PHILOSOPHIE DES SCIENCES

Causes, probabilités, inférences

Isabelle Drouet

Préface de Max Kistler

Vuibert

Causes, probabilités, inférences

Causes, probabilités, inférences

Isabelle Drouet

Préface de Max Kistler

Collection « Philosophie des sciences »
dirigée par Thierry Martin

Vuibert

On trouvera en fin de volume, extrait du catalogue Vuibert, des dizaines d'autres livres d'histoire et de philosophie des sciences.

www.VUIBERT.fr

Relecture et correction : Alain Rossignol
Maquette intérieure : Sébastien Mengin/Edilibre.net
Composition et mise en page de l'auteur
Couverture : Isabelle Paisant

ISBN Vuibert 978-2-311-00356-7

Registre de l'éditeur : 596

Table des matières

Préface **IX**

Introduction **1**

1 Les théories probabilistes de la causalité **7**
1.1 L'idée fondatrice 9
1.2 Des corrélations trompeuses à la théorie de Suppes 11
1.3 Limites de la théorie de Suppes 13
1.4 Les théories de la causalité postérieures à celle de Suppes 23
1.5 Théories probabilistes et inférence causale 29

2 La voie hypothético-déductive **35**
2.1 Modèles causaux probabilistes 36
2.2 Une procédure hypothético-déductive 45
2.3 Caractéristiques et limites 53

3 Une voie inductive ? **63**
3.1 Un aperçu sur les méthodes RB 64
3.2 Vue détaillée des méthodes RB 71
3.3 Peut-on parler d'induction ? 80
3.4 Possibilité de l'inférence inductive 84

4 Limites et portée de l'utilisation des réseaux bayésiens **99**
4.1 Hypothèses véhiculées par les méthodes RB 100
4.2 Limites des méthodes RB 108
4.3 Portée des méthodes RB 119
4.4 Pour une méthodologie mixte 131

Conclusion **139**

Appendice **141**

Bibliographie **145**

Index **153**

Préface

IL EST DE TRADITION que les manuels de méthodologie des sciences sociales et les manuels d'argumentation (ou de « logique informelle ») consacrent un chapitre à mettre le lecteur en garde contre le sophisme qui consiste à prendre des corrélations pour des relations causales. Aristote lui-même, l'auteur des premières études systématiques de l'argumentation, présente un sophisme que l'on baptisera plus tard « argument du *cum hoc ergo propter hoc* », c'est-à-dire « cela va ensemble, donc cela est causalement lié ».

« Autre lieu, celui qui tient à ce que l'on considère comme une cause ce qui n'en est pas une, par exemple ce qui s'est passé en même temps que la chose ou après la chose, car les gens confondent l'après la chose avec l'à cause de la chose, et surtout les hommes politiques, tel Démade faisant de l'action politique de Démosthène la cause de tous les malheurs, car c'est après elle que la guerre intervint ». [1]

Ce n'est pas parce que A est accompagné de B, ou parce que B suit A, que A cause B. Aristote parle, il est vrai, de deux événements individuels plutôt que de deux types d'événements ou de deux « facteurs » qui se répètent et auxquels on peut appliquer des méthodes statistiques. Isabelle Drouet a décidé de ne s'intéresser qu'aux relations de cette dernière sorte ; son analyse se limite à ce qu'on appelle le niveau générique, à l'exclusion du niveau individuel. Le domaine ainsi couvert exclut l'histoire à laquelle appartient l'exemple d'Aristote, mais il contient l'ensemble des sciences sociales, de la sociologie et de l'économie à la démographie et à l'épidémiologie.

Les arguments sophistiques qui concluent de manière abusive d'une corrélation entre deux facteurs à l'existence d'une relation causale, sont légion, à la fois dans la vie de tous les jours, dans la presse, dans les débats politiques, etc., mais aussi en sciences, et tout particulièrement dans les sciences sociales.

Prenons l'épidémiologie. L'un des objectifs principaux de l'épidémiologie consiste à découvrir l'existence de relations causales entre des

[1] Aristote, *Rhétorique*, Livre 2, ch. 24, 1401b29 ; trad. P. Chiron, Paris, Flammarion, 2007, p. 409.

substances qui affectent le corps et certaines maladies. Pour détecter de telles influences, on commence par faire des estimations statistiques à partir d'échantillons considérés comme représentatifs de la population concernée. Dans la mesure où l'exposition aux substances en question est susceptible d'être contrôlée, le résultat de telles recherches épidémiologiques est, bien entendu, de la plus grande importance : s'il est établi que le contact des poussières d'amiante avec les poumons est un facteur qui cause le cancer pulmonaire, il est moralement et légalement obligatoire d'empêcher toute possibilité de contact avec l'amiante. Le problème est que l'existence de corrélations statistiques selon lesquelles l'exposition à la substance A augmente ou diminue, en comparaison avec le reste de la population qui n'est pas exposée à la substance A, la probabilité de contracter la maladie B, ne suffit pas à elle seule à établir que l'exposition à A est une cause de la maladie B. Voici un cas récent et spectaculaire qui a fait l'objet d'une controverse scientifique[2] : plusieurs études épidémiologiques effectuées dans les années 1990 ont montré que les femmes qui subissaient une thérapie de remplacement hormonal risquaient moins de souffrir de cardiopathies coronariennes et, en particulier, d'être victimes d'un infarctus du myocarde. Bien entendu, s'il existait vraiment une relation causale entre ce traitement et la diminution du risque d'infarctus, et tant qu'on ne découvre pas d'effets secondaires négatifs, cela fournirait une excellente raison de donner accès à cette thérapie à toutes les femmes dont le niveau hormonal est naturellement faible (notamment après la ménopause). Cependant, d'autres enquêtes, menées avec une méthode plus sophistiquée, appelée « étude randomisée en double aveugle », furent par la suite mises en œuvre, et conduisirent au résultat opposé : selon ces études, l'hormonothérapie s'accompagne d'une augmentation petite mais significative du risque de cardiopathie coronarienne. L'hypothèse avancée pour expliquer cette divergence est que les femmes qui se soumettent à une thérapie hormonale appartiennent en moyenne à des groupes sociaux plus aisés où le régime alimentaire est meilleur et la pratique d'activités sportives plus répandue. Cela suggère que le recours à la thérapie hormonale et le risque diminué d'infarctus sont des effets d'une cause commune, à savoir le fait d'appartenir à un groupe social aisé, plutôt qu'ils ne sont reliés comme cause et effet.

Ce récit d'une erreur et de sa correction est assez clair, et on peut trouver beaucoup de comptes-rendus de ce genre dans les manuels de raisonnement informel et de méthodologie des sciences sociales. Malheureusement, sa clarté est en partie illusoire. Car de tels récits reposent sur une distinction, entre corrélation et causalité, dont l'analyse ne fait l'objet

[2] Voir D.A. Lawlor, G.D. Smith et S. Ebrahim, « Commentary : The hormone replacement-coronary heart disease conundrum : is this the death of observational epidemiology ? », *International Journal of Epidemiology*, 33 (2004), p. 464-467.

d'aucun consensus. Il n'existe à ce jour aucune analyse communément admise des conditions nécessaires et suffisantes pour considérer que le facteur A est une cause du facteur B.

Le livre d'Isabelle Drouet s'attaque à ce problème central de la philosophie des sciences en général, mais dont la solution est particulièrement urgente pour assurer le fondement des sciences sociales. Sa stratégie tout à fait originale distingue son travail des traités purement philosophiques qui ont souvent pour objet la seule analyse conceptuelle et utilisent pour critère principal d'adéquation l'accord de l'analyse philosophique avec l'intuition du sens commun [3]. Mais il diffère encore plus des traités techniques d'analyse de la causalité qui s'y attaquent en développant des algorithmes permettant de découvrir les relations causales à partir de données statistiques [4]. Prenant du recul par rapport à ces deux démarches, son livre jette un nouvel éclairage à la fois sur l'analyse philosophique de la causalité en termes de corrélations et sur les méthodes développées dans les sciences sociales et en intelligence artificielle pour découvrir de manière systématique les relations causales, à partir de l'observation de corrélations statistiques.

Il faut dire que l'entreprise est difficile : la littérature sur chacun de ces domaines est considérable. Surtout, s'agissant en grande partie de travaux qui ont recours à des outils mathématiques sophistiqués, il est passablement difficile de se frayer un chemin pour parvenir aux questions philosophiquement cruciales. Isabelle Drouet a trouvé le moyen de ne rien sacrifier de la rigueur dans l'exposition des différentes analyses et méthodes qu'elle présente, et surtout des difficultés qu'elles rencontrent, sans exiger du lecteur la maîtrise d'outils formels qui vont au-delà de la notion de probabilité conditionnelle. Elle a organisé la présentation de façon à donner au lecteur le choix entre deux parcours : dans un premier temps, il est possible d'omettre les quelques paragraphes qui sont d'un plus grand niveau de technicité et qui sont signalés par un astérisque ; cependant, ces paragraphes plus techniques permettent une compréhension plus aiguë des problèmes conceptuels posés par les méthodes de découverte des causes.

L'ouvrage que l'on va lire n'a pas la prétention de clore le débat sur la possibilité de faire des inférences sur la causalité à partir de données statistiques. Mais il a le mérite considérable de proposer une cartographie intellectuelle de ce domaine de recherche d'une redoutable complexité. Isa-

[3] Un excellent ouvrage de ce genre qui prend pour objet la distinction qui intéresse Isabelle Drouet est E. Eells, *Probabilistic Causality*, Cambridge, Cambridge University Press, 1991.

[4] Les ouvrages de référence sont : P. Spirtes, C. Glymour et R. Scheines, *Causation, Prediction and Search*, 2e édition, Cambridge (Mass.), MIT Press, 2000, et J. Pearl, *Causality. Models, Reasoning, and Inference*, Cambridge, Cambridge University Press, 2000.

belle Drouet situe clairement la place logique des analyses philosophiques, d'une part, et des méthodes algorithmiques, d'autre part. Ne serait-ce que par sa classification de ces dernières en méthodes hypothético-déductives et bayésiennes, elle apporte un gain réel de clarification ; mais elle va plus loin en identifiant clairement les possibilités et les limites de chacune de ces méthodes, telles qu'elles existent aujourd'hui.

La clarification des fondements conceptuels de ces différentes approches et de leurs limites respectives est particulièrement bienvenue dans un contexte où certains modèles de la causalité prétendent à leur tour jouer un rôle fondationnel dans bon nombre de domaines de recherche. Les défenseurs de l'analyse de la causalité par les réseaux bayésiens soutiennent que leurs modèles permettent non seulement de comprendre la recherche scientifique des causes, mais aussi différents aspects de nos mécanismes cognitifs [5]. Parmi les processus cognitifs dont l'analyse bayésienne promet de révéler les fondements logiques, figurent le raisonnement mathématique, l'explication intuitive des raisons d'agir d'autrui, le raisonnement contrefactuel, la prise de décision, de nombreux biais et erreurs découverts par la psychologie de raisonnement, souvent censés établir des limitations intrinsèques de notre rationalité, la catégorisation, l'induction, la sémantique de nombreuses structures linguistiques et surtout l'apprentissage.

L'un des plus grands mérites d'un ouvrage scientifique ou philosophique consiste à énoncer ses propres limites. Le travail d'Isabelle Drouet est à ce titre exemplaire : pour ne donner qu'un exemple, son ouvrage fait clairement état d'une limite qui passe facilement inaperçue et qui affecte les meilleures analyses philosophiques de la causalité en termes de probabilité et les méthodes formelles d'analyse de données en termes de réseaux bayésiens. Toutes ces méthodes reposent sur ce qu'on appelle la condition de Markov : quand un facteur A cause deux autres facteurs B et C, de sorte qu'il est leur cause commune, il est possible de détecter cette situation grâce à un critère d'abord proposé par Reichenbach en 1956 [6], en vertu duquel la cause commune A « fait écran » de façon à faire disparaître la corrélation positive entre les deux effets B et C. Dans le cas de la corrélation trompeuse entre la thérapie hormonale et la probabilité d'un infarctus, la cause commune est l'appartenance à un groupe social aisé (A), et les effets sont le fait de suivre une thérapie hormonale (B) et d'avoir un infarctus (C). Si la situation obéit à la condition de Markov, la corrélation statistique entre B et C disparaît dès qu'on restreint les données aux personnes appartenant au même groupe social. Le facteur A d'appartenance à un groupe social déterminé est dit « faire écran »,

[5] Voir par exemple S. Sloman, *Causal Models*, New York, Oxford University Press, 2005.
[6] H. Reichenbach, *The Direction of Time*, Berkeley, University of Los Angeles Press, 1956.

faisant disparaître toutes les corrélations dont il est responsable et qui ne correspondent pas à des relations causales.

Cependant, un certain nombre d'auteurs ont récemment présenté des situations qui semblent ne pas respecter la condition de Markov. Isabelle Drouet soumet ces « contre-exemples » à la condition de Markov à un examen scrupuleux, introduisant des distinctions pertinentes regardant tant les contre-exemples que les différentes « sous-hypothèses », généralement subsumées sous la condition de Markov. Le résultat peut paraître décevant, mais il est tout à fait constructif : certains de ces contre-exemples prouvent effectivement que les méthodes d'analyse à ce jour disponibles ne sont pas complètes et que nous ne maîtrisons, pour le moment, le lien entre corrélation statistique et causalité que pour une part, il est vrai très importante, des situations. Mais on aurait grandement tort de tenir ce résultat négatif pour un défaut de l'ouvrage d'Isabelle Drouet. Au contraire, il est parfois au moins aussi important de savoir ce qu'on ne sait pas que de bien comprendre ce qu'on sait.

Max Kitsler

Introduction

LA CAUSALITÉ joue un rôle central dans nos explications, qu'elles soient scientifiques ou non. Elle entretient également un lien privilégié avec l'action, puisque l'efficacité de nos actions, individuelles comme collectives, dépend en particulier de nos connaissances causales. Pour autant, nous n'en avons pas fini avec l'analyse de la notion de cause. Cela ne signifie pas que la causalité ait été négligée ; au contraire, elle est un objet d'intérêt philosophique depuis l'Antiquité et il est admis qu'une tâche scientifique majeure consiste à identifier des relations causales. Cela signifie plutôt qu'il existe aujourd'hui un nombre considérable de travaux et d'analyses portant sur la causalité, que la façon dont ces analyses s'articulent n'est pas toujours claire, et surtout qu'aucune d'entre elles n'a réussi à former autour d'elle un consensus.

Parmi les analyses philosophiques portant sur la causalité, on peut distinguer deux grandes familles. D'un côté, certains auteurs développent des théories « physiques » de la causalité, visant à la définir en référence aux caractéristiques physiques des relations de cause à effet. Ces théories trouvent leur origine dans Russell, 1948, et plus précisément dans la notion russellienne de « ligne de monde ». Elles consistent aujourd'hui à penser la causalité soit en termes de processus qui transmettent de l'énergie, transmettent une quantité de mouvement ou plus généralement transmettent ou manifestent une grandeur conservée (Aronson, 1971 ; Fair, 1979 ; Salmon, 1984 ; Salmon, 1994 ; Kistler, 1999 ; Dowe, 1992b ; Dowe, 2000), soit au moyen de la notion de mécanisme causal (Glennan, 1996 ; Machamer *et al.*, 2000). De l'autre côté, certains pensent plutôt les causes comme des « facteur de différence » (*difference-makers*) et cherchent à caractériser la différence qu'elles font relativement à leurs effets. Pour analyser cette différence, certains recourent aux probabilités (Suppes, 1970 ; Cartwright, 1979 ; Skyrms, 1980 ; Cartwright, 1989 ; Eells, 1991), d'autres aux énoncés contrefactuels (Lewis, 1973 ; Lewis, 2000), d'autres encore à des énoncés conditionnels relatifs à ce qui se passerait si l'on intervenait de manière chirurgicale sur un système causal (Price, 1991 ; Woodward, 2003).

Ainsi, de manière générale, ce qu'on appelle « philosophie de la causalité » s'est profondément renouvelé depuis le milieu du vingtième siècle

et il s'agit aujourd'hui d'un champ très dynamique de la philosophie. De façon analogue, la réflexion méthodologique relative à la causalité prend une part croissante dans la littérature scientifique, en même temps que les méthodes utilisées dans les sciences empiriques pour identifier les relations causales sont sans cesse améliorées. Dans un tel contexte, il est frappant de constater que la littérature philosophique et la littérature scientifique portant sur la causalité se développent de manière largement indépendante. Plus exactement, pour autant que la causalité est concernée, l'analyse conceptuelle et la métaphysique, d'une part, et la méthodologie et la philosophie des sciences, d'autre part, sont peu et mal connectées. L'ouvrage qu'on s'apprête à lire vise en premier lieu à articuler l'analyse conceptuelle de la causalité et la méthodologie de l'inférence causale dans les sciences empiriques. Parmi les analyses philosophiques de la causalité aujourd'hui disponibles, ce sont les approches probabilistes qui sont l'objet de notre attention, et pour lesquelles nous nous attelons à cette tâche.

D'une part, en effet, l'apparition des théories probabilistes de la causalité dans les années 1960 a joué un rôle moteur pour le renouveau de la philosophie de la causalité, et les probabilités jouent un rôle central dans le cadre de théories de la causalité développées depuis. En particulier, il existe des versions probabilistes de la plupart des théories contemporaines de la causalité (Lewis, 1986 ; Price, 1991 ; Schaffer, 2001 ; Woodward, 2003, chapitre 7). D'autre part, sur le front scientifique, la méthodologie de l'inférence aux causes à partir de données statistiques ou de connaissances probabilistes est particulièrement développée et bien renseignée. La raison en est que des domaines entiers de la science ont pour principal matériel empirique des données statistiques, véhiculant des informations relatives aux probabilités. C'est le cas, en particulier, des sciences sociales quand elles se veulent quantitatives. « Sciences sociales » peut être entendu ici en un sens large, qui ne correspond ni à la seule sociologie ni à l'ensemble qu'elle forme avec l'économie, mais recouvre également la démographie ou l'épidémiologie.

Que nous nous concentrions, du côté philosophique, sur les théories probabilistes de la causalité n'implique pas que nous les considérions comme les seules analyses correctes du concept de cause. En particulier, accepter l'une ou l'autre des théories contemporaines de la causalité ne suppose pas de rejeter les autres. En effet, l'une des positions aujourd'hui les plus répandues parmi les philosophes de la causalité est sans doute le « pluralisme causal », c'est-à-dire la position selon laquelle les théories contemporaines de la causalité sont, au moins dans une certaine mesure, dans un rapport de complémentarité plutôt que de concurrence. Ici, nous ne prenons parti ni en faveur d'une analyse unitariste du concept de cause, ni en faveur d'une analyse pluraliste. *A fortiori* n'accordons-nous aucun privilège à l'une parmi les différentes formes de pluralisme aujourd'hui disponibles en philosophie de la causalité (Anscombe, 1981 ; Skyrms,

1984 ; Sober, 1985 ; Hall, 2004 ; Cartwright, 2004 ; Psillos, 2009 ; Reiss, 2009).

De façon similaire, nous ne nous prononçons pas ici sur le rapport entre la causalité générique et la causalité singulière. Du point de vue des énoncés, la différence entre causalité générique et causalité singulière est la différence qui existe entre « Fumer cause le cancer du poumon » et « Le tabagisme de Pierre a causé son cancer du poumon », ou entre « Les chutes causent des fractures » et « Ma chute dans l'escalier ce matin a causé la fracture de mon poignet droit ». Ainsi, la causalité singulière est une relation entre des événements singuliers, qui sont effectivement advenus, là où la causalité générique est une relation entre des propriétés – par exemple la propriété de chuter et la propriété de souffrir d'une fracture[7]. La distinction entre causalité générique et causalité singulière soulève à la fois la question de savoir quels rapports elles entretiennent et la question, différente, de savoir si la causalité générique et la causalité singulière doivent être analysées en termes analogues ou bien si, au contraire, elles peuvent, ou même doivent, recevoir des analyses différentes (Good, 1961 et 1961b ; Carroll, 1991 ; Hitchcock, 1995). Nous n'abordons pas ces questions parce que, plus généralement, seule la causalité générique a à nous intéresser ici.

En premier lieu, l'inférence causale probabiliste, c'est-à-dire l'inférence aux causes à partir de connaissances relatives aux probabilités, vise directement la causalité générique, et non la causalité singulière. En effet, ainsi que nous l'avons déjà indiqué, les données empiriques dont il est possible de tirer des connaissances probabilistes sont des données statistiques portant, en tant que telles, sur des propriétés dans des populations. En outre, en second lieu, on peut considérer que les théories probabilistes de la causalité générique constituent une analyse satisfaisante du concept de cause générique. Plus précisément, les théories développées dans Cartwright, 1989, et dans Eells, 1991, rendent compte de manière satisfaisante de tous les cas identifiés comme problématiques pour les théories probabilistes qui les précèdent. D'ailleurs, le débat portant sur les théories probabilistes de la causalité générique s'est tari à la suite de la publication de ces deux ouvrages. À l'inverse, les cas qui semblent résister à une analyse probabiliste sont des cas de causalité singulière, et à l'heure actuelle il n'est clair ni si la causalité singulière peut recevoir une analyse probabiliste, ni quelle analyse probabiliste pourrait être adéquate. Or, l'articulation de l'analyse conceptuelle de la causalité à la méthodologie de l'inférence causale est considérablement plus difficile si l'on s'intéresse

[7] On peut rejeter l'idée selon laquelle des propriétés pourraient entrer dans des relations qui seraient, à proprement parler, causales. Nous laissons à ceux qui rejettent cette idée le soin de nommer la relation que, après les auteurs du domaine, nous appelons « causalité générique », ainsi que celui d'analyser le rapport que cette relation entretient avec la causalité.

à des théories de la causalité insatisfaisantes plutôt qu'à des théories satisfaisantes. Surtout, les difficultés supplémentaires qui apparaissent alors ne concernent pas tant l'articulation de l'analyse conceptuelle à la méthodologie, que l'analyse conceptuelle elle-même. Nous nous en tiendrons donc à la causalité générique.

Ainsi, nous commençons dans le chapitre 1 par présenter les théories probabilistes de la causalité générique. À la fin de ce chapitre, nous montrons que l'analyse de « A cause B » par nos meilleures théories probabilistes de la causalité ne peut pas être utilisée comme un critère permettant de reconnaître si A cause B. En d'autres termes, les théories probabilistes de la causalité ne peuvent en aucun cas être exportées sans modification du domaine de l'analyse conceptuelle, où elles sont satisfaisantes, au domaine de la méthodologie de l'inférence causale probabiliste. Dans ces conditions, et parce que les théories probabilistes sont nos meilleures analyses du rapport entre les concepts de cause et de probabilité, il devient étonnant que, du côté méthodologique, des connaissances causales soient effectivement inférées avec succès à partir de données statistiques ou de connaissances probabilistes. Le problème, dès lors, se précise : il s'agira d'expliquer pourquoi et selon quelles voies les obstacles qui apparaissent si l'on prétend utiliser les théories probabilistes pour inférer des connaissances causales sont effectivement contournés en sciences quand l'analyse causale est menée à partir de données statistiques.

Nous envisageons deux voies selon lesquelles ils peuvent l'être. La première est hypothético-déductive et c'est elle qui est traditionnellement empruntée pour l'analyse causale quand elle prend des données statistiques pour tout matériel empirique. La seconde a été mise au jour beaucoup plus récemment : depuis le milieu des années 1990, des méthodes d'inférence causale probabiliste se sont développées, qui se définissent par le recours à des outils formels qu'on appelle « réseaux bayésiens », et dont les partisans prétendent qu'elles sont inductives. Nous explorons ces deux voies tour à tour, dans les chapitres 2 et 3, et pour chacune nous expliquons pourquoi et à quel prix elle permet d'inférer des connaissances causales à partir de données statistiques d'observation.

En prenant pour objet les méthodes d'inférence causale probabiliste fondées sur les réseaux bayésiens, nous nous trouvons confrontée à de nouveaux enjeux. D'une part, du point de vue philosophique, l'idée selon laquelle ces méthodes permettraient d'*induire* des connaissances portant sur la causalité générique éveille immédiatement la méfiance. En effet, on sait depuis le *Traité de la nature humaine* que l'induction, et plus particulièrement l'induction causale qui est l'objet précis de l'analyse de Hume, soulèvent d'importants problèmes de justification. D'autre part, parce que ces méthodes sont récentes, et même si elles ont attiré l'attention des philosophes des sciences et si ceux-ci leur ont consacré quelques articles, on n'a pas encore fini de déterminer quelles sont leur

signification, leur portée et leurs limites. Dans ces conditions, l'ouvrage vise en second lieu à contribuer à l'analyse philosophique systématique de ces nouvelles méthodes. Ainsi, si le chapitre 3 s'en tient à ce qui est strictement nécessaire pour l'articulation de ces méthodes aux théories probabilistes de la causalité, l'analyse qui y est menée se poursuit dans le chapitre 4. Dans ce dernier chapitre, nous proposons une analyse systématique des limites et de la portée de l'utilisation des réseaux bayésiens pour l'inférence causale probabiliste. Nous y faisons apparaître ce que les réseaux bayésiens changent effectivement pour l'inférence causale probabiliste en général, et en particulier pour la façon dont celle-ci s'articule à l'analyse philosophique du concept de cause.

La méthodologie de l'inférence causale probabiliste, et en particulier certains développements portant sur les statistiques, peuvent se faire relativement techniques. Pour cette raison, nous rappelons dans l'Appendice quelques principes, définitions et résultats, du calcul des probabilités. En outre, de manière plus générale, nous nous sommes efforcée de faire en sorte que la technicité ne puisse en aucun cas constituer un obstacle important à la lecture de l'ouvrage. Ainsi, nous avons regroupé les considérations les plus techniques dans un petit nombre d'unités de texte (paragraphes, sous-sections, sections), dont nous avons fait suivre le titre d'une étoile. Le lecteur qui le souhaiterait peut faire l'économie de la lecture des unités dont le titre est étoilé. En effet, le texte a été conçu afin que la lecture de ces unités ne soit pas requise pour suivre l'argument philosophique principal. D'un autre côté, bien entendu, lire ces passages un peu plus techniques permet non seulement d'acquérir des connaissances plus approfondies concernant la méthodologie de l'inférence causale probabiliste, mais encore de comprendre mieux et plus exactement les arguments que nous développons. Signalons finalement que, pour les citations, leurs références renvoient aux textes qui figurent dans la bibliographie et que c'est nous qui traduisons quand ces textes ne sont pas rédigés en français.

Remerciements

Cet ouvrage est en partie tiré de ma thèse de doctorat, il doit donc beaucoup à Jacques Dubucs, qui a dirigé cette thèse et à qui je tiens à exprimer ma gratitude. Je remercie en outre l'Institut d'Histoire et de Philosophie des Sciences et des Techniques (CNRS / Paris 1 / ENS) pour m'avoir offert un environnement idéal à la rédaction de ma thèse, le Fonds National de la Recherche Scientifique pour m'avoir accordé un mandat post-doctoral qui m'a permis d'écrire cet ouvrage dans de bonnes conditions, ainsi que l'Institut Supérieur de Philosophie de l'Université Catholique de Louvain pour m'avoir accueillie durant cette période. Pour leur lecture attentive d'une version antérieure de ce texte et pour leurs nombreux commentaires, je tiens à remercier Mikaël Cozic et Max Kistler,

et tout particulièrement le second pour avoir également accepté de rédiger la préface de cet ouvrage. Merci aussi à Anouk Barberousse, Donald Gillies, Paul Humphreys et Philippe Mongin pour m'avoir lue et conseillée sur différents points que j'aborde dans le texte, ainsi qu'à Thierry Martin, Delphine Marchand et Sébastien Mengin pour leur disponibilité et leur aide aux différentes étapes du processus éditorial.

CHAPITRE 1

Les théories probabilistes de la causalité

ON PEUT CONSIDÉRER que la philosophie contemporaine de la causalité trouve son origine chez Hume. En effet, les différentes théories de la causalité aujourd'hui disponibles, qu'elles soient des théories physiques ou des théories de la causalité comme facteur de différence, peuvent être considérées comme autant de réactions, très différentes les unes des autres, à l'analyse de la causalité qui est développée dans le *Traité de la nature humaine*. Selon cette analyse, la relation de causalité se caractérise de la façon suivante :

1. les causes et leurs effets sont contigus : « En premier lieu, je constate que tous les objets que l'on considère comme causes ou comme effets sont *contigus* » (Hume, 1739, p. 134) ;
2. les causes précèdent leurs effets : « La seconde relation dont j'observerai qu'elle est essentielle aux causes et aux effets [...] est celle d'*antériorité* temporelle de la cause par rapport à l'effet » (Hume, 1739, p. 135) ;
3. les causes sont régulièrement suivies par leurs effets : « Des objets semblables ont toujours été placés dans des relations semblables de contiguïté et de succession » (Hume, 1739, p. 150).

C'est la clause 3 qui est essentielle ici. D'une part, elle ouvre une tradition d'analyse de la causalité en termes de co-occurrence des causes et de leurs effets ; d'autre part, elle implique qu'une cause est toujours suivie de ses effets et que, en ce sens, elle les rend *nécessaires*. Ainsi que le souligne Anscombe, Hume conçoit la causalité comme une relation de nécessitation, et cela est indépendant de la façon dont il conçoit la nécessité : « En ce qui concerne l'identification de la causalité avec la nécessitation, la pensée de Hume ne l'a pas affaiblie mais, curieusement, l'a renforcée » (Anscombe, 1981, p. 89-90).

La théorie humienne de la causalité se heurte à un certain nombre de difficultés. Parmi celles-ci, on trouve d'abord le fait que la plupart des causes ne sont pas toujours suivies de leurs effets, mais le sont seulement en présence de certains autres facteurs. Ainsi, gratter une allumette n'est pas

toujours suivi de l'embrasement de cette allumette, mais l'est seulement quand il y a de l'oxygène à l'endroit où l'allumette est grattée, quand l'allumette est sèche ... Un raffinement minimal de la théorie humienne visant à prendre en compte les situations de ce type est proposé par Mackie. Mackie considère qu'un effet donné peut être produit de plusieurs façons différentes, chacune étant suffisante mais non nécessaire pour l'effet, et il propose d'appeler « cause » tout élément qui, bien que non suffisant pour l'effet, est essentiel à l'une de ces façons différentes de produire l'effet. Il exprime cette idée de la façon suivante : une cause est « une partie *insuffisante* mais *non redondante* d'une condition *non nécessaire* mais *suffisante* » [1] (Mackie, 1974, p. 62). Ici, « condition » désigne un ensemble de facteurs. Selon la conception développée par Mackie, gratter une allumette est bien une cause de son embrasement : il existe une façon (parmi d'autres) de produire l'embrasement de l'allumette dont le fait de gratter l'allumette est un élément essentiel (« non redondant ») quoiqu'insuffisant – puisqu'il faut aussi, en particulier, qu'il y ait de l'oxygène dans l'air et que l'allumette soit sèche. Ainsi, plus généralement, la conception de Mackie implique qu'il est faux qu'une cause est toujours suivie de son effet ; elle en est suivie seulement quand sont réunis les autres composants de la condition suffisante dont elle fait partie.

Chez Mackie, toutefois, la causalité reste liée à la nécessitation. En effet, les « conditions » qu'il fait entrer dans sa définition de la causalité sont suffisantes pour l'effet considéré ; elles le rendent nécessaire. Or, si l'on peut considérer qu'elle n'est pas avérée au sens strict, l'hypothèse selon laquelle il existe des effets qui ne sont pas rendus nécessaires par l'ensemble de leurs causes est aujourd'hui très plausible. Historiquement, cette hypothèse a pris consistance à mesure que les phénomènes quantiques ont été mieux connus. Ainsi, à titre d'illustration, nous ne connaissons pas d'ensemble de facteurs qui rendrait nécessaire la désintégration du noyau d'un atome radioactif donné dans un intervalle de temps fixé.

Les théories probabilistes s'inscrivent dans la tradition humienne d'analyse de la causalité en termes de co-occurrence des causes et de leurs effets et, au sein de cette tradition, elles visent à faire une place à l'hypothèse selon laquelle certains effets ne sont pas rendus nécessaires par l'ensemble de leurs causes. Pour le dire autrement, il s'agit, au sein d'une certaine tradition humienne d'analyse de la causalité, de rompre le lien qui attache la causalité à la nécessitation. Plus précisément, les théories probabilistes se développent à partir de l'idée, que nous appelons ici « fondatrice », qui consiste caractériser une cause par ce qu'elle rend ses effets plus probables. Une telle caractérisation n'implique ni qu'une cause rende ses effets nécessaires, ni qu'un effet soit rendu nécessaire par l'ensemble de ses causes.

[1] Les italiques sont dans le texte original.

Telle que nous venons de la décrire, la raison d'être des théories probabilistes de la causalité est conceptuelle : il s'agit de donner une analyse de la causalité qui soit plus adéquate que celles auxquelles elle fait suite, c'est-à-dire une analyse qui soit correcte dans un plus grand nombre de cas. Corrélativement, les théories probabilistes se sont développées de manière largement indépendante de considérations relatives à la pratique scientifique, et indépendante en particulier des techniques d'inférence causale probabiliste auxquelles nous nous proposons de les confronter.

La présentation que nous proposons maintenant s'en tient largement au point de vue conceptuel depuis lequel ont été motivés l'apparition des théories probabilistes de la causalité, puis leur développement. Cette approche conduit à adopter un mode de présentation d'inspiration historique, pour partie similaire à celui qui est mis en œuvre dans Hitchcock, 2002. En effet, nous soutenons que, à partir de l'idée consistant à caractériser une cause par ce qu'elle rend ses effets plus probables, les théories probabilistes de la causalité se sont succédé comme autant de raffinements visant à rendre compte correctement de classes de contre-exemples à l'analyse initiale qui soient de plus en plus nombreuses. Les quatre premières sections du chapitre constituent ainsi une présentation d'inspiration historique des principales théories probabilistes de la causalité générique. Les conséquences pour l'inférence causale sont abordées dans une cinquième et dernière section.

1.1 L'idée fondatrice

Les théories probabilistes se sont développées à partir de l'idée qui consiste à caractériser une cause par ce qu'elle rend ses effets plus probables. La notion probabiliste de conditionalisation bayésienne [2] permet de donner à cette idée une formulation précise.

Si l'on note $p(B)$ la probabilité de B, alors la probabilité conditionnelle $p(B|A)$ qu'on obtient en conditionalisant par A est la probabilité de B quand A est le cas. Ainsi, si $p(6)$ est la probabilité d'obtenir un six en lançant un certain dé, la probabilité conditionnelle $p(6|pair)$ est la probabilité d'obtenir un six quand on obtient un nombre pair. Si le dé considéré est équilibré, la première probabilité vaut $1/6$ et la seconde probabilité, conditionnelle, vaut $1/3$. De façon plus générale, par définition, la probabilité conditionnelle $p(B|A)$ est égale au rapport $p(B \wedge A)/p(A)$. Ici, « $B \wedge A$ » désigne la propriété qui est instanciée exactement quand B et A sont instanciées toutes les deux ; dans la suite, nous emploierons les connecteurs $\vee$ et $\neg$ de façon analogue.

[2] « Conditionalisation » est un terme philosophique ; les statisticiens utilisent plutôt « conditionnement » pour parler de la même chose. De manière analogue, j'emploierai « conditionaliser » là où un statisticien préférerait « conditionner ».

En utilisant la conditionalisation et les probabilités conditionnelles, on peut exprimer de la façon suivante l'idée à partir de laquelle les théories probabilistes de la causalité se sont développées :

Théorie probabiliste de la causalité 1.1 (Idée fondatrice :) A cause B si et seulement si $p(B|A) > p(B)$.

On notera que cette proposition n'a pas de sens si la probabilité de A est nulle. Dans ce cas, en effet, la probabilité conditionnelle $p(B|A)$ n'est pas définie. Ici comme dans la suite, nous supposons que la probabilité de A *n'*est *pas* nulle, et nous ne mentionnerons plus cette hypothèse. Dans le contexte présent, cette hypothèse est assez peu substantielle : l'inférence causale probabiliste ne porte généralement pas sur des propriétés de probabilité nulle (telles, par exemple, des propriétés correspondant à des résultats précis sur un *continuum*). Notons, en outre, que si la probabilité de A n'est ni nulle ni égale à 1, alors la proposition $p(B|A) > p(B)$ est équivalente aux propositions $p(B|A) > p(B|A \vee \neg A)$ et $p(B|A) > p(B|\neg A)$. L'idée fondatrice est donc exprimée par n'importe laquelle de ces trois propositions. Ici, A et B sont des propriétés et leur probabilité dépend de la population qu'on considère, puisqu'elle est la probabilité qu'un individu quelconque de cette population les instancie. En outre, nous notons $\neg A$ la propriété qui est complémentaire de A dans la population considérée, c'est-à-dire la propriété qu'un individu de cette population instancie si et seulement s'il n'instancie pas A.

En première approche, la théorie 1.1 est plausible. Ainsi, elle revient à considérer que c'est au sens où elle la rend plus probable que la propriété d'être obèse cause le diabète de type 2. Or, il nous semble clair que dire que l'obésité cause le diabète de type 2[3], c'est bien en particulier dire que, précisément, la probabilité de souffrir de diabète de type 2 est plus grande pour les personnes obèses que pour celles qui ne le sont pas. Plus généralement, il semble bien que la probabilité d'une propriété est plus élevée si l'une de ses causes est présente, que si elle ne l'est pas : comme le souligne Cartwright, « les causes produisent leurs effets ; elles les font advenir » (Cartwright, 2001, p. 255).

Qu'elle augmente la probabilité de ses effets ne suffit toutefois pas à caractériser une cause. Ainsi, il existe des cas d'augmentation de probabilité qui ne sont pas à mettre au compte d'une relation de cause à effet. On parle alors de « corrélations trompeuses » (*spurious correlations*)[4]. L'existence de corrélations trompeuses n'a pas échappé aux tenants des

[3] Par là, nous voulons dire exactement que *la propriété* d'être obèse cause *la propriété* de souffrir de diabète de type 2. Dans la suite, il nous arrivera à nouveau de faire l'économie de la référence aux propriétés au moment de discuter des exemples. Nous le ferons afin de rendre la lecture plus aisée et seulement si cela n'entraîne pas d'ambiguïté relativement à ce que nous disons de la causalité entre propriétés.

[4] À proprement parler, « corrélation trompeuse » désigne toute dépendance probabiliste,

théories probabilistes de la causalité, qui n'ont jamais prétendu faire de l'augmentation de probabilité une condition nécessaire et suffisante de causalité. En d'autres termes, si les théories probabilistes trouvent bien leur origine dans l'idée fondatrice que nous venons d'expliciter, aucun de leurs défenseurs (même parmi les plus précoces) n'a soutenu que la proposition 1.1 était vraie.

1.2 Des corrélations trompeuses à la théorie de Suppes

Parmi les situations dans lesquelles une propriété augmente la probabilité d'une autre propriété qu'elle ne cause pas, deux types sont aisément repérables et d'ailleurs pris en compte dès les premières théories probabilistes de la causalité. Commençons, ici, par présenter ces deux types de situations ou, plus exactement, les deux types de corrélations trompeuses auxquelles elles donnent lieu.

1.2.1 Corrélations trompeuses entre effets et causes

Un premier type de corrélations trompeuses procède du caractère symétrique de la relation d'augmentation de probabilité : si A augmente la probabilité de B, alors B augmente la probabilité de A.

Preuve*. Soit A et B deux propriétés de probabilité non nulle.
Les propositions suivantes sont équivalentes :

- $p(B|A) > p(B)$
- $p(B \wedge A)/p(A) > p(B)$
- $p(B \wedge A)/p(B) > p(A)$
- $p(A|B) > p(A)$.

Ainsi, en particulier, $p(B|A) > p(B)$ équivaut à (et donc implique que) $p(A|B) > p(A)$ □

Il en découle, en particulier, que toute cause qui augmente la probabilité de ses effets voit sa propre probabilité augmentée par chacun de ces effets.

Or la causalité générique n'est pas, en général, symétrique en ce sens : le plus souvent, quand A cause génériquement B, B *ne* cause *pas* génériquement A. Par exemple, s'il est vrai que l'exposition prolongée à l'amiante cause le cancer, il semble indéniable que le cancer ne cause pas l'exposition prolongée à l'amiante. Dans ces conditions, et si l'on

positive ou négative, qui ne correspond pas à une dépendance causale. Ce sont donc les seules corrélations trompeuses *positives* qui constituent un problème relativement à l'idée fondatrice. Cela dit, à toute corrélation trompeuse négative correspond une corrélation trompeuse positive : si A diminue la probabilité sans relation de causalité correspondante, alors $\neg A$ augmente la probabilité de B sans la causer. Nous pouvons donc, de manière générale, nous en tenir à « corrélation trompeuse », sans spécifier à chaque fois s'il s'agit d'une corrélation positive ou négative.

admet que la causalité se marque dans une augmentation de probabilité, le caractère symétrique de la relation d'augmentation de probabilité engendre un premier type de corrélations trompeuses : les corrélations entre effet et cause dans le cas (dont nous verrons plus bas qu'il est le plus général) où un effet ne cause pas sa cause. Ainsi, le caractère symétrique de la relation d'augmentation de probabilité implique que souffrir d'un cancer augmente la probabilité d'être exposé à l'amiante de manière prolongée alors que, nous venons de le voir, souffrir d'un cancer ne cause pas l'exposition prolongée à l'amiante. La corrélation est trompeuse.

1.2.2 Corrélations trompeuses entre effets d'une même cause

Le second type de corrélations trompeuses aisément repérable et tôt identifié dans l'histoire des théories probabilistes de la causalité est celui des corrélations entre effets d'une même cause. Pour comprendre de quoi il s'agit, considérons la propriété d'être fumeur. Ainsi qu'il est aujourd'hui bien connu, cette propriété cause celle de souffrir d'un cancer du poumon. D'un autre côté, elle cause également la propriété d'avoir les doigts jaunis. Les propriétés de souffrir d'un cancer du poumon et d'avoir les doigts jaunis sont toutes deux produites, et donc rendues plus probables, par la propriété d'être fumeur. Surtout, il est vraisemblable que le rapport que chacune de ces propriétés entretient avec celle d'être fumeur implique qu'avoir les doigts jaunis augmente la probabilité de souffrir d'un cancer du poumon : il est plus probable de souffrir d'un cancer du poumon quand on a les doigts jaunis que dans le cas général. Pourtant, on accordera qu'avoir les doigts jaunis ne cause pas le cancer du poumon. À nouveau, nous avons affaire à une corrélation trompeuse.

Depuis les années 1950 au moins, on sait caractériser les situations dans lesquelles existent des corrélations trompeuses de ce type. Plus exactement, l'analyse causale de la direction du temps qui est développée dans Reichenbach, 1956, repose sur l'identification d'une propriété caractéristique des causes qui sont communes à plusieurs effets : quand on conditionalise par une cause de ce type, la dépendance probabiliste entre ses effets qui découle de ce qu'ils sont effets de cette même cause disparaît. Ainsi, la structure composée d'une cause C et de deux de ses effets A et B qui sont en relation de corrélation trompeuse peut être caractérisée au moyen des quatre conditions suivantes (Reichenbach, 1956, p. 161) :

- $p(A|C) > p(A|\neg C)$
- $p(B|C) > p(B|\neg C)$
- $p(A \wedge B|C) = p(A|C).p(B|C)$
- $p(A \wedge B|\neg C) = p(A|\neg C).p(B|\neg C)$.

Reichenbach parle de « fourche conjonctive ». La structure que composent la propriété de fumer, d'une part, et celles de souffrir d'un cancer du poumon et d'avoir les doigts jaunis, d'autre part, est une fourche conjonctive : une fois la propriété d'être fumeur ou la propriété de ne

pas l'être prise en compte, avoir les doigts jaunis ne change plus rien à la probabilité de souffrir d'un cancer du poumon. On dit alors que les propriétés de souffrir du cancer du poumon et d'avoir les doigts jaunis sont indépendantes relativement à la propriété d'être fumeur, ou encore que la propriété d'être fumeur « fait écran entre » (*screens off*) les propriétés de souffrir du cancer du poumon et d'avoir les doigts jaunis. Concernant les différentes notions d'indépendance probabiliste, nous renvoyons le lecteur à l'Appendice.

1.2.3 La théorie de Suppes

A Probabilistic Theory of Causality (Suppes, 1970) fonde la tradition contemporaine d'analyse probabiliste de la causalité. Dans cet ouvrage, Suppes s'appuie sur la propriété des causes communes de faire écran entre leurs effets afin de formuler une théorie probabiliste de la causalité. Plus précisément, sa stratégie est la suivante : il utilise cette propriété des causes communes pour caractériser les corrélations trompeuses du second type en termes probabilistes, et adjoint à l'idée fondatrice une clause stipulant que sont causales seulement les corrélations qui ne tombent pas sous sa caractérisation des corrélations trompeuses. Pour ce qui est des corrélations trompeuses du premier type, entre effets et causes, elles reçoivent chez Suppes un traitement temporel : une cause et son effet sont distingués par ceci que la première est antérieure au second.

La théorie alors obtenue par Suppes peut être énoncée dans les termes suivants (Suppes, 1970, chapitre 2) :

Théorie probabiliste de la causalité 1.2 (Suppes, 1970) A cause B si et seulement si :

1. $p(B|A) > p(B)$ (idée fondatrice).
2. A est antérieure à B.
3. il n'existe pas de propriété C antérieure à A telle que $p(B|A \wedge C) = p(B|C)$.

Dans le cas où les deux premières clauses sont satisfaites, Suppes parle de « cause *prima facie* » (Suppes, 1970, p. 12). Selon Suppes, sa théorie de la causalité vaut aussi bien pour la causalité singulière que pour la causalité générique. Ainsi qu'il doit être clair à ce point, c'est en tant que théorie de la causalité générique qu'elle retient notre attention ici.

1.3 Limites de la théorie de Suppes

La théorie de la causalité générique qui est développée par Suppes se heurte à cinq limites. Ces limites sont autant de types de situations dans lesquelles la théorie échoue à capturer la structure causale réelle –

c'est-à-dire qu'elle conduit à considérer comme causales des relations qui ne le sont pas, ou comme non causales des relations qui le sont.

1.3.1 Corrélations trompeuses entre effets d'une même cause *interactive*

La théorie proposée par Suppes se heurte d'abord à ceci que sa clause 3 ne suffit pas à prendre en compte toutes les corrélations trompeuses entre effets d'une même cause. Plus exactement, il existe des effets d'une même cause qui sont dépendants seulement parce qu'ils sont effets de cette cause – et sont, donc, en relation de corrélation trompeuse –, mais entre lesquels la cause qui leur est commune ne fait pas écran. On trouve des exemples de causes de ce type dans Salmon, 1980 (p. 150-151) ; Cartwright, 1999 (p. 7-8) ; Davis, 1988 (p. 156) ; Salmon, 1984 (p. 168-169) ou Salmon, 1998 (p. 223). Citons ici Salmon (1980, p. 150-151) :

> « Des boules de billard reposent sur le tapis, de telle sorte que le joueur peut mettre la boule noire dans le filet à un bout de la table si et presque seulement si sa boule de choc va dans le filet à l'autre bout de la table. Étant relativement novice, le joueur ne réalise pas ce fait ; par ailleurs, son habileté est telle qu'il a seulement 50 % de chances de mettre la boule noire dans le filet s'il essaie. Supposons, en outre, que si les deux boules tombent dans leurs filets respectifs, la boule noire tombera avant la boule de choc. Soit A l'événement que le joueur tente le tir, B la chute de la boule noire dans le filet du bout de la table, C la chute de la boule de choc dans le filet de l'autre bout de la table. [...] L'événement A, qui doit assurément être considéré comme une cause directe à la fois de B et de C, ne fait pas écran entre B et C, puisque $P(C|A) = 1/2$ alors que $P(C|A \wedge B) = 1$. »

Parce qu'elles comportent des causes communes qui échouent à faire écran entre leurs effets, les structures similaires à celle qui est décrite dans cet extrait se distinguent au premier chef des fourches conjonctives de Reichenbach. Salmon propose de parler de « fourches interactives ». Aussi, nous appellerons « causes interactives » les causes qui, dans une telle structure, jouent un rôle analogue à celui qui est joué ici par A.

Avant d'aller plus loin dans la caractérisation des fourches et des causes interactives, il convient de remarquer à la fois que dans l'extrait que nous venons de citer, Salmon parle d'*événements* là où nous avons jusqu'à présent parlé de *propriétés*, et que la description de l'exemple proposé par Salmon se transpose sans difficulté dans le langage de la causalité entre propriétés. On considérera alors que des propriétés susceptibles d'être instanciées par le système composé de la table de billard, des boules et du joueur, et on dira que la propriété d'avoir sa boule de choc tirée par le joueur :

- est une cause à la fois de la propriété de voir sa boule noire chuter et de la propriété de voir sa boule de choc chuter.
- échoue à faire écran entre ces deux propriétés dépendantes en probabilité.

Revenons maintenant à l'exemple de Salmon. Il semble consister essentiellement en ce que la cause interactive A produit l'un de ses effets si et seulement si elle produit l'autre. Mais Salmon écrit plus précisément ceci : « Le joueur peut mettre la boule noire dans le filet à un bout de la table si et *presque* seulement si sa boule de choc va dans le filet à l'autre bout de la table » [5] (Salmon, 1980, p. 150). L'adverbe « presque » est important ici. En effet, il indique qu'une cause commune A échoue à faire écran entre ses effets B et C non pas seulement si elle produit B exactement quand elle produit C, mais dès que sa production de B n'est pas indépendante de sa production de C. En termes épistémiques, la classe des contre-exemples n'est pas caractérisée par les deux propositions suivantes :

1. savoir que A est instanciée dans un cas particulier n'est suffisant ni pour conclure que B est instanciée, ni pour conclure que C est instanciée dans ce cas particulier ;
2. sachant que A est instanciée dans un cas particulier, apprendre que B (resp. C) est instanciée dans ce cas particulier permet de conclure que C (resp. B) est également instanciée,

mais plutôt par :

1. savoir que A est instanciée dans un cas particulier n'est suffisant ni pour conclure que B est instanciée, ni pour conclure que C est instanciée dans ce cas particulier ;

2'. sachant que A est instanciée dans un cas particulier, apprendre que B (resp. C) est instanciée dans ce cas augmente la probabilité que C (resp. B) soit également instanciée dans ce même cas.

Dans les cas ainsi caractérisés, A n'est pas suffisante pour que la dépendance entre B et C (quand elle existe) soit mise au nombre des corrélations trompeuses au titre de la clause 3 de la définition de Suppes. Si en outre il n'existe pas une autre cause commune à B et C qui, elle, soit telle que la dépendance entre B et C doive être considérée comme trompeuse au titre de la clause 3 [6], alors la théorie de Suppes conduit à considérer que B est une cause de C ou que C est une cause de B (selon la façon dont B et C sont temporellement ordonnées). Or, tel n'est généralement pas le cas – ainsi qu'il apparaît en particulier dans le cas envisagé par Salmon. Il en découle que la théorie de Suppes n'est

[5] Nous ajoutons les italiques.

[6] Dans le cas considéré par Salmon, il semble bien qu'une telle cause n'existe pas – en tout cas pour le grain de description adopté par Salmon.

pas adéquate quand des causes sont interactives. Cela étant établi, dans la suite de la section nous ne parlons que de causes qui *ne* sont *pas* interactives. Plus exactement, nous présentons en termes de causes non interactives les autres situations qui sont problématiques pour la théorie de Suppes – et il convient de garder à l'esprit que les problèmes se dédoublent si, dans les types de situations que nous allons présenter, certaines causes communes sont interactives.

1.3.2 Traitement des corrélations entre effets et causes

Pas plus qu'elle ne résout complètement le problème des corrélations trompeuses entre effets d'une même cause, la théorie 1.2 de Suppes ne résout correctement le problème des corrélations trompeuses entre les effets et leurs causes. En d'autres termes, le critère proposé par Suppes pour le sens des relations de cause à effet est insatisfaisant.

Pour commencer, Suppes suppose qu'on peut ordonner temporellement les *relata* de la causalité générique, c'est-à-dire des propriétés. Or, ni ce que cela peut vouloir dire exactement, ni comment cela peut être fait effectivement n'est clair. On pourrait toutefois envisager de le faire en se référant aux relations de cause à effet singulières qui correspondent à une relation de cause à effet générique donnée. Plus précisément, il s'agirait de :

a) réduire l'ordre de la causalité générique à l'ordre des relations de cause à effet singulières correspondantes ;
b) réduire l'ordre de la causalité singulière à l'ordre temporel.

Ainsi, on dirait que la propriété d'être exposé à l'amiante précède temporellement la propriété de souffrir d'un cancer (et donc la cause génériquement, plutôt que l'inverse) si a) dans les cas singuliers, ce sont des expositions prolongées à l'amiante qui causent des cancers (plutôt que l'inverse), et b) l'ordre de la causalité singulière s'analyse en termes d'antériorité, pour les relations de cause à effet singulières concernées, de l'exposition à l'amiante sur le développement du cancer.

Une telle stratégie est motivée par ceci que les *relata* de la causalité singulière, qu'on conçoit généralement comme des événements, sont temporellement ordonnables. Toutefois, sa seconde composante – qui consiste à rendre compte du sens de la causalité singulière en termes temporels – se heurte à deux limites clairement identifiées :

- elle exclut *a priori* la possibilité de la causalité à rebours, c'est-à-dire la possibilité qu'existe une cause singulière qui serait postérieure à l'un de ses effets. Cette limite est peu embarrassante pour qui considère, comme c'est notre cas, qu'il est très peu probable qu'on identifie jamais une telle cause. Plus problématique, en revanche, est le fait que la solution temporelle au problème du sens de la causalité singulière exclut également *a priori* la possibilité de la causalité simultanée. La thèse selon laquelle il existe des relations

de cause à effet entre événements simultanés a des partisans au moins depuis Kant, et, dans tous les cas, la question de l'existence de telles relations demande à être débattue indépendamment de la question de savoir comment l'ordre de la causalité singulière doit être analysé ;
- elle implique de renoncer au projet de réduire l'asymétrie du temps à celle de la causalité singulière, et l'ordre temporel à l'ordre causal. Or, ce projet a attiré plusieurs auteurs, comme en témoigne Reichenbach, 1956, en particulier.

Selon cette analyse, la clause 2 de la théorie proposée par Suppes n'est même pas convenable dans le cas singulier. Il en découle que, même si la stratégie que nous avons envisagée pour l'ordonnancement temporel des *relata* de la causalité générique pouvait être acceptée, la théorie de Suppes ne résoudrait pas de manière satisfaisante le problème posé par les corrélations trompeuses entre les effets génériques et leurs causes. La théorie de Suppes ne traite donc correctement aucun des deux types de corrélations trompeuses qu'elle prend en compte – c'est-à-dire que, précisément, elle vise à traiter.

Une autre difficulté, sensiblement différente, consiste en ceci qu'il existe des corrélations trompeuses ne relevant ni de l'un ni de l'autre des deux types que la théorie de Suppes prend en compte. Ces corrélations ne sont pas envisagées par Suppes, *a fortiori* ne sont-elles pas traitées correctement dans le cadre de la théorie 1.2. En nous appuyant sur l'analyse développée dans Cartwright, 2001, p. 254 et suiv., nous considérons que les corrélations trompeuses non prises en compte dans le cadre de la théorie de Suppes relèvent de deux grands types. Avant de les présenter, rappelons que nous concentrons maintenant notre attention sur les contextes qui ne sont pas interactifs, c'est-à-dire les contextes tels que les causes communes font écran entre leurs effets.

1.3.3 Corrélations trompeuses entre effets de *plusieurs* causes

En premier lieu, la théorie de la causalité 1.2 proposée par Suppes ne tient pas compte de la possibilité qu'existent des corrélations trompeuses entre effets de *plusieurs* causes. Plus précisément, elle ne tient pas compte de l'existence de corrélations trompeuses entre des effets *communs* à plusieurs causes, qui sont engendrées par cette communauté de causes, et telles qu'aucune de ces causes ne suffit, à elle seule, à faire écran entre ces effets. Pour qu'une corrélation entre A et B soit identifiée comme trompeuse au titre de la clause 3 de la théorie de Suppes, il faut qu'*une* cause commune C suffise à faire écran entre A et B. Or, il existe des situations telles qu'il faut conditionaliser par *plusieurs* causes communes pour que la dépendance probabiliste entre leurs effets disparaisse. Plus précisément, ces situations sont telles que la façon la plus évidente de les décrire conduit à distinguer plusieurs causes communes et qu'il faut condi-

tionaliser par plusieurs de ces causes pour faire disparaître la corrélation trompeuse.

Considérons, à titre d'illustration, la population des familles de deux enfants et admettons l'hypothèse simplificatrice selon laquelle un seul gène code pour la couleur des yeux et un allèle récessif de ce gène code pour les yeux bleus. La propriété de compter un aîné aux yeux bleus et celle de compter un cadet aux yeux bleus sont dépendantes en probabilité : la probabilité de l'une augmente la probabilité de l'autre. Relativement à la propriété de compter une mère qui porte l'allèle codant pour les yeux bleus sur l'un des chromosomes pertinents, cette dépendance diminue mais elle continue d'exister. Un aîné aux yeux bleus continue d'augmenter la probabilité d'un cadet aux yeux bleus, mais dans une mesure moindre qu'initialement. Il en va de même si l'on conditionalise par la propriété de compter une mère qui porte l'allèle codant pour les yeux bleus sur l'autre des deux chromosomes pertinents, ou par les propriétés de compter un père qui porte l'allèle codant pour les yeux bleus sur l'un ou sur l'autre des deux chromosomes. Pour que la dépendance trompeuse disparaisse, il faut conditionaliser par ces quatre propriétés *prises ensemble.* En termes épistémiques, c'est seulement quand on sait quels sont les deux allèles portés par la mère et quels sont les deux allèles portés par le père, qu'apprendre que l'aîné a les yeux bleus ne modifie pas le degré auquel on croit la proposition selon laquelle le cadet a les yeux bleus (ceci, bien sûr, dans le cas où on ne sait pas si le cadet a effectivement les yeux bleus). La clause 3 de la théorie 1.2 est donc trop faible, quand elle conduit à reconnaître une corrélation comme trompeuse seulement si une cause commune suffit à faire écran entre les deux propriétés corrélées.

De façon similaire, la théorie de Suppes ne tient pas compte de l'existence de corrélations trompeuses entre propriétés qui, d'une part, ont chacune *plusieurs* causes et, d'autre part, n'entretiennent aucune forme de rapport causal — au sens où aucune n'est cause de l'autre et où elles n'ont pas de cause en commun. Pour comprendre ce qui est en jeu ici, nous adaptons un exemple introduit par Sober dans un contexte sensiblement différent. À la suite de Sober (Sober, 1988, p. 215),

> « Considérons le fait que le niveau de la mer à Venise et le coût du pain en Grande-Bretagne ont été tous deux à la hausse au cours des deux siècles passés. Disons que tous deux ont augmenté de façon monotone. Imaginons que nous mettions ces informations sous la forme d'une liste chronologique. Pour chaque date, nous relevons le niveau de la mer à Venise et le prix courant du pain britannique. »

Par construction de la situation, le prix élevé du pain britannique et le niveau important des eaux à Venise — ou, pour celui qui jugerait trop vagues les termes « élevé » et « important », la propriété du prix du pain britannique et du niveau des eaux à Venise de dépasser un

seuil adéquatement choisi – sont corrélés positivement. En outre, nous admettons avec Sober à la fois que cette corrélation est trompeuse et qu'il n'y a pas de cause commune au prix élévé du pain britannique et au niveau important des eaux à Venise.

Imaginons maintenant que le prix du pain en Grande-Bretagne ait pour seule cause le prix du blé en Grande-Bretagne. Dans ce cas, le prix élevé du blé britannique suffit à faire écran entre le prix élevé du pain britannique et le niveau important des eaux à Venise : en conditionalisant par la propriété du prix du blé d'être élevé, on rend complètement compte du prix élevé du pain, de telle sorte que le niveau élevé des eaux à Venise devient non pertinent. On notera au passage qu'apparaît ici un type de corrélations trompeuses que la théorie de Suppes prend en charge alors même qu'elle ne le vise pas explicitement : les corrélations trompeuses entre propriétés qui ne sont pas en relation de cause à effet, qui n'ont pas de cause commune, et entre lesquelles une propriété suffit à faire écran.

Mais imaginons maintenant que le prix du pain en Grande-Bretagne dépende *à la fois* du prix du blé et de la demande de pain, que la propriété pour le niveau des eaux d'être important ait également plusieurs causes et que, dans les deux cas (prix du pain britannique et niveau des eaux vénitiennes), il faille prendre en compte plusieurs causes pour rendre compte de la probabilité de l'effet. Alors, il n'existe pas *une* cause C suffisant à faire écran entre les propriétés pour le prix du pain britannique d'être élevé et pour le niveau des eaux à Venise d'être important. La théorie de Suppes implique donc à tort que les deux propriétés entretiennent une relation de cause à effet.

1.3.4 Le paradoxe de Simpson

En second lieu, la théorie 1.2 proposée par Suppes ne prend pas en compte les corrélations trompeuses qui participent du phénomène connu par les statisticiens sous le nom de « paradoxe de Simpson ». À proprement parler, le paradoxe de Simpson se manifeste quand une corrélation est positive (resp. négative) dans une population, mais négative (resp. positive) dans toutes les sous-populations de cette population, pour une certaine partition. Ainsi, il est possible que deux propriétés soient corrélées positivement (resp. négativement) dans une population, mais que la corrélation devienne négative (resp. positive) quand on conditionalise par la propriété d'appartenir, pour une partition donnée, à une des sous-populations de la population considérée. L'exemple le plus fameux est relatif à l'admission en troisième cycle (*graduate studies*) à l'université de Berkeley. De façon générale, le taux d'admission est sensiblement moins élevé pour les filles que pour les garçons. Ainsi pour l'année 1973, 35 % des candidates contre 44 % des candidats ont été admises en troisième cycle. Mais, cette même année, le taux de filles admises est plus élevé que le taux de garçons admis dans presque tous les départements de

l'université quand on les considère séparément [7]. En d'autres termes, pour presque tous les départements de l'université de Berkeley, conditionaliser par la propriété de se porter candidat dans ce département, inverse le sens de la dépendance probabiliste entre être une fille et être admis en troisième cycle. Soulignons que « presque » n'est pas essentiel d'un point de vue mathématique : on pourrait construire une situation dans laquelle le taux d'admission des garçons reste plus élevé que celui des filles au niveau global, mais est moins élevé que lui dans *tous* les départements. Ce qui joue en fait est que les filles se portent massivement candidates dans les départements dont les taux d'admission sont les plus faibles.

L'exemple de l'admission en troisième cycle à Berkeley nous intéresse ici pour la raison suivante : sous l'hypothèse selon laquelle les taux d'admission observés sont de bons indicateurs des probabilités, la théorie de Suppes conduit à conclure qu'être une fille cause la non-admission. En effet, être une fille augmente la probabilité de ne pas être admis et il n'existe pas de propriété qui fasse écran entre les deux premières. En particulier, se porter candidat dans tel département particulier ne satisfait aucune des deux conditions qui permettraient de rejeter la corrélation comme non causale au titre de la clause 3 :

a. la propriété de se porter candidat dans tel département particulier ne peut sans doute pas être considérée comme antérieure (quoi que cela puisse signifier exactement) à celle d'être une fille ;
b. elle ne fait pas écran entre les deux propriétés, mais seulement inverse le sens de la corrélation (presque toujours, c'est-à-dire pour presque tous les départements).

Pourtant, si la théorie de Suppes conduit à considérer qu'être une fille cause la non-admission en troisième cycle à Berkeley, ce qu'on obtient en conditionalisant par l'une ou par l'autre des propriétés de se porter candidat dans tel département particulier suggère que cette conclusion n'est pas satisfaisante. Positivement, ces résultats suggèrent que la relation de cause à effet n'est pas entre être une fille et ne pas être admis en troisième cycle à Berkeley, mais bien plutôt entre être une fille et être admis en troisième cycle à Berkeley.

Le paradoxe de Simpson engage des corrélations trompeuses différentes de toutes celles que nous avons envisagées jusqu'à présent. Plus exactement, dans les instances du paradoxe, les corrélations ne sont pas trompeuses parce qu'elles ne correspondraient pas à des relations de cause à effet, mais plutôt parce qu'elles indiquent mal quelles propriétés sont en relation de cause à effet. Malgré cette différence, elles partagent avec les corrélations trompeuses envisagées plus haut la caractéristique de ne pas être prises en compte par la théorie de Suppes. En effet, pour

[7] Pour les chiffres exacts, on se reportera à Bickel *et al.*, 1975, dans lequel l'exemple est présenté.

ce qui est des instances du paradoxe de Simpson, la théorie de Suppes implique à tort que les corrélations repérées dans la population globale correspondent à des relations de cause à effet.

1.3.5 Indépendances trompeuses

S'il est possible qu'une propriété A diminue la probabilité d'une propriété B dans une population $\mathbf{P}$, mais l'augmente dans toutes les sous-populations de $\mathbf{P}$ pour une partition donnée, il est également possible que A laisse la probabilité de B inchangée dans $\mathbf{P}$, mais l'augmente dans toutes les sous-populations de $\mathbf{P}$ pour une partition donnée. Ainsi, on peut imaginer qu'il existe entre être fumeur et pratiquer régulièrement une activité physique, une corrélation telle que fumer devienne indépendant de son effet qu'est la propriété de souffrir de problèmes cardiaques. Pour dire les choses plus clairement, on peut imaginer que, pour les individus d'une population donnée, l'effet néfaste du tabac sur la santé cardiaque soit annulé par l'effet bénéfique d'une activité physique régulière.[8] Il n'en resterait pas moins que fumer augmente la probabilité de souffrir de problèmes cardiaques à la fois parmi ceux qui pratiquent régulièrement une activité physique et parmi ceux qui n'en pratiquent pas, et que, en ce sens indiscutablement légitime, elle la cause. L'indépendance probabiliste entre la propriété de fumer et celle de souffrir de problèmes cardiaques serait alors trompeuse, en un sens analogue à celui où nous avons parlé plus haut de « corrélations trompeuses ».

Le cas que nous venons d'envisager est fictif et il peut sembler bien peu plausible que les propriétés de fumer et de pratiquer régulièrement une activité physique se trouvent être distribuées de telle sorte que l'effet de la seconde sur la santé vienne exactement contre-balancer l'effet de la première. Toutefois, dans d'autres cas, l'indépendance probabiliste de deux propriétés qui entretiennent une relation de cause à effet est soutenue (et, donc, rendue bien plus plausible) par certains faits réels. Hesslow le premier a mis en avant des faits de ce type : « On a soutenu, par exemple, que la pilule contraceptive (C) peut causer la thrombose (T) [...] Mais la grossesse peut aussi causer la thrombose, et C diminue la probabilité de grossesse » (Hesslow, 1976, p. 291). Une situation similaire a été plus récemment mise en évidence par Steel : « Bien que des routes améliorées, plus sûres, contribuent à diminuer le nombre des accidents de la circulation, elles ont également le regrettable effet secondaire d'augmenter la vitesse, qui est une cause importante de la mortalité sur la route » (Steel, 2006, p. 312). D'un côté, la propriété, pour une région donnée, d'avoir des routes en bon état cause la propriété d'enregistrer une

[8] À notre connaissance, Cartwright a été la première à utiliser cet exemple (Cartwright, 1979, p. 421).

faible mortalité sur la route ; de l'autre côté, et en tant qu'elle agit sur la vitesse, la même propriété cause celle d'enregistrer une *forte* mortalité sur la route. Dans ces conditions, on peut tout à fait envisager que les conséquences probabilistes associées à l'une et à l'autre de ces deux relations de cause à effet se compensent et que la propriété d'avoir des routes en bon état soit indépendante en probabilité de la propriété d'enregistrer une faible mortalité sur la route et ce alors même que, encore une fois, la première propriété cause la seconde.

Une telle indépendance trompeuse diffère de celle que nous envisagions initialement (entre tabagisme et problèmes cardiaques) non seulement par l'existence de relations réelles qui la rendent plausible, là où le premier cas était complètement fictif, mais encore par sa structure. Dans le premier cas, l'influence causale du tabagisme sur les problèmes cardiaques est univoque : il tend à les causer. La question est alors seulement celle de savoir si les effets, dans la population, de cette première tendance sont annulés par ceux d'une autre tendance, associée à une autre propriété : la tendance de la propriété de pratiquer régulièrement une activité physique à causer la propriété de *ne pas* souffrir de problèmes cardiaques. De façon sensiblement différente – et sensiblement plus complexe –, le second exemple est caractérisé par l'existence de deux chemins causaux différents : un chemin direct selon lequel le bon état des routes tend à causer la propriété d'enregistrer une faible mortalité sur la route, et un chemin indirect selon lequel elle tend à causer la propriété de *ne pas* enregistrer une faible mortalité sur la route. Graphiquement, ces deux situations se représentent de manière différente, comme en témoigne la figure 1.1.

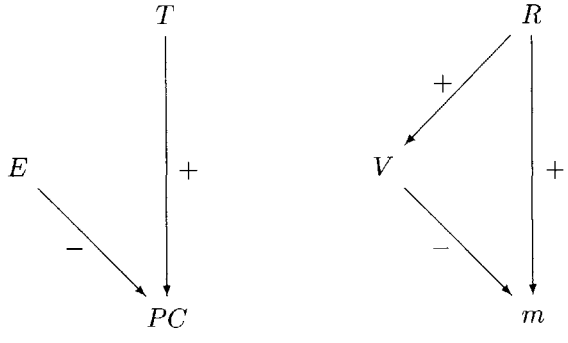


FIGURE 1.1. Le graphe (a) représente les relations causales entre exercice physique régulier (E), tabagisme (T) et problèmes cardiaques (PC), et le graphe (b) représente les relations causales entre bon état des routes (R), vitesse élevée (V) et faible mortalité sur la route (m). Une flèche de A vers B marquée $+$ signifie que A est une cause de B, une flèche de A vers B marquée $-$ signifie que A est une cause de $\neg B$.

Parce qu'ils ont des structures différentes, les exemples que nous avons présentés – et plus généralement les deux types de cas auxquels ils ressortissent respectivement – diffèrent également par les moyens qui permettent de les traiter correctement dans le cadre d'une analyse probabiliste de la causalité. Les deux exemples partagent toutefois la caractéristique de ne pas être analysés correctement dans le cadre de la théorie de Suppes. Plus exactement, la théorie de Suppes ne permet de conclure ni que fumer cause des problèmes cardiaques, ni que le bon état des routes cause une faible mortalité sur la route. La difficulté, toutefois, est sensiblement différente de celle que la théorie rencontrait à l'endroit de la relation entre être une fille et être admis en troisième cycle à Berkeley. Pour ce qui est de l'admission en troisième cycle à Berkeley, la difficulté venait de ce que que la corrélation est trompeuse mais que la clause 3 de la théorie de Suppes ne permet pas de l'identifier comme telle. Dans le cas des indépendances trompeuses, le problème est plutôt que la relation de cause à effet n'est signalée en première instance par aucune dépendance probabiliste (trompeuse ou non) : la clause 2 de la définition de Suppes n'est pas satisfaite et, en conséquence, la théorie de Suppes implique qu'il n'y a pas de relation de cause à effet.

1.4 Les théories de la causalité postérieures à celle de Suppes

1.4.1 Une idée pour dépasser deux limites de la théorie de Suppes

Les théories probabilistes de la causalité qui succèdent immédiatement à celle de Suppes visent à prendre en considération :

- les corrélations trompeuses qui sont des instances du paradoxe de Simpson. Ces corrélations trompeuses sont du même type que celle qui existe entre la propriété d'être une fille et celle de ne pas être admis en troisième cycle à Berkeley ;
- les indépendances trompeuses structurellement similaires à ces corrélations trompeuses, et qui découlent de l'existence d'une propriété dont l'influence sur l'effet considéré vient s'opposer à celle de la cause considérée (et non pas de l'existence de deux chemins causaux menant d'une propriété donnée à une autre). Ces indépendances trompeuses sont similaires à l'indépendance de la propriété de fumer et de la propriété de souffrir de problèmes cardiaques quand cette dernière est corrélée, selon certaines modalités très spécifiques, avec la propriété de pratiquer une activité physique régulière.

Dans l'un comme dans l'autre de ces deux types de cas, la difficulté vient pour la théorie de Suppes de ce qu'elle fait de l'augmentation de probabilité dans une population, une condition nécessaire pour la causalité dans cette population. Contrairement à ce que voudrait la théorie 1.2, il est possible que A cause B dans une population sans que A augmente la

probabilité de B dans cette population. Toutefois, les deux cas que nous avons présentés suggèrent que, si l'on considère non plus la population étudiée mais certaines de ses sous-populations, la causalité peut bien être analysée comme une augmentation de probabilité. Selon cette suggestion, A cause B dans une population P si et seulement si A augmente la probabilité de B dans certaines sous-populations, adéquatement définies, de P. Pour l'admission en troisième cycle à Berkeley, il semble clair que ces sous-populations sont définies par les différents départements dans lesquels les individus peuvent se porter candidats. Pour presque tous départements au sein de la sous-population composée des individus qui se portent candidats dans ce département, la probabilité d'être une fille augmente la probabilité d'être admis en troisième cycle. Pour l'indépendance trompeuse entre la propriété d'être fumeur et celle de souffrir de problèmes cardiaques, ces sous-populations sont, d'une part, la sous-population composée des individus qui pratiquent régulièrement une activité physique et, d'autre part, la sous-population composée des individus qui ne pratiquent pas régulièrement une activité physique. Dans chacune de ces deux sous-populations considérées isolément, la propriété de fumer augmente la probabilité de souffrir de problèmes cardiaques.

Dans les deux cas, les sous-populations pertinentes pour l'analyse probabiliste de la causalité sont caractérisées chacune par une propriété. Pour le premier cas, les propriétés pertinentes sont celles de se porter candidat dans tel ou tel département particulier ; pour le second cas, il s'agit des propriétés de pratiquer et de ne pas pratiquer régulièrement une activité physique. Un moyen de considérer les probabilités dans l'une de ces sous-populations plutôt que dans la population tout entière est de conditionaliser par la propriété caractérisant cette sous-population. L'idée qui se fait jour alors est d'analyser la causalité non pas en termes d'augmentation de la probabilité absolue de l'effet, mais en termes d'augmentation de ses probabilités conditionnelles relatives aux propriétés caractérisant les sous-populations pertinentes. Plus explicitement, il s'agirait de ne plus considérer des relations du type $p(E|C) > p(E)$, mais plutôt des relations du type $p(E|C \wedge A) > p(E|A)$, où A caractérise une sous-population pertinente.

Au-delà des cas particuliers que nous avons choisis comme illustrations, il reste à donner une définition générale des sous-populations pertinentes – et donc des propriétés par lesquelles il faut conditionaliser – pour pouvoir proposer une théorie de la causalité comme augmentation de probabilité. L'examen des deux exemples que nous avons analysés ouvre la voie menant à une telle définition. En effet, dans ces deux cas, il apparaît que les sous-populations à définir sont caractérisées par leur homogénéité relativement à des propriétés différentes de la cause envisagée, mais qui comme elle peuvent avoir une influence sur l'effet. Ainsi, la propriété de pratiquer régulièrement une activité physique est différente de la propriété de fumer, mais influence celle de souffrir de problèmes cardiaques. De façon

analogue, à Berkeley, se porter candidat dans tel département plutôt que dans tel autre est une propriété différente de celle d'être une fille et qui influence l'admission en troisième cycle. L'influence en question est, dans les deux cas, causale. On notera, par ailleurs, qu'elle peut être positive ou négative. Il existe des départements tels que se porter candidat dans ces départements cause l'admission en troisième cycle à Berkeley et des départements tels que se porter candidat dans ces départements-ci cause la *non*-admission en troisième cycle à Berkeley. De même, la propriété de pratiquer régulièrement une activité physique cause non pas la propriété de souffrir de problèmes cardiaques, mais plutôt celle de *ne pas* souffrir de problèmes cardiaques.

Plus généralement, il apparaît alors que les sous-populations qu'il s'agit ici de définir sont des sous-populations d'individus homogènes relativement à toutes les propriétés qui causent soit l'effet considéré soit sa négation, et qui diffèrent de la cause considérée. La population composée de tous les candidats à l'admission en troisième cycle à Berkeley n'est pas homogène en ce sens : elle réunit les candidats de tous les départements, alors même que le département de candidature a une influence causale sur le taux d'admission. Positivement, l'idée qui se fait jour est ici celle d'analyser la causalité comme une augmentation de probabilité toutes choses étant égales par ailleurs et de considérer que toutes choses sont égales par ailleurs quand est fixée la situation relativement aux causes de l'effet ou de sa négation qui diffèrent de celle qui fait l'objet de l'analyse.

1.4.2 Les théories de Cartwright (1979) et de Skyrms (1980)

La ligne d'analyse que nous venons de tracer a abouti à l'idée qui est développée dans Cartwright, 1979 : « C *cause* E si et seulement si C augmente la probabilité de E dans toute situation qui est par ailleurs (*otherwise*) causalement homogène par rapport à E » (Cartwright, 1979, p. 423). En termes plus formels, la proposition est la suivante (Cartwright, 1979, p. 423) :

Théorie probabiliste de la causalité 1.3 (Cartwright, 1979)
A cause B si et seulement si $p(B|A \wedge \mathbf{S}_i) > p(B|\mathbf{S}_i)$ pour tout $\mathbf{S}_i$, où les $\mathbf{S}_i$ sont les descriptions d'états sur l'ensemble des symboles représentant des propriétés qui causent B ou sa négation $\neg B$, mais qui ne sont pas causées par A [9].

Sous cette définition des $\mathbf{S}_i$, chacun décrit une sous-population de la population considérée qui est minimale parmi celles qui sont homogènes

[9] Nous substituons aux notations de Cartwright une description non formelle des $\mathbf{S}_i$, mais la définition que nous donnons ici est bien celle que propose Cartwright dans l'article de 1979.

relativement aux causes de B ou de $\neg B$ qui diffèrent de A et ne sont pas causées par A. En d'autres termes, chaque $\mathbf{S}_i$ fixe la situation relativement aux causes de l'effet considéré qui sont différentes de celle qui fait l'objet de l'analyse. La définition proposée par Cartwright reprend donc bien l'idée que nous avons énoncée un peu plus haut.

La théorie de la causalité 1.3 introduit toutefois un complément à l'idée énoncée. Elle comporte, en effet, la précision selon laquelle les propriétés qui doivent être prises en compte quand on s'intéresse à « A cause B » ne sont pas elles-mêmes causées par A. La raison de cette précision est que les intermédiaires causaux font écran entre une cause et son effet : en règle générale, en conditionalisant par une cause de B qui est causée par A, on rend A et B indépendantes en probabilité. Ainsi, conditionaliser par le coût élevé de la main-d'œuvre dans la boulangerie rend la rareté de la main-d'œuvre en boulangerie indépendante du prix élevé du pain – faisant disparaître la dépendance probabiliste positive qui correspond à ceci qu'une pénurie de main-d'œuvre en boulangerie cause la cherté du pain. D'une cause de B qui est distincte de A et qui n'est pas causée par elle, on dira qu'elle est « causalement indépendante de A ».

La théorie 1.3 n'est pas la seule à reposer sur une analyse du type de celle que nous avons développée dans la sous-section 1.4.1. Plus explicitement, Skyrms suggère que « un affaiblissement intéressant et plausible de la condition [énoncée dans 1.3] [...] est une condition du type Pareto-dominance » (Skyrms, 1980, p. 108). Selon un tel affaiblissement, il n'est pas nécessaire, pour que A cause B, que A augmente la probabilité de B dans *toutes* les situations homogènes relativement aux causes de B et de $\neg B$ qui sont causalement indépendantes de A. Selon Skyrms, pour que A cause B, il suffit que A augmente la probabilité de B dans *une* situation de ce type et ne la diminue dans aucune (Skyrms, 1980, p. 108) :

Théorie probabiliste de la causalité 1.4 (Skyrms, 1980) A cause B si et seulement si :
- $p(B|A \wedge \mathbf{S}_i) \geq p(B|\mathbf{S}_i)$ pour tout $\mathbf{S}_i$
- $p(B|A \wedge \mathbf{S}_i) > p(B|\mathbf{S}_i)$ pour au moins un $\mathbf{S}_i$

où les $\mathbf{S}_i$ sont les descriptions d'états sur l'ensemble des symboles représentant des propriétés qui causent B ou sa négation $\neg B$, mais qui ne sont pas causées par A.

Cette théorie et la théorie défendue dans Cartwright, 1979, ont la même forme, et cette forme leur confère une propriété remarquable : il s'agit d'analyses de la causalité qui sont circulaires, au sens où l'analyse de « A cause B » mobilise le concept de cause.

Si la théorie 1.4 est seulement évoquée dans Skyrms, 1980, Sober (Sober, 1984) et surtout Dupré (Dupré, 1984) donnent des raisons de la

préférer à la théorie 1.3 de Cartwright. Plus précisément, Dupré, 1984, vise en fait à invalider l'idée même de s'intéresser à des sous-populations de la population pour laquelle on veut analyser la causalité en termes probabilistes, mais il nous semble que, si on l'adapte légèrement, le principal exemple de Dupré fait apparaître la supériorité de la théorie 1.4 sur la théorie 1.3. Imaginons, en effet, que des scientifiques découvrent une condition physiologique très rare et, surtout, telle que les individus qui la possèdent voient leur probabilité de souffrir d'un cancer du poumon inchangée par leur éventuel tabagisme. La théorie 1.3 implique que ces scientifiques découvrent en fait que, après tout, fumer ne cause pas le cancer du poumon. À l'inverse, selon la théorie 1.4, cette découverte ne vient pas remettre en cause la thèse selon laquelle fumer cause le cancer du poumon. Il nous semble que cette dernière façon de voir les choses rend plus adéquatement compte de nos intuitions causales concernant le cas en question et que, en ce sens, la théorie 1.4 est préférable.

Cette façon d'argumenter en faveur de la théorie 1.4 au détriment de la théorie 1.3 pourrait suggérer que la théorie 1.4 est encore trop forte ou, d'une certaine manière, arbitraire. Imaginons en effet que, dans la sous-population des individus ayant la condition physiologique très rare imaginée par Dupré, le tabagisme ne se contente pas de laisser la probabilité de souffrir d'un cancer du poumon inchangée, mais qu'elle la diminue. Ne continuerions-nous pas alors de considérer que fumer cause le cancer du poumon ? Et cela n'établirait-il pas qu'il est excessif et injustifié d'imposer qu'une cause ne diminue jamais la probabilité de son effet, c'est-à-dire qu'elle ne la diminue dans aucune des sous-populations pertinentes de la population considérée ? Nous ne le pensons pas. Plus précisément, nous ne pensons pas que, dans la situation que nous venons de décrire, nous continuerions de dire que le tabagisme cause le cancer du poumon *dans la population considérée*. Nous dirions plutôt que le tabagisme cause le cancer du poumon *dans la sous-population des individus qui n'ont pas cette condition physiologique particulière* et qu'elle ne le cause pas dans la sous-population des individus qui l'ont – aussi faible soit la proportion des individus appartenant à cette dernière sous-population. Autrement dit, nous défendons la thèse selon laquelle une condition nécessaire pour que A cause B dans une population est que A ne diminue la probabilité de B dans aucune des sous-populations du type de celles que définissent $\mathbf{S}_i$. La théorie 1.4 suggérée par Skyrms n'est donc ni trop forte ni arbitraire. Nous avons montré par ailleurs qu'elle est préférable à la théorie 1.3 défendue par Cartwright. C'est donc sur elle que nous concentrons maintenant notre attention.

1.4.3 La théorie de Skyrms et les limites de la théorie de Suppes

De même que la théorie 1.3, la théorie 1.4 suggérée par Skyrms vise à prendre en charge deux types de situations qui échappent à l'analyse

de Suppes : les corrélations trompeuses qui constituent des instances du paradoxe de Simpson et les indépendances trompeuses qui leur sont structurellement similaires – c'est-à-dire qui découlent de l'existence de propriétés dont l'influence sur l'effet envisagé vient concurrencer celle de la cause envisagée et qui disparaissent quand on conditionalise par ces propriétés. Restent maintenant les autres difficultés auxquelles nous avons vu que la théorie de Suppes se heurtait. Concernant, d'abord, la distinction entre cause et effet (sous-section 1.3.2), elle est assurée dans le cadre de la théorie 1.4. Plus précisément, l'analyse proposée n'est pas symétrique. Là où l'analyse de « A cause B » mobilise la notion de cause de B (ou de $\neg B$) causalement indépendante de A, celle de « B cause A » mobilise la notion, différente, de cause de A (ou de $\neg A$) causalement indépendante de B. Le caractère symétrique de la relation d'augmentation de probabilité (sous-section 1.2.1) n'implique donc pas la symétrie de la causalité telle qu'analysée par la théorie 1.4. En outre, la théorie 1.4 ne fait pas appel à la notion d'antériorité entre des propriétés, qui est utilisée par Suppes mais dont nous avons vu qu'elle était problématique (sous-section 1.3.2).

Concernant, ensuite, les corrélations entre effets de plusieurs causes qui sont trompeuses mais que la théorie de Suppes échouent à identifier comme telles (sous-section 1.3.3), elles sont correctement analysées dans le cadre de la théorie suggérée par Skyrms. En effet, cette analyse requiert, pour « A cause B », une augmentation de probabilité alors que la situation est homogène relativement à *toutes* les causes de B et de $\neg B$ qui sont causalement indépendantes de A. Il n'est donc pas nécessaire, pour qu'une corrélation entre A et B soit correctement caractérisée comme trompeuse, qu'*une* cause suffise à faire écran entre A et B – ce qui était exactement le point problématique dans le cadre de la théorie de Suppes.

En revanche, la théorie de Skyrms, pas plus que celle de Suppes, ne permet d'identifier comme trompeuses les corrélations entre effets d'une cause commune interactive (sous-section 1.3.1). En effet, la théorie suggérée par Skyrms repose sur l'idée selon laquelle les relations de cause à effet correspondent à des dépendances probabilistes positives dans certains contextes causaux bien précis, et la définition de ces contextes repose au premier chef sur l'idée selon laquelle les causes font écran entre leurs effets si ceux-ci ne sont pas causalement reliés par ailleurs. Or, ce qui caractérise les causes communes interactives est précisément qu'elles ne font pas écran entre leurs effets.

La théorie de Skyrms échoue également à identifier comme causale la relation entre bon état des routes et faible mortalité sur la route dans le cas où les deux propriétés correspondantes sont en relation d'indépendance probabiliste (alors trompeuse). Nous avons vu que, dans ce cas, la difficulté découle de l'existence de deux chemins causaux menant du bon état des routes à une faible mortalité sur la route : le chemin direct selon lequel le bon état des routes cause une faible mortalité sur la

route est concurrencé par un chemin indirect. Or, conditionaliser par les causes des propriétés d'avoir et de ne pas avoir des routes en bon état qui sont causalement indépendantes de la propriété de bénéficier d'une faible mortalité sur la route, ainsi que la théorie de Skyrms le suggère, ne fait pas disparaître les caractéristiques probabilistes problématiques qui sont associées à l'existence du chemin causal indirect. En effet, parce qu'elles sont causées par la propriété d'avoir des routes en bon état, les propriétés qui figurent le long de ce chemin causal indirect ne sont pas causalement indépendantes de cette première propriété. La politique de conditionalisation recommandée par Skyrms ne permet donc pas de prendre en compte l'existence de ce chemin causal indirect.

Deux stratégies principales ont été envisagées afin de résoudre le problème que constituent, pour l'analyse probabiliste de la causalité, les situations dans lesquelles plusieurs chemins causaux se font concurrence. Certains auteurs (Eells et Sober, 1983 ; Cartwright, 1989, ch. 4) ont proposé de raffiner la définition des sous-populations à considérer quand il s'agit d'analyser les relations de cause à effet pour une certaine population. D'autres ont plutôt cherché à distinguer plusieurs notions de causalité, et ont en particulier distingué la causalité (ou « causalité nette ») de la causalité selon tel ou tel chemin causal (Eells, 1991, partie 1 ; Hitchcock, 1993 ; Hitchcock, 2001). Dans les deux cas, le problème constitué par les indépendances trompeuses induites par l'existence de deux chemins causaux est résolu et, de manière plus générale, les théories les plus récentes semblent fournir une analyse correcte de la causalité entre propriétés au sein d'une population. Concernant les théories probabilistes de la causalité postérieures à la théorie 1.4, nous nous contenterons d'ajouter que, au même titre que les théories 1.3 et 1.4, elles sont circulaires. Mais nous ne nous intéresserons pas à leur détail. En effet, la présentation et l'examen de ce détail ne sont pas nécessaires pour confronter aux théories probabilistes les méthodes permettant d'inférer des relations de cause à effet à partir de données statistiques d'observation. Pour le dire autrement, nous verrons que ces méthodes, si elles engagent une conception de la causalité, n'engagent pas une conception qui correspondrait à une théorie de la causalité plus récente que celle qui est envisagée par Skyrms.

1.5 Théories probabilistes et inférence causale

Les théories probabilistes visent à analyser le concept de cause et à préciser ses rapports avec le concept de probabilité. En particulier, elles n'ont pas été développées en vue de donner des critères qui permettraient de reconnaître les relations de cause à effet et de les distinguer de relations non causales. Pour le dire plus concisément, les théories probabilistes de la causalité relèvent de l'analyse conceptuelle, et non pas de la méthodologie de l'inférence causale.

1.5.1 Les théories probabilistes comme principes pour l'inférence causale ?

Les théories probabilistes de la causalité peuvent-elles être exportées du champ de l'analyse conceptuelle de la causalité à celui de la méthodologie de l'inférence causale ? Plus précisément, peuvent-elles servir de principes quand il s'agit d'inférer des connaissances causales de données statistiques d'observation, ou de connaissances probabilistes qu'on peut tirer de ces données ? Dans la sous-section qui commence, nous soutenons qu'elles ne le peuvent pas. Autrement dit, nous soutenons qu'il n'est pas possible d'utiliser les théories probabilistes de la causalité comme si elles énonçaient des critères de reconnaissance pour les relations de cause à effet. Pour la théorie 1.4 qui nous intéresse particulièrement, nous tirons cette conclusion de l'examen de deux de ses caractéristiques.

La première de ces caractéristiques est assez couramment mise en avant quand il s'agit de montrer qu'on ne peut pas considérer que les théories probabilistes énoncent des critères de reconnaissance des causes, mais le raisonnement menant de cette caractéristique à la conclusion visée n'est généralement pas explicité. Cette caractéristique consiste dans ce que nous avons appelé la « circularité » de la théorie de la causalité de Skyrms. Plus exactement, selon la théorie 1.4 envisagée par Skyrms, le concept de cause figure au cœur de l'analyse de « A cause B » : selon cette théorie, « A cause génériquement B » s'analyse en termes d'augmentation de la probabilité de B par A dans des sous-populations caractérisées par leur homogénéité relativement aux *causes* génériques de B ou de $\neg B$ qui sont *causalement* indépendantes de A, c'est-à-dire qui sont distinctes de A et non *causées* par elle. Dans ces conditions, faire de la théorie probabiliste de la causalité envisagée par Skyrms un principe pour l'inférence causale aurait la conséquence qu'il serait nécessaire de connaître toutes les causes de B ou de $\neg B$ causalement indépendantes de A afin de déterminer si A cause B. Il en découle (au moins) deux difficultés conceptuellement distinctes au moment de passer de l'analyse conceptuelle à la mise au jour des relations causales.

La première se rencontrerait si, à partir d'une situation dans laquelle on ignorerait complètement quelles sont les causes d'une propriété B, on cherchait à le découvrir en prenant pour principe la théorie 1.4. Dans cette situation, en effet, qu'il faille connaître les causes de B ou de $\neg B$ qui sont causalement indépendantes de A afin de déterminer si A cause B, implique qu'il faut connaître les causes de B afin de déterminer quelles sont les causes de B. En ce sens, la théorie 1.4 envisagée comme principe pour l'inférence causale peut être considérée comme vide.

La seconde difficulté ne se rencontre pas seulement dans la situation radicale où, sans connaissance causale préalable, on chercherait à identifier l'ensemble des causes de B en s'appuyant seulement sur la théorie 1.4. Elle apparaît aussi dans la situation plus favorable où l'on cherche à déterminer si une propriété A donnée cause une propriété B donnée et où

éventuellement on dispose de certaines connaissances causales préalables. Afin de déterminer si A cause B en s'appuyant sur la théorie de la causalité envisagée par Skyrms, il est nécessaire d'identifier les causes de B ou de $\neg B$ qui sont causalement indépendantes de A. Par conséquent, il est en particulier nécessaire d'identifier les causes de B qui sont différentes de A et, pour chacune, de déterminer si elle est causée par A. Si l'on s'appuie sur la seule théorie 1.4 de la causalité, cela signifie que pour toute cause C de B qui est différente de A, il est nécessaire d'identifier les causes de C qui sont causalement indépendantes de A. Par conséquent, il est nécessaire d'identifier les causes de C qui sont différentes de A et, pour chacune, de déterminer si elle est causée par A. Mais si l'on s'appuie sur la seule théorie 1.4 de la causalité, cela signifie que, pour toute cause D de C qui est différente de A, il est nécessaire d'identifier les causes de D qui sont causalement indépendantes de A. En conséquence, ... Le lecteur aura compris que celui qui chercherait à déterminer si A cause B en s'appuyant exclusivement et directement sur la théorie probabiliste de la causalité que suggère Skyrms se trouverait entraîné dans une régression infinie.

La seconde des caractéristiques de la théorie 1.4 impliquant qu'il n'est pas possible de l'utiliser comme principe pour l'inférence causale est la suivante : selon la théorie 1.4, l'analyse de « A cause B » engage la conditionalisation par les descriptions d'états sur l'ensemble de *toutes* les causes de B ou de $\neg B$ qui sont causalement indépendantes de A. Cette caractéristique est problématique parce que, d'un autre côté, la théorie 1.4 n'implique aucune forme de clôture de l'espace auquel appartiennent les causes d'une propriété donnée B : n'importe quelle propriété peut être une cause de la propriété B, il suffit pour cela qu'elle soit avec B dans le rapport énoncé par la théorie 1.4. En conséquence, il n'existe aucune forme de limitation de l'espace au sein duquel des causes de B peuvent être identifiées, et donc devraient être recherchées. Cet espace est à la fois potentiellement infini et indéterminé. Dans ces conditions, il ne peut pas exister de garantie épistémique de ce qu'on a identifié *toutes* les causes de B ou de $\neg B$ qui sont causalement indépendantes de A. Il est toujours possible d'avoir omis une cause, dans une partie de l'espace des causes possibles de B qui n'a pas été explorée. Dès lors, celui qui prendrait la théorie 1.4 comme *seul* principe pour l'inférence causale ne pourrait jamais établir que A cause B. Cette difficulté se manifeste indifféremment si l'on cherche à déterminer si une propriété A donnée cause génériquement une propriété B donnée et dans la situation moins favorable où l'on cherche simplement à identifier l'ensemble des causes de B.

Chacune des deux caractéristiques que nous venons de discuter appartient à la théorie de la causalité 1.4. En conséquence, les difficultés que l'une et l'autre soulèvent relativement à l'utilisation de cette théorie comme principe pour l'inférence causale ne sont pas exclusives. Par

ailleurs, il convient de noter que ces difficultés ne sont pas spécifiques de la théorie de la causalité 1.4, mais affecteraient aussi bien toute tentative visant à fonder l'inférence causale sur la théorie 1.3 développée dans Cartwright, 1979, ou sur une théorie de la causalité introduite après Skyrms, 1980. D'une part, en effet, nous avons vu que toutes ces théories résolvent le problème des indépendances trompeuses en exigeant d'une cause qu'elle augmente la probabilité de ses effets toutes choses *causales* étant égales par ailleurs. Autrement dit, ces théories introduisent toutes le concept de cause au sein de l'analyse de « A cause B » : elles sont circulaires au sens où l'est la théorie 1.4 envisagée par Skyrms. D'autre part, aucune de ces théories n'impose de contrainte sur l'ensemble des causes possibles d'une propriété donnée et ne comporte d'élément qui viendrait limiter cet ensemble depuis l'extérieur de la théorie. De façon générale, pas plus que la théorie 1.4 sur laquelle notre analyse se concentre, les théories probabilistes de la causalité développées dans Cartwright, 1979, ou ultérieurement à Skyrms, 1980, ne peuvent être utilisées comme si elles énonçaient des critères de reconnaissance des relations de cause à effet.

1.5.2 Conséquences et perspectives

Si les théories probabilistes de la causalité ne peuvent pas servir de principes pour l'inférence causale, il n'en reste pas moins qu'une tâche majeure parmi celles que les scientifiques se sont traditionnellement assignées consiste précisément à identifier des relations causales. Surtout, c'est l'objectif revendiqué de nombreuses études se fondant seulement, au plan empirique, sur des données statistiques, de telles études relevant bien souvent du domaine des sciences sociales. C'est le cas, par exemple, de Meuret et Morlaix, 2006, ou de Jensen et Ahlburg, 2004, qui visent pour le premier à discuter l'hypothèse selon laquelle « les inégalités de richesse sont une des *causes*, au sens strict du mot, des inégalités sociales devant l'école » (Meuret et Morlaix, 2006, p. 49-50) et pour le second à établir « la conclusion selon laquelle la migration *cause* une baisse de la fertilité corrélative de l'adoption par les migrants des normes urbaines de fertilité » (Jensen et Ahlburg, 2004, p. 219) [10]. Le fait que les théories probabilistes ne peuvent pas servir de principes pour l'inférence causale soulève la question de savoir comment l'on peut, dans de telles études, établir des énoncés causaux. Autrement dit, on voit mal comment des connaissances causales peuvent être tirées de connaissances probabilistes, et *a fortiori* de données statistiques, si les théories probabilistes de la causalité, c'est-à-dire nos meilleures analyses probabilistes du concept de cause, ne peuvent pas servir de principes à de telles inférences.

[10] Nous ajoutons les italiques dans les deux cas.

La suite de l'ouvrage vise à faire apparaître comment cela est possible. Plus concrètement, nous envisageons deux voies apparemment susceptibles d'être empruntées par celui qui cherche à inférer des relations de cause à effet à partir de données statistiques ou de connaissances probabilistes. Pour chacune, nous nous attachons notamment à déterminer comment elle se situe par rapport aux théories probabilistes de la causalité. En particulier, nous montrons si et comment on peut, en l'empruntant, contourner les obstacles qui interdisent de faire des théories probabilistes le principe de l'inférence causale.

La première des voies que nous envisageons s'ouvre quand on s'attache au détail de ce qui a été établi dans la sous-section 1.5.1. En effet, ce qu'établit cette sous-section est que les théories probabilistes de la causalité ne peuvent pas être utilisées comme des principes qui permettraient de tirer des conclusions relatives à la causalité à partir de prémisses portant seulement sur les probabilités dans la population étudiée. *A fortiori*, les théories probabilistes ne peuvent pas servir de principes pour tirer des conclusions causales indépendamment de toute théorie causale préalable, à partir de prémisses qui seraient *toutes* des rapports d'observation. Autrement dit, ce que nous avons montré est qu'il n'est pas possible de s'appuyer sur les théories probabilistes de la causalité afin d'*induire* des connaissances relatives aux relations de cause à effet. Dans ces conditions, nous n'avons en particulier pas montré qu'il était impossible d'utiliser des connaissances probabilistes tirées d'observations afin d'établir des relations de cause à effet dans un cadre méthodologique qui serait hypothético-déductif. Cette première voie est d'autant plus attractive que l'analyse causale en tant qu'elle recourt à des données statistiques est généralement reconnue pour être hypothético-déductive. Cette voie fait l'objet du chapitre 2.

Pour ce qui est de la seconde voie possible pour l'inférence de relations de cause à effet à partir de données statistiques et de connaissances probabilistes, il s'agit de celle que certains auteurs ont déclaré tout récemment avoir mise au jour. Plus explicitement, de nouvelles méthodes sont apparues et se sont développées depuis les années 1990, qui reposent de manière fondamentale sur l'utilisation des réseaux bayésiens, et dont les partisans affirment qu'elles permettent d'*induire* des relations de cause à effet à partir de connaissances probabilistes. Dans le chapitre 3, nous présenterons ces méthodes et montrerons comment l'affirmation selon laquelle elles permettent d'induire des relations de cause à effet peut être articulée avec les conclusions de la sous-section 1.5.1. Le chapitre 4 visera à déterminer précisément ce que ces méthodes apportent et peuvent apporter pour l'inférence causale probabiliste, c'est-à-dire pour l'inférence de connaissances causales à partir de données statistiques.

CHAPITRE 2

La voie hypothético-déductive. Modèles causaux probabilistes et inférence causale

Le chapitre qui commence explore la possibilité d'emprunter la voie de l'hypothético-déduction pour contourner les difficultés qui s'opposent à ce que les théories probabilistes servent de principes pour l'inférence causale. En d'autres termes, il s'agit de déterminer si, et surtout comment, il est possible d'inférer des relations de cause à effet à partir de données statistiques dans un cadre méthodologique hypothético-déductif. L'hypothèse selon laquelle cette voie est praticable découle de ce que le chapitre 1 a seulement établi qu'il était impossible de prendre appui sur les théories probabilistes afin d'*induire* des connaissances causales. L'hypothèse s'inscrit, en outre, dans un contexte théorique tel que les investigations relatives à la causalité et qui recourent à des données statistiques sont très généralement et traditionnellement menées dans un cadre hypothético-déductif. Plus précisément, ces investigations sont menées dans le cadre de la modélisation causale (*causal modeling*) et elles commencent par la formulation d'hypothèses causales qui sont ensuite mises en regard des données disponibles. Afin d'autoriser cette confrontation, on donne à ces hypothèses la forme de « modèles causaux probabilistes ».

Il s'agit pour nous ici d'explorer en principe une possibilité méthodologique, avant et plutôt que de rendre compte de pratiques scientifiques. Ainsi, une partie importante du travail mené dans le présent chapitre consiste à définir une procédure hypothético-déductive prenant pour prémisses des données statistiques et permettant de conclure à l'existence de certaines relations de cause à effet. Pour cela, nous mobilisons les outils conceptuels et techniques de la modélisation causale, tels qu'ils sont explicitement identifiés et définis par la méthodologie qui lui est relative. Plus exactement, la procédure que nous définissons est la plus exigeante – et du coup, nous semble-t-il, la plus fiable – parmi les procédures hypothético-déductives que nous pouvons imaginer forger au moyen des outils habituellement mobilisés dans le cadre de la modélisation causale.

En vue de justifier une telle approche, revenons un instant sur le principal objectif de l'ouvrage. Il s'agit en effet d'articuler, pour les approches probabilistes, les analyses philosophiques du concept de cause à la méthodologie de l'inférence causale. Dans ces conditions, il semble clair que la méthodologie de l'inférence causale doit être envisagée au plan de ses principes avant de l'être au plan des pratiques. Cela nous semble d'autant plus nécessaire que les pratiques relevant de la modélisation causale sont à fois considérablement variées et, surtout, souvent très éloignées des principes méthodologiques qui les guident – ou même seulement de la mise en œuvre de procédures qui seraient réfléchies et rigoureuses. Plus précisément, nous considérons que la diversité des pratiques réelles dans le domaine de la modélisation causale se comprend en grande partie en termes d'écart aux normes pourtant reconnues dans la littérature méthodologique du domaine. Dans ces conditions, s'intéresser aux principes de l'inférence causale probabiliste permet également de retrouver la rigueur et l'unité méthodologiques qui manquent trop souvent à la modélisation causale telle qu'elle est effectivement pratiquée.

Après avoir présenté les modèles causaux probabilistes qui sont mobilisés dans le cadre de la modélisation causale (section 2.1), nous les utilisons afin de définir une procédure hypothético-déductive pour l'inférence causale probabiliste (section 2.2). Les obstacles mis en évidence dans la sous-section 1.5.1 sont ainsi contournés. Mais ce contournement est corrélatif de certaines caractéristiques et de certaines limites de l'inférence causale probabiliste hypothético-déductive, qui sont analysées finalement (section 2.3).

2.1 Modèles causaux probabilistes

Les modèles causaux probabilistes constituent les principaux outils de la modélisation causale. Au sein de la modélisation causale, nous nous tournons plus spécifiquement vers la modélisation structurelle (ou modélisation d'équations structurelles) (*structural equation modeling*), et vers ses modèles structurels. Ce sont eux qui sont adaptés pour étudier les relations causales qui sont à l'œuvre au sein d'une population *à un instant du temps donné*, et plus précisément pour les étudier à partir de données d'*observation* seulement, c'est-à-dire indépendamment de toute manipulation [1].

La justification que nous venons de donner suppose que nous entendons « modélisation structurelle » au sens large qu'elle a, par exemple,

[1] Le premier point distingue les modèles qui nous intéressent ici en particulier de ceux qui sont développés par Granger pour les relations causales quantitatives en tant qu'elles se déploient dans le temps (Granger, 1969 ; Granger, 1980). Le second point exclut du domaine de notre analyse, en particulier, les modèles du type de ceux que propose Rubin (Rubin, 1974).

dans Kline, 1998. Il convient de souligner ici que cet usage n'est pas universel. Ainsi, dans la typologie des modèles causaux probabilistes proposée au début du chapitre 3 de Russo, 2008, celle-ci considère que seuls les modèles du premier des types qu'elle envisage relèvent de la modélisation structurelle. Or au sens de Kline, c'est le cas des modèles des premier, deuxième et cinquième types identifiés par Russo c'est-à-dire, respectivement, des modèles de cheminement (*path models*), des modèles de structure de covariance (*covariance structure models*) et des modèles multi-niveaux. De manière plus générale, face au flottement qui entoure la terminologie de la modélisation causale en général et de la modélisation structurelle en particulier, nous prenons les deux partis suivants : en premier lieu, nous utilisons aussi peu de termes techniques qu'il est possible et nous veillons à toujours indiquer en quel sens nous les employons ; en second lieu, nous nous en tenons autant que faire se peut à la terminologie qui est utilisée dans Kline, 1998.

2.1.1 Modèles structurels

Nous commençons ici par présenter les principales caractéristiques générales des modèles structurels. Ces caractéristiques sont générales au sens où elles sont partagées par tous les modèles structurels.

Un modèle structurel est caractérisé d'abord par le système qu'il prend pour objet, c'est-à-dire par le système réel dont il est un modèle structurel. Par « système », nous entendons un ensemble de phénomènes qui sont à la fois causalement connectés les uns aux autres et relativement isolés, du point de vue causal, des phénomènes qui n'appartiennent pas à l'ensemble. À titre d'illustration et d'explicitation, considérons Caldwell, 1979, étude démographique classique consacrée à l'influence de l'éducation maternelle sur la mortalité infantile dans les pays en voie de développement. Caldwell prend en compte ces deux phénomènes – l'éducation de la mère et la mortalité infantile –, mais aussi l'activité de la mère, celle du père, le lieu de résidence et le type de mariage (monogame ou polygame) (Caldwell, 1979, p. 406). Cet ensemble de phénomènes constitue un système dans la mesure où il contient tous les phénomènes qui causent la mortalité infantile ou qui affectent le rapport causal que celle-ci entretient avec l'éducation maternelle. Plus exactement, cet ensemble contient tous les phénomènes que Caldwell a pu envisager comme causant la mortalité infantile ou affectant le rapport qu'elle entretient avec l'éducation maternelle. Nous ne traitons pas ici la question de savoir si, et surtout à quelles conditions, le second ensemble est identique au premier. Toutefois il convient de réaliser que les ensembles de phénomènes constituant des systèmes ne sont pas donnés comme tels, et que leur identification adéquate requiert des connaissances déjà causales.

Dans les modèles structurels, les phénomènes qui composent un système sont représentés au moyen de variables. Ainsi, dans l'article de

Caldwell, le niveau d'éducation de la mère est représenté par une variable susceptible de prendre les trois valeurs suivantes : pas de scolarisation, scolarisation primaire seulement, scolarisation au-delà de la scolarisation primaire (Caldwell, 1979, p. 399). Dans ce cas, la variable utilisée peut prendre un nombre fini de valeurs : elle est discrète. En outre, nous dirons qu'il s'agit d'une variable observable, parce qu'on peut déterminer par la seule observation la valeur qu'elle prend pour une femme donnée. Par contraste, un modèle causal probabiliste peut également porter sur des variables continues et/ou sur des variables latentes, c'est-à-dire dont les valeurs ne sont pas observables au sens où le sont les valeurs de la variable « éducation de la mère » utilisée par Caldwell.

La variable « statut socio-économique » telle qu'est utilisée dans un nombre important d'études relevant des sciences sociales est typiquement une variable latente. La valeur de cette variable ne peut pas être observée directement ; elle est déterminée, selon une méthode qui peut varier, à partir de diverses observations – qui portent en général sur les revenus, le niveau d'éducation et l'activité exercée. De façon similaire, dans un article consacré à l'éveil (*arousal*) chez les enfants âgés de 10 à 12 ans (Williams *et al.*, 2002), les auteurs n'observent pas directement la valeur de la variable représentant l'éveil, mais recourent à un test (le « *Visual Similes Test II* ») qui permet d'évaluer l'éveil affectif, d'une part, et l'éveil cognitif, de l'autre[2]. En revanche, les résultats obtenus aux différentes parties du test sont représentés au moyen de variables qui, elles, sont observables. De manière générale, si un système donné est représenté au moyen d'un modèle structurel qui comporte des variables latentes, alors ce modèle comporte également des variables observables. À l'inverse, il est possible d'utiliser seulement des variables observables. Notons finalement que dans un même modèle structurel peuvent figurer des variables représentant des phénomènes qui relèvent de différents niveaux de réalité. Ainsi, les travaux de Courgeau relatifs aux comportements migratoires individuels (par exemple Courgeau, 2003) témoignent de la fécondité de l'idée consistant à prendre en compte des variables de niveau(x) supérieur(s), relatives en l'espèce aux régions de résidence.

Ce que nous venons de décrire, concernant les variables, vaut pour tous les modèles causaux probabilistes, et nous en venons seulement maintenant à ce qui concerne spécifiquement les modèles structurels qui nous intéressent ici. Étant donné un ensemble de variables, un modèle structurel représente une hypothèse relative aux relations de cause à effet qui existent entre les variables de cet ensemble, ou entre les variables d'un sous-ensemble de cet ensemble (auquel cas, les variables de ce sous-ensemble sont généralement les variables latentes de l'ensemble de variables considéré). Les hypothèses représentées par les modèles

[2] L'article est mentionné et commenté dans Kline, 1998, p. 70-74.

structurels sont qualitatives, c'est-à-dire qu'elles portent sur l'existence de relations de cause à effet avant de porter, éventuellement, sur l'intensité ou la forme de ces relations. En outre, elles prennent en compte le caractère multivarié de la causalité, c'est-à-dire le fait qu'une même cause peut avoir plusieurs effets et le fait qu'un même effet peut avoir plusieurs causes. Ainsi, selon le modèle construit dans Meuret et Morlaix, 2006, pour représenter l'influence de l'origine sociale sur la performance scolaire, la compréhension de l'écrit est causée par le statut professionnel du père à la fois directement et indirectement, passant dans ce dernier cas, d'un côté, par des « facteurs externes à l'établissement » dans lequel l'élève est scolarisé (éducation de la mère, possessions culturelles de la famille, temps de travail scolaire à la maison ...), et, de l'autre côté, par des « facteurs internes à l'établissement » (taux d'encadrement en enseignants, sentiment des élèves qu'ils sont soutenus par leurs professeurs, discipline qui règne dans la classe ...)[3].

Une hypothèse de ce type se représente naturellement au moyen d'un graphe orienté dont les nœuds sont les variables considérées et dont les flèches représentent des relations de cause à effet directes entre ces variables. Une relation de cause à effet entre deux variables X et Y d'un ensemble $\mathbf{V}$ donné est directe si l'influence de X sur Y ne se réduit pas à l'influence sur Y de variables de $\mathbf{V}$ elles-mêmes causées par X. Soulignons que la propriété, pour une relation de cause à effet, d'être directe est relative à l'ensemble de variables qu'on considère. Soit, en effet, les variables correspondant, pour l'une, à la propriété d'être porteur d'un gène de susceptibilité pour le cancer du sein et, pour l'autre, de recevoir une chimiothérapie, et notons ces variables G et C, respectivement. Relativement à un ensemble de variables $\mathbf{V}$ auquel appartient la variable A correspondant à la propriété d'être atteint d'un cancer, G ne cause pas directement C. En effet, l'influence de G sur C se réduit à l'influence de A sur C. Mais si ni A ni aucune autre variable causalement intermédiaire entre G et C n'appartient à $\mathbf{V}$, alors G cause directement C dans $\mathbf{V}$.

Mais revenons à l'article de Meuret et Morlaix et à l'influence du statut professionnel du père sur la compréhension de l'écrit. Sous l'hypothèse simplificatrice selon laquelle les possessions culturelles et le taux d'encadrement en enseignants sont les seuls facteurs pertinents[4], l'hypothèse causale considérée est celle que représente le graphe de la figure 2.1. Parce qu'il représente une hypothèse causale, un graphe du type de celui qui constitue la figure 2.1 est généralement appelé « graphe causal ». Le graphe causal que nous venons de tracer présente la particularité

[3] Notons que toutes ces variables sont considérées ici comme des variables individuelles. Une approche multi-niveaux aurait également pu être envisagée.

[4] Ils sont en fait, respectivement, le facteur externe et le facteur interne pour lesquels l'effet indirect du statut professionnel du père sur la compréhension de l'écrit est le plus important selon Meuret et Morlaix (Meuret et Morlaix, 2006, p. 61).

d'être acyclique, c'est-à-dire qu'il n'est pas possible, en suivant les flèches du graphe, de revenir à une variable dont on serait parti. De manière équivalente, on dit que le modèle auquel il est associé est récursif.

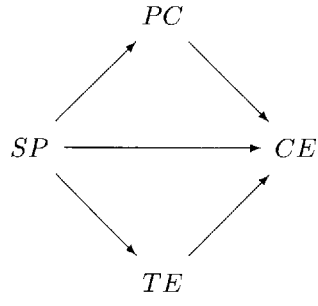


FIGURE 2.1. SP, CE, PC et TE représentent respectivement le statut professionnel du père, la compréhension de l'écrit, les possessions culturelles de la famille et le taux d'encadrement dans l'établissement fréquenté.

L'hypothèse causale que représente un graphe causal peut également être représentée au moyen d'un ensemble d'équations, ce mode de représentation étant d'ailleurs souvent privilégié dans les travaux relevant de la méthodologie de la modélisation structurelle. Les équations sont alors en nombre égal à celui des variables du modèle et telles que chaque variable figure du côté gauche de l'une d'entre elles exactement. Cette équation fait apparaître la variable en question comme une fonction des variables qui la causent directement selon l'hypothèse causale représentée par le modèle, ainsi que d'un terme d'erreur. Pour l'hypothèse causale adaptée de Meuret et Morlaix, une équation appartenant à la représentation du système fait de CE une fonction de SP, PC, TE et d'un terme d'erreur ε_{CE}, ce qu'on peut noter : $CE = f_{CE}(SP, PC, TE, \varepsilon_{CE})$. À la suite de nombreux auteurs, soulignons que le signe « = » qui figure dans une équation de ce type n'a pas sa signification usuelle, puisqu'il véhicule non seulement l'affirmation selon laquelle deux grandeurs sont égales, mais encore l'idée selon laquelle les variables qui figurent à sa droite sont des *causes* directes de celle qui figure à sa gauche. Pour reprendre les termes de Pearl, « dans les modèles structurels, le signe d'égalité véhicule la relation asymétrique "est déterminé par" et, dès lors, se comporte plus comme un symbole d'assignation (:=) dans les langages de programmation que comme une égalité algébrique » (Pearl, 2000, p. 211). En ce sens, on parle d'« équations structurelles ».

Outre un graphe causal et/ou un ensemble d'équations structurelles lui correspondant, un modèle structurel se compose d'un ensemble d'hypothèses relatives aux individus de l'échantillon pour lequel on recueille les valeurs des variables observables du modèle, aux variables du modèle elles-mêmes, et enfin à la forme du modèle. Ce sont ces hypothèses qui

permettent d'utiliser les données disponibles pour quantifier le modèle et, au-delà, de confronter à ces données l'hypothèse causale représentée par le modèle et, plus généralement, de tester cette hypothèse.

La quantification d'un modèle structurel se manifeste d'abord sous la forme de coefficients qui sont associés aux flèches des graphes causaux, chacun mesurant l'influence causale qui s'exerce selon la flèche à laquelle il est associé. Quoi qu'il en soit des hypothèses relatives aux variables et aux équations d'un modèle structurel, cette quantification ne peut être correcte qu'à certaines conditions concernant les individus pour lesquels on va recueillir la valeur des variables du modèle. Ces individus doivent être comme tirés au sort à l'occasion de tirages indépendants : les observations sont indépendantes. En outre, les variables individuelles correspondant aux variables du modèle doivent être identiques entre elles et être distribuées comme les variables du modèle. Enfin, ces individus doivent être nombreux – en d'autres termes, l'échantillon doit être suffisamment vaste. Nous ne nous attardons pas ici sur ces hypothèses qui, implicitement ou explicitement, sont toujours sous-jacentes à la modélisation causale. Il convient toutefois de réaliser qu'elles sont contraignantes. Pour une explicitation de ces hypothèses (et aussi pour la distinction de plusieurs types d'hypothèses dans les modèles structurels), nous renvoyons le lecteur intéressé à Freedman, 1987.

Contrairement aux hypothèses relatives aux individus de l'échantillon qu'on considère, les hypothèses portant sur les variables et sur la forme du modèle ne sont pas les mêmes pour tous les modèles structurels. Nous nous en tenons ici à un cas simple parmi ceux que permet d'aborder la modélisation structurelle : le cas où 1) les variables du modèle structurel relèvent toutes du même niveau de réalité, 2) la valeur de chacune est observable, et 3) l'ensemble des variables du modèle est causalement suffisant, c'est-à-dire que toutes les variables qui sont des causes communes à au moins deux variables du modèle sont elles-mêmes des variables du modèle. Intégrer à notre analyse des cas plus complexes la rendrait plus technique et ne l'enrichirait pas significativement : se limiter aux cas simples suffit à apporter une réponse satisfaisante aux questions philosophiques qui nous intéressent, au niveau de généralité où nous les abordons. En particulier, cela autorise la confrontation de l'inférence hypothético-déductive, d'une part, aux théories probabilistes de la causalité et, d'autre part, à ces méthodes d'inférence causale probabiliste récemment développées dont les partisans soutiennent qu'elles sont inductives. Un même souci de limiter les considérations techniques à ce qui est requis pour notre argument philosophique nous conduit à ne lier notre analyse ni à certains jeux d'hypothèses statistiques concernant le modèle, ni à un statut relativement à la récursivité. En effet, en même temps que les hypothèses statistiques et le statut du modèle relativement à la récursivité, varient les outils statistiques autorisés et, spécifiquement, qu'on peut mobiliser pour mettre en œuvre la procédure d'inférence causale

que nous nous apprêtons à définir. Cette procédure est donc également valide dans tous les cas de figure concernant les hypothèses statistiques et concernant le caractère récursif ou non du modèle – ou, pour le dire autrement, concernant le caractère acyclique ou non du graphe causal correspondant au modèle.

Si la procédure définie dans la section 2.2 l'est à un haut niveau de généralité, il nous semble néanmoins utile de présenter dans un plus grand détail technique certains modèles structurels : les modèles de cheminement. Cette démarche admet deux justifications, qui ne sont pas indépendantes l'une de l'autre. D'une part, une présentation de l'analyse de cheminement devrait contribuer à rendre notre propos plus concret. D'autre part, les modèles de cheminement sont les plus simples parmi les modèles structurels, l'analyse de cheminement qui les mobilise se trouve au fondement de l'ensemble des méthodes de la modélisation structurelle et, en conséquence, comprendre son fonctionnement et son principe permet de se repérer dans pratiquement toutes les études relevant de ce domaine.

2.1.2 L'analyse de cheminement*

L'analyse de cheminement (également appelée « analyse de dépendance », ou « analyse en pistes causales ») a été introduite par Sewall Wright dès les années 1920, Wright, 1921 et Wright, 1934 étant les articles fondateurs. Si Wright s'intéressait à des questions de génétique, l'économétrie a également vu émerger de précoces et importantes contributions à la méthode, dues en particulier à Haavelmo (Haavelmo, 1943). Comme telle (c'est-à-dire indépendamment du fait qu'elle sert de fondement à la modélisation structurelle tout entière), l'analyse de cheminement continue d'être utilisée dans le cadre de certaines études relevant de la modélisation structurelle. On peut mentionner ici Hope, 1980[5] ou, plus récemment, Meuret et Morlaix, 2006, que nous avons évoqué plus haut.

Parmi les modèles structurels, les modèles de cheminement sont caractérisés par leur récursivité, ainsi que par les hypothèses statistiques suivantes :

1. linéarité : les équations structurelles sont linéaires, c'est-à-dire que la variable qui figure à gauche du signe d'égalité est la somme de multiples des variables qui figurent à droite du signe d'égalité ;
2. standardisation : les variables du modèle ont une espérance nulle et une variance égale à un ;
3. indépendance des termes d'erreur entre eux, et indépendance de chacun des termes d'erreur avec toutes les variables du modèle

[5] Cet article est pris comme exemple et discuté dans Freedman, 1987.

différentes de celle à laquelle il est attaché. Cette hypothèse implique que l'ensemble de variables choisi pour représenter le système considéré est causalement suffisant ;

4. espérance nulle des termes d'erreur ;
5. variance finie et uniforme (relativement aux différentes valeurs des variables du modèle) pour les termes d'erreur.

L'hypothèse 1 porte sur la forme des équations structurelles. Les hypothèses 2 à 5, sur les variables qui figurent dans ces équations, qu'il s'agisse des variables utilisées pour représenter le système considéré ou des termes d'erreur, et sur le rapport qu'elles entretiennent.

À titre d'illustration, reprenons la version simplifiée du modèle développé par Meuret et Morlaix pour représenter l'influence du statut professionnel du père sur la compréhension de l'écrit. Rappelons que dans cette version simplifiée, l'hypothèse causale de Meuret et Morlaix est représentée par le graphe de la figure 2.1. Ainsi que nous l'avons déjà indiqué, ce graphe est acyclique. Les modèles dont il participe sont donc récursifs.

Supposons maintenant avec Meuret et Morlaix que le modèle structurel correspondant à ce graphe causal est un modèle de cheminement. Pour le dire autrement, admettons que les hypothèses 1 à 5 sont satisfaites – ce que les deux auteurs ne se soucient d'ailleurs jamais de justifier. Dans ces conditions, et exactement en vertu de l'hypothèse 1, les équations structurelles associées au modèle peuvent s'écrire de la façon suivante :

$$SP = \varepsilon_{SP} \tag{2.1}$$

$$PC = a'SP + \varepsilon_{PC} \tag{2.2}$$

$$TE = a''SP + \varepsilon_{TE} \tag{2.3}$$

$$CE = aSP + b'PC + b''TE + \varepsilon_{CE} \tag{2.4}$$

a, a', a'', b' et b'' sont les coefficients associés aux différentes relations de cause à effet dont l'existence est postulée par l'hypothèse que le modèle représente. Ils quantifient ces relations. Une représentation usuelle est donnée par la figure 2.2.

Sous les hypothèses 1 à 4, les coefficients a, a', a'', b' et b'' peuvent être estimés par régression linéaire multiple. En outre, ils permettent de composer des coefficients plus complexes, qui quantifient non plus les relations de cause à effet directes que représentent les flèches du graphe, mais les influences causales indirectes qui correspondent aux chemins indirects du graphe. Par exemple, l'influence de SP sur CE suivant le facteur externe PC correspond au chemin indirect qui mène de SP à CE en passant par PC, et le coefficient qui lui est associé vaut $a'.b'$.

Pour le cas que nous considérons, l'idée qu'on trouve au fondement de l'analyse de cheminement peut s'exprimer de la façon suivante : l'effet

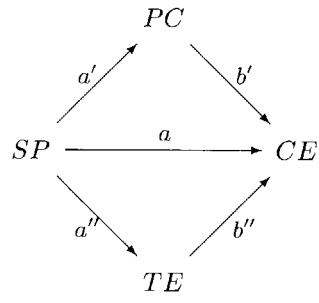


FIGURE 2.2

total de SP sur CE est la somme des coefficients correspondant chacun à un chemin, direct ou indirect, menant de SP à CE dans le graphe causal. En d'autres termes, l'effet total de SP sur CE est égal à $a + a'.b' + a''.b''$. Plus exactement, l'effet total de SP sur CE tel qu'il est mesuré par le coefficient de corrélation $r_{SP.CE}$ est égal à cette somme :

$$r_{SP.CE} = a + a'.b' + a''.b''. \tag{2.5}$$

Cette égalité est une conséquence directe du résultat plus général établi par Wright (Wright, 1921, p. 568) :

Théorème 2.1 (Règle du tracé (*tracing rule*)) *Le coefficient de corrélation entre deux variables d'un modèle de cheminement est égal à la somme des coefficients associés aux différents chemins, directs ou indirects, qui mènent de l'une à l'autre de ces variables.*

L'intérêt principal de l'analyse de cheminement telle qu'elle s'appuie sur ce théorème est le suivant : pour un modèle de cheminement donné, il est possible d'écrire un certain nombre d'équations, et ces équations permettent en particulier d'inférer les coefficients de cheminement à partir des coefficients de corrélation entre les variables du modèle (dont la valeur peut être estimée). En effet, la récursivité des modèles de cheminement implique que le système composé de ces équations, et dont les inconnues sont les coefficients de cheminement, admet une solution unique (Kline, 1998, p. 105 et 107).

Dans ces conditions, l'analyse de cheminement se présente d'abord, pour reprendre les termes de Wright, comme « une méthode pour mesurer l'influence directe le long de chaque chemin séparé dans un tel système, et donc pour trouver le degré auquel la variation d'un effet donné est déterminée par chaque cause particulière » (Wright, 1921, p. 557). Ce point est également souligné par Freedman (Freedman, 1987, p. 112). Toutefois, Wright signale dès le début du texte introduisant l'analyse de cheminement qu'un autre usage de cette méthode est envisageable :

« Dans le cas où les relations causales sont incertaines, la méthode peut être utilisée afin de trouver les conséquences logiques de toute hypothèse particulière concernant ces relations » (Wright, 1921, p. 557). Autrement dit, Wright envisage explicitement la possibilité d'utiliser les modèles de cheminement dans le cadre d'une démarche hypothético-déductive, qui pourrait viser à identifier les relations de cause à effet entre les variables du modèle. Le but de la prochaine section est de donner corps à ce qui reste chez Wright à l'état de suggestion.

2.2 Une procédure d'inférence causale probabiliste hypothético-déductive

2.2.1 Hypothético-déduction

Afin de définir une procédure d'inférence causale probabiliste hypothético-déductive, nous commençons par rappeler ce qu'il faut entendre, précisément, par « hypothético-déduction ». Une caractérisation classique est donnée par Popper (Popper, 1934, p. 28-29) :

> « La méthode qui consiste à mettre les théories à l'épreuve dans un esprit critique et à les sélectionner conformément aux résultats des tests, suit toujours la même démarche : en partant d'une nouvelle idée, avancée à titre d'essai et nullement justifiée à ce stade – qui peut être une prévision, une hypothèse, un système théorique ou tout ce que vous voulez –, l'on tire par une déduction logique des conclusions. L'on compare alors ces conclusions les unes aux autres et à d'autres énoncés relatifs à la question de manière à trouver les relations logiques (telles l'équivalence, la déductibilité, la compatibilité ou l'incompatibilité) qui les unissent. »

En suivant cet extrait, on peut considérer que les inférences hypothético-déductives se déroulent en quatre temps :

i) une hypothèse est formulée ;

ii) des conséquences en sont tirées ;

iii) l'hypothèse est mise à l'épreuve sur la base de ces conséquences ;

iv) l'hypothèse est rejetée ou elle est renforcée – qu'on la considère confirmée ou seulement corroborée.

C'est ce schéma que nous prenons ici comme définition de l'hypothético-déduction. Cela signifie que, ainsi que nous l'avons déjà annoncé, nous nous situons à un niveau d'abstraction élevé. En particulier, nous considérons que le caractère hypothético-déductif d'une inférence est indépendant de la nature des moyens mobilisés pour mener à bien les étapes i à iv.

Notons toutefois que, pour ce qui est de la mise à l'épreuve de l'hypothèse (à l'étape iii), Popper distingue en particulier deux choses : d'une

part, les tests portant sur les conséquences tirées de l'hypothèse (qui peuvent prendre différentes formes) ; d'autre part, « la comparaison de la théorie [c'est-à-dire de l'hypothèse] à d'autres théories, dans le but principal de déterminer si elle constituerait un progrès scientifique au cas où elle survivrait à nos divers tests » (Popper, 1934, p. 29). Autrement dit, les conséquences tirées de la théorie peuvent non seulement servir à tester la théorie de manière isolée, mais encore être comparées aux conséquences tirées de théories concurrentes en vue de déterminer si la théorie discutée est préférable à ces théories concurrentes. Il nous semble clair que la logique de l'hypothético-déduction n'impose pas un ordre dans lequel ces différents examens devraient avoir lieu.

2.2.2 Schéma d'inférence causale probabiliste hypothético-déductive

Étant donné un ensemble de variables $\mathbf{V}$ et des données probabilistes relatives à $\mathbf{V}$, les étapes suivantes définissent une procédure d'inférence causale probabiliste hypothético-déductive mobilisant les outils et les techniques de la modélisation structurelle :
Étape A : Spécifier un modèle structurel M.
Étape B : Estimer M, c'est-à-dire estimer la valeur des paramètres qui sont associés aux différentes relations de cause à effet représentées par M, et mesurer son adéquation (*fit*) aux données.
Étape C : Tester M et décider s'il doit être rejeté.
Étape D : Réitérer les étapes A à C pour des modèles différents de M.
Étape E : Identifier celui des modèles non rejetés à l'issue de l'étape C qui a la meilleure adéquation aux données, et noter $M*$ ce modèle.
Étape F : Engendrer des modèles équivalents à $M*$ et identifier, parmi les modèles engendrés, celui dont il est le plus plausible qu'il représente adéquatement la structure causale sur $\mathbf{V}$, et noter $MI_{\mathbf{V}}$ ce modèle.
$MI_{\mathbf{V}}$ est le résultat de la procédure d'inférence causale.

Sans entrer dans un détail de nature technique, il convient d'expliciter ici certains termes :
Spécifier un modèle structurel, c'est définir un graphe orienté dont on considère qu'il pourrait représenter adéquatement les relations de cause à effet directes au sein de l'ensemble de variables considéré.

Estimer un modèle structurel, c'est attribuer, sur la base des données disponibles, une valeur aux coefficients qui mesurent l'effet de la variation de la valeur d'une cause supposée sur la valeur de l'un de ses effets supposés. Dans le cas linéaire et où les variables sont continues quantitatives, cela revient à estimer la différence, sur la valeur de l'effet, d'une augmentation d'une unité de la valeur de la cause.

Mesurer l'adéquation d'un modèle aux données, c'est évaluer le degré auquel le modèle est capable de reproduire les données. Le fait que l'adéquation vient par degré ou, pour le dire autrement, le caractère

gradué de l'adéquation rend possible la comparaison des modèles entre eux sur le critère de leur adéquation aux données (à l'étape E).

Tester un modèle, c'est déterminer si ce modèle reste plausible après que ses paramètres ont été estimés. Plus précisément, c'est répondre à la question (fermée) de savoir si l'hypothèse selon laquelle le modèle envisagé représente adéquatement les relations de cause à effet sur l'ensemble de variables considéré ne se révèle pas incohérente à la lumière de l'estimation des paramètres causaux.

Engendrer un modèle équivalent à un modèle donné, c'est construire un modèle différent du modèle initial mais qui prédit les mêmes corrélations que lui. Des modèles équivalents ont la même adéquation aux données, pour toutes les mesures d'adéquation.

Ces précisions étant données, il nous est possible d'expliquer en quoi la procédure proposée peut être considérée comme conforme au canon de l'hypothético-déduction. Clairement, l'étape A est le moment où est formulée une hypothèse. L'étape B est le moment où des conséquences sont tirées de cette hypothèse : les valeurs estimées des différents paramètres ainsi que son degré d'adéquation aux données sont impliqués par le modèle spécifié à l'étape A, les hypothèses statistiques émises et les données disponibles. L'étape C est celui où ces conséquences sont utilisées pour tester l'hypothèse de manière isolée.

Cela signifie que, à ce point, le rapport entre l'hypothèse causale formulée à l'étape A et les données disponibles n'est pas un rapport de confrontation qui serait susceptible de conduire à rejeter l'hypothèse. Plutôt, les données sont utilisées en conjonction avec l'hypothèse en vue de lui donner une forme testable. Pour cette raison, la procédure que nous venons de décrire peut sembler relever du *bootstrap* tel qu'il est défini dans Glymour, 1980, plutôt que de l'hypothético-déduction. D'un côté, en effet, Glymour propose de considérer que des données confirment une hypothèse théorique si ces données et des hypothèses auxiliaires impliquent conjointement une instanciation de l'hypothèse à confirmer. De l'autre côté, on considère généralement que la place naturelle des données dans l'hypothético-déduction est celle d'un vis-à-vis auquel confronter des conséquences observables déduites de l'hypothèse théorique qu'il s'agit de discuter. Cependant, il nous semble que rien dans la définition abstraite de l'hypothético-déduction que nous avons citée et sur laquelle nous nous appuyons ici n'impose de donner cette place particulière aux données. En outre, la stratégie définie par Glymour est plus généralement la suivante : utiliser les données en conjonction avec une partie de l'hypothèse théorique discutée pour tester une autre partie de cette hypothèse et, en renouvelant cette opération pour des parties différentes de l'hypothèse, « se sortir » des difficultés associées à la confirmation d'hypothèses qui sont théoriques de part en part. Or, ainsi précisée, cette stratégie apparaît clairement comme non implémentée

dans la procédure d'inférence causale probabiliste que nous avons définie. Enfin, positivement, nous verrons plus bas que la modélisation structurelle offre des outils et méthodes permettant effectivement de tester un modèle causal hypothétique sur la base des conséquences qui peuvent être tirées de ce modèle en conjonction avec les données.

Une fois que le modèle a été testé (à l'étape C), il est rejeté si et seulement si ces conséquences sont, d'une façon ou d'une autre, considérées comme problématiques en regard des informations disponibles. Puis, seulement ensuite et d'abord à l'étape E, a lieu la comparaison entre hypothèses dont nous avons vu qu'elle est explicitement prescrite par Popper. Plus précisément, la comparaison est alors entre les modèles qui n'ont pas été rejetés à l'issue des tests individuels menés à l'étape C. Mener la comparaison entre modèles *après* que les modèles ont été testés isolément permet de comparer entre eux seulement des candidats déjà sérieux au titre de modèle adéquat. Ces candidats sérieux sont comparés sur le critère de leur adéquation aux données. Plus exactement, on retient à l'issue de l'étape E celui parmi les modèles comparés qui a la meilleure adéquation aux données. Une fois ce modèle identifié, on considère l'ensemble des modèles qui lui sont équivalents (lui-même appartenant à cet ensemble) et on retient le plus plausible d'entre eux.

Les étapes E et F prises ensemble constituent une inférence à la meilleure explication au sens où Harman définit cette notion : « On infère, de la prémisse qu'une hypothèse donnée fournirait une "meilleure" explication que n'importe quelle autre hypothèse, la conclusion selon laquelle l'hypothèse donnée est vraie » (Harman, 1965, p. 89). D'une part, en effet, l'adéquation d'un modèle aux données est généralement considérée comme sa capacité à expliquer les données. L'étape E vise donc à choisir celle qui *explique le mieux* les données parmi les hypothèses formulées (à l'étape A) et qui n'ont pas été rejetées (à l'issue de l'étape C). D'autre part, l'étape F vise à identifier celle qui est la plus plausible, et en ce sens *la meilleure*, parmi des hypothèses maximalement et également explicatives. Les critères conduisant ici à préférer $MI_{\mathbf{V}}$ à des modèles concurrents sont des critères théoriques, puisque les modèles envisagés sont tous équivalents et ont donc tous le même score pour toutes les mesures d'adéquation aux données. La simplicité ou la bonne articulation avec des modèles acceptés pour d'autres phénomènes ou avec des théories plus générales pourraient être citées ici, mais le détail de ce qui conduit à isoler $MI_{\mathbf{V}}$ dépasse le cadre de notre discussion présente. Surtout, nous considérons que le caractère ampliatif de la séquence constituée des étapes E et F n'est pas incompatible avec l'idée selon laquelle la procédure que nous avons définie serait hypothético-déductive au sens abstrait où nous l'entendons ici. Dans le contexte d'inférence causale probabiliste menée dans le cadre de la modélisation structurelle, cette séquence ampliative correspond à la façon dont des modèles sont généralement comparés, et

elle est la seule façon dont nous pouvons imaginer comparer des modèles sur la base des conséquences tirées à l'étape B.

2.2.3 Détails et précisions*

Nous revenons maintenant sur les étapes de la procédure que nous venons de définir et, pour chacune, nous apportons quelques détails et précisions dont il ne nous a pas semblé nécessaire de faire état au moment d'offrir un schéma de cette procédure. Indiquons d'emblée qu'il ne s'agit pas pour nous ici de proposer un développement portant sur les statistiques et qui consisterait, par exemple, à présenter une ou plusieurs théories du test d'hypothèses statistiques. Plutôt, nous expliquons, pour chaque étape, en quoi elle consiste plus précisément. Cette attention plus précise ne signifie pas non plus que nous prenions pour objet la pratique de l'inférence causale. Pour le dire autrement, nous continuons de concentrer notre attention sur les principes de l'inférence causale rigoureuse. En outre, nous ne portons pas d'intérêt systématique et poussé aux techniques qui doivent être utilisées afin de mener à bien les tâches définies et présentées plus en détail ici. Ces techniques dépendent des hypothèses statistiques qui composent le modèle structurel considéré et/ou du statut du modèle relativement à la récursivité.

Pour ce qui est, d'abord, de l'étape A, soulignons que le modèle spécifié doit être sur-identifié (*over-identified*). Un modèle est sur-identifié si et seulement s'il a des degrés de liberté, c'est-à-dire si et seulement si le nombre de ses paramètres est inférieur au nombre d'observations autorisées par $\mathbf{V}$. Dans ce contexte théorique, une observation, ou plus précisément une « observation autorisée par $\mathbf{V}$ » est soit la variance d'une variable de $\mathbf{V}$, soit la covariance entre deux variables de $\mathbf{V}$. Le nombre d'observations autorisées par $\mathbf{V}$ est donc égal à $\|\mathbf{V}\|(\|\mathbf{V}\|+1)/2$. Pour ce qui est d'autres notions d'identification et pour des critères d'identification, nous renvoyons le lecteur à Kline, 1998, p. 105-110.

La première raison pour laquelle il convient de spécifier un modèle sur-identifié est la suivante : pour un modèle sur-identifié, il est théoriquement possible de dériver une estimation unique de chacun des paramètres causaux. Il existe deux options principales pour l'estimation des paramètres causaux d'un modèle :

- l'estimation par régression multiple. Le principe est ici le suivant : pour chaque variable V, on considère les relations pour lesquelles elle est effet et on attribue aux coefficients associés à ces relations la valeur qui minimise la distance entre les valeurs de V qu'on observe et les valeurs de V que le modèle prédit ;
- l'estimation par maximum de vraisemblance. Il s'agit alors de maximiser la vraisemblance de l'hypothèse selon laquelle les observations données sont tirées de la population considérée.

Le choix de l'une ou de l'autre de ces options dépend, en particulier, des hypothèses qu'on émet à propos du modèle.

La seconde raison pour laquelle il convient de spécifier un modèle sur-identifié est que, précisément, pour un tel modèle il faut recourir à des critères statistiques, tels que la minimisation de la somme des carrés des différences aux observations ou la maximisation de la vraisemblance du modèle, pour parvenir à une estimation des coefficients. En effet, un modèle sur-identifié est un modèle pour lequel les observations autorisées sur-déterminent la valeur des coefficients, et ce n'est qu'en se donnant un critère de ce type qu'on peut arrêter *une* estimation des paramètres (Kline, 1998, p. 108-110). Une telle estimation ne découle pas logiquement de la valeur des observations et, pour cette raison, elle ouvre des possibilités pour tester le modèle causal hypothétique considéré.

Plus précisément, nous identifions dans la littérature méthodologique portant sur la modélisation causale trois types de tests qui peuvent (et donc, en un sens, devraient) être menés à l'étape C de la procédure décrite plus haut. Ces tests consistent respectivement à :

a. s'assurer que les signes et valeurs absolues des estimations obtenues pour les paramètres sont plausibles. Cette vérification doit être à la fois locale et globale. Localement, il s'agit de vérifier que le signe et la valeur absolue de l'estimation de chaque paramètre fait sens. En particulier, chacun de ces paramètres doit être significativement différent de zéro. En effet, si tel n'était pas le cas, alors le modèle, précisément parce qu'il postule l'existence de relations de cause à effet auxquelles les paramètres sont associés, devrait être rejeté. De façon générale, les conséquences des examens locaux aussi bien que globaux menés à ce point portent toujours sur le modèle dans son ensemble, qui est rejeté ou non ;

b. calculer les résidus de corrélation – c'est-à-dire les différences entre les corrélations impliquées par le modèle et les corrélations observées – et vérifier qu'aucun n'a une valeur absolue supérieure à 0,1[6]. Les corrélations impliquées par le modèle sont calculées, étant donné la forme du modèle, à partir des coefficients associés aux différentes relations de cause à effet. Il est en outre possible, pour certaines classes de modèles structurels, de lire sur le graphe causal quelles sont les corrélations partielles dont la forme du modèle implique qu'elles sont nulles (voir en particulier Pearl, 2000, p. 140-144). Dans ces cas, on peut également calculer des résidus de corrélation partielle. Les corrélations impliquées par le modèle peuvent différer des corrélations observées seulement pour les modèles sur-identifiés, pour lesquels nous avons vu que la valeur des coefficients estimés ne découle pas logiquement de la forme du modèle et des valeurs

[6] Il s'agit d'une valeur conventionnelle mais généralement acceptée.

des observations. Les résidus de corrélation devraient être nuls si le modèle considéré est correct ;

c. calculer les restrictions de sur-identification (*over-identification restrictions*) – c'est-à-dire la différence entre deux estimations différentes des mêmes paramètres structurels – et vérifier que l'hypothèse selon laquelle elles sont nulles ne peut pas être rejetée. L'idée est ici la suivante : pour des modèles sur-identifiés, il est parfois possible d'estimer un même paramètre de plusieurs façons différentes, qui correspondent à différents ensembles d'équations identifiés dans lesquels le paramètre apparaît. Si le modèle est correct, ces différentes méthodes doivent donner des résultats identiques. Autrement dit, si ces méthodes donnent des résultats différents (ou, plus exactement, si l'on peut rejeter l'hypothèse selon laquelle ils donnent des résultats identiques), alors le modèle peut être rejeté.

Concernant, maintenant, l'adéquation d'un modèle causal aux données disponibles, il en existe plusieurs mesures, qui portent sur des aspects différents du rapport entre les données et le modèle considéré. Plusieurs de ces grandeurs peuvent être calculées à l'étape E de la procédure que nous avons définie. Le « chi-deux » d'un modèle est sans doute la plus fondamentale d'entre elles, au sens où les mesures plus complexes mobilisent presque toujours cette grandeur. Elle représente ce que le modèle explique des corrélations entre les variables considérées. On notera que mesurer l'adéquation d'un modèle n'a de sens que si ce modèle est sur-identifié ; dans le cas contraire, le modèle estimé ne peut qu'être complètement adéquat aux données.

Pour finir, signalons l'existence d'algorithmes qui permettent d'engendrer des modèles équivalents à un modèle donné. Une présentation d'ordre général en est proposée par Kline (Kline, 1998, p. 153-156). Pour certaines classes de modèles structurels, Pearl indique comment construire des modèles équivalents à un modèle donné en pratiquant des manipulations simples de son graphe causal (Pearl, 2000, p. 146-148).

2.2.4 Mise en œuvre

Il nous reste à évoquer la mise en œuvre de la procédure que nous avons définie. Encore une fois, cela ne signifie ni que nous en venions au détail des techniques statistiques appelées par la procédure définie, ni que nous nous intéressions, finalement, à la pratique de l'inférence causale. La question qui nous intéresse ici est plutôt celle de la part qui peut être prise par des programmes informatiques dans la mise en œuvre de la procédure – ou, en d'autres termes, de la possibilité ou non d'automatiser l'inférence causale si elle se conforme à la procédure que nous avons définie. Nous abordons cette question en grande partie parce qu'elle se révélera décisive quand nous examinerons les méthodes d'inférence causale probabiliste plus récemment développées.

La plupart des tâches qui composent la procédure d'inférence causale que nous avons définie sont généralement menées à bien en recourant à des programmes informatiques. C'est le cas, en particulier, pour l'estimation des paramètres d'un modèle à partir de données statistiques, pour la plupart des vérifications et des calculs qu'il convient de mener afin de tester un modèle, et pour l'évaluation de l'adéquation d'un modèle aux données sous différentes mesures. Dans ces conditions, la procédure d'inférence causale dans son ensemble peut être menée à bien étape après étape en recourant à un programme convenable à chaque étape qui le requiert.

Même si elle comprend l'utilisation de programmes informatiques, la façon dont nous venons d'envisager la mise en œuvre de la procédure que nous avons définie n'est pas automatique. En effet, le scientifique intervient à chacune des étapes. Ce mode de mise en œuvre est celui auquel conduit naturellement le caractère hypothético-déductif de la procédure définie – et en particulier l'importance revêtue par la spécification de modèles. Nous souhaitons toutefois donner l'intuition de ce qu'automatiser la recherche des causes peut vouloir dire dans le contexte de la modélisation structurelle. Positivement, dans ce contexte, l'automatisation de la recherche des causes consiste essentiellement à automatiser l'étape A de la procédure que nous avons définie, c'est-à-dire à engendrer automatiquement et successivement des modèles structurels hypothétiques. Plus précisément, la recherche automatique des causes relevant de la modélisation structurelle est généralement menée de la façon suivante : étant donné le modèle vide (dans lequel ne figure aucune flèche) sur l'ensemble de variables considéré, on ajoute la flèche dont la présence augmente le plus l'adéquation du modèle aux données, puis la flèche dont la présence – en plus de celle de la première – augmente le plus l'adéquation du modèle aux données, et ainsi de suite jusqu'à ce qu'un critère d'arrêt de la procédure soit atteint. Il existe des ensembles (*packages*) de programmes qui intègrent des programmes réalisant ces tâches. À ces ensembles appartiennent également des programmes menant à bien les tâches classiques dans le domaine de la modélisation causale, et donc en particulier celles qui ont à être menées dans le cadre de la procédure que nous avons définie. LISREL constitue la plus ancienne et la plus connue des familles de tels ensembles de programmes. Les ensembles de programmes LISREL sont modulaires : pour chaque tâche, l'utilisateur peut choisir la façon dont elle devra être réalisée.

2.3 Caractéristiques et limites de l'inférence causale probabiliste hypothético-déductive

2.3.1 Possibilité de l'inférence causale

La voie hypothético-déductive permet de contourner les obstacles qui s'opposent à ce que les théories probabilistes soient directement utilisées comme principes pour l'inférence causale. Autrement dit, la procédure que nous avons définie est telle que, si on l'adopte, on ne rencontre aucune des deux difficultés dont nous avons discutées dans la sous-section 1.5.1, et la raison en est que cette procédure est hypothético-déductive.

En premier lieu, la procédure n'a pas la conséquence qu'il serait nécessaire de connaître toutes les causes de B qui sont indépendantes de A afin de déterminer si A cause B. En effet, les relations de cause à effet qui sont représentées par le graphe spécifié à l'étape A sont supposées, à titre d'hypothèses, mais elles n'ont pas à être connues au sens strict. En second lieu, le problème de la détermination de l'ensemble au sein duquel les causes d'une variable donnée doivent être cherchées ne se pose pas. Plus exactement, le problème est déjà résolu au moment où la procédure commence, puisque l'ensemble de variables $\mathbf{V}$ pour lequel un modèle causal est spécifié à l'étape A est déjà défini. Corrélativement, la procédure – en particulier, les relations de cause à effet qu'elle permet de mettre au jour – est relative à un ensemble de variables.

Il semble donc clair que, si la voie hypothético-déductive permet effectivement de contourner les obstacles mis en évidence et discutés dans la sous-section 1.5.5, ce contournement doit avoir un prix. La section qui commence vise essentiellement à l'évaluer. Nous nous y arrêterons successivement sur l'objet visé par l'inférence causale (2.3.2), sur les conséquences de l'hypothético-déductivité en matière d'inférence causale probabiliste (sous-section 2.3.3), et sur le recours aux tests d'hypothèses statistiques (sous-section 2.3.4).

2.3.2 Objet de l'inférence causale probabiliste hypothético-déductive

Le contournement des obstacles qui surgissent quand on souhaite s'appuyer sur les théories probabilistes de la causalité afin d'induire des relations de cause à effet correspond d'abord à une différence sensible entre l'objet visé par l'inférence causale et l'objet des théories probabilistes de la causalité. Plus précisément, les théories probabilistes présentées dans le chapitre 1 analysent un concept absolu de causalité entre propriétés. À l'inverse, la procédure que nous venons de définir – et, plus généralement, la modélisation structurelle – vise des relations de cause à effet entre variables qui sont relatives à un ensemble de variables préalablement arrêté (l'ensemble que nous avons généralement noté $\mathbf{V}$). Nous avons déjà indiqué le rôle joué par ceci que les relations de cause à effet visées par l'inférence hypothético-déductive sont relatives à un ensemble

de variables. Il permet de contourner la difficulté correspondant, du côté des théories probabilistes, à l'absence de principe de clôture de l'espace au sein duquel les causes d'une propriété donnée doivent être cherchées. Dans la sous-section qui commence, nous cherchons plutôt à expliquer en quoi le fait que l'objet de l'inférence causale probabiliste hypothético-déductive, soit la causalité 1) entre variables et 2) relativement à un ensemble de variables, peut être considéré comme une partie du prix à payer pour que l'inférence causale soit possible. Plus explicitement, nous montrons que chacune de ces caractéristiques contribue à rendre l'objet de l'inférence plus grossier que la notion de causalité visée par les théories probabilistes.

Causalité entre variables Une première différence entre la relation qui est visée par l'inférence causale probabiliste hypothético-déductive et la relation que les théories probabilistes visent à analyser concerne la nature des *relata* de la causalité, c'est-à-dire le type des entités qui sont susceptibles d'entrer dans des relations de cause à effet. D'un côté, il s'agit de variables ; de l'autre côté, il s'agit de propriétés. Or, s'il reste possible de discuter de la nature des *relata* de la causalité générique, aucun métaphysicien ne propose de considérer que ces *relata* sont des variables. Aussi, notre première tâche ici consiste à comprendre d'où viennent les variables de la modélisation causale et quels rapports elles entretiennent avec les propriétés qui sont généralement considérées comme les *relata* de la causalité générique.

On trouve dans un ouvrage récent de Williamson une défense de l'idée selon laquelle les variables permettent de représenter, d'une part, des événements singuliers et, d'autre part, des propriétés (Williamson, 2005, p. 50) :

> « Il semble que ce soit une idéalisation sans conséquence que d'interpréter la causalité comme une relation entre variables – on peut penser un événement comme une variable binaire singulière qui prend une valeur s'il advient et l'autre valeur s'il n'advient pas, on peut penser une propriété comme une variable binaire répétable qui prend une valeur quand elle est instanciée et l'autre quand elle n'est pas instanciée, etc. – et une telle idéalisation entre rarement en conflit avec les intuitions causales. »

Seule la seconde partie de l'analyse de Williamson, portant sur la causalité générique, nous intéresse ici. Nous considérons qu'elle est correcte. Plus précisément, la proposition formulée (« penser une propriété comme une variable binaire répétable qui prend une valeur quand elle est instanciée et l'autre quand elle n'est pas instanciée ») est générale : pour toute propriété, elle indique quelle variable peut la représenter. Cette variable est binaire, prenant par exemple la valeur 1 quand la propriété est

instanciée et la valeur 0 sinon. Ainsi, la représentation des propriétés au moyen de variables correctement définies repose sur ceci que, pour toute propriété et pour tout individu d'une population donnée, cette propriété ou bien est instanciée, ou bien n'est pas instanciée par cet individu.

Il apparaît, en outre, que les variables binaires représentant des propriétés selon ces modalités constituent un type particulier, limite, de variables. Pour comprendre ce dernier point, considérons à nouveau la propriété d'être fumeur. Un individu échoue à instancier cette propriété si et seulement s'il instancie la propriété de ne pas être fumeur. L'ensemble de propriétés {être fumeur, ne pas être fumeur} est donc caractérisé par ceci que tout individu – plus précisément, tout individu appartenant à la population étudiée – instancie exactement une des propriétés de l'ensemble. En d'autres termes (et si la population considérée compte au moins un individu fumeur et au moins un individu non fumeur), l'ensemble {être fumeur, ne pas être fumeur} définit une partition sur l'ensemble des individus considérés. Or, cette caractéristique n'appartient pas exclusivement à des ensembles de deux propriétés telles que l'une est définie comme la négation de l'autre. De nombreux ensembles de propriétés la possèdent, qui peuvent compter n'importe quel nombre de propriétés. Parmi ces ensembles, on trouve en particulier des ensembles de propriétés dont la définition implique une valeur numérique – par exemple, {avoir 20 ans ou moins, avoir entre 21 et 40 ans, avoir entre 41 et 60 ans, avoir 61 ans ou plus}. De même que les ensembles du type {être fumeur, ne pas être fumeur} et pour les mêmes raisons, ces ensembles sont représentés de manière adéquate par des variables. Ce sont d'ailleurs des ensembles de plus de deux propriétés qui sont le plus généralement représentés par les variables de la modélisation structurelle. Finalement, il nous semble qu'on peut étendre cette analyse des variables discrètes aux variables continues, et considérer qu'une variable continue représente un ensemble infini de propriétés.

Nous venons de montrer comment tout ensemble de propriétés tel que tout individu de la population considérée instancie exactement une de ces propriétés peut être représenté au moyen d'une variable. Ce résultat, toutefois, peut être présenté sous un jour un peu différent. On insistera alors sur ceci que les variables ne représentent pas des propriétés, mais des *ensembles de* propriétés. Dans ces conditions, une flèche $V_C \rightarrow V_E$ dans le graphe causal associé à un modèle structurel représente une relation non pas entre deux propriétés, mais entre deux ensembles de propriétés. Cette flèche seule n'indique ni quelle(s) valeur(s) de V_C est (sont) une (des) cause(s), ni quelle(s) valeur(s) de V_E elle(s) cause(nt). À titre d'illustration, une flèche entre la variable binaire correspondant à la propriété d'être fumeur et la variable binaire correspondant à la propriété de développer un cancer n'indique pas si c'est fumer ou ne pas fumer qui cause le cancer. De façon similaire, une flèche entre une variable représentant la quantité d'eau qu'on donne à une plante et l'état

de santé de cette plante n'indique pas à elle seule quelle(s) quantité(s) d'eau cause(nt) quel(s) état(s) de santé de la plante. En ce sens, si les propriétés considérées comme *relata* de la causalité générique peuvent bien être représentées au moyen de variables, la représentation des relations de cause à effet génériques qui en découle est grossière [7]. On peut donc considérer que le fait que notre procédure vise la causalité entre variables plutôt que la causalité entre propriétés constitue une partie du prix à payer pour la possibilité d'inférer des relations de cause à effet à partir de données statistiques. Il convient toutefois de nuancer cette conclusion et de remarquer ici que la procédure que nous avons définie est telle que le modèle qu'elle conduit à retenir est un modèle estimé, c'est-à-dire un modèle pour lequel on dispose d'une quantification des relations de cause à effet directes, et que l'estimation permet généralement de préciser l'affirmation selon laquelle une variable en cause une autre, et d'identifier des relations de cause à effet entre propriétés.

Causalité relative à un ensemble de variables La seconde caractéristique par où l'objet visé par la procédure d'inférence causale de la section 2.2 se distingue de l'objet analysé par les théories probabilistes consiste dans son caractère relatif. Alors que les théories probabilistes – et avec elles presque toutes les théories philosophiques de la causalité – sont des analyses du concept de causalité tout court – ou, en d'autres termes, de la causalité absolue –, la procédure de la section 2.2 vise la causalité *relativement à un ensemble de variables*. De même que la différence abordée dans le dernier paragraphe, mais pour une raison sensiblement différente, celle-ci contribue à rendre l'objet visé par l'inférence causale hypothético-déductive plus grossier que celui qu'analysent les théories probabilistes. Il est apparu déjà qu'une relation de cause à effet entre variables peut représenter sans les distinguer plusieurs relations de cause à effet entre propriétés. Nous nous apprêtons à montrer que la relation de causalité absolue peut être considérée comme une relation de causalité relative particulière.

Plus précisément, la causalité absolue peut être considérée comme la causalité relative à un ensemble complet de variables (Spohn, 2001, p. 11) :

> « Si la notion de dépendance causale se présente d'abord comme relative à un cadre (*frame relative*), nous pouvons éliminer ce caractère relatif seulement en en venant au cadre qui embrasse tout (*all-embracive frame*), contenant toutes

[7] Cette représentation est grossière en d'autres sens que celui que nous venons de mettre au jour. En particulier, une flèche entre deux variables ne renseigne ni sur le mode d'action de la cause, ni sur les caractéristiques de la relation de cause à effet. Ces points, toutefois, ne nous intéressent pas directement ici.

> les variables nécessaires pour une description complète de la réalité empirique. »

Ainsi, la seconde différence entre les objets visés par l'inférence causale hypothético-déductive et par les théories probabilistes de la causalité se réduit sous l'hypothèse selon laquelle les relations de cause à effet absolues sont les relations de cause à effet relatives à un ensemble de variables qui suffit à décrire la réalité empirique. Cette hypothèse nous semble nécessaire pour qu'il y ait sinon peut-être un sens, en tout cas un intérêt, à s'intéresser aux relations de cause à effet relatives à un ensemble de variables – et donc aussi bien à la modélisation causale qu'aux méthodes d'inférence causale plus récemment développées. Aussi l'émettons-nous.

Il peut toutefois sembler que la notion d'ensemble « contenant toutes les variables nécessaires pour une description complète de la réalité empirique » ne soit pas tenable. Par là, on entendrait essentiellement qu'elle ne peut pas être effectivement manipulée. À cette objection, nous répondons que l'utilisation effective des concepts causaux ne requiert pas de prendre en compte à tout moment toutes les variables d'un ensemble qui suffit à décrire la réalité empirique. Pour analyser la structure causale d'un système donné, il suffit de (et, en général, il faut) considérer un sous-ensemble d'un tel ensemble qui soit tel que connaître les relations de cause à effet entre les variables de ce sous-ensemble permet de comprendre les phénomènes qui composent le système considéré. De cette façon, la notion de relation de cause à effet entre toutes les variables d'un ensemble qui suffit à décrire la réalité empirique devient superflue en pratique et donc légitime en théorie : la notion est légitime en théorie précisément parce que le fait de s'intéresser à la causalité relative à un ensemble de variables qui suffit à décrire la réalité empirique n'implique pas de prendre en compte toutes les variables d'un tel ensemble quand il s'agit de pratiquer l'analyse causale. Surtout, le fait que les relations de cause à effet absolues puissent être considérées comme des relations de cause à effet relatives à un ensemble de variables bien particulier (un ensemble de variables suffisant à décrire la réalité empirique) constitue un second sens auquel l'objet de l'inférence causale hypothético-déductive est plus grossier que l'objet visé par les théories probabilistes de la causalité.

2.3.3 Conséquences de l'hypothético-déductivité

Intéressons-nous maintenant à la procédure d'inférence causale elle-même – et non plus seulement aux types de relations de cause à effet qu'elle vise à mettre au jour. Ainsi qu'il doit être clair à ce point, la principale caractéristique, logique, de cette procédure est que les inférences qui s'y conforment sont hypothético-déductives. De cette caractéristique, il découle que la conclusion de l'inférence n'est pas une conséquence logique de ses prémisses : il est possible que la conclusion soit fausse alors que les prémisses sont vraies.

De manière générale, il est possible de distinguer au moins trois raisons, qui ne sont pas indépendantes, pour lesquelles la conclusion d'une inférence hypothético-déductive (quelle qu'elle soit) peut être fausse alors que ses prémisses sont vraies. En premier lieu, l'hypothético-déduction n'est pas de nature à garantir la vérité des conclusions des inférences. L'hypothèse qui est retenue au titre de conclusion d'une inférence hypothético-déductive n'est pas démontrée à proprement parler ; elle a seulement « provisoirement réussi son test : nous n'avons pas trouvé de raisons de l'écarter » (Popper, 1934, p. 29). En deuxième lieu, il est possible que l'hypothèse qui est vraie n'ait été envisagée à aucun moment par l'investigateur. Dans ce cas, elle ne peut pas être retenue au titre de conclusion de l'inférence et, partant, la conclusion de l'inférence est fausse. En troisième lieu, il est logiquement possible qu'une hypothèse vraie soit rejetée même dans le cas où elle est envisagée. En effet, ainsi que Duhem le premier l'a expliqué clairement, ce ne sont jamais des hypothèses isolées, mais toujours des théories entières qui entrent en contact avec l'expérience (Duhem, 1906). Dans ces conditions, l'échec à passer un test n'affecte pas la seule hypothèse principale explicitement visée par l'inférence, mais l'ensemble formé de cette hypothèse et des hypothèses auxiliaires qui permettent la confrontation avec l'expérience. La fausseté de l'hypothèse principale est inférée de l'échec à passer le test seulement parce que cette inférence est autorisée par la règle de décision adoptée en réponse au problème mis en évidence par Duhem – dans les termes de Mongin : parce qu'elle est autorisée par la « solution duhémienne » adoptée (Mongin, 2007, section II.3[8]). La réfutation n'est alors pas logique ; elle est méthodologique.

Tout cela, que nous venons de rappeler rapidement, est bien connu, et il n'est pas nécessaire que nous nous y attardions. Il convient plutôt que nous en venions aux limites de l'hypothético-déduction telles qu'elles se formulent plus spécifiquement pour le cas qui nous occupe. Autrement dit, il s'agit maintenant pour nous d'expliciter, concernant l'inférence causale probabiliste hypothético-déductive telle que nous proposons de la mener, d'une part la thèse selon laquelle la conclusion de l'inférence n'est pas une conséquence logique de ses prémisses, d'autre part les trois raisons pour lesquelles cette thèse est vraie.

Concernant l'inférence causale telle qu'elle peut être menée en mobilisant les outils de la modélisation structurelle, dire que la conclusion de l'inférence n'est pas une conséquence logique de ses prémisses revient à dire ceci : même si les données traitées qui constituent les prémisses de l'inférence sont correctes, le modèle qui constitue la conclusion de l'inférence peut ne pas être une représentation correcte du système réel

[8] Le texte de Mongin, et en particulier le ch. II., offre une présentation précise et une discussion fouillée de la thèse de Duhem.

qui est étudié[9]. Par « données traitées », nous entendons ici les corrélations partielles entre les variables du modèle telles qu'elles sont calculées à partir des données recueillies pour un échantillon de la population considérée. Ces données traitées sont correctes si et seulement si elles sont égales aux mêmes corrélations partielles, mais pour la fonction de probabilités dans l'ensemble de la population visée par l'analyse. Ainsi, du caractère hypothético-déductif des inférences qui nous intéressent ici, il découle qu'une évaluation correcte des corrélations partielles au sein de la population étudiée n'implique pas que les flèches qui figurent dans la conclusion $MI_{\mathbf{V}}$ de la procédure représentent toutes et exactement les relations de cause à effet entre les variables de cet ensemble.

La première raison en est qu'il est possible que des modèles causaux incorrects passent les tests de l'étape C, et même qu'ils soient retenus à l'issue des étapes E et F. C'est sur le second point que nous nous attardons ici. Nous avons vu plus haut que les étapes E et F prises ensemble constituent une inférence à la meilleure explication. Or, on peut soutenir que rien ne garantit que l'explication la meilleure soit correcte. Surtout (et de manière non indépendante), il est très difficile de définir ce que c'est, pour une explication, qu'être meilleure qu'une explication concurrente (Harman, 1965, p. 89). Cette difficulté prend une forme particulièrement aiguë aux étapes E et F de la procédure définie dans la section 2.2. D'une part, « il existe des douzaines de mesures d'adéquation des modèles aux données [...] et de nouvelles mesures sont développées en permanence » (Kline, 1998, p. 133) et, parce qu'elles ne visent pas toutes le même aspect du rapport entre un modèle et les données traitées, ces différentes mesures n'ordonnent pas les modèles de la même façon. D'autre part, ainsi que nous l'avons déjà suggéré, il est notoirement difficile de donner un contenu précis à la notion de plausibilité d'un modèle causal hypothétique. En conséquence, à chaque étape, le critère de sélection d'un modèle parmi des modèles concurrents n'est pas tel que le modèle causal correct est nécessairement sélectionné s'il figure parmi ceux entre lesquels il s'agit de discriminer. Encore une fois, même si les données qui constituent les prémisses de l'inférence causale sont correctes, rien ne garantit la correction du modèle qui constitue sa conclusion.

En deuxième lieu, de même que dans le cas général, la conclusion d'une inférence qui se conformerait à la procédure de la section 2.2 dépend dans ce cas particulier des modèles structurels que le scientifique pratiquant l'inférence aura été capable de, ou porté à, spécifier à l'étape A.

[9] Ici comme dans la fin du chapitre, nous parlons de correction ou d'incorrection d'un modèle ou de la représentation d'un système par un modèle. En effet, il n'y a pas de sens propre à parler de vérité dans ces cas. Toutefois, nous considérons que ce qu'on peut dire de la vérité et de la fausseté pour l'hypothético-déduction en général peut se dire de la correction ou de l'incorrection dans le cas particulier de l'inférence causale probabiliste hypothético-déductive.

Ainsi que plusieurs auteurs l'ont souligné (en particulier Freedman, 1987, p. 120-121 et Freedman, 1991, p. 303-304 et 309), cette conclusion ne dépend donc pas des seules données traitées, *a fortiori* n'en est-elle pas une conséquence logique.

Pour ce qui est, en troisième lieu, du problème soulevé par Duhem, les termes dans lesquels il se pose relativement à la procédure de la section 2.2 sont les suivants : les modèles rejetés à l'étape C ne sont pas réfutés à la rigueur logique du terme, mais seulement en un sens méthodologique. Autrement dit, du point de vue logique, rien ne garantit que c'est l'hypothèse causale examinée qui doit être rejetée en cas d'échec aux tests de l'étape C. Un tel échec peut également être le fait des hypothèses auxiliaires qui permettent de confronter cette hypothèse aux données disponibles. En particulier, il peut être le fait des hypothèses statistiques qui entrent dans la composition des modèles structurels – et dont dépend le détail des techniques qui sont mobilisées. Le problème de Duhem est d'autant plus prégnant dans le cas qui nous occupe que ces hypothèses statistiques sont elles-mêmes plus difficiles à tester (et que d'ailleurs elles ne le sont généralement pas).

2.3.4 Inférence statistique

Concernant l'inférence causale probabiliste hypothético-déductive, l'analyse a porté jusqu'à présent sur sa seule logique. Nous nous tournons maintenant vers une autre caractéristique de la procédure d'inférence causale définie dans la section 2.2, qui a aussi des conséquences relativement à sa validité : les inférences qui s'y conforment portent sur des hypothèses et des énoncés probabilistes, et les données qui en constituent les prémisses ne sont pas relatives à la population considérée, mais à un échantillon de cette population. En d'autres termes, ces inférences relèvent du domaine de l'inférence statistique. Il en découle de nouvelles raisons pour lesquelles il est possible que la conclusion d'une telle inférence soit incorrecte alors même que les prémisses seraient correctes.

À l'étape B de la procédure qui est définie dans la section 2.2, cette possibilité prend plus précisément la forme suivante : les paramètres associés au modèle considéré ne sont pas déduits des données d'observation disponibles – mais seulement *estimés* à partir d'elles. C'est toutefois principalement à l'étape C qu'émergent les conséquences que la nature des prémisses de l'inférence peut avoir sur la correction de sa conclusion. À cette étape, en effet, le caractère probabiliste des hypothèses testées implique que les modèles rejetés ne sont pas réfutés, mais plus exactement font l'objet d'une décision de rejet.

Il convient ici d'être clair. Ce qui est en jeu ici n'est ni la possibilité pour une hypothèse incorrecte de passer un certain nombre de tests, ni le fait que la réfutation méthodologique est compatible avec l'attribution à une hypothèse correcte de la responsabilité de l'échec à passer un

test donné. En d'autres termes, ce qui est en jeu ici ne se réduit pas au problème de Duhem. Positivement, ce qui est en jeu est *propre au contexte statistique* et il s'agit de ceci que, *les hypothèses auxiliaires étant satisfaites*, il existe une probabilité non nulle qu'une hypothèse incorrecte passe le test et il existe une probabilité non nulle qu'une hypothèse correcte échoue à le passer.

Concernant des hypothèses probabilistes, la réfutation (fût-elle méthodologique) cède la place aux décisions de rejet fondées sur la définition de zones critiques (Mongin, 2007, ch. III ; Hacking, 1965). Popper parle de « falsifiabilité pratique » et de « falsification pratique », et il n'a de cesse de souligner leurs limites (Popper, 1934, p. 192-193) :

> « Il est clair que cette "falsification pratique" ne peut résulter que de la décision méthodologique de considérer des événements hautement improbables comme exclus – prohibés. Mais de quel droit peut-on les considérer ainsi ? Où devons-nous tracer la ligne de séparation ? Où commence cette "haute improbabilité" ?
>
> [...] D'un point de vue purement logique, il ne peut y avoir de doute : c'est un fait que les énoncés de probabilité ne peuvent pas être falsifiés. »

Pour ce qui concerne spécifiquement les tests menés à l'étape C de notre procédure d'inférence causale, ils peuvent conduire à rejeter des hypothèses correctes et à ne pas rejeter des hypothèses incorrectes. Plus précisément, même dans le cas où les hypothèses statistiques qui entrent dans la composition d'un modèle sont satisfaites par le système considéré, l'étape C peut avoir pour issue le rejet d'un modèle correct ou l'absence de rejet d'un modèle incorrect.

2.3.5 Conclusion

En nous appuyant sur la modélisation causale, nous avons défini dans la section 2.2 une procédure pour l'inférence causale probabiliste hypothético-déductive. Plus précisément, la section 2.2 définit la procédure d'inférence causale probabiliste hypothéthico-déductive la plus exigeante parmi celles que les outils de la modélisation structurelle permettent de définir. Cette procédure permet d'inférer des relations de cause à effet à partir de données statistiques ou, pour le dire autrement, de données d'observation de nature probabiliste – et donc de contourner les obstacles dont on a vu qu'ils surgissent si l'on prétend fonder l'inférence causale sur les théories probabilistes de la causalité. Ce contournement se paie d'un prix dont nous avons discuté dans la section 2.3 : l'inférence causale a pour objet une notion relative de causalité entre variables, elle se heurte aux limites logiques génériquement attachées à l'hypothético-déduction, mais aussi aux difficultés qui sont plus spécifiquement attachées à l'inférence statistique et aux tests d'hypothèses probabilistes.

CHAPITRE 3

Une voie inductive ? Réseaux bayésiens et inférence causale probabiliste

APRÈS AVOIR MONTRÉ (dans la sous-section 1.5.1) que les théories probabilistes ne peuvent pas être utilisées comme principes pour l'inférence causale, nous venons d'envisager la possibilité d'une approche hypothético-déductive. Cette approche nous a été suggérée par ceci que les arguments de la sous-section 1.5.1 semblent viser la seule induction. Plus précisément, la thèse que nous avons défendue dans la sous-section 1.5.1 est la suivante : il est impossible de considérer les théories probabilistes comme énonçant un critère qui permettrait de tirer des conclusions causales seulement à partir de connaissances probabilistes issues de données d'observation.

Dans ces conditions, il est étonnant que les partisans de certaines méthodes d'inférence causale récemment développées soutiennent que, précisément, ces méthodes récentes permettent d'*induire* des connaissances causales à partir de connaissances probabilistes. Sur quoi une telle prétention peut-elle être fondée, si nos meilleures analyses probabilistes du concept de cause ne peuvent pas servir de principes à de telles inductions ? Telle est la question qui doit nous occuper maintenant.

Les méthodes d'inférence causale probabiliste dont il est question ici reposent de manière essentielle sur le recours à des objets formels appelés « réseaux bayésiens ». Aussi, par commodité, nous désignerons ces méthodes au moyen de l'expression suivante : « méthodes RB ». En outre, de même que dans le chapitre 2, pour les mêmes raisons mais également pour autoriser la confrontation des méthodes étudiées dans l'un et l'autre chapitre, nous limitons ici notre analyse au cas simple où : 1) les variables du modèle structurel relèvent toutes du même niveau de réalité, 2) la valeur de chacune est observable, et 3) l'ensemble des variables du modèle est causalement suffisant.

3.1 Un aperçu sur les méthodes RB

La section qui commence vise à rendre compte du principe de l'inférence causale probabiliste fondée sur les réseaux bayésiens. Nous n'entrons donc pas dans le détail des méthodes RB. En particulier, nous ne définissons pas la notion de réseau bayésien.

On peut considérer que les méthodes RB pour l'inférence causale probabiliste reposent sur deux piliers de nature différente :

1. l'hypothèse selon laquelle le graphe causal sur un ensemble de variables donné entretient un certain rapport avec la fonction de probabilités sur cet ensemble ;
2. la possibilité, une fonction de probabilités sur un ensemble de variables étant donnée, de construire l'ensemble des graphes entretenant avec elle le rapport qu'elle est supposée (selon 1) avoir avec le graphe causal.

3.1.1 Hypothèse concernant le rapport entre graphe causal et fonction de probabilités

Nous avons vu dans la sous-section 2.1.1 que le graphe causal sur un ensemble de variables $\mathbf{V}$ est le graphe orienté dont les nœuds sont les variables de $\mathbf{V}$ et dont les flèches représentent les relations de cause à effet directes entre ces variables. De l'autre côté, une fonction de probabilités sur $\mathbf{V}$ attribue un nombre réel compris entre 0 et 1 à chaque « valeur » de $\mathbf{V}$, c'est-à-dire à chaque combinaison d'une valeur pour chacune des variables de $\mathbf{V}$. La somme des nombres réels attribués aux différentes valeurs de $\mathbf{V}$ vaut 1. D'une fonction de probabilités sur $\mathbf{V}$, il suit une probabilité pour toute valeur de tout sous-ensemble non vide de $\mathbf{V}$, en même temps que toutes les probabilités conditionnelles engageant les variables de $\mathbf{V}$ et leurs différentes valeurs. De manière générale, la fonction de probabilités sur un ensemble de variables donne donc toutes les probabilités, absolues et conditionnelles, concernant les variables de cet ensemble.

L'hypothèse qui constitue le premier des piliers sur lesquels reposent les méthodes RB porte sur le rapport entre graphes causaux et fonctions de probabilités. Plus exactement, elle porte sur le rapport entre le graphe causal sur un ensemble de variables donné et les indépendances probabilistes relatives au sein du même ensemble. Une indépendance probabiliste relative est une indépendance probabiliste qui vaut relativement à un ensemble de variables par lequel on conditionalise. Ainsi, pour une fonction de probabilités p, deux variables A et B sont indépendantes relativement à un ensemble de variables $\mathbf{C}$ si et seulement si : pour toute valeur a de A, toute valeur b de B et toute valeur $\mathbf{c}$ de $\mathbf{C}$ telle que $p(\mathbf{c}) \neq 0$, $p(a \wedge b|\mathbf{c}) = p(a|\mathbf{c}).p(b|\mathbf{c})$. Dans le cas où $p(b|\mathbf{c}) \neq 0$, cela équivaut à $p(a|b \wedge \mathbf{c}) = p(a|\mathbf{c})$. Soulignons qu'il y a indépendance probabiliste relative

seulement si les égalités énoncées valent pour *toutes* les valeurs de A, de B et de $\mathbf{C}$; il s'agit donc d'une relation particulièrement exigeante.

L'hypothèse sur laquelle reposent les méthodes RB est en fait double. Pour un ensemble de variables $\mathbf{V}$, si GC est le graphe causal sur $\mathbf{V}$ et p la fonction de probabilités sur $\mathbf{V}$, l'hypothèse veut que :

1. toute variable de $\mathbf{V}$ est indépendante pour p, relativement à ses parents dans GC, de toute variable de $\mathbf{V}$ qui n'appartient pas à l'ensemble de ses descendants dans GC. Autrement dit et puisque CG est le graphe causal sur $\mathbf{V}$, pour toute variable de $\mathbf{V}$, l'ensemble de ses causes directes dans $\mathbf{V}$ fait écran entre cette variable et toutes les variables de $\mathbf{V}$ qu'elle ne cause pas, à l'exception d'elle-même. C'est ce qu'on appelle la « condition de Markov causale » ;
2. toutes les indépendances probabilistes entre variables de $\mathbf{V}$ relatives à des (ensembles de) variables de $\mathbf{V}$ sont impliquées par la clause 1, c'est-à-dire par la condition de Markov causale. C'est ce qu'on appelle l'hypothèse de fidélité causale.

Par commodité, nous proposons de noter R le rapport que graphes causaux et fonctions de probabilités entretiennent selon l'hypothèse que nous venons d'énoncer.

À titre d'illustration, considérons l'ensemble de variables

$$\mathbf{V} = \{A, B_1, B_2, C, D, E, F\}$$

et admettons que le graphe causal sur cet ensemble est donné par la figure 3.1 :

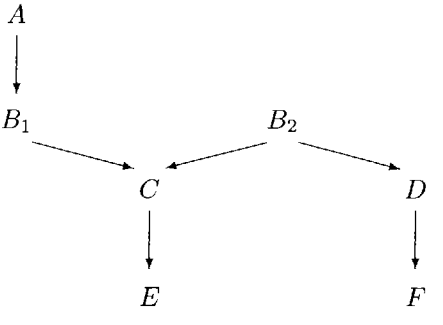


FIGURE 3.1

Ce graphe et la fonction de probabilités sur $\mathbf{V}$ sont dans le rapport R si et seulement si, pour p :

- A est indépendante de B_2, D et F relativement à l'ensemble vide, c'est-à-dire qu'elle en est absolument indépendante ;
- B_1 est indépendante de B_2, D et F relativement à A ;

- B_2 est absolument indépendante de B_1 et de A ;
- C est indépendante de A, D et F relativement à $\{B_1, B_2\}$;
- D est indépendante de A, B_1, C et E relativement à B_2 ;
- E est indépendante de A, B_1, B_2, D, F relativement à C ;
- F est indépendante de A, B_1, B_2, C et E relativement à D ;
- toutes les indépendances probabilistes entre variables de $\mathbf{V}$ relatives à des variables de $\mathbf{V}$ sont impliquées par ces indépendances probabilistes relatives.

Nous ne nous attardons pas maintenant sur la question de savoir si graphes causaux et fonctions de probabilités sont effectivement dans le rapport R. Cette question sera traitée plus bas, notamment dans la sous-section 3.4.1 et dans la section 4.2. Nous pouvons toutefois déjà indiquer que l'analyse conduira à la fois à considérer comme fort plausible l'hypothèse selon laquelle ils le sont et à envisager la possibilité que cette hypothèse soit, parfois, violée.

Notons, pour en finir avec R, que nous avons évoqué jusqu'ici *le* graphe causal et *la* fonction de probabilités sur un ensemble de variables donné, suggérant que l'un comme l'autre sont uniques. Cela n'implique pas que nous considérions que la causalité et les probabilités sont des réalités du monde physique, et que la tâche d'inférence causale probabiliste consiste à inférer des connaissances sur la causalité physique à partir de la connaissance des probabilités physiques. Plus simplement mais de manière déjà non triviale, cela signifie que, à l'instar des tenants des méthodes RB, nous supposons que les informations disponibles fixent un ensemble de dépendances et d'indépendances probabilistes relatives et que la causalité est une réalité objective au sens où deux individus rationnels disposant des mêmes informations ne peuvent pas être en désaccord concernant la causalité. Williamson, 2005, déploie complètement, pour les probabilités comme pour la causalité, une position objective de ce type, aux détails de laquelle nous estimons ne pas être tenus ici.

3.1.2 Graphes construits à partir d'une fonction de probabilités

Le second des piliers sur lesquels reposent les méthodes RB est le suivant : étant donné un ensemble de variables et les indépendances probabilistes relatives qui valent au sein de cet ensemble, on sait construire les graphes qui sont dans le rapport R avec la fonction de probabilités sur cet ensemble. Plus précisément, il existe des algorithmes qui permettent de mener cette tâche à bien.

Exactement, étant donné l'ensemble des indépendances probabilistes relatives entre les variables d'un ensemble $\mathbf{V}$ donné, ces algorithmes construisent sur $\mathbf{V}$ un graphe partiellement orienté – c'est-à-dire dans lequel ne figurent pas seulement des flèches, mais également des liens non orientés. Ce graphe, généralement appelé « patron » (*pattern*), représente la classe de tous les graphes acycliques et orientés qui sont dans le rapport R avec la fonction de probabilités considérée. Il est composé de tous et

seulement les liens qui, orientés ou non, sont partagés par l'ensemble de ces graphes. Notons qu'on sait énoncer des contraintes graphiques simples pour l'orientation des flèches en vue d'obtenir des graphes complètement orientés et acycliques parmi ceux que le patron représente.

3.1.3 Schéma d'inférence causale probabiliste

Il nous semble que le principe de l'inférence causale probabiliste menée selon les méthodes apparaît clairement à ce point. Étant donné un ensemble de variables **V** et les indépendances probabilistes relatives entre les variables de **V**, les algorithmes évoqués dans la sous-section 3.1.2 permettent d'inférer le graphe causal sur **V**. En effet, ces algorithmes construisent l'ensemble des graphes orientés acycliques qui sont dans le rapport R avec la fonction de probabilités sur **V** et, par hypothèse, le graphe causal sur **V** entretient précisément ce rapport avec cette fonction de probabilités.

Cette formulation demande à être nuancée ou, à tout le moins, précisée sur deux points. En premier lieu, les algorithmes mobilisés ont pour résultat le graphe causal *parmi d'autres graphes*. Rappelons, en effet, que ces algorithmes construisent un patron représentant *l'ensemble* des graphes orientés et acycliques qui sont dans le rapport R avec la fonction de probabilités sur **V**. Nous avons déjà indiqué que les liens qui figurent dans ce patron sont exactement ceux que partagent tous les graphes orientés acycliques que ce patron représente. Dès lors, le patron représente exactement les connaissances causales qu'il est possible de tirer de la fonction de probabilités sur **V** sous l'hypothèse selon laquelle cette fonction entretient le rapport R avec le graphe causal sur **V**.

En second lieu, l'expression « étant donné un ensemble de variables **V** et la fonction de probabilités sur **V** » est trompeuse :

- les cas dans lesquels on infère des connaissances causales de connaissances probabilistes ne sont généralement pas tels que la fonction de probabilités sur l'ensemble de variables considéré est *donnée*. En effet, ainsi que nous l'avons indiqué plus haut, la fonction de probabilités sur un ensemble de variables véhicule une quantité d'informations très importante, puisqu'elle donne toutes les probabilités (absolues et conditionnelles) concernant toutes les variables de l'ensemble considéré. En outre, dans le contexte qui nous intéresse, on ne dispose presque jamais d'informations concernant les probabilités *dans la population considérée* : on dispose plutôt de données statistiques, portant sur des fréquences relatives observées dans un échantillon de cette population ;
- nous avons vu que, du côté probabiliste, le rapport R n'engage pas des fonctions de probabilités tout entières, mais plutôt les ensembles d'indépendances probabilistes relatives qui valent pour ces fonctions.

Dans ces conditions, construire les graphes qui sont dans le rapport R avec la fonction de probabilités pour laquelle on a des données suppose que soient d'abord inférées, à partir des données disponibles, les indépendances probabilistes relatives au sein de l'ensemble de variables considéré, et pour la population étudiée.

Selon les méthodes RB, l'inférence causale probabiliste se conforme donc au schéma très général qui suit. Initialement, l'hypothèse selon laquelle le graphe causal et la fonction de probabilités sur l'ensemble de variables considéré entretiennent le rapport R, est acceptée. Puis, les indépendances probabilistes relatives au sein de l'ensemble de variables considéré sont identifiées à partir des données disponibles, et [1] un algorithme est utilisé afin d'inférer, à partir de ces indépendances, l'ensemble des graphes orientés acycliques qui entretiennent le rapport R avec la fonction de probabilités concernée. Finalement, la conclusion autorisée veut que le graphe causal sur l'ensemble de variables considéré appartienne à l'ensemble des graphes orientés et acycliques construit par l'algorithme. Le patron représente donc toute l'information causale qu'on peut tirer des indépendances probabilistes relatives sous l'hypothèse selon laquelle graphe causal et fonction de probabilités entretiennent le rapport R.

3.1.4 Illustration

À titre d'illustration, imaginons que nous soyons intéressés par les relations de cause à effet qui existent, dans une population donnée, entre consommation de thé, consommation de tabac, couleur des dents et qualité de leur brossage. Plus précisément, imaginons que, concernant cette population, nous disposions de données statistiques concernant les variables C, T,D et B, représentant respectivement la consommation de café, la consommation de tabac, la blancheur des dents et la qualité du brossage des dents.

Afin d'inférer les relations de cause à effet au sein de l'ensemble de ces variables, les méthodes RB demandent qu'on commence par accepter l'hypothèse selon laquelle le graphe causal GC sur cet ensemble entretient le rapport R avec la distribution de probabilités p sur cet ensemble :

1. toute variable de $\{C, T, D, B\}$ est indépendante pour p, relativement à l'ensemble de ses descendants dans GC, de toute variable de $\{C, T, D, B\}$ qui n'appartient pas à l'ensemble de ses descendants dans GC ;
2. toutes les indépendances probabilistes au sein de $\{C, T, D, B\}$ sont impliquées par 1.

[1] Signalons que « et » ne doit pas nécessairement être interprété comme désignant une succession temporelle. Nous reviendrons sur ce point plus bas.

Une fois cette hypothèse acceptée, on utilise les données statistiques disponibles afin d'identifier les indépendances probabilistes au sein de l'ensemble $\{C, T, D, B\}$. Imaginons que ces données nous conduisent à accepter les affirmations suivantes :

- C et B sont indépendantes relativement à tous les sous-ensembles de $\{T, D\}$;
- T et B sont indépendantes relativement à $\varnothing$ mais dépendantes relativement à $\{D\}$.

On recourt ensuite aux algorithmes évoqués dans la sous-section 3.1.2. Ils permettent de construire les graphes orientés acycliques sur $\{C, T, D, B\}$ qui entretiennent le rapport R avec les distributions de probabilités sur $\{C, T, D, B\}$ que caractérise l'ensemble de ces indépendances probabilistes. L'ensemble de ces graphes orientés acycliques est représenté par le patron de la figure 3.2 :

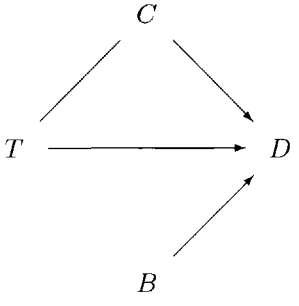


FIGURE 3.2

Ce patron représente une classe comprenant exactement deux graphes causaux acycliques : les graphes (a), (b) et (c) de la figure 3.3.

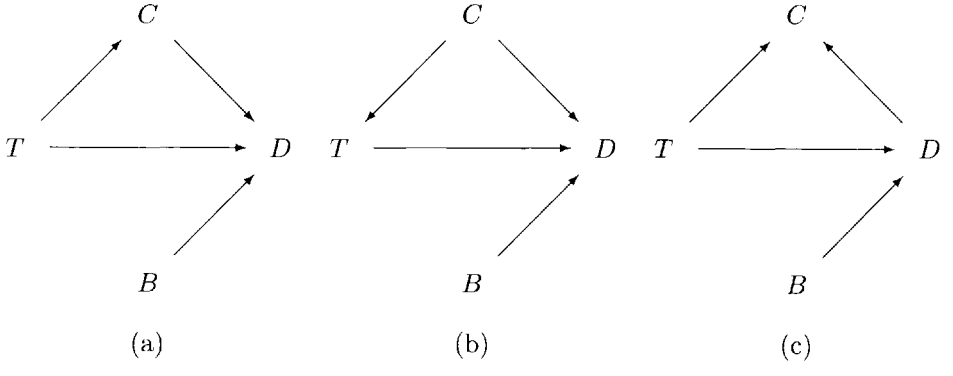


FIGURE 3.3

Chacun de ces trois graphes entretient la relation R avec les fonctions de probabilités sur $\{C, T, D, B\}$ que caractérise l'ensemble des indépendances probabilistes que nous avons énoncées. Considérons, par exemple,

le graphe (a). Dans ce graphe, T n'a pas de parent et il n'y a que B qui ne soit pas l'un de ses descendants. La première clause du rapport R implique donc que T et B sont absolument indépendantes. Or, c'est bien le cas pour l'ensemble d'indépendances probabilistes que nous avons envisagé. La forme de (a) et la première clause de R prises ensemble impliquent de la même façon exactement deux autres indépendances : celle de C et B absolument, d'une part, et relativement à T, d'autre part. Il s'agit là encore d'indépendances que nous avons stipulées. Mais les trois indépendances impliquées par la forme de (a) et la première clause de R n'épuisent pas l'ensemble des indépendances probabilistes que nous avons acceptées. Plus exactement, nous avons également accepté les indépendances suivantes :

– T et B sont indépendantes relativement à C, à D et à $\{C, D\}$.

Or, toutes ces indépendances supplémentaires sont impliquées pour les premières par l'indépendance absolue de C et de B, et pour les secondes par l'indépendance absolue de T et de B. Dans ces conditions, il est bien vrai que (a) entretient le rapport R avec la distribution de probabilités sur p. C'est le cas également des graphes (b) et (c), et d'aucun autre graphe orienté acyclique sur $\{C, T, D, B\}$.

La conclusion de l'inférence causale probabiliste menée pour $\{C, T, D, B\}$ selon les méthodes RB est la suivante : le graphe causal sur $\{C, T, D, B\}$ est soit (a) soit (b) soit (c). Il s'agit là de toute l'information qu'on peut tirer des indépendances probabilistes au sein de $\{C, T, D, B\}$ sous le rapport R. Selon cette information, dans $\{C, T, D, B\}$, T et B sont toutes deux causes directes de D dans $\{C, T, D, B\}$ et il existe une relation de cause à effet directe entre C et T et entre C et D. La méthode adoptée ne permet toutefois pas de déterminer si c'est T (resp. D) qui est une cause directe de C dans $\{C, T, D, B\}$, ou si c'est l'inverse.

3.1.5 Mise en œuvre

Telles que nous venons de les introduire, les méthodes RB reposent de manière essentielle sur l'existence de certains algorithmes. Dans ces conditions, l'inférence causale en tant qu'elle se conforme à ces méthodes se prête particulièrement bien à l'automatisation. En fait, elle en est même inséparable, puisque les méthodes RB ne sont jamais mises en œuvre que de façon automatique.

Les membres de la famille TETRAD [2] sont des ensembles de programmes (des *packages*) qui, entre autres tâches, mènent à bien la tâche d'inférence causale conformément aux méthodes RB. Ils mènent à bien cette tâche de manière modulaire : le programme utilisé pour l'identification des indépendances probabilistes relatives au sein de l'ensemble de va-

[2] Il ne nous est pas nécessaire ici de distinguer entre ces différents membres.

riables considéré est indépendant du programme qui construit l'ensemble des graphes entretenant le rapport R avec la fonction de probabilités sur cet ensemble. Selon les hypothèses qu'il émet à propos du système qu'il étudie, l'utilisateur choisit tel ou tel programme de l'un et de l'autre type. La modularité des ensembles de programmes TETRAD implique qu'il est possible, pour le cas où les indépendances probabilistes relatives au sein de l'ensemble de variables considéré sont connues, d'utiliser seulement un programme permettant de construire les graphes qui entretiennent le rapport R avec la fonction de probabilités sur cet ensemble.

3.2 Vue détaillée des méthodes RB*

Nous venons de proposer un aperçu sur les méthodes récentes pour l'inférence causale probabiliste. Cet aperçu a été conçu afin de donner de ces méthodes une idée suffisamment précise pour permettre de comprendre les arguments philosophiques développés dans la suite de l'ouvrage. Toutefois, il ne présente dans le détail technique ni les outils ni les procédures qui sont mobilisés. Dans ces conditions, précisément, la section qui débute offre une vue un peu plus détaillée. Sans prétendre à l'exhaustivité, elle permettra au lecteur intéressé d'acquérir des connaissances plus détaillées et plus précises concernant une famille de méthodes d'inférence causale présentant un intérêt philosophique et méthodologique certain. Rappelons, en effet, que les partisans de ces méthodes soutiennent qu'elles permettent d'*induire* des relations de cause à effet, à partir de connaissances probabilistes tirées de données statistiques d'observation. En outre, les outils et procédures mobilisés dans le cadre de ces méthodes présentent un intérêt intrinsèque pour qui s'intéresse aux développements scientifiques récents. Particulièrement, le principal outil théorique mobilisé est le concept de réseau bayésien, et les réseaux bayésiens sont de plus en plus utilisés pour représenter et traiter l'incertitude, dans des domaines scientifiques de plus en plus variés.

3.2.1 Réseaux bayésiens

Les réseaux bayésiens sont des outils de représentation et, conséquemment, de traitement de l'incertitude qui sont apparus en intelligence artificielle au début des années 1980. La représentation offerte est numérique, par opposition à la représentation logique véhiculée en particulier par les logiques non monotones. Elle est en outre probabiliste, c'est-à-dire que les nombres utilisés pour représenter l'incertitude sont des probabilités au sens mathématique du terme. En cela, la représentation de l'incertitude au moyen des réseaux bayésiens se distingue de représentations numériques et non probabilistes. À titre d'illustration, le système expert MYCIN (Shortliffe et Buchanan, 1984) repose sur une représentation de ce dernier type.

Un réseau bayésien est un couple composé d'un graphe orienté acyclique et d'une fonction de probabilités, tous deux définis sur un même ensemble de variables, et qui entretiennent un certain rapport. En se servant du vocabulaire de la parenté pour désigner les relations entre les variables d'un graphe orienté acyclique et en convenant qu'une variable appartient à l'ensemble de ses descendants dans un tel graphe, ce rapport est défini de la façon suivante :

Définition 3.1 (Condition de Markov) Soit $\mathbf{V}$ un ensemble de variables, G un graphe orienté acyclique sur $\mathbf{V}$ et p une fonction de probabilités sur $\mathbf{V}$.
(G, p) satisfait la condition de Markov si et seulement si chaque variable de $\mathbf{V}$ est indépendante pour p de tous ses non-descendants dans G relativement à ses parents dans G.

On reconnaît ici la première des clauses qui définit le rapport R de la sous-section 3.1.1. Un réseau bayésien, maintenant, se définit comme suit :

Définition 3.2 (Réseau bayésien) Soit $\mathbf{V}$ un ensemble de variables. Un réseau bayésien sur $\mathbf{V}$ est un couple (G, p) tel que :

1. G est un graphe orienté acyclique sur $\mathbf{V}$;
2. p est une fonction de probabilités sur $\mathbf{V}$;
3. (G, p) satisfait la condition de Markov.

Étant donné une fonction de probabilités p sur un ensemble de variables $\mathbf{V}$, d'un graphe orienté acyclique G tel que (G, p) est un réseau bayésien sur $\mathbf{V}$, on dira qu'il *représente* p.

Un réseau bayésien est caractérisé par ceci que, pour toute variable figurant dans le graphe qui le compose, la valeur de cette variable dépend seulement de celle de ses parents dans le graphe. De façon similaire, dans une chaîne de Markov, un état suffit à rendre non pertinents tous les états qui le précèdent relativement à l'état qui lui succède immédiatement. On comprend alors la référence à Markov dans le cadre de la définition des réseaux bayésiens.

On comprend surtout que le graphe G qui constitue l'une des composantes d'un réseau bayésien (G, p) sur $\mathbf{V}$ donne à voir, pour chaque variable V de $\mathbf{V}$, un ensemble de variables dont la valeur suffit à déterminer la probabilité des différentes valeurs de V. Corrélativement, G indique les variables dont la valeur de V ne dépend pas et que, en conséquence, il n'est pas nécessaire de prendre en compte au moment de définir la probabilité des différentes valeurs possibles de V : « Le truc, dès lors, est d'encoder les connaissances de telle sorte que ce qu'on peut ignorer est reconnaissable (*the ignorable is recognizable*) ou, mieux encore, que ce qu'on peut ignorer est identifié rapidement et accessible facilement »

(Pearl, 1988, p. 12-13). Dans ces conditions, connaître un graphe G sur $\mathbf{V}$ dont on sait qu'il représente la fonction de probabilités p sur $\mathbf{V}$ simplifie la tâche consistant à définir p. Plus précisément, connaître un tel graphe réduit significativement le nombre de paramètres qu'il faut spécifier afin de définir complètement p. La réduction est d'autant plus importante que le graphe compte un plus grand nombre de liens[3].

Les variables qui n'ont pas à être prises en compte au moment de définir les probabilités des différentes valeurs de V n'ont pas non plus à l'être au moment d'actualiser ces probabilités. Ainsi, le graphe G qui compose un réseau bayésien (G, p) représente des informations qui non seulement simplifient la tâche qui consiste à définir p, mais encore facilitent l'actualisation de cette fonction de probabilités. L'actualisation des probabilités est d'ailleurs une tâche pour laquelle les réseaux bayésiens ont été utilisés de manière précoce et avec profit. Dès Pearl, 1988, Pearl a développé un algorithme d'actualisation des probabilités dans les graphes appartenant à des réseaux bayésiens et qui sont des arbres (c'est-à-dire qui sont tels que pour toute variable du graphe sauf une, appelée racine, il existe exactement une flèche qui pointe vers elle). À partir de ce travail fondateur, le problème de l'actualisation automatique des probabilités a été résolu pour des classes de graphes de plus en plus vastes. Une solution pour le cas général est présentée dans Lauritzen et Spiegelhalter, 1988. Notons que l'actualisation des probabilités qu'autorisent les réseaux bayésiens porte toujours sur les résultats qu'on obtient par conditionalisation bayésienne ; c'est l'une des raisons pour lesquelles on parle de réseaux *bayésiens*.

Signalons finalement que le graphe qui compose un réseau bayésien (G, p) donne sur les indépendances probabilistes relatives pour p des informations qui ne se réduisent pas à celles que la condition de Markov mentionne explicitement. En effet, pour un réseau bayésien (G, p), on peut lire sur G *toutes* les indépendances probabilistes relatives pour p qui sont impliquées par le fait que (G, p) est un réseau bayésien. Pour le dire autrement, dans les réseaux bayésiens, il existe une condition *graphique* qui est suffisante pour l'indépendance probabiliste relative. Cette condition, appelée « d-séparation », s'énonce de la façon suivante :

Définition 3.3 (d-séparation) Soit un graphe orienté acyclique G sur un ensemble de variables $\mathbf{V}$ et soit $\mathbf{W}$, $\mathbf{X}$ et $\mathbf{Y}$ trois sous-ensembles de $\mathbf{V}$.
$\mathbf{W}$ et $\mathbf{X}$ sont d-séparés par $\mathbf{Y}$ dans G si et seulement si, dans G, tout chemin c d'une variable de $\mathbf{W}$ à une variable de $\mathbf{X}$ est tel que l'une des

[3] Sur ce point, une discussion plus précise et des illustrations se trouvent dans Williamson, 2005.

deux propositions suivantes est vraie :

1. c contient une chaîne $V_i \longrightarrow V_j \longrightarrow V_k$ ou une fourche $V_i \longleftarrow V_j \longrightarrow V_k$ telle que V_j appartient $\mathbf{Y}$;
2. c contient une fourche inversée $V_i \longrightarrow V_j \longleftarrow V_k$ telle que ni V_j ni aucun de ses descendants n'appartient à $\mathbf{Y}$.

À titre d'illustration, considérons le graphe G de la figure 3.4 :

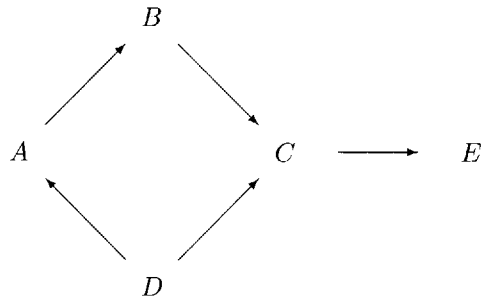


FIGURE 3.4

Dans ce graphe, $\{B\}$ et $\{D\}$ sont d-séparés par $\{A\}$. En effet, il existe seulement deux chemins entre l'unique variable de $\{B\}$ et l'unique variable de $\{D\}$:

1. $B \longleftarrow A \longleftarrow D$ où A appartient à $\{A\}$;
2. $B \longrightarrow C \longleftarrow D$ où ni C ni son unique descendant E n'appartiennent à $\{A\}$.

En revanche, $\{B\}$ et $\{D\}$ ne sont pas d-séparés par $\{C\}$: le chemin $B \longrightarrow C \longleftarrow D$, en particulier, ne satisfait ni 1 ni 2 :

1a. il ne contient pas de chaîne ;
1b. il ne contient pas de fourche ;
2. il contient la fourche inversée $B \longrightarrow C \longleftarrow D$, mais C appartient à $\{C\}$.

De la même façon, $\{B\}$ et $\{D\}$ ne sont pas d-séparés par $\{E\}$. En effet, le chemin constitué par la seule fourche inversée $B \longrightarrow C \longleftarrow D$ ne satisfait toujours pas 2 puisque C a un descendant, E, qui appartient à $\{E\}$. Il apparaît alors que, de façon plus générale, tout ensemble auquel appartiennent C ou E ne d-sépare pas $\{B\}$ et $\{D\}$.

La d-séparation est définie dans Verma et Pearl, 1988, et les auteurs montrent que, dans un réseau bayésien, elle est suffisante pour l'indépendance probabiliste relative. Plus exactement, ils établissent le théorème suivant :

Théorème 3.4 (Verma et Pearl, 1988) *Soit G un graphe acyclique orienté sur un ensemble de variables $\mathbf{V}$ et soit $\mathbf{W}$, $\mathbf{X}$ et $\mathbf{Y}$ trois sous-ensembles de $\mathbf{V}$.*

W *et* **X** *sont* d-*séparés par* **Y** *dans* G *si et seulement si* **W** *est indépendant de* **X** *relativement à* **Y** *pour toute fonction de probabilités* p *sur* **V** *qui est représentée par* G.

Nous renvoyons le lecteur intéressé à Verma et Pearl, 1988. De manière plus générale, divers ouvrages offrent sur les réseaux bayésiens des développements plus poussés (en particulier, au plan technique) que ceux qui ont leur place ici. Pearl, 2000, Naïm *et al.*, 2004, et Williamson, 2005, procèdent de points de vue et de perspectives différents, et qui nous semblent complémentaires.

3.2.2 Réseaux bayésiens causaux

Nous venons de présenter les réseaux bayésiens considérés comme des objets formels. Pourtant, ils sont rarement mobilisés en tant que tels, mais le sont presque toujours en référence à une interprétation – plus précisément, à une interprétation des flèches qui figurent dans les graphes composant les réseaux bayésiens. Plusieurs interprétations sont envisageables mais, ainsi qu'il est sans doute déjà clair, c'est l'interprétation causale qui retiendra notre attention ici.

L'idée qu'on trouve au fondement de l'interprétation causale des réseaux bayésiens consiste à considérer que les flèches du graphe orienté qui compose un réseau bayésien représentent (toutes et exactement) les relations de cause à effet directes entre les variables de l'ensemble sur lequel le réseau est défini. Ainsi, un réseau bayésien causal est un réseau bayésien tel que le graphe qui le compose est causal.

Il découle de cette caractérisation que la notion même de réseau bayésien causal va de pair avec l'hypothèse suivante : pour l'ensemble de variables sur lequel est défini le réseau bayésien qu'on veut interpréter causalement, la condition de Markov est satisfaite par le couple composé du graphe causal et de la fonction de probabilités sur cet ensemble. Rappelons que la condition de Markov est satisfaite par un couple (G, p) sur **V** si toute variable de **V** est indépendante pour p de tous ses non-descendants dans G relativement à l'ensemble de ses parents dans G. Dès lors, la notion de réseau bayésien causal suppose que l'ensemble de variables considéré satisfait la condition de Markov causale :

Définition 3.5 (Condition de Markov causale) Soit **V** un ensemble de variables.
V satisfait la condition de Markov causale si et seulement si toute variable de **V** est indépendante en probabilité de toutes les variables de **V** qu'elle ne cause pas relativement à l'ensemble de ses causes directes dans **V**.

La condition de Markov causale n'est rien d'autre que la condition de Markov en tant qu'elle porte sur un couple (G, p) tel que G est causal : dans ce cas, un ancêtre dans G est une cause, un parent une cause

directe, un descendant un effet. La condition de Markov causale peut être considérée comme une propriété non pas du couple, mais de l'ensemble de variables $\mathbf{V}$ sur lequel G et p sont définis. En effet, nous l'avons déjà indiqué, on peut considérer que G et p sont univoquement déterminés pour $\mathbf{V}$ donné.

De manière générale, l'utilisation des réseaux bayésiens causaux repose sur ceci qu'il est plausible que graphes causaux et fonctions de probabilités soient effectivement dans le rapport énoncé par la condition de Markov causale. Plus précisément, on peut considérer que les utilisations des réseaux bayésiens causaux relèvent de deux grands types et il convient, pour les distinguer, que nous nous arrêtions un moment sur le problème suivant : étant donné un ensemble de variables $\mathbf{V}$ et un ensemble d'indépendances probabilistes relatives sur $\mathbf{V}$, construire un graphe orienté acyclique sur $\mathbf{V}$ qui représente les fonctions de probabilités pour lesquelles ces indépendances probabilistes relatives sont satisfaites. Il s'agit d'un problème très difficile du point de vue computationnel[4]. On sait toutefois :

1. le résoudre de manière approchée pour un espace de recherche restreint aux graphes orientés acycliques ayant une certaine forme. Les algorithmes classiques qui résolvent le problème de cette façon sont présentés dans Naïm *et al.*, 2004, section 6.2, et dans Williamson, 2005, sections 3.5 à 3.11 ;
2. le résoudre pour un espace de recherche restreint aux graphes orientés acycliques qui entretiennent avec la fonction de probabilités considérée non seulement la condition de Markov, mais encore l'hypothèse de fidélité (parfois également appelée « hypothèse de stabilité », par Pearl en particulier) :

   Définition 3.6 (Fidélité) Soit (G, p) un réseau bayésien sur $\mathbf{V}$. (G, p) satisfait l'hypothèse de fidélité si et seulement si toutes les indépendances probabilistes relatives pour p sont impliquées par ceci que (G, p) est un réseau bayésien.

En vue d'illustrer ce que signifie l'hypothèse de fidélité, on peut revenir au graphe G que nous considérions dans la sous-section 3.2.1. Étant donné une fonction de probabilités p sur $\{A, B, C, D, E\}$ telle que (G, p) est un réseau bayésien, (G, p) satisfait l'hypothèse de fidélité si toutes les indépendances relatives pour p découlent de ce que la condition de Markov est satisfaite, et donc s'il n'existe pas d'indépendance probabiliste relative pour p ne correspondant pas à une d-séparation dans G. Ainsi, en particulier, si p était telle que B et D étaient indépendantes

[4] Dans la plupart des cas, le problème est NP-complet. Sur ce point, voir par exemple Chickering, 1996.

relativement à C, cela impliquerait que (G, p) ne satisfait pas l'hypothèse de fidélité.

Nous pouvons maintenant en venir à la distinction de deux types d'utilisations de la notion de réseau bayésien causal. Le premier repose sur l'inférence suivante : de ce que la condition de Markov causale semble plausible (un point sur lequel nous aurons à revenir plus bas, en particulier dans la section 4.1), on tire l'idée selon laquelle le graphe causal sur un ensemble de variables donné représente la fonction de probabilités sur cet ensemble. Une méthode rivale de celles que nous venons de présenter pour construire un graphe orienté acyclique représentant la fonction de probabilités sur un ensemble de variables donné consiste alors à construire le graphe causal sur l'ensemble de variables considéré. Cette méthode est à la fois exacte et telle qu'il n'est pas nécessaire d'émettre une hypothèse supplémentaire concernant le rapport entre graphes causaux et fonctions de probabilités. Dans le cadre de ce premier type d'utilisations, la causalité est liée aux réseaux bayésiens selon la modalité suivante : connaître les relations de cause à effet directes au sein d'un ensemble de variables permet de construire un graphe orienté acyclique qui représente une fonction de probabilités sur cet ensemble. En d'autres termes, la causalité directe est considérée comme connue et joue le rôle de guide pour la construction de graphes qui composent des réseaux bayésiens. On trouve une présentation détaillée et une discussion de ce type d'utilisations des réseaux bayésiens causaux dans Gillies, 2002, section 6.

Dans le cadre du second type d'utilisations, la causalité directe n'est pas supposée connue, mais elle est, au contraire, ce qu'il s'agit d'apprendre, l'objet de l'inférence. Plus exactement, le second type d'utilisations des réseaux bayésiens causaux consiste à interpréter causalement les graphes orientés acycliques qu'on construit au moyen des méthodes évoquées en 1 et en 2. Ces utilisations sont en partie justifiées par le caractère plausible de la condition de Markov causale. Cependant, ainsi qu'il doit être clair, il ne suffit pas que l'ensemble de variables qu'on considère satisfasse la condition de Markov causale pour que la conclusion qu'on tire selon les méthodes évoquées en 1 et en 2 puisse être causalement interprétée. Si l'on utilise les méthodes évoquées en 1, il faut également que le graphe causal sur cet ensemble de variables ait une des formes imposées par ces méthodes – le plus souvent, la forme d'un arbre. Si l'on utilise les méthodes évoquées en 2, il faut également que le graphe causal et la fonction de probabilités sur l'ensemble de variables considéré satisfassent l'hypothèse de fidélité. Pour le dire autrement, il faut que l'ensemble de variables considéré satisfasse l'hypothèse de fidélité causale :

Définition 3.7 (Fidélité causale) Soit $\mathbf{V}$ un ensemble de variables qui satisfait la condition de Markov causale.
$\mathbf{V}$ satisfait l'hypothèse de fidélité causale si et seulement si toutes les indépendances probabilistes relatives entre des variables de $\mathbf{V}$ sont impliquées par ceci que $\mathbf{V}$ satisfait la condition de Markov causale.

Cette hypothèse est privilégiée dans le cadre des méthodes RB pour l'inférence causale probabiliste et, même, nous ne connaissons pas d'exemple tel que l'inférence causale probabiliste reposerait sur une hypothèse concernant la forme du graphe causal sur l'ensemble de variables considéré. L'hypothèse de fidélité causale signifie qu'il n'y a pas d'indépendance probabiliste gratuite, ne procédant pas de la structure causale sur $\mathbf{V}$ *via* la condition de Markov causale. Elle serait, par exemple, violée dans le cas où deux variables de $\mathbf{V}$ dont l'une cause l'autre seraient absolument indépendantes en probabilité : une telle indépendance n'est pas impliquée par la condition de Markov causale. De même que la condition de Markov causale, mais pour des raisons sensiblement différentes sur lesquelles nous aurons également l'occasion de revenir (dans la sous-section 4.1.2), l'hypothèse de fidélité causale est une hypothèse plausible. Elle correspond à la seconde clause de la définition du rapport R dans la sous-section 3.1.1. Dans ces conditions, les algorithmes auxquels le point 2 fait référence sont ce que nous avons identifié plus haut (dans la section 3.1) et plus particulièrement la sous-section 3.1.2 comme étant le second des piliers sur lesquels reposent les méthodes RB pour l'inférence causale probabiliste.

3.2.3 Réseaux bayésiens et inférence causale probabiliste

Les algorithmes évoqués dans le point 2 ci-dessus exploitent les résultats suivants : si (G, p) défini sur $\mathbf{V}$ satisfait la condition de Markov et l'hypothèse de fidélité, alors :

- étant donné deux variables V et W appartenant à $\mathbf{V}$, V et W sont directement liées dans G si et seulement s'il n'existe pas de sous-ensemble de $\mathbf{V}$ relativement auquel V et W sont indépendantes pour p. Autrement dit, les flèches qui figurent dans G représentent toutes et exactement les dépendances probabilistes qui ne disparaissent pas par conditionalisation par les sous-ensembles de $\mathbf{V}$;
- étant donné trois variables X, Y, Z appartenant à $\mathbf{V}$, s'il existe dans G un chemin $X - Y - Z$ et aucun lien direct entre X et Z, alors l'orientation de ce chemin est $X \rightarrow Y \leftarrow Z$ si et seulement s'il existe un sous-ensemble $\mathbf{W}$ de $\mathbf{V}$ auquel n'appartient pas Y et relativement auquel X et Z sont indépendants.

Il nous reste maintenant à montrer comment ces résultats sont effectivement exploités à des fins d'inférence causale.

Les algorithmes auxquels il est fait référence au point 2 ci-dessus ont été développés en deux séries parallèles : d'un côté, les algorithmes IC et IC* de Verma et Pearl, présentés en particulier dans Pearl, 2000, sections 2.5 et 2.6 ; de l'autre, les algorithmes SGS, PC et PC*, puis CI et FCI de Spirtes, Glymour et Scheines présentés dans Spirtes *et al.*, 1993, sections 5.4 et 6.7. Parmi ces algorithmes, nous présentons le seul algorithme PC de Spirtes, Glymour et Scheines. La raison pour laquelle nous pouvons présenter un seul parmi les algorithmes que nous avons

nommés est la suivante : dans tous les cas, le principe de l'inférence causale – ce qui nous intéresse ici, finalement est le même. Le choix de présenter un algorithme de la série développée par Spirtes, Glymour et Scheines repose sur deux considérations non indépendantes. D'une part, ces algorithmes sont ceux qui sont utilisés par les ensembles de programmes TETRAD mentionnés dans la sous-section 3.1.5, et les ensembles de programmes de la famille TETRAD sont les plus connus parmi ceux qui mettent en œuvre les méthodes d'inférence causale probabiliste fondées sur les réseaux bayésiens. D'autre part, dans la littérature philosophique consacrée à l'inférence causale fondée sur les réseaux bayésiens, ce sont les algorithmes de Spirtes, Glymour et Scheines qui ont donné lieu aux débats les plus nourris. En particulier, la querelle qui se déroule au long de Humphreys et Freedman, 1996 ; Spirtes *et al.*, 1997 ; Korb et Wallace, 1997, et Freedman et Humphreys, 1999, concerne l'algorithme PC et les programmes de la famille TETRAD. Finalement, nous nous attachons spécifiquement à PC, car il vise l'inférence aux causes dans les cas simples auxquels notre analyse se limite. Ces cas sont caractérisés en particulier par l'hypothèse selon laquelle l'ensemble de variables considéré est causalement suffisant ; l'algorithme CI, en particulier, permet de faire l'économie de cette hypothèse (Spirtes *et al.*, 1993, p. 139-140).

PC est introduit par Spirtes, Glymour et Scheines dans Spirtes *et al.*, 1991, et il est également présenté dans Spirtes *et al.*, 1993, auquel nous nous référons (Spirtes *et al.*, 1993, p. 84-85). PC prend pour entrée (*input*) les indépendances probabilistes relatives sur un ensemble de variables $\mathbf{V}$ et son résultat (*output*) est un patron sur $\mathbf{V}$. PC se compose de quatre instructions, qui découlent immédiatement des deux résultats que nous avons énoncés au début de la présente sous-section. Étant donné l'ensemble des indépendances probabilistes relatives sur un ensemble de variables $\mathbf{V}$, PC procède en effet de la façon suivante :

Étape 1 : Former sur $\mathbf{V}$ le graphe non orienté complet (c'est-à-dire tel que chaque variable de $\mathbf{V}$ est reliée à chaque autre variable de $\mathbf{V}$ par une arête, non orientée), et noter C_1 ce graphe.

Étape 2 : Retirer de C_1 toutes les arêtes entre deux variables X et Y pour lesquelles il existe un sous-ensemble de $\mathbf{V} \setminus \{X, Y\}$ relativement auquel elles sont indépendantes, et noter C_2 le graphe obtenu.

Étape 3 : Pour tout sous-graphe $X - Y - Z$ de C_2 tel que X et Z ne sont pas (directement) liées dans C_2, remplacer $X - Y - Z$ par $X \rightarrow Y \leftarrow Z$ si Y n'est pas nécessaire pour l'indépendance relative de X et Z pour p.

Étape 4 : Orienter toutes les arêtes de C_3 qui peuvent l'être sans qu'aucun sous-graphe de la forme $X \rightarrow Y \leftarrow Z$, ni aucun cycle ne soit créé, et noter $GI_{\mathbf{V}}$ le graphe obtenu.

Comme nous l'avons déjà noté, l'utilisation d'un algorithme du type de PC suppose que soient identifiées les indépendances probabilistes relatives sur l'ensemble de variables considéré. Précisons ici qu'elles sont

généralement recherchées selon la stratégie suivante : pour tout couple de variables de l'ensemble considéré, on détermine d'abord si elles sont absolument indépendantes ; puis, si et seulement si elles ne le sont pas, on détermine si elles sont indépendantes relativement à *une* variable ; puis, si et seulement si elles ne le sont pas, on détermine si elles sont indépendantes relativement à *deux* variables... Cette façon de procéder prend en compte ceci qu'il suffit que deux variables de **V** soient indépendantes relativement à un sous-ensemble de **V** pour qu'elles ne soient pas liées causalement dans **V**. Elle permet de mener à bien l'étape 2 de l'algorithme PC.

Pour ce qui est, maintenant, de l'identification des indépendances probabilistes relatives elle-même, il convient de distinguer deux cas. Dans le cas extrêmement rare où on connaît la fonction de probabilités sur l'ensemble de variables et dans la population considérés, alors identifier les indépendances probabilistes revient à comparer des valeurs de probabilités conditionnelles. Dans le cas où l'on dispose seulement de fréquences relatives sur un échantillon de la population considérée, on recourt à des tests statistiques d'hypothèses de la forme « X est indépendant de Y relativement à $\mathbf{V}'$ *dans la population considérée* », où $\mathbf{V}'$ est un sous-ensemble de **V** auquel n'appartient ni X ni Y. La nature exacte des tests qu'on utilise dépend de la forme qu'on suppose à la distribution des différentes propriétés considérées dans la population étudiée. Ainsi que nous l'avons indiqué dans la sous-section 3.1.5, la mise en œuvre de ces tests, ainsi que celle de PC ou d'un algorithme similaire, sont assurées, en particulier, par les ensembles de programmes de la famille TETRAD.

Venons-en finalement à $GI_{\mathbf{V}}$, c'est-à-dire au résultat de PC. Ainsi que nous l'avons déjà expliqué, il s'agit d'un graphe partiellement orienté, ou « patron », représentant la classe de tous les graphes orientés acycliques qui, avec p, composent un couple satisfaisant la condition de Markov et l'hypothèse de fidélité. On obtient un graphe de la classe représentée par $GI_{\mathbf{V}}$ en orientant les liens non orientés du patron d'une manière qui respecte deux contraintes graphiques : ne pas créer de cycle et ne pas créer de sous-graphe de la forme $X \rightarrow Y \leftarrow Z$ où X et Z ne sont pas adjacentes. Sous l'hypothèse selon laquelle l'ensemble de variables qu'on considère satisfait la condition de Markov causale et l'hypothèse de fidélité causale, le graphe causal sur **V** est l'un de ces graphes orientés acycliques. Pour le dire autrement, la condition de Markov causale et l'hypothèse de fidélité causale autorisent une interprétation causale des flèches qui figurent dans $GI_{\mathbf{V}}$. Ce patron est la conclusion de l'inférence causale probabiliste menée selon les méthodes RB et il représente toute l'information causale relative à **V** que, sous ces hypothèses, on peut tirer de la fonction de probabilités sur **V**.

3.3 Peut-on parler d'induction ?

L'inférence causale probabiliste menée selon les méthodes RB est-elle inductive ? Il nous faut répondre à cette question en vue de déterminer

si et en quel sens ces méthodes permettent effectivement de contourner les obstacles qui s'opposent à ce que les théories probabilistes soient directement utilisées comme des principes pour l'inférence causale.

3.3.1 Induction ou déduction

En première approche, l'inférence causale probabiliste menée selon les méthodes RB se présente effectivement comme inductive : la connaissance causale acquise suivant ces méthodes est tirée de données d'observation seulement, indépendamment de toute théorie concernant la causalité au sein de l'ensemble de variables considéré. Pour le dire plus précisément, l'inférence causale est inductive parce que des connaissances portant sur la causalité générique sont tirées de données issues de la seule observation de cas particuliers. Elle est donc inductive au sens où des conclusions générales sont tirées de prémisses particulières, et en sont tirées directement – c'est-à-dire, en particulier, indépendamment de toute théorie concernant la causalité au sein de l'ensemble de variables considéré. L'inductivité est alors *a-théoricité* ; elle se situe au plan de la *nature* des prémisses et de la *nature* de la conclusion.

Mais il existe un autre plan pour lequel le qualificatif « inductif » fait sens : celui du *rapport logique* entre les prémisses du raisonnement et sa conclusion[5]. En l'espèce, ce plan est celui du rapport logique entre, d'une part, l'ensemble d'indépendances probabilistes relatives $\mathbf{I_V}$ que les algorithmes mobilisés dans le cadre des méthodes d'inférence causale RB prennent pour entrée et, d'autre part, le patron $GI_{\mathbf{V}}$ que ces algorithmes donnent comme résultat.

Nous savons déjà que $GI_{\mathbf{V}}$ représente toute et exactement l'information causale relative à $\mathbf{V}$ que l'on peut tirer des indépendances probabilistes relatives au sein de $\mathbf{V}$ moyennant l'hypothèse émise concernant le rapport entre graphe causal et fonction de probabilités. Ainsi, sous l'hypothèse selon laquelle graphes causaux et fonctions de probabilités entretiennent le rapport R, le rapport entre $\mathbf{I_V}$ et $GI_{\mathbf{V}}$ est de *nécessité* : si les indépendances probabilistes relatives entre les variables de $\mathbf{V}$ sont bien celles qui appartiennent à $\mathbf{I_V}$, alors *nécessairement* les liens qui figurent dans $GI_{\mathbf{V}}$ représentent adéquatement des relations de cause à effet. Pour être plus concis, on dira que l'information causale véhiculée par le résultat des algorithmes d'inférence causale est nécessairement vraie si leur entrée probabiliste l'est. Dans la mesure où leur conclusion découle nécessairement des données traitées ou, en d'autres termes, des connaissances probabilistes tirées des données statistiques disponibles, il faut reconnaître que les inférences causales menées selon les méthodes RB sont, en un sens, déductives.

[5] L'idée selon laquelle l'induction fait l'objet de plusieurs caractérisations classiques qui ne coïncident pas toujours est suggérée dans Vickers, 2006, section 1 en particulier.

Il existe donc un sens auquel les inférences causales suivant les méthodes RB sont inductives, et un sens auquel elles sont déductives. Elles sont inductives parce que des connaissances portant sur la causalité générique sont tirées directement de données d'observation et qui, en conséquence, portent nécessairement sur des cas singuliers (en nombre fini). Or, nous avons mis en évidence dans la sous-section 1.5.1 que les théories probabilistes de la causalité, c'est-à-dire nos meilleures analyses du rapport entre la causalité et les probabilités, ne peuvent pas servir de principes à des inférences causales qui seraient inductives précisément en ce sens. Comment cette affirmation est-elle compatible avec les analyses que nous venons de mener ? Il nous reste à le comprendre. Avant cela, toutefois, il nous semble utile de compléter notre présentation des méthodes RB par une exploration rapide des conséquences qu'a le fait que les inférences qui se conforment à ces méthodes sont, en un sens, déductives.

3.3.2 Conséquences de la déductivité

Les inférences causales probabilistes menées selon les méthodes RB sont déductives au sens où les algorithmes mobilisés dans le cadre de ces méthodes donnent des résultats qui sont une conséquence logique des entrées qu'ils prennent. Dans la sous-section 2.3.3, nous avons montré que les inférences qu'on peut mener en utilisant des modèles causaux probabilistes ne sont pas déductives en ce sens. Dans la même sous-section, nous avons défendu que la non-déductivité de ces inférences plus traditionnelles découlait de ce qu'elles procédaient de la formulation d'hypothèses. Si tel est bien le cas, alors, par contraposition, la déductivité des inférences causales menées selon les méthodes RB implique que ces inférences ne procèdent pas de la formulation d'hypothèses causales ou, pour le dire autrement, qu'elles sont a-théoriques. Une conséquence méthodologique s'ensuit : si la déductivité des inférences causales menées selon les méthodes RB est inséparable de leur a-théoricité, alors il semble difficile d'isoler l'effet de l'a-théoricité de celui de la déductivité.

Nous entrevoyons toutefois un moyen de le faire. D'une part, nous avons vu que l'inférence causale probabiliste hypothético-déductive est mise en œuvre le plus naturellement de manière non automatique (sous-section 2.2.4) et qu'à l'inverse l'a-théoricité a pour pendant l'automatisation de la recherche des causes fondée sur les réseaux bayésiens (sous-section 3.1.5). Toutefois, nous avons indiqué, d'autre part, qu'il existait des méthodes automatiques pour l'inférence causale traditionnelle : les ensembles de programmes de la famille LISREL. Ces programmes, étant automatiques, ne supposent pas qu'une hypothèse théorique soit formulée par le scientifique qui mène l'inférence. Bien au contraire, leur apport essentiel consiste à automatiser l'étape initiale de formulation d'une hypothèse théorique. Du coup, en s'intéressant aux inférences causales menées

par LISREL, on peut prétendre isoler le caractère non déductif des inférences causales probabilistes hypothético-déductives, de leur caractère théorique. Nous proposons donc de comparer les inférences menées par LISREL aux inférences causales RB telles qu'elles sont menées par TETRAD. Cette comparaison devrait permettre sinon d'évaluer précisément, du moins de se faire une idée, de ce qu'est l'effet net de la déductivité des inférences causales menées selon les méthodes RB.

La comparaison des inférences causales menées par TETRAD aux inférences causales menées par LISREL est largement en faveur de TETRAD (voir Spirtes *et al.*, 1990, par exemple). À cela, il existe plusieurs raisons non indépendantes. Arrêtons-nous sur deux d'entre elles, qui résultent clairement de ce que les inférences menées selon les méthodes RB sont déductives là où les inférences causales probabilistes plus traditionnelles ne le sont pas. En premier lieu, TETRAD n'a pas pour résultat un modèle (comme c'est le cas de LISREL), mais un patron représentant une classe de modèles équivalents. Il s'agit d'un avantage de TETRAD, présenté comme tel par Spirtes, Glymour et Scheines (Spirtes *et al.*, 1993, p. 77-78). En effet, dans le contexte de recherche *automatique* des causes, il n'y a pas de raison de préférer à un autre un modèle dont il est indiscernable par les données. Cette caractéristique de TETRAD découle de ce que son résultat n'est ni plus ni moins que la représentation de toute l'information causale qui peut être *déduite* des données traitées. À l'inverse, LISREL, procédant par spécification puis test de modèles, a pour résultat *un* modèle. En second lieu, la sortie donnée par LISREL dépend du chemin emprunté pour y parvenir. Pour le dire plus simplement, LISREL procède de manière séquentielle, et (surtout) une flèche ajoutée ne peut plus être retirée. Il s'agit clairement d'une limite de LISREL : il n'y a pas de raison que des flèches ajoutées successivement et parce que, pour chacune, son ajout est optimal au moment où il a lieu, constituent ensemble un modèle globalement optimal. TETRAD ne se heurte pas à cette limite et, à nouveau, cet avantage tient au caractère déductif des inférences causales menées selon les méthodes RB : quand la conclusion de l'inférence est une conséquence nécessaire de ses prémisses, les voies empruntées pour arriver à la conclusion ne peuvent pas avoir d'importance au plan logique.

À ce point, nous en avons fini avec l'exploration des conséquences de la déductivité des inférences causales menées selon les méthodes RB. Nous pouvons donc en revenir au caractère inductif de ces inférences, c'est-à-dire au fait qu'elles prennent pour prémisses des données issues de la seule observation de cas particuliers. Ce caractère inductif, considéré comme propriété d'une inférence causale probabiliste, est précisément ce que la sous-section 1.5.1 semble exclure. Comment il se peut que les inférences causales menées selon les méthodes RB soient inductives, c'est ce qu'il nous reste à déterminer.

3.4 Possibilité de l'inférence causale probabiliste inductive

3.4.1 Conception de la causalité véhiculée par les méthodes RB

Si les méthodes RB permettent d'inférer des relations de cause à effet, c'est qu'elles véhiculent une certaine conception (ou, à tout le moins, un critère de reconnaissance) de la causalité. En mettant cette conception en évidence, puis en la comparant aux théories probabilistes de la causalité, nous espérons expliquer que les méthodes RB permettent d'induire des relations de cause à effet. Il s'agit donc ici, plus précisément, de mettre au jour une condition qui est nécessaire et suffisante pour que « X cause Y » soit reconnu comme vrai dans le cadre de l'inférence causale probabiliste menée selon les méthodes RB.

Deux voies semblent s'offrir à nous pour faire apparaître la conception de la causalité que mettent en œuvre les méthodes RB pour l'inférence causale probabiliste. D'un côté, on peut se concentrer sur les algorithmes évoqués dans la sous-section 3.1.2 et décrits dans la sous-section 3.2.3, et s'attacher à déterminer à quelle(s) condition(s) ils ont pour résultat un graphe dans lequel figure une flèche partant d'une variable A et dirigée vers une variable B de l'ensemble de variables considéré. De l'autre côté, on peut revenir à l'hypothèse concernant le rapport entre causalité et probabilités sur laquelle reposent les méthodes récentes, et s'intéresser à la contre-partie probabiliste d'une flèche dans un graphe qui entretient le rapport R avec la fonction de probabilités sur l'ensemble de variables considéré.

En vue de mieux qualifier l'alternative que nous venons de décrire, rappelons que les algorithmes mobilisés dans le cadre des méthodes RB ne donnent pas pour résultat un graphe, mais un patron représentant la classe des graphes orientés acycliques qui entretiennent le rapport R avec la fonction de probabilités considérée. Les flèches qui figurent dans un tel patron causal représentent celles des relations de cause à effet directes qu'il est possible d'inférer de la fonction de probabilités. Cela conduit à privilégier la seconde des deux voies d'analyse que nous venons de distinguer. En effet, cela implique que les conditions que nous pourrons mettre en évidence en nous concentrant sur les algorithmes d'inférence causale sont des conditions *suffisantes* de causalité. Or, ni le projet qui est le nôtre dans la présente section, ni la voie qui conduit à examiner l'hypothèse concernant le rapport entre causalité et probabilités sur laquelle les méthodes RB reposent, n'impose une telle focalisation. D'ailleurs, c'est précisément sur le point des conditions *nécessaires* de causalité que la comparaison avec les théories probabilistes de la causalité s'avérera la plus instructive. En outre, il nous semble qu'emprunter la voie qui consiste à s'intéresser directement aux algorithmes d'inférence causale risque de rendre délicate la distinction entre ce qui relève de la conception

de la causalité sur laquelle l'inférence aux causes repose et qui constitue l'objet spécifique de notre analyse – et ce qui relève de la méthodologie de l'inférence causale. À l'inverse, la distinction sera maintenue sans effort si nous choisissons d'analyser l'hypothèse même sur laquelle repose l'inférence causale probabiliste menée selon les méthodes RB.

Les méthodes RB reposent sur l'hypothèse selon laquelle graphes causaux et fonctions de probabilités entretiennent le rapport R. Rappelons qu'un graphe orienté acyclique G et une fonction de probabilités p, tous deux définis sur l'ensemble de variables $\mathbf{V}$:

- satisfont la première clause de la définition du rapport R si et seulement si toute variable de $\mathbf{V}$ est indépendante, relativement à l'ensemble de ses parents dans G, de toutes les variables de $\mathbf{V}$ qui n'appartiennent pas à l'ensemble de ses descendants dans G. C'est la condition de Markov ;
- satisfont la seconde clause de la définition du rapport R si et seulement si toutes les indépendances probabilistes relatives au sein de $\mathbf{V}$ sont impliquées par la première clause. C'est l'hypothèse de fidélité.

Les définitions de la condition de Markov causale et de l'hypothèse de fidélité causale impliquent que si une variable X d'un ensemble $\mathbf{V}$ qui satisfait la condition de Markov causale *ne* cause *pas* une variable Y du même ensemble, alors X et Y sont indépendantes relativement à l'ensemble des causes directes de X dans $\mathbf{V}$. Par ailleurs, si $\mathbf{V}$ satisfait également l'hypothèse de fidélité, alors *seules* les variables Y qui ne sont pas des effets de X sont indépendantes de X relativement à l'ensemble de ses causes directes dans $\mathbf{V}$. Dans ce cas, que X et Y soient indépendantes relativement à l'ensemble des causes directes de X dans $\mathbf{V}$ est à la fois une condition nécessaire et une condition suffisante pour que X ne cause pas Y. Il en découle le résultat suivant :

Théorème 3.8 *Soit* $\mathbf{V}$ *un ensemble de variables, et soit* X *et* Y *deux variables de* $\mathbf{V}$.
Si $\mathbf{V}$ *satisfait la condition de Markov causale et l'hypothèse de fidélité causale, alors* X *cause* Y *dans* $\mathbf{V}$ *si et seulement si* X *et* Y *sont dépendantes relativement à l'ensemble des causes directes de* X *dans* $\mathbf{V}$.

Le théorème 3.8 suit immédiatement des définitions de la condition de Markov causale et de l'hypothèse de fidélité causale. Surtout, il implique que la conception de la causalité que véhiculent les méthodes RB – ce que nous appellerons la « conception RB » – est la suivante :

Proposition 3.9 (Conception RB de la causalité) X cause Y dans $\mathbf{V}$ si et seulement si X et Y sont dépendantes relativement à l'ensemble des causes directes de X dans $\mathbf{V}$.

C'est cette conception de la causalité qu'il nous faut comparer avec les théories probabilistes.

En vue de cette comparaison, il nous sera utile d'avoir établi ceci :

Théorème 3.10 (Condition nécessaire de causalité) *Soit* $\mathbf{V}$ *un ensemble de variables qui satisfait la condition de Markov causale et l'hypothèse de fidélité causale, et soit* X *et* Y *deux variables de* $\mathbf{V}$.
Si X *cause* Y, *alors* X *et* Y *sont dépendantes en probabilité (relativement à l'ensemble vide).*

Preuve*. Soit $\mathbf{V}$ un ensemble de variables qui satisfait la condition de Markov causale et l'hypothèse de fidélité causale, et soit GC le graphe causal sur $\mathbf{V}$, p la fonction de probabilités sur $\mathbf{V}$, et X et Y deux variables de $\mathbf{V}$ telles que X cause Y.
On remarque que $\{X\}$ et $\{Y\}$ ne sont pas d-séparés par l'ensemble vide dans le graphe causal sur $\mathbf{V}$. En effet, le chemin de GC qui correspond à ceci que X est une cause de Y ne satisfait aucune des deux conditions énoncées dans la définition 3.3.
Dans ces conditions, le théorème 3.4 implique qu'il existe une fonction de probabilités représentée par GC pour laquelle X et Y sont dépendants. Autrement dit, la satisfaction de la condition de Markov causale par $\mathbf{V}$ n'implique pas l'indépendance de X et de Y.
Or, nous avons supposé que $\mathbf{V}$ satisfait non seulement la condition de Markov causale, mais encore l'hypothèse de fidélité causale : toutes les indépendances probabilistes au sein de $\mathbf{V}$ et relativement à des variables de $\mathbf{V}$ sont impliquées par la condition de Markov causale. Il en découle que X et Y sont dépendants pour p. □

Ainsi, sous la conception RB, la dépendance probabiliste relative à l'ensemble vide (c'est-à-dire la dépendance probabiliste absolue) est une condition nécessaire de causalité. Signalons que cette condition n'est pas suffisante, et venons-en à l'examen du rapport que la conception RB entretient avec les théories probabilistes de la causalité.

3.4.2 Conception RB et théories probabilistes de la causalité

Comparaison des objets

Si la conception RB et les théories probabilistes de la causalité ont un air de famille, elles ne sont pas immédiatement commensurables. En effet, elles n'ont pas exactement le même objet. Plus précisément, les relations de cause à effet visées par les inférences qui se conforment aux méthodes RB sont des relations entre variables, et relatives à un ensemble de variables. À l'inverse, les relations de cause à effet que les théories probabilistes visent à analyser sont des relations entre propriétés et qui valent de manière absolue.

Nous avons déjà abordé la différence que nous pointons ici. En effet, la procédure d'inférence causale qui fait l'objet du chapitre 2 se distingue elle aussi des théories probabilistes de la causalité par ce qu'elle prend pour objet la causalité entre variables, et relativement à un ensemble de variables. Nous avons montré que ces caractéristiques rendent l'objet de l'inférence causale probabiliste plus grossier que la notion de causalité visée par les théories probabilistes. Nous avons indiqué que cette conclusion peut être nuancée pour ce qui concerne l'inférence causale probabiliste hypothético-déductive, si l'on prend en compte le fait que la modélisation structurelle va de pair avec une quantification des relations de cause à effet qui permet généralement de distinguer quelle(s) propriété(s) cause(nt) quelle(s) autre(s) propriété(s). Tel n'est pas le cas pour ce qui concerne l'inférence causale probabiliste menée selon les méthodes RB. Le fait que les *relata* de la causalité sont des variables a alors d'importantes conséquences pratiques. En effet, ce qu'on perd d'information en représentant les propriétés au moyen de variables, et qu'on ne peut pas recouvrer par la quantification dans le cadre des méthodes RB, est souvent un guide indispensable pour l'action. Est-ce fumer ou ne pas fumer qui préserve du cancer ? Faut-il arroser la plante ou ne pas l'arroser si l'on vise sa survie ? Des flèches causales entre les variables correspondantes ne nous le disent pas. Le prix à payer pour la possibilité de l'inférence causale probabiliste que constitue le fait que l'objet de l'inférence est la causalité entre variables est donc plus élevé dans le cadre des méthodes RB que dans le cadre de la modélisation structurelle.

En vue, maintenant, de comparer l'analyse de la causalité qui sous-tend les méthodes RB aux théories probabilistes de la causalité – et non plus seulement de comparer son objet à celui des théories probabilistes de la causalité –, il nous faut rendre ces objets commensurables. Pour cela, nous proposons de définir un objet commun aux théories probabilistes et à l'analyse de la causalité qui sous-tend les méthodes RB pour l'inférence causale probabiliste. Dans cette optique, rappelons en premier lieu que dans la sous-section 2.3.2 nous avons accepté l'idée selon laquelle les relations de cause à effet absolues peuvent être considérées comme relatives à un ensemble de variables qui suffit à décrire la réalité empirique. En second lieu, on peut considérer qu'une variable représente un ensemble de propriétés, chaque propriété de l'ensemble étant représentée par une valeur de la variable. Surtout, il existe un rapport entre la causalité entre variables et la causalité entre propriétés : une variable X cause une variable Y si et seulement s'il existe une valeur x de X et une valeur y de Y telles que la propriété qui correspond à x cause la propriété qui correspond à y. Il en découle que l'énoncé « X cause Y » est équivalent à la disjonction des énoncés de la forme « La propriété qui est représentée par la valeur x_i de X cause la propriété qui est représentée par la valeur y_j de Y », où x_i et y_j sont des valeurs possibles de X et de Y.

Considérons maintenant deux propriétés A et B, et notons V_A et V_B les variables binaires qui les représentent : V_A peut prendre les deux valeurs A et $\neg A$, et V_B peut prendre les deux valeurs B et $\neg B$. Nous venons de voir que « V_A cause V_B » est équivalent à la disjonction des énoncés de la forme « La propriété qui est représentée par v_A cause la propriété qui est représentée par v_B » où v_A est une valeur de V_A et v_B est une valeur de V_B. Cela signifie que « V_A cause V_B » est équivalent à la disjonction « A cause B, ou A cause $\neg B$, ou $\neg A$ cause B, ou $\neg A$ cause $\neg B$ ». Par ailleurs, nous avons accepté que V_A cause V_B si et seulement si V_A cause V_B relativement à un ensemble de variables qui suffit à décrire la réalité empirique. Dès lors, « V_A cause V_B relativement à un ensemble de variables qui suffit à décrire la réalité empirique » est équivalent à « A cause B, ou A cause $\neg B$, ou $\neg A$ cause B, ou $\neg A$ cause $\neg B$ ». Le premier terme de cette équivalence est un objet pour la conception RB, tandis que son second terme est un objet d'analyse pour les théories probabilistes de la causalité.

Comparaison des analyses

Commençons ici par adopter deux conventions lexicales. D'une part, nous abrégerons la disjonction « A cause B, ou A cause $\neg B$, ou $\neg A$ cause B, ou $\neg A$ cause $\neg B$ », en « $(\neg)A$ cause $(\neg)B$ ». D'autre part, nous ferons l'économie de la précision selon laquelle la causalité est relative à un ensemble de variables qui suffit à décrire la réalité empirique, même quand cette précision est, à la rigueur, nécessaire.

Stratégie Les analyses de « $(\neg)A$ cause $(\neg)B$ » par les théories probabilistes et la caractérisation RB de « V_A cause V_B » peuvent être comparées selon plusieurs critères. Parmi ceux-ci, on distingue en particulier la forme de l'analyse[6], la stratégie utilisée pour tenir compte de tel ou tel type de contre-exemples à l'idée de caractériser une cause par ce qu'elle rend ses effets plus probables, ou plus simplement ceux de ces contre-exemples que l'analyse prend en charge.

Nous retenons le dernier de ces critères, pour trois raisons non indépendantes. D'abord, les contre-exemples à une analyse de « $(\neg)A$ cause $(\neg)B$ » sont les mêmes, moyennant la disjonction et éventuellement la négation, que les contre-exemples pour l'analyse correspondante de « A cause B ». Or, des contre-exemples aux analyses proposées pour « A cause B » ont gouverné le développement réel et, en conséquence, notre présentation des théories probabilistes de la causalité. Dès lors, il apparaît naturel de comparer les théories probabilistes et la conception RB sur le

[6] Selon ce critère, la caractérisation RB semble plus proche des théories postérieures à celle de Suppes que de la théorie de Suppes.

critère des contre-exemples. Ensuite, l'étendue du domaine au sein duquel elle est correcte – ou, en d'autres termes, les types de contre-exemples qu'elle prend en charge – est ce qui permet d'évaluer une analyse (et non seulement de la comparer à une autre), et ce qu'on en retient *in fine*. Enfin, procéder à une comparaison selon le critère des contre-exemples à l'idée fondatrice qui sont pris en charge nous permettra d'expliquer pourquoi les méthodes RB permettent d'induire des relations de cause à effet à partir de connaissances probabilistes.

En choisissant ce critère pour la comparaison de la conception RB et des théories probabilistes de la causalité, nous choisissons aussi une façon d'organiser la comparaison. Plus précisément, le troisième critère ayant été choisi, il semble naturel de reprendre chacun à son tour les types de contre-exemples que nous avons discutés dans le chapitre 1. Pour chacun de ces types, on déterminera si la conception RB de la causalité permet de le prendre en charge de manière satisfaisante. Pour une plus grande clarté de l'analyse, nous *ne* reprenons *pas* l'ordre dans lequel les types de contre-exemples ont été introduits dans le chapitre 1. De cette analyse, nous attendons donc qu'elle détermine quelle est la place qu'occupe la conception RB au sein du champ des analyses, par les théories probabilistes, de « $(\neg)A$ cause $(\neg)B$ ».

Indépendances trompeuses Dans la sous-section 3.4.1, nous avons établi qu'une condition nécessaire pour qu'une variable X cause une variable Y selon la condition RB est que X et Y soient dépendantes en probabilité (théorème 3.10). En particulier, une variable V_A cause une variable V_B seulement si V_A est dépendante de V_B. Or, sous la signification que nous avons donnée plus haut à la notation V_P, P étant une propriété, on a le résultat suivant :

Théorème 3.11 (Dépendance probabiliste entre variables et entre propriétés) *Soit A et B deux propriétés.*
V_A est dépendante de V_B si et seulement si A est dépendante de B.

Preuve*. Soit A et B deux propriétés.
Si A est dépendante de B, alors une valeur de V_A (en l'occurrence, A) est dépendante d'une valeur de V_B (en l'occurrence, B) et, par la définition de l'indépendance probabiliste entre variables, V_A est dépendante de V_B.

Supposons maintenant que V_A est dépendante de V_B. Quatre cas, non exclusifs, sont possibles :

1. A est dépendante de B – c'est-à-dire qu'on a immédiatement la conclusion recherchée ;
2. A est dépendante de $\neg B$.
   On a alors $p(A|\neg B) \neq p(A)$.

Par le théorème des probabilités totales, on a :
$p(A) = p(A|\neg B).p(\neg B) + p(A|B).p(B)$.
L'inégalité $p(A|\neg B) \neq p(A)$ implique alors : $p(A) \neq p(A).p(\neg B) + p(A|B).p(B)$, soit :
$p(A) \neq p(A).p(\neg B) + p(B|A).p(A)$, et enfin :
$p(A) \neq p(A)[p(\neg B) + p(B|A)]$.
En divisant par $p(A)$ [7], il vient :
$1 \neq p(\neg B) + p(B|A)$, soit :
$p(B) \neq p(B|A)$, qui est une expression de la dépendance de A et B ;

3. $\neg A$ est dépendante de B.
On a alors $p(\neg A|B) \neq p(\neg A)$, et la dépendance de A et B découle immédiatement de ce que $p(\neg A|B) = 1 - p(A|B)$ et $p(\neg A) = 1 - p(A)$;

4. $\neg A$ est dépendante de $\neg B$.
On a alors $p(\neg A|\neg B) \neq p(\neg A)$. Pour des raisons analogues à celles qui ont été évoquées pour le cas 3, il en découle que $p(A|\neg B) \neq p(A)$. Or, on a montré pour le cas 2 que cette inégalité implique la dépendance de A et de B

Nous avons ainsi établi que V_A est dépendante de V_B si et seulement si A est dépendante de B. □

Des théorèmes 3.10 et 3.11 pris ensemble, il découle que, sous la conception RB, « $(\neg)A$ cause $(\neg)B$ » n'est vrai que si A et B sont dépendantes en probabilité. Sous cette conception, il n'y a pas de relation de cause à effet sans dépendance probabiliste. La conception RB échoue donc à prendre en compte la possibilité qu'existent des indépendances trompeuses. Cela vaut pour les deux types d'indépendances trompeuses que nous avons identifiés dans la sous-section 1.3.5 : d'une part, les indépendances trompeuses qui découlent de ce que les effets d'une propriété peuvent être contre-balancés par ceux d'une autre propriété à laquelle la première est corrélée – c'est-à-dire les indépendances trompeuses qui peuvent être considérées comme des instances du paradoxe de Simpson (exemple du tabagisme et de la pratique régulière d'un exercice physique) ; d'autre part, les indépendances trompeuses qui découlent de l'existence de deux routes causales, l'une positive et l'autre négative, menant d'une propriété à une autre (exemple du bon état des routes et de la faible mortalité sur la route). Dans ces conditions, la condition nécessaire énoncée au titre du théorème 3.10 est à la conception RB de la causalité ce que sa clause 1 est à la théorie 1.2 développée par Suppes.

[7] Rappelons que nous ne considérons que des propriétés de probabilité non nulle, en particulier parce qu'il n'y aurait pas de sens clair à parler d'une cause ou d'un effet de probabilité nulle.

Revenons, maintenant, sur les théories probabilistes de la causalité postérieures à celle de Suppes. Nous avons montré que, parmi les difficultés auxquelles se heurte la théorie de Suppes, les seules sur lesquelles achoppent les théories de Cartwright (théorie 1.3) et de Skyrms (théorie 1.4) concernent, pour l'une, les indépendances trompeuses découlant de l'existence de deux routes causales menant d'une propriété à une autre, et, pour l'autre, les causes interactives. En outre, les théories encore ultérieures sont strictement plus satisfaisantes que les théories 1.3 et 1.4, au sens où l'ensemble de contre-exemples à l'idée fondatrice qu'elles traitent correctement inclut strictement l'ensemble des contre-exemples pris en compte par les théories 1.3 et 1.4. Dès lors, ce que nous avons établi dans le dernier paragraphe suffit à établir pour le cas non interactif que l'analyse RB de « $(\neg)A$ cause $(\neg)B$ » est strictement moins satisfaisante que l'analyse du même énoncé par n'importe laquelle d'entre les théories probabilistes de la causalité postérieures à celle de Suppes. Dans ces conditions, nous nous tournons maintenant vers les corrélations trompeuses qui découlent de l'existence de causes interactives. Si la conception RB ne prend pas en charge les contre-exemples à l'analyse fondatrice qu'elles constituent, alors elle est strictement moins satisfaisante que toutes les théories probabilistes de la causalité qui sont postérieures à celle de Suppes.

Corrélations trompeuses avec causes interactives Ici comme dans la section 1.3, nous nous en tenons, pour ce qui est des corrélations trompeuses induites par des causes interactives, aux corrélations entre effets communs à *une* cause interactive, c'est-à-dire à une cause qui échoue à faire écran entre eux (exemple des boules de billard donné par Salmon). Nous supposons, en outre, que chacun des deux effets a cette cause commune pour seule cause. Les situations ainsi caractérisées sont les plus simples parmi celles pour lesquelles le caractère interactif fait problème, mais, en même temps, elles posent déjà un problème substantiel pour l'analyse probabiliste de la causalité.

Ces corrélations trompeuses sont-elles correctement traitées dans le cadre des méthodes RB pour l'inférence causale probabiliste ? En vue de répondre à cette question, commençons par remarquer que le théorème 3.11 peut être généralisé :

Théorème 3.12 (Généralisation de 3.11) *Soit A, B deux propriétés et $\mathbf{C}$ un ensemble de propriétés.*
V_A est dépendante de V_B relativement à l'ensemble $\mathbf{V_C}$ des variables qui représentent les propriétés de $\mathbf{C}$ si et seulement si A est dépendante de B relativement à $\mathbf{C}$.

Ce résultat admet une preuve analogue à celle que nous avons donnée pour le théorème 3.11.

Considérons maintenant trois propriétés A, B et C telles que C est une cause interactive commune à A et à B qui induit une corrélation trompeuse entre elles. Dans ce cas, la corrélation entre A et B étant trompeuse (par hypothèse), la proposition « $(\neg)A$ cause $(\neg)B$ » est fausse. La question, alors, est de savoir si la conception RB conduit bien à cette conclusion ou, de façon équivalente, si elle conduit bien à conclure que V_A ne cause pas V_B. Pour répondre à cette question, rappelons que la conception RB a la conséquence suivante : V_A cause V_B si et seulement si les deux variables sont dépendantes relativement à l'ensemble des causes directes de V_A. Dans le cas qui nous occupe, par hypothèse sur la situation, V_A a une seule cause directe : V_C. Dans ces conditions, la question est de savoir si V_A est dépendante de V_B relativement à V_C. Par le théorème 3.12, V_A est dépendante de V_B relativement à V_C si et seulement si A est dépendante de B relativement à C. Or, tel est bien le cas, par hypothèse sur la situation. La conception RB conduit à la conclusion erronée que V_A cause V_B, c'est-à-dire que « $(\neg)A$ cause $(\neg)B$ » est vraie.

La conception RB ne permet pas de traiter correctement les cas de corrélation trompeuse entre effets d'une même cause qui ne fait pas écran entre eux. Ainsi que nous l'avons annoncé, il en découle que la conception RB se heurte à strictement plus de contre-exemples que n'importe laquelle des théories probabilistes de la causalité développées après la théorie de Suppes. Dès lors, c'est avec la théorie proposée par Suppes qu'il convient maintenant de comparer la conception RB. Nous nous attachons d'abord à déterminer si cette conception traite correctement au moins autant de cas que la théorie de Suppes. En d'autres termes, nous déterminons si elle permet d'identifier comme trompeuses les corrélations trompeuses identifiées comme telles par la théorie de Suppes. Il est apparu dans le chapitre 1 que ces corrélations trompeuses sont précisément des corrélations entre effets communs à une cause qui n'est pas interactive.

Corrélations trompeuses entre effets d'une cause non interactive Envisageons une corrélation trompeuse entre deux effets A et B d'une propriété C qui fait écran entre eux. À titre de rappel, A et B pourraient être les propriétés de développer un cancer du poumon et d'avoir les doigts jaunis, et C la propriété d'être fumeur. Dans une telle situation, l'énoncé « $(\neg)A$ cause $(\neg)B$ » est faux, et encore une fois la question est de savoir si la conception RB conduit bien à cette conclusion. En termes de variables, la question est de savoir si la conception RB conduit bien à la conclusion selon laquelle V_A ne cause pas V_B.

Sous la conception RB (proposition 3.9), V_A cause V_B si et seulement si V_A et V_B sont dépendantes relativement à l'ensemble des causes directes de V_A. Si C est une cause de A, alors parmi les causes directes de V_A figure soit V_C, soit un effet I de V_C. Dans le premier cas, l'hypothèse selon laquelle C fait écran à la dépendance entre A et B implique que V_A

et V_B sont indépendantes relativement à l'ensemble des causes directes de V_A, auquel appartient V_C. Dans le second cas, il faut utiliser le théorème des probabilités totales pour obtenir la même conclusion :

Preuve*. Soit $\mathbf{CD}_A$ l'ensemble des causes directes de V_A qui sont différentes de I. Soit aussi v_A une valeur de V_A, i une valeur de I, v_B une valeur de V_B et $\mathbf{cd}_A$ une valeur de $\mathbf{CD}_A$.
Par le théorème des probabilités totales, on a :

$$p(v_A|v_B \wedge i \wedge \mathbf{cd}_A) = p(v_A|v_B \wedge i \wedge \mathbf{cd}_A \wedge (V_C = C)).p(V_C = C)+ \\ p(v_A|v_B \wedge i \wedge \mathbf{cd}_A \wedge (V_C = \neg C)).p(V_C = \neg C).$$

Par hypothèse, A et B sont indépendantes relativement à C. Il en découle, par le théorème 3.12, que V_A et V_B sont indépendantes relativement à V_C.
En conséquence, pour toute valeur v_C de V_C, on a :
$p(v_A|v_B \wedge i \wedge \mathbf{cd}_A \wedge v_C) = p(v_A|i \wedge \mathbf{cd}_A \wedge v_C)$.
Du coup, on a :

$$p(v_A|v_B \wedge i \wedge \mathbf{cd}_A) = p(v_A|i \wedge \mathbf{cd}_A \wedge (V_C = C)).p(V_C = C)+ \\ p(v_A|i \wedge \mathbf{cd}_A \wedge (V_C = \neg C)).p(V_C = \neg C).$$

En utilisant à nouveau le théorème des probabilités totales, il vient :
$p(v_A|v_B \wedge i \wedge \mathbf{cd}_A) = p(v_A|i \wedge \mathbf{cd}_A)$.
En conséquence, V_A et V_B sont indépendantes relativement à l'ensemble des causes directes de V_A. □

Ainsi, V_A est indépendante de V_B relativement à l'ensemble des causes directes de V_A dans tous les cas – c'est-à-dire que V_C soit ou non l'une de ces causes *directes*. Dès lors, V_A ne cause pas V_B selon la conception RB de la causalité.

De façon plus générale, il apparaît que la conception RB est adéquate pour le type de cas qui fait l'objet du présent paragraphe et, surtout, qu'elle prend correctement en charge tous ceux des contre-exemples à l'idée fondatrice que la théorie de Suppes traite correctement. Reste à déterminer si la conception RB prend en charge certains des contre-exemples sur lesquels la théorie de Suppes achoppe, ou si au contraire elle se heurte exactement aux difficultés que nous avons mises en évidence plus haut pour la théorie de Suppes (section 1.3). Celles de ces difficultés qui n'impliquent pas de cause interactive correspondent aux corrélations trompeuses entre effets de plusieurs causes d'abord, aux corrélations entre effets et causes ensuite, et au paradoxe de Simpson enfin.

Corrélations trompeuses entre effets de plusieurs causes Nous avons vu plus haut (sous-section 1.3.3) que ces corrélations sont de deux sous-types : d'une part, les corrélations entre des effets qui ont *plusieurs*

causes en commun ; d'autre part, les corrélations entre deux variables causalement indépendantes et dont l'une a plusieurs causes directes dont aucune ne suffit à déterminer la valeur qu'elle prend.

La distinction entre ces deux sous-types n'est pas importante ici. En effet, nous avons mis en évidence que la difficulté rencontrée par la théorie de Suppes découlait en fait de ce qu'elle n'identifiait une corrélation entre deux propriétés comme trompeuse que si *une* troisième propriété suffisait à faire écran entre les deux premières. Or, la difficulté ainsi décrite est clairement surmontée dans le cadre de la conception RB : selon cette conception, une relation de cause à effet est une relation de dépendance probabiliste qui résiste à la conditionalisation par l'ensemble de *toutes* les causes directes de la cause (théorème 3.8). La conception RB permet donc de traiter correctement ce premier type de cas non interactifs posant problème pour la théorie de Suppes.

Corrélations trompeuses entre effets et causes De même que les corrélations trompeuses entre effets de plusieurs causes, les corrélations trompeuses entre les effets et leurs causes posent problème dans le cadre de la théorie de Suppes. Plus précisément, la solution apportée par Suppes au problème constitué par les corrélations trompeuses entre effets et causes n'est pas satisfaisante. Pour commencer, il s'agit d'une solution temporelle dont le sens n'est pas complètement clair pour ce qui concerne la causalité *générique*. Surtout, nous avons vu qu'il n'était pas souhaitable de réduire l'ordre causal à l'ordre temporel.

Pas plus que les théories probabilistes postérieures à celle de Suppes, la conception RB ne conduit à une solution temporelle du problème des corrélations trompeuses entre effets et causes. En effet, la proposition 3.9 ne fait aucune référence à des instants du temps. Pourtant, le problème est résolu, puisque la proposition 3.9 n'implique pas que la relation de causalité est symétrique. En effet, deux variables peuvent être dépendantes relativement à l'ensemble des causes directes de l'une sans être dépendantes relativement à l'ensemble des causes directes de l'autre. Bien sûr, cela n'implique pas que les méthodes d'inférence causale probabiliste fondées sur les réseaux bayésiens identifient toujours correctement, pour deux variables dont l'une cause l'autre, laquelle est la cause et laquelle est l'effet.

Paradoxe de Simpson Il nous reste à déterminer si la conception RB de la causalité traite correctement les corrélations trompeuses qui sont des instances du paradoxe de Simpson, c'est-à-dire les corrélations dont le *sens* est trompeur[8]. Pour cela, revenons aux termes de la comparaison

[8] Le paradoxe de Simpson implique également des indépendances trompeuses, correspondant à ceci que les effets d'une propriété peuvent être contre-balancés par les effets

que nous menons : d'une part, l'analyse de « $(\neg)A$ cause $(\neg)B$ » par les théories probabilistes ; d'autre part, celle de « V_A cause V_B » dans le cadre de la conception PR. Par ailleurs, rappelons que le paradoxe de Simpson soulève en fait le problème de l'identification de celle d'une propriété ou de sa négation qui doit compter comme cause d'une troisième propriété. Le paradoxe de Simpson, et avec lui les corrélations trompeuses qu'il implique, ne peut donc s'énoncer qu'en termes de propriétés. Dans ces conditions, et parce que la conception RB est une conception de la causalité entre variables et non de la causalité en termes de propriétés, le paradoxe de Simpson ne *peut* pas être énoncé relativement à la conception RB de la causalité. Plus précisément, l'objet de cette conception est trop grossier pour qu'elle soit sensible aux distinctions engagées par le paradoxe de Simpson. Le paradoxe de Simpson ne peut donc pas même être un problème pour la conception RB de la causalité.

Ainsi, non seulement les contre-exemples à l'idée fondatrice pris en charge par la théorie de Suppes le sont également par la conception RB, mais encore celle-ci traite correctement certains des contre-exemples sur lesquels celle-là achoppe : les corrélations trompeuses entre effets de plusieurs causes, d'une part, et, d'autre part, les corrélations trompeuses entre effets et causes. De l'autre côté, il est apparu que la conception RB achoppe sur strictement plus de contre-exemples à l'idée fondatrice que n'importe laquelle des théories probabilistes postérieures à celle de Suppes. Dans ces conditions, l'analyse RB de l'énoncé « V_A cause V_B » (qui, rappelons-le, doit être compris comme relatif à un ensemble de variables qui suffit à décrire la réalité physique) prend place juste après, mais strictement après, la théorie 1.2 de Suppes dans l'histoire des analyses de « A cause B » telle que nous l'avons rationnellement reconstruite dans le chapitre 1.

Cela étant montré, nous pouvons ouvrir la sous-section conclusive dans laquelle nous expliquerons que les méthodes RB permettent d'induire des connaissances causales à partir de connaissances probabilistes alors que nos meilleures analyses du rapport entre les concepts de causalité et de probabilité, les théories probabilistes de la causalité, ne peuvent pas servir de principes à des inférences de ce type.

3.4.3 Conclusion

Dans la sous-section 1.5.1, nous avons mis en évidence deux caractéristiques des théories probabilistes de la causalité qui ont pour conséquence que ces théories ne peuvent pas être transposées directement du domaine de l'analyse conceptuelle à celui de l'épistémologie. D'une part, les théories probabilistes de la causalité sont circulaires ; d'autre part,

d'une autre propriété à laquelle elle est corrélée. Ces cas ont été envisagés plus haut dans la sous-section, dans le paragraphe consacré aux indépendances trompeuses.

elles font de la conditionalisation sur *toutes* les causes de B qui sont indépendantes de A, un ingrédient de l'analyse de « A cause B ». Nous revenons ici sur ces deux aspects successivement et dans cet ordre, et nous montrons comment les obstacles à la reconnaissance des causes qui leur correspondent sont contournés dans le cadre des méthodes RB pour l'inférence causale probabiliste.

Commençons, donc, par la circularité des théories probabilistes de la causalité. Il s'agit plus exactement de ceci que celles de ces théories qui sont postérieures à la théorie de Suppes sont telles que la notion de cause figure dans l'analyse de « A cause B ». Nous avons indiqué dans la section 1.4 que la circularité des théories probabilistes apparaît afin de traiter correctement le paradoxe de Simpson, à la fois en tant qu'il donne lieu à des indépendances trompeuses et en tant qu'il donne lieu à des corrélations trompeuses. Or, nous venons de montrer que, en tant qu'il donne lieu à des *corrélations* trompeuses, le paradoxe de Simpson ne peut pas être formulé pour ce qui est de la causalité entre variables et, surtout, que la conception RB ne permet en aucun cas de traiter correctement les situations engageant des indépendances trompeuses (de quelque type qu'elles soient). En conséquence, il n'y a pas de raison pour que la conception RB de la causalité se heurte aux difficultés qui découlent, pour les théories probabilistes, de leur caractère circulaire.

Positivement, il est bien vrai que la conception RB de la causalité n'est pas circulaire à proprement parler. Selon la proposition 3.9, en effet, c'est la notion de cause *directe*, et non pas la notion de cause elle-même, qui apparaît dans l'analyse de « A cause B ». Surtout, le théorème 3.10, dont nous avons vu qu'il implique que la conception RB ne traite pas correctement les indépendances trompeuses, implique également que la dépendance probabiliste est une condition nécessaire pour la causalité, et donc *a fortiori* pour la causalité directe. Or, ce fait est fondamental pour les algorithmes qui sont mobilisés dans le cadre des méthodes d'inférence causale RB [9]. Plus précisément, la possibilité d'induire des relations de cause à effet en se conformant aux méthodes RB dépend de manière essentielle de ceci : sous l'hypothèse selon laquelle graphe causal et fonction de probabilités entretiennent le rapport R, deux variables X et Y sont en relation de cause à effet directe dans l'ensemble $\mathbf{V}$ seulement si elles sont dépendantes relativement à tous les sous-ensembles de $\mathbf{V} \setminus \{X, Y\}$ (Spirtes *et al.*, 1993, p. 82, théorème 3.4.).

[9] En particulier, nous avons vu dans la sous-section 3.2.3 que l'algorithme PC de Spirtes, Glymour et Scheines commençait par identifier tous les couples de variables indépendantes pour $n = 0$ (Spirtes *et al.*, 1993, p. 84-85) et conclure que les variables qui composent un tel couple ne sont pas en relation de cause à effet – par où nous entendons qu'aucune ne cause directement l'autre dans l'ensemble de variables considéré.

On pourrait reprendre une démonstration du résultat que nous venons d'énoncer [10] et en produire une analyse qui ferait apparaître la façon dont le résultat articule, exactement, ce que nous avons dit plus haut de la conception RB de la causalité avec la possibilité d'induire des connaissances causales à partir de connaissances probabilistes. Nous ne nous livrons pas ici à une telle analyse, pour deux raisons.

En premier lieu, les conclusions qu'on peut espérer tirer de cette enquête ne sont pas à la hauteur de l'investissement qu'elle représente. En effet, pour les raisons que nous avons exposées au début de la présente sous-section, il nous semble *déjà* clair que, parmi les caractéristiques de la conception RB que nous avons mises en évidence, c'est avec le fait qu'elle ne prend pas en compte les indépendances trompeuses que le résultat auquel nous faisons allusion ici a le plus directement à voir. Conformément à cette analyse, nous considérons que la conséquence de la conception RB qui est énoncée dans le théorème 3.10 est essentielle à la possibilité, ouverte par les méthodes RB, d'induire des relations causales à partir de connaissances probabilistes et, en deçà, de données statistiques d'observation.

En second lieu, nous n'analysons pas la preuve du résultat que nous venons de mentionner parce que ce résultat lui-même ne serait d'aucune utilité pour l'induction causale si la conception RB ne venait pas à bout des conséquences, pour les théories probabilistes de la causalité, de leur seconde caractéristique problématique. Cette caractéristique consiste à faire de la conditionalisation par toutes les causes de B qui sont indépendantes de A, un ingrédient de l'analyse de « A cause B ». En l'absence d'aucune forme de délimitation de l'espace auquel ces causes appartiennent et au sein duquel elles devraient donc être cherchées, il en découle qu'il n'est jamais possible de *savoir* que A cause B si l'on s'en tient aux théories probabilistes.

De même que la procédure hypothético-déductive définie dans le chapitre 2, les méthodes RB contournent cet obstacle en visant un concept de causalité qui est relatif à un ensemble de variables. Que l'inférence causale soit menée relativement à un ensemble de variables donné signifie que, pour toute variable de l'ensemble, on s'intéresse seulement à ses causes *dans cet ensemble*. C'est donc seulement au sein de l'ensemble de variables considéré que doivent être recherchées les causes d'une variable donnée de l'ensemble. Plus spécifiquement, les analyses menées dans la sous-section 3.4.1 et dans la sous-section 3.4.2 font apparaître que, dans le cadre de la conception RB, une relation de cause à effet est une relation de dépendance probabiliste qui résiste à la conditionalisation. Dès lors, l'inférence aux causes devient la recherche de dépendances probabilistes qui ne disparaissent pas par conditionalisation. Dans ces

[10] Une telle démonstration est donnée dans Spirtes *et al.*, 1990.

conditions, ce qui compte, et qui permet de rompre en pratique la circularité, est la possibilité d'envisager *toutes* les dépendances et indépendances probabilistes relatives. C'est exactement ce que permet le caractère relatif de la causalité telle qu'elle est analysée par la proposition 3.9 et visée par les méthodes RB.

Ainsi, la comparaison de la conception RB aux théories probabilistes de la causalité permet de comprendre comment il est possible que les inférences menées selon les méthodes RB soient inductives quand les théories probabilistes ne peuvent être directement utilisées comme principes pour des inductions causales. Plus précisément, la sous-section qui s'achève fait apparaître que les différences concernant l'objet de l'analyse aussi bien que les différences concernant l'analyse elle-même doivent être prises en compte pour expliquer que les méthodes RB autorisent l'induction causale. Nous savons déjà (depuis la sous-section 2.3.2) que l'objet qu'on vise en suivant les méthodes RB, c'est-à-dire la causalité entre variables et relativement à un ensemble de variables, est plus grossier que l'objet des théories probabilistes de la causalité. C'est là une première partie du prix à payer pour la possibilité de l'induction. La seconde partie de ce prix correspond à ceci que la position occupée par la conception RB au sein du champ des analyses probabilistes de la causalité implique que cette conception se heurte à un certain nombre de contre-exemples. Plus précisément, nous avons montré qu'elle ne prend pas en compte la possibilité d'indépendances trompeuses et qu'elle ne traite pas correctement les cas de corrélation trompeuse entre effets d'une même cause interactive. Il en découle que les méthodes RB ne sont pas de nature à assurer un résultat causal correct si le système considéré donne lieu à des indépendances trompeuses ou à des corrélations trompeuses entre effets d'une cause commune interactive. La seconde partie du prix à payer pour la possibilité d'induire des connaissances causales à partir de connaissances probabilistes tirées de données statistiques consiste donc dans le caractère restreint du domaine au sein duquel l'inférence causale est valide. Quelle est, plus exactement, la mesure de cette restriction ? Quelle en est la portée, et qu'en découle-t-il relativement à l'utilisation et à l'utilité des méthodes d'inférence causale probabiliste que nous venons de présenter ? Telles sont les principales questions traitées dans le prochain chapitre.

CHAPITRE 4

Limites et portée de l'utilisation des réseaux bayésiens

Il existe des systèmes pour lesquels les inférences causales probabilistes menées selon les méthodes RB ne sont pas valides. Tel est le cas, notamment, des systèmes au sein desquels certaines indépendances, ou certaines dépendances entre effets d'une même cause interactive, sont trompeuses. Une telle description, toutefois, est abstraite et ne donne une mesure précise ni de la restriction dont l'existence a été établie, ni de la portée qu'elle a dans le contexte d'inférence causale probabiliste qui nous intéresse. Il s'agit maintenant de tracer les limites exactes et, corrélativement, de mesurer la portée de l'utilisation des réseaux bayésiens pour l'inférence causale probabiliste. Nous avons vu que, en utilisant les réseaux bayésiens et, plus précisément, en suivant les méthodes RB, on peut, à partir des indépendances probabilistes relatives sur l'ensemble de variables considéré, induire des relations de cause à effet sur cet ensemble.

De même que dans la sous-section 3.4.1, et pour les mêmes raisons, nous privilégions l'examen des conditions de causalité qui sont véhiculées par les méthodes d'inférence causale probabiliste fondées sur les réseaux bayésiens, à celui des algorithmes qu'elles mobilisent. Toutefois, notre but est maintenant d'analyser autant qu'il est possible, et de critiquer, la conception de la causalité sur laquelle les méthodes RB reposent. Dès lors, nous concentrons notre attention directement sur les hypothèses, relatives à la causalité et aux probabilités, qui constituent cette conception, plutôt que sur cette conception elle-même.

Nous commençons par présenter ces hypothèses et par expliquer en quoi elles sont plausibles (section 4.1) puis, partant de là, nous pouvons mettre évidence les limites des méthodes RB (section 4.2). En effet, ces limites correspondent en premier lieu à l'existence de situations telles que ne sont pas satisfaites les hypothèses que véhiculent ces méthodes. Une fois mis au jour les limites des méthodes RB, nous faisons apparaître leurs conséquences pour l'inférence causale probabiliste (section 4.3). Finalement, nous formulons une proposition méthodologique (section 4.4).

Le lecteur aura compris que, dans ce chapitre conclusif, nous dépassons le cadre que définit à la rigueur la question de savoir pourquoi et comment les méthodes RB permettent d'induire des relations de cause à effet à partir de connaissances probabilistes si les théories probabilistes de la causalité ne peuvent pas servir de principes à de telles inductions. Plus précisément, en achevant de répondre à cette question-ci, nous abordons dans toute sa généralité une autre question : celle de la portée des méthodes RB, c'est-à-dire la question de savoir ce que ces méthodes changent à l'inférence causale probabiliste. Répondre à cette dernière question, c'est en particulier déterminer ce par quoi les méthodes RB diffèrent des méthodes plus traditionnelles qu'elles viennent concurrencer. Dans ces conditions, la comparaison des méthodes RB et des méthodes hypothético-déductives présentées dans le chapitre 2 constitue un thème qui traverse le chapitre. Cette comparaison fait l'objet explicite de l'analyse seulement dans la sous-section 4.2.2 et dans un moment de la sous-section 4.3.2.

4.1 Hypothèses véhiculées par les méthodes RB : mise au jour et plausibilité

4.1.1 Hypothèses concernant la causalité

Étant donné un ensemble de variables, les méthodes RB pour l'inférence causale probabiliste visent à construire un graphe orienté acyclique dont les flèches représentent les relations de cause à effet directes entre les variables de cet ensemble. Dans ces conditions, les méthodes RB reposent, en premier lieu, sur l'hypothèse suivante :

- **hypothèse 1**, ou hypothèse de représentation : les relations de cause à effet directes au sein de l'ensemble de variables considéré peuvent bien être représentées par les flèches qui figurent dans un graphe orienté acyclique défini sur cet ensemble.

L'hypothèse 1 comporte au moins deux sous-hypothèses :

- **sous-hypothèse 1.1** : les *relata* de la causalité peuvent être représentés par des variables.
- **sous-hypothèse 1.2** : en représentant par des flèches les relations de cause à effet directes au sein de l'ensemble de variables considéré, on obtient un graphe acyclique.

Parce qu'elle sous-tend également les méthodes hypothético-déductives qui ont été discutées dans le chapitre 2, la sous-hypothèse 1.1 a déjà été présentée et discutée dans la sous-section 2.3.2. Nous avons alors montré que, si une relation entre variables peut bien représenter une relation entre propriétés, la représentation n'en est pas moins grossière.

La sous-hypothèse 1.2 porte explicitement sur la causalité *directe*. Elle implique que celle-ci est asymétrique. En outre, admettons avec les tenants des méthodes RB que la causalité (tout court) se définit comme la clôture transitive de la causalité directe. Pour le dire autrement, admettons qu'on

peut considérer que A cause B si et seulement s'il existe une chaîne de relations de cause à effet *directes* menant de A à B. Dans ce cas, la sous-hypothèse 1.2 équivaut à l'asymétrie de la causalité. Sous la définition de la causalité comme clôture transitive de la causalité directe, l'acyclicité du graphe représentant la causalité directe est équivalente à l'asymétrie de la causalité.

Preuve*. Soit un graphe orienté G représentant les relations de cause à effet directes sur un ensemble de variables $\mathbf{V}$.
Les propositions suivantes sont alors équivalentes :

- la causalité sur $\mathbf{V}$ n'est pas une relation asymétrique ;
- il existe deux variables V_i et V_j de $\mathbf{V}$ telles que chacune cause l'autre ;
- G contient un sous-graphe qui a la forme du graphe de la figure 4.1 ci-dessous ;
  G n'est pas acyclique.

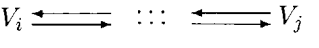


FIGURE 4.1

□

Il nous reste ici à ajouter que, en première approche, la causalité se présente effectivement comme une relation asymétrique : un effet ne cause pas sa cause. Dans ces conditions, la première composante de l'hypothèse de représentation ne semble pas soulever de difficulté particulière. Son bien-fondé découle des propriétés de la causalité elle-même.

4.1.2 Hypothèses concernant le rapport entre la causalité et les probabilités

Outre l'hypothèse de représentation, les méthodes RB reposent sur l'hypothèse selon laquelle la causalité et les probabilités au sein d'un ensemble de variables donné entretiennent un certain rapport. Plus exactement, étant donné un ensemble de variables $\mathbf{V}$ auquel elles sont appliquées, les méthodes RB supposent que le graphe causal GC sur $\mathbf{V}$ et la fonction de probabilités p sur $\mathbf{V}$ entretiennent le rapport R défini dans la sous-section 3.1.1. Ainsi que nous l'avons déjà expliqué, cela revient à émettre les deux hypothèses suivantes :

- **hypothèse 2**, ou condition de Markov causale : toute variable V de $\mathbf{V}$ est indépendante, relativement à ses causes directes dans $\mathbf{V}$, de toutes les variables de $\mathbf{V}$ qu'elle ne cause pas ;
- **hypothèse 3**, ou hypothèse de fidélité causale : toutes les indépendances probabilistes entre des sous-ensembles de $\mathbf{V}$ relativement à des sous-ensembles de $\mathbf{V}$ sont impliquées par l'hypothèse 2.

Nous avons déjà indiqué que si la condition de Markov causale est satisfaite, alors la structure causale fait peser un grand nombre de contraintes

sur la structure probabiliste – ou, plus exactement, sur la structure des indépendances probabilistes. Nous ne voyons pas comment construire une typologie exhaustive des indépendances probabilistes qui sont imposées par la structure causale dans le cas où la condition de Markov causale est satisfaite. Dans ces conditions, nous concentrons notre attention sur les indépendances probabilistes auxquelles la condition de Markov causale fait *explicitement* référence.

Étant donné un ensemble de variables **V**, la lettre de la condition de Markov causale veut que toute variable V de **V** soit indépendante, relativement à l'ensemble de ses causes directes dans **V**, des variables de **V** qu'elle ne cause pas. Or, une variable de **V** non causée par V peut être en particulier :

- une cause directe de V dans **V** ;
- une variable qui n'a aucune sorte de rapport causal avec V ;
- une cause de V qui n'est pas une cause *directe* de V dans **V** ;
- un effet d'une cause directe de V qui n'est pas également un effet de V.

En conséquence, l'hypothèse selon laquelle **V** satisfait la condition de Markov causale implique les quatre sous-hypothèses suivantes :

- **sous-hypothèse 2.0** : toute variable de **V** est indépendante, relativement à l'ensemble de ses causes directes dans **V**, de ses causes directes dans **V** ;
- **sous-hypothèse 2.1** : toute variable de **V** est indépendante, relativement à l'ensemble de ses causes directes dans **V**, de toute variable avec laquelle elle n'entretient aucun type de rapport causal ;
- **sous-hypothèse 2.2** : toute variable de **V** est indépendante, relativement à l'ensemble de ses causes directes dans **V**, de ses causes indirectes dans **V** ;
- **sous-hypothèse 2.3** : toute variable de **V** est indépendante, relativement à l'ensemble de ses causes directes dans **V**, des effets de ces causes qu'elle ne cause pas.

La sous-hypothèse 2.0 est analytique, au sens où la définition des probabilités conditionnelles implique qu'elle est satisfaite. En effet, on montre facilement la proposition suivante :

Proposition 4.1 Soit p une fonction de probabilités sur un ensemble de variables **V**.
Pour tout $\mathbf{W} \subset \mathbf{V}$, tout $W \in \mathbf{W}$ et tout $V \in \mathbf{V}$, V est indépendante de W relativement à **W**.

En revanche, les sous-hypothèses 2.1 à 2.3 ne sont pas analytiques. Nous les discutons donc, tour à tour, avant d'en venir à l'hypothèse de fidélité causale.

Condition de Markov causale, sous-hypothèse 2.1. Indépendance causale et probabilités Selon la sous-hypothèse 2.1,

une variable V de $\mathbf{V}$ est indépendante, relativement à l'ensemble de ses causes directes dans $\mathbf{V}$, de chacune des variables de $\mathbf{V}$ avec lesquelles elle n'entretient aucune forme de relation causale. En termes graphiques, V est indépendante, relativement à l'ensemble de ses parents dans le graphe causal sur $\mathbf{V}$, de toute variable à laquelle elle n'est pas connectée dans ce graphe. Par contraposition, cette première sous-hypothèse peut s'énoncer de la façon suivante : toutes les dépendances probabilistes renvoient à des dépendances causales.

Ainsi reformulée, la sous-hypothèse 2.1 se présente comme stipulant qu'il est possible d'expliquer [1] les variations, ou plutôt les co-variations, en termes causaux. Or, une telle hypothèse est inséparable du projet scientifique lui-même, conçu comme projet d'explication des phénomènes observables. Plus précisément, le principe selon lequel les phénomènes observables peuvent être expliqués est constitutif de la science. L'hypothèse selon laquelle toute co-variation peut être expliquée causalement instancie ce principe, pour des phénomènes observables qui sont des phénomènes de co-variation et pour une conception classique de l'explication comme explication causale. Mill donne une formulation explicite de cette instanciation : « Quelque phénomène que ce soit qui varie de quelque façon que ce soit quand un autre phénomène varie d'une certaine façon soit est une cause ou un effet de ce phénomène, soit est connecté avec lui par quelque fait causal » (Mill, 1853, p. 287) [2]. Dans ces conditions, la première sous-hypothèse de la condition de Markov causale semble ne pas être problématique.

Condition de Markov causale, sous-hypothèse 2.2. Causalité indirecte et probabilités En deuxième lieu, la condition de Markov causale implique que toute variable V de $\mathbf{V}$ est indépendante, relativement à l'ensemble de ses causes directes dans $\mathbf{V}$, de ses autres causes dans $\mathbf{V}$. Pour le dire autrement, la sous-hypothèse 2.2 veut que l'influence des antécédents causaux se résume à celle des antécédents causaux immédiats. Ainsi reformulée, elle apparaît comme la partie proprement markovienne de la condition : seules les causes directes sont pertinentes pour l'état d'une variable, c'est-à-dire, ici, pour la fonction de probabilités sur cette variable.

La sous-hypothèse 2.2 décrit une propriété qui semble bien appartenir au rapport entre la causalité et les probabilités. Plus précisément, elle peut être considérée comme une version probabiliste de la thèse selon laquelle une cause et son effet sont contigus. La thèse de la contiguïté des

[1] La notion d'explication est utilisée ici dans un sens peu précis, mais qui nous semble suffisamment clair. Cette notion, et en particulier les rapports précis qu'elle entretient avec celle de causalité, n'est pas l'objet de notre analyse.

[2] Cette référence est empruntée à Williamson, 2005, p. 51.

causes et de leurs effets se présente comme une évidence : il semble clair qu'une action causale est toujours d'un proche à un proche. Pour cette raison, Hume fait de la contiguïté une caractéristique essentielle de la relation entre les causes et leurs effets : « Tous les objets qu'on considère comme causes et comme effets sont *contigus* » (Hume, 1739, p. 134).

Plus récemment, la contiguïté des causes et de leurs effets est celle des caractéristiques classiques de la causalité que visent au premier chef les analyses de la causalité au moyen de la notion de transmission. Selon ces théories, ce qui fait que A cause B est ceci que quelque chose est transmis de A à B. Cette thèse générale est spécifiée en particulier dans Reichenbach, 1956, chapitre 23, et dans Salmon, 1984, et devient alors l'idée selon laquelle les processus causaux se caractérisent par leur capacité à transmettre une marque. Elle a été défendue ensuite sous la forme de la thèse selon laquelle une quantité présente dans la cause est conservée dans l'effet (Aronson, 1971 ; Fair, 1979 ; Dowe, 1992a ; Dowe, 1992b ; Kistler, 1999).

Au-delà de l'attrait que la thèse de la contiguïté des causes et des effets exerce manifestement sur les théoriciens de la causalité, et ainsi que le souligne Kistler, cette thèse est étayée par des résultats physiques (Kistler, 1999, p. 40-41) :

> « Aujourd'hui, nous pouvons compter sur le résultat scientifique selon lequel toutes les forces fondamentales n'agissent à distance que grâce à une propagation préalable qui, elle, a une vitesse finie. [...] En ce qui concerne le mécanisme de l'action à distance des forces fondamentales, la physique nous donne des arguments solides pour soutenir qu'elle est le résultat de processus causaux sous-jacents qui obéissent strictement à la contiguïté. »

Nous n'entrerons pas dans l'analyse des points mentionnés par Kistler, mais retenons que la thèse de la contiguïté des causes et de leurs effets n'est pas seulement attirante d'un point de vue intuitif. De même que Kistler un peu plus loin dans son texte (Kistler, 1999, p. 43), nous considérerons donc que cette thèse est vraie.

Il reste que la vérité de la thèse de la contiguïté des causes et de leurs effets n'établit pas directement la vérité de la sous-hypothèse 2. La raison principale en est que les causes et les effets dont nous venons de conclure qu'ils sont contigus – et avec eux la causalité qu'on analyse en termes de transmission – sont *singuliers*. Or, les relations de cause à effet auxquelles nous nous intéressons dans le présent travail sont des relations *génériques*. En outre, et surtout, la thèse de la contiguïté ne peut pas s'entendre au sens strict dans le cas générique, puisqu'elle suppose que les *relata* de la causalité soient situés dans l'espace et dans le temps, et elle ne fait pas explicitement référence à des probabilités. On comble d'un même mouvement les deux fossés que nous venons de mettre en évidence en

considérant que la thèse de la contiguïté des causes et de leurs effets est l'expression, pour le cas singulier, de la thèse plus générale selon laquelle les influences causales se propagent de proche en proche.

Nous soutenons que cette thèse peut se concevoir comme la thèse de la contiguïté des causes et de leurs effets en tant qu'elle est étendue du niveau singulier auquel elle trouve son sens premier, spatio-temporel, au niveau générique qui nous intéresse. En effet, la thèse selon laquelle les influences causales se propagent de proche en proche peut trouver une expression relative à la causalité générique et, dans le cas qui nous intéresse, l'influence causale se comprend en termes de dépendance probabiliste. Dire que les influences causales se propagent de proche en proche, c'est alors en particulier dire que la dépendance probabiliste qu'une variable entretient à l'égard de ses causes est tout entière contenue dans la dépendance probabiliste qu'elle entretient à l'égard de ses causes *directes*. Il en découle que conditionaliser par ses causes directes rend la variable considérée indépendante de toutes ses autres causes. C'est là le contenu exact de la sous-hypothèse 2.2.

Nous venons de soutenir deux idées distinctes : d'une part, la thèse de la contiguïté des causes singulières et de leurs effets peut être acceptée ; d'autre part, cette thèse et la sous-hypothèse 2.2 sont deux expressions de la thèse plus générale selon laquelle l'influence causale se propage de proche en proche. Bien sûr, ces deux affirmations prises ensemble ne suffisent pas à établir la vérité de la sous-hypothèse 2.2. Toutefois, elles contribuent significativement à la rendre plausible. En particulier, la propagation des influences causales n'existe physiquement que comme propagation des influences causales singulières et dès lors la valeur de vérité de la thèse de la contiguïté ne semble pas devoir différer de celle de la thèse plus générale selon laquelle les influences causales se propagent de proche en proche.

Condition de Markov causale, sous-hypothèse 2.3. Causes communes et probabilités Selon la sous-hypothèse 2.3, une variable V est indépendante relativement à l'ensemble de ses causes directes dans $\mathbf{V}$ de toute variable de $\mathbf{V}$ qui est un effet de l'une de ces causes mais pas de V elle-même. À titre d'illustration, considérons le graphe de la figure 4.2. En admettant que ce graphe est le graphe causal sur $\{V_1, V_2, V_3, V_4, V_5\}$, la sous-hypothèse 2.3 implique que :

- V_2 est indépendante de V_5 relativement à $\{V_1, V_4\}$;
- V_5 est indépendante de V_2 et de V_3 relativement à V_4.

On peut considérer la sous-hypothèse 2.3 comme énonçant une version d'un principe plus général, selon lequel l'indépendance causale impliquerait l'indépendance probabiliste. Or, il existe un principe bien connu selon lequel, de manière contraposée, la dépendance probabiliste implique la dépendance causale. Il s'agit du principe de la cause commune tel qu'il est formulé explicitement et discuté par Reichenbach (Reichenbach, 1956,

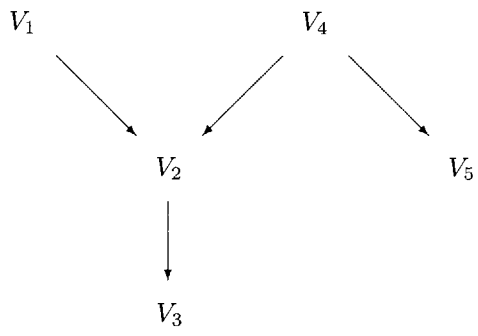


FIGURE 4.2

p. 158-159). Concernant des variables, il peut être formulé de la façon suivante :

Proposition 4.2 (Principe de la cause commune (PCC)) Si deux variables V_1 et V_2 sont dépendantes en probabilité et qu'aucune des deux n'est une cause de l'autre, alors il existe une variable V_3 qui est une cause de V_1 et une cause de V_2, et qui fait écran entre elles.

Cette formulation révèle des différences entre le PCC et la sous-hypothèse 2.3 ; nous les discutons en vue de rendre compte du rapport exact qu'entretiennent les deux énoncés.

En premier lieu, la sous-hypothèse 2.3 fait référence à des causes directes, là où le PCC porte sur la causalité en général, non spécifiquement directe. Cette première différence se réduit à partir de la remarque suivante : sous la sous-hypothèse 2.2, s'il existe V_3 qui est à la fois une cause commune à V_1 et à V_2 et une cause directe de V_1, alors qu'il existe une cause commune à V_1 et V_2 qui les rend indépendantes, implique que V_3 les rend indépendantes. Cette propriété résulte de ce que les influences causales se propagent de proche en proche si la sous-hypothèse 2.2 est satisfaite. Elle implique que, sous la sous-hypothèse 2.2, le PCC a la conséquence suivante : étant donné une variable V_1, V_3 une cause directe de V_1, et V_2 un effet de V_3 qui n'est pas causé par V_1, V_1 et V_2 sont indépendantes relativement à V_3. Autrement dit, le PCC implique l'indépendance des effets relativement à une cause commune qui est une cause directe de l'un d'entre eux.

Ce n'est pas là exactement la sous-hypothèse 2.3. En second lieu, en effet, la sous-hypothèse 2.3 mentionne la conditionalisation par *l'ensemble* des causes directes d'une variable donnée, là où le principe de la cause commune ne mentionne que la conditionalisation par *une* cause. Ainsi que le note Arntzenius prendre en compte l'ensemble des causes plutôt que l'une d'elles seulement est une modification qui

- « clairement ne viole pas l'esprit du principe de la cause commune de Reichenbach » (Arntzenius, 2005, section 1.1) ;
- étend le domaine de validité du principe. En effet, nous avons vu (sous-section 1.3.3) qu'il existe des couples d'effets ayant au moins deux causes communes, qui ne sont indépendants relativement à aucune de ces causes communes prises séparément, mais qui sont indépendants relativement à l'ensemble qu'elles constituent.

Nous avons montré d'abord que, moyennant la sous-hypothèse 2.2, le PCC implique des indépendances relatives qui relèvent de la sous-hypothèse 2.3 – précisément, les indépendances entre deux effets relativement à une cause commune qui est une cause directe de l'un des effets considérés. Dans ces conditions, les indépendances de ce type héritent de la plausibilité et de l'utilité généralement reconnues au PCC. Nous avons montré ensuite qu'une seconde différence entre la sous-hypothèse 2.3 et le PCC est telle que la première est plus souvent satisfaite que le second. Notre conclusion en sort renforcée : la sous-hypothèse 2.3, selon laquelle toute variable est indépendante, relativement à l'ensemble de ses causes directes dans $\mathbf{V}$, des effets de ces causes qu'elle ne cause pas, semble bien décrire adéquatement un aspect du rapport entre la causalité et les probabilités.

Pour en finir avec le traitement de cette sous-hypothèse, il convient de mentionner le rapport suivant entre la condition de Markov causale (et non plus la seule sous-hypothèse 2.3) et le PCC :

Proposition 4.3 La condition de Markov causale implique le PCC.

Cette proposition s'établit aisément : on peut se référer ici à Williamson, 2005, p. 51-52. On notera, en outre, que la réciproque de l'implication énoncée n'est pas vraie. Ainsi, il est faux que la condition de Markov causale est satisfaite dès lors que l'est le PCC. Considérons, par exemple, un ensemble $\{V_1, V_2, V_3\}$ de trois variables, tel que le graphe représentant les relations de cause à effet directes est $V_1 \longrightarrow V_2 \longrightarrow V_3$. Si V_1 n'est pas indépendante de V_3 relativement à V_2, alors le PCC est (trivialement) satisfait alors que la condition de Markov causale est violée. La proposition 4.3 n'est pas mise au premier plan par la décomposition que nous avons adoptée pour la condition de Markov causale. Nous nous en tenons toutefois à cette décomposition, en raison de son caractère littéral et de la valeur de clarification que nous lui reconnaissons.

Hypothèse de fidélité causale. Indépendances probabilistes et causalité Selon l'hypothèse de fidélité causale pour un ensemble de variables $\mathbf{V}$, il n'existe pas d'indépendance entre variables de $\mathbf{V}$ et relativement à un sous-ensemble de $\mathbf{V}$ qui ne soit pas impliquée par la condition de Markov causale. Pour le dire autrement, toutes les indépendances

probabilistes au sein de $\mathbf{V}$, absolues comme relatives, découlent de la condition de Markov causale.

Dans le cadre des méthodes RB, cette hypothèse est inséparable du projet même d'inférer des connaissances causales, ici à partir de connaissances probabilistes. En effet, en termes très généraux, ces méthodes consistent à utiliser les indépendances probabilistes comme signes de ce que nous pourrions appeler des « indépendances causales ». La condition de Markov causale indique selon quelles modalités ces indépendances causales impliquent des indépendances probabilistes, ouvrant ainsi la possibilité de reconduire celles-ci à celles-là. Mais pour que l'inférence causale soit possible, et même seulement envisageable, il faut que les indépendances probabilistes indiquent effectivement des indépendances causales, qu'elles ne soient pas trompeuses. Il apparaît ainsi que l'hypothèse de fidélité causale peut être considérée comme une hypothèse méthodologique fondamentale pour l'inférence causale probabiliste.

Nous avons vu dans le chapitre 1 (sous-section 1.3.5) que certaines indépendances probabilistes sont en fait trompeuses et nous avons montré dans le chapitre 3 (sous-section 3.4.2) que l'inférence causale suivant les méthodes RB est prise en défaut par de telles indépendances. Dans ces conditions, il apparaît non seulement que l'hypothèse de fidélité causale a été déjà, en un sens, discutée plus haut, mais encore que, dans le cadre des méthodes RB pour l'inférence causale probabiliste, elle doit être considérée comme une hypothèse méthodologique plutôt que comme une hypothèse théorique. Dès lors, il nous semble inutile de la discuter à nouveau ou plus avant. Dans la suite, nous ne revenons pas sur les violations de l'hypothèse de fidélité causale.

4.2 Limites des méthodes RB

4.2.1 Violations des hypothèses véhiculées par les méthodes RB

Les limites des méthodes RB pour l'inférence causale probabiliste tiennent principalement à ceci que les hypothèses que nous venons de mettre au jour ne sont pas toujours satisfaites.

Cycles causaux

Selon la première des hypothèses véhiculées par les méthodes RB, le graphe représentant les relations de cause à effet directes au sein d'un ensemble de variables est acyclique. Nous avons vu dans la sous-section 4.1.1 que cela équivalait à l'asymétrie de la causalité, et nous avons expliqué que la causalité se présentait en effet comme asymétrique : un effet ne cause pas sa cause. Si l'on veut aller au-delà des slogans et justifier plus avant la thèse selon laquelle la relation de causalité est asymétrique, on fait généralement valoir que les causes précèdent temporellement leurs

effets, qu'elles leur sont antérieures. Cette antériorité peut être considérée comme un élément de la définition de la causalité (notablement chez Hume) ou comme une propriété, éventuellement contingente, de toutes les relations de cause à effet que nous connaissons. Dans les deux cas, elle implique bien l'asymétrie de la relation de causalité : une cause précédant temporellement son effet, celui-ci lui succède et, ne la précédant pas, il ne peut la causer.

Cet argument, toutefois, vaut dans le seul cas singulier : s'il y a du sens à dire qu'un événement en précède un autre dans le temps, il n'y en a guère à dire qu'une propriété C est antérieure à une autre propriété E. *A fortiori*, il n'y a pas de sens à affirmer d'une variable qu'elle est antérieure à une autre variable. En ce qui concerne les propriétés, seules leurs instanciations par des individus singuliers sont temporellement ordonnées. En ce qui concerne les variables, nous avons vu (sous-section 2.3.2) qu'on pouvait considérer qu'elles représentaient des ensembles de propriétés, et seules sont temporellement ordonnées les instanciations, par des individus singuliers, des propriétés appartenant à ces ensembles. Ainsi, notre meilleur argument en faveur de l'asymétrie de la causalité ne peut pas être utilisé pour établir l'asymétrie de la causalité générique. Plus spécifiquement, il ne peut pas être utilisé pour établir l'asymétrie de la relation de causalité entre variables qui fait l'objet de l'inférence causale probabiliste quand elle est menée selon les méthodes présentées aux chapitres 2 ou 3.

Nous venons de voir que l'argument temporel en faveur de l'asymétrie de la causalité n'est pas disponible pour le cas générique. Il y a, cependant, pire que cela : à y regarder de plus près, il semble bien que la relation de causalité générique ne soit pas, finalement, asymétrique. Ainsi Williamson formule-t-il la remarque suivante (Williamson, 2005, p. 50) :

> « Les cycles causaux sont en fait répandus : la pauvreté cause le crime qui cause plus de pauvreté ; un système immunitaire faible conduit à la maladie qui peut encore affaiblir le système immunitaire ; les augmentations des prix de l'immobilier causent un empressement à acheter qui en retour cause de nouvelles augmentations des prix »[3].

Certains auteurs maintiennent la thèse de l'asymétrie de la causalité générique en dépit de contre-exemples de ce type. Ces auteurs doivent principalement montrer que ce n'est qu'en apparence que la causalité générique n'est pas asymétrique dans les cas du type de ceux que Williamson mentionne. Eells est celui qui a entrepris cette tâche avec le plus de rigueur et pour le résultat le plus abouti (Eells, 1991, p. 48 et suivantes). Or, son analyse suppose de considérer 1) que tout individu a est un ensemble de tranches temporelles de substance individuelle : $a = \{a_t\}$ où

[3] Davis formule une remarque du même type (Davis, 1988, p. 146).

tout t est un instant de la durée de a, 2) qu'il existe des propriétés indexées sur le temps, définies de la façon suivante : pour tout F, tout t, tout a, $F_t(a) =_{def} F(a_t)$, 3) que les énoncés causaux singuliers font référence à de telles propriétés, et 4) que les affirmations causales génériques sont relatives à des distances temporelles et s'analysent comme des quantifications universelles, sur l'ensemble des couples ordonnés d'instants temporels adéquatement distants, d'énoncés causaux singuliers. Chacune de ces quatre propositions rompt avec certaines de nos intuitions et impliquent qu'il existe une différence entre la façon dont nous parlons des choses (en particulier, de la causalité) et ce qu'elles sont réellement. Accepter ces quatre propositions ensemble nous semble être un prix trop élevé pour rétablir l'asymétrie de la causalité générique. Nous nous en tiendrons donc à la thèse selon laquelle la causalité générique n'est pas, après tout, asymétrique. Les cas mentionnés par Williamson seront considérés comme des contre-exemples à l'hypothèse selon laquelle elle l'est. De manière équivalente, ils sont des contre-exemples à l'hypothèse selon laquelle le graphe représentant les relations de cause à effet directes au sein d'un ensemble de variables donné est acyclique.

Violations de la condition de Markov causale

L'essentiel des critiques qui ont été soulevées contre les méthodes RB pour l'inférence causale probabiliste porte en fait, précisément, sur la condition de Markov causale (voir en particulier Cartwright, 1999 ; Cartwright, 2001, section 4 ; Freedman et Humphreys, 1999, p. 31-34 ; Williamson, 2001, §2 ; Williamson, 2005, section 4.2). Ces critiques de la condition de Markov causale consistent le plus souvent à lui opposer des contre-exemples. De ces contre-exemples, donc, la littérature critique sur ces méthodes regorge. Ici, nous ne prétendons ni tracer en théorie les limites du domaine de vérité de la condition de Markov causale[4], ni même proposer une liste ou une typologie exhaustive des contre-exemples connus à la condition. Positivement, le traitement que nous donnons de ces contre-exemples est guidé par le principe suivant : donner un exemple classique pour chacun des types de violations de la condition de Markov causale que nous aurons pu identifier. C'est que l'argument que nous proposons dans cette sous-section prétend valoir de manière générale et indépendamment de la diversité réelle des contre-exemples à la condition de Markov causale.

Nous nous appuyons ici sur la typologie des indépendances fondamentales impliquées par la condition de Markov causale que nous avons dressée dans la sous-section 1.1.2. Toutefois, nous ne nous conformons pas à l'ordre adopté dans cette sous-section. En effet, c'est d'abord et principalement la troisième composante de la condition de Markov cau-

[4] Sur ce point, le lecteur intéressé peut se reporter en particulier à Steel, 2005, et à Drouet, 2009.

sale, c'est-à-dire la sous-hypothèse 2.3, qui a été remise en cause, les contre-exemples à la sous-hypothèse 2.1 dérivant des contre-exemples à la sous-hypothèse 2.3. Nous envisageons donc d'abord les violations de la sous-hypothèse 2.3, puis les contre-exemples à la sous-hypothèse 2.1, et enfin nous abordons la sous-hypothèse 2.2.

Fourches interactives Selon la sous-hypothèse 2.3, toute variable V de l'ensemble $\mathbf{V}$ considéré est indépendante, relativement à l'ensemble de ses causes directes dans $\mathbf{V}$, de toute variable de $\mathbf{V}$ qui soit un effet de l'une de ces causes mais pas de V elle-même. En d'autres termes, les causes communes font écran à la dépendance probabiliste entre leurs effets. Nous avons vu (sous-section 1.3.1) que certaines causes communes, ou certains ensembles de causes communes, échouent à faire écran à la dépendance entre les effets qu'ils ont en commun. Ainsi que nous l'avons indiqué, on parle alors de « fourches interactives ». Les exemples de fourches interactives sont nombreux dans la littérature relative à la condition de Markov causale (en particulier : Cartwright, 1999, p. 7-8 ; Davis, 1988, p. 156 ; Salmon, 1980, p. 150-151 ; Salmon, 1984, p. 168-169).

Parmi les exemples de fourches interactives, une classe numériquement et conceptuellement importante met en scène des causes indéterministes. Nous n'abordons pas ici la question de savoir quelle est la place exacte de cette sous-classe dans la classe de tous les contre-exemples à la troisième composante de la condition de Markov causale, mais nous contentons de présenter celui qui est canonique parmi les contre-exemples indéterministes. Il est présenté dans les termes suivants (Cartwright, 1999, p. 7) :

> « Deux usines sont en concurrence pour produire un produit chimique qui est consommé immédiatement par une usine d'épuration proche. La ville fait une étude pour décider à laquelle faire appel. Certains jours, les produits chimiques sont achetés à Clean / Green, d'autres à Cheap-but-Dirty. Cheap-but-Dirty emploie un processus véritablement probabiliste pour produire le composant. La probabilité d'obtenir le composant désiré un jour quelconque d'activité de l'usine est de 0,8. Donc, dans environ un cinquième des cas où le composant est acheté à Cheap-but-Dirty, les eaux usées ne sont pas traitées. Mais la méthode est si bon marché que la ville est prête à s'accommoder de cela. En revanche, il existe une autre raison pour laquelle elle ne veut pas acheter à Cheap-but-Dirty : elle refuse les polluants qui sont émis en même temps que le composant dès que celui-ci est produit. »

Pour comprendre que ce passage introduit un contre-exemple à la troisième composante de l'hypothèse de Markov causale, considérons les variables X, Y et C représentant respectivement la production du compo-

sant chimique par Cheap-but-Dirty, l'émission de produits polluants par Cheap-but-Dirty et le fonctionnement même de l'usine Cheap-but-Dirty. Les seules relations de cause à effet sont de C à X et de C à Y. Les relations de cause à effet directes sur $\{X, Y, C\}$ sont représentées par la figure 4.3 ci-dessous. Or X et Y ne sont pas indépendantes relativement

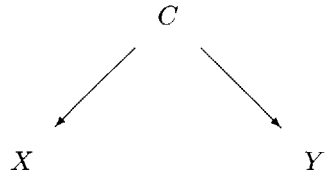


FIGURE 4.3

à C : $p(X = 1|Y = 1 \wedge C = 1) = 1$ tandis que $p(X = 1|C = 1) = 0,8$. De même que dans le cas introduit par Salmon que nous avons présenté dans la sous-section 1.3.1, la cause commune ne fait pas écran à la dépendance probabiliste entre ses effets parce qu'elle ne les produit pas indépendamment l'un de l'autre. L'apport spécifique de l'exemple de Cartwright est double : il lie les phénomènes de ce type, d'une part, à la condition de Markov causale et, d'autre part, à la notion de cause « véritablement probabiliste » (Cartwright, 1999, p. 7) – ou, en d'autres termes, indéterministe.

Indépendance causale et dépendance probabiliste Selon la sous-hypothèse 2.1, toute variable est indépendante, relativement à l'ensemble de ses causes directes dans l'ensemble de variables $\mathbf{V}$ qu'on considère, de toutes les variables de $\mathbf{V}$ dont elle est causalement indépendante. Pour comprendre comment l'existence de causes qui ne suffisent pas à déterminer la valeur de leurs effets (c'est-à-dire de causes indéterministes au sens envisagé dans le dernier paragraphe) implique une classe de contre-exemples à cette première sous-hypothèse, il convient de commencer par rappeler que deux variables causalement indépendantes peuvent être dépendantes en probabilités. Citons Sober de nouveau – et un peu plus longuement que nous ne l'avons fait dans la sous-section 1.3.3 (Sober, 1988, p. 215) :

> « Considérons le fait que le niveau de la mer à Venise et le coût du pain en Grande-Bretagne ont été tous deux à la hausse dans les deux siècles passés. Disons que tous deux ont augmenté de façon monotone. Imaginons que nous mettions ces informations sous la forme d'une liste chronologique. Pour chaque date, nous relevons le niveau de la mer à Venise et le prix courant du pain britannique. Parce que les deux quantités ont augmenté régulièrement avec le temps, il est vrai que les niveaux de la mer plus élevés que la moyenne tendent à être

> associés à des prix du pain plus élevés que la moyenne. Les deux quantités sont très fortement corrélées positivement.
>
> Il me semble que nous ne nous sentons pas conduits à expliquer cette corrélation par une cause commune. Plutôt, nous considérons les niveaux de la mer à Venise et les prix du pain britannique comme augmentant tous deux pour des raisons endogènes et quelque peu isolées. Les conditions locales vénitiennes ont accru le niveau de la mer et des conditions locales assez différentes ont poussé à la hausse le coût du pain en Grande-Bretagne. Ici, postuler une cause commune est simplement peu plausible, étant donné le reste de ce que nous croyons. »

Ainsi qu'il est explicite dans cet extrait, l'exemple de Sober est conçu d'abord comme un contre-exemple au principe de la cause commune. Néanmoins, pas plus qu'une cause commune, nous ne sommes prêts à reconnaître un quelconque autre lien causal entre le niveau de la mer à Venise et le prix du pain britannique. Nous considérons donc que le cas envisagé illustre la possibilité qu'une dépendance probabiliste ne corresponde à aucune dépendance causale, de quelque forme qu'elle soit. En ce sens, le contre-exemple présenté ici est plus radical que celui que nous présentions dans le paragraphe précédent et qui illustrait la possibilité qu'une dépendance probabiliste ne disparaisse pas complètement quand on prend en compte la dépendance causale à laquelle elle correspond.

Imaginons que le prix du pain en Grande-Bretagne est suffisamment déterminé par les valeurs que prennent les variables d'un ensemble $\mathbf{V}$ – auquel appartiendraient, en particulier, les variables représentant le prix du blé et la demande de pain en Grande-Bretagne. Dans ce cas, le prix du pain britannique est indépendant du niveau de la mer à Venise relativement à l'ensemble de ces variables. Mais nous venons de voir qu'il existe des causes dont la valeur ne suffit pas à déterminer celle de leurs effets. Émettons maintenant l'hypothèse selon laquelle l'ensemble des variables qui contribuent à déterminer le prix du pain en Grande-Bretagne est de ce type. Dans ce cas, celui-ci reste dépendant du niveau de la mer à Venise, même quand on conditionalise par l'ensemble de ses causes. Ainsi, en s'appuyant à la fois sur l'existence de dépendances probabilistes auxquelles ne correspondent pas de dépendances causales et sur l'existence de causes qui ne suffisent pas à déterminer la valeur de leurs effets, on fait apparaître une classe de contre-exemples à la première sous-hypothèse issue de la condition de Markov causale.

Processus non markoviens Venons-en, pour finir, à la sous-hypothèse 2.2. Selon cette sous-hypothèse, une variable est indépendante de ses causes indirectes relativement à ses causes directes : les causes directes d'une variable font écran entre cette variable et ses ancêtres causaux. C'est

là la composante proprement markovienne de la condition de Markov causale, aussi semble-t-il naturel de se tourner du côté des processus non markoviens afin d'identifier, éventuellement, des contre-exemples. Nous entendons ici « processus non markoviens » au sens que cette expression a eu d'abord, historiquement. En ce sens, un processus affectant un système est non markovien s'il est tel que l'état du système à un instant donné ne suffit pas à déterminer les probabilités que le système soit dans tel ou tel état à l'instant immédiatement postérieur. En d'autres termes, les probabilités à l'instant t_{n+1} ne dépendent pas seulement de ce qui a eu lieu à l'instant t_n, elles dépendent aussi de ce qui a eu lieu avant. Les phénomènes d'apprentisssage, qu'ils soient individuels ou collectifs, sont généralement non markoviens en ce sens [5].

En vue de comprendre si et en quel sens les processus qui ne sont pas markoviens au sens classique constituent des contre-exemples à la condition de Markov causale, considérons un exemple. Celui que nous choisissons est extrêmement simple et épuré. Soit une urne qui contient une boule blanche et une boule noire à l'instant initial t_0, et dans laquelle on effectue une suite de tirages avec double remise : à la suite du tirage, on remet dans l'urne *deux* boules qui ont la même couleur que la boule tirée. La suite des résultats de tirages effectués dans l'urne est un processus non markovien : la probabilité de tirer une boule blanche (resp. noire) à l'instant $t_{n+1}(n > 1)$ dépend non seulement du résultat du tirage effectué à l'instant t_n, mais encore des résultats de tous les tirages antérieurs. En effet, la règle de remise est telle que le nombre de boules blanches et le nombre de boules noires qui sont contenues par l'urne à l'issue du tirage effectué à l'instant t_n dépendent de l'ensemble des résultats des tirages effectués jusqu'à cet instant. Partant, les probabilités à l'instant t_{n+1}, si elles sont classiques [6], dépendent elles aussi de l'ensemble des résultats des tirages effectués jusqu'à l'instant t_n. Un peu plus précisément, la variable pour laquelle les probabilités ne sont pas complètement déterminées par la valeur que prennent ses parents est la variable binaire T_{n+1} qui représente le résultat du tirage effectué à l'instant t_{n+1}. Cette variable a un seul parent : la variable binaire T_n qui représente le résultat du tirage effectué à l'instant t_n. Or, conditionaliser par T_n ne suffit pas à rendre T_{n+1} indépendante des variables T_i, pour $1 \leq i < n$.

Il apparaît alors que le processus que nous proposons de considérer constitue un contre-exemple à la deuxième sous-hypothèse de la condition de Markov causale si l'on représente ce processus de la façon suivante :

$$T_1 \longrightarrow T_2 \longrightarrow \ldots \longrightarrow T_n \longrightarrow T_{n+1} \longrightarrow \ldots$$

où T_i est la variable binaire qui prend la valeur B ou la valeur N selon que

[5] Pour une analyse des phénomènes sociaux de renforcement, voir Skyrms, 2004.

[6] Par là, nous entendons que la probabilité d'un événement est égale au rapport entre le nombre de cas favorables et le nombre de cas possibles.

la boule tirée à l'instant t_i est blanche ou noire. Le couple composé de ce graphe orienté acyclique et des probabilités classiques sur les variables qui y figurent constitue un contre-exemple à la deuxième sous-hypothèse de la condition de Markov causale. Plus précisément, il constitue un contre-exemple à la deuxième sous-hypothèse de la condition de Markov causale seulement si l'on considère que les flèches du graphe que nous venons de tracer sont causales. En d'autres termes, notre suite de tirages avec double remise invalide la condition de Markov causale si et seulement si chaque variable T_{n+1} a T_n pour seule cause directe dans $\mathbf{T}$.

Or, il nous semble que nous ne serions pas prêts à dire cela. En effet, si l'on reconnaît que T_n est une cause directe de T_{n+1}, ce ne peut qu'être au sens où la valeur que prend T_n affecte la composition de l'urne, dont dépendent les probabilités sur T_{n+1}. Mais alors il n'y a aucune raison de ne pas considérer les variables T_1 à T_{n-1} aussi comme des causes directes de T_n et cela, clairement, suffit à rétablir la condition de Markov causale. Ainsi, l'exemple que nous envisageons ne donne pas de contre-exemple à la condition de Markov causale si celle-ci est correctement comprise, comme portant spécifiquement sur le rapport entre graphes *causaux* et fonctions de probabilités. C'est seulement à une version *temporelle* de la condition de Markov qu'il pourrait servir de contre-exemple.

De façon plus générale, les processus qui ne sont pas markoviens au sens classique du terme ne constitueraient des contre-exemples à la sous-hypothèse 2.2 qu'à la condition que la seule cause directe de ce qui advient à un instant donné est ce qui advient à l'instant immédiatement précédent. Or, justement, les processus non markoviens sont tels qu'il existe une influence de type causal du passé vers le futur qui ne peut pas être réduite à l'influence du présent sur le futur ; c'est même là ce qui caractérise ces processus. Dans ces conditions, les processus qui ne sont pas markoviens au sens classique du terme ne sont pas des contre-exemples à la deuxième sous-hypothèse issue de la condition de Markov causale. En outre, nous ne connaissons pas d'autres candidats au titre de contre-exemples à cette sous-hypothèse 2.2. Nous nous en tiendrons donc, pour ce qui concerne la condition de Markov causale, aux contre-exemples aux sous-hypothèses 2.1 et 2.3. Ces contre-exemples constituent des limites pour l'inférence causale probabiliste menées selon les méthodes RB, en même temps qu'une partie du prix à payer pour la possibilité d'induire des relations de cause à effet à partir de connaissances probabilistes tirées de données statistiques d'observation.

4.2.2 Limites spécifiques et limites non spécifiques

Le présent chapitre ne vise pas seulement à définir les limites des méthodes RB ; il vise aussi, plus généralement, à mesurer la portée de ces méthodes pour l'inférence causale probabiliste. Pour cette raison, nous nous attachons maintenant à déterminer si les limites que nous venons de

mettre en évidence affectent spécifiquement les méthodes RB, ou bien si elles affectent également les méthodes plus traditionnelles que nous avons présentées dans le chapitre 2. Cette question se comprend plus simplement comme la question de savoir si les hypothèses sur lesquelles reposent les méthodes RB (et dont nous venons de voir qu'elles sont parfois violées) sont spécifiques de ces méthodes, ou bien si elles sous-tendent également les méthodes hypothético-déductives. Nous commençons par répondre à cette question, et cela nous amène à aborder dans un second temps la question inverse : les limites que nous avons mises en évidence pour les méthodes traditionnelles leur sont-elles spécifiques, ou bien affectent-elles également les méthodes RB ? En traitant cette dernière question, nous achèverons de tracer les limites des méthodes RB pour l'inférence causale probabiliste.

Des hypothèses spécifiques des méthodes RB ?

Les méthodes hypothético-déductives présentées dans le chapitre 2 reposent-elles elles aussi sur l'hypothèse d'acyclicité des graphes causaux et sur la condition de Markov causale et l'hypothèse de fidélité causale ? C'est évidemment pour l'hypothèse d'acyclicité que la réponse est la plus évidente. En effet, il semble clair que le caractère cyclique ou acyclique du graphe causal ne change rien ni à la possibilité de mener une inférence se conformant à la procédure définie dans la sous-section 2.2.2, ni à la validité qui serait celle d'une telle inférence. Plus précisément, toutes les tâches que nous avons explicitées dans la sous-section 2.2.3 peuvent être menées à bien pour des graphes cycliques aussi bien qu'acycliques ou, en d'autres termes que nous avons introduits dans le chapitre 2, pour des modèles structurels récursifs aussi bien que non récursifs. L'hypothèse d'acyclicité est donc spécifique des méthodes RB, et avec elle les limites que les violations de ces hypothèses constituent pour l'inférence causale probabiliste.

Pour ce qui est de la condition de Markov causale et de l'hypothèse de fidélité causale, notre procédure hypothético-déductive ne requiert pas à proprement parler qu'elles soient satisfaites. Toutefois, ces deux hypothèses dessinent ensemble une conception du rapport entre causalité et probabilités dont nous soutenons qu'elle est similaire à celle qui sous-tend, au moins en pratique, la modélisation causale. Plus précisément, nous avons montré dans la sous-section 3.4.3 que les méthodes RB véhiculent une conception de la causalité directe comme corrélation qui ne disparaît pas par conditionalisation, et nous soutenons maintenant qu'une telle conception est également à l'œuvre quand des modèles structurels sont utilisés dans le domaine des sciences de la société. Ainsi, on considérera qu'une variable C cause (directement) une variable E si et seulement si 1) C et E sont dépendantes en probabilité et 2) les autres variables de l'ensemble considéré ne font pas écran entre C et E.

D'une part, seules des variables dépendantes en probabilité sont susceptibles d'être considérées comme cause et effet directs dans le cadre de la modélisation causale : à l'étape A de la procédure que nous avons décrite dans la sous-section 2.2.2, on ne spécifie un modèle dans lequel figure une flèche entre deux variables données que si ces deux variables sont corrélées. En outre, si la valeur estimée du paramètre associé à l'une des flèches d'un modèle ne diffère pas significativement de zéro, alors ce modèle est rejeté à l'étape C, et il l'est au profit d'un modèle identique sauf pour ceci que cette flèche n'y figure pas. Autrement dit, parmi les dépendances probabilistes, on ne considère finalement comme causales que celles pour lesquelles on peut rejeter l'hypothèse selon laquelle elles disparaissent par conditionalisation.

D'autre part, une corrélation entre variables est interprétée causalement si elle ne disparaît pas par conditionalisation. Plus précisément, nous soutenons qu'un modèle conduisant à ne pas interpréter causalement une dépendance qui ne disparaît pas par conditionalisation est rejeté à l'étape C de la procédure que nous avons définie dans la section 2.2. Pour justifier cette affirmation, il nous faut entrer dans le détail de cette procédure tel qu'il est présenté dans la sous-section 2.2.3 (et dont le lecteur non intéressé a pu faire l'économie). Considérons plus précisément les tests de type [b.] parmi ceux que nous avons identifiés comme pouvant être menés à l'étape C. Ces tests consistent à vérifier que les corrélations impliquées par un modèle ne diffèrent pas significativement des corrélations observées. Or, les corrélations impliquées par le modèle se calculent à partir des coefficients associés aux différentes flèches qui figurent dans le modèle – par sommation, dans les cas simples. Les tests de type [b.] reposent donc sur l'idée selon laquelle la dépendance entre deux variables qui ne se causent pas l'une l'autre s'épuise dans la dépendance qu'impliquent les différents chemins causaux menant de l'une à l'autre. Réciproquement, une dépendance que n'épuiserait pas la prise en compte de ces chemins demande à être interprétée causalement. Les tests de type [b.] mettent exactement en œuvre l'idée selon laquelle un modèle qui n'interpréterait pas causalement une telle dépendance doit être rejeté.

De façon générale, il apparaît que les méthodes hypothético-déductives pour l'inférence causale probabiliste impliquent une conception du rapport entre la causalité et les probabilités qui est similaire à celle que véhiculent les méthodes RB. Dès lors, les violations de la condition de Markov causale et de l'hypothèse de fidélité causale ne constituent pas des limites à l'inférence causale probabiliste qui seraient spécifiques des méthodes RB.

Spécificité des limites des méthodes hypothético-déductives

La question se pose en retour de savoir si les limites qui affectent l'inférence causale hypothético-déductive, affectent également l'inférence causale selon les méthodes RB. Ainsi que nous l'avons déjà indiqué, répondre à

cette dernière question nous permet d'achever de tracer les limites des méthodes RB pour l'inférence causale probabiliste.

Les limites analysées dans la section 2.3 sont de deux types : elles tiennent en premier lieu au caractère hypothético-déductif de l'inférence elle-même, et en second lieu à son caractère statistique. Clairement, les limites du premier type sont spécifiques à l'approche hypothético-déductive. Ainsi, l'une des principales caractéristiques logiques des inférences causales menées selon les méthodes RB consiste en ce qu'elles sont inductives, au sens (mis en évidence dans la section 3.3) où elles sont a-théoriques, où elles permettent de tirer des conclusions causales *directement* de connaissances probabilistes – c'est-à-dire, précisément, au sens où elles *ne* sont *pas* hypothético-déductives.

Les inférences causales probabilistes hypothético-déductives sont statistiques au sens suivant : elles portent sur des hypothèses et des énoncés probabilistes et les données qui en constituent les prémisses ne sont pas relatives à la population considérée, mais à un échantillon de cette population. Nous avons montré que cette caractéristique impliquait la possibilité que la conclusion d'une inférence causale probabiliste soit fausse alors même que les prémisses seraient vraies. Nous expliquons maintenant pourquoi la propriété d'être statistique n'est pas spécifique des méthodes hypothético-déductives, mais caractérise également les méthodes RB.

Certes, nous avons montré dans la section 3.3 que l'inférence causale menée selon les méthodes RB pouvait être considérée comme déductive : la conclusion d'une telle inférence est une conséquence logique de ses prémisses. Plus précisément, la conclusion de l'inférence est une conséquence logique de l'entrée probabiliste qui est donnée à l'algorithme mis en œuvre dans le cadre de l'inférence. Mais cette entrée consiste exactement dans l'ensemble des indépendances probabilistes relatives au sein de l'ensemble de variables considéré. Or, si l'on prétend utiliser les méthodes RB pour inférer des connaissances causales à partir de données statistiques d'observation, ces indépendances ne sont pas données. Seules sont alors données des fréquences relatives ou des corrélations au sein d'un échantillon de la population étudiée. C'est pourquoi une étape précoce de la procédure d'inférence causale schématisée dans la sous-section 3.1.3 consiste à utiliser les données statistiques disponibles pour mettre au jour les indépendances probabilistes relatives qui valent pour la population considérée. Les indépendances ainsi identifiées ne sont rien d'autre que celles que décrivent des hypothèses statistiques qui n'ont pas été rejetées. En ce sens, les inférences causales probabilistes qui se conforment aux méthodes RB sont elles aussi des inférences statistiques, et il en découle qu'elles sont affectées par les limites que nous avons vu être associées à cette caractéristique (sous-section 2.3.4).

L'idée selon laquelle les inférences menées selon les méthodes RB sont statistiques au même titre que les inférences causales probabilistes plus

traditionnelles n'est pas nouvelle. Le point est soulevé en particulier dans Humphreys et Freedman, 1996, p. 117. Néanmoins, il nous semble qu'il reste encore souvent – trop souvent – négligé. Aussi y insistons-nous : les méthodes RB ne donnent aucun moyen de résoudre les problèmes qui se posent quand on veut inférer des connaissances portant génériquement sur une population à partir de données relatives aux fréquences et, éventuellement, aux corrélations dans un échantillon de la population considérée. Nous voyons deux explications non exclusives l'une de l'autre, ni sans doute même indépendantes, au manque d'attention dont ce point souffre le plus souvent. D'une part, c'est pour la logique des inférences auxquelles elles donnent lieu que les méthodes RB renouvellent l'inférence aux causes génériques à partir de données statistiques et des connaissances probabilistes qu'on peut tirer de telles données. Ce qui distingue ces méthodes RB, c'est le caractère a-théorique des inférences auxquelles elles donnent lieu, ainsi que le caractère déductif des inférences qu'on mène au moyen des algorithmes qu'elles mobilisent. La question de savoir comment sont engendrées les entrées probabilistes de ces algorithmes passe alors au second plan. D'autre part, le débat relatif à ces méthodes s'est focalisé sur les hypothèses que nous avons mises au jour dans la section 4.1 et discutées dans la section 4.2, c'est-à-dire sur les hypothèses sous lesquelles les conclusions causales sont effectivement des conséquences logiques des indépendances probabilistes relatives au sein de la population considérée.

Au total, il apparaît que, dans le cadre de l'inférence causale probabiliste, les limites spécifiques sont celles qui découlent, pour les inférences menées selon les méthodes RB, de l'existence de cycles causaux et celles qui tiennent, pour les inférences menées selon les voies explorées dans le chapitre 2, précisément à leur caractère hypothético-déductif. En revanche, les difficultés qui sont attachées au caractère statistique des inférences qui nous intéressent dans cet ouvrage, de même que celles que constituent les violations de la condition de Markov causale ou de l'hypothèse de fidélité causale, affectent également les méthodes RB et les méthodes hypothético-déductives plus traditionnelles. À ce point de notre étude, il nous est donc impossible de tirer une conclusion claire et univoque relativement à la portée des méthodes RB, à ce qu'elles changent pour l'inférence causale probabiliste. Pourtant, c'est seulement une fois que nous aurons tiré une telle conclusion que nous aurons déterminé exactement quelle est la contre partie de la possibilité, ouverte par les méthodes RB, d'induire des connaissances causales à partir de données statistiques d'observation.

4.3 Portée des méthodes RB

Ce que les méthodes fondées sur la notion de réseau bayésien changent le plus visiblement à l'inférence causale probabiliste touche à la logique

des inférences : les inférences menées selon les méthodes RB sont à la fois a-théoriques et telles que leur conclusion est une conséquence logique des indépendances probabilistes relatives qui sont identifiées au sein de l'ensemble de variables considéré. Ces caractéristiques sont telles que les méthodes RB sont à l'abri des difficultés que nous avons vu être associées à l'hypothético-déduction et, en ce sens, elles impliquent que ces méthodes constituent une avancée dans le domaine de l'inférence causale probabiliste. Toutefois, nous avons montré que le fait que la causalité directe n'était pas toujours acyclique affectait spécifiquement les méthodes RB. En outre, et surtout, nous ne nous sommes pas encore intéressés à ce que changent réellement, pour l'inférence causale probabiliste menée selon les méthodes RB, les limites que nous avons mises en évidence. C'est par cet angle que nous nous attaquons à la question de savoir quelle est la portée des méthodes d'inférence causale probabiliste qui mobilisent les réseaux bayésiens.

Ce qui suit du caractère statistique d'une inférence est bien connu en général et, dans la sous-section 2.3.4, nous l'avons formulé pour le cas particulier de l'inférence causale probabiliste. Sur ce point, nous ne voyons rien qui distinguerait les méthodes RB des méthodes hypothético-déductives. En conséquence, nous concentrons notre attention sur les autres limites de l'inférence causale probabiliste menée selon les méthodes RB : celles que constituent les violations de l'hypothèse d'acyclicité et de la condition de Markov causale.

4.3.1 Des artifices contre les violations de l'hypothèse d'acyclicité ou de la condition de Markov causale ?

Selon une première hypothèse, l'existence de systèmes pour lesquels l'hypothèse d'acyclicité et la condition de Markov causale sont violées n'a en fait aucune conséquence pour l'inférence causale probabiliste, parce que ces hypothèses peuvent être rétablies artificiellement dans les cas où elles sont violées.

Rétablir artificiellement l'acyclicité

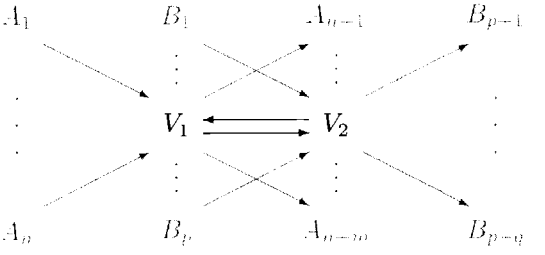
Nous avons montré (sous-section 4.2.1) que l'acyclicité du graphe causal sur un ensemble de variables équivaut à l'asymétrie de la relation de causalité sur cet ensemble. Cette équivalence, toutefois, n'a été établie que sous l'hypothèse, restée implicite, selon laquelle une même variable ne figure qu'une fois dans le graphe représentant les relations de cause à effet directes au sein d'un ensemble de variables donné.

Williamson explique comment le fait de lever cette dernière hypothèse permet de représenter la causalité directe au sein d'un ensemble de variables $\mathbf{V}$ au moyen d'un graphe acyclique même quand il existe dans $\mathbf{V}$ deux variables V_1 et V_2 qui se causent directement l'une l'autre. Pour

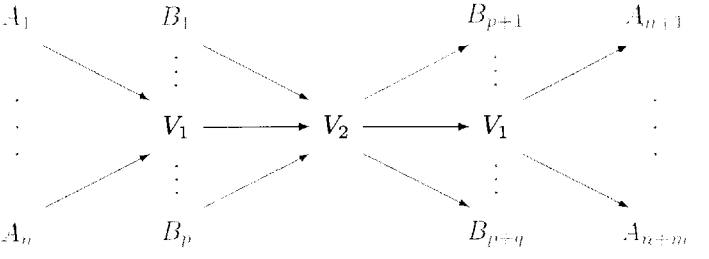
faire cela, il suffit de distinguer V_1 en tant qu'il cause V_2, et V_1 en tant qu'il est causé par V_2 (Williamson, 2005, p. 50) :

> « Il est possible d'éradiquer ces cycles en considérant différentes instanciations des causes et des effets comme différentes variables : si la première augmentation des prix de l'immobilier est une variable différente de la seconde augmentation des prix, alors il existe une chaîne de connexions causales d'une augmentation à l'autre, plutôt qu'un cercle entre augmentation des prix et empressement à acheter. »

Plus formellement, on remplacera, dans le graphe dans lequel chaque variable figure exactement une fois et où chaque relation de cause à effet directe est représentée par une flèche, tout sous-graphe de la forme :



par un graphe de la forme :



Cette solution est disponible indépendamment de l'interprétation en termes d'instanciation qu'en donne Williamson – et qui, telle qu'il la formule, semble beaucoup mieux adaptée à la causalité singulière qu'à la causalité générique, que pourtant elle vise aussi.

L'artifice envisagé par Williamson vise le cas où la causalité directe est acyclique. Cependant, nous avons vu que, en l'absence de tout artifice, l'hypothèse de représentation véhiculée par les méthodes RB implique que non seulement la causalité directe mais encore la causalité (tout court) sont asymétriques. Pour que cela n'interdise jamais de représenter les relations de cause à effet directes au sein d'un ensemble de variables au moyen d'un graphe acyclique, il suffit que la proposition de Williamson

puisse être généralisée. Or, tel est bien le cas, la proposition généralisée étant la suivante : étant donné un cycle causal, on choisit une des variables de ce cycle et on distingue cette variable en tant qu'origine du cycle de cette variable en tant qu'aboutissement du cycle. Selon les voies empruntées par Williamson pour le cas où la causalité directe est asymétrique, on peut alors faire disparaître le cycle en le déployant.

Nous venons de montrer que, si la causalité générique n'est pas asymétrique, l'hypothèse de représentation véhiculée par les méthodes RB pour l'inférence causale probabiliste n'implique pas, finalement, qu'elle le soit. Si nous avons pu montrer plus haut (sous-section 4.2.1) qu'elle l'impliquait, c'est que nous avons considéré qu'une même variable ne pouvait figurer qu'une fois dans le graphe représentant les relations de cause à effet directes sur un ensemble de variables auquel elle appartient. En termes plus généraux, nous avons négligé ceci que l'hypothèse véhiculée par les méthodes RB relativement à la causalité est une hypothèse de *représentation*, qu'elle porte sur la façon dont la causalité est représentée, plutôt que sur la causalité elle-même. C'est seulement si la causalité générique était asymétrique, que l'hypothèse serait satisfaite pour le mode de représentation le plus immédiat pour la causalité directe sur un ensemble de variables (le seul que nous ayons envisagé avant le présent paragraphe).

Rétablir artificiellement la condition de Markov causale

De même que l'hypothèse d'acyclicité, la condition de Markov causale porte sur les représentations de la causalité plutôt que sur la causalité elle-même. Plus précisément, elle porte sur le rapport entre la fonction de probabilités sur un ensemble de variables et le graphe représentant les relations de cause à effet directes au sein de cet ensemble. Corrélativement, la condition de Markov causale peut être rétablie artificiellement, quand elle est violée, selon des voies analogues à celles que Williamson préconise pour l'acyclicité. Afin de le montrer, présentons d'abord une manière simple de prendre en charge certains contre-exemples à la condition de Markov causale. Ces contre-exemples sont du type de celui-ci, introduit par Gillies (2002, p. 80) :

> « Le second contre-exemple pourrait être appelé exemple de la vache pleine et est donné dans Jensen, 1996, p. 36-37. Il vient du domaine de la science vétérinaire, et est relatif à un test destiné à déterminer si une vache est pleine. [...]
> Ici Pl est une variable qui prend la valeur 1 si la vache est pleine et la valeur 0 sinon. TS représente le résultat d'un test sanguin, TU le résultat d'un test d'urine, et Sc le résultat d'un scanner. [...] Les conditions d'indépendance [imposées par la structure causale] sont ici les suivantes : les variables TS, TU et Sc doivent être indépendantes relativement à Pl.

> Or, Sc est en effet indépendante de TS et de TU relativement à Pl, mais TS et TU sont corrélées relativement à Pl[7]. »

La structure causale sur l'ensemble de variables envisagé par Gillies représentée par le graphe de la figure 4.4 :

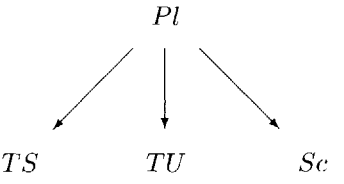


FIGURE 4.4

La difficulté vient ici de ce que, contrairement à ce qu'implique la condition de Markov causale, TU et TS ne sont pas indépendantes relativement à Pl. Toutefois, continue Gillies, chacune de ces deux variables est un effet de la variable Ho représentant le niveau de production d'hormones par la vache et cette variable fait écran entre TU et TS. On rétablit donc la condition en incluant Ho dans l'ensemble de variables considéré. Ainsi, le graphe de la figure 4.5 est avec la fonction de probabilités sur l'ensemble des variables qui figurent dans le graphe, dans le rapport que décrit la condition de Markov causale :

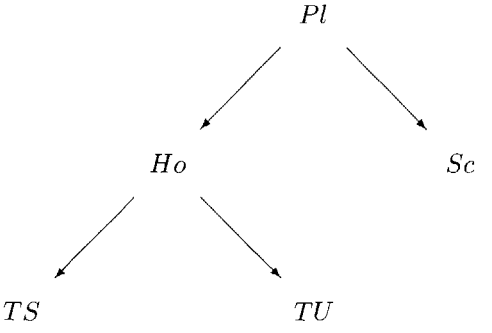


FIGURE 4.5

Dans l'exemple de la vache pleine, la condition de Markov causale est violée parce que la variable Ho n'a pas été prise en compte, et on rétablit la condition en intégrant Ho à l'ensemble de variables considéré.

Il existe toutefois, nous l'avons vu, des contre-exemples à la condition de Markov causale qui ne sont pas du même type que l'exemple de la vache pleine. Plus précisément, nous avons vu dans la sous-section 4.2.1 que

[7] Le texte original est en anglais et nous avons changé le nom des variables afin que le nom de chaque variable fasse écho au mot français qui désigne ce qu'elle représente.

les contre-exemples à la troisième composante de l'hypothèse de Markov causale ne dépendent pas de l'omission d'une variable, mais du caractère « véritablement probabiliste » (pour reprendre les mots de Cartwright) de l'action de certaines causes sur leurs effets. En conséquence, ces contre-exemples continuent d'exister même quand toutes les variables pertinentes pour la structure causale de la situation considérée sont prises en compte.

Plus précisément, ces contre-exemples continuent d'exister même quand toutes les variables *observables* pertinentes pour la structure causale de la situation considérée sont prises en compte. Mais on peut imaginer revenir sur cette propriété – d'être observable. Pour le dire autrement, on peut imaginer rétablir la condition de Markov causale dans certains des cas qui nous intéressent en introduisant une variable qui ne serait pas directement observable et qui jouerait un rôle analogue à celui qui est joué par Ho dans l'exemple de la vache pleine. Dans le cas de l'usine Cheap-but-Dirty, on introduirait ainsi une variable E représentant l'efficacité du fonctionnement de l'usine. De la même façon que Ho fait écran à la dépendance entre TS et TU dans l'exemple de la vache pleine, cette variable fait écran à la dépendance entre X et Y dans l'exemple de Cheap-but-Dirty.

Cette stratégie peut être étendue de façon à prendre en charge le contre-exemple que nous avons construit plus haut pour la première sous-hypothèse de la condition de Markov causale. Dans ce cas, en effet, il semble suffisant d'introduire une variable représentant le temps pour rendre indépendants le prix du pain en Grande-Bretagne et le niveau de la mer à Venise. Notons toutefois que cette variable n'est pas du même type que celles qui figurent habituellement dans l'analyse causale menée selon les méthodes qui mobilisent la notion de réseau bayésien : contrairement à ces variables, elle ne représente pas un phénomène s'intégrant dans un système réel et susceptible d'entrer dans des relations génériques de cause à effet. En outre, on voit mal d'abord comment une variable représentant le temps pourrait permettre de prendre en charge l'exemple de fourche interactive initialement proposé par Salmon et que nous avons présenté dans la sous-section 1.3.1. Rappelons ici que la fourche interactive envisagée par Salmon est celle que forment un tir au billard, la chute de la boule noire dans un filet de la table, et la chute de la boule de choc dans l'autre filet sous deux hypothèses : 1) « le joueur peut mettre la boule noire dans le filet à un bout de la table si et presque seulement si sa boule de choc va dans le filet à l'autre bout de table » et 2) le tireur « a 50 % de chances de mettre la boule noire dans le filet s'il essaie » (Salmon, 1980, p. 150-151[8]). Dans ce cas, on voit mal quelle variable pourrait être

[8] La formulation de Salmon fait intervenir des événements comme *relata* de la causalité. Nous avons montré (sous-section 1.3.1) qu'on peut en proposer une reformulation qui fait intervenir des propriétés. Au vu de ce que nous avons dit ensuite (sous-section 3.4.2) concernant le rapport entre causalité entre propriétés et causalité entre variables,

introduite qui, à la manière de *Ho* dans l'exemple de la vache pleine ou de *E* dans le cas de Cheap-but-Dirty, suffirait à rétablir la condition de Markov causale. Au-delà, il semble bien que rien ne vient garantir que la stratégie envisagée puisse être généralisée, et permette de prendre en charge tous les contre-exemples à la condition de Markov causale.

L'apparence selon laquelle la stratégie que nous avons envisagée avec Gillies ne peut pas être généralisée se dissipe quand on dissocie identification d'une variable et définition d'une variable. Dans les deux cas relativement auxquels nous avons soutenu que la stratégie envisagée était efficace, la variable à introduire s'identifie facilement parce qu'elle se définit aisément dans le langage naturel – à défaut d'être observable. Mais la propriété d'être aisément identifiable n'est pas co-extensive, pour une variable, de celle d'être définissable. Surtout, même quand nous ne savons pas identifier la variable qui fait disparaître une violation donnée de la condition de Markov causale, il est possible de définir une telle variable. On dit alors qu'on introduit un « nœud caché ». Kwoh et Gillies, 1996, introduisent une méthode pour définir les probabilités sur les nœuds cachés[9]. Plus précisément, considérant un ensemble de variables et admettant qu'un nœud caché a été introduit pour toute dépendance probabiliste relative qui constitue une violation de la condition de Markov causale, ils indiquent comment définir sur l'ensemble comprenant les variables initiales et les nœuds cachés, une fonction qui 1) conserve les probabilités sur l'ensemble de variables initial ; 2) étend la fonction de probabilités conditionnelles initiale ; 3) est telle que la condition de Markov est satisfaite pour le graphe et la fonction de probabilités étendus (Kwoh et Gillies, 1996, section 3).

Ainsi, les contre-exemples à la condition de Markov causale peuvent tous être réintégrés au domaine de pertinence des méthodes RB pourvu qu'on remette en cause l'hypothèse selon laquelle le graphe représentant les relations de cause à effet directes sur un ensemble de variables directement observables a pour sommets seulement des variables de ce type. Il en découle que, si l'introduction de variables a souvent été proposée comme réponse aux violations de la condition de Markov causale (en particulier : Spirtes *et al.*, 1993, p. 32-37 et Pearl, 2000, p. 62), cette réponse a généralement été critiquée pour ce que les variables ainsi introduites ne représentent rien qui soit susceptible d'entrer dans une relation de cause à effet (en particulier : Cartwright, 2001, p. 259 ; Williamson, 2001, §2). Formulée autrement, la critique consiste à faire valoir que le sur-graphe qu'on considère finalement n'est plus causal. Il nous semble que cette critique n'est pas rédhibitoire si l'on tient compte du fait que le graphe

il en découle que l'exemple peut également être formulé de telle façon que les *relata* causaux considérés sont des variables.

[9] Notons que l'auteur dont il est question ici est Duncan Gillies, et non Donald Gillies à qui nous avons emprunté l'exemple de la vache pleine.

causal sur un ensemble de variables donné n'est rien d'autre, encore une fois, qu'une *représentation* des relations de cause à effet directes au sein d'un ensemble $\mathbf{V}$ de variables observables. En effet, on ne voit pas pourquoi ces relations ne pourraient pas être représentées par un graphe sur $\mathbf{V}$ éventuellement augmenté d'un ensemble $\mathbf{V}'$ de variables non observables, selon les modalités suivantes :

- une flèche d'une variable de $\mathbf{V}$ à une variable de $\mathbf{V}$ représente une relation de cause à effet directe entre ces deux variables ;
- une flèche d'une variable V' de $\mathbf{V}'$ à une variable V_1 de $\mathbf{V}$
  - représente une relation de cause à effet directe entre V_2 et V_1 s'il existe V_2 de $\mathbf{V}$ qui est un parent de V',
  - ne représente aucune relation de cause à effet s'il n'existe pas de tel V_2.

De même que pour l'hypothèse d'acyclicité, on peut, en renonçant au mode le plus immédiat de représentation de la causalité directe au moyen d'un graphe orienté, rétablir la condition de Markov causale quand cette hypothèse est violée. En termes plus généraux, il semble bien que les limites que constituent, pour les méthodes RB, les violations de l'hypothèse d'acyclicité et de la condition de Markov causale se dépassent aisément. Il nous reste toutefois à montrer si et comment les artifices que nous venons de décrire peuvent être effectivement utilisés.

Utilisation effective des artifices

L'objet de l'inférence causale probabiliste est de construire, à partir de connaissances relatives aux probabilités au sein d'un ensemble de variables donné, un graphe orienté représentant les relations de cause à effet directes sur cet ensemble. Ce graphe n'est donc pas connu initialement ; il est ce qu'on prétend inférer. Or, les graphes qui sont construits dans le cadre des méthodes RB sont, par définition de ces méthodes, des graphes acycliques dans lesquels chaque variable de l'ensemble de variables considéré apparaît une et une seule fois. Dès lors, l'artifice proposé par Williamson afin de garantir l'acyclicité du graphe représentant les relations de cause à effet au sein de l'ensemble de variables considéré, ne peut pas être utilisé. En effet, cet artifice consiste essentiellement à faire figurer certaines variables deux fois dans les graphes qui représentent la causalité au sein d'ensembles de variables pour lesquels elle n'est pas asymétrique. Si la causalité directe n'est pas *effectivement* asymétrique sur l'ensemble de variables qu'on considère, les méthodes fondées sur les réseaux bayésiens ne peuvent pas construire une représentation adéquate de la causalité directe au sein de cet ensemble. L'artifice de Williamson n'est d'aucune pertinence relativement au contexte inférentiel qui nous intéresse ici.

En première approche, le cas de la condition de Markov causale semble plus complexe. D'un côté, un argument analogue à celui que nous venons de développer concernant l'hypothèse d'acyclicité se présenterait ainsi :

les méthodes RB construisent des graphes dans lesquels les seules variables de l'ensemble considéré figurent, et il en découle que les artifices permettant de rétablir la condition de Markov causale quand elle est violée, ne peuvent pas être effectivement utilisés dans ce cadre. Mais, de l'autre côté, parmi les algorithmes qui peuvent être utilisés dans le cadre des méthodes RB, certains, plus perfectionnés que celui que nous avons présenté en détail dans la sous-section 3.2.3, prennent en compte la possibilité qu'existent des causes communes à des variables de l'ensemble considéré qui n'appartiennent pas elles-mêmes à cet ensemble [10]. Cela suggère que le raisonnement que nous venons d'esquisser n'est pas correct. En effet, les artifices que nous avons présentés consistent précisément à introduire de nouvelles variables qui viennent jouer le rôle de causes communes à des variables de l'ensemble initial. Positivement, nous sommes conduits à envisager que ces artifices sont effectivement mis en œuvre par les algorithmes perfectionnés auxquels nous faisions allusion.

L'hypothèse, toutefois, ne résiste pas à l'examen, en particulier parce qu'artifices et algorithmes perfectionnés ne visent pas exactement les mêmes cas. Les artifices définis dans le paragraphe précédent visent, en particulier, à intégrer au domaine de ce qui est représentable moyennant la condition de Markov causale, les cas tels que soit il n'existe aucune relation causale entre deux variables pourtant dépendantes, soit deux variables dont aucune des deux n'est cause de l'autre, ne sont pas indépendantes relativement à une cause qui leur est commune. Or, les algorithmes perfectionnés ne visent qu'à prendre en compte le fait que certaines causes communes à des variables de l'ensemble de variables observées $\mathbf{V}$ peuvent ne pas appartenir elles-mêmes à $\mathbf{V}$ – exactement de la même façon que la variable *Ho* avait été initialement omise dans l'exemple de la vache pleine. Les algorithmes perfectionnés prennent donc en compte un seul type de violations de la condition de Markov causale, là où les artifices évoqués dans le paragraphe précédent rétablissent la condition de Markov causale dans tous les cas. On ne peut donc pas considérer que les artifices que nous avons présentés sont effectivement utilisés dans le cadre des algorithmes perfectionnés. Surtout, de façon plus générale, les algorithmes d'inférence causale mobilisant les réseaux bayésiens mènent des inférences valides seulement quand l'hypothèse d'acyclicité et la condition de Markov causale sont effectivement satisfaites pour la représentation graphique naturelle des relations de cause à effet directes sur $\mathbf{V}$. Les artifices présentés ci-dessus ne peuvent donc pas être effectivement utilisés dans le contexte d'inférence causale probabiliste.

[10] Il s'agit, en particulier, des algorithmes IC*, CI, FCI.

4.3.2 Conséquences des violations de l'hypothèse d'acyclicité et de la condition de Markov causale

Nous venons de montrer que l'existence des artifices que nous avons présentés n'impliquait pas que les violations de l'hypothèse d'acyclicité et de la condition de Markov causale sont dépourvues d'impact sur l'inférence causale probabiliste menée selon les méthodes RB. Afin de mesurer la portée de ces méthodes, il nous faut maintenant déterminer quelles sont, positivement, les conséquences de ces violations relativement à l'inférence causale probabiliste menée selon les méthodes RB.

Quelle conclusion tirer de l'existence des violations ?

Le résultat d'une inférence causale menée selon les méthodes RB est une conséquence logique de l'ensemble des indépendances probabilistes relatives que l'algorithme mobilisé dans le cadre de l'inférence prend pour entrée seulement si l'hypothèse d'acyclicité et la condition de Markov causale sont satisfaites. Or, nous avons vu qu'elles ne le sont pas toujours : il existe des systèmes réels et des ensembles de variables représentant des aspects observables de ces systèmes pour lesquels elles ne le sont pas. Il semble en découler que les méthodes d'inférence causale probabiliste fondées sur les réseaux bayésiens doivent être réservées aux systèmes et aux ensembles de variables pour lesquels ces hypothèses sont satisfaites. En d'autres termes, il semble que l'existence de violations de ces hypothèses limite le domaine dans lequel on peut recourir aux méthodes RB. La portée de ces méthodes serait alors décrite au moyen de la proposition suivante : elles autorisent des inférences causales probabilistes déductives (et a-théoriques) pour les ensembles de variables satisfaisant l'hypothèse d'acyclicité et la condition de Markov causale.

Cette première analyse, toutefois, est trompeuse. Plus précisément, l'hypothèse d'acyclicité et la condition de Markov causale ne peuvent pas être simplement considérées comme les limites d'un domaine à l'intérieur duquel l'inférence causale probabiliste pourrait être menée selon les méthodes RB et, donc, être a-théorique et déductive au sens que nous avons dit. La raison en est simple : *puisque les inférences causales menées selon les méthodes RB sont a-théoriques*, l'enquête qui viserait à déterminer avant coup si les hypothèses requises pour la validité de l'inférence sont satisfaites n'a simplement pas d'objet. En effet, si l'on adopte les méthodes d'inférence causale mobilisant les réseaux bayésiens, on n'a pas, avant de mener l'inférence causale probabiliste, de graphe causal dont on pourrait se demander s'il est acyclique et s'il est avec les probabilités dans un rapport tel que la condition de Markov causale soit satisfaite.

En vue d'expliciter cette thèse, considérons à nouveau la condition de Markov causale. Nous connaissons des conditions suffisantes auxquelles un ensemble de variables donné la satisfait. En particulier, la condition de

Markov causale est prouvablement satisfaite par les ensembles de variables déterministes et qui sont tels que les sous-ensembles non vides et disjoints de l'ensemble des variables exogènes sont deux à deux indépendants en probabilité. Toutefois, dans le contexte d'inférence causale a-théorique que dessinent les méthodes RB, ce résultat ne peut pas être utilisé comme un critère *d'identification* d'ensembles de variables qui satisfont la condition de Markov causale, parce que le critère envisagé est causal à deux titres : d'une part, les variables exogènes d'un ensemble de variables **V** sont celles qui n'ont pas de *cause* dans **V** ; d'autre part, un ensemble de variables **V** est déterministe si la valeur des variables non exogènes de **V** est déterminée par celle de leurs *causes* directes dans **V**. Dès lors, décider si **V** est déterministe et tel que les sous-ensembles non vides et disjoints de l'ensemble de ses variables exogènes sont deux à deux indépendants requiert de connaître les relations de cause à effet au sein de **V**. Or, c'est précisément là ce que vise l'inférence causale probabiliste menée selon les méthodes RB.

De façon plus générale, les hypothèses mises en évidence dans la section 4.1 ont un contenu causal, et il en découle qu'on ne peut pas déterminer si elles sont satisfaites par un ensemble de variables **V** si l'on ne dispose pas d'un graphe causal, même hypothétique, sur **V**. En outre, l'absence de connaissance (et même seulement d'hypothèse) concernant la causalité au sein de **V** n'implique pas seulement qu'il est impossible de déterminer si **V** satisfait les hypothèses. Elle a également pour corrélat qu'il est impossible d'estimer l'écart maximum qui peut exister, pour une mesure de distance entre graphes, entre le graphe qu'on obtient en appliquant les méthodes RB à **V** et le graphe causal correct sur **V**. Dans ces conditions, entre la conclusion d'une inférence causale menée selon les méthodes RB et le graphe causal sur l'ensemble de variables considéré, il n'existe aucune forme d'adéquation qui soit garantie. L'hypothèse d'acyclicité et la condition de Markov causale sont donc bien plus que ce qui limite le domaine à l'intérieur duquel les inférences causales probabilistes peuvent être menées selon les méthodes RB et, de ce fait, jouir de la propriété d'être déductives. Le fait qu'elles soient, parfois, violées a pour conséquence que les méthodes d'inférence causale fondées sur la notion de réseau bayésien ne sont jamais fiables – au sens précis où on ne peut jamais être assuré qu'elles donnent un résultat correct, même quand c'est le cas. Ainsi, il semble que le prix à payer pour la possibilité d'induire des relations causales à partir de données statistiques est tel que, en fait, il annule cette possibilité. En d'autres termes, les conséquences à tirer de l'existence de violations de l'hypothèse d'acyclicité et de la condition de Markov causale sont de nature à limiter drastiquement, sinon peut-être à annihiler, la portée des méthodes RB pour l'inférence causale probabiliste. Dans le prochain paragraphe, nous verrons que ce que nous venons de montrer est spécifique des méthodes RB. Si ces méthodes sont *grosso modo* limitées par les mêmes faits que les méthodes traditionnelles, les conséquences

qu'ont ces limitations sont différentes dans l'un et l'autre cas – et ces différences sont fondamentales pour ce qui est de la portée des méthodes fondées sur les réseaux bayésiens.

L'inférence causale probabiliste face à ses limites : spécificité des méthodes RB

Si l'existence de violations de la condition de Markov causale n'a pas les mêmes conséquences pour les méthodes RB et pour les méthodes traditionnelles, c'est d'abord parce que ces hypothèses elles-mêmes n'ont pas le même statut dans l'un et dans l'autre cas. Pour le comprendre, revenons à l'idée selon laquelle les méthodes traditionnelles véhiculent, au moins en pratique, une conception du rapport entre la causalité et les probabilités qui est similaire à celle qui sous-tend les méthodes RB. Pour ce faire, nous avons indiqué en quels points de la procédure d'inférence définie dans la section 2.2, cette conception est mobilisée. À l'inverse, dans les procédures d'inférence causale fondées sur les réseaux bayésiens, nous ne pouvons pas localiser le recours à la condition de Markov causale. En effet, ce recours constitue le *principe* même de l'inférence causale menée selon les méthodes RB, qui en cela se distinguent des méthodes plus traditionnelles.

Par ailleurs, nous savons (section 3.3) que les inférences menées selon les méthodes RB sont a-théoriques et nous venons de montrer que cela implique qu'elles sont toujours suspectes de donner des résultats incorrects. De l'autre côté, les inférences causales probabilistes plus traditionnelles ne sont pas a-théoriques ; elles reposent sur la spécification de modèles structurels. Il en découle qu'on peut tester si un modèle donné satisfait la condition de Markov causale. Plus précisément, il est possible de procéder à un test statistique de l'hypothèse selon laquelle la condition de Markov causale serait satisfaite dans le cas où le modèle spécifié serait correct. Il en va d'ailleurs de même pour l'hypothèse de fidélité causale dont nous avons soutenu que, dans le cadre des méthodes fondées sur les réseaux bayésiens, elle doit être considérée comme une hypothèse méthodologique dont les violations ne peuvent pas être envisagées. Ainsi, de ce que méthodes RB et méthodes traditionnelles véhiculent des conceptions similaires du rapport entre la causalité et les probabilités, il ne découle pas que les contre-exemples à ces conceptions aient, dans l'un et dans l'autre cas, les mêmes conséquences relativement à l'inférence causale probabiliste. Si les limites que constituent les violations de la condition de Markov causale et de l'hypothèse de fidélité causale ne sont pas spécifiques des méthodes RB, le statut de ces hypothèses et les conséquences de ces violations le sont.

Avant de conclure relativement à la portée des méthodes recourant aux réseaux bayésiens, nous souhaitons souligner qu'il n'y a pas d'analogue pour les méthodes traditionnelles de ce que sont, pour ces méthodes

récentes, les violations des hypothèses mises en évidence dans la section 4.1. Plus précisément, nous avons mentionné dans la sous-section 2.1.1 que l'utilisation de l'une ou l'autre technique pour l'estimation des paramètres d'un modèle ou l'évaluation de son adéquation aux données dépend des hypothèses statistiques qu'on formule relativement au modèle étudié (voir par exemple Kline, 1998, ou Kenny, 1979, pour une présentation, et Freedman, 1987, p. 101-116, ou Clogg et Haritou, 1997, pour une analyse). Ces hypothèses sont contraignantes et le fait qu'elles soient parfois émises à tort conduit à des mésusages des outils de la modélisation structurelle. Ces mésusages ont été décrits aussi bien dans les ouvrages méthodologiques (par exemple, Kline, 1998, chap. 12), que dans la littérature critique (par exemple, Freedman, 1991) ; on a même cherché à les expliquer (Blalock, 1991). Pourtant, aussi dommageables ces mésusages soient-ils, leurs conséquences ne sont pas comparables à celles, décrites plus haut, de l'utilisation dans le cas général des procédures d'inférence causale mobilisant la notion de réseau bayésien. En effet, les hypothèses que ces mésusages consistent à négliger ne sont pas du même type que l'hypothèse d'acyclité, la condition de Markov causale ou l'hypothèse de fidélité causale. Contrairement à celles-ci, elles ne sont pas causales, mais portent seulement sur les fonctions de probabilités. Dès lors, elles peuvent être testées avant d'utiliser ou après avoir utilisé les outils de la modélisation structurelle qui requièrent qu'elles soient satisfaites. La difficulté qu'elles constituent pour les méthodes traditionnelles n'est donc en rien analogue à la difficulté théorique constituée, pour les méthodes RB, par les violations de l'hypothèse d'acyclicité ou de la condition de Markov causale.

Finalement, ce qui distingue les méthodes RB des méthodes hypothético-déductives plus traditionnelles peut être résumé ainsi : d'une part, les inférences qu'elles guident sont déductives ; d'autre part, l'existence de violations de l'hypothèse d'acyclicité et de la condition de Markov causale a pour conséquence que ces inférences ne sont jamais fiables. Le second point implique-t-il que les méthodes fondées sur le recours aux réseaux bayésiens ne changent, en fait, rien à l'inférence causale probabiliste, qu'elles ne peuvent pas être utilisées pour induire des connaissances causales, et qu'en dépit du premier point leur portée peut être considérée comme nulle ? Nous soutenons que tel n'est pas le cas et que les méthodes RB peuvent contribuer à l'inférence causale probabiliste si on les intègre à ce que nous appelons une « méthodologie mixte ».

4.4 Pour une méthodologie mixte

La fin du présent chapitre est essentiellement prescriptive. Il s'agit de déterminer comment les méthodes RB *pourraient* être utilisées, et la déductivité qu'elles autorisent être mise à profit, dans le cadre de l'inférence causale probabiliste. Étant donné ce que nous venons de montrer,

il semble clair que cela n'est possible qu'à la condition d'intégrer les algorithmes caractéristiques des méthodes RB, et sur lesquels leur déductivité repose, à une procédure qui soit hypothético-déductive. C'est en ce sens que les méthodes d'inférence causale que nous nous apprêtons à envisager sont mixtes. Par ailleurs, nous avons également établi que les méthodes traditionnelles reposent sur une conception du rapport entre causalité et probabilités qui est similaire à celle que véhiculent la condition de Markov causale et l'hypothèse de fidélité causale, dont nous savons qu'elles peuvent être violées. Ainsi, au total, la présente section vise à proposer une procédure d'inférence causale probabiliste hypothético-déductive 1) qui utilise les algorithmes RB et, en ce sens, bénéficie de leurs intéressantes propriétés et 2) telle que les violations éventuelles de la condition de Markov causale et de l'hypothèse de fidélité causale soient prises en compte.

4.4.1 Discussion d'une proposition existante

Il existe déjà une proposition de méthodologie mixte, qui ménage dans un cadre hypothético-déductif une place aux « algorithmes RB », c'est-à-dire aux algorithmes qui, à partir des indépendances probabilistes au sein d'un ensemble de variables, construisent l'ensemble des graphes orientés acycliques entretenant avec ces indépendances le rapport R qui caractérise les réseaux bayésiens. Selon cette proposition et pour ce qui concerne plus précisément la procédure hypothético-déductive que nous avons définie dans la section 2.2, les algorithmes caractéristiques des méthodes RB seraient utilisés à l'étape A. Ils servent donc à formuler des hypothèses causales. La proposition a déjà été évoquée dans Glymour *et al.*, 1988, p. 428-429, mais c'est Williamson qui l'a développé avec le plus de soin, dans Williamson, 2002. La proposition alors formulée est la suivante : utiliser les algorithmes RB afin de formuler une hypothèse, déduire de cette hypothèse des prédictions, tester ces prédictions, amender l'hypothèse en fonction du résultat des tests, et mener les mêmes tâches pour la nouvelle hypothèse (Williamson, 2002, p. 6-7).

Le principal avantage de cette proposition est qu'elle utilise les algorithmes RB afin de combler une lacune patente de la méthodologie hypothético-déductive. En effet, Popper ne propose pas de méthode pour formuler des hypothèses – considérant d'ailleurs qu'il ne saurait exister une méthode générale à cet effet. En particulier, il ne propose pas de méthode pour formuler des hypothèses causales. En utilisant les algorithmes RB à l'étape A de la procédure d'inférence causale hypothético-déductive, on se donne une telle méthode.

La première limite de cette proposition est qu'elle ne comble pas *complètement* la lacune qu'elle vise à combler. Rappelons, en effet, que le résultat d'un algorithme RB n'est pas un modèle, mais un patron représentant une classe de modèles. Williamson le mentionne, mais considère

que cela ne doit pas être considéré comme problématique : « Ici, les techniques de l'intelligence artificielle sont utilisées pour engendrer un modèle structurel (ou un ensemble de modèles structurels, auquel cas plusieurs hypothèses sont évaluées simultanément – j'emploierai le singulier dans la suite pour des raisons de simplicité) [*for simplicity's sake*] »(Williamson, 2002, p. 7). Nous ne soutenons pas que Williamson ait tort de considérer que plusieurs hypothèses puissent être testées simultanément. Il suffit pour cela de tester une conséquence qui est commune à ces hypothèses. Toutefois, le fait que les algorithmes RB ont pour résultats des patrons, et non des graphes orientés acycliques, interdit de leur attribuer le rôle précis que Williamson envisage. Si, d'un côté, la conclusion de l'inférence causale probabiliste doit bien être *un* modèle structurel et si, de l'autre côté, l'inférence doit bien être hypothético-déductive, alors il faut à un moment ou à un autre formuler une hypothèse qui soit *un* modèle structurel. En conséquence, les algorithmes RB ne peuvent pas être le seul outil pour la spécification d'hypothèses causales, contrairement à ce que Williamson laisse entendre. La spécification d'hypothèses causales repose alors toujours, *in fine*, sur des considérations théoriques et, du coup, il est difficile de distinguer exactement le gain méthodologique qu'il convient d'associer à la proposition de Williamson.

En outre, cette proposition souffre à nos yeux de ce qu'elle ne prend pas en compte la possibilité que la condition de Markov causale et l'hypothèse de fidélité causale soient violées. D'une part, tels que Williamson propose de les intégrer à l'inférence causale hypothético-déductive, les algorithmes RB sont utilisés de la manière aveugle que nous avons critiquée dans la sous-section 4.3.2. D'autre part, la proposition de Williamson ne fait pas cas de ceci que la condition de Markov causale et l'hypothèse de fidélité causale sont sous-jacentes à la pratique de l'inférence causale hypothético-déductive et supposées par certains de ses outils. Ce point ne peut vraisemblablement pas être opposé à Williamson lui-même, dont la proposition est formulée à un niveau d'abstraction plus élevé que celui auquel nous nous plaçons ici, et tel que la nature des outils de la modélisation causale importe peu. Il n'en reste pas moins que ce point constitue une raison pour nous de rejeter l'idée selon laquelle les algorithmes RB pourraient simplement servir à formuler des hypothèses causales.

4.4.2 Une nouvelle proposition

Si les algorithmes RB ne peuvent pas être utilisés au moment de formuler les hypothèses causales, il reste à envisager de s'en servir pour tester de telles hypothèses. Sous sa spécification la plus naturelle, la proposition consiste à employer les algorithmes RB pour fonder un nouveau test qui viendrait prendre place à l'étape C de la procédure d'inférence causale probabiliste que nous avons définie dans la section 2.2.

Un nouveau test à l'étape C

Plus précisément, la suggestion est la suivante : pour les hypothèses causales 1) qui ont passé tous les tests usuels (ceux que nous avons détaillés dans la sous-section 2.2.3 à l'étape C et 2) telles qu'on ne peut pas rejeter l'hypothèse selon laquelle la condition de Markov causale et l'hypothèse de fidélité causale sont satisfaites, vérifier qu'elles font bien partie de la classe de modèles qui est représentée par la sortie de l'algorithme RB qu'on aura choisi. Dans le cas où elle ne l'est pas, une hypothèse causale de ce type doit être rejetée ; dans le cas où elle l'est, elle s'en trouve corroborée.

À l'appui de cette proposition, commençons par souligner qu'il est effectivement possible de tester si la condition de Markov causale et l'hypothèse de fidélité causale seraient satisfaites dans le cas où un modèle structurel donné serait adéquat. C'est ce que nous faisions valoir, déjà, dans le paragraphe de la sous-section 4.3.2 consacré au statut des limites de l'inférence causale probabiliste menée selon les méthodes RB. Plus important, la proposition prend explicitement en compte la possibilité que la condition de Markov causale et l'hypothèse de fidélité causale soient violées. Finalement, s'il est vrai que le test concernant la condition de Markov causale et l'hypothèse de fidélité causale ne serait pas complètement fiable parce que l'hypothèse à tester est statistique, cela ne saurait compter contre la proposition que nous discutons. En effet, nous avons vu (sous-section 4.2.2.) que les difficultés associées au fait d'inférer des conclusions relatives à une population à partir de données portant sur un échantillon de cette population sont générales, non spécifiques d'un type d'inférence causale probabiliste. Ces difficultés ne peuvent pas être évitées, et le fait qu'elle ne donne pas de moyen de les éviter ne saurait donc compter contre la proposition que nous discutons.

Il nous reste toutefois à examiner en quel sens le test proposé serait instructif. Or, à ce point, les choses deviennent moins favorables. En effet, nous savons maintenant que l'inférence causale probabiliste hypothético-déductive repose sur une conception du rapport entre la causalité et les probabilités qui est similaire à celle que définissent ensemble la condition de Markov causale et l'hypothèse de fidélité causale. Nous avons indiqué que cette conception est mobilisée au moment où les hypothèses sont spécifiées, puis à l'occasion de différents types de tests qui peuvent être menés à l'étape C. Dans ces conditions, une hypothèse qui aurait été spécifiée en A et ne serait pas rejetée à l'issue des tests usuels pour l'étape C aurait ce statut en grande partie parce qu'elle garantirait que le rapport entre la causalité et les probabilités est bien celui que dessinent la condition de Markov causale et l'hypothèse de fidélité causale. Elle ne pourrait donc pas manquer d'appartenir à la classe des graphes orientés acycliques qui constituerait le résultat d'un algorithme. Aussi, un test dont l'issue dépend de ce qu'une hypothèse non rejetée à l'issue des tests

usuels pour l'étape C appartient ou non à cette classe ne saurait être instructif.

Contre l'objection ainsi exprimée, on peut revenir sur le statut de la conception du rapport entre causalité et probabilités dans le cadre de l'inférence causale probabiliste hypothético-déductive et rappeler que, dans ce cadre, la conception du rapport entre causalité et probabilités n'est mobilisée que localement. Cela semble impliquer qu'elle est dispensable : elle n'intervient que dans des zones délimitables de l'inférence, et n'en constitue pas le principe. Dans ces conditions, on pourrait imaginer accompagner la proposition que nous avançons, de l'injonction de ne pas recourir à des hypothèses relatives au rapport entre la causalité et les probabilités *avant* le nouveau test. C'est ce qu'on fait si

- on ne fait pas dépendre la formulation d'hypothèses causales (à l'étape A) de considérations probabilistes – mais seulement de considérations théoriques ;
- on réalise le nouveau test non pas après, mais *avant* les tests usuels pour l'étape C.

Toutefois, cela est sans compter que la façon dont les paramètres structurels sont estimés (à l'étape B) repose elle aussi sur la condition de Markov causale. Plus précisément, le paramètre mesurant l'effet d'une variable C sur une variable E est estimé en tenant fixée la valeur des autres causes supposées de E. À proprement parler, il n'y a rien ici qui requière que la condition de Markov causale soit satisfaite. Aussi n'avons-nous pas mentionné ce point au début de la sous-section 4.2.2. Toutefois, ce mode d'estimation des paramètres ne fait sens que moyennant l'idée selon laquelle ses causes prises ensemble suffisent à expliquer les variations d'une variable et ses corrélations avec les autres variables.

Dans ces conditions, prendre au sérieux l'injonction d'ignorer le rapport supposé entre la causalité et les probabilités jusqu'à la mise en œuvre du test mobilisant les algorithmes RB, conduit à renoncer à la fois à l'interprétation causale des paramètres estimés à l'étape B et aux tests usuels qui les concernent. En ce sens, la proposition de mobiliser les algorithmes RB à l'étape C avant d'avoir recouru à des hypothèses relatives au rapport entre causalité et probabilités revient pratiquement à renoncer à l'inférence causale probabiliste hypothético-déductive. La prise en compte, dans le cadre hypothético-déductif, de la possibilité que la condition de Markov causale et l'hypothèse de fidélité causale soient violées n'est pas compatible avec le fait d'utiliser les algorithmes RB si tard dans la procédure d'inférence causale.

Un test entre l'étape A et l'étape B

L'idée que nous envisageons maintenant est la suivante : utiliser les algorithmes RB pour tester les hypothèses causales dès après leur formulation. Plus précisément, la proposition n'a de sens que si les hypothèses causales

sont formulées indépendamment de considérations probabilistes (mais plutôt, en particulier, sur la base de considérations théoriques d'arrière-plan). En outre, le test ne peut être mené que pour les modèles structurels acycliques et tels qu'on ne peut pas rejeter l'hypothèse selon laquelle la condition de Markov causale et l'hypothèse de fidélité causale seraient satisfaites dans le cas où ils seraient corrects. Pour de tels modèles, le test consisterait comme plus haut à vérifier que le modèle structurel envisagé appartient bien à la classe des graphes orientés acycliques qui constitue le résultat des algorithmes RB. Si tel n'était pas le cas, ce modèle pourrait être rejeté dès l'étape que nous envisageons maintenant, et que nous proposons de noter A'.

La proposition que nous venons de décrire présente les mêmes avantages que celle que nous venons d'examiner pour finalement la rejeter. Elle permet d'utiliser les algorithmes RB dans le cadre de l'inférence causale probabiliste hypothético-déductive, et de les utiliser seulement quand les hypothèses requises pour ces algorithmes sont satisfaites ou, plus précisément, seulement quand elles seraient satisfaites au cas où le modèle envisagé serait correct. Mais, de l'autre côté, elle n'a pas les inconvénients qui nous ont arrêtés plus haut. En premier lieu, le test est instructif et, pourvu que la formulation de modèles causaux hypothétiques à l'étape A ne repose pas sur des considérations probabilistes, il doit permettre de rejeter certains de ces modèles. En second lieu, la nouvelle proposition tient compte de la possibilité que la condition de Markov causale et l'hypothèse de fidélité causale soient violées. Plus, elle contribue positivement à la prise en compte de cette possibilité, puisqu'elle requiert qu'on recherche dès la formulation d'un modèle si ces hypothèses seraient satisfaites dans le cas où le modèle envisagé serait adéquat.

Notre proposition ne sera complète qu'après qu'on aura précisé ce qu'il convient de faire si l'on ne peut pas rejeter l'hypothèse selon laquelle la condition de Markov causale ou l'hypothèse de fidélité causale serait violée dans le cas où le modèle structurel envisagé serait adéquat. À cette question, nous répondons différemment pour ce qui concerne la condition de Markov causale et pour ce qui concerne l'hypothèse de fidélité. Ces réponses différentes correspondent aux rôles différents que jouent les deux hypothèses dans le cadre de l'inférence causale probabiliste hypothético-déductive. Ainsi, prendre en compte le fait que l'hypothèse de fidélité causale serait violée dans le cas où le modèle envisagé serait correct implique seulement de ne pas mobiliser cette hypothèse quand il s'agit de tester ce modèle à l'étape C. Plus précisément et en entrant dans le détail qui a fait l'objet de la sous-section 2.2.3, cela implique seulement que, pour rejeter un modèle structurel à l'issue des tests de type [a.], il n'est pas suffisant de ne pas pouvoir rejeter l'hypothèse statistique selon laquelle tous les paramètres sont significativement différents de zéro. En effet, une violation de l'hypothèse de fidélité causale correspond à l'existence d'une relation de cause à effet qui ne se marque pas dans les probabilités et donc pour laquelle le paramètre associé est nul.

Le cas de la condition de Markov causale est différent, puisque nous avons vu qu'elle était sous-jacente à l'estimation des paramètres causaux à l'étape B. Dans ces conditions, prendre en compte une violation de la condition de Markov causale, c'est (encore une fois) pratiquement renoncer à l'inférence causale probabiliste hypothético-déductive elle-même. Ne pas rejeter immédiatement un modèle tel que la condition de Markov causale serait violée dans le cas où il serait correct requiert donc d'avoir de très bons arguments théoriques en faveur de ce modèle.

Pour ce qui est de la question computationnelle, elle se pose et ne peut recevoir une réponse complètement satisfaisante : les algorithmes RB étant ce qu'ils sont, leur complexité ne varie pas. À l'attention du lecteur légitimement inquiet, nous formulons néanmoins trois remarques. En premier lieu, nous avons pris soin de ne pas imposer que le nouveau test soit mené pour chaque hypothèse causale. Il est mené seulement pour les modèles acycliques et tels qu'on ne peut pas rejeter l'hypothèse selon laquelle la condition de Markov causale et l'hypothèse de fidélité causale seraient satisfaites dans le cas où ils seraient adéquats. En deuxième lieu, remarquons que l'algorithme RB qu'on aura choisi n'a pas à être utilisé à nouveaux frais à chaque fois qu'est mené le nouveau test que nous envisageons, c'est-à-dire pour chaque nouveau modèle structurel. En effet, le résultat d'un algorithme RB ne dépend que de l'ensemble de variables qu'on considère et des indépendances probabilistes relatives au sein de cet ensemble ; il est donc inchangé par le modèle structurel hypothétique qu'il y a à tester. En troisième lieu, enfin, notons que si des hypothèses causales peuvent être rejetées à l'issue de l'étape A' que nous envisageons, alors il n'est nul besoin de mener les étapes B et C pour ces hypothèses – ce qui entraîne des gains en termes de computation.

Reste, finalement, la question de l'articulation avec les théories probabilistes de la causalité. Puisqu'elle prend en compte la possibilité que la condition de Markov causale et l'hypothèse de fidélité causale soient violées, la méthode que nous proposons n'est pas liée à une conception de la causalité qui serait plus fruste que les théories de la causalité postérieures à celle de Suppes. Plus précisément, ce n'est pas parce qu'elle reposerait sur une conception de la causalité particulièrement fruste que cette méthode permet d'inférer des connaissances causales à partir de données statistiques d'observation et des connaissances probabilistes qu'on peut tirer de telles données. Les obstacles qui apparaissent si l'on prétend utiliser les théories probabilistes comme principes d'inférence causale sont contournés selon les mêmes voies qu'emprunte l'inférence causale probabiliste hypothético-déductive traditionnelle : d'une part, une notion de causalité qui est relative à un ensemble de variables, et, d'autre part l'hypothético-déductivité elle-même. Enfin, rappelons une dernière fois que l'inférence causale probabiliste est toujours une inférence statistique, avec les limites que cela comporte.

Conclusion

NOTRE OBJECTIF dans cet ouvrage était double. Il s'agissait, en premier lieu, d'articuler l'analyse philosophique du concept de cause à la méthodologie de l'inférence causale menée à partir de données statistiques, et, en second lieu, de contribuer à porter un éclairage philosophique systématique sur ces méthodes récemment apparues et dont les partisans ont prétendu qu'elles étaient inductives.

Concernant le premier point, nous avons montré que les méthodes relevant de l'inférence causale probabiliste recourent à une conception de la causalité qui est plus fruste que toutes les théories probabilistes qui ont été développées après 1970. D'une part, elles prennent pour objet une notion de causalité qui est plus grossière que celle que visent les théories probabilistes ; d'autre part, l'analyse de la causalité qu'elles véhiculent se heurte à strictement plus de contre-exemples que toutes les théories postérieures à celle qui est développée par Suppes dans Suppes, 1970. Mais, en même temps, c'est là exactement la raison pour laquelle ces méthodes permettent d'inférer des relations de cause à effet à partir de données statistiques d'observation et des connaissances probabilistes qu'on peut tirer de telles données. En effet, nous avons expliqué pourquoi les théories probabilistes plus récentes que celle de Suppes, et en premier lieu la théorie 1.3 proposée par Cartwright et la théorie 1.4 suggérée par Skyrms, ne peuvent pas servir de principes pour l'inférence causale.

Plus précisément, ce sont les méthodes hypothético-déductives traditionnellement utilisées pour la modélisation causale qui permettent l'inférence causale probabiliste. Pour ce qui est des méthodes récemment introduites et qui reposent sur la notion de réseau bayésien, nous avons défendu la thèse selon laquelle elles ne peuvent pas être légitimement utilisées seules à des fins d'inférence causale. Même si elles visent la même notion de causalité que les méthodes hypothético-déductives et véhiculent la même notion de causalité qu'elles, elles ne sont pas fiables.

Nous avons tiré cette conclusion de l'examen systématique des méthodes mobilisant les réseaux bayésiens auquel nous nous sommes livré en réponse au second objectif de l'ouvrage. D'un côté, nous avons montré que ces méthodes sont a-théoriques : l'inférence causale prend pour seules prémisses des données d'observation, indépendamment de toute

hypothèse théorique concernant la causalité au sein du système étudié. Nous avons également montré qu'elles sont déductives, au sens où les algorithmes qu'elles mobilisent de manière centrale ont pour résultat un modèle causal adéquat si les énoncés probabilistes qu'ils prennent pour entrée sont vrais. Plus exactement, elles sont déductives moyennant l'hypothèse d'asymétrie de la causalité et deux hypothèses concernant le rapport entre la causalité et les probabilités : la condition de Markov causale et l'hypothèse de fidélité causale. Or, de l'autre côté, nous avons analysé l'hypothèse d'asymétrie de la causalité, la condition de Markov causale et l'hypothèse de fidélité causale, et pour chacune nous avons montré qu'il existe des systèmes pour lesquels elle est violée. Nous avons fait apparaître que, parce que les méthodes mobilisant les réseaux bayésiens sont a-théoriques, l'existence de violations des hypothèses sur lesquelles elles reposent a pour conséquence qu'on n'a jamais aucune forme de garantie épistémique de ce que la conclusion qu'elles délivrent est vraie, même quand elle l'est : ces méthodes récentes ne sont pas fiables.

L'a-théoricité et la déductivité, au sens où nous les avons définies, sont pourtant des propriétés intéressantes d'une inférence causale. Pour cette raison, nous avons consacré la dernière section du corps de l'ouvrage à proposer une méthode d'inférence causale probabiliste qui ménage une place pour les algorithmes qu'on trouve au cœur des méthodes mobilisant les réseaux bayésiens – et donc bénéficie des propriétés intéressantes des inférences menées selon ces méthodes –, mais évite que ces algorithmes soient utilisés dans des conditions telles que l'inférence causale probabiliste n'est pas fiable. La méthode proposée est hypothético-déductive et, parce que les hypothèses causales envisagées y font l'objet d'un test supplémentaire, elle améliore l'inférence causale probabiliste hypothético-déductive telle qu'elle est menée traditionnellement. Toutefois, si elle évite les écueils méthodologiques associés aux méthodes fondées sur les réseaux bayésiens quand elles sont utilisées seules, cette méthode a les limites des méthodes traditionnelles : les inférences qu'elles guident sont hypothético-déductives, statistiques, elles visent une notion de causalité qui est plus grossière que celle qui fait l'objet des théories probabilistes, et reposent sur une conception de la causalité plus fruste que celle que véhiculent les théories probabilistes contemporaines. Ces limites sont le prix à payer pour la possibilité de l'inférence causale probabiliste, et plus spécifiquement pour que soient contournés les obstacles à l'utilisation des théories probabilistes comme principes pour l'inférence causale.

Appendice

L'OBJECTIF de cet appendice est de présenter les principes du calcul des probabilités. Il ne s'agit donc ni de proposer des définitions au niveau mathématique le plus fondamental, ni d'énoncer (et encore moins de démontrer) des résultats élaborés. Nous renvoyons le lecteur intéressé à des manuels de calcul des probabilités ou, par exemple, à l'appendice de Suppes, 1970.

Dans cette optique, nous considérons qu'une fonction de probabilités est définie sur un ensemble de propositions clos par conjonction, disjonction et négation. Le fait que les éléments d'un ensemble sur lequel une fonction de probabilités est définie soient des propositions autorise l'utilisation des symboles $\wedge$, $\vee$ et $\neg$. Par ailleurs, il implique que des tautologies appartiennent à cet ensemble.

Dans toute la suite, $\mathbf{A}$ désigne un ensemble de propositions clos par conjonction, disjonction et négation.

Définitions

Définition A.4 (Fonction de probabilités) Une fonction de probabilités p sur $\mathbf{A}$ est une fonction de $\mathbf{A}$ dans l'ensemble des réels telle que :

1. pour tout A de $\mathbf{A}$, $p(A) \geq 0$;
2. pour toute tautologie $\top$ de $\mathbf{A}$, $p(\top) = 1$;
3. pour tout A et B de $\mathbf{A}$, si A et B sont incompatibles (c.-à-d. si $A \wedge B$ est une contradiction), alors $p(A \vee B) = p(A) + p(B)$.

Définition A.5 (Indépendance probabiliste) Soit p une fonction de probabilités sur $\mathbf{A}$ et soit A et B deux éléments de $\mathbf{A}$.
A et B sont indépendants pour p si et seulement si $p(A \wedge B) = p(A).p(B)$.

Ces définitions portent sur les probabilités absolues, de la forme $p(A)$. Cependant, une fonction de probabilités absolues p permet également de définir des probabilités conditionnelles, de la forme $p(A|B)$. Le rapport entre probabilités absolues et probabilités conditionnelles s'énonce de la façon suivante :

Proposition A.6 (Probabilités conditionnelles) Soit p une fonction de probabilités sur $\mathbf{A}$.
Pour tout A et B de $\mathbf{A}$ tels que $p(B) \neq 0$, $p(A|B) = \frac{p(A \wedge B)}{p(B)}$.

Définition A.7 (Indépendance probabiliste relative) Soit p une fonction de probabilités sur $\mathbf{A}$ et soit A, B et C trois éléments de $\mathbf{A}$.
A et B sont indépendants pour p relativement à C si et seulement si $p(A \wedge B|C) = p(A|C).p(B|C)$.

Quelques résultats

Dans cette section, p est une fonction de probabilités sur $\mathbf{A}$.

Théorème A.8 *Pour tout A de $\mathbf{A}$, $0 \leq p(A) \leq 1$.*

Théorème A.9 *Pour tout A de $\mathbf{A}$, $p(\neg A) = 1 - p(A)$.*

Théorème A.10 *Pour toute contradiction $\perp$ de $\mathbf{A}$, $p(\perp) = 0$.*

Théorème A.11 *Pour tout A et B de $\mathbf{A}$, $p(A) = p(B)$ si A et B sont logiquement équivalents.*

Théorème A.12 *Pour tout A et B de $\mathbf{A}$, $p(A) \leq p(B)$ si A implique B.*

Théorème A.13 (Théorème des probabilités totales) *Soit A un élément de $\mathbf{A}$ et soit $\mathbf{B}$ un sous-ensemble de $\mathbf{A}$.*
Si les éléments de $\mathbf{B}$ sont deux à deux incompatibles et si $\mathbf{B}$ est exhaustif (c.-à-d. si $\sum_{B \in \mathbf{B}} p(B) = 1$), alors $p(A) = \sum_{B \in \mathbf{B}} p(A \wedge B)$.

Une conséquence directe du théorème des probabilités totales s'énonce de la façon suivante :

Théorème A.14 *Soit A un élément de $\mathbf{A}$ et $\mathbf{B}$ un sous-ensemble de $\mathbf{A}$.*
Si les éléments de $\mathbf{B}$ sont deux à deux incompatibles, si $\mathbf{B}$ est exhaustif et si tout $B \in \mathbf{B}$ est tel que $p(B) \neq 0$, alors $p(A) = \sum_{B \in \mathbf{B}} p(A|B).p(B)$.

Théorème A.15 (Théorème de Bayes) *Soit A et B deux éléments de $\mathbf{A}$.*
Si $p(A) \neq 0$ et $p(B) \neq 0$, alors $p(A|B) = \frac{p(B|A).p(A)}{p(B)}$.

Tous ces théorèmes se démontrent aisément à partir de la définition A.4 et de la proposition A.6. Nous renvoyons le lecteur intéressé par les démonstrations à Howson et Urbach, 2006, p. 16-21 par exemple.

Probabilités et variables aléatoires

Il est courant de parler de probabilités d'événements, de probabilités de propriétés, ou encore d'associer probabilités et variables aléatoires. Une connexion avec les définitions que nous avons données dans la section **4.4.2**, et selon lesquelles les fonctions de probabilités ont pour arguments des propositions, s'établit toutefois aisément. D'abord, on peut associer à tout événement E la proposition $prop(E)$ selon lequel il est advenu et dans tous les contextes, c'est-à-dire pour tout ensemble de propositions clos par conjonction, disjonction et négation auquel $prop(E)$ appartient et pour toute fonction de probabilités p sur cet ensemble, identifier la probabilité de E à la probabilité de $prop(E)$. Ensuite et de la même façon, nous identifions, dans tous les contextes, la probabilité d'une propriété P dans une population donnée à la probabilité de la proposition $prop(P)$ selon laquelle un individu quelconque de cette population instancie P.

Enfin, pour toute variable V, on peut associer à chacune de ses valeurs possibles v la proposition $V = v$ selon laquelle V prend la valeur v et dans tous les contextes identifier à $p(V = v)$ la probabilité que V prenne la valeur v. Mais la connexion entre les définitions que nous avons données et la notion de variable aléatoire est plus profonde. En effet, étant donné un ensemble de variables aléatoires $\mathbf{V}$, on peut considérer la clôture par conjonction, disjonction et négation de l'ensemble des propositions de la forme $V_i = v_i$ et introduire la définition suivante :

Définition A.16 (Fonction de probabilités sur V) Une fonction de probabilités sur $\mathbf{V} = \{V_1, V_2, \ldots V_n\}$ est une fonction de probabilités sur la clôture par conjonction, disjonction et négation de l'ensemble des propositions de la forme $V_i = v_i$ pour $1 \leq i \leq n$.

Si $\mathbf{V} = \{V_1, V_2, \ldots V_n\}$, on appelle « valeurs de $\mathbf{V}$ » les conjonctions de la forme $(V_1 = v_1) \wedge (V_2 = v_2) \wedge \ldots \wedge (V_n = v_n)$. Pour définir une fonction de probabilités p sur $\mathbf{V}$, il suffit d'associer un réel positif à chaque valeur de $\mathbf{V}$ de telle façon que la somme des réels attribués vaut 1. En effet, toutes les autres propositions auxquelles p attribue une valeur sont équivalentes à des propositions de cette forme ou à des disjonctions de propositions de cette forme et, en conséquence, leur probabilité pour p peut être calculée à partir de celles des propositions de la forme $(V_1 = v_1) \wedge (V_2 = v_2) \wedge \ldots \wedge (V_n = v_n)$. En particulier, pour toute valeur $\mathbf{x}$ d'un sous-ensemble $\mathbf{X}$ de $\mathbf{V}$, sa probabilité est égale à la somme de probabilités des valeurs de $\mathbf{V}$ qui l'impliquent.

De même que la notion de fonction de probabilités telle que nous l'avons définie dans la section **4.4.2**, la notion de fonction de probabilités sur un ensemble de variables s'étend des probabilités absolues aux probabilités conditionnelles. En outre, on définit pour les variables les notions d'indépendance probabiliste suivantes :

Définition A.17 (Indépendance probabiliste pour des variables) Soit $\mathbf{V}$ un ensemble de variables, soit p une fonction de probabilités sur $\mathbf{V}$ et soit $\mathbf{X}$ et $\mathbf{Y}$ deux sous-ensembles de $\mathbf{V}$.
$\mathbf{X}$ et $\mathbf{Y}$ sont indépendants pour p si et seulement si :
pour toute valeur $\mathbf{x}$ de $\mathbf{X}$ et toute valeur $\mathbf{y}$ de $\mathbf{Y}$, $p(\mathbf{x} \wedge \mathbf{y}) = p(\mathbf{x}).p(\mathbf{y})$.

Définition A.18 (Indépendance probabiliste relative pour des variables) Soit $\mathbf{V}$ un ensemble de variables, soit p une fonction de probabilités sur $\mathbf{V}$ et soit $\mathbf{X}$, $\mathbf{Y}$ et $\mathbf{Z}$ trois sous-ensembles de $\mathbf{V}$.
$\mathbf{X}$ et $\mathbf{Y}$ sont indépendants relativement à $\mathbf{Z}$ pour p si et seulement si :
pour toutes valeurs $\mathbf{x}$, $\mathbf{y}$ et $\mathbf{z}$ de $\mathbf{X}$, $\mathbf{Y}$ et $\mathbf{Z}$ respectivement,

$$p(\mathbf{x} \wedge \mathbf{y}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z}).p(\mathbf{y}|\mathbf{z}).$$

Bibliographie

ANSCOMBE, E. (1993/1981). « Causality and determination ». *In* SOSA, E. et TOOLEY, M., éditeurs : *Causation*, pages 88–104. Oxford University Press, New York.

ARNTZENIUS, F. (2005). « Reichenbach's common cause principle ». *In* ZALTA, E. N., éditeur : *The Stanford encyclopedia of philosophy*. Stanford University.

ARONSON, J. (1971). « On the grammar of "Cause" ». *Synthese*, 22(3/4): 414–430.

BICKEL, P. J., HAMMEL, E. A. et O'CONNELL, J. W. (1975). « Sex bias in graduate admissions : Data from Berkeley ». *Science*, 187(4175):398–404.

BLALOCK, H. (1991). « Are there really constructive alternatives to causal modeling? ». *Sociological Methodology*, 21:325–335.

CALDWELL, J. C. (1979). « Education as a factor in mortality decline. An Examination of Nigerian data ». *Population Studies*, 33(3):395–413.

CARROLL, J. W. (1991). « Property-level causation ». *Philosophical Studies*, 63:245–270.

CARTWRIGHT, N. (1979). « Causal laws and effective strategies ». *Noûs*, 13(4):419–437.

CARTWRIGHT, N. (1989). *Nature's capacities and their measurement*. Oxford University Press, New York.

CARTWRIGHT, N. (1999). « Causal Diversity and the Causal Markov Condition ». *Synthese*, 121(1):3–27.

CARTWRIGHT, N. (2001). « What is wrong with Bayes nets? ». *The Monist*, 84(2):242–264.

CARTWRIGHT, N. (2004). « Causation : One word, many things ». *Philosophy of Science*, 71:805–819.

CHICKERING, D. (1996). « Learning Bayesian networks is np complete ». *In* FISHER, D. et LENZ, H.-J., éditeurs : *Learning from data*, pages 121–130. Springer, Berlin.

CLOGG, C. et HARITOU, A. (1997). « The regression method for causal inference and a dilemma confronting this method ». *In* MCKIM, V. et TURNER, S., éditeurs : *Causality in crisis ? Statistical methods and the search for causal knowledge in the social sciences*, pages 83–112, Notre Dame. University of Notre Dame Press.

COURGEAU, D. (2003). « From the macro-micro opposition to multilevel analysis in demography ». *In* COURGEAU, D., éditeur : *Methodology and epistemology of multilevel analysis. Approaches from different social sciences*, pages 43–92. Kluwer academic publishers, Boston, Dordrecht, Londres.

DAVIS, W. (1988). « Probabilistic theories of causation ». *In* FETZER, J., éditeur : *Probability and causality : Essays in honor of Wesley C. Salmon.* Reidel, Dordrecht.

DOWE, P. (1992a). « Process causality and asymmetry ». *Erkenntnis*, 37(2):179–196.

DOWE, P. (1992b). « Wesley Salmon's process theory of causality and the conserved quantity theory ». *Philosophy of Science*, 59(2):195–216.

DOWE, P. (2000). *Physical causation.* Cambridge University Press, Cambridge.

DROUET, I. (2009). « Is determinism more favorable than indeterminism for the Causal Markov Condition ? ». *Philosophy of Science (Proceedings of the 2008 Biennial Meeting of the Philosophy of Science Association. Part I : Contributed Papers)*, 76(5):662–675.

DUHEM, P. (1914/1906). *La théorie physique, son objet, sa structure.* Librairie Vrin, Paris, deuxième édition.

DUPRÉ, J. (1984). « Probabilistic causality emancipated ». *Midwest Studies in Philosophy*, 9(1):169–175.

EELLS, E. T. (1991). *Probabilistic causality.* Cambridge University Press, Cambridge.

EELLS, E. T. et SOBER, E. (1983). « Probabilistic causality and the question of transitivity ». *Philosophy of Science*, 50(1):35–57.

FAIR, D. (1979). « Causation and the flow of energy ». *Erkenntnis*, 14(3):219–250.

FREEDMAN, D. et HUMPHREYS, P. (1999). « Are there algorithms that discover causal structure ? ». *Synthese*, 121(1/2):29 54.

FREEDMAN, D. A. (1987). « As others see us : A case study in path analysis ». *Journal of Educational Statistics*, 12(2):101–128.

FREEDMAN, D. A. (1991). « Statistical models and leather shoe ». *Sociological Methodology*, 21:291–313.

GILLIES, D. (2002). « Causality, propensity, and Bayesian networks ». *Synthese*, 132(1/2):63–88.

GLENNAN, S. (1996). « Mechanisms and the nature of causation ». *Erkenntnis*, 44:49–71.

GLYMOUR, C. (1980). *Theory and evidence*. Princeton University Press, Princeton.

GLYMOUR, C., SCHEINES, R. et SPIRTES, P. (1988). « Exploring causal structure with the Tetrad program ». *Sociological Methodology*, 18:411–448.

GOOD, I. J. (1961). « A causal calculus (I) ». *The British Journal for the Philosophy of Science*, XI(44):305–318.

GRANGER, C. W. J. (1969). « Investigating causal relations by econometric models and cross-spectral methods ». *Econometrica*, 37:424–438.

GRANGER, C. W. J. (1980). « Testing for causality : A personal viewpoint ». *Journal of Economic Dynamics and Control*, 2:329–352.

HAAVELMO, T. (1943). « The statistical implications of a system of simultaneous equations ». *Econometrica*, 11:1–12.

HACKING, I. (1965). *Logic of statistical inference*. Cambridge University Press, Cambridge.

HALL, N. (2004). « Two concepts of causation ». *In* COLLINS, J., HALL, N. et PAUL, L., éditeurs : *Counterfactuals and causation*, pages 225–276. MIT Press, Cambridge.

HARMAN, G. (1965). « The Inference to the Best Explanation ». *The Philosophical Review*, 74:88–95.

HESSLOW, G. (1976). « Two notes on the probabilistic approach to causality ». *Philosophy of Science*, 43(2):290–292.

HITCHCOCK, C. (1995). « The Mishap at Reichenbach's Fall : Singular vs. General Causation ». *Philosophical Studies*, 78(3):257–291.

HITCHCOCK, C. R. (1993). « A generalized probabilistic theory of causal relevance ». *Synthese*, 97(3):335–364.

HITCHCOCK, C. R. (2001). « A tale of two effects ». *Philosophical Review*, 110(3):361–396.

HITCHCOCK, C. R. (2002). « Probabilistic causation ». *In* ZALTA, E. N., éditeur : *The Stanford encyclopedia of philosophy*. Stanford University.

HOPE, K. (1980). « A geometrical approach to sociological analysis ». *Quality and Quantity*, 14:309–325.

HOWSON, C. et URBACH, P. (2006). *Scientific reasoning. The Bayesian approach*. Open Court, Chicago et LaSalle, troisième édition.

HUME, D. (1995/1739). *Traité de la nature humaine. Livre 1 : L'entendement*. GF–Flammarion, Paris. Traduction Ph. Baranger et Ph. Saltel.

HUMPHREYS, P. et FREEDMAN, D. (1996). « The grand Leap. Review of Peter Spirtes, Clark Glymour, and Richard Scheines [1993] : Causation, prediction, and search ». *The British Journal for the Philosophy of Science*, 47(1):113–123.

JENSEN, E. R. et AHLBURG, D. A. (2004). « Why does migration decrease fertility ? Evidence from the Philippines ». *Population Studies*, 58(2):219–231.

JENSEN, F. V. (1996). *An Introduction to Bayesian networks*. UCL Press, London.

KENNY, D. A. (1979). *Correlation and causality*. Wiley, New-York. Édition révisée téléchargeable en ligne.

KISTLER, M. (1999). *Causalité et lois de la nature*. Vrin, Paris.

KLINE, R. B. (2005/1998). *Principles and practice of structural equation modeling*. Guilford Press, New York et Londres, deuxième édition.

KORB, K. et WALLACE, C. (1997). « In Search of the Philosopher's Stone : Remarks on Humphreys and Freedman's Critique of Causal Discovery ». *The British Journal for the Philosophy of Science*, 48(3):543–553.

KWOH, C. K. et GILLIES, G. F. (1996). « Using hidden nodes in Bayesian networks ». *Artificial Intelligence*, 88:1–38.

LAURITZEN, S. et SPIEGELHALTER, D. (1988). « Local computations with probabilities in graphical structures and their applications to expert systems (with discussion) ». *Journal of the Royal Statistical Society, Series B*, 50(2):157–224.

LEWIS, D. (1973). « Causation ». *Journal of Philosophy*, 70:172–213.

LEWIS, D. (1986). *Philosophical papers, volume II*, chapitre Postscripts to « Causation », pages 172–213. Oxford University Press, New York.

LEWIS, D. (2000). « Causation as Influence ». *Journal of Philosophy*, 97:182–197.

MACHAMER, P., DARDEN, L. et CRAVER, C. (2000). « Thinking about Mechanisms ». *Philosophy of Science*, 67(1):1–25.

MACKIE, J. L. (1980/1974). *The Cement of the universe : A study of causation*. Oxford University Press, Oxford.

MEURET, D. et MORLAIX, S. (2006). « L'influence de l'origine sociale sur les performances scolaires : par où passe-t-elle ? ». *Revue Française de Sociologie*, 47(1):49–79.

MILL, J. S. (1900/1853). *A System of logic, ratiocinative and inductive, Being a connected view of the principles of evidence and the methods of scientific investigation*. Harper and Brothers, New-York, huitième édition.

MONGIN, P. (2007). « La réfutation et la réfutabilité en science. Applications à l'économie ».

NAÏM, P., WUILLEMIN, P.-H., LERAY, P., POURRET, O. et BECKER, A. (2004). *Réseaux bayésiens.* Eyrolles, Paris.

PEARL, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo (Californie).

PEARL, J. (2000). *Causality. Models, reasoning and inference.* Cambridge University Press, Cambridge.

POPPER, K. R. (1990/1934). *La logique de la découverte scientifique.* Payot, Paris. Traduction A. Thyssen-Rutten et P. Devaux.

PRICE, H. (1991). « Agency and Probabilistic Theory ». *The British Journal for the Philosophy of Science*, 42:156–176.

PSILLOS, S. (2009). « Causal pluralism ». *In* VANDERBEEKEN, R. et D'HOOGHE, B., éditeurs : *Worldviews, science and us : Studies of analytical metaphysics : A Selection of topics from a methodological perspective.* World Scientific Publishers.

REICHENBACH, H. (1956). *The Direction of time.* University of California Press, Berkeley, Los Angeles.

REISS, J. (2009). « Causation in the Social Sciences ». *Philosophy of the Social Sciences*, 39(1):20–40.

RUBIN, D. (1974). « Estimating causal effects of treatments in randomized and non randomized studies ». *Journal of Educational Psychology*, 66:688 701.

RUSSELL, B. (1992/1948). *Human knowledge, its scopes and limits.* Routledge, Londres.

RUSSO, F. (2008). *Causality and causal modelling in the social sciences. Measuring variations.* Springer, Dordrecht.

SALMON, W. C. (1980). « Probabilistic causality ». *In* SOSA, E. et TOOLEY, M., éditeurs : *Causation*, pages 137–153. Oxford University Press, Oxford.

SALMON, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton University Press, Princeton.

SALMON, W. C. (1994). « Causality without Counterfactuals ». *Philosophy of Science*, 61:297–312.

SALMON, W. C. (1998). *Causality and explanation.* Oxford University Press, New York.

SCHAFFER, J. (2001). « Causes as Probability-Raisers of Processes ». *Journal of Philosophy*, 67:285–300.

SHORTLIFFE, E. et BUCHANAN, B. (1984). « A model of inexact reasoning in medicine ». *In* BUCHANAN, B. et SHORTLIFFE, E., éditeurs : *Rule-based expert systems : The MYCIN experiments of the Stanford heuristic programming project*, pages 233–262. Addison-Wesley, Reading.

SKYRMS, B. (1980). *Causal necessity : A pragmatic investigation of the necessity of laws.* Yale University Press, New-Haven et Londres.

SKYRMS, B. (1984). « EPR : Lessons for metaphysics ». *In* FRENCH, P. et UEHLING, T., éditeurs : *Causation and causal theories*, volume IX de *Midwest studies in philosophy*, pages 245–255. University of Minnesota Press.

SKYRMS, B. (2004). *The Stag hunt and the evolution of social structure.* Cambridge University Press, Cambridge.

SOBER, E. (1984). *The Nature of selection. Evolutionary theory in philosophical focus.* MIT Press, Cambridge (MA).

SOBER, E. (1985). « Two Concepts of Cause ». *Philosophy of Science (Proceedings of the biennial meeting of the Philosophy of Science Association 1984)*, 2:405–424.

SOBER, E. (1988). « Probabilistic theories of causation ». *In* FETZER, J., éditeur : *Probability and causality : Essays in honor of Wesley C. Salmon*, pages 211–288. Reidel, Dordrecht.

SPIRTES, P., GLYMOUR, C. et SCHEINES, R. (1990). « Simulation studies of the reliability of computer aided specification using TETRAD II, EQS, and the LISREL programs ». *Sociological Methods and Research*, 19(1):3–66.

SPIRTES, P., GLYMOUR, C. et SCHEINES, R. (1991). « An algorithm for fast recovery of sparse causal graphs ». *Social Science Computer Review*, 9(1):62 72.

SPIRTES, P., GLYMOUR, C. et SCHEINES, R. (1997). « Reply to Humphreys and Freedman's review of *Causation, prediction, and search* ». *The British Journal for the Philosophy of Science*, 48(4):555–568.

SPIRTES, P., GLYMOUR, C. et SCHEINES, R. (2001/1993). *Causation, prediction, and search.* MIT Press, Cambridge (MA), deuxième édition.

SPOHN, W. (2001). « Bayesian nets are all there is to causal dependence ». *In* GALAVOTTI, M. C., SUPPES, P. et COSTANTINI, D., éditeurs : *Stochastic causality*, pages 157–172. CSLI Publications, Stanford.

STEEL, D. (2005). « Indeterminism and the Causal Markov Condition ». *The British Journal for the Philosophy of Science*, 56(1):3–26.

STEEL, D. (2006). « Homogeneity, selection and the faithfulness condition ». *Minds and Machines*, 16:303–317.

SUPPES, P. (1970). *A Probabilistic theory of causality.* North Holland Publishing Company, Amsterdam.

VERMA, T. et PEARL, J. (1988). « Causal networks : Semantics and expressiveness ». *In Proceedings of the 4th annual conference on uncertainty in artificial intelligence (UAI-88)*, pages 69–78, New York. Elsevier science.

VICKERS, J. (2006). « The problem of induction ». *In* ZALTA, E. N., éditeur : *The Stanford encyclopedia of philosophy*. Stanford University.

WILLIAMS, T. O., EAVES, R. C. et COX, C. (2002). « Confirmatory factor analysis of an instrument designed to measure affective and cognitive arousal ». *Educational and Psychological Measurement*, 62(2):264–283.

WILLIAMSON, J. (2001). « Foundations for Bayesian networks ». *In* CORFIELD, D. et WILLIAMSON, J., éditeurs : *Foundations of bayesianism*, Applied logic series, pages 75–116. Kluwer.

WILLIAMSON, J. (2002). Learning causal relationships. Rapport technique 02/02, London School of Economics, Center for Philosophy of Natural and Social Science.

WILLIAMSON, J. (2005). *Bayesian nets and causality. Philosophical and computational foundations*. Oxford University Press, New York.

WOODWARD, J. (2003). *Making things happen : A Theory of causal explanation*. Oxford University Press, New York.

WRIGHT, S. G. (1921). « Correlation and causation ». *Journal of Agricultural Research*, 20(7):557–585.

WRIGHT, S. G. (1934). « The Method of Path Coefficients ». *Annals of Mathematical Statistics*, 5(3):161–215.

Index

Ahlburg, D., 32
Anscombe, E., 2, 7
Arntzenius, F., 106
Aronson, J., 1, 104

Becker, A., 75, 76
Bickel, P.J., 20
Blalock, H., 131
Buchanan, B., 71

Caldwell, J., 37, 38
Carroll, J.W., 3
Cartwright, N., 1, 3, 10, 14, 17, 21, 25-27, 29, 32, 91, 110-112, 124, 125, 139
Carver, C., 1
Chickering, D., 76
Clogg, C., 131
Courgeau, D., 38
Cox, C., 38

Darden, L., 1
Davis, W., 14, 109, 111
Dowe, P., 1, 104
Drouet, I., 110
Duhem, P., 58, 60, 61
Dupré, J., 26, 27

Eaves, R.C., 38
Eells, E., 1, 3, 29, 109

Fair, D., 1, 104
Freedman, D., 41, 42, 44, 60, 79, 110, 119, 131

Gillies, Donald, 77, 122, 123, 125
Gillies, Duncan, 125
Glennan, S., 1
Glymour, C., 47, 78, 79, 83, 96, 97, 125, 132
Good, I.J., 3
Granger, C., 36

Haavelmo, T., 42
Hacking, I., 61
Hall, N., 3
Hammell, E., 20
Haritou, A., 131
Harman, G., 48, 59
Hesslow, G., 21
Hitchcock, C., 3, 9, 29
Hope, K., 42
Hume, D., 4, 7, 104, 109
Humphreys, P., 79, 110, 119

Jensen, E.R., 32

Kenny, D.A., 131
Kistler, M., 1, 104
Kline, R., 37, 38, 44, 49-51, 59, 131
Korb, K., 79
Kwoh, C., 125

Lauritzen, S., 73
Leray, Ph., 75, 76
Lewis, D., 1, 2

Machamer, P., 1
Mackie, J., 8
Meuret, D., 39
Meuret, D., 32, 39, 40, 42, 43
Mill, J.S., 103
Mongin, Ph., 61
Mongin, Ph., 58

Morlaix, S., 32, 39, 40, 42, 43

Naïm, P., 75, 76

O'Connell, J.W., 20

Pearl, J., 40, 50, 51, 73-76, 78, 125
Popper, K., 45, 46, 48, 58, 61, 132
Pourret, O., 75, 76
Price, H., 1, 2
Psillos, S., 3

Reichenbach, H., 12, 14, 17, 104, 105, 107
Reiss, J., 3
Rubin, D., 36
Russell, B., 1
Russo, F., 37

Salmon, W., 1, 14, 15, 91, 104, 111, 112, 124
Schaffer, J., 2
Scheines, R., 78, 79, 83, 96, 97, 125, 132
Shortliffe, E., 71
Simpson, E., 19-21, 28, 93-96
Skyrms, B., 1, 3, 25-32, 91, 114, 139
Sober, E., 3, 18, 19, 26, 29, 112, 113
Spiegelhalter, D., 73
Spirtes, P., 78, 79, 83, 96, 97, 125, 132
Spohn, W., 56
Steel, D., 21, 110
Suppes, P., 1, 11, 13, 15-20, 23, 27, 28, 88, 90-96, 137, 139

Verma, T., 74, 75, 78
Vickers, J., 81

Wallace, C., 79
Williams, T.O., 38
Williamson, J., 54, 66, 73, 75, 76, 103, 107, 109, 110, 120-122, 125, 126, 132, 133
Woodward, J., 1, 2
Wright, S., 42, 44, 45
Wuillemin, P.H., 75, 76

Achevé d'imprimer en janvier 2012 par EMD S.A.S. (France)
N° éditeur : 2012/607 - Dépôt légal : janvier 2012
N° d'imprimeur : 25975

La définition de la causalité est une question centrale en philosophie des sciences qui, si elle suscite l'intérêt des philosophes depuis l'Antiquité, s'est vu profondément renouvelée depuis le milieu du XX^e siècle. Ainsi, la philosophie de la causalité constitue aujourd'hui un domaine très dynamique. Néanmoins, les avancées dans l'analyse du concept de cause sont restées largement indépendantes des méthodes utilisées dans les sciences expérimentales pour identifier les relations causales.

Le principal objectif de cet ouvrage est donc de reconnecter l'analyse philosophique du concept de cause et la méthodologie scientifique. Pour cela, il montre la place majeure occupée aujourd'hui par les approches probabilistes. D'un côté, en effet, les théories probabilistes ont joué un rôle moteur dans le renouveau récent de la philosophie de la causalité. De l'autre, des avancées méthodologiques récentes parmi les plus remarquables concernent l'identification des relations causales à partir de données statistiques.

Titulaire d'un doctorat en philosophie des sciences de l'université Paris I, **Isabelle Drouet** est actuellement maître de conférences à l'université Paris-Sorbonne (Paris IV). Elle est également chercheuse associée à l'équipe Logiques de l'agir de l'université de Franche-Comté.

Collection « Philosophie des sciences » dirigée par Thierry Martin

ISBN 978-2-311-00356-7

9 782311 003567

www.Vuibert.fr