
Liens causaux et explications. Problèmes de modélisation – une discussion préliminaire

Didier Dubois, Henri Prade

I.R.I.T. - UMR CNRS 5505
Université Paul Sabatier
118 Route de Narbonne
31062 Toulouse Cedex 4
{dubois, prade}@irit.fr

RESUME. La notion de causalité joue un rôle essentiel dans l'expression de notre perception des phénomènes de notre environnement et de leur description. Elle s'avère difficile à modéliser pleinement, malgré nombre de 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. Cette note, tout en se situant par rapport à certains travaux récents, suggère quelques éléments de modélisation, ou du moins met en évidence des questions auxquelles toute modélisation devrait s'affronter.

MOTS-CLÉS causalité, explication,

1 Introduction

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. Cela a été aussi le cas en «physique qualitative» où la notion de causalité intervient dans la simulation qualitative de systèmes dynamiques au travers de la propagation de contraintes dans des graphes d’influence [dKB 86].

Dans ces approches, l’identification des causes s’appuie sur des connaissances reliant “causes” et “effets”. Ces relations sont souvent modélisées par des probabilités conditionnelles de la forme $\text{Prob}(\text{effets}|\text{causes})$, les effets n’étant pas

toujours connus avec précision et certitude. Le cadre bayésien [P 88] permet notamment 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). Ce genre de processus peut aussi être appréhendé dans des cadres plus qualitatifs comme celui de la théorie des possibilités (e.g., [SCDP 96]). Ceci peut aussi être relié à l'idée de privilégier les ensembles minimaux de causes susceptibles d'expliquer à eux seuls tous les effets observés (comme l'ont fait Peng et Reggia [PR 90]).

De fait, les réseaux bayésiens, qui fournissent une représentation graphique orientée d'une distribution de probabilité conjointe sont quelquefois appelés "réseaux causaux", de manière éventuellement abusive puisque l'orientation des flèches est induite par l'ordre dans lequel on considère les variables, sans que cette orientation ne reflète nécessairement une relation de causalité qu'on chercherait à modéliser. Voir aussi, par exemple, [DP 90], [RC 03] pour des discussions soulignant également le caractère injustifié d'une lecture causale du théorème de Bayes.

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 [H 1758] qui a souligné qu'on ne pouvait pas déductivement inférer les effets des causes seules, ni les causes des effets. Voir aussi Salmon [Sa 98] pour des essais sur la causalité en relation avec l'idée d'explication scientifique, ou Kyburg [K 90] qui plaide pour une réponse négative à la question « Est-ce que la dynamique classique requiert que la chute des feuilles soit causée par quoique ce soit? ».

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, notamment Shafer [S 96], Pearl [P 98], [P 00], et Halpern et Pearl [HP 01a], [HP 01b]. Cette préoccupation est effectivement naturelle pour l'IA, car il importe 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. Or l'idée de causalité est étroitement liée à celle d'explication (et peut-être encore plus particulièrement à l'idée d'explication négative » [Sa 90], répondant à des questions de type « Pourquoi pas? »).

Nous passons maintenant en revue les principales idées qui ont été avancées, et les difficultés qui ont été soulevées, avant de suggérer quelques autres éléments de modélisation. Ce court article ne prétend pas offrir une (nouvelle) définition de la causalité, mais plutôt recenser différents aspects qui sont à l'œuvre dans l'idée intuitive de causalité, de manière à dégager des éléments utiles, dont la validité psychologique puisse être testée, de manière à pouvoir envisager à terme la réalisation de systèmes d'explications causales qui soient en accord avec la perception humaine de la causalité.

2 Quelques points de vue récents sur la causalité

Différentes analyses ou formalisations de processus causaux, en particulier dans le cadre de modèles à base de probabilités conditionnelles, ou du moins susceptibles d'être "probabilisés" ont été proposés. Nous les rappelons dans cette section. Mentionnons brièvement aussi qu'il existe des propositions de formalisation basées sur les logiques modales (voir Demolombe [De 00] pour une discussion d'ensemble), autour d'une modélisation initiale de G. H. von Wright. Cette modélisation vise à capturer l'idée que l'action d'un agent i est la cause du fait que p soit vrai si et seulement si 1) p était faux avant l'action, 2) si i n'avait pas fait l'action, alors p ne serait pas devenu vrai, ou encore si i maintient p vrai contre le cours normal des choses. Des extensions de cette approche à des actions «non-déterministes» n'ayant pas d'effet unique, ou à des actions multiples ont été développées.

2.1 Shafer et les relations entre événements

Shafer ([S 96], [S 98]) a développé une "logique causale" à partir de l'idée d'arbres d'événements (où chaque nœud correspond à un choix possible, à une situation), qu'on peut ensuite équiper de mesures d'incertitude. Cette approche par arbre d'événements permet de représenter l'enchaînement temporel entre événements, tandis que des "espaces d'événements" permettent de manipuler des situations qui correspondent à des nœuds dans des arbres d'événements plus ou moins détaillés, reliées entre elles non seulement par des relations de précédence, mais aussi par des relations de spécificité et d'implication. Cinq relations de base possibles sont ainsi identifiées entre deux événements (instantanés) S et T . Un événement S peut

- i) être une spécialisation de T (si S arrive, T aussi du même coup)
- ii) requérir T pour sa réalisation (si S arrive, T est déjà arrivé)
- iii) annoncer (être certainement suivi de) T (si S arrive, T doit arriver après)
- iv) être possiblement suivi de T (si S arrive, il est possible que T arrive après)
- v) exclure T (si S arrive, T n'arrivera pas).

Shafer [S 01], sans prétendre à proposer une définition universelle de la causalité, discute le sens très spécifique du concept classique de cause, du point de vue du

droit une action qui est nécessaire et suffisante pour un effet. Il insiste sur le fait que dans la relation "A cause B", A est une action (qui peut être produit par un agent animé, ou même inanimé comme par exemple un orage) tandis que B est un événement instantané, à la différence des cinq relations ci-dessus qui sont des relations entre événements instantanés.

2.2 Halpern et Pearl et la définition de la cause réelle

Plus récemment, dans une approche différente, Halpern et Pearl [HP 01a] 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"). Un exemple typique de ce problème considère deux feux A et B avançant vers une maison. Si le feu A brûle la maison avant le feu B, alors le feu A est considéré comme étant "la cause réelle" pour les dommages, même si on sait que la maison aurait certainement brûlé par le feu B.

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 "est vrai que si A n'avait pas eu lieu, B ne se serait pas produit"). Plus précisément, ces auteurs considèrent que le fait A qu'un sous-ensemble de variables endogènes 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.

Cette approche offre un modèle raisonnable de l'idée de causalité et permet de traiter des exemples qui posent problèmes dans d'autres approches de la causalité. La cause réelle est importante dans les applications en intelligence artificielle. 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. 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.

Par ailleurs, un autre problème important a été récemment étudié par Pearl [P 01]. Il concerne la distinction entre les effets directs et les effets indirects d'un

événement. Cette distinction est en effet d'une grande importance pratique dans beaucoup de domaines, y compris des décisions de politique, et les définitions légales.

2.3 La meilleure explication privilégie ce qui s'est passé d'anormal

Notons cependant que l'idée de contrefactuelle, seule, ne suffit peut-être pas toujours pour rendre compte pleinement de l'idée de causalité (e.g., Miller et Turnbull [MT 90]), ou du moins du choix d'une meilleure explication. En effet, en cas d'événements survenus dans une situation anormale, il semble que l'explication doive privilégier les conditions "anormales" parmi les circonstances qui ont présidé à la réalisation de l'événement (Hart et Honoré [HH 85]). L'idée de sélectionner parmi les causes qui contribuent à la réalisation d'un événement celles qui peuvent être jugées les plus pertinentes, essentielles, déterminantes, etc n'est pas sans rappeler les maximes de Grice quant à l'économie d'un dialogue bien maîtrisé (où seuls les faits pertinents doivent être énoncés par des agents), comme le soulignent Hilton *et col.* [HMSBM 01]. Hilton *et col.* établissent aussi dans leurs expériences que les sujets humains retiennent rarement les causes immédiates comme explications, que les explications vont en général de pair avec des conditions suffisantes pour la réalisation de l'événement (plutôt qu'avec des conditions nécessaires de réalisation), et que les actions humaines sont préférées comme explications des événements.

2.4 La causalité serait-elle indéfinissable

Zadeh [Z 01] insiste sur le caractère quasiment indéfinissable de l'idée de causalité, du moins si on prétend à une définition générale (non restreinte à des classes de cas particuliers), précise et non ambiguë susceptible d'être formalisée logiquement, qui permette de toujours déterminer s'il y a une relation causale entre deux événements (et éventuellement quelle est sa force), et qui soit cohérente avec l'intuition. Il rappelle, après de nombreux auteurs, un certain nombre de difficultés inhérentes à l'idée de causalité, comme les problèmes

- des causes conjointes (à quel "degré" chaque cause contribue-t-elle à la situation résultante) voir à ce sujet (Chockler et Halpern [CH 03]).
- de la distinction entre corrélation statistique et causalité même incertaine (ainsi la perte de l'acuité auditive est-elle due à l'âge, ou plutôt à des expositions répétées au bruit, dont bien sûr la répétition s'accumule avec l'âge),
- ou des chaînes d'événements (la relation de causalité est-elle transitive), comme l'illustre l'exemple suivant [Z 01] "Un ami m'appelle. Il a besoin de mon aide et me demande d'accourir chez lui. Je saute dans ma voiture et je conduis aussi vite que je peux. À un carrefour, je suis percuté par une autre voiture, et je suis tué. Qui a causé ma mort? Mon ami, moi, le chauffeur de la voiture qui m'a percuté. Se posant la question quelle est la cause de ma mort, on pourrait ajouter "Ma trop

grande émotivité qui a nui à mes capacités à éviter l'accident□, “Il fait que je n'avais pas attaché ma ceinture□, ou même “Il fait que le téléphone marchait et que j'étais là pour recevoir l'appel□, etc...

3 Autres éléments de modélisation. Quelques pistes

Dans ce qui suit, on suggère d'une part que les effets attendus d'une cause sont souvent imprécis et incertains, ce qui semble mettre en évidence l'intérêt d'une représentation bipolaire, et on souligne d'autre part quelques spécificités des problèmes d'explication et de détermination de « causes réelles » (voire de responsabilités) où ce qui s'est produit est connu, par rapport aux problèmes de diagnostic où il s'agit de découvrir ce qui s'est produit.

3.1 Effets imprécis et bipolarité

En considérant l'exemple ci-dessus de la personne tuée à la suite de l'appel de son ami, il semble que ce qui fait problème, dans la perception causale de l'enchaînement des faits□“Mon ami appelle □ j'y vais rapidement □ j'ai un accident □ je suis tué”, est que

- pour les différents enchaînements, “Il □ aurait pu se passer autrement. Par exemple, “J'ai un accident□ est une conséquence *possible* de□j'y vais rapidement□, pas une conséquence *certaine*, dans la mesure on n'a pas toujours un accident quand on roule (trop) vite□
- certains actes ne conduisent pas *intentionnellement* à leurs effets□ “J'y vais rapidement□ est intentionnel, mais pas “J'ai un accident□

De fait, si on considère l'exemple suivant, “Il jette une pierre avec force dans la vitre, la vitre se brise□, la perception de la causalité ne semble pas faire de problème, car le résultat apparaît comme une conséquence certaine de mon geste, et qui plus est, va de pair avec un acte qui semble intentionnel. Si au contraire, “Il lance une pierre dans une vitre, pour attirer l'attention de quelqu'un à l'intérieur, et que la vitre se brise (parce que j'ai mal calibré mon geste, ou que la vitre était frêle)□, on est dans une situation où ce qui est survenu n'est qu'une conséquence *possible* de mon geste. Dans ce dernier cas, même si mon geste a provoqué le résultat, la relation de causalité est perçue comme moins immédiate, et s'intégrerait plus difficilement dans une chaîne transitive de relations causales. Ainsi si “A cause B” et que “B cause C”, il semble qu'on ne puisse pas toujours dire que “A cause C”.

Ceci semble pouvoir être mis en rapport avec des problèmes délicats posés par la description de relations de causalité imprécise. En effet, on peut vouloir dire par exemple, que “La grippe cause une fièvre□qui se situe nécessairement dans un intervalle I”, au sens où les valeurs de température en dehors de I sont incompatibles avec la grippe (i.e., □patient, grippe(patient) => □t □ I tel que temp(patient) = t), et où *toutes* les valeurs dans I sont compatibles avec l'éventualité d'une grippe (qui

exprime plutôt que $\exists t \in I, \exists \text{patient tel que } \text{grippe}(\text{patient}) \text{ et } \text{temp}(\text{patient}) = t$. Soit alors un intervalle I' contenant I . Quoique " $\text{fièvre dans } I$ " implique " $\text{fièvre dans } I'$ ", on ne peut pas dire que " $\text{grippe cause fièvre dans } I'$ ", comme l'ont souligné Besnard et Cordier [BC 99]). Au contraire, " $\text{grippe ne cause fièvre}$ " que dans des intervalles qui sont des sous-ensembles de I . Ceci montre que la relation de causalité fonctionne dans une certaine mesure à l'inverse de la relation de conséquence classique.

La représentation de telles relations de causalité imprécises semble requérir une représentation de type bipolaire (comme celle décrite par Dubois *et col.* [DPS 01]), pour exprimer d'une part i) que les valeurs en dehors d'un ensemble S ne sont pas des effets possibles, et d'autre part ii) que toute valeur dans un sous-ensemble $G \subseteq S$ est un effet garanti possible de la cause considérée, la grippe dans notre exemple. Notons que si ce qui est observé n'est pas dans S , on peut écarter pour incohérence la cause associée à l'effet imprécis (G, S) , tandis qu'on pourra considérer la cause associée comme une explication candidate, par abduction, si ce qui est observé est dans G . Quand on a une information *complète* sur les effets possibles de cette cause, alors $G = S (= I)$. A partir du moment où au niveau de description considéré, I contient plus d'une valeur, l'effet observé n'est qu'un des effets possibles de la cause qu'on sait présente, et on se trouve dans la situation évoquée plus haut. Au total on peut ainsi distinguer entre trois situations en termes de scénarios: 1) un événement j suit *nécessairement* un autre événement i ; 2) un événement j *peut pour sûr* suivre un événement i ; 3) il n'est *pas impossible* qu'un événement j suive un événement i (i.e. rien ne l'interdit). La différence entre les situations 2 et 3, est que dans la situation 2, j appartient à un ensemble d'effets garantis possibles de i , tandis que dans la situation 3, j n'appartient qu'à un ensemble d'effets qui ne sont pas impossibles. La différence entre les situations 1 et 2 est que dans la situation 2, j est parmi les effets possibles avérés de i , tandis que dans la situation 1 l'événement j est le seul effet possible de i .

3.2 Diagnostic vs. Explication

Il convient de distinguer entre plusieurs problèmes où l'idée de causalité joue un rôle central. Le problème discuté ici n'est pas celui du diagnostic, lequel concerne la recherche de causes possibles ayant pu plausiblement conduire à des effets observés. C'est celui de discerner dans une chaîne d'événements connus ceux qui entretiennent des relations que l'on peut qualifier de «*causales*», et fournir des explications cognitivement pertinentes à un utilisateur humain, compte tenu des informations disponibles pour l'agent. Les croyances de l'agent doivent aussi être prises en compte, en particulier dans la situation où l'agent croit que c'est A la cause (pour des raisons abductives, étant donné ce qu'il sait), alors que la vraie cause est A' compte tenu des informations dont dispose le système qui doit fournir l'explication. Un autre problème, une fois les relations causales établies, est de déterminer les rôles

éventuels d'agents et leur responsabilités d'un point de vue déontique.

Ainsi, on est supposé disposer à la fois i) d'une description, nécessairement partielle, (en termes logiques, ou en termes de formalismes de représentation de l'incertain) du monde considéré à différents moments successifs, et ii) sous forme de conditionnelles, de transitions potentiellement applicables qui décrivent la manière dont peut évoluer le monde soit de manière spontanée et progressive, soit sous l'effet de l'action d'agents. On passe ainsi d'une description K_t à une description K_{t+1} telles que par exemple K_t implique B et K_{t+1} implique $\neg B$, en sachant quelles transitions se sont déclenchées, ou ont été déclenchées, entre t et $t + 1$. Il reste alors à déterminer quelle est celle, ou quelles sont celles qui individuellement ou collectivement ont provoqué le basculement de B à $\neg B$. Considérant une transition A (qui peut n'être que partiellement connue dans ses effets), et par analogie avec la situation en révision de croyance (voir Dubois *et col.* [DFHP 99]), on peut avoir en fait cinq situations de base par rapport à un énoncé B

- i) B est inconnu en t ; sachant que A s'est produit, B reste inconnu en $t + 1$ (A n'apporte pas d'information sur B)
- ii) B est inconnu en t ; sachant que A s'est produit, B est tenu pour vrai en $t + 1$ (A justifie B)
- iii) B est tenu pour vrai en t ; sachant que A s'est produit, B reste tenu pour vrai en $t + 1$ (B est indépendant de A)
- iv) B est tenu pour vrai en t ; sachant que A s'est produit, B devient inconnu en $t + 1$ (A fait perdre de l'information sur B)
- v) B est tenu pour vrai en t ; sachant que A s'est produit, B est maintenant tenu pour faux en $t + 1$ (A réfute B).

La situation i) ne veut pas dire que A est sans effet sur B (il se peut que si B est vrai dans la réalité, B devient faux quand A se produit), seulement que la venue de A ne permet pas d'apprendre quelque chose sur B . Ces relations, à la différence de celles proposées par Shafer (voir la section 2.1) lient une transition (associée à une action) à la connaissance qu'on peut avoir d'un événement instantané plutôt qu'à son occurrence proprement dite.

Enfin, on peut se demander dans le cas où un effet observé a plusieurs causes possibles, si le fait que la cause réelle soit inattendue, alors qu'on pouvait s'attendre à la venue d'une autre cause et ainsi essayer de se prémunir contre ses effets, peut influencer notre perception du processus causal. Ceci est illustré par l'exemple suivant. «Une personne descend à sa cave. La lumière, qu'elle a allumée en s'engageant dans l'escalier, s'éteint brusquement. La personne tombe. On apprend ensuite qu'il y a eu une coupure d'électricité générale quand la personne était dans l'escalier». Il semble là qu'on puisse dire que la coupure d'électricité a contribué à faire chuter la personne (peut-être de manière conjointe avec l'état physique de la

personne et l'état de l'escalier). On a donc l'enchaînement "coupure → lumière s'éteint → chute". Supposons qu'en fait le système d'éclairage de l'escalier ait été défectueux depuis longtemps (c.-à-d., que la lumière quand on parvenait à l'allumer pouvait se ré-éteindre toute seule), et que *cela était connu* de la personne. Dans ce dernier cas, peut-on encore dire que la coupure d'électricité est en partie la cause de sa chute, ou que c'est plutôt l'inattention de la personne qui en est la cause, puisque la personne devait s'attendre à tout moment à une extinction assurément possible de la lumière → à ses conséquences selon le schéma "Installation défectueuse → lumière s'éteint → chute possible", et donc prendre des précautions ? Ceci illustre le rôle, déjà souligné plus haut, des circonstances « anormales » dans une explication causale, non pas tant peut-être parce qu'elles sont inattendues et que donc elles frappent l'esprit, que parce que ce qui est attendu est prévisible et peut éventuellement être évité.

4 Conclusion

Les propositions existantes en Intelligence Artificielle pour modéliser l'idée de causalité restent partiellement insuffisantes dans leur capacité à rendre compte du phénomène de la perception causale. Des modèles plus satisfaisants doivent prendre en compte le caractère toujours incomplet, et éventuellement incertain de l'information disponible, le niveau de granularité des descriptions, et tirer parti des développements actuels des travaux en IA notamment en matière de mise à jour, de logiques de l'action et de modélisation de l'intentionnalité. Ils pourront alors constituer à la fois des référents susceptibles d'être validés sur la base de tests expérimentaux, et permettre le développement de modèles plus complets pour l'explication causale, dans des cadres éventuellement plus qualitatifs que les probabilités.

Références

- [SCDP 96] S. Benferhat, D. Cayrac, D. Dubois et H. Prade, Explaining 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.
- [BC 99] Ph. Besnard, M. O. Cordier, Explications causales *Journées Nationales sur la Modélisation du Raisonnement (JNMR-1999)*, Paris, 22-23 mars 1999, <http://www.irit.fr/GDRI3-ModRais/articlesJNMR.html>.
- [CH 03] H. Chockler, J. Y. Halpern, Responsibility and blame: A structural-model approach. *Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI 2003)*, 147-153, 2003.
- [dKB 86] J. de Kleer, J. S. Brown, Theories of causal ordering, *Artificial*

Intelligence, 29, 33-61, 1986.

- [D 00] R. Demolombe, Actions et causalité—essais de formalisation en logique, in *Le Temps, l'Espace et l'Evolutif en Sciences du Traitement de l'Information*, (H. Prade, R. Jeansoulin, C. Garbay, eds.), Cépaduès, Toulouse, 2000, 209-223.
- [DFHP 99] D. Dubois, L. Farinas, A. Herzig, H. Prade, Qualitative relevance and independence: A roadmap. *Proc. 15th Int. Joint Conf. on Artificial Intelligence (IJCAI'97)*, Nagoya, 62-67. Extended version: A roadmap of qualitative independence. In: *Fuzzy Sets, Logics and Reasoning about Knowledge* (Dubois, D., Prade, H., Klement, E.P., eds.), Kluwer Academic Publ., 325-350, 1999.
- [DP 90] D. Dubois, H. Prade, Probability theory in Artificial Intelligence, book review of Judea Pearl's Probabilistic Reasoning in Intelligent Systems, *J. of Mathematical Psychology*, 34, 472-482, 1990
- [DPS 01] D. Dubois, H. Prade, P. Smets, Not impossible vs. guaranteed possible in fusion and revision. *Proc. of the 6th European Conference (ESCQARU 2001)*, Toulouse, Springer-Verlag, LNAI, 522-531, 19-21 sept. 2001.
- [HP 01a] J. Y. Halpern, J. Pearl, Causes and explanations: A structural model approach. Part I—Causes. *Proc. of the 17th Conf. on Uncertainty in Artificial Intelligence (UAI'2001)*, (J. Breese, D. Koller, eds.), Seattle, Wa., Aug. 2-5, Morgan Kaufmann Publ., San Francisco, 194-202, 2001.
- [HP 01b] J. Y. Halpern, J. Pearl, Causes and explanations: A structural model approach. Part II—Explanations. *Proc. of the 17th Inter. Joint Conf. on Artificial Intelligence (IJCAI'2001)*, Seattle, Wa., Aug. 4-10, Morgan Kaufmann Publ., San Francisco, 27-34, 2001.
- [HH 85] H. L. A. Hart, A. M. Honoré, *Causation in Law*. Oxford University Press. 2nd édition, 1985.
- [HMSBM 01] D. J. Hilton, J. McClure, R. Sutton, A. Baroux, I. Magaorou, Selecting “the” explanation from a causal chain—A comparison of abnormality, intentionality and statistical principles. Univ. Toulouse Le Mirail, 2001.
- [H 1758] D. Hume, *Enquête sur l'Entendement Humain*. Trad. A. Leroy. GF Flammarion n°343, 1983.
- [K 90] H. E. Kyburg, *Science and Reason*, Oxford University Press, Oxford, 1990.

- [MT 90] D. T. Miller, W. Turnbull, The counterfactual fallacy—Confusing what might have been with what ought to have been. *Social Justice Research*, 4, 1-19, 1990.
- [P 88] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988.
- [P 98] J. Pearl, Graphical models for probabilistic and causal reasoning, in: *Quantified Representation of Uncertainty and Imprecision* (P. Smets, ed.), Vol. 1 of the *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, (D. M. Gabbay, P. Smets, series eds.), Kluwer Acad. Publ., 367- 389, 1998.
- [P 00] J. Pearl, *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, 2000.
- [P 01] J. Pearl, Direct and Indirect Effects. *Proc. of the 17th Conference on Uncertainty in Artificial Intelligence (UAI'2001)*, (J. Breese, D. Koller, eds.), Seattle, Wa., Aug. 2-5, Morgan Kaufmann Publ., San Francisco, 411-420.
- [PR 90] Y. Peng, J. A. Reggia, *Abductive Inference Models for Diagnostic Problem-Solving*, Springer Verlag, 1990.
- [RC 03] J.-D. Rouault, P. Cappy, Bayes theorem and causality, Note Tech. Lab. Neurobiologie de l'Apprentissage, de la Mémoire et de la Communication, Univ. Paris-Sud, 8 p., 2003
- [Saf 90] B. Safar, Répondre à des questions de type *Pourquoi pas ?*, *Revue d'Intelligence Artificielle*, 4(2), 101-112, 1990.
- [Sa 98] W. C. Salmon, *Causality and Explanation*, [Collection de 26 essais écrits entre 1971 et 1995], Oxford University Press, Oxford, 1998.
- [S 98] G. Shafer, Causal logic. *Proc. of the 13th European Conf. on Artificial Intelligence (ECAI'98)*, Brighton, UK, Aug. 23-28, (H. Prade, ed.), John Wiley & Sons, Chichester, 711-719.
- [S 96] G. Shafer, *The Art of Causal Conjecture*. MIT Press, Cambridge, 1996.
- [S 01] G. Shafer, Causality and responsibility. *Cardozo Law Review*, 22, 101-123.
- [Z 01] L. A. Zadeh, Causality is undefinable. (Résumé long) Séminaire, Univ. of California, Berkeley, 18 Janvier 2001.