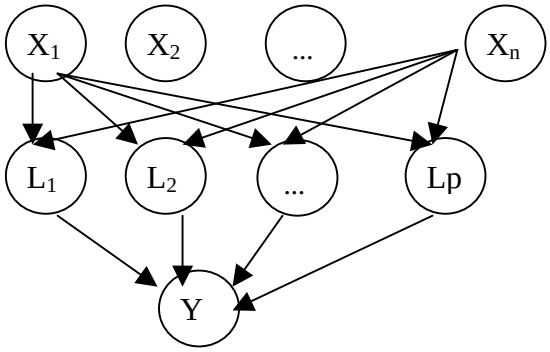


Figure4 : Réseau bayésien pour le problème 3SAT

Nous définissons de même les n+p+1 variables aléatoires binaires $(X_i)_{1\leq i \leq n}$, $(L_i)_{1\leq i \leq p}$ et Y.
Les probabilités sur ce réseau sont définies comme suit :

- $P(X_i=x^j_i)=0,5$ pour $1\leq i \leq n$; $j=1,2$;
- $P(L_i=l^{ki}_i | X_{l1}=x^{k1}_{l1},\ldots, X_{ln}=x^{kn}_{ln})=1$, si $L(x^{k1}_l,\ldots,x^{kn}_n)=V$;
- $P(L_i=l^{ki}_i | X_{l1}=x^{k1}_{l1},\ldots, X_{ln}=x^{kn}_{ln})=0$, sinon ;
- $P(Y=V|L_1=l^{k1}_1,,,,L_p=l^{kp}_p)=1$, si $\forall i$, $1\leq i \leq p$, $l^{ki}_i=V$.
- $P(Y=V|L_1=l^{k1}_1,,,,L_p=l^{kp}_p)=0$, sinon.
-

On a donc :
$P(X_1,\ldots,X_n,L_1,\ldots,L_p,Y)=\Pi^n_{i=1}P(X_i).\Pi^p_{j=1}\ P(L_j|C(L_i)).P(Y|C(Y))$.
Ce réseau ainsi défini, calculer P(Y)=V revient à décider s'il existe un jeu de valeurs$(x_1,x_2,\ldots,x_n)$ qui satisfait simultanément toutes les clauses de L.
L'inférence peut être interprétée comme la propagation de certaines observations dans le réseau. Cette propagation peut être effectuée directement dans les cas de réseaux sans boucle. Dans le cas où les réseaux présentent des boucles, on est amené d'abord à créer un macro-réseau dont chaque nœud représente un ensemble de variables. Le même type de propagation peut alors être appliqué. En effet, une fois que sont définies les différentes probabilités du réseau bayésien, il faut savoir déterminer les nouvelles probabilités sachant les certitudes qui sont en notre possession. Il a été prouvé que c'est un problème qui est NP-difficile (non résolvable en un temps polynomial).

Toute application mettant en œuvre des connaissances peut donc relever de l'utilisation des réseaux bayésiens, qu'il s'agisse de formaliser la connaissances d'experts, d'extraire la connaissance contenue dans des bases de données, ou d'utiliser le plus rationnellement possible l'un ou l'autre type de connaissances. On utilise les réseaux bayésiens pour leur capacité d'effectuer des inférences dans un contexte d'incertitude, en quelque sorte comme alternative aux systèmes experts.

On les utilise aussi pour leurs algorithmes d'apprentissage, comme alternative aux autres méthodes de modélisation quantitative, en les considérant comme des modèles de régression.

VI. Avantage et limites des réseaux bayésiens [Becker99]

VI.1 Avantages des réseaux bayésiens :

Les réseaux bayésiens fournissent des outils intuitifs et naturels pour traiter des problèmes dans lesquelles l'incertitude et la complexité des données joue un rôle important. Les réseaux causaux bayésiens sont des modèles graphique pour la représentation de la relation de causalité [Dubois03].

- ✓ La représentation des connaissances causales utilisées dans des réseaux bayésiens est la plus intuitive possible : simplement relier des causes et des effets par des flèches. Pratiquement toute représentation graphique d'un domaine de connaissances peut être présentée sous cette forme [Geslin03].

- ✓ Un formalisme unificateur : la plupart des applications qui relèvent des réseaux bayésiens sont des applications d'aide à la décision. Par nature, ces applications intègrent un certain degré d'incertitude, qui est assez bien pris en compte par le formalisme probabiliste des réseaux bayésiens.

- ✓ Une représentation de la connaissance lisible : les deux propriétés fondamentales des réseaux bayésiens sont, d'abord, d'être des graphes orientés, c'est à dire de représenter des causalités et non des simples corrélations, et, ensuite, de garantir une correspondance entre la distribution de probabilité sous-jacente et le graphe associé.

- ✓ Une gamme de requêtes très complète : les possibilités offertes par les algorithmes d'inférence permettent d'envisager une gamme de requêtes assez complète, et qui peut être extrêmement intéressante dans certains types d'applications. Tout d'abord, il n'y a aucune réelle contrainte sur les informations nécessaires pour être en mesure de calculer la probabilité d'un fait. Dans tous les cas, l'inférence est possible, et la nouvelle information permet de raffiner les conclusions[Fabiani97][Populaire00].

Il n'y a pas d'entrée ni de sortie dans un réseau bayésien. Le réseau peut donc être utilisé pour déterminer la valeur la plus probable d'un nœud en fonction d'informations données (prévoir, au sens entrées →sorties), mais également pour connaître la cause la plus probable d'une information donnée (expliquer, ou sens sorties → entrées) cette requête s'appelle *explication la plus probable*.

Le cadre bayésien permet de mettre en évidence des phénomènes de renversement d'explication " explaining away ". Ils correspondent à l'influence négative d'une observation sur la plausibilité d'une cause potentielle, suggérée par une première observation. C'est-à-dire qu'une observation complémentaire conduit à écarter une cause pressentie au profit d'une autre cause maintenant plus plausible (compte tenu de la nouvelle observation)[Dubois03].

VI.2 Les limites des réseaux bayésiens [Becker99]

Les réseaux bayésiens sont essentiellement des modèles d'évaluation des risques. Ce type de modélisation ne se prête pas naturellement à la représentation de systèmes dynamiques puisqu'elle ne permet aucun feedback (le graphe doit être acyclique) et ne traite que de probabilités et non de quantités physiques ou monétaires (stocks et flux). La dimension temporelle peut cependant n'être pas tout à fait absente si l'on considère comme variables différentes les états successifs d'une même variable.

Les faiblesses tiennent surtout au fait que le nombre de parents directs que peut recevoir une même variable, bien que théoriquement illimité, est pratiquement limité par le fait qu'il est nécessaire de disposer de la distribution conditionnelle conjointe de cette dernière par rapport à l'ensemble de ses parents.

- ✓ Utilisation des probabilités : l'utilisation des graphes de causalités est une approche très intuitive. L'utilisation des probabilités pour rendre ces modèles quantitatifs était justifiée. Il reste cependant que la notion de probabilité, est, au contraire, assez peu intuitive. Il est en effet assez facile de construire des paradoxes fondés sur des raisonnements probabilistes.
- ✓ Les variables continues : l'essentiel des algorithmes développés pour l'inférence dans les réseaux bayésiens utilisent des variables discrètes. Même s'il est théoriquement possible de généraliser les techniques développées aux variables continues, il semble que la communauté de recherche travaillent sur les réseaux bayésiens n'a pas encore vraiment intégré ces problèmes. Cela pénalise cette technologie en particulier pour des applications de data-mining où variables continues et discrètes cohabitent.

- La généralité des formalismes des réseaux bayésiens aussi bien en termes de représentation que d'utilisation les rend difficiles à manipuler à partir d'une certaine taille. La complexité des réseaux bayésiens ne se traduit pas seulement en termes de compréhension par les utilisateurs. Les problèmes sous-jacents sont pratiquement tous de complexité non polynomiale, et conduisent à développer des algorithmes approchés, dont le comportement n'est pas garanti pour des problèmes de grande taille [Cooper92].

Une des raisons du boom actuel des réseaux bayésiens, outre leur convivialité et leur efficacité, est la multiplicité d'applications dans les domaines de l'industrie, du marketing, de la santé, de la banque, de la finance, du droit, etc. Le "système" dont on représente la connaissance au moyen d'un réseau bayésien peut être aussi bien le contenu du caddie d'un client de supermarché, un navire de la Marine, le patient d'une consultation médicale, le moteur d'une automobile, un réseau électrique, l'utilisateur d'un logiciel, etc.
Les réseaux bayésiens apportent des solutions rapides et intuitives à différentes sortes de problématiques d'évaluation, de décision, de prévision, de diagnostic, etc.

Les réseaux bayésiens sont actuellement une des techniques les plus intéressantes de l'intelligence artificielle car ils permettent la représentation de la connaissance par un graphe causal intuitif et compréhensible. De plus, comme ils sont basés sur des probabilités, ils intègrent l'incertitude dans le raisonnement. Malheureusement, il s'agit d'un domaine de recherche récent et l'offre logicielle est encore pauvre et incomplète.

VII.Conclusion

Les réseaux bayésiens sont un mode de représentation des connaissances fondé sur une description des relations entre les variables d'un domaine donné. Les arcs orientés du réseau représentent des relations de causalité et donc de dépendance entre les variables associées aux nœuds ainsi reliés[Populaire00]. Ce mode de représentation sous forme de distribution de probabilité permet l'utilisation la plus rationnelle possible de ces connaissances dans la plus part des situations, en particulier quand certaines des informations qui permettraient de décider ne sont pas connues.

Les réseaux causaux bayésiens sont des modèles graphiques pour la représentation de la relation de causalité [Dubois03]. Leur utilisation essentielle est de calculer des probabilités conditionnelles d'événements reliés les uns aux autres par des relations de cause à effet. Cette utilisation s'appelle Inférence. Le problème de l'inférence est uniquement un problème de calculs. Les méthodes de calculs sont plus au moins complexes selon la complexité du graphe. Il a été prouvé que c'est un problème qui est NP-difficile (non résolvable en un temps polynomial). Mais comme il s'agit de graphe on peut donc décomposer le problème en problèmes moins complexes.

Dans le chapitre suivant nous allons voir une approche pour la modélisation de la relation causale, c'est une approche qui a été influencée par les idées de Shoham proposé par Mokhtari, une approche normative sur la causalité [Mokhtari97a]. Cette théorie causale est une approche qui permet de maintenir les points du 'frame problem' : la Qualification, la Persistance et enfin la Ramification [Khelfallah01].

I. Introduction

Le concept de la causalité est central dans notre raisonnement de tous les jours et intuitivement banal. Pourtant diverses disciplines s'y intéressent, certains depuis des siècles. Une nouvelle approche pour représenter la causalité a été proposée par Mokhtari, une approche normative de la causalité[Mokhtari94][Mokhtari97a]. L'idée principale est que la déduction non-monotone devrait être modélisée naturellement par la donnée de certaines normes dans le monde. Il est plus naturel d'avoir à notre disposition un ensemble de *normes* qu'un ordre préférentiel, et d'expliquer comment ces normes peuvent aider à offrir une définition pratique de la causalité. Par ailleurs, une conception interventionniste de la causalité est considérée i.e. une préférence de sélectionner les causes parmi un ensemble d'*actions* qu'un agent a la possibilité d'effectuer ou pas. La représentation de la notion de volonté libre a exigé une structure de temps avec un branchement dans le futur. Un branchement dans le passé a été aussi exigé pour examiner les différents événements qui mènent à la même situation [Mokhtari94].

Nous présentons dans ce chapitre, premièrement les notions de base qui ont motivées cette modélisation (La non-monotonie, Action, Norme, Persistance,..), qui sera suivit par les définitions participant à la formalisation du langage utilisé pour la modélisation d'une théorie causale et ses caractéristiques. Nous présentons ensuite la modélisation de la notion d'explication ainsi que la prédiction. Enfin, nous présentons une solution du problème de la ramification [Khelfallah01].

II. Action et causalité

Plusieurs approches ont été développées sur le concept d'actions, en particulier pour trouver une solution au «frame problème »[McCarty69] qui a suscité de nombreuses discussions. Rappelons que le «frame problème » englobe trois principaux problèmes :

- La qualification,
- La persistance, et
- La ramification.

Sachant qu'une bonne représentation des informations temporelles est un pré-requis pour le raisonnement causal. De nombreux auteurs n'ont pas réussi à atteindre leurs objectifs, essentiellement dans la théorie du changement, pour avoir négligé le temps (voir notamment les travaux de McCarthy [McCarthy69] et de Lewis [Lewis73]).

II.1 Le temps dans la causalité [Mokhtari97a]

Il y a deux idées clés à représenter :

1. la continuité du temps,
2. l'ouverture du futur et du passé.

Le temps a été défini explicitement au moyen de points du temps.

Définitions [Mokhtari97a] : Soit T l'ensemble des points de temps.

- Un *point du temps t,* où $t \in T$, est un état instantané de l'univers définit par un sous-ensemble de propositions vraies à une certaine date et par cette date. La *date d'un point du temps* est définie par la fonction Date qui associe à chaque élément de T sa date le projetant sur l'axe des réels :

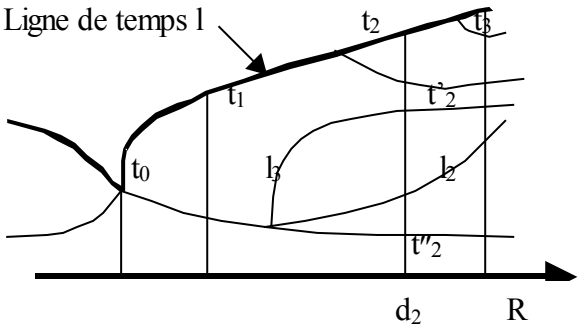
Date :$T \rightarrow R$ où *date (t)* = d (noté d_t).

Signifie que la date du point de temps t est le réel d.

-Une *ligne de temps l* (ou 'chronique', selon la terminologie de McDermott [McDermott82b]) c'est une succession de points du temps, qui sont en bijection avec un ensemble de dates. Il représente une évolution possible de l'univers. Un point de temps de cette succession répond à la règle que «*il n'y a pas d'effet sans cause*».

Soit L l'ensemble de toutes les lignes de temps. Les points de temps d'une ligne du temps sont totalement rangés par la relation de la précédence notée " ≤"; tel que $t_1 \leq t_2$ veut dire que t_2 ne précède pas t_1. Si $t_1 <= t_2$ alors $d_{t1} <= d_{t2}$. La relation de la précédence exprime le principe *"aucun effet ne précède sa cause "*.

Parmi les points de temps, des points de temps particuliers sont distingués appelés «*points de choix* ». Ces points de choix sont des points de temps qui sont associés aux actions.



$t_1<t_2<t_3$ et $t_1<t'_2$, t'_2 et t_3 sont non reliés
Mais date(t'_2)< date(t_3), et date(t_2)=date(t'_2)

Figure 1 : Structure de branchement

II.2 Langage[Mokhtari97a][Kayser98][Khelfallah01]

Le langage proposé est défini à deux niveaux :

1. Le premier niveau représente les informations statiques. C'est un langage propositionnel ordinaire avec : P un ensemble des propositions qui nous intéressent, A l'ensemble des actions et E l'ensemble d'effets(faits, événements…), avec $A \cap P = \varnothing$ et $P = A \cup E$.
   Une formule du premier niveau est soit une formule d'action ou une formule d'effet..
   Soit FOR(A) (respectivement FOR(E)) l'ensemble des formules d'actions (respectivement d'effet).
   Une littéral d'effet est soit un effet, ou la négation de l'effet. L'ensemble des littéraux d'effet est noté LIT(E), tel que :

$$LIT(E) = E \cup \{\neg e : e \in E\}$$

Soit *obp* un des opérateurs binaires classiques : $\wedge$, $\vee$ ou $\supset$, a, a'$\in$A et e, e'$\in$E.
Soit a et a' des formules, $\neg$a et a *obp* a' sont aussi des formules.
Soit e et e' des formules, $\neg$e et e *obp* e' sont aussi des formules.
Notez qu'une formule de ce langage est un regroupement particulier dans lequel la composition des actions avec des événements n'est pas permise.

2. Le deuxième niveau exprime les informations dynamiques, les informations temporelles ; représentées par des formules de la forme :

   - v(p,l,d) qui signifie que la proposition p est vraie dans la ligne du temps l à date d.
   - nocc(p,l,d_t,Δ) qui signifie que la proposition p n'apparaît pas dans la ligne de temps l à partir de la date du point de temps t sur toute la durée Δ.

- occ(e,l,d) qui signifie que l'effet e à été généré dans la ligne de temps l à la date t. Nous avons : $\forall e,l,d$; $occ(e,l,d) \supset v(e,l,d)$.

La théorie des modèles associée est une généralisation de la sémantique Kripéenne des mondes possibles [Kripke63]. Dans ce modèle, une interprétation est définie comme une fonction I telle que :

$$I : L \times R \rightarrow 2^P \text{ où}$$

- L est l'ensemble des lignes de temps,
- R est l'ensemble des réels,
- $t \in l$ est vrai dans le modèle ssi t est précisément l'ensemble $I(l,d_t)$ (rappelons qu'un point de temps est l'ensemble de propositions vraies à la date de ce point de temps),
- $v(p,l,d_t)$ est vraie dans le modèle ssi $p \in I(l,d_t)$, i.e., la proposition p est vraie au point de temps t déterminé par la ligne de temps l et la date d_t.
- $nocc(p,l,d_t,\Delta)$ est donc vraie dans l modèle ssi $(\forall t')\ t' \in l \wedge d_t \leq d_{t'} \leq d_t+\Delta \supset p \notin I(l,d_{t'})$

Les interprétations des formules $v(p,l,d_t)$ ou $nocc(p,l,d_t,\Delta)$ permettent facilement les définitions des formules :

- $v(p \wedge p,l,d_t)$,$v(p \vee q,l,d_t)$ et $v(\neg p,l,d_t)$, ou
- $nocc(p \wedge p,l,d_t,\Delta)$,$nocc(p \vee q,l,d_t,\Delta)$ et $nocc(\neg p,l,d_t,\Delta)$.

Les actions peuvent causer des événements, des faits ou autres. Comme dans [Shoham88], aucune distinction n'est faite entre ces objets, ceci permet d'éviter de définir plusieurs types de causes telles que les 'Ecause' et 'Pcause' de McDrmott [McDermott82b] ou 'Ecause' et 'Acause' de J. Allen [Allen84].

La Causalité est exprimée par l'opérateur de causalité normal '$\Rightarrow$'. '$a \Rightarrow e[\Delta]$' exprime que l'action **a** implique normalement l'effet **e** dans le délai Δ, à moins qu'il y ait l'occurrence d'un événement qui inhibe l'effet **e**. Pour la formalisation de telle notion, nous avons besoin d'un raisonnement non-monotone exprimé au moyen de la norme de l'action et les événements inhibant.
Pour donner un sens à cette notion de normalité, il faut raisonner avec des informations incertaines, dans l'absence d'informations spécifiques, on suppose que les choses se comportent normalement, nous devons donc prendre en compte :

1. Les pré-conditions d'une action, ou qualification de l'action correspondant à la question de savoir quand est-ce qu'il est raisonnable de supposer qu'une action peut s'effectuer ?

2.Les post-conditions d'une action correspondant à la question de savoir si les effets de l'action ne sont pas inhibés par d'autres faits, et enfin

3. La persistance, correspondant au fait qu'un événement demeure vrai pendant sa durée de persistance à moins qu'un événement externe entraîne sa fausseté.

Tous ces aspects nécessitent généralement l'utilisation d'un raisonnement non-monotone.

II.3 Non-monotonie dans la causalité[Mokhtari94][Mokhtari97a][Kayser98] [Khelfallah01]

La norme d'une action est définie comme l'ensemble de propositions qui doivent être normalement vraies pour pouvoir exécuter l'action. Formellement, la norme est définie comme une fonction :

$$\text{norme}:A \rightarrow 2^{FOR(E).}$$

Où *norme(a)* contient les qualifications de l'action a.

Vu l'importance de définir les empêchements, on suppose aussi l'existence de la fonction appelée inhibe telle que

$$\text{inhibe} : E\times A \longrightarrow 2^{E}$$

inhibe(e, a) est le sous-ensemble E' de E avec e'$\in$ E' si à chaque fois que e' apparaît durant le délai, après a où e devrait devenir vrai, e n'est plus à vrai dans tous les futurs préférés (e' inhibe l'effet e de l'action a).
Il s'agit maintenant de définir ce que nous entendons par 'est normalement vrai' : cette notion est généralement attachée à une notion d'ordre de préférence. Mais l'utilisation d'une relation de préférence pose la difficulté de trouver le ''bon critère d'ordre''[Mokhtari97b]. L'approche normative de la causalité se base sur une nouvelle définition des *lignes préférées.* Afin de définir cette notion il faut définir la notion de *lignes de temps qui coïncident.*

- deux lignes de temps l_1 te l_2 coïncident jusqu'au point de temps t, notée coïncide(l_1,l_2,t) ssi $\forall$t' t'<t I(l_1,date(t'))= I(l_2,date(t')). Cette définition signifie que toutes les interprétations sur la ligne de temps l_1 jusqu'au point de temps t sont identiques aux interprétations de la ligne de temps l_2 jusqu'au même point de temps t.
- l'ensemble des lignes de temps préférées de la ligne l à la date d_t dénoté par Lp(l,d_t) comme une fonction : Lp : L×R$\rightarrow 2^L$ telle que : ($\forall$l') (l'$\in$ L_P(l, d_t)$\supset$ coïncide(l, l', t)). Lp ne dépend que du passé.

II.4 Implication normale [Mokhtari97a]

Ayant toutes ces définitions, alors la définition de la règle causale a$\Rightarrow$e[Δ]. «une action a implique normalement l'effet e dan un délai Δ» se définit comme suit :

'a$\Rightarrow$e[Δ]' ssi : ($\forall$l,t)(t$\in$l $\supset$(C1$\wedge$C2))

C1 : {[v(a,l,d_t) $\wedge\forall$p(p$\in$norme(a) $\supset$ v(p,l,d_t))] $\supset$($\forall$l')(l'$\in$Lp(l,d_t) $\supset$C11$\vee$C12)}

C11 :{($\exists$t')(t'$\in$l'$\wedge d_t \leq d_{t'} \leq d_t+\Delta\wedge$ v(e,l',d_t))}

C12 :{($\exists$ e',t'')(e'$\in$inhibe(e,a)$\wedge$t''$\in$l'$\wedge$v(e',l',$d_{t''}$)$\wedge d_t \leq d_{t''} \leq d_t+\Delta$)}

C2 : {v($\neg$a,l,d_t)$\supset$($\exists$l')(l'$\in$Lp(l,d_t)$\wedge$nocc(e,l',d_t, Δ))}.

L'action a, est exécutée sous des conditions normales (i.e., toutes les propositions dans *norme(a)* sont vraies) 'e' se produit dans tous les futurs préférés pendant le délai Δ, ou bien il y a eu l'occurrence de l'événement e' tel que (e' inhibe a) après t et dans le délai Δ.
Si l'action a n'est pas exécutée, alors il existe au moins un futur préféré où e n'a pas lieu pendant le délai Δ. Cette condition prend en compte l'idée implicite de contrefactualité toujours présente dans la causalité.
Les règle 'a$\Rightarrow$e[Δ]' «implique normalement» et 'a$\rightarrow$e[Δ]' «implique strictement» sont appelées «règles causales ». Les règles causales sont assemblées dans une base de règles appelée BR. Alors que les liens statiques seront regroupés dans une base de connaissances statiques appelés BS.

En utilisant ces deux opérateurs, les extensions des définitions de ces deux opérateurs aux cas où a et e sont des fbf utilisant les opérateurs $\wedge$, $\vee$ et $\neg$.

- La composition des actions reflète la possibilité de les effectuer simultanément,
- La disjonction, la capacité pour l'agent de choisir parmi elles ;
- La négation, la décision de ne pas l'effectuer.

En conséquence il semble raisonnable d'étendre l fonction *norme* par [Mokhtari97a]:

1. norme (a $\wedge$a')=norme(a)$\cup$norme(a') (pour exécuter a$\wedge$a', l'union de leurs normes doit être vraie).
2. norme(a$\vee$a')=S avec norme(a)$\subset$S$\vee$ norme(a')$\subset$S (S est un ensemble minimal de pré-condition permettant à l'agent d'exécuter au moins l'une des deux conditions), et
3. norme($\neg$a)=$\emptyset$. (Ne pas faire une action ne nécessite pas de pré-conditions spécifiques).

L'extension pour la fonction inhibe [Mokhtari97a] :

1. inhibe(e $\wedge$e',a)=inhibe(e,a)$\cup$inhibe(e',a),
2. inhibe(e$\vee$e',a)=inhibe(e,a)$\cap$inhibe(e',a)(pour bloquer e$\vee$e', il faut un événement qui bloque les deux) et
3. inhibe(e,a$\vee$a')=M avec inhibe(e,a)$\subset$M$\vee$ inhibe(e,a')$\subset$M (si le choix de l'agent est d'exécuter a ou a' entraîne normalement e, le fait de bloquer au moins l'une de ces causalité conduit à ne plus pouvoir garantir que ce choix entraîne e).

Remarques

1. La première extension de la fonction norme donne un résultat contre-intuitif. Par exemple norme(a$\wedge\neg$a)=norme(a)$\cup$norme($\neg$a)=norme(a) si l'on se base sur les définitions 1 et 3 sur norme alors que norme de la contradiction devrait être tout le langage. Pour cela il est imposé de réduire les formules à leur forme la plus simple (une forme canonique où les formules sont dan leur forme normale conjonctive ou disjonctive avec un minimum de symboles).
2. Norme(a$\vee$a')=S avec norme(a)$\subset$S$\vee$ norme(a')$\subset$S. Ayant un certain S on ne nous permet pas de garantir à la possibilité de l'exécution de l'une des deux actions.

Considérant les règles suivantes :

1. si $a_1 \Rightarrow e[\Delta]$ (resp $a_1 \rightarrow e[\Delta]$) $\in$BR et $a_2 \supset a_1$ est un théorème, alors $a_1 \Rightarrow e[\Delta]$ (resp $a_2 \rightarrow e[\Delta]$) $\in$BR.
2. $a \Rightarrow e_1[\Delta]$ (resp $a \rightarrow e_1[\Delta]$) $\in$BR et $e_1 \supset e_2$ est un théorème, alors $a \Rightarrow e_2[\Delta]$ (resp $a \rightarrow e_2[\Delta]$) $\in$BR.
3. (Le ou logique).
   - si $a \Rightarrow e[\Delta] \in$BR et $a' \Rightarrow e[\Delta'] \in$BR, alors $a \vee a' \Rightarrow e[\sup(\Delta,\Delta')] \in$BR.
   - si $a \Rightarrow e[\Delta] \in$BR et $a' \rightarrow e[\Delta'] \in$BR, alors $a \vee a' \Rightarrow e[\sup(\Delta,\Delta')] \in$BR.
   - si $a \rightarrow e[\Delta] \in$BR et $a' \rightarrow e[\Delta'] \in$BR, alors $a \vee a' \Rightarrow e[\sup(\Delta,\Delta')] \in$BR.

La fermeture (BR) est la base qui associe à chaque règle de BR sa clôture par application des règles (1)-(3).

Proposition 1 :

Tout modèle de BR est un modèle de fermeture(BR).(Preuve voir [Mokhtari97a]).

En introduisant cette notion de fermeture(BR) et la proposition précédente, une solution est obtenue pour un des problèmes important du «frame problème » qui est *la Ramification.*
Dans ce qui suit, le traitement du dernier problème lié au «frame problème » et qui est très important du point de vue du raisonnement non-monotone temporel : *La Persistance normale.*

II.4.1 La persistance normale [Mokhtari97b]

Deux problèmes sont à considérer pour introduire cette notion de persistance :

1. le problème lié à la non-monotonie temporelle correspondant à la possibilité qu'un événement externe puisse empêcher e de continuer à être vrai, et
2. le problème lié à la durée de persistance d'un fait (à quel moment ce fait a commencé à être vrai ? Et qu'arrivera-t-il au-delà du délai de persistance ?).

Soit δ la durée normale de e, la persistance se définit comme suit :

$$\text{Persiste}(e,\delta)\equiv[(\forall l,t)((t\in l)\wedge v(e,l,d_t)\wedge(\exists t'')(\ t''\in l\wedge nocc(e,l,d_{t''},d_t\text{-}_{t''}))\supset (\forall l',t')(t'\in l'\wedge(C1\wedge C2\wedge C3)\supset v(e,l',d_{t'})].$$

Où C1,C2, C3 abrègent les conditions suivantes :

$C1 : l'\in Lp(l,d_t)$

$C2 : d_t \le d_{t'} \le d_t+ \delta$

$C3 : (\forall a',\Delta)[a'\Rightarrow\neg e[\Delta]\in BR)\vee(a'\rightarrow\neg e[\Delta]\in BR)\supset nocc(a',l',d_t\text{-}\Delta,d_{t'}+\Delta)]$.

Les formules de persistances sont regroupées dans une table de persistance appelée P. t est le point de temps à partir duquel e devient vrai dans la ligne de temps l.

C1 exprime le fait que la persistance est prédite au moins dans les futurs préférés ; C2 exprime que la persistance dure au moins durant δ, s'il n'y a pas une action a' sur cet intervalle qui puisse l'annuler (cette notion est définit sans faire référence à la fonction inhibe, pour ne pas faire référence à la cause de cet événement).

La persistance d'un événement est relative à son début de réalisation. La durée de persistance varie selon le type d'événement et peut être :

- En nombre d'années, c'est le cas du pistolet qui reste chargé longtemps après son chargement si aucune personne ne le décharge volontairement ou ne tire avec,
- De quelques minutes, c'est le cas de la durée pendant laquelle un chat resterait assis sur un fauteuil.

Les aspects définis jusqu'à présent sont inhérents à toute théorie d'action qui est une notion essentielle pour cette théorie causale puisqu'elle s'intéresse aux causes volontaires.

II.4.2 Rôle de l'agent

Une propriété qui a un rôle déterminant dans la définition de la causalité est la propriété de l'*intention* chez l'agent, que Shoham dans [Shoham89] relie à la notion de «libre arbitre ». Cette notion nécessite la définition de point de choix qui est le point temporel où a eu lieu le choix d'exécution de l'action ou pas (du quel résulte de moins deux différentes lignes de temps, l'une sur laquelle l'action est performée, l'autre pas).

Définition 1 :

Un point de temps t est un point de choix relativement à l'action a entre les lignes de temps l_1, l_2 noté : Pchoix(a, l_1,l_2,t) ssi $(\forall t')(d_{t'} \leq d_t \subset (\text{coïncide}(l_1,l_2,t') \wedge v(a,l_1,d_t) \wedge v(\neg a, l_2,d_t)))$.
L'ensemble BR plus l'ensemble des définitions données jusque là nous permettent de définir la (ou les) cause(s) volontaires a d'un effet e observé sur une ligne de temps l' au point de temps t'.

Définition 2 : Les causes volontaires sont définies comme une fonction partielle :

$$\text{Cause}_v : E \times L \times T \rightarrow 2^A$$

Définie seulement si v(e, l', t') et t' $\in$l' est vrai. Alors Cause_v(e, l', t') est le sous-ensemble A', A' $\subset$A et a$\in$ A' ssi a satisfait soit les contraintes C1-C4 soit C'1, C2, C'3, C4 :

C1 : $(\exists \Delta)$ (a$\Rightarrow$e[Δ]$\in$BR)
C'1 : $(\exists \Delta)$ (a$\rightarrow$e[Δ]$\in$BR)
C2 : $(\exists t, l, l'')$ (pchoix (a, l, l'', t)$\wedge$ $d_t \leq d_{t'} \leq d_t$ +DelaiPertinent), où **Si** $(\exists \delta)$(Persist (e, δ)$\in$P) **Alors** DelaiPertinent=Δ+δ **sinon** DelaiPertinent=Δ.
C3 : l'$\in L_p(l,d_t)$;
C'3 : coïncide (l, l', t)
C4 : $v(\neg e,l,d_t)$.

C1-(C'1) sélectionne l'ensemble des règles causales de BR contenant l'effet e dans leur partie droite, C2 veut dire que l'agent avait (en un temps qui est pertinent pour l'observation de e le choix entre performer ou pas l'action a (libre arbitre), et il a choisit de performer a.
C3-(C'3) dans la ligne de temps l, pour laquelle la ligne de temps l', où l'événement e a été réalisé, était parmi les futurs préférés au moment où le choix a été fait,
C4 spécifie la pertinence de l'action a pour l'événement e qui ne devrait pas être vrai au moment de l'exécution de a dans l.

III. Théorie causale

III.1 Caractérisation de la théorie.

Une théorie causale décrivant un scénario quelconque sur l'action a besoin d'un certain nombre d'informations et de suppositions :

1. Des informations générales décrivant le comportement de certains aspects causaux dans le monde (règles causales), des formules de la forme 'a$\rightarrow$e[Δ]' ou 'a$\Rightarrow$ e[Δ]' qui sont regroupées dans une base notée BR,
2. Des informations sur des situations particulières à des moments précis, des formules de la forme $v([\neg]e,l,d_t)$, ou bien $nocc([\neg]e,l,d_t,\Delta)$,
3. Des informations sur des liens statiques dans le monde, des formules de la forme a$\supset$a' ou e$\supset$e', ces formules sont regroupées une base notée BS,
4. D'une norme associée a chaque action décrite dans BR,
5. De l'ensemble E'$\supset$ E contenant les événements inhibant associés à chaque couple (e,a) d'une règle de BR, et enfin
6. L'ensemble P contenant les délais de persistance δ associés aux événements des règles causales.

Soit DL un ensemble de formules de la forme $v([\neg]e,l,d_t)$ ou $nocc([\neg]e,l,d_t,\Delta)$, c'est description des lignes de temps (un ensemble d'observations du monde à des moments particuliers). Sur cet ensemble un ensemble de supposition est fait :

1. Consistance : toute formule de DL a une interprétation dans un des modèles de BR,
2. Complétude sur les actions : pas d'autres actions que ceux mentionnées dans DL aux dates de leur point de temps,
3. Normalité positive : par défaut toutes les pré-conditions d'une action sont vérifiées,
4. Normalité négative : par défaut il n'y a pas d'événements qui inhibent l'effet d'une action.

Etant donné toutes ces informations et ces suppositions, la théorie causale se définit alors comme suit :

Définition 3 : une théorie causale notée $\mathcal{T}$ est la donnée de l'ensemble des informations décrites ci-dessous.

A l'aide de cette théorie causale nous pouvons

- Expliquer ce qui a pu arriver dans le passé, ou
- Prédire ce qui pourrait arriver dans le futur préféré,
- Traitement du problème de la Ramification par une extension du langage proposé avec d'autres règles causales.

III.2 Explication

L'explication est un raisonnement temporel qui s'effectue par des projections dans le passé. Pour simplifier, la théorie causale considérée est finie, i.e.,

- une description des lignes de temps DL, comportant un ensemble de formules, celles qui sont vraies à des moments précis dans les lignes de temps précises, l'ensemble DL est trié par ordre croissant de points de temps, et
- une description d'une base de règles BR qui est également composée d'un ensemble fini de règles,
- les ensembles BS, norme(a), inhibe(e, a) et P sont aussi finis.

Etant donné cette théorie causale finie $\mathcal{T}$, nous pouvons expliquer ce qui est arrivé dans le passé.

Proposition 2 :

Etant donné une théorie causale finie $\mathcal{T}$, un événement e observé au point de temps t' est le résultat d'un ensemble unique d'actions.

Un algorithme a été proposé pour calculer l'ensemble des causes de l'événement observé e au point de temps t' sur l'ensemble L' des lignes de temps l_i, cet algorithme est correct et son résultat est un ensemble unique.

III.2.1. Algorithme général de l'explication

Données :
Les ensembles DL, BR, BS, pour chacune des actions a de BR, la norme associée, pour chacune des formules de BR, l'ensemble des événements inhibant l'effet de l'action, les durées de persistance de chaque événement et l'événement e observé au point de temps t' sur un ensemble L' des lignes de temps l_i.

Résultat :

L'ensemble A_c des actions effectuées dans le passé et qui sont les causes de l'événement observé.

Début

Etape1 :

{Construction de l'ensemble A_1 composé des couples (cause possible a_i, délai d'apparition de son effet Δ_i) obtenu à partir de BR**}.**
Initialement A_1 est vide ;
{On parcourt BR pour sélectionner les (a_i, Δ_i) des règles ayant e à droite des opérateurs des règles de BR}.
Pour toute les règles i de BR : si $a_i \rightarrow e[\Delta_i] \in BR \vee a \Rightarrow e[\Delta_i] \in BR$ alors $A_1 = A_1 \cup (a_i, \Delta_i)$;

Etape2 :

{Complétude de DL, i.e., on cherche toutes les coïncidences jusqu'au point de temps t et on reporte les informations (les vérités sur les actions et sur les événements) sur les lignes qui coïncident, pour ne considérer qu'une ligne jusqu'en t**}.**
On parcourt DL par ordre croissant des points de temps t_j : si $\exists$ a, l tel que v(a,l, d_{tj}) alors
Pour toutes les lignes l_k tel que coïncide(l,l_k,t_j) faire :
Reporter toutes les informations vraies dans l_k et non mentionnées dans l, dans l à partir de t_0 jusqu'au t_j (d_{t0} est la date minimale apparaissant dans DL (qui est finie)).

Etape3 :

{Construction de l'ensemble A_c des causes de l'événement e en t'_j**}.**
Initialement A_c est vide ;
Parcourt de DL par ordre décroissant des points de temps à partir de $d_{ti} \leq d_{t'}$
Si $\exists a_i$ $a_i \in A_1 \wedge v(a_i, l_i, d_{ti}) \in DL \wedge d_{ti} < d_{t'} < d_{ti} + \Delta_i$ et Precond(a_i,l_i,d_{ti})et Postcond(a_i,e,l_i,d_{ti}) alors
$A_c = A_c \cup (a_i, l_i)$;
Fnsi ; finpour
$d_{ti} := d_{ti} - 1$
Jusqu'à $d_{ti} < d_{t0}$.

Fin .

Precond(a_i,l_i,d_{ti}) ssi $\forall$e', t'' $d_{t0} \leq d_{t''} \leq d_{ti} \wedge [(v(e', l_i, d_{t''}) \in DL \supset \neg e' \notin norme(a_i)) \vee \in norme(a_i) \wedge persiste(e', \delta) \wedge \delta > d_{ti} - d_{t''})] \vee [\forall (d,d') \wedge d \leq d_{ti} \wedge d' \geq d_{t''} \wedge nocc(e', l_i, d_{t''}, (d-d')) \in DL \supset e' \notin norme(a_i)]$
{Tout événement e' vrai dans DL entre t_0 et t_i ne doit pas être contraire à la norme, et s'il appartient à la norme il doit rester vrai jusqu'en t_i. Si maintenant on sait qu'il n'est pas vrai dans DL (nocc..) alors il ne doit pas faire partie de la norme.}
Postcond(a_i,e,l_i,d_{ti})ssi $\forall$ t'' $d_{ti} \leq d_{t''} \leq d_{t'} \wedge [(v(e', l_j, d_{t''}) \in DL \supset e' \notin inhibe(a_i, e)) \vee ((v(e', l_j, d_{t''}) \in DL \wedge (\neg e' \in inhibe(a,e)) \supset persiste(e', \delta) \wedge \delta > d_{t''} - d_{ti}] \vee [\exists (d,d') \wedge d \leq d_{ti} \wedge d' \geq d_{t''} \wedge nocc(e', l_i, d_{t''}, (d-d')) \in DL \supset \neg e' \notin inhibe(a,e)].\}$

{ si un événement e' est vrai entre le moment de l'effectuation de l'action et l'apparition de e alors soit il ne doit pas faire partie des événements inhibants ou bien c'est ¬e' qui fait partie de inhibe, et e' doit persister jusqu'en t_i. Si maintenant on sait que e' est faux jusqu'en t_i, alors ¬e' ne doit pas faire partie des inhibants de (a, e).}
L'algorithme s'arrête et calcule bien le résultat attendu, i.e., fournit un ensemble unique d'actions préférées causes de e.

III.2.2. Complexité

Si - n est la taille de DL,

- m est la taille de BR, et
- p est la taille de BS.

L'algorithme général est en **$O(mp+2n^2)$**.

III.3 Prédiction

La prédiction est également un raisonnement temporel non-monotone qui s'effectue par des projections. Mais les projections se font cette fois vers l'avant. Le but est de prédire les évolutions des lignes de temps dans les futurs préférés si une action est effectuée.

Proposition 3 :

Etant donné une théorie causale finie $\mathcal{T}$, et une action *a* effectuée en un point de temps t, il en résulte, dans les futurs préférés et dans un délai fini, un ensemble fini d'effets de cette action.

III.3.1 Algorithme de la prédiction

Un algorithme a été proposé pour calculer l'ensemble des effets de l'action *a* effectuée au point de temps t, cet algorithme est correcte et son résultat est un ensemble fini d'événements, les effets préférés de l'action *a*.

Données
Les ensembles DL, BR, BS, pour chacune des actions a de BR, la norme associée, pour chacune des formules de BR, l'ensemble des événements inhibant l'effet de l'action, les durées de persistance de chaque événement et l'action a applicable en un point de temps t sur l'ensemble L' des lignes de temps l_j qui passent par ce point.

Résultats :
E_c un ensemble fini d'effets de l'action *a,* observés dans les futurs préférés des lignes de temps où *a* est effectuée.

Début :

Etape1 :
{ Construction de l'ensemble E_1 des couples, (effet possible e, délai d'apparition Δ de cet effet) obtenus à partir de BR et de BS.**}**
Initialement E_1 est vide ;
{On parcourt BR pour sélectionner les (e_i,Δ_i) des règles de BR ayant *a* comme prémisse}
Pour toutes les règles i de BR : si $a \Rightarrow e_i[\Delta_i] \in BR \vee a \rightarrow e_i[\Delta_i] \in BR$ alors $E_1 = E_1 \cup (e_i,\Delta_i)$
Posons $\Delta = \sup[\Delta_i]$(Δ va servir comme borne supérieure pour l'itération de recherche des effets de a).

Etape 2 :

{de la même manière que l'explication, nous saturons DL par la connaissance des coïncidences**}**

Etape 3 :

{ Construction de E_c des effets de l'action a en t**}**

Initialement E_c est vide ;

{Construction de E_0 correspondant à l'ensemble des événements observés (vrai dans DL) jusqu'au moment de l'effectuation de l'action.}

$E_0=\{e_j$ tel que $v(e_j,l_j,d_{tj})\in DL \wedge e_j \in E \wedge persite(e_j,\delta_j) \wedge \delta_j \geq t\text{-}t=j\}$;

Parcourt de DL par ordre croissant des points de temps à partir de $t_i \geq t$

Pour chaque ligne $l_i \in L$' vérifier si $\exists e_i$, $e_i \in E_1 \wedge v(e_i,l_i,d_{ti}) \in DL \wedge d_{ti} < d_t+\Delta_i \wedge (e_i \notin E_0 \vee \neg e_i \in E_0\}$

{Cette dernière condition exprime le fait qu'il y a un changement du à l'action a, i.e., e n'était pas vrai avant l'effectuation de l'action.}et Postcond(a_i,e_i,l_i,d_{ti}) alors $E_c=E_c \cup \{e_i,l_i\}$ finsi.

finpour

$d_{ti} := d_{ti}+1$

Jusqu'à $d_{ti}=d_t+\Delta$.

Fin.

III.3.2. Complexité :

De la même manière que l'explication, si n est la taille de DL et m est la taille de BR et p est la taille de BS, la complexité de l'algorithme est en **$O(mp+2n^2)$.**

IV. Ramification [Khelfallah01]

Le problème de la ramification a éveillé l'intérêt des chercheurs vers la fin des années 90's. Il a été défini, dans le contexte de raisonner au sujet d'actions, comme le problème de décrire tous les effets indirects d'actions. Par les effets indirects nous signifions les effets du débordement qui paraissent après avoir exécuté une action. Plusieurs méthodes ont été proposées dans la littérature, la plupart d'eux ont utilisé les contraintes de domaine, qui sont des formules qui représentent des dépendances statiques qui existent entre les composants mondiaux. Ces formules sont vérifiées dans chaque état valide du monde. L'usage de contraintes de domaine n'est pas suffisant pour produire les effets indirects attendus. La solution proposée permet de calculer tous les effets indirects d'une action en utilisant une relation appelée «rapport des effets indirects». Ces rapports sont inspirés des rapports de causalité de Thielscher [Thielscher97], qui se basent sur les contraintes de domaine et les renseignements d'influence. Cependant la méthode proposée par Mokhtari [Mokhtari97b][Khelfallah01] diffère par la notion du 'délai de l'effet indirect'. De plus, les rapports des effets indirects sont appliqués selon l'ordre de génération des effets.

La méthode normative de causalité est basée sur le concept interventionniste de causalité où un agent a le choix pour exécuter (ou pas) une action (volonté libre). C'est une approche permettant un raisonnement non-monotone basé sur l'utilisation des Normes, parmi ces principes : «une action peut en causer un ou plusieurs effets » et « aucun effet ne précède sa cause ».

La représentation de la notion de volonté libre a exigé une structure de temps avec un branchement dans le futur. Un branchement dans le passé a été aussi exigé pour examiner les différents événements qui mènent à la même situation [Kayser98].

IV.1 Ramification dans la méthode normative de causalité [Khelfallah01]

Pour calculer les effets indirects, des relations dirigées entre deux effets appelées 'rapports des effets indirects' ont été introduites.

Définition 4 :

"Implication indirect" noté $\longmapsto$, se définit comme suit :
« $(e_1,e_2)\longmapsto e[\Delta]$ » : exprime que l'occurrence du fait e_1, dans une situation où la formule e_2 est vraie, ceci implique indirectement l'occurrence de l'événement e dans un délai Δ. Formellement l'implication indirecte est définie comme suit :

$$(e_1,e_2) \longmapsto e[\Delta] : \text{Ssi } (\forall l,d)(occ(e_1,l,d)\wedge v(e_2,l,d)\supset(\exists d')(d\leq d'\leq d+\Delta\wedge occ(e,l,d'))).$$

Le rapport des effets indirects est exprimé par la règle '$(e_1,e_2) \longmapsto e[\Delta]$'
Tel que : e_1 et e sont des effets littéraux, e_2 est une formule d'effet, Δest un réel.
Si $\Delta=0$, l'effet indirect généré est instantané. Les rapports des effets indirects sont rassemblés dans une base notée BEI.
Soit : $D = \{v(e,l,d): e\in FOR(E)\wedge l\in L\wedge d\in R\}$ et $F=\{occ(e,l,d) : e\in FOR(E)\wedge l\in L\wedge d\in R\}$.

Définition 5 :

La fermeture d'un ensemble d'observations DL est définie par la fonction:

$$Closure: 2^{D\cup F}\rightarrow 2^{D}.$$

Cette fonction permet d'avoir une connaissance complète du monde à partir d'un ensemble d'observations. Intuitivement, DL contient les informations pertinentes et la fermeture de DL «Closure(DL)» représente toutes les informations que nous pouvons déduire de DL en exploitant les délais de persistances.

Soit $DL\in 2^{D\cup F}$ un ensemble d'observations, Closure(DL) est définie par :

1. $\forall\ e\in LIT(E),l\in L\ ,d\in R : v(e,l,d)\in Closure(DL)$ Ssi l'une des conditions suivantes est vérifiées :
   i. $v(e,l,d)\in DL$,
   ii. $\exists d',\delta[\delta>d-d'\wedge v(e,l,d')\in Closure(DL)\wedge persist(e,\delta)\wedge \forall d''(d'\leq d''<d\supset v(\neg e,l,d)\notin Closure(DL))]$,
   iii. $\exists l'(l'\in L\wedge v(e,l',d)\in Closure(DL)\wedge Coincide(l,l',d))$,
   iv. $occ(e,l,d)\in Closure(DL)$.
2. $\forall\ e\in FOR(E)\ ;l\in L\ ;d\in R : v(e,l,d)\in Closure(DL)$ Si $v(e,l,d)\in DL$,
3. $\forall\ e_1,e_2\in FOR(E)\ ;l\in L\ ;d\in R :\ (v(e_1\wedge e_2,l,d)\in Closure(DL)$ Si$(v(e_1,l,d)\in Closure(DL)$ and $v(e_2,l,d)\in Closure(DL))$.

Point (1) traite seulement le cas d'effet littéral. Intuitivement, condition(1.i) signifie que si un effet est vrai dans DL il sera aussi dans Closure(DL). (1.ii) signifie que si un effet est vrai dans DL, il sera vrai dans Closure(DL), durant sont délai de persistance, sauf si le contraire se produit. (1.iii) copies dehors contenu de la ligne sur toutes les lignes qui coïncident avec lui (1.iiii) utilise la relation entre occ et v.
Point (2) et (3) traitent le cas des formules d'effets (l'utilisation d'opérateurs logique peut être transformée par des formules contenant seulement l'opérateur $\neg$ et $\wedge$).

Le calculer DL ne dépend que des actions exécutées, puisqu'elles sont les seules qui peuvent changer l'état du monde, donc,

Définition 6 :

Une action *a* est exécutée dans une ligne de temps *l* dans une date *t*. L'ensemble des effets (directes et indirects) générés sont regroupés dans un ensemble EG.

Le contexte d'exécution d'une action a est définie par le tuple < DL,l,d,EG> tel que :

- DL est l'ensemble des formules v(e,l,d) : les observations sur l'état du monde quand l'action a est exécutée,
- $l \in L$, est la ligne de temps ou a est exécutée,
- $d \in R$,la date d'exécution de a.

EG est l'ensemble des formules de la forme occ(e,l,d), il représente les effets générés dans la ligne de temps l à la date d.

La génération d'effets de l'action a consiste à appliquer séquentiellement les règles de causalité et les rapports des effets indirects associés à l'action a. L'ordre de leurs applications n'est pas arbitraire. Il dépend de la date de la génération des effets, i.e., une règle ou un rapport est appliqué si est seulement si tous les effets qui devraient se produire avant son effet avaient été produits. D'après ce critère, les effets sont produits dans l'ordre exact de leur apparence. La génération s'arrête si aucuns nouveaux effets ne sont produits.

Définition 7 :

La règle $a \Rightarrow e[\Delta]$ est *applicable* dans le contexte <DL,l,d,EG> avec d_e est la date d'apparence de l'effet e si les conditions suivantes sont vérifiées :

1. v(a,l,d),
2. $d_e = d+\Delta$,
3. $\forall p(p \in Norm(a) \supset v(p,l,d) \in Closure(DL \cup EG))$,
4. $(\forall e')(e' \in Inhibit(e,a) \supset (\forall d')(d \leq d' \leq d+\Delta : v(e',l,d') \notin Closure(DL \cup EG)))$,
5. $(a \Rightarrow e'[\Delta] \in BR \{respectivement, ((e_1,e_2) \longmapsto e'[\Delta]) \in BEI\}$ Applicable dans le contexte <DL,l,d,EG> avec d' la date d'apparence de l'effet $e' \supset d' \geq d$,
6. Cette règle n'est pas toujours appliquée dans le même contexte.

Définition 8 :

La règle$(e_1,e_2) \longmapsto e[\Delta]$ est applicable dans le contexte <DL,l,d,EG> avec d_e la date d'apparence de e, si les conditions suivantes sont vérifiées:

1. $(\forall l',d')(l' \in Lp(l,d) \wedge d \leq d' \supset occ(e,l',d') \notin EG)$,
2. $(\exists l',d')(l' \in Lp(l,d) \wedge d \leq d' \wedge occ(e_1,l',d') \in EG \wedge v(e_2,l',d') \in Closure(DL \cup EG))$,
3. $d_e = d+\Delta$,
4. $\forall (a \Rightarrow e'[\Delta]) BR \{respectivement, ((e_1,e_2) \longmapsto e'[\Delta] \in BEI\}$ Applicable dans le contexte <DL,l,d,EG> et d' la date d'apparence de e $\supset d' \geq d$.

IV.2 Génération systématique des rapports des effets indirects

Une extension de la méthode normative proposée dans [Khelfallah01] calcule avec succès les actions de ramifications (i.e., ne génère pas les effets inattendus) grâce à un ensemble adéquat de rapports des effets indirects. Ces derniers ne sont pas écrits par un système de raisonnement mais ils sont produits automatiquement à partir des contraintes de domaine et des informations d'influences. Les contraintes de domaine sont des formules d'effets qui représentent des dépendances statiques qui existent entre les effets du domaine. En plus du fait qu'ils sont concis, ils sont écrits de manière naturelle et facile. Les contraintes de domaine sont rassemblées dans un ensemble noté CD.

Définition 9 :

Les informations d'influence sont définis sur LIT(E)*LIT(E)*R, tel que $(e_1,e_2,\Delta)\in I$, elles expriment que les changements de la valeur de $e_1$change potentiellement la valeur de e_2 dans un délai maximum Δ. L'ensemble de toutes les informations d'influences est noté I.

IV.3 Algorithme de génération automatique des rapports des effets indirects [Khelfallah01]

Input : ensemble des informations d'influences I ;

La forme normale disjonctive $D_1\wedge\ldots\wedge D_n$ de la conjonction des contraintes de domaine de CD, CNF(CD).

Out put : L'ensemble BEI des rapports des effets indirects.

Début

Initialement, BEI := {};

Pour chaque contrainte $D_i=e_1\vee\ldots\vee e_{mi}\in CNF(CD),i=1,\ ,n$

Pour $j=1,\ldots,m_i$

Pour $k=1,\ldots,m_i,k\neq j$

Si $(\neg e_j,e_k,\Delta)\in I$, Ajouter à BEI la règle $(\neg e_j, \underset{l=1,\ldots,m_j,\ l\neq j,l\neq k.}{\wedge} \neg e_l) \longmapsto e_k[\Delta]$.

Fin.

Cet algorithme est une extension de l'algorithme de Thielscher pour la génération automatique de rapports de causalité [Thielscher97]. La première extension consiste dans l'introduction de la notion de délai dans les informations d'influences qui a permit de maintenir les effets indirects. La deuxième extension, est que les informations d'influences sont définies sur l'ensemble des effets littéraux au lieu de l'ensemble des effets [Castilho99].

Il y' a quelques problèmes soulevé dans cette approche, certains d'ordre techniques, d'autres d'ordre conceptuels. Nous citons par exemple [Mokhtari97]:

Un problème technique est souligné dans l'algorithme d'explication (seules les formules atomiques sont considérées) dans les cas des formules non-atomiques, cet algorithme est insuffisant ; il faudrait le coupler avec un démonstrateur de théorèmes ; aucune idée n'est donnée sur la complexité de l'élaboration de ce «couplage ».

Un autre problème moins technique concerne la partition de P en des actions A et des événements E. Cette partition n'est pas une question triviale. Dans des exemples de la vie réelle, il n'est pas toujours évident de décider, précisément, quelle proposition rentre dans quelle catégorie.

Enfin, un problème d'ordre conceptuel concerne les actions à effets aléatoires. C'est le cas, par exemple, du résultat non déterministe de l'action «lancer une pièce », qui est pile ou face.

La méthode normative permet de traiter les effets retardés. Mais elle ne permet pas de traiter les actions avec des effets non déterministes ni les actions complexes. Un sous-ensemble d'actions concurrentes est traité, c'est l'ensemble des actions avec des effets indépendants (la seule condition, est la non-existence d'actions qui génèrent des effets opposés qui s 'exécutent simultanément).

V. Conclusion

Nous avons présenté dans ce chapitre un formalisme pour une représentation de la causalité dont le langage sous-jacent est un langage propositionnel étendu à un langage *L* à l'aide de deux opérateurs, l'opérateur '$\rightarrow$' pour «implique dans tous les cas » et l'opérateur '$\Rightarrow$' pour «implique normalement», et d'un certain nombre de prédicats et de fonctions. Ces fondements sont nécessaires pour une approche sur l'inférence non-monotone.

Cette dernière est modéliser par une certaine relation de normalité entre les mondes qui s'articule essentiellement autour de trois idées clés qui se rattache au «frame problème» [McCarthy69]. Il s'agit de :

- La supposition d'une norme correspondant à un sous-ensemble de pré-conditions jugé pertinent dans le point de vue où on se place (problème qualification),
- La possibilité qu'un événement explicitement mentionné puisse inhiber l'effet d'une action,
- La supposition qu'un fait donné persiste jusqu'à ce qu'une action externe l'en empêche ou qu'il cesse naturellement (problème de la persistance des faits).

Ce formalisme a été utilisé pour définir une théorie causale d'explication, qui s'applique aussi bien à la prédiction qu'à l'explication à l'aide d'algorithmes généraux d'explication causale et prédiction causale. Ces algorithmes sont corrects et calculables (preuve voir [Mokhtari97b]).

Le chapitre suivant sera dédié à une approche récente proposée par Halpern et Pearl[Halpern02][Halpern05]. Les auteurs proposent une modélisation contre-factuelle de la causalité, via un modèle d'équations structurelles. Ces modèles peuvent être représentés graphiquement par des réseaux causaux en s'inspirant de réseaux causaux bayésiens [Pearl00][Halpern05]. La définition d4e la causalité dans ce cadre reste étroitement liée à l'idée de conditionnelle "contre-factuelle" (c'est-à-dire que "A cause B" dans la mesure où "il est vrai que si A n'avait pas eu lieu, B ne se serait pas produit"). Cette approche offre un modèle raisonnable de l'idée de causalité et permet de traiter des exemples qui posaient problèmes aux approches précédentes.

I. Introduction

Que signifie qu' "un événement C est la cause réelle de l'événement E?» La question de la définition de "cause réelle" a été depuis longtemps l'objet de spéculations de la part des philosophes.

Un exemple typique [Wright88], est de considérer deux incendies d'une maison. Si l'incendie A détruit la maison avant l'incendie B, on va considérer que l'incendie A est la "cause réelle" des dommages, même si la maison aurait pu être totalement détruite par l'incendie B, si l'incendie A n'avait pas lieu. La notion de "cause réelle" est une notion très importante dans les applications d'IA et c'est une notion nécessaire pour l'explication. La génération automatique des explications est une tache essentielle dans la planification, le diagnostique, pour cela nous avons besoin d'une analyse formelle du concept de "cause réelle".

Le problème de définition de la causalité avec la notion de dépendance contre-factuelle a été initialement étudié par [Hume1739] qui défini la causalité comme suit :

> *On peut définir la cause comme étant un objet suivi d'un autre,…,tel que, si le premier n'avait pas eu lieu, le second n'aura pas eu lieu.*

Dans ce chapitre nous présentons une nouvelle approche contre-factuelle de la relation causale, nous commençons par présenter le modèle de cette approche structurelle[Halpern02][Halpern05], puis une définition préliminaire de la causalité est donnée selon ce modèle. Afin de bien clarifier cette approche, quelques exemples bien connus dans la littérature sont traités. Dans la quatrième section un raffinement de la définition précédente est donné ainsi que d'autres exemples qui rendent compte de cette définition. L'avant dernière partie de ce chapitre sera consacré à une comparaison qui montre les points communs et les divergences entre cette approche et l'approche normative de la causalité et nous clôturons le chapitre par une conclusion.

Il existe grand nombre de travaux en IA sur la théorie des actions [Lin95][Sandwall94] [Reiter01], relatifs à la manière d'introduire les relations causales dans la base de connaissances, pour guider l'action. Le but de ce travail est assez différent. Il consiste en l'extraction de la cause réelle à partir de telles bases de connaissances couplée d'un scénario spécifique.

II. Modèle causal

Une voie d'étude très intéressante pour le lien causal est la logique contre-factuelle. Cette logique proposée par [Lewis86] utilise une variante de l'implication stricte pour représenter la dépendance entre événements. Un événement **e** est causalement dépendant d'un événement **e'** si "le fait que **e** a eu lieu ou pas dépend contre-factuellement du fait que **e'** a lieu ou pas". Cette notion de dépendance contre-factuelle continue à recevoir une grande attention. Lewis a donné une formulation moins appropriée des dépendances contre-factuelles dans la sémantique du monde possible et a fait le lien entre de la cause réelle et la fermeture transitive des dépendances contre-factuelles. **C** est une cause réelle de **E** ssi **C** est liée à **E** par une chaîne d'événements, chacun dépend de ses prédécesseurs. Cependant la théorie de dépendance de Lewis a présenté quelques difficultés. En effet, les effets ne sont pas toujours dépendants contre-factuellement de leurs causes, ni directement ni indirectement, voir l'exemple précédent. De plus, la causalité n'est pas toujours transitive, comme le prétend Lewis[Lewis86].

Une idée récente de J. Y Halpern et J. Pearl [Halpern05] basée sur la définition de la cause réelle utilisant les équations structurelles, est en fait une extension de la notion de dépendance contre-factuelle afin de permettre " la dépendance sous certaines contingences". En d'autres termes, puisque les effets ne sont pas toujours dépendants contre-factuellement de leurs causes dans la situation actuelle, ils y dépendront sous certaines contingences. Dans le cas des deux feux, par exemple, la destruction complète de la maison dépend du feu A sous la contingence que les flammes du feu détruisent la maison à n'importe quel moment entre l'arrivée réel du feu A et celle du feu B. Sous cette contingence, si le feu A n'avait pas eu lieu, la maison n'aurait pas été détruite complètement. La destruction complète de la maison dépend aussi du feu A sous la contingence que le feu B n'ait pas commencé. Mais ceci nous amène à un état évident : la destruction complète dépend aussi du feu B sous la contingence que le feu A n'ait pas commencé. Cette dernière contingence n'est pas désirée, et le modèle proposé permet seulement les contingences qui n'interfèrent pas avec le processus causal actif.

Selon cette approche, la vérité de chaque affirmation doit être évaluée relativement à un modèle particulier du monde; Donc, on permet, selon cette définition de dire que " **C** cause **E**" dans un modèle structurel particulier (un contexte particulier). Il est possible de construire deux modèles structurels tel que dans le premier **C** cause **E** alors que dans le deuxième **C'** cause **E.** Dans ces conditions le modélisateur joue un rôle très important dans le choix des variables et des événements sur lesquels il va travailler et dans le choix du modèle du monde qui permet une meilleure représentation.

Dans la partie suivante, nous présentons la définition de base du modèle causal, en terme d'équations structurelles, ainsi que syntaxe et la sémantique du langage pour le raisonnement sur la causalité.

II.1 Modèle causal [Halpern05][Halpern01][Halpern02]

L'idée de base est que le monde est décrit par des variables aléatoires, et qu'on s'intéresse à leurs valeurs. Si X est une variable aléatoire, un événement typique a la forme X=x (En terme de mondes possibles, cela représente l'ensemble des mondes possibles où X prend la valeur x). Certaines variables aléatoires peuvent avoir une influence causale sur d'autres. Cette influence est modélisée par un ensemble d'équations structurelles. Chaque équation représente un mécanisme distinct (une loi) dans le monde, chacun peut être modifié (par une action externe) sans altérer les autres. Les variables aléatoires sont divisées en deux ensembles, les variables **exogènes,** leurs valeurs sont déterminées par des facteurs hors du modèle, et les variables **endogènes,** ce sont ces variables qui sont évaluées par des équations structurelles.

Formellement, une signature S est un tuple (U,V,R) où U est un ensemble de variables exogènes, V est un ensemble de variables endogènes (on suppose que V est un ensemble fini), R associe à chaque variable $Y \in U \cup V$ un ensemble $R(Y)$ non-vide de valeurs possibles de Y.

Un modèle causal (modèle structurel) selon une signature S est un tuple M=(S,F), où F associe à chaque variable $X \in V$ une fonction dénotée F_X tel que :

$$F_X : (\times_{U \in U} R(U)) \times (\times_{Y \in V - \{X\}} R(Y)) \rightarrow R(X)$$

F_X détermine la valeur de X ayant les valeurs de toutes les autres variables dans $U \cup V$. La fonction F définit un ensemble d'équations structurelles (modifiables), reliant les valeurs des variables.

Comme F_X est une fonction, il y a une valeur unique de X une fois qu'on a établi les valeurs des autres variables. Notons que nous disposons de la fonction F que pour les variables endogènes. Les valeurs des variables exogènes étant données, et c'est leurs effets sur les variables endogènes (et les effets de variables entre elles) que les auteurs tentent de modéliser par des équations structurelles.

L'interprétation contre-factuelle et l'asymétrie causale associées aux équations structurelles sont plus claires lorsqu'on considère les interventions externes (ou bien les changements spontanés), sous lesquels certaines équations sont modifiées. Une équation tel que $x=F_X(\vec{u}, y)$ signifie que dans un contexte où les variables exogènes ont les valeurs du vecteur $\vec{u}$, si Y prend la valeur y d'une manière quelconque i.e., une manière non spécifiée dans le modèle, alors X devrait avoir la valeur x, tel que cela est dicté par F_X. Cela ne peut pas se produire si on intervient directement sur X ;

Exemple 2.1.1

Supposons que l'on veuille raisonner sur un incendie de forêt. Celui-ci aurait pu être causé par un éclair ou par le fait qu'un pyromane ait allumé une allumette. Dans ce cas le modèle causal aurait les variables endogènes suivantes (ou peut être d'autres !).

- F : pour décrire le feu (1 si le feu a eu lieu, 0 si non).
- L : pour décrire l'éclair (1 si l'éclair a eu lieu, 0 si non).
- ML : pour allumette allumée (1 si elle a eu lieu, 0 si non).

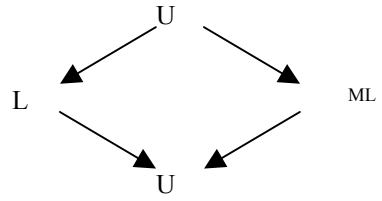


Figure1 : Réseau causal de l'exemple

L'ensemble *U* des variables exogènes inclut les conditions suffisantes pour rendre toutes les relations déterministes (le bois est sec, il y a suffisamment d'oxygène,…). Supposons que $\vec{u}$ est l'attribution de valeurs des variables exogènes qui fait que l'incendie de la forêt soit possible. Alors, par exemple, $F_F(\vec{u}$,L,ML) est tel que F=1 si soit L=1 ou ML=1. Notez que bien que F dépende des valeurs de L et ML, la valeur de L ne dépend pas des valeurs de F et ML.

Un modèle causal dispose des ressources pour déterminer les effets contre-factuels. Etant donné un modèle causal M=(*S,F)*, un vecteur $\vec{X}$ (éventuellement vide) de variables dans *V*, et des vecteurs de valeurs ($\vec{x},\vec{u}$) des variables dans $\vec{X}$ et U, respectivement, on peut définir un nouveau modèle causal dénote $M_{\vec{X}\leftarrow\vec{x}}$ sur la signature $S_{\vec{X}}$=($U,V-\vec{X}$,$R|_{V-\vec{X}}$). $M_{\vec{X}\leftarrow\vec{x}}$ est appelé un sous-modèle de M [Pearl00]. Intuitivement, c'est un modèle causal qui résulte quand les variables dans $\vec{X}$ sont associées à $\vec{x}$ par une certaine action externe qui affecte uniquement les variables dans $\vec{X}$; on ne modélise ni l'action ni leurs causes explicitement.

$R|_{V-\vec{X}}$ Est une restriction de *R* aux variables dans, *V*- $\vec{X}$.

Formellement $M_{\vec{X}\leftarrow\vec{x}}$ =($S_{\vec{X}}, F^{\vec{X}\leftarrow\vec{x}}$), ou $F_Y^{\vec{X}\leftarrow\vec{x}}$ est obtenu à partir de F_Y en attribuant la valeur $\vec{x}$ aux variables dans $\vec{X}$.

Par exemple si M est le modèle causal associé à l'exemple précédent, alors le modèle $M_{L=0}$ a l'équation F=ML. L'équation pour F dans $M_{L=0}$ n'impliquera plus L, plutôt, elle est déterminer en attribuant à L la valeur 0 dans l'équation de F dans M. De plus il n y'aura plus d'équation pour L dans $M_{L=0}$.

Ce raisonnement n'est pas circulaire. Le but n'est pas de réduire la causalité à des concepts non-causaux, mais d'interpréter les questions sur les causes d'événements spécifiques dans un scénario spécifié en terme de connaissances causales génériques tel que ce que nous pouvons obtenir à partir d'équations physiques. Le modèle causal représente un fond de connaissances sur les tendances de certains types d'évènements à causer d'autres types d'évènements. On utilise les modèles pour déterminer les causes d'évènements, tel que si c'était un pyromane qui a causé l'incendie de la forêt le 10 Juin 2004, étant donné ce qui est connu ou supposé sur ce feu particulier.

Pour des raisons de simplicité, nous n'allons nous restreindre uniquement aux équations récursives (acycliques). Ceci est un cas particulier où il y a un certain ordre total « < » sur des variables dans *V* tel que si X<Y, alors F_X est indépendant de la valeur de Y. $F_X(\ldots,y,\ldots)=F_X(\ldots,y',\ldots)$ Pour tous y, y'$\in$R(Y). Intuitivement, si une théorie est récursive, il n'y a pas de retour arrière. Si X<Y, alors la valeur de X peut affecter la valeur de Y, mais la valeur de Y n'a aucun effet sur la valeur de X.

On peut décrire un modèle causal M en utilisant un réseau causal[Halpern01][Halpern05]. C'est un graphe avec les nœuds correspondants aux variables aléatoires dans *V* et tel qu'un arc va d'un nœud X à un nœud Y si F_Y dépend de la valeur de X. Ce graphe est un DAG(Directed Acyclique Graphe). Acyclique du fait que les équations sont récursives. Intuitivement, les variables peuvent avoir un effet causal uniquement sur leurs descendants dans le réseau causal. Ces réseaux causaux, qui sont similaires aux réseaux bayésiens [Pearl88] déjà présenté dans le chapitre 3, représentent des relations fonctionnelles arbitraires, néanmoins que la nature exacte des fonctions est spécifiée dans les équations structurelles alors qu'elle n'est pas représentée dans le diagramme.

Les équations structurelles englobent toutes les informations nécessaires pour le raisonnement causal, incluant toutes les informations sur les croyances, les causalités, les interventions et les comportements contre-factuelles

Il y a plusieurs décisions non-triviales à faire pour effectuer le choix du modèle causal, les variables exogènes dans certaines mesures représentent le fond de la situation, que l'on veut prendre en considération. D'autres suppositions implicites sont représentées dans les équations structurelles, qui peuvent représenter quelques mécanismes causaux. Il n'est pas facile de décider ce qui est le «bon» modèle causal étant donnée une situation, ni de quel est le meilleur des deux models[Halpern01][Halpern05].

II.2 Syntaxe et sémantique

Pour que la définition de la cause réelle soit précise, il est utile d'avoir une logique avec une syntaxe formelle. Ayant une signature S=(U,V,R). Une formule de la forme X=x, pour X$\in$V et x$\in$R(X), est appelée « événement primitif ». Une formule causale de base (à travers S) est de la forme $[Y_1 \leftarrow y_1 ,..,Y_k \leftarrow y_k]\varphi$, où

- φ est une combinaison d'événements primitifs ;
- $Y_1,..Y_k$ sont des variables distinctes de V ;
- $y_i \in R(Yi)$.

Une abréviation de cette formule est $[\vec{Y} \leftarrow \vec{y}]\varphi$. Si k=0 alors l'abréviation sera φ. Intuitivement $[Y_1 \leftarrow y_1, .., Y_k \leftarrow y_k]\varphi$, Signifie que φ est vraie dans le monde contre-factuel qui apparaît si $Yi = y_i$, i=1,..,k.

Une « formule causale » est une combinaison de formules de base. Une formule causale ψ est vraie ou fausse dans un modèle causal selon un contexte. On écrit . $(M, \vec{u}) \models \psi$, si ψ est vraie dans le modèle causal M et étant donné le contexte $\vec{u}$. $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]$ (X=x) si la variable X à la valeur x dans l'unique solution aux équations dans $M_{\vec{Y} \leftarrow \vec{y}}$ dans le contexte $\vec{u}$. Notez que les équations structurelles sont déterministes.

II.3 Définition de la cause

Sur la base de ces notations, on peut donner une définition préliminaire de la cause réelle. On veut donc un sens hors des expressions de la forme « un événement A est la cause réelle de l'événement φ (dans le contexte $\vec{u}$) ». Le contexte représente le fond des informations et les équations structurelles représentent le fond des connaissances. Ayant le contexte (tous les événements pertinents sont connus) et les équations structurelles, la seule question est de savoir lesquels sont les causes de φ ou alternativement, tester si un ensemble donné d'événements peut être considéré comme cause de φ.

Les types d'événements qui peuvent être les causes réelles ont la forme $X_1 = x_1 \wedge \ldots \wedge X_k = x_k$. C'est une conjonction d'événements primitifs ; une abréviation typique est $\vec{X} = \vec{x}$. Les événements qui peuvent être causés sont des combinaisons booléennes arbitraires d'événements primitifs. Une généralisation doit se faire pour permettre les causes disjonctives.

Pour la causalité on considère que le modèle structurel et tous les effets pertinents sont connus, et que la seule définition raisonnable de « A ou B cause φ » semble être « soit A cause φ ou B cause φ ».

Définition 2.3.1

$\vec{X} = \vec{x}$ est une cause réelle de φ dans $(M, \vec{u})$ si les trois conditions suivantes sont vérifiées :

- AC1 : $(M, \vec{u}) \models \vec{X} = \vec{x} \wedge \varphi$, ($\vec{X} = \vec{x}$ et φ sont vrais dans le monde réel) ;
- AC2 : il existe une partition $(\vec{Z}, \vec{W})$ de *V*, avec $\vec{X} \subseteq \vec{Z}$ et une attribution $(\vec{x}', \vec{w}')$ des variables dans $(\vec{X}, \vec{W})$ tel que si $(M, \vec{u}) \models Z = z^*$ pour $Z \in \vec{Z}$, alors les deux conditions suivantes sont vraies :
  - a) $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}']\neg\varphi$. Dans le monde, le changement de $(\vec{X}, \vec{W})$ de $(\vec{x}, \vec{w})$ à $(\vec{x}', \vec{w}')$ change φ de vrai à faux.
  - b) $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w}', \vec{Z}' \leftarrow \vec{z}^*]\varphi$, pour tous les sous-ensembles $\vec{Z}'$ de $\vec{Z}$. Dans le monde, mettre $\vec{W}$ *à* $\vec{w}'$ ne doit pas affecter la valeur de φ aussi longtemps que $\vec{X}$ *est* à $\vec{x}$, même si toutes les variables dans les sous-ensembles arbitraires de $\vec{Z}$ ont leurs valeurs originales dans le contexte $\vec{u}$.

- AC3 : $\vec{X}$ est minimale ; aucun sous-ensemble de $\vec{X}$ satisfait les conditions AC1 et AC2. La condition de minimalité assure que seul les éléments de la conjonction $\vec{X} = \vec{x}$ qui sont essentiels pour le changement de φ dans AC2(a) sont considérés comme faisant parties de la cause ; les éléments non-essentiels sont éliminés.

L'essentiel de la définition est dans la condition AC2. Informellement les variables dans $\vec{Z}$ doivent décrire le «processus causal actif» de $\vec{X}$ jusqu'à φ (processus intrinsèque de Lewis [Lewis86]). Ce sont les variables qui se trouvent sur le chemin entre $\vec{X}$ et φ.

En fait on peut définir le processus causal actif de $\vec{X} = \vec{x}$ à φ comme étant l'ensemble minimal de $\vec{Z}$ qui satisfait AC2. La condition AC2(a) signifie qu'il existe une attribution $\vec{x}'$ de $\vec{X}$ qui change φ à ¬φ, aussi longtemps que les variables non-impliquées dans le processus causal, W ont la valeur w' (ceci constitue un évocateur du critère contre-factuel traditionnelle, φ doit être fausse si $\vec{X}$ n'est pas à $\vec{x}$)[Lewis86]. Cependant la condition AC2(a) est plus tolérante que le critère traditionnel, car elle permet de tester la dépendance de φ sur $\vec{X}$ sous des circonstances spéciales où les variables dans $\vec{W}$ ont les attributions $\vec{w}'$. Cette modification a été proposée par Pearl [Pearl98][Pearl00] et a été nommé «contingence structurelle » c'est une altération du modèle M qui implique l'arrêt dans quelques mécanismes (par des actions externes) mais qui ne change pas le contexte $\vec{u}$.

AC2(b) est une tentative de limiter la tolérance de AC2(a) selon les contingences structurelles. Elle assure essentiellement que $\vec{X}$ est suffisante pour apporter le changement de φ à ¬φ ; Alors que l'attribution de $\vec{W}$ *à* $\vec{w}'$ élime simplement les effets qui tentent de masquer l'action de $\vec{X}$. Le changement de $\vec{W}$ *de* $\vec{w}$ à $\vec{w}'$ n'a pas d'effet sur la valeur de φ. Encore, malgré que les variables de $\vec{Z}$ impliquées dans le processus causal peuvent être perturbées, cette perturbation n'a pas d'effet sur la valeur de φ. Le but essentiel de ce besoin est que nous n'avons pas la liberté de construire des tests contre-factuels de AC2(a) sous une altération arbitraire du modèle.

L'altération considérée ne doit pas affecter le processus causal. Clairement, si les contingences considérées sont limitées à remettre les valeurs à leurs valeurs réelles (restriction utilisée par Hitchcock [Hitchcock99]) alors, (M, $\vec{u}$) $\models \vec{W} = \vec{w}'$, et AC2(b) est satisfaite automatiquement.

La notion de cause contributive [Pearl00] est définie comme la cause réelle. Si AC2(a) est vraie seulement avec $\vec{W} = \vec{w}' \neq \vec{w}$, alors $\vec{X} = \vec{x}$ est une cause contributive de φ ; la cause réelle est vraie seulement pour $\vec{W} = \vec{w}$.

AC2(c) : (M, $\vec{u}$) $\models [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w}'']\varphi$ pour toutes les attributions de $\vec{W}$.

AC2(c) signifie que l'attribution $\vec{X} \leftarrow \vec{x}$ est suffisante pour forcer φ à être vraie, indépendamment des attributions de $\vec{W}$. On dit que $\vec{X} = \vec{x}$ est une *cause forte* de φ si AC2(c) est vraie en plus des autres conditions.

AC3 est une condition de minimalité, elle force la cause à être une conjonction singulière de la forme $\vec{X} = \vec{x}$. Ceci dépend de la supposition que l'ensemble $\vec{V}$ des variables endogènes est fini [Hopkins01] ; elle dépend aussi du fait que l'on utilise la causalité plutôt que la causalité forte.
On dit que $\vec{X} = \vec{x}$ est une *cause faible* de φ si AC1 et AC2 sont vérifiées sans la condition de minimalité AC3.

Exemple explicatif 2.3.2

Supposons que deux pyromanes jettent deux allumettes dans deux endroits différents d'une forêt sèche et les deux causent le fait que les arbres commencent à brûler. Considérons deux scénarios. Dans le premier cas disjonctif, chacune des allumettes suffit à elle seule à brûler l'ensemble de la forêt. Dans le second cas conjonctif, les deux allumettes sont nécessaires pour brûler la forêt. Si l'une d'elle seulement est allumée, l'incendie s'arrête avant que la forêt brûle.

On modélise le scénario par un modèle structurel contenant quatre variables :

- Une variable exogène U qui détermine, parmi d'autres choses, les motivations et l'état mentale des pyromanes, on suppose que $R(U)=\{U_{00},U_{10},U_{01},U_{11}\}$; Tel que si $U=U_{ij}$, alors le premier pyromane à l'intention de mettre du feu si i=1 et le deuxième à l'intention de mettre du feu si j=1, dans notre scénario $U=U_{11}$.
- Des variables endogènes ML_1 et ML_2, ou $ML_i=0$ si le pyromane i n'as pas allumé l'allumette et MLi=1 s'il l'a allumé, i :=1,2 ;
- Une variable endogène FB pour la forêt a brûlée, avec FB=0 si la forêt n'a pas brûlée et FB=1 sinon.

Les deux scénarios ont le même réseau causal, ils diffèrent uniquement dans l'équation structurelle de FB. Pour le scénario disjonctif on a $F_{FB}=(u,1,1) =F_{FB}(u,1,0)= F_{FB}(u,0,1)=1$ et $F_{FB}(u,0,0)=0$ (où $u \in R(U)$). Pour le cas conjonctif on a $F_{FB}(u,1,1)=1$ et $F_{FB}(u,1,0)= F_{FB}(u,0,1)= F_{FB}(u,0,0)=0$. On suppose que tous les effets pertinents sont donnés et qu'il n'y a pas d'autres causes pour le feu de la forêt. Soient M1 et M2 les (portions) modèles causaux associés aux scénarios disjonctifs et conjonctifs (respectivement).
On suppose que tous les faits pertinents sont donnés (il n y a pas de feu de camping et pas d'éclair).
Le réseau causal est le suivant :

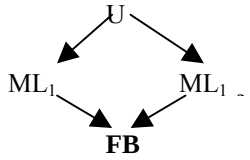


Figure 2 : Le réseau causal pour M1 et M2

Malgré les différences entre les deux modèles, chacune de $ML_1=1$ et $ML_2=1$ est une cause de FB=1 dans les deux scénarios. Nous allons l'illustrer pour $ML_1=1$. Soit $Z=\{ML_1,FB\}$; Alors $W=\{ML_2\}$. Il est facile de voir que la contingence $ML_2=0$ satisfait les deux conditions dans AC2. (Notez qu'on a besoin de mettre $ML_2=0$, contrairement au fait, pour montrer la dépendance cachée entre FB et ML_1 de telle attribution constitue un changement structurel dans le modèle original, puisque ceci implique d'enlever quelques équations structurelles).

Pour voir que ML_1 est aussi une cause dans M2, alors $Z=\{ML_1,FB\}$ et $W=\{ML_2\}$. $(M2,U_{11}) \models [ML_1 \leftarrow 0, ML_2 \leftarrow 1](FB=0)$ AC2(a) est satisfaite. Plus encore, comme la valeur de ML_2 nécessaire pour AC2(a) est la même que sa valeur originale i.e., (w'=w). AC2(b) est satisfaite trivialement.

Cet exemple illustre aussi le besoin de la condition de minimalité AC3 par exemple, si allumer l'allumette est qualifiée comme la cause de l'incendie alors allumer l'allumette et éternuer pourrait également vérifier AC1 et AC2 et maladroitement être qualifiée comme la cause de l'incendie. La minimalité élime l'éternuement. Cette condition sert à enlever les parties non-pertinentes de la cause.

Le fait de permettre les causes disjonctives est utile pour pouvoir faire la distinction entre M1 et M2. Une définition contre-factuelle proprement dite, doit faire en sorte que « $ML_1=1 \vee ML_2=1$ » est la cause de FB (si $ML_1=1 \vee ML_2=1$ n'est pas vraie alors FB ne doit pas l'être), mais ne doit pas permettre ni ML1=1 ni $ML_2=1$ seule comme cause de FB (puisque si $ML_1=1$ n'est pas vrai dans M_1, FB=1 pourrait encore être vraie). Cette définition ne permet pas de capturer cette intuition.

Cet exemple montre que la causalité et la causalité forte ne coïncident pas toujours, il n'est pas difficile de voir que ML_1 et ML_2 sont des causes de l'incendie de la forêt dans les deux scénarios. Cependant pour que ML_1 soit une *cause forte* de FB dans le scénario conjonctif, on doit inclure ML_2 dans Z ($\vec{W}$ est vide) ; si ML_2 est dans $\vec{W}$, alors AC2(c) (n'est pas vérifiée). Ainsi dans le cas de la causalité forte, ce n'est plus le cas de considérer que Z consiste uniquement en des variables sur le chemin entre la cause et l'effet.

Enfin, une remarque concernant une extension complémentaire de la définition de cause. Quand on cherche la cause de φ, on est souvent, non seulement, intéressé par l'occurrence ou la non occurrence de φ, mais aussi par la manière dans laquelle φ apparaît ; S'opposant à une certaine alternative par laquelle φ aurait pu apparaître [Hitchcock96]. On dit, par exemple « X=x a causé l'incendie en Juin par opposition à l'incendie en mai ». Si on suppose qu'il y a assez de bois dans la forêt pour uniquement un seul incendie de la forêt, les deux événements contradictoires « incendie en mai » et « incendie en Juin » ces deux événements ne sont pas complémentaires mais chacun exclut l'autre. La définition précédente peut être étendue pour s'accommoder de la causalité complémentaire. On définie « x cause φ qui s'oppose à φ' », où φ et φ' sont incompatibles mais non-exhaustives, en remplaçant ¬φ par φ' dans la condition AC2(a) de la définition.

Le complémentaire s'appliquer aussi à l'antécédent, comme par exemple « le fait que Susan a couru plutôt que marcher vers la salle de musique est la cause du fait qu'elle soit tombée ». On peut capturer les phrases de la forme « X=x, plutôt que X=x' pour x≠x', cause φ ». Ceci exprime que

1) X=x cause φ et
2) AC2(b) est vrai pour X=x' et φ. i.e., la seule raison pour laquelle X=x' n'est pas la cause de φ est que X=x' ne s'est pas produite dans le monde réel.

Complémenter les composants de l'antécédent et le conséquent nous permet d'interpréter les phrases de le forme «Susan a couru plutôt que marcher vers la salle de musique» a causé «le fait que «elle ait passé la nuit à l'hôpital», qui s'oppose au fait que «elle ait passé la nuit dans l'appartement de son petit ami».

III. Exemples de modélisations

Dans cette section nous allons présenter la modélisation de quelques exemples qui ont posé des problèmes aux autres approches.

Exemple 3.1

Le premier exemple est du à Bennett (et apparaît aussi dans [Sosa93]). Supposons qu'il y eu une forte pluie en Avril et des tempêtes avec éclairs dans les deux mois suivants ; qu'en Juin il y eu un éclair. S'il n'y avait pas eu une forte pluie en Avril, la forêt aurait pris feu en Mai. La question est : est-ce la pluie d'Avril qui a causé l'incendie de la forêt ? Selon une analyse contre-factuelle Naïve, c'est le cas, s'il n'avait pas plu, il n'y aurait pas eu d'incendie en Juin. Bennet dit : « C'est inacceptable, un assez bon historique des événements et de causation doit nous donner raison d'accepter des choses qui semble intuitivement fausses, mais aucune théorie ne peut nous persuader que le fait de retarder l'incendie de la forêt d'un mois (ou d'une minute) va causer l'incendie de la forêt »

Selon les auteurs [Halpern05], c'est en effet faux dire que les pluies d'Avril ont causé l'incendie de la forêt, mais elles étaient une cause du fait qu'il y eu un incendie en Juin et pas à Mai, par opposition en Mai. Ceci nous semble intuitivement correct. Pour mieux comprendre cette situation, il suffit d'utiliser un modèle avec trois variables aléatoires endogènes.

- AS : pour représenter les averses d'Avril, avec deux valeurs ; 0 pour représenter qu'il n'a pas plu fortement en Avril et 1 pour représenter qu'il a plu fortement en Avril ;
- ES : pour représenter les tempêtes avec éclairs, avec quatre valeurs possibles : (0,0,) (pas de tempêtes ni en Mai ni en Juin), (1,0) (une tempête en Mai et pas en Juin), (0,1) (pas de tempête en Mai mais en Juin), (1,1)(une tempête en Mai et en Juin) ;
- F : pour représenter l'incendie de la forêt, avec trois valeurs possibles : 0 pas d'incendie ; 1 en mai ; 2 en Juin.

On ne décrit pas le contexte explicitement, et on suppose que les valeurs de U sont de telle sorte qu'elles permettent qu'il y ait une averse en Avril, il y avait les tempêtes dans les deux mois suivants, qu'il y a suffisamment d'oxygène, et qu'il n'y a pas d'autres causes potentielles (allumettes..) et des causes inhibantes (alerte,..). Le réseau causal est simple ; il y a un arc de AS vers F et de ES vers F. il est facile de vérifie que ce qui suit est vrai :

- AS=1 est la cause de l'incendie en Juin (F=2) (en choisit W={ES}et Z={AS,F}) mais pas de l'incendie tout court (F=2$\vee$F=1).
- ES=(1,1) est la cause de F=2 et de (F=1$\vee$F=2). Ayant les tempêtes avec éclairs en mai et Juin causent le fait d'avoir l'incendie.
- AS=1$\wedge$ES=(1,1) n'est pas a cause de F=2, car violation de la condition de minimalité AC3 ; chaque élément de la conjonction est une cause de F=2. De façon similaire AS=1$\wedge$ES=(1,1) n'est pas la cause de (F=1$\vee$F=2).

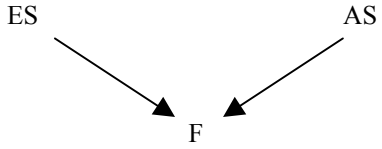


Figure 3 : L'incendie de la forêt

Il nous semble important de faire la distinction entre le fait que les pluies en Avril aient causé l'incendie (qui n'est pas le cas, selon cette analyse) et le fait que les pluies d'Avril aient causé l'incendie en Juin (qui le sont)[Halpern05]. Cette distinction n'est pas faite par Lewis[Lewis02].

Exemple 3.2

L'exemple suivant montre l'importance du choix des variables modélisant le processus causal et sa structure. L'histoire suivante prise de [Hall03] est un exemple de préemption, ou il y a deux causes potentielles d'un événement, un qui interrompe l'autre. Une définition adéquate de la causalité doit traiter la préemption dans toutes ses formes.
Suzy et Billy prennent chacun d'eux une pierre et tirent sur une même bouteille. La pierre de Suzy arrive en premier, et elle casse la bouteille. Alors que les deux tirs ont bien eu lieu, le tir de Billy aurait pu casser la bouteille s'il n a pas été interrompu par le tir de Suzy.
Le sens commun suggère que le tir de Suzy sont la cause du fait que la bouteille soit cassée, mais que celui de Billy ne l'est pas. Ceci est vrai dans cette approche, mais uniquement si on fait une modélisation appropriée de l'histoire.
Considérons en premier, un modèle causal, avec trois variables endogènes :

- ST pour le tir de Suzy, avec la valeur 0 (si elle ne tire pas), 1 si elle le fait ;
- BT pour le tir de Billy, avec la valeur 0 (s'il ne tire pas), 1 s'il le fait ;
- BS pour la bouteille cassée, avec la valeur 0 (si elle n'est pas cassée) et 1 si elle l'est.

Le réseau causal est simple, un arc de ST vers BS et un autre arc de BT vers BS. BT et ST jouent un rôle symétrique, avec BS=ST $\vee$BT ; il n'y a rien pour distinguer BT de ST. Sans aucune surprise, c'est le tir de Billy et celui de Suzy qui sont classés, dans ce modèle, comme cause du fait que la bouteille soit cassée. Le problème avec ce modèle est qu'on ne peut pas distinguer entre le cas où les deux pierres atteignent la bouteille en même temps de celui où la pierre de Suzy arrive en premier lieu. Pour cela le modèle doit être raffiné afin de modéliser cette distinction.

Une solution est de faire appel au modèle dynamique[Pearl00]. Une manière peut être plus simple est d'attribuer à BS trois valeurs, 0 pour bouteille non cassée, 1 pour bouteille cassée par le tir de Suzy, 2 pour bouteille cassée par le tir de Billy. Donc ST=1 est la cause de BS=1, mais BT=1 ne l'est pas (si Suzy n'a pas tiré mais Billy si, alors BS=2). Ceci résout le problème, mais la solution est presque programmée dans le modèle en invoquant la relation «comme résultat de », qui nécessite l'identification de la cause réelle.
Un choix plus judicieux est d'ajouter deux nouvelles variables dans le modèle :

- BH pour représenter le fait que la pierre de Billy atteint la bouteille, avec 0 (elle ne l'a pas atteint), 1 (elle l'atteint);
- SH pour représenter le fait que la pierre de Suzy atteint la bouteille, avec 0 (elle ne l'a pas atteint), 1 (elle l' atteint) ;

Dans ce modèle, on aura le réseau causal de la figure ci-dessous, avec l'arc de SH$\rightarrow$BH représentant le fait que SH inhibe l'effet de BT; BH=BT$\wedge\neg$SH (donc, BH=1 SSi BT=1 et SH=0). Dans le contexte où la pierre de Billy atteint en premier la bouteille on doit avoir un arc de BH vers SH.

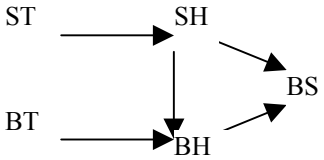


Figure4 : L'exemple de tir de la

Dans ce cas ST=1 est la cause de BS=1. Pour satisfaire AC2, on choisit $\vec{W}$ = {BT} et w'=0 et notons que, puisque BT=0, BS aura l'attribution de ST. Notons aussi que BT=1 n'est pas la cause de BS=1 ; il n'y a pas de partition $\vec{Z} \cup \vec{W}$ qui satisfait AC2. Un choix symétrique $\vec{W}$ ={BT} et w'=0 violerait la contrainte AC2(b) (avec $\vec{Z}$ '={BT}, puisque φ devient fausse lorsque ST=0 et BH restaure sa valeur originale 0.

Cet exemple montre le besoin d'invoquer les sous-ensembles de $\vec{Z}$ dans la condition AC2(b). (M, $\vec{u}$) $\models [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w'}]\varphi$ est vraie si on prend $\vec{Z}$ ={BT,BH}et $\vec{W}$ ={ST,SH,}, et sans avoir le besoin que AC2(b) soit vraie pour tous les sous-ensembles de $\vec{Z}$, BT=1 est qualifiée comme cause de BS=1. Le fait d'insister sur le fait que φ reste inchangé quand on met $\vec{W}$ à w' et $\vec{Z}$ ' à z* (pour un sous-ensemble arbitraire de $\vec{Z}$) nous empêche de choisir les contingences qui interfèrent avec le chemin causal actif de $\vec{X}$ vers φ.

De plus si on veut expliquer que la préemption $\vec{X} = \vec{x}$ est la cause de φ plutôt que $\vec{Y} = \vec{y}$, alors on doit disposer d'une variable aléatoire (BH dans ce cas) tel que sa valeur dépend de la cause réelle si c'est $\vec{X} = \vec{x}$ ou $\vec{Y} = \vec{y}$. Si le modèle ne contient pas une telle variable, alors il ne serait pas possible de déterminer laquelle est en fait la cause réelle. Il semble que par l'introduction des variables intermédiaires SH et BH dans l'histoire nous avons programmé la réponse désirée de ce problème ; Après tout, c'est le fait de tirer sur la bouteille, et non SH, qui empêche BH. Pearl analyse de façon similaire le problème de préemption retardée par des modèles d'équations structurelles dynamiques [Pearl00], où les variables sont indexées par le temps, et montre que la sélection de la première action comme une cause réelle de l'effet vient des conditions similaires à AC1-AC3 sans même spécifier celui qui a tiré la pierre. Nous présentons une adaptation simplifiée de cette analyse.

Soient t_1, t_2 et t_3, respectivement, pour le temps du tir de Suzy, de Billy et le temps où la bouteille est cassée. Soient H_i et BS_i les variables indiquant si la bouteille est atteinte (H_i), et cassée (BS_i) en temps t_i (avec i=1,2,3 et $t_1<t_2<t_3$), avec la valeur 1 si la bouteille est atteinte (respectivement cassée), 0 sinon. Si T_i est la variable représentant «quelqu'un tir une pierre en temps t_i et mettre BS_0 à vrai (toujours 1), on suppose que les équations suivantes sont vraies pour tous les temps t_i(pas seulement t_1, t_2 et t_3) :

$H_i = T_i \wedge \neg BS_{i-1}$.

$BS_i = BS_{i-1} \vee Hi$.

i. e., la bouteille est frappée au temps t_i si quelqu'un tir une pierre au temps t_i et qu'elle n'était pas déjà cassée au temps t_i, de même la bouteille est cassée en temps t_i si elle était déjà cassée au temps t_{i-1} ou bien elle a été atteinte au temps t_i.

Pour ce cas nous considérons uniquement les temps t_1, t_2 et t_3, nous aurons les équations structurelles suivantes,

$H_1 = ST$,

$BS_1 = H_1$,

$H_2 = BT \wedge \neg BS_1$,

$BS_2 = BS_1 \vee H2$,

$H_3 = T_3 \wedge \neg BS_2$,

$BS_3 = BS_2 \vee H_3$.

Le contexte est tel que ST=1, BT=1, T_3=0. Le réseau causal est le suivant :

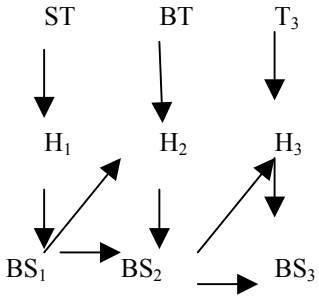


Figure 5 : le tir de la pierre

Il n'est pas difficile de voir que ST=1 est la cause de BS_3=1 (avec W={BT} dans AC2 et w'=0). BT=1 n'est pas la cause de BS_3 ; aucune partition de Z∪W ne satisfait la condition AC2(b). Pour établir la dépendance contre-factuelle entre BS_3 et BT, on doit mettre H_2 dans Z, et BS_1 dans W, et imposer la contingence BS_1=0. Mais cette contingence viole la condition AC2(b), alors qu'elle résulte de BS_3=0 quand on restore H_2 à 0 (sa valeur courante).

Deux points sont mis en relief. Premièrement, le tir de Suzy est déclaré comme la cause de la réalisation de l'événement BS_3=1 même si le tir n'est pas affirmé, retarde ou change n'importe quelle propriété de cette réalisation. Ceci pourrait être plus clair si on considère une autre réalisation, J_4 : «Joe était incapable de boire son chocolat favori de cette bouteille la nuit de jeudi». Comme une conséquence de BS_3, J_4 sera aussi classée comme étant causée par le tir de Suzy, non de celui de Billy, bien que J_4 aurait pu se réaliser de la même manière et en même temps si Suzy n'avait pas atteint la bouteille. Ceci implique que l'avancement ou le retardement du résultat ne peut pas être pris comme base pour décider de la cause réelle, un principe recommandé par [Pearl98].

Deuxièmement, le tir de Suzy est déclaré comme cause de BS_3 bien qu'il n'y ait pas de chaîne de dépendance contre-factuelle entre les deux. L'existence d'une telle chaîne a été proposée par Lewis comme critère nécessaire de la causation dans le cas de préemption[Lewis86]. Dans le contexte actuel, BS_2 ne dépend (contre-factuellement) ni de BS_1 ni de H_2 ; la bouteille pourrait être cassée au temps t_2 même si elle n'a pas été cassée au temps t_1 (parce que la pierre de Billy l'aurait atteinte).

Exemple 3.3

Ne pas avoir exécuté une action est (partie) d'une cause ? Considérons l'exemple suivant, pris de [Hall0 3].

> Billy étant resté dehors dans le froid assez longtemps à jeter des pierres, il a attrapé une maladie sérieuse, mais pas fatale. Il a été hospitalisé et traité le lundi, alors il s'est rétabli mardi matin.

Est-ce que l'omission de traiter Billy par le docteur le lundi qui est la cause de la maladie de Billy le mardi? Ceci semble vrai. Il semble raisonnable de considérer un modèle avec deux variables :

- MT pour «traitement le lundi », avec la valeur 0 (le médecin ne traite pas Billy le lundi) et 1 (il l'a fait) ;
- BMC pour «les conditions médicales de Billy», avec la valeur 0 (rétablit), 1 (encore malade).

Il est clair que, dans ce modèle, MT=0 est la cause de BMC=1. Supposons qu'il y ait 100 médecins dans l'hôpital. Mais qu'un seul est affecté à Billy (et qu'il oubli de lui donner son traitement), en principe, n'importe quel autre médecin des 99, aurait pu donner le traitement à Billy. Alors, est-ce le fait qu'ils n'aient pas donné le traitement à Billy est aussi une partie de la cause qu'il soit encore malade le mardi ?
Dans ce modèle la faille des autres médecins n'est pas une cause, puisque nous n'avons pas de variables aléatoires pour modéliser les actions des autres médecins, leur omission d'action est une partie du contexte. Si nous avions introduit des variables aléatoires correspondant aux autres médecins, alors leur faille aurait fait partie de la cause que Billy soit malade le mardi.

Supposons que le médecin du lundi fait son travail, et que la première chose qu'il fait est de donner le traitement aux patients la matinée, donc Billy est rétablit l'après midi du mardi. Le médecin de mardi fait aussi son travail, et donne le traitement à Billy si le médecin du lundi ne la fait pas. Ajoutons une autre contrainte : une dose de médicament est suffisante, mais deux sont fatales.

Est-ce le fait que le médecin du mardi n'ait pas traité Billy qui est la cause qu'il soit vivant (et guéri) le mercredi ? Cette histoire peut être modélisée par trois variables aléatoires : MT pour le traitement de lundi (1 Billy est traité le lundi, 0 sinon), TT pour le traitement de mardi (1 Billy est traité le mardi, 0 sinon) et BMC pour les conditions médicales de Billy (0 il est bien le mardi matin et mercredi matin,1 s'il est malade le mardi matin et en bon état le mercredi matin, 2 il est malade le mardi matin et mercredi matin,3 il est bien le mardi matin et mort le mercredi matin). On peut décrire les conditions médicales de Billy par une fonction à quatre combinaisons possibles traitement/non-traitement le lundi et mardi. Le réseau causal est simple il y a un arc de MT vers TT, le traitement de mardi dépend de celui de lundi, et des arcs de MT et de TT vers BMC, les conditions médicales dépendent des deux traitements.

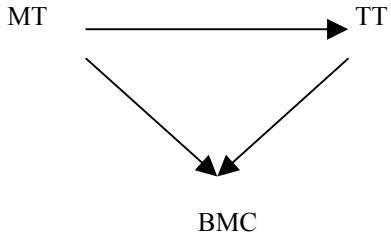


Figure6 : Le traitement de Billy

Dans ce modèle causal, il est vrai que MT=1 est la cause de BMC=0, Billy est traité le lundi, et il n'est pas traité le mardi matin, donc il se trouve en bon état le mercredi matin. MT=1 est aussi la cause de TT=0, et TT=0 est la cause que Billy soit en vie (BMC=0∨*BMC*=1∨ BMC=2). Cependant, MT=1 n'est pas la cause que Billy est en vie. La condition AC2(a) n'est pas vérifiée : mettre MT=0 mène toujours au résultat que Billy est en vie (avec W=∅). Notez que le fait de prendre W={TT} ne nous aide pas. Si TT=0, alors Billy est en vie quel que soit MT, si TT=1, alors Billy meurt si MT a sa valeur originale, donc AC2(b) est violée (avec Z'=∅).

Ceci montre que la causalité n'est pas transitive, selon cette définition. Bien que MT=1 est la cause de TT=0 et TT=0 est la cause de BMC=0∨BMC=1∨ BMC=2, MT=1 n'est pas la cause de BMC=0∨BMC=1∨ BMC=2. C'est la causalité fermée sous certaines faiblesse : MT=1 est la cause de BMC=0, et BMC=0 implique logiquement BMC=0∨BMC=1∨ BMC=2, ce qui n'est pas causé par MT=1.

Hall discute de l'issu de la transitivité de la causalité, et suggère qu'il y ait un lien entre le fait que la causalité soit transitive et la causalité due à la non-réalisation de quelques événements. Il suggère qu'il existe deux concepts de la causation : un correspondant aux dépendances contre-factuelles et l'autre correspondant à «la production », où A cause B si A aide à produire B. La causalité par production est transitive ; alors que la causalité par dépendance ne l'est pas[Hall03].

L'approche contre-factuelle arrive à capturer les deux concepts : AC2(a) capture quelques intuitions des dépendances contre-factuelles (Si A n'as pas lieu alors B ne doit pas avoir lieu si W=w') et AC(b) capture quelques traits de la production (A force B d'être vraie, même si W=w'). Néanmoins, on n'a pas besoin de deux notions séparées pour traiter ces rapports.

D'après hall, l'échec de la transitivité est lié à la distinction entre la présence et l'absence d'événements, mais selon Halpern, la notion de transitivité cause des problèmes indépendamment du fait de permettre ou pas la causalité due à la faille de quelques événements.

Lewis insiste sur le fait que la causalité est transitive [Lewis86][Lewis02], en partie pour pouvoir traiter la préemption. La définition proposée par Halpern permet de traiter les exemples standard de la préemption sans avoir besoin d'invoquer la transitivité, qui, comme les exemples de Lewis le montrent, mènent vers des conclusions contre-intuitives.

IV. Définition plus raffinée

Cette définition est dite 'préliminaire', car il y 'a quelques situations qu'elle ne peut pas traiter. Les exemples suivants illustrent le problème [Halpern05].

Exemple 4.1

Considérant l'exemple 3.2, où Suzy et Billy jettent chacun d'eux une pierre sur une même bouteille. La pierre de Suzy atteint la bouteille en premier. Supposons qu'il y a bruit qui cause un retard au tir de Suzy, mais qui reste néanmoins le premier à atteindre la bouteille. Nous allons ajouter deux variables au modèle précédant, N (où N=0 signifie qu'il n'y a pas de bruit et N=1 s'il y en a) et $BS_{1.5}$ (où $BS_{1.5}$=1 si la bouteille est cassée au temps $t_{1.5}$, où $t_1<t_{1.5}<t_2$, et $BS_{1.5}$=0 sinon). Dans la situation réelle, il y a un bruit, la bouteille est cassée en $t_{1.5}$, donc N=1 et $BS_{1.5}$=1. Comme dans l'exemple 4.2 on peut montrer que le tir de Suzy est la cause du fait que la bouteille soit cassée et le tir de Billy non. Pas de surprise, N=1 est la cause $BS_{1.5}$=1 (sans bruit la bouteille aurait été cassée au temps 1). N=1 est aussi la cause que la bouteille soit cassée. Donc, N=1 est la cause de BS_3=1.

Ceci semble irraisonnable, car intuitivement, la bouteille aurait pu être cassée qu'il ait eu du bruit ou pas. Cependant, cette intuition n'est pas prise en compte dans ce modèle. Considérons la contingence BS_1=0. Sous cette contingence, la bouteille n'est pas cassée au temps 1, même si le tir de Suzy l'a touché. Cependant, si N=1 et BS_1=0, alors la bouteille est cassée au temps 1.5. Ayant ceci, il s'en suit facilement que, selon la définition, N=1 est une cause de BS_3=1.

Le problème ici est causé par ce qui peut être considéré comme un scénario extrêmement irraisonnable : la bouteille ne s'est pas cassée malgré qu'elle soit atteinte par le jet de pierre de Suzy. Désirons-nous considérer ce scénario ? Le choix revient au modélisateur. Intuitivement, si on considère ces scénarios, alors le bruit devrait être une cause ; sinon, il ne devait pas l'être.

J.Pearl et JY. Halpern proposent une modification de la définition préliminaire pour qu'elle capture cette intuition. On prend un *modèle causal étendu* qui sera le tuple (S,F,ε), où (S,F) est le modèle causal précédent, et ε est l'ensemble des *attributions possibles* pour les variables endogènes. Donc, si les variables $X_1,\ldots,X_n$ sont endogènes, alors $(x_1,..,x_n)\in\varepsilon$ si $X_1=x_1,\ldots,X_n=x_n$ est une attribution possible. On dit que cette attribution, d'un sous-ensemble de variables endogènes, est possible si elle peut être étendue à une attribution de ε. Alors on modifie légèrement la condition AC2(a) et AC2(b) dans la définition de la causalité pour restreindre les attributions possibles. Dans le cas spécial où ε consiste en toutes les attributions, la définition se réduit à la définition déjà donnée.

L'exemple précédent peut se représenter par un modèle causal étendu, en enlevant les attributions où $BS_1=0\wedge H_1=1$. Ceci nous retourne aux attributions initiales. L'exemple suivant illustre aussi le besoin d'être capable de considérer les attributions irraisonnables.

Exemple 4.2

Fred s'est gravement blessé du doigt lors de son travail dans l'usine (FS=1). Fred est transporté en urgence à l'hôpital, où son doigt est soigné. Un mois après, le doigt est complètement rétabli (FF=1). Dans cette histoire, nous ne voulons pas dire que FS=1 est causé par FF=1. En effet, selon cette nouvelle définition, ce n'est pas le cas, puisqu'on a FF=1 indépendamment du fait que FS=1 ou pas(dans toutes les contingences satisfaisant AC2(b)).

Cependant, supposons qu'on introduit deux nouveaux éléments à l'histoire, représentant les contingences structurelles non réelles : Larry attend Fred a la sortie de l'usine avec l'intention de couper lui le doigt, comme un avertissement pour qu'il lui rembourse ce qu'il lui doit rapidement. Soit LL la variable qui représente si Larry attend Fred ou pas, et LC représente le fait que Larry coupe le doigt de Fred ou pas. Si Larry coupe le doigt de Fred, il va le jeter, donc Fred ne peut pas pouvoir le récupérer pour le remettre en place. Dans cette situation, LL=LC=0 ;

Supposons que, Fred s'est blessé le doigt à l'usine, dans ce cas Larry ne pourra pas le lui couper le doigt (puisque Fred sera hospitalisé). Maintenant FS=1 devient la cause de FF=1. Pour la contingence structurelle où LL=1, si FS=0 alors FF=0 (Larry coupa le doigt de Fred et le jeta pour qu'il ne soit plus fonctionnel). De plus, si FS=1, alors LC=0 et FF=1, juste comme dans la situation réelle.

Si on veut réellement considérer que «Larry coupe le doigt de Fred» est un fait totalement imaginaire ('irraisonnable'), alors il suffit d'enlever les attributions où LL=1. Par ailleurs, si le fait d'avoir un doigt coupé est un cas souvent réalisé, alors il semble raisonnable de dire que : « l'accident est la cause du fait que le doigt soit fonctionnel, quelques temps après l'accident».

Exemple 4.3

Hall et Paul [Hall03], donnent un exemple du à Sarah McGrath suggérant qu'il y ait une différence entre la causalité par omission et la causalité par commission :

> Supposant que Suzy part en vacance et laisse ses plantes entre les mains de Billy, qui lui avait promis de les arroser. Cependant Billy ne les arrose pas, et les plantes meurent, ce qui n'aurait pas été le cas, si Billy les avait arrosé. Le fait que Billy n'ait pas arrosé les plantes est la cause de leur mort. Mais Vladimir Putin non plus n'a pas arrosé les plantes de Suzy, et s'il les avait arrosé, elles ne seraient pas mortes. Pourquoi dans ce cas ne pas considéré son omission aussi comme la cause de la mort des plantes?

Il est clair que Billy est la cause. Vladimir l'est aussi, si on garde toutes les attributions. Cependant, si on enlève l'attribution où Vladimir arrose les plantes, alors la faille de Billy est la cause mais pas celle de Vladimir.

Est-ce que les auteurs se donnent trop de flexibilité ? C'est à celui qui modélise de défendre son choix de modèle. Un modèle qui ne nous permet pas de considérer 'Vladimir arrosant les plantes' peut être défendu de manière évidente : c'est un scénario ridicule à considérer. Considérons d'autre part que Maggie est la sœur de Suzy (qui a les clefs de la maison) vient pour contrôler les choses, il semble raisonnable, que Suzy soit fâchée contre de sa sœur de ne pas avoir arrosé les plantes même si elle ne l'avait pas responsabilisé. Intuitivement, il semble raisonnable de ne pas enlever l'attribution où Maggie arrose les plantes.

Supposons que l'on permet uniquement les attributions jouant des rôles significatifs et non les attributions imaginaires. Dans l'exemple suivant, ceci aide à clarifier la relation entre les différents modèles de l'histoire.

Exemple 4.4

Cet exemple concerne ce que Hall appelle la distinction entre la causation et la détermination [Hall03].

> Vous êtes assis devant un switcher de pistes de chemin de fer. Un train qui arrive, si vous changez le switcher vous envoyez ce train vers la piste à droite, si vous le laissez tel qu'il est, le train va prendre la piste à gauche. Dans tous les cas, le train converge vers la même station. Votre action n'est pas la cause de son arrivée, simplement elle détermine la piste d'où il est passé (via la piste à droite ou via la piste à gauche).

Une fois encore, le modèle prend en compte ce problème. Supposons qu'on ait trois variables aléatoires :

- F : pour «changer le switcher », avec 0 si vous ne changez pas le switcher et 1 si non.
- T : pour la piste, avec 0 le train prend la piste de gauche et 1 si le train prend la piste de droite.
- A : pour l'arrivée du train, avec la valeur 0 (si le train n'arrive pas à la station de convergence et 1 si non.

Il est facile de voir que F=1 cause le fait que le train prend la piste de droite (T=0), mais ne cause pas l'arrivée du train «le changement du switcher ne cause pas l'arrivée du train ».

Supposons qu'on modélise les pistes en utilisant deux variables :

- LT pour la piste à droite, avec la valeur 1 si le train prend cette piste et 0 sinon.
- RT pour la piste à droite, avec la valeur 1 si le train prend cette piste et 0 sinon.

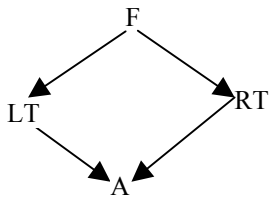


Figure 7 : Le changement du switcher

Le diagramme causal est isomorphe à la classe des problèmes que Pearl appelle «Switching causation»[Pearl00]. Il semble raisonnable d'enlever l'attribution LT=RT=1, un train ne peut prendre plus d'une piste. Si on n'enlève pas les autres attributions, cette représentation considère F=1 comme cause de A. A premièrement vue, elle semble contre-intuitive : un changement dans la représentation peut-il changer une «non cause» en une cause ?

C'est possible et elle doit l'être! Un changement vers un modèle avec deux variables n'est pas simplement syntactique, mais représente un changement profond dans l'histoire. Le modèle avec deux variables décrit les deux pistes comme deux mécanismes indépendants, donc permettant à une piste d'avoir la valeur vraie (ou faux) sans affecter l'autre. Spécifiquement, ceci conduit à un accident désastreux, de changer le switcher alors que la piste gauche ne fonctionne pas. Formellement, ceci permet une attribution où F=1 et LT=0. De telles attributions sont possibles qu'on puisse exprimer dans un modèle à deux variables, mais pas dans un modèle à une variable. Bien sure, si nous enlevons les attributions où F=1 et LT=0,ou bien F=0 et RT=0, alors nous reviendrons essentiellement à notre premier modèle. Le potentiel de telles attributions est précisément ce qui rend F=1 la cause de A dans le modèle de la figure précédente.

Est-ce le changement du switcher qui est la cause de l'arrivée du train ? Non ; dans des situations idéales, où tous les mécanismes fonctionnent. Mais ceci n'est pas de la causalité (modélisation causale). Les modèles causaux gagnent de leurs intérêts dans des circonstances anormales où une piste ne fonctionne pas, des travaux récents existent, qui portent sur la modélisation de la causalité par la violation des normes[Kayser04][nouioua05]. C'est cette possibilité qu'il faut avoir en tête une fois qu'on décide de désigner chaque piste comme un mécanisme (équation) séparé dans le modèle et, gardant en tête cette contingence, il ne doit pas être étrange de dire que la position du switcher est la cause de l'arrivée du train (ou pas).

L'exemple 4.4 nous donne quelques indications sur le processus de construction du modèle. Alors qu'il n'y a pas une manière de prouver qu'un modèle donné est le bon modèle, il est clairement important pour un modèle d'avoir suffisamment de variables pour pouvoir exprimer ce que le modélisateur considère vraie dans des situations raisonnables.

Par ailleurs, le fait de permettre la restriction de l'ensemble des attributions possibles dans la définition de la causalité, n'est pas pénalisant pour le modélisateur.

Exemple 4.5

Une balle est attrapée par un gardien, un peu plus loin sur son chemin il y a un mur, derrière celui ci il y a une fenêtre[McDermott95] [Lewis02]. Est-ce le fait que le gardien ait attrapé la balle qui est la cause du fait que la fenêtre ne se soit pas cassée ? Comme dit Lewis [Lewis02] :

On peut penser : que oui, le gardien et le mur ont sauvé la fenêtre, mais le mur n'as rien fait, puisque la balle n'a jamais atteint le mur. Donc ça doit être le gardien. Ou bien on peut penser que : non le mur protège la fenêtre, quel que soit ce que le gardien a fait ou n'a pas fait.

Lewis argumente que notre ambivalence dans ce cas peut être respectée, et les deux solutions devraient être tolérées. Nous pouvons donner à cette ambivalence une expression formelle dans cette approche. Si on prend le mur et le gardien comme des variables endogènes alors, le mur est la cause que la fenêtre soit en bon état, sous la supposition que le gardien n'arrive pas à attraper la balle et que le mur n'est pas présent, est un scénario qu'on ne peut considérer comme raisonnable. Dans ces circonstances, le gardien est aussi la cause du fait que la fenêtre soit en bon état [Halpern05].

Par ailleurs, si on considère la présence du mur comme un fait garanti : soit en considérant le mur comme une variable exogène, i. e., ne pas l'inclure dans le modèle. Ou bien ne pas permettre des situations où le mur n'arrête pas la balle si le gardien ne l'a pas attrapé.
Dans ce cas, le fait que le gardien attrape la balle n'est pas une cause du fait que la fenêtre ne soit pas cassée. Elle serait restée en bon état quel que soit le résultat de l'action du gardien.
Cet exemple met aussi l'accent sur l'importance du choix du modèle, et de réfléchir sur ce que nous voulons faire varier ou bien garder fixe.

V. Comparaison entre l'Approche contre-factuelle et l'approche Normative[Boutouhami 05]

Ces deux approches proposent une modélisation de la notion de causation et fournissent des solutions assez satisfaisantes à certains problèmes bien connus dans la littérature, mais ces deux approches se basent sur des raisonnements différents, si on applique donc les deux méthodes sur un même exemple, on n'aura pas le même résultat. Nous proposons dans la suite une comparaison qui permet de trouver quelques similitudes des cas traités par ces deux approches ainsi que les points de divergences.

V.1. Points communs

- "Une explication est relative à l'état épistémologique d'un agent, ce qui est une explication pour un agent peut ne pas l'être pour un autre agent", ce besoin est exprimé dans l'approche contre-factuelle par l'ensemble des contextes considérés possibles par l'agent, il est traité dans l'approche normative grâce à la notion de norme, la norme change selon celui qui explique et selon le contexte. Exemple [Mokhtari97a], Serge, grand fumeur devant l'éternel, est à peine rentré dans son appartement qu'il s'empresse d'allumer une cigarette. A cet instant une explosion détruit l'appartement. I y avait une fuite de gaz.." qu'est-ce qui sera interprété, par les assurances notamment, comme ayant causé l'explosion destructrice de l'appartement ? Ce n'est pas l'événement "cigarette allumée", bien qu'il soit une condition nécessaire pour provoquer l'explosion. En effet, il est normal que serge fume chez lui. En revanche si serge se promène dans une usine pétrochimique où il est naturel de trouver du gaz en suspension dans l'atmosphère, et qu'il allume une cigarette, la cause de l'explosion qui s'ensuivra sera attribuée à l'événement "allumer une cigarette". Cet événement n'est pas ici une description légitime du cours naturel des choses.
- "La relation de causalité n'est pas Transitive". Les deux définitions s'accordent sur cette propriété : la non-transitivité de la causation. Halpern et Pearl l'ont illustré à travers l'exemple 'Maladie de Billy". Dans l'approche normative, ceci est modéliser par le fait que seules les actions peuvent causer un fait et un fait ne peut pas causer un autre fait, et aussi par la notion de norme. Dans l'exemple de la maladie de Billy, si le traitement du Lundi a eu lieu, il aura comme effet le traitement du mardi n'aura pas lieu. La méthode normative se base sur des liens statiques, ceci ne pose pas de problème parce qu'elle impose la réalisation de toutes les normes et la non existence de faits inhibants.
- La notion d'action est bien formulée dans l'approche normative par la notion de points de choix, dans l'approche structurelle dans le déroulement de l'histoire. Les deux approches agréent avec l'idée que seules les actions peuvent être des causes dans le modèle structurel, seules les variables endogènes peuvent faire partie de la cause.

- "La Préemption": les deux méthodes permettent de traiter la préemption. Cette propriété est illustrée dans l'exemple de Billy et Suzy dans l'approche contre-factuelle. Ceci s'exprime facilement dans l'approche normative à l'aide de la notion de faits inhibants, le tire de Suzy inhibe l'effet souhaité du tire de Billy, la norme de la réalisation de l'action de Billy est incomplète car "la bouteille est déjà cassée". Donc l'approche normative permet de faire différence entre le cas où les deux tires cassent la bouteille en même temps et le cas où le tire de Suzy qui casse la bouteille.

- Difficulté de choix du type des variables": Dans les deux approches il n y a pas de formalisme pour le choix du type des variables, c'est un travail qui implique des experts dans le domaine à modéliser et du choix du modélisateur. Pour la méthode normative, il n'est pas facile de faire la différence entre un événement, un fait et une action. De même pour l'approche contre-factuelle in n'est pas facile de décider si une variable donnée est endogène ou bien exogène or les exemples donnés dans ce chapitre ont bien montré l'importance du choix du type des variables (exemple N° 4.4 du switcher).

- Les deux méthodes ne permettent pas de traiter les causes disjonctives. Dans l'approche structurelle, la seule explication de "A$\vee$B cause C" est soit "A cause C" ou bien "B cause C". puisque si je sais que A$\vee$B cause C, mais je ne sais pas lequel. Dans l'approche normative, le problème est dans la définition de la disjonction des normes.

- Les deux approches permettent la modélisation de faits inexplicables. Dans le cas de l'explication, selon l'approche structurelle il y a des faits inexplicables (exemple de l'image de la télévision), de même pour l'approche normative où l'algorithme de l'explication peut générer un ensemble vide (c'est le cas, par exemple, des causes naturelles).

- "La causalité par production est transitive", les deux approches partagent cette idée. Dans l'approche contre-factuelle, la condition AC2(b) capture certains traits de la production (A force B à être vraie, même si W=w'). Pour l'approche normative, ceci est fait en appliquant la fermeture transitive par rapport aux règles statiques et causales (et le fait que les faits ne sont pas des causes et seules les actions modifient l'état du monde) [Khelfallah01].

- Les deux approches font du raisonnement non-monotone. Dans l'approche normative ce point est modélisé par la notion de faits qui inhibent la réalisation d'une action ou bien des effets souhaités d'une action e par la notion de persistance et la notion de norme. Pour ce qui est de l'approche contre-factuelle une cause est évaluée relativement à un contexte particulier, par rapport à modèle causal donné, la cause change donc d'un contexte à un autre.

V.2. Divergences

Les deux approches ont néanmoins quelques divergences. Reprenons l'exemple du jet de pierre sur la bouteille. Avec la première modélisation où il y a un arc de TS vers BC et un autre arc de TB vers BC, la méthode contre-factuelle ne permet pas de faire distinction entre le cas où Suzy frappe la bouteille et la cas où Billy et Suzy frappe en même temps la bouteille, avec cette modélisation les deux tire jouent un rôle symétrique, et pour cela qu'il ont re-modéliser cette histoire en ajoutant deux autres variables afin de pouvoir faire la distinction. Or la méthode de normative permet de traiter facilement ce cas et de faire la distinction, simplement car la norme associée à la deuxième action n'est pas vérifiée ce qui enlève le tire de Billy d'être la cause du fait que la bouteille soit cassée.

Le temps joue un rôle crucial dans la perception de la causalité, mais dans l'approche contre-factuelle cette notion n'as pas été prise en compte, les équations structurelles ne font pas de référence à ce facteur, alors que dans l'approche normative de la causalité le temps est représenté explicitement comme un paramètre des prédicats utilisés et tout le raisonnement se base sur un enchaînement de réalisation de fait et d'action (pour rechercher la cause ou l'explication une recherche est effectuée sachant que les faits et les actions sont ordonnées selon leur date de réalisation).

L'approche normative est donc plus riche sur cet aspect. Le temps est représenté explicitement en prenant en compte,

- La durée, sachant qu'une action prend du temps,
- Le délai, sachant que les effets ne sont pas toujours immédiats, et
- Les effets d'actions concurrentes.

L'approche normative de causalité permet de traiter les actions concurrentes (la seule condition c'est qu'elles ne génèrent pas d'effets contradictoires).

L'approche normative nécessite un bon fond des connaissances historiques (détaillé sur le passé). Elle nécessite aussi la prise en compte de délai d'exécution des effets qui ne sont pas nécessairement instantanés.

Le problème avec l'utilisation des modèles à équations structurelles est que le langage des modèles structurels n'est pas suffisamment expressif pour capturer certaine relation internes qui sont nécessaire pour la relation causale, avec le modèle structurel nous manipulons des attributions de variables aléatoires ce qui ne permet pas d'exprimer les points suivants [Hopkins02]:

- La distinction entre la condition et la transition
- La distinction entre la présence et l'absence d'un événement
- Le temps et les relations causales.

L'approche contre-factuelle nécessite l'aide d'un expert du domaine pour la modélisation de l'histoire, comme on a déjà vu, le choix des variables est très important et un changement des variables peut tout changer. Cette approche propose des solutions à plusieurs problèmes, mais elle reste néanmoins intuitivement peu commode.

VI. Conclusion

Nous avons présenté une représentation formelle de la connaissance causale et une manière de déterminer les causes réelles à partir de ces connaissances. Cette approche montre qu l'approche contre-factuelle pour la causalité, dans la tradition de Hume et Lewis, ne doit pas être abandonnée; Le langage contre-factuel une fois augmenté d'une sémantique structurelle, peut produire des points de vue plausibles et élégants de la causalité réelle et résout plusieurs problèmes rencontrés dans les définitions traditionnelles.

Les principes essentiels de cette approche sont :

- L'utilisation d'équations structurelles pour modéliser les mécanismes causaux et contre-factuels ;

- L'utilisation de la notion de contre-factuelle uniforme pour coder et distinguer les faits, les actions, les réalisations, les processus et les contingences ;

- Recherche soigneuse des contingences pour éliminer les altérations avec le processus causal actif.

Cette approche met l'accent sur l'importance du choix de la granularité du modèle causal. Ceci peut être présenté comme un manque dans cette approche. Elle montre que les structures internes du processus supposé pour prendre en compte l'histoire causale joue un rôle crucial dans les jugements sur la cause réelle, et donc il est important de projeter proprement de telles histoires dans un langage qui représente ses structures explicitement. Cette approche est construite sur un tel langage. Comme déjà souligner l'approche proposée dépend du choix de l'ensemble « le bon » des variables qui modélisent la situation, lesquelles mettre comme exogènes, et celles qu'il faut mettre comme endogènes. Il n'y a pas une théorie pour faire ces choix. Ceci est une bonne direction pour la recherche.

Nous avons également effectué une comparaison entre cette approche et l'approche normative afin de mettre en relief les points compatibles et les problèmes qui peuvent être résolu par les méthodes, ainsi que les points de divergences.

J. Y. Halpern et J. Pearl ont proposé une définition formelle de l'explication en terme de causalité. Il n'y a pas beaucoup définitions formelles de l'explication en terme de causalité dans la littérature. Une des petites exceptions est celle de Lewis [Lewis86], qui défend l'idée que «pour expliquer un événement il faut fournir quelques informations sur sont histoire causale». Ce point de vue est compatible avec la définition de Halpern et Pearl, cependant il n'y a aucune définition formelle donnée pour permettre une comparaison entre les approches. Dans le chapitre suivant nous allons voir cette modélisation proposée pour l'explication causale ainsi qu'une étude sur la complexité de cette dernière.

I.Introduction

La génération automatique d'une explication adéquate est une tache essentielle pour la planification, le diagnostique et le traitement de langage naturel. Un système faisant l'inférence doit être capable d'expliquer ses résultats et ses recommandations pour pouvoir acquérir la confiance de ses utilisateurs [Dubois03]. L'idée de base selon Halpern et Pearl est qu'une explication est un fait qui n'est pas connu pour certains, mais, s'il s'en trouve vrai, il constituera une cause réelle de ce que nous voulons expliquer, sans tenir compte de l'incertitude initiale de l'agent [Halpern01][Halpern05].

Cette définition implique la causalité et la connaissance. En suivant l'idée de Gärdenfor [Gärdenfors88], une explication est relative à l'état épistémologique de l'agent. Ce qui est une explication pour un agent peut ne pas être une explication pour un autre agent. Suivant encore Gärdenfor, en permettant aux explications de contenir (des parties) d'un modèle causal. Un point important est que la notion de l'explication causale proposée par Halpern et Pearl [Halpern05] est très différente du concept des explications causales proposées dans d'autres travaux en IA dans [Geffner92][Lifschitz97][Turner99][Eiter04][Mokhtari97][Kayser98].
Dans ce chapitre nous allons voir la modélisation de l'explication causale qui se base essentiellement sur la définition de la cause réelle déjà vu dans le chapitre précédent. La notion d'explication est une conséquence logique de l'étude de la causalité, Halpern et Pearl proposent de munir l'explication avec des probabilités [Halpern05], nous présentons dans la troisième partie la définition de l'explication partielle et la puissance d'une explication qui sera accompagnée de quelques exemples. Nous terminons ce chapitre par une généralisation de la définition de l'explication pour le cas des situations. Nous présentons ensuite une étude de complexité de l'explication [Eiter04].

II. Explication

La génération automatique d'explications qui s'avère une tâche essentielle dans la planification, le diagnostic et le langage naturel, ceci exige une analyse formelle du concept de la cause réelle. L'explication a souvent été étudiée en tant qu'entité, et dans ce cas comme une entité à part. L'entité en question est réifiée en une loi scientifique [Hempel66], un acte de langage [Achinstein83], une structure mnésique [Schank88], un graphique à finalité didactique [Balacheff90]. Lorsque nous devons fournir une explication, quelles informations allons-nous chercher et sélectionner ? Dans quel ordre, de quelle façon les produisons-nous ? Expliquer, c'est faire comprendre par un développement oral ou écrit ou par des gestes (Petit l'arrousse 1995), mais c'est aussi éclaircir, exposer, commenter, formuler une justification, rendre explicite, argumenter, solutionner. Une explication concerne toute opération impliquée dans la constitution du sens d'un phénomène.

Un exemple de Gärdenfor : un agent qui cherche une explication de «pourquoi ne pas considérer le fait que M^r Johanson ayant travaillé pendant des années dans une usine d'amiante comme une partie de l'explication de sa maladie 'le cancer du poumon', s'il est déjà au courant de ce fait». Pour un tel agent, une explication de la maladie de johanson peut contenir un modèle causal décrivant la connexion entre les fibres d'amiante et le cancer du poumon. Par ailleurs, pour quelqu'un qui connaît le modèle causal, mais n'est pas au courant que $M^.$ Johanson a travaillé dans une usine d'amiante, l'explication doit impliquer le travail de $M^.$ Johanson mais ne va pas mentionner le modèle causal.

La définition d'Halpern et Pearl diffère de celle de Gärdenfor, (et des autres définitions dans la littérature) par la façon dont elle est formulée pour représenter la notion de causalité. La définition n'est pas basée sur des corrélations probabilistiques ou des pertinences statistiques, et de cette façon elle est capable de traiter les difficultés dues aux explications ordinaires.

La définition de la causalité suppose que le modèle causal est donné et que tous les effets pertinents le sont aussi ; le problème était de trouver lesquels de ces faits étaient les causes. Par contre, le rôle de l'explication est de fournir les informations nécessaires pour établir la causation. Une explication dépend de ce que nous connaissons (ou bien de nous croyons). Comme conséquence, la définition de l'explication est relative à l'état épistémologique de l'agent (comme dans Gärdenfor [Gärdenfors88])[Halpern01][Halpern05]. Il sera naturel, à partir de ce point de vue, qu'une explication va contenir des fragments du modèle causal, ou bien qu'elle fasse référence aux lois physiques représentant les connexions entre la cause et l'effet.

La définition de l'explication est motivée par les intuitions suivantes. Un agent dans un état épistémologique K se pose la question : pourquoi φ a lieu ? Une bonne réponse doit fournir des informations qui vont avec K et que l'agent peut voir que c'est une cause de φ.

Comment capturer l'état épistémologique d'un agent ? Premièrement, Halpern et Pearl considèrent le cas où le modèle structurel est connu et que seul le contexte est incertain. Dans ce cas, une façon de décrire l'état épistémologique d'un agent est par la description de l'ensemble des contextes que l'agent considère possible.

II.1 Définition

Etant donné un modèle structurel M, $\overrightarrow{X} = \vec{x}$ est une explication de φ relativement à un ensemble K de contextes si les conditions suivantes sont vérifiées :

- EX1 : $(M, \vec{u}) \models \varphi$ pour chaque contexte $\vec{u} \in K$ (φ doit être vraie dans tous les contextes que l'agent considère possible- l'agent considère ce qu'il essaie d'expliquer comme un fait établi).
- Ex2 : $\overrightarrow{X} = \vec{x}$ est une cause suffisante de φ dans $(M, \vec{u})$ pour chaque $\vec{u} \in K$ tel que $(M,u) \models \overrightarrow{X} = \vec{x}$.
- Ex3 : $\overrightarrow{X}$ est minimal, pas de sous-ensemble de $\overrightarrow{X}$ satisfait la condition Ex2.
- Ex4 : $(M, \vec{u}) \models \neg(\overrightarrow{X} = \vec{x})$ pour un certain $\vec{u} \in K$ et $(M, \vec{u}') \models \overrightarrow{X} = \vec{x}$ pour un certain $\vec{u}' \in K$. (Ceci dit simplement que l'agent considère un contexte possible où l'explication est fausse, donc une explication n'est pas connue au début et considère un contexte possible là où l'explication n'est pas vraie, ce n'est pas trivial).

La condition Ex4 exprime que l'explication n'est pas connue même si elle peut sembler incompatible avec l'usage linguistique. Une personne découvre un effet A et dit «Ah » ceci explique pourquoi B a eu lieu. En fait, A n'est pas une explication de B relativement à l'état épistémologique après que A est découverte, puisqu'en ce moment là A est connu. Cependant, A peut être considéré comme une cause de B relativement à l'état épistémologique avant la découverte de A. Bien qu'il y ait une cause pour chaque événement φ (par fois elle semble être une cause triviale de φ), un des effets de la condition Ex4 est qu'il y a des événements sans explication 'événements inexplicables'.

II.2 Exemple explicatif [Halpern05][Sosa93]

Reprenons l'exemple 3.1 du chapitre précédent, supposons qu'il y ait une forte pluie en Avril et des tempêtes avec des éclairs pendant les deux mois suivants ; et en Juin il y a eu le feu de la forêt. S'il n'y avait pas eu une forte pluie en Avril, la forêt aurait pris feu en Mai. La première question est, est-ce la pluie d'Avril qui a causé le feu de la forêt ? Selon une analyse contre-factuelle naïve oui ; s'il n'y avait pas eu de pluie, il n'aurait pas eu de feu en Juin. Selon cette approche, ce n'est pas le cas : la pluie d'Avril n'a pas causé le feu, par contre elle a causé le feu en Juin ; par opposition en Mai.

Pour modéliser cette situation nous avons les variables suivantes :

- AS : pour décrire la pluie d'Avril (0 s'il n'a pas plu, 1 sinon) ;
- ES : pour décrire les deux tempêtes avec éclairs en Mai et Juin avec, (0,0) pas d'éclair ; (1,0) éclair en Mai et pas en Juin ; (0,1) pas d'éclair en Mai, mais en Juin ; (1,1) des éclairs en Mai et en Juin ;
- F : pour décrire le feu de la forêt avec 0 s'il n'y a pas de feu ; 1 s'il y a du feu en Mai ; 2 s'il y a du feu en Juin.

On ne décrit pas le contexte explicitement, on suppose que les valeurs de $\vec{u}$ sont telles que : il y' a de la pluie en Avril, il y a les deux tempêtes avec éclair en Mai et Juin, il y a suffisamment d'oxygène et il n'y a pas d'autres causes potentielles pour le feu de la forêt. Le réseau causal est très simple il y a un arc de AS vers F et un autre de ES vers F.

Nous avons :

- AS=1 est une cause du feu en Juin (F=2),
- AS=1 n'est pas une cause de feu(F=1 $\vee$F=2=). Si ES a l'une des valeurs (0,1), (1,0) ou (1,1) alors il y aura du feu (soit en mai ou en Juin) quelle que soit la valeur de AS. Par ailleurs, si ES=(0,0), alors il n'y a pas du feu quelle que soit la valeur de AS,
- ES=(1,1) est une cause de F=2 et de (F=1 $\vee$F=2=), le fait d'avoir une tempête en Mai et Juin cause de l'incendie.
- AS=1$\wedge$ES=(1,1) est une cause suffisante de F=2, chaque conjonction est une cause réelle.

Considérons le problème de l'explication. Supposons que l'agent sait qu'il y a eu une tempête, mais ne sait pas quand, et n'est pas au courant s'il y a eu de la pluie en Avril ou pas. Dans ce cas, K comprend six contextes, chacun correspond à l'une des valeurs (1,0),(0,1),(1,1) de ES avec soit AS=0 ou AS=1. Il est facile de voir que AS=1 n'est pas une explication du feu (F=1$\vee$F=2) ; puisque ce n'est pas une cause du feu dans aucun des contextes de K, d'une façon similaire si AS=0, alors que ES=(1,1),ES=(1,0) et ES=(0,1) chacune d'elle est une explication du feu.

Si on suppose que l'on cherche une explication pour le feu en Juin, alors l'ensemble des contextes K va contenir uniquement ceux compatibles avec la notion du feu en Juin. Supposons que K contient trois contextes {AS=1 et ES=(0,1)}, {AS=1 et ES=(1,1)} et {AS=0 et ES=(0,1)}. Dans ce cas, AS=1, ES=(0,1) et ES=(1,1) chacune d'elle est une explication du feu en Juin (dans le cas de AS=1 on considère l'attribution ES=(1,1).

Finalement, si l'agent sait qu'il a eu une tempête en Mai et Juin et une forte pluie en Avril (K contient un seul contexte), alors il n'y a pas d'explication ni du feu, ni du feu en Juin. Car il est impossible de satisfaire la condition Ex4. Informellement, car l'agent est toujours au courant du pourquoi il y a eu du feu en Juin.

Notez que, comme pour les causes, les explications disjonctives ne sont pas permises, car le fait de dire que φ est causée soit par A soit par B (mais je ne sais pas lequel) n'a pas beaucoup de sens.

Considérons l'exemple du scénario conjonctif des deux pyromanes. Supposons que le modèle structurel est tel que les seules causes sont les pyromanes, l'éclair ou un feu de camp inattendu. Si on permet l'explication disjonctive, ce qu'est une explication de φ ? Une explication candidate est «il y avait les deux pyromanes ou l'éclair ou bien le feu de camp inattendu». Mais cette explication ne satisfait pas la condition Ex4, puisque la disjonction est vraie dans chaque contexte dans K. Par ailleurs, la disjonction de deux clauses de ces trois clauses constitue une explication (satisfait Ex4), si on ajoute d'autres causes potentielles ceci constituera une explication, la seule chose et qu'on ne fait pas la disjonction de toutes les causes possibles.

On croit que, dans le cas de l'explication disjonctive il semble approprié de capturer ceci directement dans le modèle causal par des variables qui représentent cette disjonction. Par exemple, dans l'exemple des deux pyromanes, où il y a d'autres causes potentielles du feu de la forêt, si on veut accepter l'idée qu'un pyromane est une explication sans mentionner qui est-t-il, alors on peut facilement le faire en remplaçant les deux variables ML_1 et ML_2 par une variable ML, dans ce cas ML=1 devient une explication, sans avoir besoin d'une explication disjonctive.

Un autre point par rapport à cette définition : Quand on cherche une explication de φ, non seulement elle explique φ, mais on suppose de plus qu'elle est vraie dans le 'monde actuel'.

La définition précédente ne fait pas de référence au monde réel, uniquement à l'état épistémologique de l'agent. Mais il n'y a pas de difficultés d'ajouter et d'imposer que l'explication soit vraie dans ce monde. On parlera donc, d'une explication relativement à la paire $(K,\vec{u})$, où K est l'ensemble des contextes. Intuitivement, $\vec{u}$ décrit le contexte dans le monde actuel. Pour que l'explication soit vraie dans le monde actuel il faut avoir $(M,\vec{u}) \models \overrightarrow{X} = \vec{x}$. Même si ceci n'avait été mentionné dans la définition précédente, le fait de l'ajouter ne pose pas de contradictions. Une fois qu'on a le monde réel comme une partie du modèle, on a besoin que $\vec{u} \in K$. Cette condition fait en sorte que K représente la connaissance d'un agent plutôt que ses croyances, le contexte actuel est l'un des contextes que l'agent considère possible.

III. Explication partielle et la puissance explicative

Les explications ne sont pas toutes bonnes. Quelques explications sont plus appropriées que d'autres. Une manière de définir la *Qualité* d'une explication est d'ajouter des probabilités au modèle. Supposons qu'un agent a une probabilité sur l'ensemble K des contextes possibles. Dans ce cas, on considère la probabilité de l'ensemble des contextes où l'explication ($\overrightarrow{X} = \vec{x}$) est vraie.

Formellement, supposant qu'il y a une probabilité Pr sur l'ensemble K des contextes possibles. Alors la probabilité de l'explication $\overrightarrow{X} = \vec{x}$ est Pr($\overrightarrow{X} = \vec{x}$). La probabilité d'une explication capture clairement une partie des aspects importants du degré de la *Qualité* d'une explication. L'autre partie concerne le degré pour lequel une explication remplit bien son rôle (relativement à φ) dans les différents contextes considérés. Ceci devient plus clair quand on considère les *explications partielles*. L'exemple suivant, pris de [Gärdenfor88] est un où l'explication partielle joue un rôle.

Exemple 3.1

Supposons que je vois victoria bronzée et je cherche une explication. Supposons que le modèle causal contient des variables pour «victoria prend des vacances aux Iles canaris», «elle s'est exposée au soleil aux Iles canaris », «elle est allée dans un salon de bronzage ». L'ensemble K contient les contextes pour toutes les attributions des variables compatibles au fait que Victoria soit bronzée. Notons que, en particulier, il y a un contexte où Victoria est partie aux Iles canaris (elle ne s'est pas bronzée là-bas, car elle ne s'est pas exposée au soleil) et elle a visité le salon de bronzage. Selon la définition de l'explication de Gärdenfor, le fait que Victoria soit partie en vacance aux Iles canaris est une explication acceptable de son bronzage. Mais selon la définition proposée, ce n'est pas une explication (relativement aux contextes dans K), puisqu'il y a un contexte $\vec{u}^* \in K$ où Victoria part en vacance aux Iles canaris mais il n'y avait pas de soleil, dans $\vec{u}^*$ la cause réelle de son bronzage est le fait qu'elle ait visité le salon de bronzage, et non pas les vacances. La condition EX2 n'est pas satisfaite. Cependant, intuitivement, elle est presque satisfaite, elle est satisfaite dans tous les contextes dans K, où Victoria prend des vacances aux Iles canaris mais pas dans $\vec{u}^*$.

La seule explication complète est ' Victoria prend des vacances aux Iles canaris *et* il y avait de soleil'. 'Victoria prend des vacances aux Iles canaris' est une *explication partielle ;* dans cet exemple, l'*explication partielle* peut devenir une explication complète en ajoutant une conjonction. Mais toute explication partielle ne pourrait être étendue à une explication complète.

Exemple 3.2

Supposons que le son d'une télévision fonctionne alors qu'il n'y a pas d'image. De plus, la seule cause du fait qu'il n'y a pas d'image que l'agent sait que le tube ne fonctionne pas. Cependant, l'agent sait aussi que par fois, malgré que le tube fonctionne, l'image n'est pas présente sur l'écran. Intuitivement, il n'y a pas d'image «pour des raisons inexplicables ». Ceci est représenté dans le réseau causal par la figure1 avec :

- T : décrit si le tube fonctionne ou pas (1 s'il fonctionne, 0 si non),
- P décrit s'il y a image ou pas (1 si l'image est présente,0 si non),
- Les deux variables exogènes U_0 détermine l'état du tube : $T=U_0$. La variable exogène U_1 représente les autres causes possibles ' causes mystérieuses'.

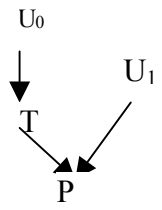


Figure 1 : La télévision sans images

Si $U_1=0$, alors qu'il y a ou pas d'images dépend seulement du statut du tube, i. e., P=T. Par ailleurs, si $U_1=1$, alors il n'y a pas d'image (P=0) indépendamment du statut du tube T. Donc, dans le contexte où $U_1=1$, T=0 n'est pas une cause de P=0. Maintenant supposons que K inclut un contexte $\vec{u}_{00}$ où $U_0=U_1=0$ et $\vec{u}_{10}$ où $U_0=1$ et $U_1=0$. La seule cause de P=0 dans les deux contextes est P=0 elle-même.

(T=0 n'est pas l'explication de P=0 dans $\vec{u}_{10}$, puisque P=0 même si T=1). Comme résultat il n'y a pas d'explication de P=0 relativement à un état épistémologique K qui inclut $\vec{u}_{00}$ et $\vec{u}_{10}$. (Ex4 exclut les explications triviales, P=0).

Autrement, T=0 est une cause de P=0 dans tous les autres contextes dans K satisfaisant T=0 autre que $\vec{u}_{00}$. Si la probabilité de $\vec{u}_{00}$ (capture l'intuition qu'il est moins probable qu'il y est quelque chose créant le non-fonctionnement de la télévision) alors on sera ramené à autoriser T=0 comme une bonne explication partielle de P=0.

Notez que, si on modifie le modèle causal, en ajoutant une variable endogène I, correspondant aux causes inexplicables U_1 (avec $I=U_1$), alors I=0 est une cause de P=0 dans le contexte $\vec{u}_{10}$ (I=0 et T=0 est une cause de P=0 dans le contexte $\vec{u}_{00}$). Dans ce modèle, I=0 est une explication de P=0.

Les exemples précédents illustrent un point important : l'ajout d'une variable endogène correspondant à une variable exogène pour faire en sorte qu'il y ait une explication alors qu'il n'y en avait pas avant. Ce phénomène d'ajouter des 'names' pour créer des explications est souvent utilisé.

III.3 Définition

Soit $K^{\wp}_{X=\chi,\varphi}$ le plus grand sous-ensemble K' de K tel que $\overrightarrow{X}=\vec{x}$ est une explication de φ relativement à $K^{\wp}_{X=\chi,\varphi}$ (il est facile de voir que le plus grand sous-ensemble consiste en tous les contextes de K sauf ceux où $\overrightarrow{X}=\vec{x}$ est vraie mais ce n'est pas une cause suffisante de φ) alors $\overrightarrow{X}=\vec{x}$ est explication partielle de φ avec une *Qualité* $\Pr(K^{\wp}_{X=\chi,\varphi} \mid \overrightarrow{X}=\vec{x})$.

Dans l'exemple 3.2, si l'agent croit qu'il fait beau aux Ils canaris avec une probabilité 0,9 (i. e., la probabilité qu'il fait beau étant donnée, que Victoria soit bronzée et soit allée aux Ils canaris est 0,9), alors le voyage de Victoria aux Ils canaris est une explication partielle du fait qu'elle soit bronzée avec une *Qualité* 0,9. L'ensemble pertinent K' consiste à considérer des contextes où il fait beau aux Iles canaris.

Par ailleurs, dans l'exemple 3.2, si l'agent croit que la probabilité que le tube soit défaillant et qu'il y ait d'autres causes mystérieuses est 0.1, alors T=0 est une explication partielle de P=0 avec une *Qualité* de 0.9 (K' correspond à tous les contextes où $U_1=0$).

Il est clair qu'une explication complète est une explication partielle avec une *Qualité* =1, mais on est souvent satisfait par une explication partielle $\overrightarrow{X}=\vec{x}$ avec une grande probabilité ($\Pr(\overrightarrow{X}=\vec{x})$ est grande). Notez qu'il y a une tension entre la qualité d'une explication et sa probabilité.

Ces idées conduisent à la définition de *la Puissance d'explication.* Reprenons l'exemple du feu de la forêt, supposons qu'il y ait une variable endogène O correspondant à la présence de l'oxygène. Si O=1 est vrai dans tous les contextes, donc O=1 ne peut être la cause du feu selon la condition Ex4. Si l'agent sait qu'il y a de l'oxygène alors la présence de l'oxygène ne peut faire partie d'une explication, si on suppose qu'il y a un contexte où O=0(c'est moins probable), dans ce cas O=1 devient une bonne explication partielle du feu. Néanmoins, c'est intuitivement une explication avec, une petite *Puissance d'explication*.

Supposons qu'il y ait une distribution de probabilité Pr^- sur l'ensemble K^- des contextes qui inclut K. Pr^- représente la probabilité de 'pré-observation' de l'agent, i. e., la probabilité à priori de l'agent sur φ, avant qu'elle soit observée ou découverte. Ainsi, Pr est le résultat de conditionner P^-r sur φ et K correspondant aux contextes dans K^- satisfaisant φ. Gärdenfor identifie la *Puissance d'explication* d'une explication (partielle) $\overrightarrow{X} = \vec{x}$ de φ par $Pr^-(\varphi | \overrightarrow{X} = \vec{x})$. Si cette probabilité est plus grande que $Pr^-(\varphi)$, alors cette explication rend φ plus probable. Notons que, puisque K consiste en tous les contextes de K^- où φ est vraie, la notion de *Puissance d'explication* selon Gärdenfor est équivalente à $Pr^-(K | \overrightarrow{X} = \vec{x})$[Gardenfor 88]. (Cependant la définition de Gärdenfor a des problèmes elle confond corrélation et causation).

Selon Halpern, la meilleure mesure *de la Puissance d'explication* de $\overrightarrow{X} = \vec{x}$ est $P^-r(K^{\wp}_{X=\vec{x},\varphi} | \overrightarrow{X} = \vec{x})$. Notons que les deux définitions s'entendent sur le fait que $\overrightarrow{X} = \vec{x}$ est une explication complète (alors $K^{\wp}_{X=\vec{x},\varphi}$ est juste K, l'ensemble des contextes dans K^- où φ est vraie). En particulier, elles s'entendent que O=1 a une *puissance d'explication* petite tandis que ML_1=1 a une grande *puissance d'explication.* La différence entre ces définitions apparaît s'il y a des contextes où φ et $\overrightarrow{X} = \vec{x}$ sont les deux vrais, mais $\overrightarrow{X} = \vec{x}$ n'est pas une cause de φ. Dans l'exemple 3.2; le contexte $\vec{u}^*$ est l'un des contextes où victoria est partie aux Iles canaris, mais cela n'était pas une explication de son bronzage, puisqu'il ne faisait pas beau.

Cette définition s'accorde sur quelques traits avec celle de Gärdenfor : une explication est relative à l'état épistémologique d'un agent. Gärdenfor considère l'état épistémologique 'constracted' caractérisée par la distribution Pr^-. Intuitivement, Pr^- décrit les croyances de l'agent avant de découvrir de φ. (Plus précisément, elle décrit un état épistémologique aussi proche que possible de Pr où l'agent n'attribue pas la probabilité 1 à φ). Si l'état épistémologique de l'agent est le résultat d'une observation de φ, alors on peut prendre Pr comme étant le résultat de conditionner Pr^- sur φ.

Cependant Gärdenfor ne suppose pas qu'il y a dans tous les cas de telles connexions entre Pr et Pr^-. Pour Gärdenfor, $\overrightarrow{X} = \vec{x}$ est une explication de φ relativement à Pr si :

1) $P(\varphi) = 1$,
2) $0 < Pr(X=x) < 1$,
3) $P^-r(\varphi | X=x) > P^-r(\varphi)$.

La condition (1) est la probabilité analogue à Ex1. La condition (2) est la probabilité analogue de Ex4, et la condition (3) explique que connaître l'explication augmente le degré de vraisemblance de φ.

Gärdenfor insiste sur *la Puissance d'explication* d'une explication, mais ne prend pas en compte ces probabilités à priori. (La définition de Gärdenfor souffre d'un autre problème : puisqu'il n'y a pas de condition de minimalité comme Ex3, si $\overrightarrow{X} = \vec{x}$ est une explication de φ, alors $X=x \wedge Y=y$ est aussi une explication.

Contrairement à la définition de Gärdenfor, l'approche dominante de l'explication dans la littérature en IA, i.e., l'approche du Maximum à posteriori (MAP) insiste sur la probabilité d'une explication, ce que nous avons noté $Pr(\overrightarrow{X} = \vec{x})$. L'approche MAP est basée sur l'intuition suivante : La meilleure explication d'une observation est l'état du monde qui est le plus probable étant donnée l'évidence. Il y a de nombreux problèmes liés à ces approches basées sur le MAP. Le plus important est : elles ignorent la notion de puissance d'explication (Voir [Chajewska97]).

IV. Généralisation de l'explication

En général, un agent a des incertitudes quant au modèle causal, donc une explication doit contenir des informations sur ce dernier. Il est relativement simple de faire une extension de la définition précédente de l'explication pour s'accommoder avec cette prévision. Maintenant un état épistémologique K consiste non-seulement en des contextes, mais en des pairs(M, $\vec{u}$), où M est un modèle causal et $\vec{u}$ un contexte, (cette paire est une situation). Intuitivement, maintenant une explication doit contenir des informations causales et des faits qui sont vrais.

Une explication (générale) à la forme ($\psi, \vec{X} = \vec{x}$), où ψ est une formule arbitraire dans le langage causale et $\vec{X} = \vec{x}$ est une conjonction d'événements primitifs. Le composant ψ contient des informations causales. Le premier composant dans une explication générale est vue comme une restriction de l'ensemble des modèle causaux. Plus précisément, étant donné un modèle causal M, on dit que ψ est valide dans M, et on écrit $M \models \psi$ si $(M, \vec{u}) \models \psi$, pour tous les contextes $\vec{u}$ consistant avec M. Avec ces connaissances, il est facile d'établir la définition générale.

IV.1 Définition [Halpern05]

($\psi, \vec{X} = \vec{x}$) est une explication de φ relativement à un ensemble K de situations si les conditions suivantes sont vraies :

- Ex1 : $(M, \vec{u}) \models \varphi$ pour chaque paire situation $(M, \vec{u}) \in K$.
- Ex2 : $\vec{X} = \vec{x}$ est une cause suffisant de φ dans $(M, \vec{u}) \in K$, tel que $(M, \vec{u}) \models \vec{X} = \vec{x}$ et $M \models \psi$. $\vec{X} = \vec{x}$ est une cause suffisante de φ dans $(M, \vec{u})$.
- Ex3 : ($\psi, \vec{X} = \vec{x}$) est minimale, il n'y a pas une paire (ψ', $\vec{X}$'=$\vec{x}$')$\neq$(ψ, $\vec{X} = \vec{x}$) satisfaisant Ex2, tel que $\{M'' \in M(K) : M'' \models \psi'\} \supseteq \{ M'' \in M(K) : M'' \models \psi\}$,où $M(K)=\{M \in (M, \vec{u}) \in K$, pour un certain u$\}$, $\vec{X}' \subseteq \vec{X}$, et $\vec{x}'$ est la restriction de $\vec{x}$ pour les variables dans $\vec{X}$'. informellement, il n'y a pas un sous-ensemble de $\vec{X}$ qui fournit une cause suffisante de φ dans plus de contextes que ceux où ψ est vraie.
- Ex4 : $(M, \vec{u}) \models \neg(\vec{X} = \vec{x})$ pour certains $(M, \vec{u}) \in K$ et $(M', \vec{u}') \models \vec{X} = \vec{x}$ pour certains $(M', \vec{u}') \in K$.

Notons que, dans Ex2, maintenant on fait une restriction aux situations $(M, \vec{u}) \in K$, qui satisfait les deux parties de l'explication ($\psi, \vec{X} = \vec{x}$), donc $M \models \psi$ et $(M, \vec{u}) \models \vec{X} = \vec{x}$. De plus, bien que les deux composants d'une explication soient des formules dans le langage causal, ils jouent des rôles très différents.

Le premier composant sert à faire une restriction de l'ensemble des modèles causaux considéré (à ceux avec la structure appropriée) ; le second décrit la cause de φ dans l'ensemble résultant des situations. La définition 2.1 est un cas spécial de la définition 4.1 où il n'y avait pas d'incertitude sur la structure causale.

Exemple 4.2

En utilisant la définition générale de la causalité, considérons le fameux exemple du *Paressis*, qui a causé tant de problèmes à plusieurs formalismes [Scriven59].

> *La Paressis* se développe uniquement chez les patients qui sont *Syphilitique* depuis longtemps, mais un petit nombre seulement de patients qui sont *Syphilitique* développe en fait le *Paressis*. De plus, selon Scriven, il n'y a pas d'autres facteurs qui sont connus comme pertinents au développement de *Paressis*. Cette description est capturée par un simple modèle causal M_p. Il y a deux variables endogènes S (pour *Syphilitique*) et P (pour *Paressis*), et deux variables exogènes U_1, les facteurs connus qui déterminent S et U_2 qui représente intuitivement la disposition de *Paressis*', i. e., les facteurs qui déterminent, en conjonction avec *Syphilis*, si le *Paressis* se développe actuellement ou pas.

Un agent qui connaît ce modèle causal et qui sait que le patient a la *Paressis* n'as pas besoin d'une explication : il sait sans que personne ne le lui dise que le patient doit avoir la *Syphilis* et que U_2=1.

Par ailleurs, un agent qui ne connaît pas le modèle causal (i. e., considère un nombre de modèles causaux de *Paressis* possibles), (φ_p,S=1) est une explication du *Paressis*, où φ_p est une formule qui caractérise M_P.

La définition 4.1 peut être étendue pour considérer les probabilités. Actuellement la probabilité joue un rôle à deux niveaux. Premièrement, il y a une probabilité sur les situations dans K, analogues aux probabilités dans les contextes discutés dans la section 3. En utilisant ces probabilités, il est possible de parler de la *Qualité* d'une explication partielle et de parler sur *la Puissance d'explication*. En plus, les probabilités peuvent aussi être ajoutées aux modèles causaux pour avoir un modèle causal probabiliste. Un modèle causal probabiliste est un tuple M=(S,F,Pr) ou M=(S,F) est un modèle causal et Pr est une mesure de probabilité sur les contextes définis par la signature S de M.

Une fois qu'on permet des modèles causaux probabilistes, l'explication peut contenir des expressions de la forme ' avec la probabilité 0.9, travailler avec de l'amiante cause le cancer du poumon'. Les modèles causaux probabilistes peuvent capturer des informations statistiques de ceux considérées par Gärdenfor et Hempel. Pour introduire de telles expressions dans ce travail, il faut faire une extension du langage pour permettre des expressions sur les probabilités. Par exemple, Pr([X←3](Y=1))=0.9, «la probabilité de mettre X à 3 résulte du fait que Y est égal 1, pourrait être une formule dans le langage. De telle formule peuvent devenir alors une partie du premier composant ψ dans l'explication générale[Halpern05].

V. Complexité

Avant de présenter une étude de la complexité de cette approche effectuée par [Eiter04] nous faisons un bref rappel sur les notions de base sur la complexité.

V.1 Notions de base

1. Les problèmes de la classe P

Les problèmes de cette classe sont faciles à résoudre. Ce sont des problèmes solubles par des algorithmes efficaces. La classe P regroupe tous les problèmes pour lesquels il existe des algorithmes dont le temps d'exécution est borné par une fonction polynomiale. Cette classe se définit par rapport aux machines de Turing déterministe.
Un problème appartient à P si toutes les instances de ce problème appartiennent à cette classe.
Un problème est traitable s'il existe un algorithme qu'il le résout en un temps polynomial.

Un problème est intractable : s'il n'existe pas un algorithme polynomial qu'il le résout.

2. Les problèmes de la classe NP

La classe NP (Nom Polynomiale) regroupe les problèmes solubles par des algorithmes non déterministes.
La classe co-NP est la classe complémentaire des problèmes de type A (leur réponse est oui si la solution de A pour x est non) ce sont des problèmes NP. Cette classe n'est pas équivalente à P ni à NP, ils sont souvent représentés par le diagramme suivant :

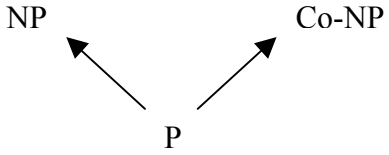


Figure2 : Relation entre les trois classes

La classe P est contenue dans les deux classes NP et co-NP.

3. Réduction Polynomiale

Soient A un problème NP, une réduction polynomiale du problème A au problème B est une fonction tel que

$$x \in A \text{ Ssi } f(x) \in B$$

Et tel que f(x) se calcule en un temps polynomial en fonction de la taille de x.
S'il existe une réduction polynomiale entre A et B, on dit que A peut être réduit polynomialement à B (noté $A \leq^p B$).
Si la fonction f est polynomiale, chaque instance de A peut être résolue en la réduisant à B, f est la réduction de A à B.

Définition1 :

Si, pour chaque problème $A \in NP$ on a $A \leq^p B$, alors B est un problème NP-hard.

Définition2 :

Si pour chaque problème $A \in NP$ et est NP-hard alors c'est un problème NP-complet.

Pour prouver qu'un problème C est NP-hard, prouver que $B \leq^p C$, et B est NP-complet. Puisque $\leq^p$ est transitive alors ceci implique que C est NP-hard.
Si A se réduit au SAT alors A est NP sinon A est NP-complet.

4. Les problèmes NP-complets

La classe des problèmes NP-complet est incluse dans la classe des problèmes NP. Cette classe regroupe les problèmes pour lesquels il n'existe pas d'algorithme de résolution en un temps polynomiale. Ce sont des problèmes intraitables difficiles à résoudre car leurs algorithmes de résolution prennent un temps exponentiel. Cependant, pour ce type de problème, on peut vérifier en un temps polynomial si une solution est proposée ou nom.
Les problèmes NP-complet sont plus difficiles que les problèmes à temps polynomiaux, dans le sens que c'est difficile de prouvé qu'ils sont tractables. Cependant, plusieurs problèmes sont NP-hard mais pas NP-complet.

5. Les problèmes NP-difficiles

Les problèmes NP-difficiles n'englobent pas seulement les problèmes de décisions mais aussi les problèmes d'optimisation. Un problème NP-difficile est au moins aussi difficile qu'un problème NP-complet. Les problèmes NP-complet sont les seuls problèmes NP-difficiles de NP. Les problèmes d'optimisation dont les problèmes de décision associés sont NP-complets, sont NP-difficiles.

Il y a deux problèmes qui restent difficiles à être capturé par ces deux classes. Pour cette raison une hiérarchie polynomiale et introduite. Cette hiérarchie est définie en terme d'Oracle et de machine de Turing avec Oracles. Un Oracle est un mécanisme qui permet de résoudre un problème sans besoin d'un temps extra.
Un Oracle-NP est un mécanisme capable de résoudre un problème sans perte de temps.
Une machine de Turing ave Oracle est une machine avec accès à l' Oracle.

Soient les deux classes avec termes d'Oracles :

- La classe Δ^p_2 est une classe des problèmes qui peuvent être résolus en un temps polynomial avec une machine de Turing augmenté d'OracleNP. Essentiellement le problème doit être résolu en un temps polynomial. Clairement l'ensemble des problèmes NP est un sous-ensemble de la classe des problèmes Δ^p_2, puisque chaque problème NP peut être résolu par une machine de Turing augmenté d'Oracle en un temps fini.
- La classe Σ^p_2 est similaire, mais l'ensemble des problèmes qui peuvent être résolu en temps polynomial et par une machine de Turing non-déterministe par l'appel d'Oracle NP.
- La classe Π^p_2 est définie par l'ensemble des problèmes A qui sont complémentaire aux problèmes dans Σ^p_2. Π^p_2=Co-Σ^p_2

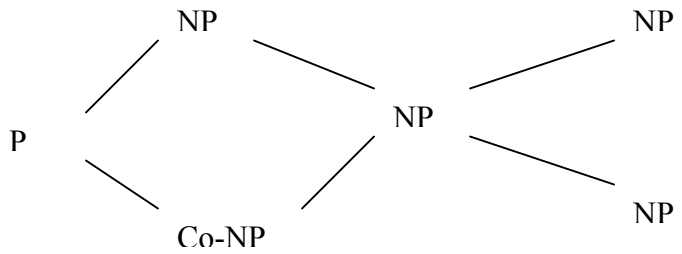


Figure3 : hiérarchie polynomiale

La hiérarchie polynomiale inclut les classes nommées Σ^p_i, Π^p_i et Δ^p_i pour tous les entiers positifs. La classe Σ^p_0, Π^p_0 et Δ^p_0 sont P, alors que Σ^p_i, Π^p_i sont NP et co-NP, respectivement.
La hiérarchie polynomiale est définie comme l'union de ces classes :

$$PH=\cup_{i\geq 0} \Sigma^p_i$$

Soit f une formule propositionnelle. Le problème $\forall\exists$QBF est définit comme : vérifier la validité de la formule de la forme $\forall X\exists Y$. F(X,Y). Ceci est un problème du prototype Π^p_2 complet.
Similairement $\exists\forall$QBF est le problème de vérification de validité de la formule $\exists X\forall Y$. F(X,Y) est un problème Σ^p_2 complet.
La classe générique Π^p_i, est un problème complet pour rechercher si $\forall X_1\exists X_2 \ldots X_i$ F($X_1,\ldots,X_i$) est vraie.
La classe générique Σ^p_i, est un problème complet pour rechercher la validité d'une formule qui doit commencer par le quantificateur existentiel.

L'utilisation d'Oracle peut être utilisée en instances de différents problèmes pour une suite de problèmes dans la classe, ce qui n'affecte pas la puissance de la complexité, un appel à l'oracle coûte une unité de temps.
$\Delta^p_2 = P^{NP}$ (respectivement $\Sigma^p_2 = P^{NP}$) dénote la classe des problèmes décidable en un temps polynomiale à l'aide de l'oracle-NP sur une machine de Turing déterministe (respectivement non déterministe).
Intuitivement, la puissance de complexité des classes dans la hiérarchie polynomiale augmente avec chaque niveau k.
Une autre classe est introduite $D^p_k = \{L \times L' \mid L \in \Sigma^p_k, L' \in \Pi^p_k\}$, k≥1, c'est la conjonction de Σ^p_k et Π^p_k. De tels problèmes peuvent être facilement résolus avec deux appels à l'oracle Σ^p_k, mais qui sont intuitivement facile puisque les appels peuvent être faits indépendamment l'un de l'autre. Autrement dit, les problèmes sont aussi difficiles que les problèmes Σ^p_k complet et Π^p_k complet.

La classe $P_{\parallel}^{\Sigma_k^p}$, k≥1 contient les problèmes de décision qui peuvent être résolus en un temps polynomial avec un seul appel circulaire d'appels parallèle à l'oracle NP plutôt que des appels en un ordre arbitraire.
Pour les problèmes de classification qui calculent une valeur de sortie (exemple : l'ensemble des atomes dans une formule classiqueφ). Une classe de fonctions similaires aux classes présentées précédemment. Parmi ces classes FP, $FP_{\parallel}^{NP}$, et $FP_{\parallel}^{\Sigma_{kk}^p}$ qui sont des classes analogues à ceux de P, $P_{\parallel}^{NP} = P_{\parallel}^{\Sigma_{11}^p}$ et $P^{\Sigma_{1k}^p}$, respectivement.

Toutes les classes dans la figure(3) sont résolues sous une réduction en temps polynomiale i.e., le problème Π à une transformation en temps polynomiale i.e., si le problème Π à une transformation en temps polynomiale Π′ de C, alors Π appartient aussi à la classe C.
Décider la satisfiabilité d'une formule booléenne φ, ce problème est bien connu NP-complet, alors que sont complément, décidé l'insatisfiabilité de ϕ est co-NP-complet. Décider si ϕ_1 et satisfaite et ϕ_2 est insatisfaite est D^p complet.
Généralement un problème complet pour la classe Σ^p_k(respectivement Π^p_k), k≥1, est de décider la satisfiabilité d'une Formule Booléenne Quantifiée(QFB) $Q_1X_1Q_2X_2\ldots Q_kX_k\ \phi$, tel que $X_1,X_2,\ldots,X_k$ un ensemble de variables, $Q_1Q_2\ldots Q_k$ est une séquence alternée de quantificateurs ∃ et ∀ tel que $Q_1=\exists$ (respectivement $Q_1=\forall$), et ϕ est une formule booléenne à travers les variables $X_1 \cup X_2 \cup \ldots \cup X_k$. Ayant deux FBQs ϕ_1 et ϕ_2 de la forme $Q_1X_1Q_2X_2\ldots Q_kX_k\ \phi$ avec $Q_1=\exists$, décider que ϕ_1 est valide et que ϕ_2 est non valide, est un problème complet pour la clase D_k^p. Finalement, ayant l FBQs : $\phi_1,..,\phi_l$, le problème de décider si un nombre de formules est valide parmi $\phi_1,..,\phi_l$, (respectivement calculer l'ensemble de toutes les formules valides parmi $\phi_1,..,\phi_l$) est complet pour $P_{\parallel}^{\Sigma_{1k}^p}$ (respectivement $FP_{\parallel}^{\Sigma_{1k}^p}$).
Les classe de complexité sont dans la figure 4. Ce sont des classes d'une hiérarchie polynomiale(PH), ou bien dérivées de celle-ci.

Rappelons que : $NP=\Sigma^P_1$, co-$NP=\Pi^P_1$, $\Sigma^P_{k+1} = NP^{\Sigma_k^p}$, et Π^P_k=co-Σ^P_k, k≥1 sont des classes dans PH et $P_{\parallel}^{NP} = P_{\parallel}^{\Sigma_1^p}$ et $FP_{\parallel}^{NP}$ $FP_{\parallel}^{\Sigma_1^p}$.

Pour plus de détails voir [Papadimitriou 94][Jonhson 90].

V.2 Résultats de Complexité [Eiter04]

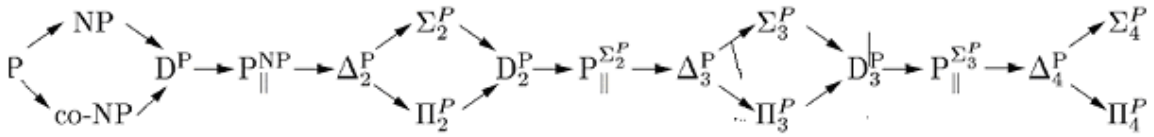


Figure 4: Classes de complexité

Eiter et Lukasiewicz proposent une étude de la complexité de cette approche. La notion de l'explication dans cette approche est basée sur la définition de la cause suffisante.

Le problème de la complexité pour l'explication est du à la définition elle-même, alors que le problème de l'existence de l'explication est associé à une tâche très importante qui est la recherche d'une explication pour un événement φ (la recherche d'un ensemble de variables).

Ayant un modèle M est un modèle causal récursif, avec la relation d'ordre < sur les éléments de V, telle que Y<X, si F_x dépend de la valeur de Y. Dans un tel modèle, chaque attribution de valeurs aux variables exogènes U=u détermine une attribution de valeur unique y pour chaque variable Y⊆Z.
X=x est un événement primitif, un ensemble d'événements est un ensemble d'événements primitifs sous les deux opérateurs logiques ¬ et ∧.
Une formule φ est vraie dans un modèle M sous un contexte u∈D(u), dénoté (M,u) ⊨φ abrégé φ(u) et ($M_{X\leftarrow x}$,u) ⊨φ abrégé $\varphi_X(u)$.

Proposition 1

Soit X⊆V et x∈D(X), u∈D(U) et un événement φ. Décider φ(u), $\varphi_X(u)$ est vraie ce fait en un temps polynomial.

V.2.1 Complexité d'une cause faible

Proposition 2

Soient un modèle causal M=(U,V,F),X⊆V, x∈D(X), u∈D(U), et un événement φ. Soit $X_0 \in X$ tel que dans le graphe causal correspondant X_0 n'est prédécesseur d'aucune variable dans φ. Soit X′=X\{X_0} et x′=x\X′.
X=x est cause faible de φ sous le contexte u ssi :

- $X_0(u)=x_0$
- X′=x′ est une cause faible de φ.

Décider si X=x est cause faible de φ sous le contexte u est complet pour Σ^p_2 (respectivement Π^p_2) dans le cas général (respectivement le cas binaire, ou D(X)={0,1}).

V.2.2 Complexité de l'explication

Explication :
Ayant un modèle causal M=(U,V,F),X⊆V, x∈D(X), un événement φ et un ensemble de contexte u∈K. Décider si X=x est une explication de φ relativement à u.

L'existence d'explication :
Ayant un modèle causal M=(U,V,F),X⊆V, x∈D(X), un événement φ et un ensemble de contexte u∈D(u). Décider si X′∈X et x′∈D(X) existe tel que X′=x′ est une explication de φ relativement à u.

Problème	Cas général	Cas binaire
L'explication	D_2^p -complet	D^p -complet
L'existence d'explication	$\sum_3^p$ -complet	$\sum_2^p$ -complet

Tableau 1 : La complexité des explications

Théorème1 : L'explication est un problème D_2^p -complet

La complexité de l'explication provient directement de la définition de l'explication. Le problème de l'existence d'explication est associé à la tâche importante de trouver une explication pour un événement φ.
Trouver une explication est au troisième niveau de la hiérarchie PH. Les explications sont plus difficiles à calculer que la cause faible, qui est du deuxième niveau de la hiérarchie PH, la reconnaissance d'une explication est légèrement difficile que la reconnaissance d'une cause faible.
Le problème de l'explication est D_2^p -complet, la condition EX2 revient à une conjonction d'un nombre linéaire du problème Σ^p_2, et la condition Ex3 est la négation de tels problèmes.
La condition Ex1 et Ex4 sont facilement vérifiées.
La vérification de l'explication, peut être réduite à une conjonction de problème en Σ^p_2, et Π^p_2. La difficulté de réduction D_2^p est vu comme une réduction à un problème de D_2^p -complet de décision, ayant un pair(ϕ_1,ϕ_2) deux QBFs, si ϕ_1 est valide et que ϕ_2 n'est pas valide.

Preuve

X=x est une explication de φ relativement à u Ssi Ex1-Ex4 sont vérifiées.
La condition Ex1, est la vérification de φ(u) pour tous les contextes dans K. La condition Ex4, vérifier que X(u)=x et X(u′)≠x pour quelque u, u′∈D(u). La vérification de ces deux conditions se fait en temps polynomial.
Dans la condition Ex2, l'ensemble K' de K pour tous les contextes u∈K tel que X(u)=x est calculé en un temps polynomial. C'est une vérification par rapport au conjonction polynomiale de « X=x est une cause faible de φ ». La condition Ex3 prend X′⊂X et vérifie que X′=x | X′ est une cause faible de φ sous tous les contextes u∈K tel que X′(u)=x | X′ est Σ^p_2. Décider la condition Ex3 est Π^p_2.

Théorème 2 : l'existence d'une explication est un problème Σ^p_3-complet.

La borne supérieure Σ^p_3 est due simplement à la borne supérieure Σ^p_2 de la reconnaissance de l'explication. La complexité provient de la satisfaction des conditions Ex1, Ex2 et Ex4 pour X=x, on ne peut pas conclure qu'il existe X′=x′ contenu dans X=x tel qu'il satisfait Ex1-Ex4 ; La minimisation de X=x afin de satisfaire la condition Ex3, le résultat X′=x′ de la minimisation risque de violé la condition Ex4. C'est cette condition qui rend le problème difficile.

Preuve
Le problème peut se réduire, en la supposition de $X' \subseteq X$ et $x' \in D(X)$, et la vérification que $X'=x'$ est une explication de φ relativement à K
La difficulté Σ^p_3 est montrée par une réduction au problème de décision d'une QFB $\phi=\exists B \forall C \exists D Y$, où *Y* est une formule propositionnelle sur les variables dans $B \cup C \cup D$. l'idée est de voir la quantification « $\exists B$ » comme la recherche de $X'=x'$, et « $\forall C \exists D Y$ » comme la vérification du complément de la cause faible dans la condition Ex3.

Dans le cas binaire, la complexité de tous les problèmes considérés est réduite d'un niveau dans PH. La complexité pour la cause faible se réduit de Σ^p_2 à NP.

V.2.3 Complexité de l'explication partielle et la puissance d'explication

L'idée de l'explication partielle est une généralisation de la définition de l'explication en plus d'attribution additionnelle de distributions de probabilités à travers un ensemble donné de contextes.
Ayant un modèle causal $M=(U,V,F), X \subseteq V$, $x \in D(X)$, un événement φ et un ensemble de contexte $u \in K$. $K'=K_{X \leftarrow x,\varphi}$ dénoté de plus grand ensemble de contexte tel que X=x est une explication de φ relativement à K'. $K_{X \leftarrow x,\varphi}$ est définie, si le sous-ensemble K' de K existe tel que X=x est une explication de φ relativement à K'.
Soit P une distribution de probabilité sur K, $P(K_{X \leftarrow x,\varphi} \mid X=x) = \sum_{u \in K^{\varphi}_{X=x}, X(u)=x} P(u) \,/\, \sum_{u \in K, X(u)=x} P(u)$.

X=x est une α-partielle explication de φ relativement à K ssi $K_{X \leftarrow x,\varphi}$ est définit et $P(K_{X \leftarrow x,\varphi} \mid X=x) \geq \alpha$. X=x est une explication partielle si X=x est une α-partielle explication de φ relativement à K ssi X=x est une α-partielle explication pour $\alpha>0$.

Supposition 3:
On suppose que les fonctions de probabilités P sont calculées en un temps polynomial.

α-partielle explication : Ayant un modèle causale $M=(U,V,F), X \subseteq V$, $x \in D(X)$, un événement φ et un ensemble de contexte $u \in K$. une fonction de probabilité P sur K, et $\alpha \geq 0$. Décider si X=x est α-partielle explication de φ relativement à K.

L'existence de α-partielle explication : Ayant un modèle causal $M=(U,V,F), X \subseteq V$, $x \in D(X)$, un événement φ et un ensemble de contexte $u \in K$. une fonction de probabilité P sur K, et $\alpha \geq 0$. Décider si $X' \subseteq X$ et $x' \in D(X)$ existe tel que $X'=x'$ est une α-partielle explication de φ relativement à K.

Partielle explication : Ayant un modèle causal $M=(U,V,F), X \subseteq V$, $x \in D(X)$, un événement φ et un ensemble de contexte $u \in K$. Une fonction de probabilité P sur K. Décider si X=x est une partielle explication de φ relativement à K.
Puissance d'explication : Ayant un modèle causal $M=(U,V,F), X \subseteq V$, $x \in D(X)$, un événement φ et un ensemble de contexte $u \in K$. Une fonction de probabilité P sur K, et $\alpha \geq 0$. Décider si $X' \subseteq X$ et $x' \in D(X)$ existe tel que $X'=x'$ est une α-partielle explication de φ relativement à K.

La complexité des ces problèmes est résumée dans le tableau suivant :

Problème	Cas général	Cas binaire
α-partielle explication	$P_{\|\|}^{\Sigma_2^p}$ complet	$P_{\|\|}^{NP}$ -complet
L'existence de α-partielle explication	Σ^p_3-complet	Σ^p_2-complet
Partielle explication	$P_{\|\|}^{\Sigma_2^p}$ complet	$P_{\|\|}^{NP}$ -complet
Puissance d'explication	$FP_{\|\|}^{\Sigma_2^p}$ -complet	$FP_{\|\|}^{NP}$ -complet

Tableau2 : La complexité de l'explication partielle et la puissance d'explication

Proposition2

Ayant un modèle causal M=(U,V,F),X⊆V, x∈D(X), un événement φ et un ensemble de contexte K.

Si X=x est une explication de φ relativement à K'⊆K, alors $K_{X\leftarrow x,\varphi}$ est l'ensemble de tous les contextes de K, tel que l'une ou l'autre de deux conditions est vraie:

i. X(u)≠x ou bien
ii. X(u)=x et X=x est une cause faible de φ sous le contexte u.

La reconnaissance de α-partielle explication est $P_{\|\|}^{\Sigma_2^p}$ -complet, cette reconnaissance nécessite

La reconnaissance de $K_{X=x,\varphi}$. En exploitant la proposition précédente, le calcule peut se faire par des appels parallèles de l'oracle Σ^p_2. Une fois l'ensemble $K_{X\leftarrow x,\varphi}$ calculé, le reste du travail est la vérification que X=x est une explication par rapport à cet ensemble. La complexité est liée au problème de calcule du $K_{X\leftarrow x,\varphi}$.

Théorème 3 : α-partielle explication est $P_{\|\|}^{\Sigma_2^p}$ -complet.

X=x est une α-partielle explication si :

a) X=x est explication de φ par rapport à $K_{X=x,\varphi}$.
b) $P(K_{X=x,\varphi}. \mid X=x)\geq\alpha$.

Le calcul de $K_{X=x,\varphi}$. est $FP_{\|\|}^{\Sigma_2^p}$ -complet. Une fois $K_{X=x,\varphi}$.calculé, la vérification de (a) se fait avec deux appels parallèles de l'oracle Σ^p_2, la vérification de (b) est fait en un temps polynomial. Il à été montré dans [Bus91] que deux appels parallèles à l'oracle Σ^p_2, dans un calcule en temps polynomial peut être remplacé par un seul. Le problème alors sera dans la classe $P_{\|\|}^{\Sigma_2^p}$.

La difficulté est vue comme une réduction au problème de décision suivant : ayant k QBFs $\phi_i=\exists A_i \forall B_i Y_i$ avec i∈{1,…,k}. Tel que chaque Y_i est une formule propositionnelle sur les variables $A_i=\{A_{i,1},\ldots,A_{i,mi}\}$et $B_i=\{B_{i,1},\ldots,B_{i,ni}\}$, décider du nombre de formules valide parmi les formules $\phi_1,..,\phi_k$.

Théorème 4: L'existence de α-partielle explication est Σ^p_3 complet.

Selon le théorème précédent, décider si X′=x′ est une α-partielle explication de φ relativement à K est $P^{\Sigma^p_2}_{\|}$.

Rechercher X′⊆X et x′∈D(x), et décider si X′=x′ est une α-partielle explication de φ relativement à K est Σ^p_3.
La difficulté Σ^p_3 est réduite à une réduction de l'existence d'explication.

Théorème 5: Partielle explication est $P^{\Sigma^p_2}_{\|}$ -complet.

X=x est une partielle explication de φ relativement K ssi :

a) X=x est une explication de φ relativement à $K_{x=x,\varphi}$.
b) $K_{x=x,\varphi}$. contient quelques contextes tel que X(u)=x et P(u)>0.

Le calcule de $K_{x=x,\varphi}$ est un problème $FP^{\Sigma^p_2}_{\|}$, la vérification de (a) est D^p_2, et la vérification de (b) se fait en temps polynomial. Donc le problème de décider que X=x est une partielle explication relativement à K est $P^{\Sigma^p_2}_{\|}$.

Théorème 6: La puissance d'explication partielle est $FP^{\Sigma^p_2}_{\|}$ complète.

Soit X=x une partielle explication de φ relativement à K. Pour le calcule de la puissance d'explication, on calcule $K_{X\leftarrow x,\varphi}$. et ensuite calculer $P(K_{X\leftarrow x,\varphi} \mid X=x)$. Le dernier point est $FP^{\Sigma^p_2}_{\|}$ alors que le dernier point ce fait en un temps polynomial. En résumé le problème est $FP^{\Sigma^p_2}_{\|}$.

La difficulté pour $FP^{\Sigma^p_2}_{\|}$, est vue par la réduction au problème de calcule, ayant k QBFs $\phi_i=\exists A_i \forall B_i Y_i$ avec $i\in\{1,\ldots,k\}$, tel que chaque Y_i est une formule propositionnelle sur les variables $A_i=\{A_{i,1},\ldots,A_{i,mi}\}$ et $B_i=\{B_{i,1},\ldots,B_{i,ni}\}$, décider du nombre de formules valide parmi les formules $\phi_1,..,\phi_k$.
Pour plus de détails sur les démonstrations des calculs proposées par Eaiter et Lukasiewiczs voir [Eiter04]. Les travaux de Eiter et Lukasiewicz sur la recherche d'algorithmes efficaces pour la recherche d'explication partielle dans le cadre de l'approche de Halpern et Pearl ne sont pas encore achevés, leurs recherches sont en cours.

VI. Conclusion

J. Y. Halpern et J. Pearl ont proposé une définition formelle de l'explication en terme de causalité. Comme cela a déjà été mentionné, il n'y a pas plusieurs définitions formelles de l'explication en terme de causalité dans la littérature. Excepté peut être l'approche de Lewis [Lewis86], qui défend l'idée que «pour expliquer un événement il faut fournir quelques informations sur son histoire causale». Ce point de vue est compatible avec la définition de Halpern et Pearl, cependant il n'y a aucune définition formelle donnée pour permettre une comparaison entre les approches.

Un problème a été signalé dans la définition : Traiter l'explication disjonctive. La disjonction cause des problèmes dans la définition de la causalité, c'est pour cela que ce point n'est pas traité dans l'explication. On a vu qu'il était possible de modifier la définition de la causalité pour être capable de traiter la disjonction sans être obliger de changer la structure de la définition de l'explication. De plus, la définition proposée ne donne pas les outils pour traiter la tension entre la puissance explicative, la qualité des croyances partielles et de la probabilité de l'explication. C'est une partie qui mérite d'être approfondie.

La formalisation de l'explication de Halpern et Pearl [Halpern05] utilise la notion de probabilité. On sait que les probabilités font que cette modélisation est plutôt quantitative ce qui s'éloigne du raisonnement humain, qui est plutôt qualitatif. Nous proposons dans le prochain chapitre, une approche qualitative de l'explication dans le cadre représentationel fourni par la logique possibiliste.

I. Introduction

Le raisonnement qualitatif constitue un sujet classique de l'intelligence artificielle (IA). La raison de ceci est que les activités cognitives des êtres humains sont au niveau qualitatif. Les gens déduisent le monde à partir d'un certain nombre de modèles qualitatifs plutôt qu'en traitant des modèles quantitatifs. Le raisonnement qualitatif est un domaine d'activité de l'IA qui a pour but de modéliser, de prévoir et d'expliquer le comportement des systèmes physiques en termes qualitatifs [Yan05].

Dans le présent chapitre nous présentons notre approche sur l'explication. C'est une modélisation possibiliste (qualitative) de l'explication partielle et de la puissance d'une explication. Mais avons de présenter cette approche nous avons jugé nécessaire de faire un bref rappel sur la logique possibiliste, modélisation qualitative, ainsi que des notions de base de cette logique (mesure de nécessité, de possibilité, base de connaissance possibiliste) qui sera suivit par une comparaison entre la mesure probabilité et la mesure possibilité.

II. Logique possibiliste

La logique possibiliste est une logique de l'incertain, issue de la théorie des possibilités de Zadeh [Zadeh78], capable de formaliser le raisonnement non-monotone. Dans cette approche, les règles avec exceptions sont traduites sous forme de contraintes sur des mesures de possibilité, ce qui permet de calculer (en terme de mesures de nécessité) des niveaux de priorité pour les règles, et de coder le résultat sous forme de bases de formules de la logique possibiliste [Benferhat98]

La logique possibiliste a été proposée par Dubois, Lang et Prade [Dubois88][Dubois93] c'est une logique permettant de manipuler des connaissances incertaines sur le monde. Au niveau syntaxique, la logique possibiliste manipule des paires de la forme (p,α) avec p une formule de la logique classique et α un élément d'un ensemble totalement ordonné. La paire(p,α) exprime que la formule p est certaine au moins au niveau α, ou, plus formellement que $N(p)\geq\alpha$, où N est la mesure de nécessité associée à la distribution de possibilité exprimant la sémantique sous-jacente. Pour simplifier, nous ne considérerons que des formules propositionnelles (mais l'extension aux formules du premier ordre ne pose pas de problème particulier [Dupin96]).

La logique possibiliste est essentiellement de nature qualitative contrairement à la logique probabiliste, en ce sens que la seule propriété requise est la structure d'ordre total de l'échelle utilisée; les valeurs précises n'ont donc pas d'importance réelle ; seul compte le pré-ordre qu'elles induisent sur les formules ("p est plus certaine que q" correspondant à $N(p)>N(q)$). La logique possibiliste est une logique calculatoirement traitable puisque sa complexité de calcul est de l'ordre de logn*SAT où n est le nombre de niveaux de certitude utilisés dans la base de connaissances [Benferhat98].

Soit £ un langage propositionnel fini et Ω l'ensemble des interprétations associé à £. Soient φ, ψ,…des formules propositionnelles.
On note par $\omega \models \varphi$ pour dire ω satisfait φ ou bien φ est un modèle de ω (Les interprétations où φ est vraie).

L'élément de base de la théorie possibiliste est la distribution de possibilité π qui correspond à une application de l'ensemble des interprétations Ω vers l'intervalle [0,1]. $\pi(\omega)$ est le degré de compatibilité de l'interprétation ω avec les informations (croyances) du monde réel. Par convention [Benferhat96][Benferhat98][Benferhat02b] [Benferhat02a] [Benferhat04]:

- $\pi(\omega) = 0$ signifie que ω est impossible,
- $\pi(\omega) = 1$ signifie que ω peut être le monde réel (ω est totalement possible),
- $\pi(\omega) > \pi(\omega')$ signifie que ω est une candidate préférée à ω' pour être l'état du monde réel.

D. Dubois et H. Prade soulignent le fait que la théorie des Probabilités ne permette pas de représenter la différence entre deux états de connaissance se distinguant uniquement par une différence de confiance dans les informations disponibles, ainsi ils introduisent l'outil des *mesures de possibilité* comme outil de représentation de l'incertain plus général permettant de représenter l'ignorance et de prendre en compte la *pertinence* d'une information incertaine.

Une distribution de possibilité π est dite normalisée s'il existe une interprétation ω tel que $\pi(\omega) = 1$[Benferhat02b][Benferhat02a]. Étant donné une distribution de possibilité π, nous pouvons définir deux manières différentes d'ordonner les formules. Cela est obtenu en définissant deux applications évaluant respectivement la possibilité et la certitude d'une formule [Benferhat96] [Benferhat98] [Benferhat02b] [Benferhat02a]:

- La mesure de possibilité d'une formule φ :
  $\Pi(\varphi) = \max \{\pi(\omega) : \omega \models \varphi\}$
  $\Pi(\varphi)$ évalue le degré de compatibilité de φ avec les croyances codées par π.Le degré de possibilité satisfait : $\forall p \, \forall q \, \Pi(p \vee q) = \max(\Pi(p), \Pi(q))$[Dubois95] [Benferhat04].
- La mesure de nécessité d'une formule φ :
  $N(\varphi) = 1 - \Pi(\neg\varphi)$
  $= \min \{1 - \pi(\omega) : \omega \models \neg\varphi\}$[Benferhat04]
  $N(\varphi)$ représente le degré de certitude de φ à partir des croyances codées par π. Le degré de nécessité satisfait : $\forall p \, \forall q \, N(p \wedge q) = \min(N(p), N(q))$. Cet axiome exprime que pour être certain de $p \wedge q$ à un certain degré, il faut l'être (au moins) autant de p et q séparément [Dubois93][Dubois95] [Benferhat04].

II.1 Propriétés des mesures de nécessité et de possibilité :

Dans la théorie probabiliste, la quantité $P(\neg\varphi)$ peut être simplement déduite à partir de $P(\varphi)$ car $P(\neg\varphi) = 1 - P(\varphi)$. Donc, si φ n'est pas probable, alors $\neg\varphi$ est nécessairement probable. Par contre, l'expression « il est impossible que φ soit vraie », n'implique pas seulement l'expression « $\neg\varphi$ est possible » mais entraîne ainsi, une forte conclusion : « il est nécessaire que $\neg\varphi$ soit vrai » [Benferhat02a].
En outre, l'expression « il est possible que φ soit vraie » ne donne aucune information sur la possibilité ou l'impossibilité de φ.
Dans la théorie des possibilités, l'incertitude de l'occurrence φ est décrite par deux mesures duales [Dubois93] [Benferhat02a] : La mesure de possibilité $\Pi(\varphi)$ et la mesure de nécessité $N(\varphi) = 1 - \Pi(\neg\varphi)$.

Popriété1 :

Si une proposition n'est pas tout à fait possible, alors son contraire l'est. Soit Π une mesure de possibilité sur un espace mesurable fini de mondes possibles (Ω ,**A**) alors :

$\forall A \in \mathbf{A}, \max(\Pi(A), \Pi(\neg A)) = \Pi(\Omega) = 1$ …(1)

Propriété2:

Si une proposition est nécessaire, alors son contraire ne l'est pas. Soit N une mesure de nécessité sur un espace mesurable (Ω, **A**) alors :

$\forall$ A $\in$ **A**, min(N(A), N($\neg$A)) = N($\emptyset$) = 0 …(2)

Du fait de ces deux propriétés, on peut également voir que pour tout A $\in$ **A**, il y a deux éventualités pour chacune des propriétés 1 et 2 ci- dessus :

Soit Π(A) < 1 et dans ce cas (1) implique que $\Pi(\neg A) = 1$ et donc par définition que N(A) = 0;

Soit $\Pi(\neg A) < 1$ (c'est-à-dire aussi N(A) > 0) et dans ce cas (1) implique que Π(A) = 1;

Soit N(A) > 0 et dans ce cas (2) implique que $N(\neg A) = 0$ et donc par définition que Π(A) =1;

Soit $N(\neg A) > 0$ (c'est-à-dire aussi Π(A) < 1) et dans ce cas (2) implique que N(A) = 0.

On peut bien sûr regrouper les cas (1 et 4) d'une part et (2 et 3) d'autre part dans la propriété suivante:

Propriété3 : cohérence sémantique des degrés de nécessité-possibilité

Soit Π une mesure de possibilité sur un espace mesurable fini de mondes possibles et N sa mesure de nécessité duale, pour tout A $\in$ **A** :

1. $\Pi(A) < 1 \Rightarrow N(A) = 0$ (3)

2. $N(A) > 0 \Rightarrow \Pi(A) = 1$ (4)

En d'autres termes, tout événement A vérifie toujours au moins une des deux conditions N(A)=0 ou Π(A) =1 (soit il n'est *pas complètement possible* et alors il n'est pas du tout *nécessaire*, ou bien il est *au moins légèrement nécessaire* et dans ce cas il doit être *complètement possible*). Ceci nous amène à remarquer que pour tout A, si N(A)=0 alors $N(A) \leq \Pi(A)$ et si N(A)>0 alors comme Π(A) =1 on a encore $N(A) \leq \Pi(A)$. D'où la propriété:

Propriété4 :

Soit Π une mesure de possibilité sur un espace mesurable fini de mondes possibles (Ω,**A**) et N sa mesure de nécessité duale, alors pour tout A $\in$ **A** les degrés de croyance [N(A), Π(A)] vérifient l'inégalité :

$$N(A) \leq \Pi(A) \qquad (5)$$

On peut bien sûr avoir à la fois { N(A) =0 et Π(A) =1}, { N(A) = 1 et Π(A) =1}, ou bien { N(A) =0 et Π(A) =0} chacune de ces éventualités s'interprétant différemment .

Propriété5 :

Soit Π une mesure de possibilité sur un espace mesurable fini de mondes possibles (Ω,A) et N sa mesure de nécessité duale, alors pour tout A $\in$ **A** les degrés de croyance [Π(A), N(A)] ont toujours l'une des formes suivantes :

1. N(A) = 1 et Π(A) = 1: il est certain que A est vérifié,

2. $N(A) > 0$ et $\Pi(A) = 1$:il est possible que A soit vérifié et A est plus certainement vérifié que $\neg A$,
3. $N(A) = 0$ et $\Pi(A) = 1$: il est possible que A soit vérifié mais il est également possible que $\neg A$ le soit et aucun des deux n'est nécessaire (ignorance),
4. $N(A) = 0$ et $\Pi(A) < 1$: il est possible que A soit vérifié mais $\neg A$ est plus certainement vérifié que A,
5. $N(A) = 0$ et $\Pi(A) = 0$: il est certain que $\neg A$ est vérifié.

Nous récapitulons via les deux tableaux suivants les propriétés de la mesure les possibilités et de la mesure de nécessité[Fabiani69] [Benferhat98] [Benferhat02a]:

$\Pi(\varphi) = 1$ et $\Pi(\neg\varphi) = 0$	φ est certainement vraie
$\Pi(\varphi) = 1$ et $\Pi(\neg\varphi) \in [0, 1]$	φ est certaine à un certain degré
$\Pi(\varphi) = 1$ et $\Pi(\neg\varphi) = 1$	Ignorance totale (φ est inconnue)
$\Pi(\varphi) > \Pi(\psi)$	φ est préférée à ψ
$\text{Max}(\Pi(\varphi), \Pi(\neg\varphi)) = 1$	φ et $\neg\varphi$ ne peuvent pas être toutes les deux impossibles (la seule relation qui existe entre φ et $\neg\varphi$)
$\Pi(\varphi \vee \psi) = \max(\Pi(\varphi), \Pi(\psi))$	Axiome de disjonction
$\Pi(\varphi \wedge \psi) \leq \min(\Pi(\varphi), \Pi(\psi))$	Axiome de conjonction

La table 2: Résume des propriétés sur la mesure de possibilité

$N(\varphi) = 1$ et $N(\neg\varphi) = 0$	φ est certaine
$N(\varphi) \in [0, 1]$ et $N(\neg\varphi) = 0$	φ est sure à un certain degré
$N(\varphi) = 0$ et $N(\neg\varphi) = 0$	Ignorance totale
$\text{Min}(N(\varphi), N(\neg\varphi)) = 0$	La seule relation qui existe entre $N(\varphi)$ et $N(\neg\varphi)$
$N(\varphi \wedge \psi) = \min(N(\varphi), N(\psi))$	Axiome de conjonction

La table 3 : Résume des propriétés sur la mesure de nécessité

II.2 Base de connaissances possibilistes

Une base de connaissances possibilistes Σ est un ensemble fini de formules pondérées de la forme (φ_i, α_i), $\Sigma = \{(\varphi_i, \alpha_i), i=1, n\}$ où φ_i est une formule propositionnelle et $\alpha_i \in [0, 1]$ représente le degré de certitude minimal de φ_i. Chaque couple (φ_i, α_i) de la base de connaissances possibiliste peut être considérée comme une contrainte qui restreint l'ensemble des interprétations possibles. Les formules avec un degré zéro ne sont pas représentées explicitement dans la base de connaissance (seulement les croyances qui sont au moins un peu acceptées sont explicitement représentées). Plus la pondération est élevée, plus la formule est certaine[Garcia04]
Si une interprétation ω satisfait φ_i alors le degré de possibilité $\pi(\omega) = 1$ (ω est complètement compatible avec la croyance φ_i), sinon il est égal à $1-\alpha_i$ (plus φ_i est certaine, moins ω est possible)

En particulier, si $\alpha_i = 1$, alors toute interprétation falsifiant φ_i est complètement impossible [Benferhat96][Benferhat98][Benferhat02b][Benferhat02a]
Plus formellement, la distribution de possibilité associée à une formule élémentaire (φ_i, α_i) est définie par :

$$\forall \omega \in \Omega\ \pi_{(\varphi i \alpha i)}(\omega) = \begin{cases} 1 - \alpha_i \ \text{ si } \omega \not\models \varphi_i \\ 1 \quad \text{sinon} \end{cases}$$

Plus généralement, la distribution de possibilité associée à Σ est le résultat de la combinaison de l'ensemble des distributions de possibilité associées aux formules élémentaires (φ_i, α_i) de Σ, c'est à dire [Benferhat96][Benferhat98][Benferhat02b][Benferhat02a]:

$$\forall \omega \in \Omega\ \pi_{\Sigma}(\omega) = \square \{\pi_{(\varphi i \alpha i)}(\omega)\ (\varphi_i, \alpha_i) \in \Sigma\}$$

Où $\square$ correspond soit à l'opérateur minimum (logique possibiliste standard ou qualitative) soit à l'opérateur produit (*)(logique possibiliste quantitative).
Cette définition peut être écrite de façon équivalente :

- Cas de l'opérateur min

$$\pi_{\Sigma}(\omega) = \begin{cases} 1 \quad \text{si } \forall\ (\varphi_i, \alpha_i) \in \Sigma\ \ \omega \models \varphi_i \\ \text{Min } \{(1 - \alpha_i) : (\varphi_i, \alpha_i) \in \Sigma\ \ \omega \not\models \varphi_i\ \} \text{ sinon} \end{cases}$$

- Cas de l'opérateur produit (*)

$$\pi_{\Sigma}(\omega) = \begin{cases} 1 \quad \text{si } \forall\ (\varphi_i, \alpha_i) \in \Sigma\ \ \omega \models \varphi_i \\ * \{(1 - \alpha_i) : (\varphi_i, \alpha_i) \in \Sigma\ \ \omega \not\models \varphi_i\ \} \text{ sinon} \end{cases}$$

Dans le cadre qualitatif, une base possibiliste Σ peut être représentée par des partitions bien ordonnées $WOP(\Sigma) = S_1 \cup \ldots \cup S_n$ (WOP : Well Ordred Partition) tel que S_1 contient les formules les plus certaines dans Σ, S_n contient les moins certaines . Plus généralement, les formules de S_i sont plus certaines que celles de S_j tel que $i < j$. Pour chaque partition $S_1 \cup \ldots \cup S_n$ nous pouvons construire une base possibiliste en associant à chaque formule dans S_i un degré de certitude α_i, tel que $1 >= \alpha_1 >= \ldots >= \alpha_n >= 0$.

Définition 1
Soit Σ une base de connaissances possibiliste, et $\alpha \in [0, 1]$. On appelle α-cut (resp. strict α-cut) de Σ, et on note $\Sigma_{>=\alpha}$ (resp. $\Sigma_{>\alpha}$), l'ensemble de formules de Σ ayant un degré de certitude supérieur ou égale à (resp. strictement supérieur à) α.
La base de connaissances possibilistes Σ est dite *consistant* si son support classique, obtenu en oubliant les pondérations, est consistant. Nous dénotons par : $Inc(\Sigma)=\max\{\alpha_i : \Sigma_{\geq \alpha i}$ est inconsistant$\}$ le degré d'inconsistance de Σ. $Inc(\Sigma)=0$ signifie que $\Sigma_{\geq \alpha i}$ est consistant pour tout α_i

Une conséquence syntaxique possibiliste a été définie comme suit. Soit Σ une base de connaissances possibiliste et (p,α) une formule possibiliste, p est une conséquence de Σ au degré α dénoté par $\Sigma \vdash (p,\alpha)$ si par définition p est une conséquence de $\Sigma_{>\alpha}$. En d'autres termes, (p,α) est une conséquence possibiliste de Σ si et seulement si $Inc(\Sigma \cup \{(\neg p,1)\}) > Inc(\Sigma)$ et $Inc(\Sigma \cup \{(\neg p,1)\}) > \alpha$ [Garcia04].

Définition 2

Soit (φ, α) une formule dans ∑. Alors (φ, α) est sous-sommé par ∑ si $(\sum - \{(\varphi, \alpha)\})_{>=} \vdash \varphi$.

Lemme 1

Soit (φ, α) une formule est sous-sommé dans ∑. Alors ∑ et ∑' = ∑ - {(φ, α)} sont équivalentes.

Les règles d'inférence de base en logique possibiliste sont les suivantes [Prade03] :

- $(\neg\varphi \vee \psi, m)(\psi \vee \rho, n) \vdash (\psi \vee \rho, \min(m,n))$ **résolution**.
- $\forall n \leq m\ (\varphi,m) \vdash (\varphi,n)$ **affaiblissement de poids**.
- Si $\varphi \vdash_{CL} \psi$, alors $(\varphi,m) \vdash (\psi,,m)$ **affaiblissement logique.**
- $(\varphi,m)\ (\varphi,n)\ ) \vdash (\varphi,\max(m,n))$ **fusion de poids.**

II.3. Possibilité et probabilité

La théorie des possibilités fournit une méthode pour formaliser des incertitudes subjectives sur des événements, i.e., un moyen de dire dans quelle mesure la réalisation d'un événement est possible et dans quelle mesure on en est certain, sans toutefois avoir à sa disposition l'évaluation de la probabilité de cette réalisation, par exemple parce qu'on ne connaît pas d'événement analogue auquel se référer, ou parce que l'incertitude est la conséquence d'une absence de fiabilité des instruments d'observation ou d'un doute de l'observateur lui-même. Par exemple, considérons l'assertion "Charles viendra au cocktail chez Jeanne à Paris le lundi 6 mai", on sait difficilement à quoi se référer pour lui attribuer une probabilité: au fait que Charles n'assiste pas régulièrement aux cocktails auxquels il est invité, qu'il n'est pas toujours à Paris le lundi, qu'il est souvent en vacances au début du mois de mai, qu'il est périodiquement fâché contre Jeanne…[Bouchon95].
Cependant, celle-ci évalue subjectivement à quel point il est possible et à quel point il est certain que Charles soit présent, en se basant globalement sur ce qu'elle sait de lui.
Les mesures de possibilité et de nécessité constituent des moyens de représentation d'incertitudes sur l'occurrence d'événements, comme les mesures de probabilité. Leurs propriétés en sont différentes, mais on peut établir un parallèle lien entre elles. De même, les propriétés des distributions de possibilités peuvent être comparées à celles des distributions de probabilité.
La mesure de possibilité d'un événement A est égale à la plus grande des valeurs de la distribution de possibilité pour tous les éléments de A, alors que la probabilité de A est la somme des valeurs de la distribution de probabilité pour tous les événements de A.
La probabilité de l'union de deux événements est égale à la somme de leurs probabilités, à la condition qu'ils soient disjoints. Dans le cas d'une mesure de possibilité, le coefficient attribué à l'union de deux événements est la plus grande des deux valeurs attribuées à chacun, sans condition restrictive [Bouchon95].

En ce qui concerne les mesures de nécessité, le coefficient affecté à l'intersection de deux événements est la plus petite des deux valeurs affectées à chacun, sans condition particulière sur les événements, alors que la probabilité de l'intersection de deux événements n'est égale au produit de leurs probabilités que si les deux événements sont indépendants.
Si l'on considère le complémentaire A^c d'un sous-ensemble A quelconque A, sa probabilité est déduite de celle de A lui même. La possibilité et la nécessité qui lui sont associées ne sont pas données directement à partir de celles de A, mais des bornes, inférieure dans le cas de possibilité, supérieure dans le cas de nécessité, en sont déduites de la possibilité de A et de la nécessité de A:

$$\Pi(A^c) \geq 1-\Pi(A) \text{ et } N(A^c) \leq 1-N(A)$$

Dans le cas où $\Pi(A)=1$, la valeur de $\Pi(A^c)$ est quelconque dans l'intervalle [0,1]. De même, si $N(A) =0$, $N(A^c)$ prend une valeur quelconque. On remarque donc une grande liberté dans le choix des valeurs de la possibilité et de la nécessité du complémentaire d'un sous-ensemble de A.
On observe que les produits de la théorie des probabilités deviennent des recherches de minimum, et que les sommes deviennent des recherches de maximum dans la théorie des possibilités. Les valeurs des mesures et de distributions sont choisies avec plus de liberté, dans le cas des possibilités que dans le cas des probabilités. Ceci revient du fait que la théorie des probabilités correspond à une meilleure connaissance des événements possibles que la théorie des possibilités.

On affecte une probabilité à un événement dans le cas où on dispose des connaissances suffisantes. Si ces connaissances sont insuffisantes, on peut se contenir d'attribuer des coefficients de possibilité et de nécessité à cet événement, qui sont plus souples d'utilisation, mais qui contiennent chacun moins d'information que n'en contient la probabilité.

La capacité à représenter de façon absolue l'ignorance totale est typique de la théorie des possibilités. Nous ne pouvons pas exprimer l'ignorance totale avec une distribution de probabilité unique [Benferhat04]. Par exemple, l'ignorance sur la vie extra-terrestre s'exprimera par P(vie) = P(non-vie) = ½. Cette représentation présupposera qu'il n'y a qu'une alternative. En revanche, la théorie des possibilités n'a aucune difficulté pour exprimer l'ignorance indépendamment du comptage des événements (en mettant à 1 tous les degrés de probabilités). L'équiprobabilité exprime plutôt le hasard, l'égalité des chances, l'incertitude devant le choix d'une alternative, et non pas l'ignorance [Benferhat04].

Notons que, dans le cadre possibiliste, les deux degrés N(A) et Π(A) sont utile pour caractériser ce que nous savons de l'événement A et de son contraire, alors qu'en théorie des probabilités un seul degré suffit (P(A), car P(A) = 1 – P(¬A)). En conséquence, la théorie des probabilités ne permet pas de distinguer entre possibilité et certitude [Benferhat04].

III. Explication possibiliste :

La définition de la notion de l'explication dans [Halpern05] se base sur la notion de causalité en terme contre-factuelle; L'idée principale est qu'une explication n'est pas connue pour certains mais si elle est à vraie elle constitue une cause réelle ce que nous désirons expliquer sans tenir en compte de l'incertitude initiale de l'agent. Une explication est relative à l'état épistémologique de l'agent; l'état épistémologique est représenté par l'ensemble des interprétations que l'agent considère possibles. On sait que les probabilités font que cette modélisation est plutôt quantitative ce qui s'éloigne du raisonnement humain, qui est plutôt qualitatif.

Nous proposons une modélisation possibiliste de la notion de l'explication partielle, ces explications seront classées par rapport à leurs possibilités de réalisation, ce qui nous offre un choix sur les explications disponibles. Nous proposons une réorganisation stratifiée des explications fournies, selon l'ordre de possibilités des explications [Boutouhami05]. La logique possibiliste possède une inférence syntaxique, correcte et complète par rapport à une inférence sémantique basée sur des distributions possibilistes. De plus sa complexité est très proche de la logique classique ce qui a motivé notre choix [Benferhat02a].

Définition 3.1

Soit π une distribution de possibilité, qui correspond à une application de l'ensemble des interprétations Ω considérées possibles par l'agent (les contextes possibles)vers l'intervalle [0, 1].
Ayant le modèle structurel M, X=x est une explication de φ relativement à l'ensemble Ω des interprétations si les conditions suivantes sont vraies:

- EX1 : (M, ω) $\models\varphi$. Pour toute interprétation $\omega\in\Omega$ (φ doit être vraie dans toutes les interprétations que l'agent considère possible.
- Ex2 : $\overrightarrow{X} = \vec{x}$ est une cause suffisante de φ dans (M, ω) pour chaque $\omega\in\Omega$ tel que (M, ω) $\models \overrightarrow{X} = \vec{x}$.
- Ex3 : $\overrightarrow{X}$ est minimal, pas de sous-ensemble de $\overrightarrow{X}$ satisfait la condition Ex2.
- Ex4 : (M, ω) $\models\neg(\overrightarrow{X} = \vec{x})$ quel que soit $\omega\in\Omega$et (M, ω') $\models \overrightarrow{X} = \vec{x}$ quel que soit $\omega'\in\Omega$.(ceci signifie simplement qu'un agent considère les contextes où une explication n'est pas connue au début, et considère des contextes possibles où une explication est vraie).

II1.2 Explication partielle et la puissance d'une explication

Comme déjà mentionné les explications ne sont pas toutes de même pertinences certaines explications sont plus appropriées que d'autres. Nous proposons d'introduire des possibilités dans la définition précédente afin de créer une sorte de classification par rapport aux explications disponibles.

Reprenons l'exemple des pluies d'avril (exemple 3.1 du chapitre 5), Pour modéliser cette situation nous avons les variables suivantes :

- AS : pour décrire la pluie d'Avril (0 s'il n'a pas plu, 1 sinon) ;
- ES : pour décrire les deux tempêtes avec éclairs en Mai et Juin avec, (0,0) pas d'éclair ; (1,0) éclair en Mai et pas en Juin ; (0,1) pas d'éclair en Mai, mais en Juin ; (1,1) des éclairs en Mai et en Juin ;
- F : pour décrire le feu de la forêt avec 0 s'il n'y a pas de feu ; 1 s'il y a du feu en Mai ; 2 s'il y a du feu en Juin.

Supposons que l'agent considère qu'il y est souvent des tempêtes avec éclair dans les deux mois de mai et juin, alors la possibilité de l'ensemble des interprétations où ES=(11,) va être plus grand que l'ensemble des interprétations où ES=(1,0) ou ES=(0,1). Donc ES=(1,1) sera considérée comme une bonne explication.

Soit π une distribution de possibilité, qui correspond à une application de l'ensemble des interprétations Ω considérées possibles par l'agent (les contextes possibles)vers l'intervalle [0, 1] (les interprétations où X=x est une explication de φ).

La mesure de l'explication X=x est donnée par la mesure de possibilité $\Pi(X=x)$. Formellement $\Pi(X=x)= \max \{\pi(\omega) : \omega \models X=x \}$.

Définition 3.3

X=x est une explication partielle de φ par rapport à Ω ssi X=x est une explication de φ par rapport à un sous ensemble Ω' de Ω des interprétations possibles

Soit $\Omega_{X=x,\varphi}$ le plus grand sous-ensemble Ω' de Ω, tel que X=x est une explication de φ relativement à Ω', le plus grand sous-ensemble correspondant à toutes les interprétations moins celles où X=x et φ sont vraies alors que X=x n'est pas une cause suffisante de φ);

$$\Omega'=\Omega-\{\omega: \omega\in\Omega \mid \omega \models X=x , \omega \models \varphi \text{ et } \ll X=x \text{ n'est pas une cause suffisante de } \varphi \}.$$

- X=x est une explication partielle de φ avec une Qualité $\Pi (\Omega_{X=x,\varphi}|X=x)$. Formellement $\Pi (\Omega_{X=x,\varphi}|X=x) = \max \{\pi(\omega) : \omega \models X=x \}$(la mesure de possibilité de X=x, par rapport à toutes les interprétations dans l'ensemble Ω' satisfaisant X=x et φ).
- X=x est une α-explication partielle de φ relativement à π et Ω ssi Ω' existe et $\Pi (\Omega_{X=x,\varphi}|X=x)\geq\alpha$ (la mesure de possibilité de X=x, par rapport à toutes les interprétations dans l'ensemble Ω' soit supérieur à α satisfaisant X=x et φ) .
- X=x est une explication partielle de φ relativement à π et Ω ssi X=x est une α-explication partielle de φ relativement à π et Ω pour $\alpha\geq 0$.

La notion de α-explication partielle permet d'établir une notion de priorité dans le choix des explications en classifiant les explications partielles par rapport à la valeur du facteur α.
Soient "X=x" et "Y=y"deux explications partielles de φ par rapport à Ω, "X=x" est une explication partielle candidate préférée que "Y=y" ssi $\alpha>\alpha'$ tel que "X=x est une α-explication partielle de φ" et "Y=y est une α'-explication partielle de φ"
Les explications vont être subdivisées en un ensemble de strates ordonnées selon leur indice qui représentes le degrés de puissance des explications partielles contenues (S_1,..S_i,S_j,.., S_n) telle que :

- La strate S_1 contient les explications complètes si elles sont disponibles.
- X=x est une explication dans la strate S_α, si $\Pi(\Omega_{X=x,\varphi} \mid X=x)=\alpha$;

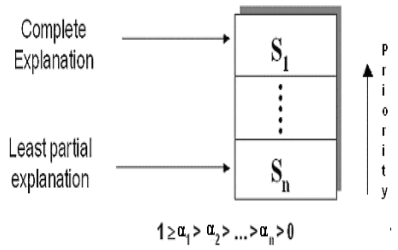


Figure 1 : Stratification des explications

Pour l'explication d'un événement φ, nous aurons un ensemble $S_1\cup\ldots\cup S_n$ contenant toutes les explications partielles de φ.
La strate S_1 contiendra l'explication compète de φ si elles sont disponibles.

Reprenons l'exemple du feu de la forêt, supposons que deux pyromanes jettent deux allumettes dans deux endroits différents d'une forêt sèche et les deux causent le fait que les arbres commencent à brûler. Considérons deux scénarios. Dans le premier cas disjonctif, chacune des allumettes suffit à elle seule à brûler l'ensemble de la forêt. Dans le second cas conjonctif, les deux allumettes sont nécessaires pour brûler la forêt. Si l'une d'elle seulement est allumée, l'incendie s'arrête avant que la forêt brûle.

On modélise le scénario par un modèle structurel contenant quatre variables :

- Une variable exogène U qui détermine, parmi d'autres choses, les motivations et l'état mentale des pyromanes, on suppose que $R(U)=\{U_{00},U_{10},U_{01},U_{11}\}$; Tel que si $U=U_{ij}$, alors le premier pyromane à l'intention de mettre du feu si i=1 et le deuxième à l'intention de mettre du feu si j=1, dans notre scénario $U=U_{11}$.
- Des variables endogènes ML_1 et ML_2, où $ML_i=0$ si le pyromane i n'as pas allumé l'allumette et MLi=1 s'il l'a allumé, i :=1,2 ;

Une variable endogène FB pour la forêt a brûlée, avec FB=0 si la forêt n'a pas brûlée et FB=1 sinon.
Supposons qu'il y a une variable endogène O correspondant à la présence de l'oxygène. Si on suppose qu'il y a une interprétation où O=0. Dans ce cas O=1 devient une explication partielle du feu, mais c'est une explication partielle d'une puissance explicative moins que celle de $ML_1=1$. Avec cette notion de stratification 'O=o' sera dans une strate inférieure à la strate contenant l'explication partielle '$ML_1=1$'.

III.4 Puissance d'une explication

Supposons qu'il existe une distribution de possibilité π^- sur l'ensemble Ω^- des interprétations incluant Ω (l'ensemble des interprétations considérées possibles par l'agent).
La distribution de possibilité π est le résultat de conditionner π^- par φ (c'est une distribution de possibilité avant que φ soit connu) $\pi=\pi^-(.|\varphi)$.
La définition usuelle du conditionnement se fait comme suit :

- $$\pi(\omega \mid \varphi)=\begin{cases} 1 \text{ if } \Pi(\varphi)=\pi(\omega) \text{ et } \omega \models \varphi, \\ \pi^-(\varphi) \text{ if } \pi^-(\varphi) < \Pi(\varphi) \text{ et } \omega \models \varphi, \\ 0 \text{ else.} \end{cases}$$

L'idée de cette définition est de voir φ comme une nouvelle connaissance complètement certaine. Le conditionnement correspond alors à la révision des degrés de possibilités associés aux différentes interprétations après la prise en compte de la connaissanceφ. Si une interprétation falsifie φ alors cette interprétation devient impossible (car l'information φ est complètement possible [Benferhat03].
La puissance d'une explication est définie comme suit :

$$\Pi^-(X=x)=\max\{\pi^-(\omega) : \omega \models X=x\}, \text{ pour les interprétations dans } \Omega^-.$$

La mesure de possibilité de X=x par rapport à la distribution de possibilité avant la découvert de φ.

Reprenons l'exemple du bronzage de Victoria, si l'agent considère qu'il faiblement possible qu'il ne fait pas beau aux Ils canaris, ce qui fait que "Victoria part au Ils canaris" est une explication partielle fortement possible de son bronzage. Néanmoins le fait " Victoria a visité un salon de bronzage" est une explication partielle moins possible.

V. Complexité

La définition de l'explication partielle a été motivée par le fait que souvent on n'arrive pas à trouver une explication complète, alors qu'on arrive à trouver des explications qui satisfont presque toutes les interprétations d'un agent mais pas toutes. Pour un événement φ, nous proposons de calculer un ensemble de strates, contenant les explications partielles afin de donner plus d'information à l'agent. Le calcul de l'ensemble des explications partielles d'un événement φ, se fait en se basant sur la définition de l'existence de l'explication partielle et du calcul de la puissance d'explication.

Proposition 2
On suppose que les mesures de possibilités sont calculées en un temps polynomial.

Le problème de la complexité pour l'explication possibiliste est pareil au problème de complexité calculée par Eiter et Lukasiewicz[Eiter04], puisque nous partons du même principe. En utilisant le théorème de la complexité de la cause faible qui est un problème Σ^p_2 [Eiter04]. En se basant sur la complexité calculée par Eiter et Lukasiewicz et de façon analogue à leur complexité nous aurons le tableau suivant :

Problème	Cas binaire
α-partielle explication	$P^{NP}_{\parallel}$-complet
L'existence de α-partielle explication	Σ^p_2-complet
Partielle explication	$P^{NP}_{\parallel}$-complet
Puissance d'explication	$FP^{NP}_{\parallel}$-complet

Tableau 4 : complexité de l'explication possibiliste.

V.1 Complexité du calcul des strates.

Le calcul de l'ensemble des strates possibles contenants toutes les explications partielles pour un événement donné φ, se base sur la définition de l'existence d'explication partielle et la puissance d'explication., est un problème Π^p_3

Le problème de l'existence d'explication partielle est le suivant :
Ayant un modèle causal M=(U,V,F), X⊆V, un événement φ et un ensemble d'interprétation Ω. Une distribution de possibilité π sur Ω. Décider si X′⊆X et x′∈D(X) existe tel que X′=x′ est une α-partielle explication de φ relativement à Ω est un problème Σ^p_2.

Le problème de décider que X′=x′ est une explication partielle de φ relativement à Ω est un problème $P^{NP}_{\|}$. Alors que le problème de recherche de X′⊆X et x′∈D(x), et décider que X′=x′ est une partielle explication de φ relativement à Ω est un problème $FP^{NP}_{\|}$, le problème de décider que X′=x′ est une partielle explication relativement à Ω est $P^{NP}_{\|}$.

En se basant sur cette définition, le calcul des strates à l'encontre du principe de l'existence des explications partielles, calculs toutes les explications partielles, ce qui fait que l'algorithme de recherche ne s'arrête pas avec la rencontre d'une réponse oui mais en dénombrant toutes les instances de oui. Le problème n'est pas de décider X′ =x′, mais de rechercher toutes les X′⊆X tel qu'il existe x′∈D(x) et X′=x′ est une explication partielle de φ relativement à Ω.

L'ensemble des strates $S=S_1\cup\ldots\cup S_{\alpha i}\cup S_{\alpha j}\cup\ldots.\cup S_{\alpha n}$, tel que Si contiens toutes les explications partielles X″=x″ tel que $\prod(\Omega_{X''=x'',\varphi}|X''=x'')=\alpha i$, et $\alpha i>\alpha j$. X″⊆X.

Le calcul des strates ne se base pas sur l'indice des strates puisque le nombre de valeurs comprises entre 0 et 1 est infini, mais plutôt sur l'exploitation de tous les sous-ensembles de X et de vérifier si ces sous-ensembles répondent bien à la définition de l'explication partielles. Chaque fois qu'une explication partielle est trouvée, l'ajouter dans la strate correspondante si elle existe déjà, sinon créer une nouvelle strate et l'ajouter.

Ayant un modèle causal M=(U,V,F), X⊆V, un événement φ et un ensemble d'interprétation Ω. Une distribution de possibilité π sur Ω. Calculer tous les sous-ensembles X′⊆X tel que x′∈D(X) existe et que X′=x′ est une α-partielle explication de φ relativement à Ω. Pour chaque X′ l'insérer dans la strate correspondante sinon créer une nouvelle strate.

Algorithme de génération de strates

Debut

Entrée {S=∅,V,φ,Ω}

X=V-{Yj}/ ∀ Yj, Yj est une variable de φ.

Pour tous X⊆′X faire

Debut

- Décider s'il existe x∈′D(X), et que X′=x′ est une explication partielle de φ relativement à Ω.
- Si (X′=x′ est une explication partielle dont $\prod(\Omega X'=x',\varphi|X'=x')=\alpha i$)

Alors

Debut

Rechercher la strate Sαi ;

Si la strate $S_{\alpha i}$ existe

Alors $S_{\alpha i}= S_{\alpha i}\cup\{X'=x'\}$

Sinon

Debut

Créer une nouvelle strate $S_{\alpha i}$

Insérer la strate selon l'ordre.

$S=S\cup S_{\alpha i}$.

Rajouter {X′=x′} dans la strate $S_{\alpha i}$.

Fin

Fin

Fin.

Sortie {$S=\cup S_{\alpha i}$}.

Fin.

La condition (b) est vérifiée en un temps polynomial, alors que la condition (a) est celle de l'existence d'une explication partielle qui se base sur réduit au problème de l'existence d'explication qui est un problème $\sum_2^p$ -complet (qui se réduit au problème de décision de validité d'une formule propositionnelle QFB ϕ= $\exists B \forall C \exists D$. L'idée de base est dans le quantificateur $\exists$ (existe-t-il $X' \subseteq X$), $\forall C \exists D$ pour la vérification de la condition Ex3.

Le calcul des strates est un problème Π_3^p. Ceci est du au fait que le problème se réduit au problème de calcul de nombre d'instances oui, pour un problème de décision de validité d'une formule. $\forall A \exists B \forall C \forall D$, vérifier si pour tout sous-ensemble de X′ de X ($\forall$A), vérifier s'il existe $x' \in D(x)$($\exists$ $x' \in D(x)$/ D(x)={0,1}), tel X′=x′ est une explication partielle de φ relativement à Ω.

VI. Conclusion

La logique possibiliste est une logique calculatoirement traitable puisque sa complexité de calcul est de l'ordre de logn*SAT où n est le nombre de niveaux de certitude utilisées dans la base de connaissances [Benferhat98]. La logique possibiliste est essentiellement de nature qualitative (contrairement à la logique probabiliste). Dans la théorie des possibilités, l'incertitude de l'occurrence φ est décrite par deux mesures duales : La mesure de possibilité $\prod(\varphi)$ et la mesure de nécessité $N(\varphi) = 1 - \prod(\neg\varphi)$.

Nous avons proposé dans ce chapitre une modélisation possibiliste de la notion de l'explication partielle où les explications générées seront classifiées par rapport au facteur α, tel que si "X=x" et "Y=y" sont deux explications partielles de φ, alors "X=x" est plus appropriée que "Y=y" ssi "X=x" est une α-explications partielles" et "Y=y" est une α-explication partielle" telle que "$\alpha > \alpha'$ ". Une autre mesure a été définie permettant de mesurer la qualité d'une explication partielle c'est la mesure de puissance d'une explication.
La complexité de notre approche se détériore légèrement mais reste néanmoins raisonnable par rapport à l'explication de l'approche structurelle.

Conclusion et Perspectives

Le concept de la causalité est central dans notre raisonnement de tous les jours qui est intuitivement banal. Pourtant diverses disciplines s'y intéressent, certaines depuis des siècles. La logique a souvent été un formalisme adopté par les auteurs pour représenter les relations causales. Plusieurs approches ont été développées pour la modélisation de la causalité.

L'objectif de cette thèse a porté sur l'étude ces approches. Nous avons présenté dans une première étape des généralités sur cette notion ; la nécessité d'un raisonnement non-monotone nous a amené à présenter trois types de modélisations.

1. Une modélisation graphique, via les réseaux causaux bayésiens ;
2. Une modélisation agentive ; via l'approche normative de la causalité et
3. Une approche contre-factuelle par une modélisation via un modèle d'équation structurel.

Les deux dernières ont fait l'objet d'une étude comparative; en soulignant les points communs ainsi que les points de divergences

La génération automatique d'une explication adéquate s'avère une tâche essentielle dans la planification, le diagnostic et le langage naturel. Un système faisant l'inférence doit être capable d'expliquer ses résultats et ses recommandations pour pouvoir acquérir la confiance de ses utilisateurs [Dubois03]. Le raisonnement qualitatif constitue un sujet classique de l'IA. La raison de ceci est que les activités cognitives des êtres humains sont au niveau qualitatif. ;

Nous avons proposé une approche qualitative de l'explication dans le cadre représentationnel fourni par la logique possibiliste. Ceci par une modification de la définition de Halpern et Pearl pour l'explication en remplaçant les probabilités par des possibilités. Une réorganisation stratifiée des explications fournies a été donnée, selon l'ordre des possibilités des explications. Le calcule de l'ensemble des strates pour un événement donné est un problèmeΠ_3^p. La logique possibiliste possède une inférence correcte et complète par rapport à une inférence sémantique basée sur des distributions possibilistes [Benferhat02]. De plus sa complexité est très proche de la logique classique ce qui a motivé notre choix.

Comme perspectives nous envisageons :

- Adapter les explications en fonction de l'évolution des croyances de l'agent ;
- Utiliser ce travail dans des applications réelles, telle que le domaine médical.

[Achinstein83] P. Achinstein. "*The Nature of Explanation"*. Oxford : Oxford University Press. 1983.

[Allen84] J.Allen. "Towards a general theory of action and time". Revue Artificial Intelligence, 23:13-154, 1984.

[Baker89] A. B. Baker. "A simple solution to the yale shooting problème". In International Conférence on Knowledge Representation and Reasoning, pp 11-20, 1989.

[Balacheff90] N. Balacheff. "Problèmes de la production d'une explication : aspects conceptuels et langagiers". *Revue d'Intelligence Artificielle*, 4(2), 149-160. 1990.

[Balmisse02] G. Balmisse. " Gestion des connaissances, outils et applications du knowledge management". Editions Vuibert ; Collection Entreprendre Informatique. ISBN : 2-7117-8697-8. 2002

[Becker99] A. Becker, P. Naïm. "Les réseaux bayésiens -Modèles Graphiques de connaissances". Edition Eyrolles, 1999.

[Becker96] A. Becker, D. Geiger. "Optimisation or Pearl's Method of Conditioning and Greedy-Like Approximation Algorithms for the Vertex FeedBacck Set Problème" .Artificial Intelligence 83, p. 167-188, 1996.

[Benferhat04] S. Benferhat, F. Khelaf et A. Mokhtari. "Product-based causal networks and quantitative possibilistic bases". Proceedings of the 17th International FLAIRS Conference (FLAIRRS'2004), may 2004.

[Benferhat03] S. Benferhat, S. Lagrue, O. Papini. "A possibilistic handling of partially ordered information". UAI 2003, pp 29-36

[Benferhat02a] S. Benferhat, S. Lagrue, O. Papini. "Revising partially ordered beliefs". In Procceeding of the 9th International Workshop on Non-Monotonic Reasoning (NMR'2002), Toulouse, France, 19 avril 21 avril 2002, p. 142-149

[Benferhat02b] S. Benferhat, D. Dubois, L. Garcia et H. Prade. 'On the transformation between possibilistic logic bases an possibilistic networks. IJAR, 2002.

[Benferhat98] S. Benferhat, D. Dubois, L. Garcia et H. Prade. "Réseaux possibilistes orientés et logique possibilistes". Acte 11ème congrés Reconnaissance des formes et Intelligence Artificielle RFIA'98 Volume2, 20-21-22 janvier 1998.

[Benferhat96] S. Benferhat, D. Cayrac, D. Dubois et H. Prade. "Explanation away in a possibilistic setting". Proc. Of the 6th Inter. Conf. On Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'1996), Granada, Spain, July 1-5, 1996, 929-934.

[Bouchon95] B. Bouchon-Meunier: "la logique floue et ses applications". Préface de Lotfi Zadeh. Edition Addison Wesly France.octobre 1995.

[Boutouhami05] S. Boutouhami, A. Mokhtari, ''Possibilictic explanation''. ICTI'S 05. Tetouan Maroc. June 2005.

[Castilho99] M. A. Castilho, O. Gasquet, Z. Hezig. "Formalizing action and change in model logic I: the frame problem". In journal of logic and computation, 9(5): 701-735, 1999.

[Chajewska97] U. Chajewska and J.Y. et Halpern. "Defining explanation in probabilistic systems". In proc. Thirteen conference on Uncertainty in Artificial Intelligence (UAI'97). Pp. 62-71. 1997.

[Cooper92] G. Cooper, E. Hersovits. "A Bayesien Method for the Induction of Probabilictic Networks from Data". Machine learning, 9, p. 393-347, 1992.

[Cooper90] G. Cooper. "The computational Complexity of Probabilistic Inference with Dependence Trees". Revue Artificial Intelligence, 42, p.393-405,1990.

[Dubois03] D. Dubois et H. Prade. “Les liens causaux et explication. Problème de modélisation : une discussion préliminaire”.Journées Nationales sur les Modèles de Raisonnement (JNMR'03) Institut Henri Poincaré, Paris, les 27 & 28 novembre 2003.

[Dubois95] D. Dubois, D. Lang, H. Prade. “Possibilistic logic”. D. Gabbay, C. Hogger, J. Robinson, EDs., Handbook of Logic in Artificial Intelligence and Logic Programming, vol 3, P 439-513, Oxford University Press, 1995.

[Dubois93] D. Dubois et H. Prade. “Logique floue”. Observatoire français des techniques avancées 93. Edition Masson.

[Dubois88] D. Dubois, H. Prade. “Possibility Theory- An approach to Computerized Processing of Uncertainty”. Plenum Press, New-York, 1988.

[Dupin96] F. Dupin de Saint-Cyr. “Gestion des croyances et de l’évolutif en logique pondérées”. Thèse de Doctorat, Université Paul Sabatier, Toulouse, décembre 1996.

[Eiter04] T. Eiter, T. Lukasiewicz. “Complexity Results for Explanations in the Structural-Model Approach”. Revue Artificial Intelligence, Volume 142. pp145-198. 2004.

[Espinoza92] M. Espinoza. “Les quatres causes de bunge à aristote”. *Revue philosophique de la France et de l’étranger*, 3 :297-316, 1992.

[Fabiani96] P. Fabiani. “Representation dynamique de l'incertain et stratégie de prise d'information pour un système autonome en environnement évolutif”. Thèse de Doctorat. Ecole Nationale Supérieur de l'Aéronautique et de l'Espace. Spécialité : Automatique, Informatique Industrielle. 17 décembre 1996.

[Finger87] J.J. Finger. “Eploiting constraints in design synthesis”. PhD thesis. Stanford University, Stanford, 1987.

[Jenson96] P. Jensen, V. Finn. “An introduction to bayesian Networks”. UCL Press, 1996.

[Jonhson 90] D.S Jonhson, “A catalog of complexity classes”, in J van Leeuwen(ED), Handbook of Theoritical Computer Science, vol, A, Elsevier, Amsterdam, 1990, Chapter2.

[Hall03] N. Hall. “Two concepts of causation”. In J. Collins, N Hall, and L. A. Paul (Eds), “Causation and Counterfactuals”. Cambridge, Mass. : MIT Press.2003.

[Halpern05] J. Y Halpern and J. Pearl. “Causes and Explanations: Astructural-model; Approach. Part I : Causes. Part II: Explanation”. Supported in part by SNF under grat IRI-96-25901. to appear. British Journal fo Philosophie of Science. 2005

[Halpern02] J. Y Halpern and J. Pearl,“Causes and Explanations: Astructural-model; Approach. Part II: Explanation”.21 August 2002. Supported in part by SNF under grat IRI-96-25901.

[Halpern01] J. Y Halpern and J. Pearl. “Causes and Explanations: Astructural-model; Approach. Part I : Causes”. IJCAI 2001.

[Halpern00] J. Y Halpern. “Axiomatizing Causal Reasoning”. Journal of A.I. Research 12, 317-337; 2000.

[Hempel66] C.G. Hempel. *Philosophy of Natural Science.* Englewood Cliffs, New Jersey : Prentice-Hall, 1966.

[Hanks87] S. Hanks and D.V. McDermott. “Nonmonotoning logic and temporel projection”. Artificial Intelligence, volume 33, pp 379-412, 1987.

[Hitchcock99] C. Hitchcock. “The intransitivity of Causation Revealed in Equations and Graphs”. Technical report, California Institute of Technology, Pasadena, Calif. 1999.

[Hitchcock96] C. Hitchcock. “The role of contrast in causal and explanatory claims”. Synthese 107, 395-419; 1996.

[Hopkins01] M. Hopkins; “A proof of the conjunctive cause conjecture”. Unpublished manuscript. 2001.

[Hopkins02] M. Hopkins. “New Challenges in the Pursuit of Causal Explanations”. UCLA Cognitive Systems Laboratory. AAAI 2002.

[Hume1739] D. Hume. “A Treatise of Humain Nature”. London : John Noon.1739.

[Hume1758] D. Hume. “Enquête sur l’Entendement Humain”. Trad. A. Leroy. GF Flammarion n°343, 1983.

[Hume75] D. Hume. “An enquiries concerning human understanding and concerning the principles of morals”. (1748, 1751). Oxford U. P., 1975.

[Gabbay84] D.M. Gabbay. “Theoretical foundation for non_monotonic reasonning in expert systems”. Technical Report RR84/11, Departement of Computing, Imperial College, London, 1984.

[Garcia04] L. Garcia. P. Nicolas, I. Stéphan. “Programmation par ensemble-réponses possibilistes”. LERIA- Université d'Angers UFR Sciences, JFPLC 2004.

[Gärdefors88] P. Gärdenfors. “Knowledge in flux”. Cambridge, Mass : MIT Press. 1988.

[Geslin03] S. Geslin, M. Chambreuil. “Introduction aux réseaux bayésiens : Apprentissage”. 17 mai 2003

[Geffner92] H. Geffner. “Default reasonning: Causal and Conditional theories”, MIT Press, Cambridge, MA, 1992.

[Gondran95] M. Gondran. “Graphes et algorithmes”. Eyrolles, 1995.

[Good93] I. Good. “A Tentative Measure of Probabilistic Causation”. Relevant to the Philosophy of the Law Static. Comput and simulation 47 pp 99_105. 1993.

[Grice75] H. P, Grice. “Logic and conversation”. IP Cole and J.L. Morgan, Edotors, Speechacts. Academic Press, 1975.

[Kayser04] D. Kayser. F. Nouioua, "Representing Knowledge About Norms " The 16th European Conference on Artificial Intelligence (ECAI'04), Valencia, Spain (August 22-27, 2004) pp. 363-367

[Kayser 98] D Kayser, A. Mokhtari. “Time in a causal theory”, Annals of Mathematics and Artificial Intelligence 22 (1998).

[Khelfallah 01] M. Khelfallah, A. Mokhtari. “Ramification in the Normative Method of Causality”. ESCARUS, pp 704-713. 2001.

[Kistler04] M. Kistler “La causalité dans la philosiphie contemporaine”: Intellictia, 2004/1, Volume 38, pp 139-185.

[Kleene71] S.C, Kleene. “Logique Mathématique”. Colection U, 19971.

[Kripke63] S.A. Kripke. “Semantical consideration on modal logique”. Acta Philosophica Fennica, 16:83-94, 1963.

[Lauritzen88] S. L. Lauritzen, D. J. Spiegelhalter. “Local Computations with probabilities on Graphical Structures and Their Applications to Expert Systems”. J. Royal Statistical Society, 50, p.157-224, 1988.

[Lewis02] D. Lewis. “Causattion as influence”. In J. Collins, N. Hall, and L. A. Paul (Eds.), *Causation and counterfactuals.* Cambridge, Mass: MIT Press. “An abridge version appears in *Journal of philosiphy 97, 2000”; 2002.*

[Lewis86] D.Lewis. “Causal Explanation”. In Philosophical Papers, volume II, pp 214-240. New york: Oxford University Press; 1986.

[Lewis73] D. Lewis. “Causation”. *Journal of philosiphy,* 70:556-567, 1973.

[Liftschitz97] V. Liftschitz. “On the logic of causal explanation”. Artificial Intelligence 96. pp 451-465. (1997)

[Lin95] F. Lin. “Embarcing causality in causality in specifying the indeterminate effects of actions”. In proc Fourteenth International Joint conference on artificial Intelligence; IJAI 95. pp 1985-1991 ; 1995.

[McCarty86] J. McCarty. “Applications of circumption to formalizing commen-sense Knowledge”. Artificial Intelligence, 28, 1986.

[McCarty77] J. McCarty. “Epistemological problems of artificial intelligence”. In IJCAI, pp 1038-1044, 1977.

[McCarthy69] J. McCarthy and P.J Hayes. “Some philosophical problems in artificial intelligence”. In B. Meltzer and D. Michie, editors, *Machine intelligence, volume 4.* Edinmburgh University Press, 1969.

[McDermott95] M. McDermott. “Redundant causation”. British Journal for the Philosophy of Science 40, 523-544. 1995

[McDermott82a] D.V. McDermott. “Non monotonic logique 2: nonmonotonic modal théories”. J.ACM, Volume 29, 1982.

[McDermott82b] D.V. McDermott. “A temporal logic for reasoning about process and plans”.Cognitive Science, vol 6,pp. 101-155. 1982.

[McDermott 80] D.V. McDermott, J. Doyle. “Non monotonic logic 1”. Revue Artificial Intelligence, 13, 1980.

[Moinard88] Y. Moinard. “Contribution à l’étude de la circonscription”. Thèse d’Université. Université rennes. 1988. I

[Mokhtari97a] A. Mokhtari. “Aspects normatifs temporels et épistémiques pour une représentation pratique de la causalité”. Thèse d’Etat. Université des Sciences et de la Technologie Houari Boumedienne. 15 février 1997.

[Mokhtari97b] A. Mokhtari. “Action Based Causal Raisoning”. The international Journal of Artificial, Neural Networks, and Complex Problème-solving technologies. Volume 7, number 2, april 1997. ISSN: 0924-669X Coden aPITE4.

[Mokhtari94] A. Mokhtari. “Apport des logiques non-classiques pour une représentation de la causalité”. Thèse de Docteur d’université. Institut Galilée. Université Paris Nord. 14 juin 1994.

[Nouioua05] F. Nouioua. “Raisonnement stratifié à base de normes pour inférer les causes dans un corpus textuel”. The seventh International Symposium on Programming and Systems (ISPS'2005), Algiers, Algeria (9-11 Mai 2005)(

[Papadimitriou94] C.H. Papadimitriou. “Computational complexity”, Addison-Wesly, Reading, MA, 1994.

[Pearl00] J. Pearl. “Causality: Models, Reasoning, and Inference”. New York : Cambridge University Press; 2000.

[Pearl98] J. Pearl. “On the Définition of Actual Cause”. Technical Report R-259, Departement of Computer Science, University of California, Los Angeles, Calif; 1998.

[Pearl88] J. Pearl. “Probabilistic Reasoning in Intelligent systems”. San Francisco : Morgan Kaufmann. 1988.

[Populaire00] S. Populaire. Introduction aux réseaux bayésiens.Technical report, Heudiasyc, Université de Technologie de Compiègne (France), 2000.

[Prade03] H. Prade, D. Dubois, S. Konieczny. “Logique quasi-possibilistes et mesures de conflit”. 2003.

[Puech90] M. Puech. “Kant et la causalité”. Librairie philosophique J. Vrin, 1990.

[Reiter01] R. Reiter. “Knowledge in Action : Logical Foundations for Specifying and Implémenting Dynamical Systems”. Cambridge, Mass : Mit Press 2001.

[Safar90] B. Safar : “Répondre à des questions de types Pourquoi pas? ”. Revue d’Intelligence Artificielle, 4(2), pp 101-112, 1990.

[Sandwall94] E. Sandwall. “Features and Fluents”. Volume1. Oxford: California Press 1994.

[Schank86] R. Schank. “*Explanation Patterns : Understanding Mechanically and Creatively”*. Hillsday, New Jersey : Lawrence Erlbaum Associates. 1986.

[Scriven59] M.J. Scriven. “Explanation and prediction in evolutionary theory”. Sience 130,477-482. 1959.

[Shimony 03] S. E. Shimony, C. Domshlak. “ Complexity of probabilistic reasoning in derected-path singly-connected Bayes networks”. Revue Artificial Intelligence, vulome 151 ; pp 213-225. 2003.

[Shoham89] Y. Shoham and A. B. Baker. “ Nonmonotonic temporal reasoning”. In D. Gabbay, editor, *The Handbook of Logic in Artificial Intelligence and Logic Programming,*1989.

[Shoham88] Y. Shoham. “Reasoning about change: time and causation from the standpoint of artificial intelligence”. Massachusetts Institute of Technologie, 1988.

[Sosa93] E. Sosa and M. Tooly. “Causation, Prediction, and Search”. New York: Springer-Verlag.1993.

[Thielscher97] M. Thielscher. “Ramification and causality”. *Revue Artificial Intelligence, 89:317-364, 1997.*

[Turner99] H. Turner. “A logic of universal causation. Revue Artificial Intelligence 113. pp 87-123. 1999.

[Wright88] Wright, R. “Causation, responsibility, risk, propability, naked satistics, and proof: pruning the brumble bush by clarifying the concepts”. Iowa Law Review 73, 1001-1077. 1988.

[Wright73] G. H. Von Wright: “On the logic and epistemology of causal relation“”. North holland, 1973.

[Yan05] Yuhong Yan. “Raisonnement qualitatif et abstraction des modèles qualitatifs”. Agente de recherche, Groupe de logique Internet, Institut de technologie de l'information du CNRC, 2005

[Zadeh78] L. Zadeh. “Fuzzy sets as basic for a theory of possibility”. Fuzzy Sets and Systems, vol.1, P 3-28, 1978.