

# Modèles structuraux et inférence causale

Jacques Juhel<sup>1</sup>

## Introduction

Les études effectuées en psychologie différentielle, le plus souvent sur la base de données non expérimentales, cherchent à conférer un sens, une « intelligibilité » (Reuchlin, 1980) aux différences de conduite entre individus, aux facteurs qui organisent ces différences et aux mécanismes qui en sous-tendent l'existence et le développement. Le processus méthodologique mis en œuvre pour atteindre cet objectif vise généralement à construire une représentation formalisée des relations entre les variables considérées au regard du phénomène étudié, l'explication recherchée tendant à se ramener à l'interprétation de cette représentation structurale. Une certaine connaissance du processus de génération des observations que les données statistiques seules ne suffisent pas à préciser est bien sûr nécessaire pour atteindre cette explication et des hypothèses doivent préciser les relations entre variables. Certaines de ces relations, non déterminées *a priori*, mesurent simplement une association. D'autres, orientées par hypothèse, font intervenir plus ou moins explicitement la notion de cause lorsqu'il est par exemple considéré que plusieurs variables sont sous la « dépendance » d'un même facteur, qu'une variable « exerce un effet » sur une autre (relation réursive entre un antécédent et un conséquent) ou que deux variables s'« influencent » réciproquement (relation non réursive). Largement employés en psychologie, les modèles d'équations structurelles linéaires (LSEM : *Linear Structural Equation Model*; Bollen, 1989) ou modèles structuraux<sup>2</sup>, permettent de modéliser avec une très grande souplesse l'organisation de ces variables et de mettre à l'épreuve des hypothèses orientées relatives à leur structure d'ensemble.

---

<sup>1</sup> Université Rennes 2, CRPCC (EA 1285), Place du recteur Henri Le Moal, CS 24307, 35043, Rennes cedex. Mèl : jacques.juhel@univ-rennes2.fr. Je remercie Isabelle Drouet (Université Paris-Sorbonne) et Jean-Luc Kop (Université de Lorraine) pour leurs remarques et demandes d'éclaircissements qui ont permis d'améliorer une précédente version de ce texte.

<sup>2</sup> L'expression « modèles structuraux » désigne dans ce travail le cadre méthodologique général d'analyse quantitative auquel se réfèrent des outils d'analyse particuliers comme les « modèles d'équations structurelles linéaires », les « modèles graphiques », etc.

Historiquement, les modèles structuraux se sont développés à partir de l'idée que si l'on dispose *a priori* d'une représentation formalisée, d'un modèle des relations entre les variables considérées, il est possible de tirer des inférences causales valides de l'observation des covariations entre ces variables. Nous rappellerons d'abord les étapes majeures du développement des méthodes d'analyse causale et de l'intégration interdisciplinaire des LSEM (Matsueda, 2012). Nous verrons ensuite que l'évolution des LSEM s'est accompagnée d'une tendance croissante à considérer qu'une équation structurelle ne traduit que des hypothèses fonctionnelles et statistiques, et donc d'une tendance à délaisser progressivement l'idée de relation causale. Le constat de ce renoncement à une utilisation causale des LSEM nous amènera alors à envisager différentes conceptions philosophiques de la causalité avec lesquelles la méthodologie des modèles structuraux, si elle n'apporte pas d'éléments de définition de la causalité, paraît pourtant conciliable. Une telle compatibilité relance l'intérêt des LSEM comme outil d'analyse causale quantitative. Nous présenterons alors certains développements récents qui en prenant appui sur une conception pluraliste de la causalité et en intégrant les modèles graphiques et les LSEM (Pearl, [2000] 2009), fournissent des outils conceptuels et des critères qui permettent, lorsque c'est possible, d'inférer des relations causales à partir de données probabilistes. En aidant à répondre à la question de savoir si l'association observée entre deux variables est ou n'est pas le résultat d'une relation de causalité, ces outils présentent donc un grand intérêt dans la modélisation des structures et dans l'identification des mécanismes causaux susceptibles d'expliquer les différences entre individus dans l'organisation, la sensibilité aux variations contextuelles et les variations temporelles des conduites.

## **Le développement des méthodes d'analyse causale**

Les méthodes d'analyse causale quantitative se sont développées principalement dans trois secteurs disciplinaires : en biométrie pour tester l'absence ou la présence de relations causales et mesurer la dépendance causale ; en économétrie où les avancées méthodologiques ont été les plus importantes ; en sociologie avec la diffusion de la logique de l'inférence causale.

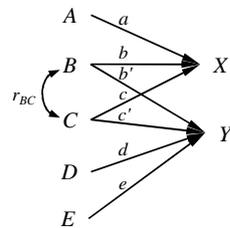
### **L'analyse de parcours en biométrie**

L'origine des SEM remonte au développement dans les années 1920 de l'analyse de parcours (*path analysis*), une méthode d'analyse statistique des corrélations mise au point par le généticien Sewall Wright (1921, 1934). Wright (1934) indique clairement que sa méthode d'analyse « n'a pas pour vocation, tâche impossible, de déduire des relations causales à partir des valeurs des coefficients de corrélation » (p. 93). Il estime par contre qu'il est possible de tirer des conclusions causales d'une combinaison d'hypothèses théoriques de

nature causale et de données d'observation. L'objectif de l'analyse de parcours est d'évaluer quantitativement ce qui est causalement déterminé dans les variations de certaines variables supposées *a priori* sous l'influence d'autres variables du système. Par exemple, Wright (1921) fait l'hypothèse d'une influence causale des facteurs  $A, B, C, D$  et  $E$  sur  $X$  et  $Y$  dans le système clos de variables  $\{X, Y, A, B, C, D, E\}$  (Figure 1). Il suppose que les variables sont mesurées sans erreur, qu'elles sont standardisées et que toutes les sources de variation de  $X$  et de  $Y$  sont connues. Dans ces conditions, l'écriture d'un nombre suffisant d'équations linéaires permet de calculer les coefficients de piste inconnus à partir des corrélations connues entre variables observées.

Représentation des relations causales au sein du système de variables observées

Exemples d'équations dérivées du système



$$\sigma_{X \cdot A} = a$$

$$\sigma_X^2 = a^2 + b^2 + c^2 + 2bc r_{BC} = 1$$

$$r_{XY} = bb' + cc' + b r_{BC} c' + c r_{BC} b'$$

**Figure 1** – Analyse de parcours. Les facteurs  $A, B$  et  $C$  contribuent à produire les variations de  $X$  ; les facteurs  $B, C, D$  et  $E$  contribuent à produire les variations de  $Y$ . Lorsque toutes les variables sont standardisées,  $\sigma_{X \cdot A}$  est l'écart-type de  $X$  lié aux seules variations de  $A$ ,  $\sigma_X^2$  est la variance de  $X$  et  $r_{XY}$  est la corrélation entre  $X$  et  $Y$  (D'après Wright, 1921, p. 565).

Wright (1921) envisage deux manières d'utiliser l'analyse de parcours. La méthode peut d'abord être employée quand les relations causales entre les variables incluses dans le modèle sont précisément connues. Cette situation est celle à laquelle Wright est confronté dans ses recherches sur les effets de l'hérédité et de l'environnement sur la couleur du pelage des cochons d'inde. Connaissant le modèle causal « vrai », la méthode permet de mesurer chaque influence causale toutes choses égales par ailleurs. Si, comme c'est le plus souvent le cas en sciences sociales ou en psychologie, le cadre théorique ne permet pas *a priori* d'explicitier en totalité les relations causales entre les variables incluses dans le modèle, l'analyse de parcours peut aussi être utilisée pour tester les conséquences logiques liées à l'absence ou à la présence de l'une ou l'autre de ces relations. Cette seconde approche implique bien sûr le respect d'un principe essentiel de la méthode scientifique selon lequel les données ne peuvent qu'infirmer ou échouer à infirmer ce que postule le système de causalité. La méthodologie de l'analyse de parcours nécessite donc de pouvoir expliciter et justifier *a priori* le système des relations causales entre les variables considérées. La causalité alléguée n'est pas inférée des données ; elle

est infirmée ou provisoirement corroborée dans le contexte de causalité défini par le chercheur.

Les hypothèses préalables qui fondent l'analyse de parcours et donc la validité de cette analyse comme outil de modélisation causale ont été contestées dès les premières publications de Wright. Le point de vue de Karl Pearson est très influent à l'époque et beaucoup d'auteurs pensent avec lui que la causalité est un concept métaphysique auquel est préférable celui d'association, de relation fonctionnelle qui résume et permet de prédire les observations. Niles (1922, 1923) par exemple estime peu réaliste de construire *a priori* un système causal décrivant les relations entre variables à partir de connaissances relatives à la question traitée. Il considère que cette méthode « est basée sur une interprétation erronée de la nature scientifique de la causalité » (Niles, 1923, p. 260) et qu'elle ne peut être justifiée simplement sur la base de résultats en accord avec la réalité observée ou avec ce que le chercheur s'attend à observer. Fisher (1925) conteste pour sa part que l'on puisse étudier les causes à partir de corrélations. Il développe un ensemble de méthodes (analyse de variance, randomisation, contrôle expérimental) visant à assurer les conditions qui permettent de distinguer les relations causales des relations non causales. L'influence de Pearson d'une part, celle de Fisher d'autre part contribuent ainsi à alimenter la suspicion à l'égard de l'analyse de parcours, ce qui en retardera la diffusion dans la communauté scientifique.

### **Les modèles à équations simultanées en économétrie**

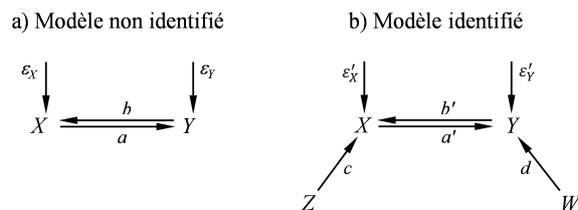
L'analyse de parcours a été appliquée en 1928 par l'économiste Phillip Wright, le père de Sewall, au problème de l'identification de systèmes comportant des variables endogènes c'est-à-dire des variables déterminées par le système. De la collaboration entre Phillip et Sewall Wright naît l'idée d'introduire des variables exogènes supplémentaires pour identifier le modèle et estimer les coefficients de parcours (figure 2), préfigurant ainsi la régression linéaire à variables instrumentales<sup>3</sup> et l'estimation par la méthode des moindres carrés

Au début des années quarante, Haavelmo (1944) et les chercheurs de la Commission Cowles introduisent les probabilités en économétrie et proposent de considérer que les données économiques sont générées par une loi de probabilité conjointe « dont l'existence est peut-être purement hypothétique » (pp. iii-iv). L'outil nécessaire pour penser le phénomène étudié est le modèle défini comme un système conçu *a priori* de mise en correspondance entre des objets théoriques et des objets empiriques. Le choix des variables et des relations entre variables est guidé par la théorie. C'est également la théorie qui

---

<sup>3</sup> Z est une variable instrumentale pour les variables X et Y avec  $X \rightarrow Y$ , si elle est corrélée avec X et est indépendante du terme d'erreur  $\varepsilon_Y$ .

décide du statut des variables considérées : endogène pour les variables dont on cherche à expliquer les variations ; exogène pour les variables qui contribuent à produire les variations des endogènes. Le système formel qui représente le phénomène réel est généralement un système d'équations linéaires dont chacune comporte un terme d'erreur. Techniquement, les variations conjointes des variables aléatoires considérées sont restreintes par le système spécifié. Le système est aussi un modèle de données, un plan d'expérience qui fait correspondre les variables observationnelles aux variables théoriques. Tester la théorie consiste alors à confronter le modèle – statistiquement, les hypothèses sur la loi de probabilité conjointe - aux observations. Les principes généraux de l'inférence statistique sont basés sur la théorie de Neyman-Pearson<sup>4</sup> et servent à estimer la valeur des paramètres du modèle, les coefficients des relations qui caractérisent la population. L'inférence statistique permet ainsi la mise à l'épreuve d'hypothèses causales par confrontation des conséquences logiques de ces hypothèses aux données d'observation.



**Figure 2** - Système à deux équations simultanées correspondant à des relations théoriques définies *a priori* ; a) les deux coefficients structurels  $a$  et  $b$  ne peuvent être estimés en connaissant seulement la corrélation entre  $X$  et  $Y$  (modèle non identifié) ; b) l'introduction des deux variables instrumentales  $W$  et  $Z$  ( $Z$  est une cause directe de  $X$  mais pas de l'erreur de  $Y$  ;  $W$  est une cause directe de  $Y$  mais pas de l'erreur de  $X$ ) résout le problème d'identification car 6 corrélations sont disponibles pour estimer 4 coefficients structurels (modèle sur-identifié).

Plusieurs autres contributions d'Haavelmo à la méthodologie des modèles à équations simultanées méritent d'être signalées. Haavelmo clarifie la distinction entre l'adéquation statistique (la robustesse de la relation entre le modèle théorique et le phénomène étudié via les données observées) et l'adéquation « substantielle » (la pertinence de l'éclairage apporté par le modèle causal au phénomène étudié). Il distingue la définition des effets causaux de leur identification<sup>5</sup> et propose une solution mathématique au problème de

<sup>4</sup> Le test du rapport de vraisemblance qui consiste à comparer les probabilités des observations sous l'hypothèse nulle et l'hypothèse alternative a été proposé par Neyman et Egon Sharpe Pearson en 1928.

<sup>5</sup> Le problème de l'identification concerne la possibilité d'effectuer des inférences sur la forme structurelle du modèle d'équations simultanées à partir de la distribution de probabilité des variables observées (voir Aldrich, 1994, pour une présentation de la théorie de l'identification de Haavelmo).

l'identification. Il établit une méthode indirecte d'estimation des coefficients du modèle qui repose sur l'estimation préalable des coefficients de la forme dite réduite du modèle c'est-à-dire de l'ensemble des équations simultanées obtenues en exprimant chacune des variables endogènes en fonction des seules variables prédéterminées (*i.e.*, les variables exogènes et, dans les séries temporelles, les variables endogènes antécédentes ou « retardées »).

Les modèles à équations simultanées seront le centre d'intérêt de la théorie économétrique jusqu'à la fin des années 1970. Ces modèles marquent une évolution sensible par rapport aux premiers développements de l'analyse de parcours. Le modèle causal (l'ensemble des équations simultanées) est considéré comme une hypothèse statistique au sens de Neyman-Pearson. La notion de facteur causal apparaît plus relative qu'absolue ; les hypothèses sont « les implications conjointes – et les seules implications testables dès lors qu'il s'agit de *données d'observation* – d'une théorie *et* d'un plan d'expérience » (Haavelmo, 1943, p. 7 ; les italiques sont de l'auteur). Les assertions du modèle compatibles avec les données deviennent des faits précis interprétés à la lumière du modèle théorique. L'interprétation conduit alors à faire, par une démarche inductive, des hypothèses sur certaines relations causales ou certains mécanismes. Bien que rien ne puisse être dit *a priori* sur la validité ultérieure (*i.e.*, pour de futures observations) d'un modèle compatible avec les observations recueillies, de telles inductions empiriques sont à l'époque jugées prometteuses par de nombreux chercheurs en économétrie.

### **Les modèles de causalité en sociologie**

Dans les années 50, les réflexions méthodologiques en sociologie quantitative conduisent à exprimer des processus causaux en termes de modèles statistiques<sup>6</sup>. Boudon (1965) rappelle ainsi que Durkheim et Lazarsfeld ont contribué à résoudre la question suivante : quand peut-on inférer que *X* est la cause de *Y* sachant *a priori* que *X* précède temporellement *Y* et qu'une corrélation est observée entre *X* et *Y* ? L'analyse causale a été popularisée par Blalock (1964) avec son ouvrage sur l'inférence causale dans la recherche non expérimentale. L'évolution de la méthode a culminé avec Duncan (1966) qui a fusionné l'analyse de parcours avec l'analyse factorielle et l'estimation de l'erreur de mesure dans la théorie psychométrique du score vrai. Le développement de l'analyse causale quantitative a été ensuite très rapide pour devenir une méthode de raisonnement sociologique intégrant la connaissance théorique et empirique.

---

<sup>6</sup> Les modèles statistiques qui représentent des processus causaux impliquent certaines contraintes structurelles dites de sur-identification. Par exemple, le processus causal  $X \rightarrow M \rightarrow Y$  implique que la corrélation partielle entre *X* et *Y* est nulle si *M* est statistiquement contrôlé.

Les sociologues ont largement contribué à poser les problèmes théoriques dont dépend l'utilisation des modèles de causalité. L'emploi du langage causal impliquant toujours une théorie, celle-ci doit être suffisamment précise pour expliquer comment la cause produit son effet. Le test de cette relation causale ne peut avoir de sens qu'au sein du modèle causal dont le chercheur fait *a priori* l'hypothèse. L'analyse causale repose ainsi sur des suppositions (*assumptions*) ou hypothèses imposées dont la vraisemblance contribue à déterminer la validité du modèle causal.

L'approche, généralement déductive, consiste à vérifier par ses conséquences logiques une structure causale établie *a priori* et à mesurer des relations causales entre variables. Boudon (1965) estime qu'« il est possible, en introduisant des hypothèses simples, de construire des modèles formels permettant de conclure aux relations de causalité à partir des covariations entre les observations » (p. 42). Duncan (1966) précise prudemment que l'analyse causale « se focalise sur le problème de l'interprétation et ne prétend pas être une méthode permettant de découvrir des causes » (p. 1). L'approche peut aussi permettre d'inférer des relations causales à partir des propriétés statistiques des données ou, comme l'écrit Boudon (1965), d'« induire la structure causale sous-jacente (p. 26). Mais le caractère essentiel des postulats et suppositions préliminaires qui fondent l'inférence causale est parfois sous-estimé face aux avantages anticipés de la méthodologie. Blalock (1964) admet ainsi que l'hypothèse de récursivité d'un modèle (termes d'erreurs non corrélés entre eux et variables indépendantes de chaque équation indépendantes des termes d'erreur) est assez peu réaliste. Pour autant, « à un moment donné, il faut arrêter et supposer, pour simplifier, que les influences laissées pour compte ne produisent pas de corrélations trompeuses entre certaines variables du modèle. Sinon, toute inférence causale serait impossible » (p. 176).

## **Perspectives offertes et difficultés rencontrées par les modèles d'équations structurelles**

### **Un cadre d'analyse cohérent**

La publication des travaux de Jöreskog (1970) sur l'analyse des structures de covariance marque le début d'une nouvelle phase de développement des modèles structurels. Celle-ci se caractérise par l'intégration de l'analyse factorielle, de l'analyse de parcours et des modèles à équations simultanées dans le cadre général des LSEM (Bollen, 1989, pour une présentation du modèle standard). Le modèle est formé de deux sous-modèles : a) le modèle de mesure relie les variables latentes aux variables observées par un modèle d'analyse factorielle confirmatoire ; b) le modèle structurel relie les variables latentes entre elles par un modèle d'équations simultanées. L'originalité des LSEM réside dans la possibilité de tester statistiquement sur une population spécifique un ensemble d'hypothèses structurelles (associations et/ou relation

causales) entre des variables latentes. Le cadre formel de modélisation est hypothético-déductif. L'approche comporte une première phase de spécification du modèle qui débouche sur une représentation graphique des relations entre les variables. Les données sont ensuite recueillies. Le modèle spécifié est alors estimé, généralement par maximum de vraisemblance. Des indices d'ajustement permettent d'évaluer la capacité du modèle à reconstruire la matrice de variance-covariance des variables observées. L'ajustement peut être amélioré en fixant et/ou libérant certains paramètres du modèle de mesure et du modèle structurel. Des hypothèses sur les relations entre variables latentes peuvent être ainsi testées. Par exemple, l'absence d'une relation causale entre deux variables latentes est testée en fixant le paramètre correspondant à cet effet à 0. Une discussion en lien avec les hypothèses initialement formulées clôt la démarche. Malgré la relative complexité de la méthode au plan statistique, le développement de logiciels spécifiques de modélisation dont LISREL (*Linear Structural Relations*) est le précurseur, en facilite la mise en œuvre.

Dans les années 1980, les LSEM apparaissent comme un outil puissant de modélisation causale (Bentler, 1980, 1998). Ils permettent en effet d'utiliser des construits mieux définis et mieux mesurés. Les hypothèses structurelles peuvent être spécifiées au sein de modèles complexes, mieux à même de correspondre au phénomène étudié. Des effets indirects entre construits peuvent être ainsi plus facilement testés. Les facteurs de confusion liés à l'erreur de mesure des variables latentes et à la variabilité des items, indépendamment de celle des variables latentes, sont aussi mieux contrôlés, ce qui conduit à des estimations moins biaisées des relations entre les variables latentes incluses dans le modèle. Confrontés à des données non expérimentales, bon nombre de chercheurs en économie, en sociologie ou en psychologie appliqueront donc cette méthodologie propre à suggérer et mettre à l'épreuve des hypothèses causales (pour une revue en psychologie, MacCallum & Austin, 2000).

Les LSEM sont aujourd'hui une technique statistique multivariée extrêmement flexible dont la sophistication ne cesse de croître. Les formalisations contemporaines dites de seconde génération (Muthén, 2001) intègrent désormais les LSEM pour données dichotomiques et catégorielles ordonnées, les LSEM avec des variables dont les distributions appartiennent à la famille exponentielle (modèles linéaires généralisés), les LSEM multi-niveaux, les LSEM à facteurs de croissance latente, les LSEM à mélange de distributions, les LSEM pour données manquantes complètement au hasard, aléatoires ou manquantes et non ignorables (liste non exhaustive...). Récemment appliquée aux LSEM, l'approche bayésienne offre aussi de nouvelles possibilités de spécification et peut contribuer à améliorer le processus de sélection des modèles (Lee, S.-Y., 2007). On trouvera dans le « *Handbook of Structural Equation Modeling* » récemment édité par Hoyle (2012) un panorama complet et actuel des possibilités de modélisation offertes par cette méthodologie.

## Difficultés et critiques

La méthodologie des LSEM a été très tôt l'objet de vives critiques de la part d'un certain nombre de statisticiens (par ex., Freedman, 1987 ; Holland, 1988). La controverse autour des SEM est en grande partie liée à la question de l'inférence causale (Bullock, Harlow & Mulaik, 1994) et à ce que laissent croire les termes de « causalité » ou de « confirmatoire » (Biddle & Martin, 1987) lorsque la méthodologie des LSEM est appliquée à des données d'observation non expérimentales. Ling (1982) considère ainsi que « le livre de Kenny (1979, *Correlation and Causality*) illustre parfaitement les lacunes statistiques et logiques qui sous-tendent ces méthodes de pseudo-magie noire » (p. 490). Baumrind (1983) dénonce « le fantasme de la justification par n'importe quel système statistique des inférences causales tirées de données corrélationnelles » (p. 1289). Cliff (1983) manifeste son inquiétude en écrivant que « tout semble se passer comme si les LSEM conduisaient à abandonner tout esprit critique normal » (p. 116). Freedman (1987) affirme que « personne ne prête beaucoup d'attention aux hypothèses et que la technologie tend à submerger le sens commun » (p. 3). Des appréciations remettant ainsi en cause la crédibilité des LSEM figurent dans beaucoup de commentaires.

Répétons d'abord ce qui est bien connu. Les données statistiques d'observation à *elles-seules* ne permettent aucune inférence causale. Utilisés dans un contexte explicite de causalité, ce qui nécessite des hypothèses supplémentaires et une approche spécifique de modélisation, les LSEM ne permettent que d'infirmar la structure causale préalablement postulée. Il s'agit donc d'une méthodologie « dis-confirmatoire » même si la compatibilité des hypothèses structurelles avec l'organisation des données peut être prise comme un argument qui renforce la plausibilité du modèle. La validité de l'interprétation du modèle doit alors être interrogée en examinant des modèles équivalents c'est-à-dire des modèles qui incluent les mêmes variables, possèdent le même nombre de paramètres, présentent la même qualité d'ajustement mais impliquent des interprétations très différentes (voir par ex., les 3 modèles équivalents de la figure 3). La critique de Cliff (1983) selon laquelle « l'incapacité à reconnaître des modèles équivalents peut conduire à des conclusions erronées » (p. 267) est donc tout à fait justifiée (mais voir sur ce point la réponse technique de Hershberger, 2006). Elle peut d'ailleurs être étendue à des modèles non équivalents et non emboîtés, éventuellement mieux ajustés mais rarement testés.

Comme d'autres techniques de modélisation statistique, les LSEM nécessitent le respect de conditions d'application qui ont trait aux données et à l'estimation des paramètres du modèle. Ces hypothèses statistiques concernent la distribution des données (par ex., la normalité multivariée pour l'estimateur du maximum de vraisemblance), la linéarité des relations entre variables, le type de données manquantes, les erreurs de spécification (par ex., une forme

fonctionnelle erronée, des erreurs de mesure, une mauvaise spécification des erreurs) ou l'exogénéité (l'effet d'une variable sur une autre ne peut être interprété car il inclut des causes non mesurées) (Kaplan, 2009). Malgré leur importance, ces conditions d'application ne sont pas toujours testées et doivent l'être avec des méthodes statistiques adaptées (voir par ex., Antonakis, Bendahan, Jacquart & Lalive, 2010).

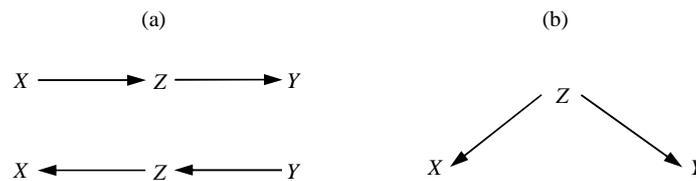
Lorsqu'ils sont employés comme outils d'analyse causale, les LSEM reposent sur diverses suppositions qui garantissent l'interprétation causale des coefficients en termes de mesure des variations d'une variable dépendante causées par les variations d'une variable indépendante. Ces suppositions que nous précisons plus loin sont des conditions idéales pour établir la causalité mais certaines d'entre elles peuvent être jugées peu vraisemblables. Freedman (1987) par exemple conteste qu'il soit possible d'identifier toutes les variables essentielles dans un système considéré<sup>7</sup>. La plausibilité de certaines suppositions peut néanmoins être théoriquement et logiquement défendue à condition de disposer d'une connaissance suffisamment précise du domaine pour pouvoir justifier par exemple l'orientation d'une relation causale ou, lorsque la condition de l'antériorité temporelle est respectée, étayer la distinction entre l'antécédent qui conditionne le phénomène et une variable intermédiaire.

Un exemple présenté par Simon (1954) illustre l'importance des suppositions non testables qui sont énoncées *a priori* dans le choix d'une structure causale. Ces suppositions reposent sur des considérations relatives au contenu (informations disponibles, appréciation de la portée d'un résultat, etc.) ou des considérations plus formelles (choix des variables observables, niveau de mesure, hypothèses statistiques, etc.) (Bacher, 1979). Simon cherche les moyens de distinguer une corrélation « authentique » d'une corrélation « trompeuse ». La première est la conséquence logique d'une relation de causalité entre deux variables  $X$  et  $Y$ . La seconde est imputable à d'autres facteurs qu'une relation causale entre  $X$  et  $Y$ . Simon propose d'introduire dans le système une troisième variable  $Z$  susceptible de rendre compte de la corrélation observée entre  $X$  et  $Y$  et de calculer la corrélation partielle entre  $X$  et  $Y$  lorsque  $Z$  a une valeur fixée (figure 3). Dans le cas par exemple où cette corrélation partielle est nulle, trois structures causales sont envisageables ; la corrélation entre  $X$  et  $Y$  est trompeuse dans un cas, authentique dans les deux autres.

---

<sup>7</sup> L'hypothèse de clôture présume que chaque variable non incluse dans le système affecte au plus une variable du système, ce qui implique des corrélations nulles entre les termes d'erreur associés à chaque variable du système. L'hypothèse de clôture ne présume par contre pas que toutes les variables susceptibles d'avoir un effet sur le phénomène que le système formel représente sont incluses dans le système.

La décision de dire que la corrélation entre  $X$  et  $Y$  est ou n'est pas trompeuse ne peut donc en général être prise que s'il est fait *a priori* l'hypothèse d'un « ordre causal » particulier des variables considérées (par ex.,  $Y$  n'est pas la cause de  $X$ ). De telles suppositions émergent de la connaissance du contexte causal de l'étude. Elles ne sont pas causales au sens où elles ne sont pas testables empiriquement comme l'est par exemple l'hypothèse d'un effet indirect de  $X$  sur  $Y$ . Ces suppositions sont également sans rapport avec d'autres suppositions d'ordre statistique.



**Figure 3** - Trois exemples de structure causale où la corrélation partielle  $r_{xy.z}$  entre  $X$  et  $Y$  à  $Z$  constant est nulle : a)  $r_{xy.z}$  est « authentique » : l'effet causal de  $X$  sur  $Y$  (ou de  $Y$  sur  $X$ ) opère via la variable intermédiaire  $Z$  ; b)  $r_{xy.z}$  est « trompeuse » : la corrélation entre  $X$  et  $Y$  résulte de l'effet causal conjoint de  $Z$  sur  $X$  et  $Y$ .

L'inaccessibilité de certaines des suppositions qui permettent l'inférence causale a une conséquence importante. Ce qui est *a priori* supposé et qui conditionne la possibilité d'une inférence relative à une hypothèse de causalité ne pouvant être empiriquement justifié, certains auteurs vont exprimer des douter sur la valeur scientifique de l'approche (par ex., Dawid, 2000). Les chercheurs s'efforcent alors d'accroître la crédibilité de la méthodologie en développant la technique statistique (Kline, 2012) et en passant corrélativement sous silence l'objectif initial de modélisation causale. Muthén par exemple considère que « ces analyses statistiques [les LSEM] ont très peu à voir avec la causalité, [même si] l'utilisateur peut trouver qu'elles confèrent une certaine plausibilité à ses spéculations causales » (Muthén, 1987, p. 180).

Les LSEM peuvent bien sûr être interprétés du seul point de vue des associations entre variables. Il s'agit alors de découvrir quelles sont les relations qui « tiennent » entre variables, d'interpréter ces relations et d'en extraire les implications théoriques. Lorsque les hypothèses statistiques standard (linéarité, normalité des termes d'erreurs, erreurs non corrélées, etc.) sont vérifiées, le modèle des associations entre variables est identifié et ses paramètres peuvent être estimés. C'est cette démarche abductive<sup>8</sup> qu'adoptera celui qui « s'intéresse à la structuration d'une pluralité de variables entre lesquelles aucun lien de causalité n'est posé *a priori* » (Reuchlin, 1999, p. 142). La structure de

<sup>8</sup> En l'absence d'un modèle théorique solide, l'abduction peut être définie comme l'inférence de la meilleure explication, en attente de tests ultérieurs.

variables identifiées est alors proposée comme modèle d'explication du phénomène étudié. Ce modèle est interprété et sa plausibilité évaluée dans un contexte scientifique plus large (Bacher, 1999 ; Reuchlin, 1995, 1998).

L'évaluation de la valeur heuristique du modèle qui prolonge la démarche en testant de nouvelles hypothèses structurelles compatibles avec l'interprétation du modèle s'inscrit dans un cadre hypothético-déductif et repose sur des modèles dont les caractéristiques sont augmentées par rapport à celles des modèles d'association (Russo, 2009). Ces modèles nécessitent en effet une connaissance du contexte causal (théories scientifiques, résultats d'études antérieures,...). Ils se réfèrent à des hypothèses qualitatives qui précisent les relations de cause à effet entre variables et tiennent compte du caractère multivarié de la causalité. Ils s'appuient sur des suppositions « extra-statistiques » qui concernent la clôture du système et les caractéristiques des relations causales (direction temporelle, relations causales asymétriques, mécanisme causal,...). Ils se fondent enfin sur des suppositions causales qui « imposent des contraintes aux variations » (*ibid*, p. 124). On suppose par exemple que la fonction causale est additivement séparable en deux composantes indépendantes : une composante déterministe (la cause) et une composante stochastique (les erreurs ou perturbations). D'autres suppositions importantes sur lesquelles nous reviendrons plus loin portent sur l'absence de variables confondantes, sur la suffisance causale (*causal sufficiency*) du système<sup>9</sup>, sur la fidélité (*faithfulness*) des associations entre variables à la structure causale postulée<sup>10</sup> ou sur l'invariance de la relation causale<sup>11</sup>. Au total, les relations de cause à effet qu'une telle procédure permet de mettre à jour sont relatives à un ensemble de variables préalablement défini. Le modèle d'explication est donc proposé dans un contexte donné de causalité, conditionnellement à la satisfaction d'hypothèses statistiques et de suppositions causales dont certaines peuvent être imposées.

---

<sup>9</sup> La condition de suffisance causale postule que toutes les causes communes de deux variables du système sont incluses dans le système.

<sup>10</sup> On peut imaginer et il peut arriver que deux effets causaux d'une variable sur une autre soient égaux en valeur absolue mais se neutralisent parce que l'un est positif et l'autre négatif. Par exemple, le stress lié au travail scolaire a un effet négatif sur le bien-être au collège. Mais l'augmentation du stress tend à accroître la probabilité d'emploi d'une stratégie de faire face centrée sur le problème et l'emploi de cette stratégie tend à améliorer le bien-être. Si l'effet direct et l'effet indirect du stress sur le bien-être se compensent exactement, aucun effet du stress sur le bien-être n'est observé. On dit alors qu'il y a infidélité (*unfaithfulness*) entre la distribution de probabilité sur la population et la structure causale postulée.

<sup>11</sup> La supposition d'invariance autorise la prédiction des effets en fonction des différentes valeurs des causes.

Comment expliquer alors le renoncement progressif à une lecture causale des LSEM quand ceux-ci peuvent être considérés, au moins dans certains contextes de modélisation, comme des outils d'analyse causale ? Pearl ([2000] 2009, 2012) y voit plusieurs raisons qui tiennent selon lui aux difficultés logiques de l'approche confirmatoire, à la confusion entre les suppositions causales et les conséquences pouvant en être déduites, à la difficulté à distinguer les suppositions causales des hypothèses statistiques en raison d'une « hégémonie de la pensée statistique en termes exclusifs de distributions de probabilité et d'associations partielles » (Bollen & Pearl, 2013, p. 308). L'abandon d'une interprétation causale des LSEM tient aussi au manque de consensus sur une définition de la causalité applicable aux LSEM ainsi que sur le cadre méthodologique et les critères nécessaires à l'identification d'une relation de cause à effet.

## **Les modèles d'équations structurelles sont compatibles avec différentes conceptions de la causalité**

### **La diversité des conceptions de la causalité**

La causalité a été et est toujours l'objet d'un profond débat en philosophie des sciences (pour revue en français : Hespel, 1994 ; Kistler, 2004, 2011). Alors qu'il y a un siècle beaucoup de philosophes semblaient être convaincus de la nécessité d'éliminer la notion de cause de leur langage (Russell, 1913), le concept de causalité est considéré aujourd'hui comme indispensable dans de nombreux domaines de la recherche scientifique. Le débat philosophique sur la causalité a été largement influencé par les réflexions de Hume pour qui la causalité était essentiellement l'occurrence régulière d'un effet qui suit sa cause. Dans le célèbre exemple de la boule de billard, Hume identifie les trois circonstances requises dans la relation entre cause et effet : la contiguïté dans le temps et dans l'espace, l'antériorité dans le temps et la conjonction constante<sup>12</sup>. Cette théorie régulariste de la causalité et la réduction de la causalité à l'explication déductive-nomologique<sup>13</sup> qui en est issue ont d'abord prévalu. L'objection que les causes ne sont pas invariablement suivies de leurs effets, celle que la régularité ne suffit pas à la causalité ou celle de la causalité redondante ont ensuite contribué à remettre en cause la conception régulariste de la causalité (Barberousse, Kistler & Ludwig, 2000 ; Kistler, 2011). Plusieurs solutions « néo-régularistes » ont été proposées pour répondre à ces objections. Par exemple, le philosophe Mackie (1980) conserve les termes logiques de

---

<sup>12</sup> Croyance en une relation entre deux phénomènes dès lors qu'ils surviennent toujours en même temps.

<sup>13</sup> Nomologique au sens où l'existence d'une relation causale particulière implique l'existence d'une loi causale.

nécessité (la cause doit produire l'effet) et de suffisance (l'effet peut se produire en l'absence de la cause mais se produit toujours en sa présence) pour dire qu'une cause est suivie de son effet quand les autres composants de la condition suffisante dont elle fait partie sont réunis<sup>14</sup>. L'épidémiologiste Rothman (1976) défend une conception comparable dans laquelle une cause dite « composée-suffisante » est constituée d'un ensemble de composants dont aucun n'est par lui-même suffisant mais est suffisante quand l'ensemble des composants est présent.

Les conceptions de la causalité ayant cherché à répondre de manière nouvelle aux limites des approches régularistes plutôt qu'à aménager celles-ci peuvent être réparties en deux grandes familles (Drouet, 2012 ; Woodward, 2007) : les conceptions « mécanistes » pour lesquelles la causalité entre  $X$  et  $Y$  est démontrée par l'identification du processus de production qui relie  $X$  à  $Y$  et les conceptions « explicatives » qui renvoient à ce qui introduit une différence dans l'effet.

### **Utilisation explicative de la cause : quel mécanisme produit l'effet ?**

Si l'intérêt explicatif des mécanismes ne fait aucun doute, il n'est pas facile de définir précisément ce qu'est un mécanisme causal ni ce qui le constitue (voir par ex., Knight & Winship, 2013, p. 279). Deux conceptions majeures se présentent dans la diversité des définitions proposées. Les mécanismes causaux ont d'abord été compris comme des processus causaux en interaction. Ils sont plutôt conçus aujourd'hui comme des systèmes organisés tout en reconnaissant que dans beaucoup de chaînes de causalité, les processus ne sont pas organisés en un système stable (par ex., Glennan, 2009). Un mécanisme causal peut donc être défini comme « i) une configuration particulière de conditions et de processus, ii) qui conduisent généralement d'un ensemble de conditions à un résultat, iii) sur la base des propriétés des processus sous-jacents » (Little, 2011, p. 277). L'idée fondamentale dans cette conception de la causalité est qu'un mécanisme causal a le pouvoir de produire du changement en transmettant une « marque » (une modification locale d'une caractéristique) par l'intermédiaire d'une (ou de plusieurs) connexion(s) causale(s). Une autre idée importante est celle de structure : le mécanisme causal produit du changement en raison des éléments qui le constituent et de leur organisation fonctionnelle. L'hypothèse de modularité<sup>15</sup> complète l'idée précédente en supposant qu'il est possible de fixer

---

<sup>14</sup> Condition INUS : la cause est un élément non-suffisant (*Insufficient*) mais non-redondant (*Necessary*) d'une condition non nécessaire (*Unnecessary*) mais suffisante (*Sufficient*) pour produire l'effet.

<sup>15</sup> Si l'hypothèse de modularité est « pratique d'un point de vue épistémique » (Cartwright, 2001, p. 68), son bien-fondé dans l'interprétation des équations structurelles est parfois discuté (par ex., Fennell, 2007, en économétrie).

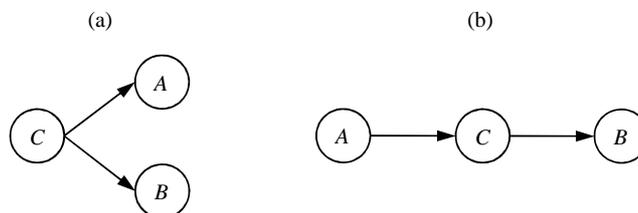
la valeur d'une variable sans changer les valeurs de ses causes ni sans altérer les relations causales au sein du modèle. Enfin, parce qu'il dépend d'éléments individuels, contextuels et temporels, l'assemblage des éléments constitutifs du système – le mécanisme causal - est transitoire tout en pouvant avoir des fonctions invariables quel que soit le contexte (voir à ce sujet Courgeau, 2012 ; Reuchlin, 1978).

### Utilisation inférentielle de la cause : quelle différence fait une cause dans son effet ?

Les conceptions qui partent de l'idée qu'une cause fait une différence dans son effet constituent une seconde famille de théories de la causalité dans laquelle on trouve l'interprétation probabiliste, l'interprétation contrefactuelle et l'interprétation agentique ou interventionniste de la causalité.

#### *Causalité et probabilités*

Les théories probabilistes de la causalité se fondent sur l'intuition que la causalité ne peut être comprise de manière satisfaisante sans être associée à des lois causales probabilistes. Elles cherchent donc à caractériser ou à analyser les relations causales en considérant que : i) le changement d'une cause produit une différence dans ses effets ; ii) cette différence se traduit par des dépendances probabilistes qui caractérisent la relation causale. Ces dépendances probabilistes fourniraient même pour certains les conditions nécessaires et suffisantes de la connexion causale.



**Figure 4** – L'événement *C* : a) est la cause des événements *A* et *B* (fourche conjonctive) ; b) est un intermédiaire causal entre les événements *A* et *B*.

Le principe de la cause commune (Reichenbach, 1956 ; cité dans Williamson, 2009) énonce ainsi que si deux événements *A* et *B* apparaissent conjointement plus fréquemment qu'attendu sur la base de leurs occurrences indépendantes c'est-à-dire si  $P(AB) > P(A)P(B)$ , *A* et *B* ont une cause commune *C* qui les précède dans le temps (figure 4a). Un autre principe dit de l'événement causal intermédiaire (*causal betweenness*) pose que *C* appartient à la chaîne causale qui relie l'événement *A* à l'événement *B* (figure 4b) si  $P(B|C) > P(B|A) > P(B)$ , si  $P(A|C) > P(A|B) > P(A)$  et si  $A \perp\!\!\!\perp B|C$ , où le symbole  $\perp\!\!\!\perp$  marque l'indépendance probabiliste. Suppes (1970) réduit

formellement la causalité à une augmentation de la probabilité :  $C$  est la cause de  $E$  si  $P(E|C) > P(E)$ , toutes choses égales par ailleurs<sup>16</sup>.

En référence à des variables aléatoires fonctionnellement reliées<sup>17</sup>, cette conception probabiliste de la causalité conduit à considérer que  $X$  n'est pas une cause de  $Y$  si la distribution de  $Y$  sachant  $X$  est la même que la distribution marginale de  $Y$  soit :  $Y \perp\!\!\!\perp X$  si  $P(Y|X) = P(Y)$ . Mais comme deux variables  $X$  et  $Y$  peuvent être dépendantes tout en étant conditionnellement indépendantes sachant une variable (ou un ensemble de variables) de « contexte »  $Z$ , la causalité probabiliste est basée sur les propriétés d'indépendance conditionnelle des distributions de probabilité. Nous verrons plus loin que les modèles graphiques causaux perpétuent cette conception probabiliste de la causalité en l'associant à la représentation graphique de relations causales.

Les conceptions probabilistes de la causalité ne sont pas sans difficultés. En l'absence d'un certain « ordre causal » défini *a priori*, il est difficile d'égaliser des relations asymétriques de cause à effet avec des relations symétriques de dépendance ou d'indépendance probabiliste. La pertinence du contrôle des variables de contexte n'est pas facile à apprécier. Certaines dépendances probabilistes peuvent être le simple fait du hasard, être dues à une cause commune ou à un autre type de relation qu'une relation de cause à effet. C'est le cas lorsque deux variables entretiennent une relation logique (par ex., la moyenne et l'écart-type d'une distribution de valeurs) ou covarient pour des raisons autres que causales (par ex., le coût de la vie et le réchauffement climatique). L'existence d'une relation causale ne s'accompagne pas non plus toujours d'une augmentation de la probabilité de l'effet (par ex., quand l'effet ne se fait sentir qu'au-delà d'un certain seuil). D'un point de vue explicatif, la différence faite par une cause dans son effet inférée à partir d'une dépendance probabiliste ne peut donc suffire.

### *Causalité et contrefactuels*

Dans les théories contrefactuelles de la causalité, juger de la dépendance causale entre  $X$  et  $Y$  consiste à évaluer l'énoncé contraire aux faits ou

---

<sup>16</sup> Si la cause  $C$  est présente chez chaque individu de la population, la force causale de  $C$  sur  $E$  est mesurée par l'augmentation  $P(E|C) - P(E)$  de la fréquence de  $E$ .

<sup>17</sup> Une relation entre deux variables est dite fonctionnelle lorsque la première est associée par une fonction donnée à la seconde. Les relations fonctionnelles sont souvent envisagées dans un contexte déterministe mais elles peuvent l'être aussi dans un contexte probabiliste. Mulaik (1986, 1993, 2009) donne une définition formelle d'une relation fonctionnelle probabiliste dans laquelle, en faisant certaines suppositions et sous certaines conditions, une fonction  $P: X \rightarrow P(Y)$  assigne à une et une seule valeur  $x$  de  $X$ , une et une seule distribution de probabilité  $P_x(y)$  de l'ensemble des distributions de probabilité des valeurs de  $Y$ .

contrefactuel « si  $X$  n'avait pas eu lieu,  $Y$  n'aurait pas eu lieu », toutes choses égales par ailleurs. La connexion entre la causalité et les énoncés conditionnels contrefactuels a été initialement discutée par Lewis (1973). Le langage contrefactuel est depuis largement utilisé dans l'étude des effets des causes ainsi que dans celle des causes des effets.

Dans l'étude des effets des causes (par ex., quel est l'effet d'un traitement ?), l'évaluation du lien de causalité consiste à comparer l'évènement d'intérêt quand l'action a eu lieu à ce qu'il aurait été si l'action n'avait pas eu lieu, toutes choses égales par ailleurs. Dans l'étude des causes des effets (par ex., quels sont les facteurs responsables de ce qui est observé ?), l'évaluation du lien de causalité consiste à comparer l'évènement observé à ce qu'il aurait été si les facteurs suspectés d'en être causalement responsables n'avaient pas varié, toutes choses égales par ailleurs.

L'approche contrefactuelle a conduit au développement dans l'étude des effets des causes d'un modèle canonique connu sous le nom de modèle des « résultats potentiels » (Rubin, 1974 ; Holland, 1986) ou modèle de Neyman-Rubin-Holland. Supposons que les individus d'une population donnée (par ex., des collégiens) soient répartis aléatoirement dans un groupe traitement (par ex., un atelier de motivation et d'orientation scolaire) et un groupe contrôle. On souhaite estimer l'effet du traitement sur les réponses des élèves à un questionnaire sur le projet personnel. Dans cet exemple, la variable causale<sup>18</sup>  $T$  prend les valeurs 1 (traitement) et 0 (contrôle) et la variable dépendante  $Y$  mesure l'effet de  $T$  sur une unité d'observation (l'individu) pour une population donnée. Chaque individu peut potentiellement donner deux réponses. La réponse potentielle de l'individu  $i$  placé dans le groupe traitement est  $y_i(1)$ , celle de même individu placé dans le groupe contrôle est  $y_i(0)$ . Au niveau de la population, la variable de résultat  $Y$  est définie par deux variables latentes de résultat ou « résultats potentiels »,  $Y(1)$  si  $t=1$  et  $Y(0)$  si  $t=0$ , soit :

$$Y = TY(1) + (1 - T)Y(0).$$

Le problème fondamental du point de vue de l'attribution de la cause est celui de savoir quelle est la valeur de l'effet causal individuel  $y_i(1) - y_i(0)$ . Or cet effet est inobservable puisqu'un seul des deux résultats potentiels est

---

<sup>18</sup> Holland pour qui il n'y a « pas de causalité sans manipulation » (1986, p. 959) considère qu'une variable n'est causale que lorsque la répartition des individus par rapport à cette variable peut être aléatoire, ce qui exclut des attributs comme le genre ou l'âge. Pour beaucoup d'auteurs cependant, définir un effet causal est une activité purement mentale distincte de l'identification et de l'estimation de l'effet causal. Les difficultés empiriques à estimer l'effet du genre par exemple n'empêchent donc pas d'envisager que le genre puisse avoir un effet causal sur la conduite.

observé pour chaque individu. Sous certaines conditions<sup>19</sup>, les paramètres de la distribution de l'effet causal peuvent néanmoins être identifiés à partir de la distribution des observables  $(Y, T)$ . Par exemple, l'effet causal moyen dans la population (*ATE* : *average treatment effect*) défini par :

$$ATE = E[Y(1) | t = 1] - E[Y(0) | t = 0]$$

et l'effet causal moyen pour les seuls participants ayant bénéficié du traitement (*ATT* : *average treatment effect on the treated*) défini par :

$$ATT = E[Y(1) | t = 1] - E[Y(0) | t = 1]$$

sont identifiables si (condition suffisante) la variable des résultats potentiels est indépendante de la variable d'accès au traitement. On dit alors que le mécanisme d'accès des individus au traitement est *ignorable* soit :

$$(Y(1), Y(0)) \perp\!\!\!\perp T.$$

Si cette condition est vérifiée, par exemple en affectant aléatoirement les individus dans les deux groupes<sup>20</sup>,  $E[Y(0) | t = 0]$  peut être utilisée comme substitut au contrefactuel non observable  $E[Y(0) | t = 1]$  et  $ATE = ATT$ . Le contrôle expérimental est donc très important dans l'analyse contrefactuelle des effets des causes.

Les approches contrefactuelles ont été critiquées sur plusieurs points. Pour certains statisticiens, la considération d'entités non observables en affaiblit les fondements empiriques (Dawid, 2000, 2007). Par ailleurs, la manipulation et la randomisation, si elles sont utiles à l'inférence de relations causales, ne sont pas des conditions nécessaires<sup>21</sup>. Les approches contrefactuelles rencontrent aussi des difficultés importantes lorsque les expériences contrôlées sont impossibles, ce qui est le cas dans la construction de modèles des causes des effets. Il faut d'abord que le cadre théorique soit suffisamment informatif pour pouvoir postuler certaines relations causales entre variables. Lorsque les données

---

<sup>19</sup> Le modèle des résultats potentiels suppose que l'effet du traitement porte sur les observations individuelles et ne déborde pas sur d'autres unités (*Stable Unit Treatment Value Assumption* ou *SUTVA*) ou que l'effet individuel du traitement est identique pour tous les individus (postulat d'homogénéité). Cette supposition est comparable à celle, plus classique en statistiques, d'observations indépendantes et identiquement distribuées (i.i.d.).

<sup>20</sup> L'affectation aléatoire des individus dans une expérimentation contrôlée supprime la possibilité d'un effet causal des unités expérimentales sur  $T$  et réduit, sans la supprimer, la vraisemblance d'une corrélation entre  $T$  et une autre cause de  $T$ .

<sup>21</sup> Les astrophysiciens ont par exemple pu démontrer, sans manipulation ni randomisation, que les forces de gravité exercées par la lune et le soleil agissent sur les marées.

d'observation sont recueillies dans des études non-randomisées, la répartition des sujets n'est que constatée par rapport à des facteurs connus (l'âge, le genre, etc.) alors que des facteurs non pris en compte peuvent intervenir. De nombreux biais (de confusion, de classement, de sélection) liés à des variables individuelles et/ou à certains facteurs contextuels doivent donc être contrôlés. Ces biais peuvent être réduits grâce à l'emploi de méthodes statistiques ou économétriques comme les techniques de conditionnement, de régression, d'appariement sur les caractéristiques observables résumées par un score de propension<sup>22</sup> ou la méthode des variables instrumentales (Angrist & Pischke, 2010 ; Morgan, 2013 ; Morgan & Winship, 2007). Les conséquences des écarts possibles à certaines hypothèses critiques (par ex., celle que toutes les variables qui influencent simultanément les réponses des individus sont observées) sur l'estimation et l'interprétation des effets causaux peuvent aussi être évaluées avec des techniques comme l'analyse de sensibilité (Rosenbaum, 2002). Il est enfin possible de s'appuyer sur différents critères (consistance de la relation sur divers échantillons, plausibilité et bien-fondé de la relation causale, etc.) pour valider les inférences d'un modèle des causes des effets. Mais si les questions contrefactuelles fournissent une aide pour inférer des relations causales, l'élément contrefactuel n'est ni nécessaire, ni suffisant et « la contrefactualité ne peut en elle-même servir de base à l'inférence de relations causales » (Mouchart, Russo & Wunsch, 2010, p. 427).

#### *Causalité et intervention formelle*

L'interprétation agentique ou interventionniste de la causalité (Woodward, 2003) repose sur l'idée que  $X$  est la cause de  $Y$  si et seulement s'il est possible de changer la valeur de  $Y$  en changeant la valeur de  $X$ . Par exemple, l'énoncé «  $X$  n'a pas d'effet sur  $Y$  » signifie que, si l'on considère différents régimes de manipulation de  $X$ , la valeur de  $Y$  (ou la distribution de probabilité de  $Y$ ) ne dépend pas de la valeur assignée à  $X$ .

Dans l'étude des effets des causes, le changement de  $X$  peut être effectué par manipulation expérimentale. Mais dire que la manipulation est la manière la plus efficace de démontrer la causalité n'est pas dire, comme Holland (1986), que ce qui n'est pas manipulable ne peut être une cause. L'action de manipuler peut en effet être comprise comme une « expérimentation par la pensée » (Aligica & Evans, 2008) en imaginant par exemple changer la valeur  $x$  de  $X$  en  $x'$ . Ce point de vue est notamment celui de Woodward (2003, 2007) qui défend une conception interventionniste de la causalité basée sur l'idée de manipulation expérimentale idéalisée.

---

<sup>22</sup> Le score de propension d'un individu  $i$  mesure sa probabilité d'être affecté au traitement, conditionnellement aux variables observées avant cette affectation et indépendamment des résultats de l'affectation (voir à ce sujet le numéro spécial de *Multivariate Behavioral Research*, 2011, 46(3)).

Cette intervention consiste à modifier fictivement l'issue d'un évènement en modifiant une de ses causes, puis à évaluer un énoncé du type «  $X$  est la cause de  $Y$  si et seulement s'il est vrai que si  $X$  variait,  $Y$  varierait, toutes choses égales par ailleurs ». Un aspect central et discuté de la théorie de Woodward concerne l'invariance de la relation causale<sup>23</sup> entre  $X$  et  $Y$  lorsque la valeur de  $Y$  change suite à une intervention formelle qui change la valeur de  $X$ . A cette forme d'invariance locale s'ajoute une forme modulaire d'invariance qui concerne l'invariance d'une équation structurelle c'est-à-dire son autonomie locale face aux changements des autres équations du système (Pearl ([2000] 2009).

La conception interventionniste de la causalité s'appuie sur deux suppositions importantes : i) le changement de la valeur de  $X$  est dû à la seule intervention, ce qui implique que les variables causalement reliées à  $X$  avant l'intervention ne le sont plus lors de l'intervention ; ii) si l'intervention change la valeur de  $Y$ , ce changement opère par le seul changement de  $X$  et non par une autre piste causale. L'exemple de la figure 5 montre en quoi l'intervention formelle se distingue de l'observation<sup>24</sup>. Supposons que  $Z$  soit la cause commune de  $X$  et  $Y$ . L'intervention formelle sur  $X$  consiste à altérer la structure causale en supprimant l'effet de  $Z$  sur  $X$ , en fixant la valeur de  $X$  à  $x_0$  tout en laissant inchangé l'effet de  $Z$  sur  $Y$ .



**Figure 5** – La causalité est une propriété de modèles hypothétiques dans lesquels les effets causaux sont générés par des variations hypothétiques de causes indépendantes les unes des autres : a) Observer :  $P(Y | X) \neq P(Y)$  ; b) Intervenir :  $P(Y | X \text{ fixé à } x_0) = P(Y)$ . L'efficacité de l'intervention formelle tient dans la rupture de la relation causale entre  $Z$  et  $X$ .

L'observation passive du système nous apprend que  $X$  et  $Y$  sont inconditionnellement dépendantes. La question « quelle est la probabilité  $P(Y=y)$  sachant qu'on observe la valeur  $x$  de  $X$  ? » est de nature factuelle. Le

<sup>23</sup> Par exemple, selon Woodward, l'équation  $Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$  décrit une relation causale si pour chaque  $X_i$ , les interventions qui changent la valeur de  $X_i$  changent  $Y$  sans que les valeurs des  $\beta_j (i \neq j)$  ne changent (invariance de niveau) et sans que la valeur ni la distribution de probabilité de  $\varepsilon$  ne changent (invariance de valeur et invariance de distribution).

<sup>24</sup> Haavelmo (1943) a été le premier à discuter la distinction entre « conditionner » par une variable (stratifier la population statistique en fonction des valeurs prises par la variable) et « fixer » une variable (maintenir la variable constante) dans les modèles à équations simultanées.

conditionnement sur  $X$  limitant l'analyse à la sous-population des individus pour lesquels  $X=x$  (autant de lois conditionnelles que de valeurs possibles pour  $X$ ), la probabilité de  $Y$  conditionnellement à  $X$  est différente de la probabilité inconditionnelle de  $Y$ .

Après l'intervention sur  $X$ , «  $X$  fixée à la valeur  $x_0$  » et  $Y$  sont indépendants. La question « quelle serait la probabilité  $P(Y=y)$  si  $X$  était fixée à la valeur  $x_0$ ? » c'est-à-dire « quelle serait la proportion d'individus pour chaque niveau de réponse de  $Y$  si les variables influençant  $X$  étaient contrôlées et si  $X$  était fixée à  $x_0$ ? » est cette fois de nature contrefactuelle. Cette question attend une réponse quantitative sous la forme d'une quantité causale qu'il est possible de calculer (si elle est identifiée) ou pas (si elle n'est pas identifiée).

L'utilisation des modèles structuraux comme outil d'évaluation d'hypothèses causales nécessite donc de pouvoir apprécier, au-delà des analyses statistiques, le potentiel de manipulation et de contrôle des causes postulées des effets observés (Kaplan, 2009 ; Mulaik, 2009 ; Woodward, 2003).

### **Un même « rationnel de variations »**

L'adhésion à une conception de la causalité plutôt qu'à une autre a évidemment un impact sur le type et sur la validité des hypothèses de modélisation employées pour inférer les causes des effets à partir de données d'observation non expérimentales. Alors que chacune des théories mécanistes, probabilistes, contrefactuelles ou interventionnistes de la causalité rencontre ses propres difficultés, ces théories sont complémentaires sous de nombreux aspects et, selon certains auteurs, peuvent être articulées dans une conception « pluraliste » de la causalité (par ex., Godfrey-Smith, 2009). Parce qu'elles sont sous-tendues par l'évaluation d'une certaine forme de variation conjointe entre des variables considérées comme formant un système, les différentes méthodologies d'analyse causale quantitative ont notamment en commun d'être gouvernées par un même « rationnel de variations » (Russo, 2009 ; Russo, Wunsch & Mouchart, 2011). Qu'il s'agisse des approches mécanistes, des approches contrefactuelles ou des approches interventionnistes, les hypothèses causales sont en effet toujours testées en mesurant une certaine forme de variations et la régularité examinée est bien celle des variations. S'agissant du test d'hypothèses causales, celles-ci peuvent être des variations contrefactuelles, des variations obtenues par contrôle statistique ou des variations obtenues par intervention sur une ou plusieurs variables de la structure causale.

L'approche structurale que les LSEM permettent de mettre en œuvre s'inscrit dans cette perspective pluraliste de la causalité. Défini comme un réseau de causes et d'effets, le modèle structural apparaît alors comme une proposition d'explication, en réponse à des questions explicitement posées, d'une (ou de plusieurs) forme(s) de variations analysées en référence à

différents systèmes de raisonnement de la causalité (mécaniste, contrefactuel, interventionniste).

## **L'émergence d'un cadre (de modélisation) structural(e) de la causalité**

Le développement en intelligence artificielle (Lauritzen, 1996 ; Pearl, 1988, [2000] 2009 ; Spirtes, Glymour & Scheines, [1993] 2000)) et en épidémiologie (Robins, 1986 ; Greenland, Robins & Pearl, 1999) des modèles graphiques probabilistes de la causalité et l'utilisation de la notion de contrefactuel dans les approches philosophiques contemporaines de la causalité (Hitchcock, 2001 ; Woodward & Hitchcock, 2003 ; Woodward, 2003) ont conduit à l'émergence d'un cadre structural de la causalité dans lequel la méthodologie des modèles d'équations structurelles trouve une place de choix. Nous nous centrerons dans cette dernière partie sur l'approche de Pearl du fait de son retentissement dans l'étude de la causalité (Dechter, Geffner & Halpern, 2010). Le modèle causal structural développé par Pearl (SCM : *Structural Causal Model* ; 1995, 1998, [2000] 2009) intègre les modèles graphiques du raisonnement probabiliste et de l'analyse causale, les modèles d'équations structurelles et la théorie des résultats potentiels de Neyman-Rubin-Holland. L'approche repose sur un langage mathématique qui clarifie la distinction entre modèles causaux et modèles purement statistiques. Elle offre une méthodologie de l'inférence causale contrefactuelle permettant de tester les prédictions logiques d'assertions causales explicites en s'appuyant sur la représentation graphique du processus de génération des données.

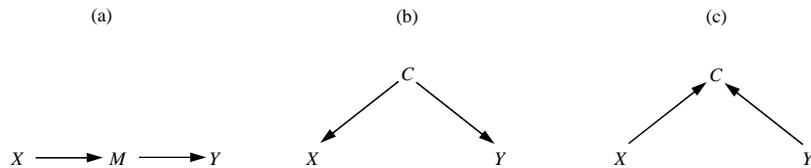
### **Les modèles graphiques probabilistes**

Les modèles graphiques et les algorithmes de calcul associés ont été développés dans les années 1980 pour « faciliter la prédiction et l'abduction en intelligence artificielle » (Pearl, [2000] 2009, p. 20). Les réseaux bayésiens qui appartiennent à cette famille de modèles reposent sur la mise en correspondance entre : i) un graphe orienté sans circuit ou DAG (*Directed Acyclic Graph*) : ii) la distribution de probabilité jointe représentée par les distributions des variables du graphe conditionnellement à leurs causes directes avec une hypothèse d'indépendance dite condition de Markov causale.

#### *Les deux fonctions d'un DAG*

Un DAG est un graphe formé de nœuds, un nœud étant la jonction de plusieurs branches, et de branches ou arcs dirigés qui représentent les relations directes mais non réciproques entre nœuds. La première fonction d'un DAG est de représenter une structure causale. Un DAG est dit causal s'il respecte la condition de suffisance causale (*causal sufficiency*) c'est-à-dire s'il inclut toutes

les causes mesurées et non mesurées de deux variables quelconques du graphe<sup>25</sup>. Lorsque les nœuds représentent des variables aléatoires, le DAG a pour autre fonction de représenter des distributions de probabilité. L'encodage des suppositions causales sur le processus de génération des données dans la population se fait cependant sans aucune hypothèse sur la distribution des variables aléatoires.



**Figure 6** – Les 3 sources structurelles d'association dans un DAG : a) chaîne  $X \rightarrow M \rightarrow Y$  ; b) piste divergente ou fourche conjonctive  $X \leftarrow C \rightarrow Y$  ; c) piste convergente  $X \rightarrow C \leftarrow Y$ .

Tout DAG comporte trois sources structurelles d'association (figure 6) : la chaîne ( $X \rightarrow M \rightarrow Y$ ), la piste divergente ( $X \leftarrow C \rightarrow Y$ ) et la piste convergente ( $X \rightarrow C \leftarrow Y$ ). A chaque source structurelle d'association correspond une source structurelle de biais (Elwert, 2013 ; Elwert & Winship, 2014).

La chaîne  $X \rightarrow M \rightarrow Y$  (figure 6a) est associée à la causalité. L'effet causal indirect de  $X$  sur  $Y$  est identifié par l'association marginale entre  $X$  et  $Y$ . Le biais dit de sur-contrôle résulte du conditionnement sur  $M$  et se traduit par la disparition de l'association entre  $X$  et  $Y$ . Dans la piste divergente ( $X \leftarrow C \rightarrow Y$ ),  $X$  et  $Y$  sont associées par leur cause commune  $C$  (figure 6b). L'effet causal direct de  $X$  sur  $Y$  est identifié par l'association conditionnelle entre  $X$  et  $Y$  sachant  $C$ . Le non conditionnement par  $C$  produit une association trompeuse entre  $X$  et  $Y$  ou biais de confusion. Enfin, dans la piste convergente ( $X \rightarrow C \leftarrow Y$ ) où  $C$  est une convergence (*collider*) (figure 6c), l'effet causal direct de  $X$  sur  $Y$  est identifié par l'association marginale entre  $X$  et  $Y$ . De manière plus contre-intuitive, si l'on conditionne par  $C$  (ou par une des variables que  $C$  influence), les deux variables  $X$  et  $Y$  deviennent dépendantes<sup>26</sup> pour au moins un niveau de  $C$  (biais de sélection dit endogène).

<sup>25</sup> Les causes non-communes exogènes (les termes d'erreur) ne sont en général pas représentées dans un DAG.

<sup>26</sup> Par exemple, il n'y a aucune association entre  $X$  et  $Y$  dans la piste convergente : ( $X$ =pluie/non pluie)  $\rightarrow$  ( $C$ =humide/sec)  $\leftarrow$  ( $Y$ =arrosage/non arrosage). Mais si le sol est humide (instanciation de  $C$ ) et qu'on sait qu'il n'a pas été arrosé, on peut en déduire qu'il a plu.

La connexion entre la structure causale et l'indépendance entre les variables aléatoires d'un DAG causal repose sur la condition de Markov causale et la  $d$ -séparation.

#### *La condition de Markov causale*

Soit une structure causale  $(X_1, X_2, \dots, X_n)$  causalement suffisante c'est-à-dire telle que l'ensemble de ses nœuds soit suffisant pour représenter toutes les relations d'indépendance conditionnelle pouvant être extraites des données d'observation. Soit  $PA_j$  l'ensemble des parents de  $X_j$  et  $ND_j$  l'ensemble des non-descendants de  $X_j$ . Dans sa version locale, la condition de Markov causale postule que si un DAG formalise une structure causale  $(X_1, X_2, \dots, X_n)$ ,

$$X_j \perp\!\!\!\perp ND_j \mid PA_j,$$

pour  $j = 1, \dots, n$ . En d'autres termes,  $X_i$  est indépendante de toutes les autres variables (sauf de celles qu'elle influence) sachant toutes ses causes directes<sup>27</sup>. L'importance de la condition de Markov causale réside dans la traduction des relations causales en relations statistiques. Ainsi, si un ensemble de nœuds obéit à la condition de Markov causale, l'observation d'une dépendance entre un nœud  $X_j$  et un de ses non-descendants  $ND_j$  implique l'existence d'une relation causale entre ce non-descendant et  $X_j$ .

#### *La $d$ -séparation*

La définition de l'indépendance entre les nœuds d'un DAG repose sur le concept de  $d$ -séparation ( $d$  pour directionnelle) qui s'applique à une relation entre trois ensembles disjoints de nœuds. Si un DAG respecte la condition de Markov causale, toutes les relations d'indépendance impliquées par les suppositions causales qu'il représente sur la population sont obtenues par la règle de  $d$ -séparation (Pearl, [2000], 2009).

Formellement, une piste  $p$  entre deux ensembles de nœuds est dite  $d$ -séparée (ou bloquée) par un ensemble  $S$  de nœuds si :

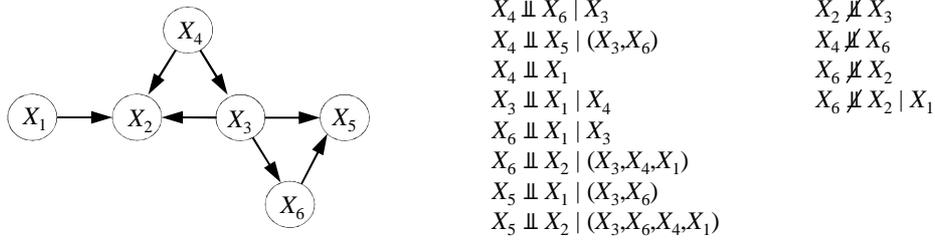
- i)  $p$  contient au moins un nœud  $m$  émettant une flèche (chaîne ou divergence) tel que  $m \in S$  ou ;
- ii)  $p$  contient une convergence  $m$  telle que  $m \notin S$  et qu'aucune variable causée directement ou indirectement par  $m$  (« descendant » de  $m$ ) n'appartienne à  $S$  (*ibid*, p. 17).

---

<sup>27</sup> La supposition de Markov causale est implicite dans les modèles linéaires comme l'analyse de régression, l'analyse de parcours ou l'analyse factorielle.

Concrètement, pour savoir si deux nœuds (ou ensembles de nœuds)  $X$  et  $Y$  sont  $d$ -séparés par un troisième nœud (ou ensemble de nœuds)  $Z$ , il faut dresser la liste de tous les chemins non dirigés entre  $X$  et  $Y$  et identifier ceux qui ne contiennent pas de convergence et  $d$ -séparent  $X$  et  $Y$  (il n’y a donc pas d’influence causale de  $X$  sur  $Y$  via chacun de ces chemins). Si une convergence sur un chemin appartient à  $Z$  ou si elle a un descendant causal qui appartient à  $Z$ , le chemin ne  $d$ -sépare pas  $X$  et  $Y$  sachant  $Z$ .

L’application du critère de  $d$ -séparation permet ainsi de retrouver toutes les indépendances conditionnelles qui découlent d’un DAG qui respecte la condition de Markov causale (figure 7). Il s’agit là d’un résultat général qui ne dépend ni des hypothèses distributionnelles, ni de la forme fonctionnelle des relations causales<sup>28</sup>.



**Figure 7** – Le DAG causal implique un ensemble d’indépendances et de dépendances marginales et conditionnelles dont certaines sont présentées à droite du graphe.

#### *Distribution de probabilité jointe générée par le DAG*

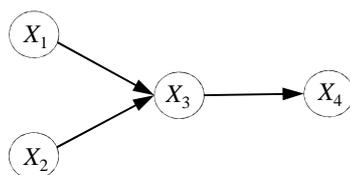
Si le DAG est causalement suffisant et que les relations d’indépendance obtenues en appliquant la  $d$ -séparation au graphe tiennent dans la distribution de probabilité de la population sur l’ensemble des variables du DAG (supposition de fidélité), la probabilité jointe de l’ensemble  $(X_1, X_2, \dots, X_n)$  sur la structure d’indépendance conditionnelle décrite par le graphe peut être factorisée et est égale à :

$$P(X_1, X_2, \dots, X_n) = \prod_{j=1}^n P(X_j \mid PA(X_j)).$$

Cette formule de décomposition permet de simplifier le calcul de la distribution de probabilité jointe de l’ensemble de variables  $(X_1, X_2, \dots, X_n)$  puisqu’il suffit de spécifier les probabilités conditionnelles des parents directs

<sup>28</sup> Dans les modèles linéaires, la  $d$ -séparation de  $X$  et  $Y$  sachant  $Z$  implique une corrélation partielle  $r_{xy.z}$  nulle.

$P(X_j | PA(X_j))$  de chaque variable et non l'ensemble des probabilités conditionnelles. L'exemple de la figure 8 montre comment une distribution de probabilité jointe qui obéit à la condition de Markov causale peut être décomposée en probabilités conditionnelles qui n'impliquent que les parents causaux des variables.



$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2)P(X_3 | X_1, X_2)P(X_4 | X_3)$$

**Figure 8** – Le DAG causal et la distribution de probabilité jointe générée par le graphe satisfait la condition de Markov causale.

#### *Critères graphiques associés aux DAG*

Les DAG causaux offrent des outils graphiques utiles à l'inférence causale probabiliste (voir par exemple les présentations des modèles graphiques causaux, sur lesquelles nous nous appuyons ici, de Elwert, 2013 ; Glymour & Greenland, 2008 ; Morgan & Winship, 2007 ; Pearl, 1998, [2000] 2009, 2010a, 2012 ; Shalizi, sous presse).

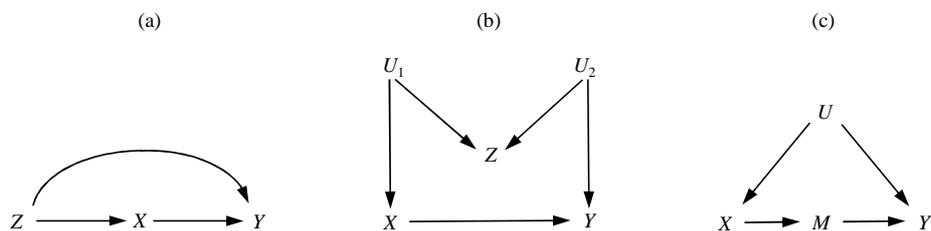
Nous avons vu plus haut que chaque condition de  $d$ -séparation correspond à un test d'indépendance conditionnelle dans l'espace probabiliste associé et que le critère de  $d$ -séparation permet d'énumérer toutes les indépendances probabilistes prédites par le DAG. Lorsqu'une relation d'indépendance probabiliste logiquement déduite du DAG n'est pas empiriquement retrouvée, le modèle causal établi *a priori* doit donc être rejeté. Lorsque les indépendances probabilistes sont en accord avec les implications logiques du DAG, la distribution de probabilités est adéquatement représentée par le graphe. Il est cependant possible que plusieurs DAG permettant de représenter les indépendances conditionnelles entre variables ne puissent être distingués sur la base des hypothèses causales préalables. La règle de  $d$ -séparation définit aussi les conditions de l'équivalence observationnelle entre DAG ayant le même squelette et le même ensemble de convergences. Par exemple, le graphe de la figure 7 peut être modifié en un graphe équivalent en changeant le sens de la relation causale entre  $X_3$  et  $X_4$  ( $X_3 \rightarrow X_4$ ) car ce changement ne modifie pas les convergences au sein du réseau (Hershberger, 2006). L'équivalence observationnelle limite ainsi les possibilités d'inférer la direction de certaines relations causales sur la base de la seule information statistique.

Plusieurs critères d'identification graphique peuvent être employés pour éliminer le biais potentiel lié à des covariables  $\{Z\}$  ou à des pistes indirectes dans l'étude de la relation causale entre deux variables  $X$  et  $Y$  (Morgan & Winship, 2007 ; Shpitser, VanderWeele & Robins, 2010). Graphiquement, l'ajustement sur les covariables  $\{Z\}$  correspond au blocage de toutes les pistes non causales entre  $X$  et  $Y$ , c'est-à-dire toutes les pistes qui commencent par une flèche pointant sur  $X$ , et au non-blocage (ou  $d$ -connexion) de toutes les pistes causales entre  $X$  et  $Y$ . Le conditionnement par  $Z$  permet ainsi d'éliminer le biais dans l'estimation de l'effet causal de  $X$  sur  $Y$  (figure 9a).

Le critère *back-door* (Pearl, 1993, [2000] 2009) permet d'identifier les pistes qui créent une association trompeuse entre  $X$  et  $Y$ . Il est satisfait par un ensemble de variables  $\{Z\}$  relativement à une paire ordonnée de variables  $(X, Y)$  si :

- i) aucun descendant de  $X$  n'est dans  $Z$  ;
- ii)  $Z$   $d$ -sépare toutes les pistes *back-door* allant de  $X$  à  $Y$ .

Par exemple, dans le DAG de la figure 9b, la variable  $Z$  est une convergence sur la piste  $X \leftarrow U_1 \rightarrow Z \leftarrow U_2 \rightarrow Y$  qui  $d$ -sépare  $X$  et  $Y$ . Le conditionnement par  $Z$  induit un biais car il ouvre la piste *back-door* qui va de  $X$  à  $Y$ .



**Figure 9** - a) La piste non causale  $X \leftarrow Z \rightarrow Y$  est ouverte. Le conditionnement par  $Z$  permet d'éliminer le biais dans l'estimation de l'effet causal de  $X$  sur  $Y$ . b) Le conditionnement par la convergence  $Z$  crée un biais dans l'estimation de l'effet causal de  $X$  sur  $Y$  ( $U_1$  et  $U_2$  non observables) ; c) L'effet causal de  $X$  sur  $Y$  n'est pas identifiable par le critère *back-door* ( $U$  non observable) mais l'est par le critère *front-door* (voir texte).

Un autre critère, le critère *front-door*, concerne les situations où un ensemble de variables  $M$  intervient comme médiateur dans la relation causale entre deux variables  $X$  et  $Y$ . L'ensemble de variables  $M$  satisfait alors le critère *front-door* relativement à une paire ordonnée de variables  $(X, Y)$  si :

- i) toutes les pistes allant de  $X$  à  $Y$  sont bloquées par  $M$  ;
- ii) aucune piste *back-door* non bloquée n'existe entre  $X$  et  $M$  ;
- iii) toutes les pistes *back-door* entre  $M$  et  $Y$  sont bloquées par  $X$ .

Supposons par exemple qu'un DAG soit constitué de la chaîne  $X \rightarrow M \rightarrow Y$  et de la piste divergente  $X \leftarrow U \rightarrow Y$  où le nœud  $U$  est non observé (figure 9c, Elwert, 2013, p. 261). L'effet causal de  $X$  sur  $Y$  n'est pas identifié car  $U$  ne satisfait pas le critère *back-door* par rapport à  $(X, Y)$ . Mais l'effet causal de  $X$  sur  $M$  est identifié en raison de l'absence de piste non causale entre  $X$  et  $M$  ( $Z = \emptyset$ ). C'est également le cas de l'effet causal de  $M$  sur  $Y$ . L'effet causal de  $X$  sur  $Y$  est donc identifié par le critère *front-door*. Il peut être calculé à partir de l'effet causal de  $X$  sur  $M$  (l'association marginale entre  $X$  et  $M$ ) et de l'effet causal de  $M$  sur  $Y$  (l'association conditionnelle entre  $M$  et  $Y$  sachant  $X$ ).

Les critères graphiques associés aux DAG causaux permettent d'éviter des biais auxquels l'ajustement classique sur les covariables par conditionnement, contrôle, ou stratification n'échappe pas toujours. Ils sont très utiles pour analyser les relations de dépendance causale entre variables en exprimant celles-ci comme des contraintes sur des fonctions de probabilité impliquant des contrefactuels. Ils permettent d'évaluer la possibilité de calculer des effets causaux et de tester la transmission d'une association sur une piste entre deux ou plusieurs variables.

### Les SEM non paramétriques

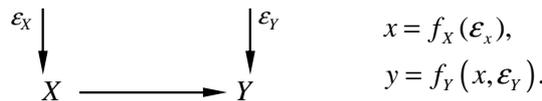
Les modèles d'équations structurelles non paramétriques (NPSEM : *Non Parametric Structural Equation Modeling*) ou modèles causaux structuraux (Pearl, [2000] 2009, p. 27) sont la généralisation non linéaire et non paramétrique des modèles d'équations structurelles compris comme des DAG causaux dans lesquels l'effet causal est défini comme « une capacité générale à transmettre des changements entre variables » (Pearl, 2010b, p. 9). Un NPSEM est défini comme un ensemble de relations fonctionnelles de la forme :

$$x_i = f_i(pa_i, \varepsilon_i), \quad i = 1, \dots, n,$$

où  $pa_i$  représente l'ensemble des variables qui déterminent directement les valeurs (les parents) de  $X_i$  et  $\varepsilon_i$  représente les erreurs structurelles dues aux facteurs implicites, non modélisés, dont la fonction de distribution conjointe est  $P(\varepsilon)$ . Les relations fonctionnelles étant considérées comme déterministes, la stochasticité découle uniquement de variables exogènes non observées. Le modèle est non paramétrique car aucune hypothèse n'est faite sur les distributions des variables ni sur la forme fonctionnelle des équations structurelles.

### Interprétation des paramètres des équations structurelles

L'interprétation des paramètres des équations structurelles n'est pas la même dans les NPSEM et dans les modèles de régression<sup>29</sup> (Pearl, 1998). Dans un NPSEM, chaque équation structurelle représente un mécanisme causal qui assigne une valeur à la variable effet en réponse à chaque combinaison de valeurs prises par  $(PA_i, E_i)$ . Par exemple, le modèle ci-dessous spécifie deux mécanismes causaux (figure 10) : l'équation structurelle  $x = f_X(\varepsilon_X)$  représente un mécanisme causal qui spécifie la valeur assignée à la variable  $X$  en réponse à chaque valeur prise par les facteurs omis  $\varepsilon_X$ ; l'équation structurelle  $y = f_Y(x, \varepsilon_Y)$  représente un second mécanisme causal qui spécifie la valeur assignée à la variable  $Y$  en réponse à chaque combinaison de valeurs prises par la variable  $X$  et les facteurs omis  $\varepsilon_Y$ .



**Figure 10** – Chaque équation structurelle représente un mécanisme causal.

On voit que l'effet causal de  $X$  sur  $Y$ , le paramètre structurel  $\beta$ , est défini de manière procédurale car le membre de droite de l'équation structurelle  $y = f_Y(x, \varepsilon_Y)$  est calculé sous l'intervention formelle «  $X$  est fixé à  $x$  » que Pearl propose de noter  $do(X = x)$  ou  $do(x)$ , puis affecté à  $Y$ . Ce paramètre mesure la sensibilité de  $E(Y)$  à l'intervention  $do(x)$  soit :

$$\beta = \frac{\partial}{\partial x} E(Y | do(x)).$$

L'erreur structurelle  $\varepsilon_Y$  (*disturbance term*) qui représente l'influence de tous les facteurs implicites pouvant affecter  $Y$  quand  $X$  est maintenu constant est définie par  $\varepsilon_Y = y - E(Y | do(x))$ . L'interprétation causale entre  $X$  et  $Y$  étant atteinte avant et indépendamment de la relation statistique entre  $X$  et  $\varepsilon$ ,

---

<sup>29</sup> Dans l'équation de régression linéaire  $y = \beta x + \varepsilon$  avec  $cov(X, \varepsilon) = 0$ ,  $\beta x + \varepsilon$  est l'espérance conditionnelle de  $Y$  sachant la valeur prise par  $X$ . Le coefficient de régression  $\beta$  mesure le changement de  $E(Y)$  par unité de changement de  $X$  soit :  $\beta = \frac{\partial}{\partial x} E(Y | X = x)$  et le terme d'erreur mesure l'écart entre  $Y$  et son espérance conditionnelle  $E(Y | x)$ .

l'interprétation des paramètres structurels est indépendante de l'existence ou non d'une corrélation entre  $X$  et  $\varepsilon_Y$ <sup>30</sup>. Les statistiques de  $\varepsilon_Y$  peuvent donc être mesurées à partir des observations, une fois  $X$  instancié. Les corrélations entre erreurs structurelles peuvent être estimées empiriquement de la même manière.

Plusieurs hypothèses d'invariance accompagnent cette interprétation. Il faut d'abord, pour pouvoir affirmer que  $X$  est la seule cause de  $Y$ , que les statistiques de  $Y$  après l'intervention  $do(x)$  restent invariantes lors de la manipulation d'une autre variable  $Z$  (ou d'un autre ensemble de variables) du modèle. Cette assertion d'invariance qui s'écrit symboliquement :

$$P(Y | do(x), do(z)) = P(Y | do(x))$$

pour tout  $Z$  disjoint de  $\{X \cup Y\}$ , quelle que soit la relation statistique entre  $\varepsilon_Y$  et  $X$  (Pearl, [2000] 2009, p. 160) peut être testée empiriquement. Une autre assertion est de nature contrefactuelle. Elle énonce que si  $X$  était fixé à  $x'$  plutôt qu'à  $x$ ,  $Y$  prendrait la valeur  $\beta x' + \varepsilon_Y$ , l'erreur structurelle  $\varepsilon_Y$  ne variant pas quand la valeur de  $X$  passe de  $x$  à  $x'$ . Cette invariance du mécanisme par lequel chaque variable est déterminée par un ensemble d'autres variables, y compris des facteurs implicites, alors que d'autres mécanismes subissent des modifications est une caractéristique fondamentale des NPSEM.

#### *Intervention formelle et probabilités conditionnelles causales*

La prédiction causale cherche à déduire par une intervention formelle une conclusion contrefactuelle de prémisses factuelles. Comme nous l'avons précédemment indiqué, l'intervention formelle consiste à libérer  $X$  de toute influence causale, à fixer (instancier)  $X$  et à laisser inchangé le reste du modèle. La question posée est alors la suivante : quelle serait la proportion d'individus pour chaque niveau de réponse de  $Y$  si les variables ayant une influence sur  $X$  étaient contrôlées et si  $X$  était fixée à  $x$  ? La quantité causale attendue est ce que Pearl appelle la distribution post-intervention de  $Y$  c'est-à-dire la probabilité que le modèle altéré  $M_x$  assigne, pour chaque réalisation  $x$  de  $X$ , à chaque niveau  $y$  de  $Y$ . Cette distribution post-intervention est formellement définie par l'équation :

$$P_M(y | do(x)) \triangleq P_{M_x}(y).$$

---

<sup>30</sup> « L'association entre  $X$  et  $\varepsilon_Y$  peut entraver la possibilité d'estimer  $\beta$  à partir de données non expérimentales mais ne change pas la définition de l'effet causal  $\beta$  » (Pearl, 2010b, p. 13).

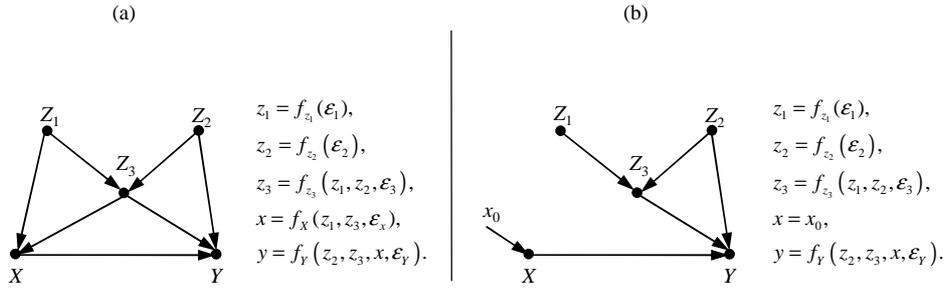
Sachant que pour tout DAG respectant la condition de Markov causale, la distribution générée par une intervention  $do(X = x_0)$  sur un ensemble  $X$  de variables endogènes dans  $M$  est donnée par la factorisation tronquée :

$$P(x_1, \dots, \hat{x}_j, \dots, x_n | do(x_j)) = \frac{P(x_1, \dots, x_n)}{P(x_j | pa(x_j))}$$

où  $P(x_j | pa(x_j))$  sont les probabilités conditionnelles avant l'intervention, il est possible de calculer des quantités causales hypothétiques à partir des observations et d'une version modifiée du modèle original.

Par exemple, dans le DAG de la figure 11a, la distribution jointe associée avant intervention est :

$$P(x, z_1, z_2, z_3, y) = P(z_1)P(z_2)P(z_3 | z_1, z_2)P(x | z_1, z_3)P(y | z_2, z_3, x).$$



**Figure 11** – Relations fonctionnelles au sein d'un DAG satisfaisant la condition de Markov causale (les erreurs structurelles, non représentées sur le graphe, sont indépendantes). a) Modèle  $M$  : le conditionnement probabiliste de  $X$  est factuel et s'appuie sur la sélection d'une sous-population telle que  $X=x_0$  ; b) Modèle  $M_x$  après intervention formelle : le conditionnement causal de  $X$  ( $do(X=x_0)$ ) est contrefactuel et conduit à un modèle altéré dans lequel les flèches qui arrivent sur  $X$  sont élaguées et l'intervention formelle ( $do(X=x_0)$ ) est encodée par l'ajout d'une flèche arrivant sur  $X$ .

La prédiction de  $Y$  par conditionnement probabiliste de  $X$  est la distribution de probabilité de  $P(y | z_2, z_3, x)$  sur la sous-population des individus pour lesquels  $X = x_0$  (figure 11a). L'intervention locale  $do(X = x_0)$  modifie le modèle (figure 11b) en remplaçant l'équation  $x = f_x(z_1, z_3, \epsilon_x)$  par chaque réalisation  $x_0$  de  $X$ . La fonction de probabilité ordinaire de  $Y$  est donc

transformée en une nouvelle fonction de probabilité<sup>31</sup> et la distribution jointe après intervention devient :

$$P(z_1, z_2, z_3, y | do(x_0)) = P(z_1)P(z_2)P(z_3 | z_1, z_2)P(y | z_2, z_3, x_0).$$

L'effet causal de  $X$  sur  $Y$  est obtenu par marginalisation des variables  $z_1, z_2$  et  $z_3$  soit :

$$P(y | do(x_0)) = \sum_{z_1, z_2, z_3} P(z_1)P(z_2)P(z_3 | z_1, z_2)P(y | z_2, z_3, x_0).$$

#### *Identification des effets causaux*

La question centrale est bien sûr celle de savoir s'il est possible d'estimer une distribution post-intervention  $P(y | do(x_0))$  à partir de données dont on connaît la distribution avant l'intervention, par exemple  $P(x, z_1, z_2, z_3, y)$  (figure 11). Cette question de l'identification des effets causaux est complexe, même pour les spécialistes qui estiment qu'« il est parfois très difficile de savoir comment spécifier un modèle statistique paramétrique pour estimer les paramètres d'un modèle causal » (Elwert, 2013, p. 147). Elle ne se pose pas de la même manière pour un paramètre statistique dans un modèle linéaire qui est identifiable lorsqu'il est possible de l'estimer de façon unique à partir des observations que pour une quantité causale<sup>32</sup>  $Q(M)$  définie de façon procédurale sachant un modèle causal  $M$ .

Savoir si une quantité causale  $Q(M) = P_{M_x}(y)$  définie par une intervention formelle  $do(X = x)$  est identifiable nécessite d'examiner les conséquences observables de la modification du modèle  $M$  sur la distribution conditionnelle de  $Y$ . Si l'altération de  $M$  a des conséquences observables sur cette distribution, l'effet causal de  $X$  sur  $Y$  est identifiable et l'association entre  $X$  et  $Y$  identifie l'effet causal. S'il est possible de changer  $P_M(Y | do(X = x))$  sans que  $P_M(Y | X = x)$  ne change parce qu'il existe d'autres voies de transmission entre la cause  $X$  et l'effet  $Y$ , l'effet causal de  $X$  sur  $Y$  n'est pas identifiable. C'est par exemple le cas lorsque l'effet causal de  $X$  sur  $Y$  se confond avec celui d'une cause commune  $U$  non observable qui crée la piste indirecte  $X \leftarrow U \rightarrow Y$ .

---

<sup>31</sup> La prédiction probabiliste  $P(Y|X=x_0)$  est identique à la prédiction causale  $P(Y|do(X=x_0))$  si  $X$  contient tous les parents de  $Y$  et aucun de ses descendants (propriété de Markov).

<sup>32</sup> Ces quantités causales (ou « requêtes » pour Pearl) sont par exemple les effets d'une intervention formelle, des contrefactuels, des effets indirects, etc.

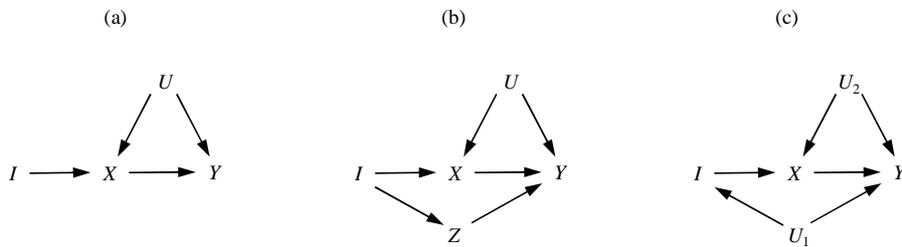
Les critères graphiques définis dans la section précédente et dont la traduction est immédiate dans le cadre des SCM (Pearl, [2000] 2009, pp. 78-83) sont particulièrement utiles pour déterminer si les quantités causales que l'on souhaite calculer sont identifiables. Ainsi, quand un ensemble de variables  $Z$  satisfait le critère *back-door* relativement à une paire ordonnée de variables  $(X, Y)$ , l'effet causal de  $X$  sur  $Y$  est identifiable et égal à :

$$P(Y | do(x)) = \sum_z P(Y | x, z)P(z).$$

De la même façon, quand un ensemble de variables  $M$  satisfait le critère *front-door* relativement à une paire ordonnée de variables  $(X, Y)$ , l'effet causal de  $X$  sur  $Y$  est identifiable et égal à :

$$P(Y | do(x)) = \sum_m P(m | x, z) \sum_{x'} P(x' | m)P(x').$$

On remarquera que les membres de droite des deux équations ci-dessus peuvent toujours être estimés avec des données d'observation non expérimentales. Il est donc toujours possible d'obtenir une estimation non paramétrique non biaisée de l'effet causal de  $X$  sur  $Y$  lorsque l'un ou l'autre de ces critères est satisfait par un ensemble de variables relativement à la paire  $(X, Y)$ .



**Figure 12** - a)  $I$  est une variable instrumentale pour l'identification de l'effet causal de  $X$  sur  $Y$  (facteur de confusion  $U$  non observable); b) En conditionnant par  $Z$ ,  $I$  est une variable instrumentale pour l'identification de l'effet causal de  $X$  sur  $Y$ ; c) Le conditionnement par  $U_1$  étant impossible,  $I$  n'est pas une variable instrumentale valide pour l'identification de l'effet causal de  $X$  sur  $Y$  ( $U_1$  et  $U_2$  non observables).  $I$  serait une variable instrumentale valide s'il était possible de conditionner par  $U_1$ .

L'effet causal de  $X$  sur  $Y$  peut être aussi identifié avec une variable instrumentale  $I$  si :

- i)  $I$  est indépendante de toutes les variables (et des erreurs structurelles) ayant une influence non médiatisée par  $X$  sur  $Y$  ;
- ii)  $I$  n'est pas indépendante de  $X$  (figure 12a et 12b).

L'effet causal de  $X$  sur  $Y$  est alors égal à :

$$P(Y | do(i)) = \sum_x P(Y | do(x))P(x | do(i)),$$

et peut être estimé si les relations entre les variables sont linéaires.

Notons qu'en dehors du cas où  $I$  est une variable d'affectation randomisée, il est difficile d'être certain, comme la supposition en est faite, qu'il n'y ait aucune autre cause commune à  $I$  et à  $Y$  que celle bloquée par  $Z$  soit :

$$I \perp\!\!\!\perp Y | Z, do(x).$$

#### *Analyse des relations contrefactuelles*

Les questions contrefactuelles du type « quelle aurait été la valeur de  $Y$  si  $X$  avait pris la valeur  $x$  pour l'unité  $U = u$  » dont la réponse attendue est la quantité de niveau unitaire  $Y_x(u) = y$ , ne peuvent pas être encodées sous la forme  $P(y | do(x))$ . Mais la stratégie d'intervention formelle précédemment décrite permet d'analyser les contrefactuels de niveau unitaire (Pearl, [2000] 2009, p. 98) en définissant la quantité  $Y_x(u)$  ou valeur de  $Y$  pour l'unité  $u$  (par ex., un individu) si  $X$  avait eu la valeur  $x$  par la solution  $Y_{M_x}(u)$  de  $Y$  dans les équations du modèle altéré  $M_x$  de  $M$  quand  $X$  est fixé à  $x$  soit :

$$Y_x(u) \triangleq Y_{M_x}(u).$$

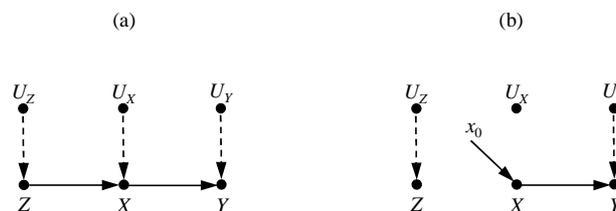
De telles quantités peuvent être dérivées des NPSEM en différenciant les relations fonctionnelles qui restent invariantes des attributs  $U = u$  qui varient d'une unité à une autre<sup>33</sup>.

Dans l'exemple ci-dessous (figure 13), l'entité contrefactuelle  $Y_{x_0}(u_x, u_y, u_z) = f_Y(x_0, u_y)$  du modèle altéré  $M_{x_0}$  mesure la manière dont un individu ayant les caractéristiques  $(u_x, u_y, u_z)$  répondrait si  $x$  était égal à  $x_0$  (figure 13b) plutôt qu'à  $f_X(z, u_x)$  (figure 13a). On pourrait de la même façon dériver  $Y_{z_0}(u)$  ou  $X_{z_0, y_0}(u)$ , ces quantités étant parfaitement définies dès lors

---

<sup>33</sup> Pearl (2010, p. 24) fait remarquer que la distinction entre causes génériques (populationnelles) et causes singulières (individuelles) ne crée aucune tension dans les NPSEM. Selon lui, la question n'est rien d'autre que celle de la spécificité de ce que  $e$  représente dans la quantité  $P(Y_x=y|e)$ . La causalité est singulière quand  $e$  inclut tous les facteurs  $u$ . Elle est générique quand  $e$  ne contient que certains attributs connus et que des probabilités sont assignées aux attributs inconnus.

que la forme des équations structurelles et les distributions des exogènes sont connues (voir Pearl, 2012, pour une illustration détaillée de la « lecture des contrefactuelles »).



**Figure 13** - a) Observer : modèle  $M$  ; b) Intervenir : modèle altéré  $M_{x_0}$  permettant de calculer l'entité contrefactuelle  $Y_{x_0}(u_Y, u_X, u_Z)$  (D'après Pearl, 2012, p. 74).

Ainsi, chaque équation structurelle, par exemple  $y = \beta x + u_Y$ , porte en elle une information contrefactuelle  $Y_x(u) = \beta x + u_Y$  avec les suppositions suivantes :

- i) linéarité de l'effet : un changement d'unité de  $X = x$  à  $X = x+1$  a le même impact sur  $Y$  qu'un changement d'unité de  $X = x'$  à  $X = x'+1$  ;
- ii) restriction d'exclusion :  $Y_{x,z}(u) = \beta x + u_Y$  où  $Z$  est n'importe quel ensemble de variables n'apparaissant pas dans le terme de droite de l'équation ;
- iii) homogénéité : l'effet causal est le même pour chaque unité de la population.

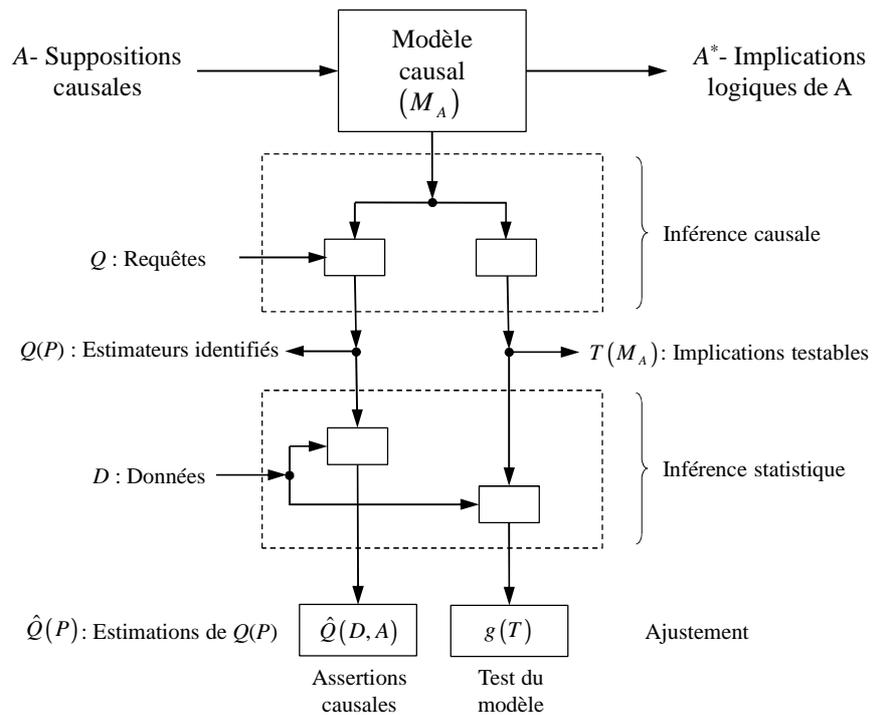
Ces suppositions contrefactuelles dont les deux premières (linéarité de l'effet, restriction d'exclusion) peuvent être empiriquement testées, sont des conséquences logiques de l'interprétation causale de l'équation structurelle  $y = \beta x + u_Y$ .

### Les modèles causaux structuraux sont des « moteurs d'inférence »

La théorie de l'inférence causale développée par Pearl enrichit et régénère la méthodologie des modèles d'équations structurelles, démontrant ainsi son intérêt dans l'analyse causale de données d'observation. Plusieurs grands principes méthodologiques, que nous reprenons ici en suivant la présentation qu'en donne Pearl (2012 ; Bollen & Pearl, 2013), guident la logique d'application des NPSEM.

Trois sources d'information contribuent à constituer le contexte de l'inférence causale (figure 14). La première source d'information provient des suppositions causales qualitatives (A) scientifiquement justifiées. Ces suppositions sont représentées sous la forme d'un modèle graphique probabiliste qui décrit l'organisation supposée des relations entre les variables considérées ( $M_A$ ). La seconde source d'information est constituée des requêtes spécifiques (Q) qui concernent les relations causales et contrefactuelles entre

variables, et auxquelles une réponse quantitative est donnée via les calculs effectués avec l'opérateur  $do(\bullet)$ . La dernière source d'information réside dans les données empiriques ( $D$ ) gouvernées par une distribution de probabilité jointe.



**Figure 14** - Contexte de l'inférence causale avec les NPSEM. Entrée : suppositions causales ( $A$ ), requêtes causales et contrefactuelles ( $Q$ ) et données empiriques ( $D$ ). Sortie : implications logiques ( $A^*$ ), implications statistiques testables ( $T$ ), assertions causales conditionnelles ( $\hat{Q}(D, A)$ ) et indices d'ajustement ( $g(T)$ ).  $P$  est la densité jointe des variables observées (d'après Pearl, 2012, p. 72).

Le traitement de ces informations conduit à trois types de résultats. Le premier ensemble de résultats ( $A^*$ ) est constitué des énoncés déduits logiquement de  $A$ , indépendamment des données ( $D$ ), par exemple «  $X$  n'a pas d'effet sur  $Y$  si  $Z$  est fixé ». Le second ensemble de résultats comporte les réponses aux requêtes ( $Q$ ) lorsqu'elles sont identifiables. Ces quantités causales notées  $\hat{Q}(D, A)$  sont des assertions empiriques conditionnelles à  $A$  qui concernent par exemple l'effet causal d'une variable sur une autre ou la probabilité d'un niveau donné de réponse chez un individu si, contrairement aux faits, il avait bénéficié d'un entraînement. Le troisième ensemble  $T$  de résultats est composé des implications statistiques testables de  $A$  et du degré d'accord de ces implications avec les données.

Si le modèle  $M_A$  décrit correctement le phénomène étudié, la méthodologie des NPSEM offre des outils formels pour lire les suppositions qui supportent chaque énoncé causal, en déduire graphiquement les implications testables, identifier les biais potentiels et les requêtes non contradictoires avec ces suppositions. La méthodologie des NPSEM fournit aussi les moyens de déterminer mathématiquement quels sont les paramètres identifiables<sup>34</sup>, d'identifier les éventuelles variables instrumentales ou de reconnaître et générer des modèles équivalents (par ex., avec TETRAD IV ; Scheines, Spirtes, Glymour, Meek & Richardson, 1998). L'intervention formelle intégrée au langage des LSEM sert ici de point d'appui au calcul causal.

La définition procédurale de la requête  $Q$  est une phase essentielle dans ce déroulement méthodologique. La requête peut être prospective c'est-à-dire concerner des états futurs sachant les conditions actuelles (par ex., si  $X$  est la cause de  $Y$  : quelle est la valeur attendue de  $Y$  pour une valeur donnée de  $X$  ? Quelle serait la valeur de  $Y$  si on intervenait pour que  $X$  prenne une valeur particulière ?). La requête peut être aussi rétrospective c'est-à-dire concerner des états passés en portant sur des attributions (importance relative de divers influences sur  $Y$ ), des effets directs et indirects ou des contrefactuels (qu'aurait-on observé sur  $Y$  si ?). Quelle que soit la requête, sa définition est celle d'un algorithme qui transforme l'information du modèle en une quantité causale  $Q(M_A)$  à partir d'équations de prédiction probabiliste. Il est important de remarquer que l'identification d'une requête ne nécessite qu'une partie (ou « noyau ») de l'ensemble des suppositions emboîtées dans le modèle  $M_A$ . L'incompatibilité entre  $M_A$  et l'organisation des données, incompatibilité pouvant être due à des parties du modèle qui n'ont rien à voir avec la dérivation de la requête  $Q$ , n'invalide donc pas nécessairement une assertion empirique en réponse à une requête  $Q$  conditionnellement aux suppositions causales ( $A$ ). Une fois les suppositions causales traduites en langage contrefactuel et si la requête est identifiée, l'expression algébrique de l'estimation de cette quantité peut être calculée. La méthode d'estimation devant être compatible avec le caractère local de celle-ci, l'estimation bayésienne (Lee, S.-Y., 2007) basée sur les données individuelles est à cet égard particulièrement adaptée.

Nous renvoyons le lecteur intéressé à Grace, Schoolmaster, Gutenspergen, Little, Mitchell, Miller & Schweiger (2012) pour une présentation des différentes étapes de l'implémentation « grapho-théorique » des LSEM et une illustration détaillant le développement d'un DAG, la description des suppositions causales et des implications logiques d'un DAG ou l'estimation bayésienne des requêtes. Signalons également que quelques articles discutant l'intérêt des DAG, du modèle des résultats potentiels et du raisonnement contrefactuel dans la modélisation de données d'observation commencent à

---

<sup>34</sup> L'identification est aussi influencée par les données et les limites des méthodes d'estimation.

apparaître en psychologie du développement (Foster, 2010) et en psychologie différentielle (Lee, J., 2012).

## Conclusion

L'explication recherchée en psychologie différentielle a besoin d'une certaine forme de pensée causale pour établir des inférences plausibles aux causes des effets. Ces inférences ne peuvent évidemment pas être effectuées sur la seule base des associations entre variables. Des suppositions complémentaires et raisonnables en substance sont nécessaires pour conclure en faveur de l'existence de relations causales. Plusieurs outils conceptuels permettent au chercheur qui emploie la méthodologie des LSEM d'explicitier les suppositions sous lesquelles les associations peuvent impliquer des relations de cause à effet, de tester empiriquement celles qui peuvent l'être et d'identifier des biais potentiels.

Le modèle des résultats potentiels est un premier outil développé en statistiques qui permet de définir clairement les effets causaux, de déterminer à quelles conditions ils sont identifiables et d'évaluer la plausibilité de ces conditions, tout cela indépendamment de la méthode d'estimation utilisée (Rubin, 1974 ; Sobel, 2009). Nous avons évoqué dans ce travail l'idée fondatrice du modèle des résultats potentiels qui fait des probabilités conditionnelles de contrefactuels le substrat des relations causales. Cette idée est largement exploitée dans l'analyse des effets des causes dans les études randomisées. Elle commence à l'être dans l'analyse avec les LSEM de données d'observation non expérimentales, par exemple pour déterminer si des effets directs et indirects dans des modèles de médiation peuvent être considérés comme des effets causaux (Muthén, 2011).

Les graphes causaux offrent également un outil conceptuel novateur et puissant pour l'inférence causale avec des données d'observation non expérimentales. Nous avons décrit brièvement quelques aspects de la théorie élaborée par Pearl ([2000] 2009) sur la base d'idées introduites par Spirtes *et al.* ([1993] 2000). L'apport de cette formalisation qui intègre les DAG causaux, les SEM et le raisonnement contrefactuel tient dans la clarté de la formulation de la notion de causalité et dans les principes méthodologiques dégagés pour représenter les suppositions causales, expliciter leurs implications logiques et estimer les effets causaux éventuellement identifiés. Une idée fondamentale dans cette théorie est que « *Y* est une cause de *Z* s'il est possible de changer *Z* en manipulant *Y* » (Pearl ([2000] 2009, p. 417). Cette idée s'accompagne de la distinction entre l'observation passive des relations entre des variables considérées comme formant un système et l'intervention formelle sur les causes supposées des effets, modifiant ainsi l'état naturel du système. Cette séparation entre interprétation probabiliste (*seing*) et interprétation interventionniste (*doing*) de la causalité n'est cependant possible qu'au prix de plusieurs

suppositions. Etant donné un DAG causal et une distribution de probabilité jointe, ni la distribution de probabilité qui résulte d'une intervention, ni les probabilités conditionnelles après intervention ne peuvent en effet être prédites si la condition de Markov causale n'est pas respectée. L'axiome d'invariance, notamment de niveau et celui de modularité ou d'autonomie locale sont également nécessaires. Les contraintes sont donc fortes d'autant que le fait de disposer de principes méthodologiques théoriquement fondés ne rend pas pour autant possible l'identification de la structure causale d'un système particulier. Les assertions causales nées de l'application de ces principes restent aussi conditionnelles au bien-fondé des suppositions causales préalables.

La théorie des modèles causaux qui prennent la forme de DAG est considérée par certains auteurs comme prototypique d'une nouvelle génération de SEM propres à l'identification d'effets causaux (Matsueda, 2012). Les équations structurelles (re)deviennent ainsi le langage naturel pour étudier et représenter les relations causales. L'utilisation des DAG pour l'inférence causale probabiliste est cependant limitée sous plusieurs aspects.

Une objection fréquente souligne l'incompatibilité entre la logique qui conduit à l'inférence causale et l'hypothèse de causalité réciproque (par ex., Drouet, 2012, en philosophie des sciences ; Borsboom, van der Sluis, Noordhof *et al.*, 2012, en psychologie). Les DAG sont en effet incomplets puisqu'ils sont récursifs. Ils peuvent néanmoins être employés avec des données temporelles pour modéliser des relations causales non récursives entre variables (Lacerda, Spirtes, Ramsey & Hoyer, 2008). Une autre approche de modélisation causale peut aussi être mise en œuvre dont l'idée de base est d'exclure et de manipuler des variables externes provenant de différentes équations afin de construire les quantités causales recherchées. L'effet causal est ainsi défini dans un système interdépendant, ce qui généralise la définition récursive de la causalité à des relations non récursives [voir les « modèles causaux (simultanés) non récursifs », Heckman, 2005, p. 38].

Beaucoup de statisticiens éprouvent également des difficultés à se défaire d'un certain scepticisme à l'égard de l'idée qu'il est possible d'identifier des relations causales avec les SEM appliqués à des données d'observation. Dans un de ses articles intitulé « Méfiez-vous des DAG ! », Dawid (2010a,b) reconnaît la fécondité de la synthèse faite par Pearl entre causalité probabiliste et causalité interventionniste mais voit un risque de réification des DAG lié à ce qu'il considère être une compréhension intuitive des termes causaux. Dawid propose de dissocier le langage formel de suppositions causales *ad hoc* et d'abandonner la supposition d'une relation entre la causalité et l'indépendance conditionnelle (condition de Markov causale). Dawid pour qui la causalité doit être comprise en termes de conséquences probabilistes des interventions, introduit la notion de « régime » (régime de repos correspondant à l'observation, régimes augmentés correspondant à des interventions) pour

explorer des conjectures causales en se libérant des modèles graphiques et en représentant explicitement les suppositions faites par Pearl. De leur côté, Lindquist et Sobel (2011) déclarent ne pas être convaincus que « les DAG (ou les LSEM) peuvent être généralement utiles pour ‘trouver des relations causales’ ou estimer des effets causaux » (p. 334). Ces auteurs avancent l’argument que les suppositions qui servent à spécifier le modèle causal représenté par le DAG constituant le modèle, les inférences causales qui en découlent ne sont pas valides si ces suppositions ne sont pas satisfaites. L’existence de ces suppositions doit d’abord être reconnue. Nous avons vu que le langage formel de la théorie des NPSEM ou le système de notation du modèle des résultats potentiels permettent d’explicitement ces suppositions. L’évaluation de la plausibilité et du bien-fondé de ces suppositions reste une question à laquelle il n’est pas possible d’apporter de réponse définitive. Cette évaluation peut être effectuée avant d’utiliser les LSEM, si l’on veut encourager le chercheur à utiliser d’autres procédures lorsque les suppositions paraissent infondées (Lindquist & Sobel, 2011). Elle peut l’être aussi avec le langage et les outils des NPSEM (Pearl, [2000] 2009). Le lecteur de ce texte aura compris que l’évaluation de la plausibilité des suppositions qui servent à spécifier le modèle causal et fondent l’inférence causale à partir de données d’observation non expérimentales reste de toute façon à la merci du jugement scientifique.

## Références

- Aldrich, J. (1994). Haavelmo’s identification theory. *Econometric Theory*, *10*, 198-219.
- Aligica, P.D., & Evans, A.J. (2008). Expérimentations par la pensée, analyses contrefactuelles et comparatives. *Revue française d’économie*, *22*, 45-71.
- Angrist, J., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, *24*, 3-30.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*, 1086–1120.
- Bacher, F. (1979). L’organisation des enquêtes et l’interprétation de leurs résultats. *Enfance*, *32(1)*, 3-13.
- Bacher, F. (1999). L’utilisation des modèles dans l’analyse des structures de covariance. *L’Année Psychologique*, *99*, 99-122.
- Barberousse, A., Kistler, M., & Ludwig, P. (2000). *La philosophie des sciences au XX<sup>ème</sup> siècle*. Paris : Flammarion.
- Baumrind, D. (1983). Specious causal attributions in the social sciences: The reformulated stepping-stone theory of heroin use as exemplar. *Journal of Personality and Social Psychology*, *45*, 1289-1298.
- Bentler, P.M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, *31*, 419-456.
- Bentler, P.M. (1998). Causal modeling: New interfaces and new statistics. In J. G. Adair, D. Belanger, & K. L. Dion (Eds.), *Advances in psychological science: Social, personal, and cultural aspects* (pp. 353-370). East Sussex, UK: Psychology Press.
- Biddle, B.J., & Martin, M.M. (1987). Causality, confirmation, credulity, and structural equation modeling. *Child Development*, *58*, 4-17.

- Blalock, H. (1964). *Causal inferences in nonexperimental research*. University of North Carolina Press, Chapel Hill, North Carolina.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York : Wiley.
- Bollen, K.A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S.L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301-328). Springer.
- Borsboom, D., van der Sluis, S., Noordhof, A. *et al.* (2012). What kind of causal modelling does personality research need ? *European Journal of Personality*, 26, 392-393.
- Boudon, R. (1965). Méthodes d'analyse causale. *Revue Française de Sociologie*, 6, 24-43.
- Bullock, H.E., Harlow, L.L., & Mulaik, S.A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling*, 1, 253-267.
- Cartwright, N. (2001). Modularity it can – and it generally does – fail. In M.C. Galavotti, P. Suppes, & D. Constantini (Eds.), *Stochastic Causality* (pp. 65-84). CSLI Publications: Stanford, California.
- Cliff, N. (1983). Some cautions regarding the application of causal modelling methods. *Multivariate Behavioral Research*, 18, 115–126.
- Courgeau, D. (2012). Peut-on parler de causalité en sciences sociales ? Les approches contrefactuelle et mécaniste. *Les rencontres de statistique appliquée. La causalité en sciences humaines et sociales : expliquer, modéliser, mesurer ?* Récupéré à <http://ined.academia.edu/DanielCourgeau/Conference-Presentations>.
- Dawid, A.P. (2000). Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association*, 95, 407–448.
- Dawid, A.P. (2007). Counterfactuals, hypotheticals and potential responses: A philosophical examination of statistical causality. In F. Russo & J. Williamson (Eds.), *Causality and Probability in the Sciences* (pp. 503-532). London: College Publications.
- Dawid, A.P. (2010a). Seeing and doing: The Pearlian synthesis. In R. Dechter, H. Gener & J.Y. Halpern (Eds.), *Heuristics, Probability and Causality: A Tribute to Judea Pearl* (pp. 309-325). London: College Publications.
- Dawid, A.P. (2010b). Beware of the DAG! In I. Guyon, D. Janzing & B. Schölkopf (Eds.), *Proceedings of the NIPS 2008 Workshop on Causality, Journal of Machine Learning Research Workshop and Conference Proceedings*, 6, 59-86.
- Dechter, R., Geffner, H., & Halpern, J.Y. (Eds.) (2010). *Heuristics, Probability and Causality. A Tribute to Judea Pearl*. London: College Publication.
- Drouet, I. (2012). *Causes, probabilités, inférences*. Paris : Vuibert.
- Duncan, O.D. (1966). Path analysis: Sociological examples. *American Journal of Sociology*, 72, 1-16.
- Elwert, F. (2013). Graphical Causal Models. In S. Morgan (ed.), *Handbook of causal analysis for social research* (pp. 245-273). New York: Sage Publications.
- Elwert, F., & Winship, C. (2014, to appear). Endogenous Selection Bias. *Annual Review of Sociology*, 40.
- Fennell, D. (2007). Causality, mechanisms and modularity: structural models in econometrics? In F. Russo & J. Williamson (Eds.), *Causality and probability in the sciences* (pp. 161-177). London: College Publication.
- Fisher, R.A. (1925). *Statistical methods for research workers*. Edinburgh, Oliver & Boyd.
- Foster, E.M. (2010). Causal inference and developmental psychology. *Developmental Psychology*, 46, 1454-1480.
- Freedman, D.A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics*, 12, 101-128.

- Glennan, S. (2009). Mechanisms. In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 315-325). Oxford University Press.
- Glymour, M.M., & Greenland, S. (2008). Causal diagrams. In K.J. Rothman, Greenland, S., & T. Lash (Eds.), *Modern Epidemiology* (3rd ed.) (pp. 183-209). Lippincott-Raven.
- Godfrey-Smith, P. (2009). Causal pluralism. In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 326-337). Oxford University Press.
- Grace, J.B., Schoolmaster, D.R., Gutenspergen, G.R., Little, A.M., Mitchell, B.R., Miller, K.M., & Schweiger, E.W. (2012). Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere*, 3, 1-44.
- Greenland, S., Robins, J., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, 14, 29-46.
- Glymour, M., & Greenland, S. (2008). Causal diagrams. In K. Rothman, S. Greenland & T. Lash (Eds.), *Modern Epidemiology* (3rd edition) (pp. 183-209). Philadelphia, PA: Lippincott Williams & Wilkins.
- Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11, 1-12.
- Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, 12 (Supplement), 1-115.
- Heckman, J.H. (2005). The scientific model of causality. *Sociological Methodology*, 35, 1-97.
- Hershberger, S.L. (2006). The problem of equivalent structural models. In G.R. Hancock, & R.O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 13-41). Information Age; Greenwich, CN.
- Hempel, B. 1994. Revue sommaire des principales théories contemporaines de la causation. In R. Franck (Ed.), *Faut-il chercher aux causes une raison ? L'explication causale dans les sciences humaines* (pp. 223-231). Paris : Librairie Philosophique J. Vrin.
- Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, 98, 273-299.
- Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945-970.
- Holland, P.W. (1988). Causal inference, path analysis, and recursive structural equations models. In C. Clogg (Eds.), *Sociological methodology* (pp. 449-484). American Sociological Association, Washington, D.C.
- Hoyle, R.H. (Ed.) (2012). *Handbook of structural equation modeling*. New York : The Guilford Press.
- Jöreskog, K.G. (1970). A general method for analysis of covariance structures. *Biometrika*, 57, 239-51.
- Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2<sup>nd</sup> edition). Thousand Oaks, CA: Sage.
- Kenny, D.A. (1979). *Correlation and causality*. New York: Wiley-Interscience.
- Kistler, M. (2004). La causalité dans la philosophie contemporaine. *Intellectica*, 38, 139-185.
- Kistler, M. (2011). La causalité. In A. Barberousse, D. Bonnay & M. Cozic (dir.), *Précis de philosophie des sciences* (pp. 100-140). Paris : Vuibert.
- Kline, R.B. (2012). Assumptions in structural equation modeling. In R.H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 111-125). New York : The Guilford Press.
- Knight, C.R., & Winship, C. (2013). The causal implications of mechanistic thinking: Identification using Directed Acyclic Graphs (DAGs). In Morgan, S.L. (Ed.), *Handbook of causal analysis for social research* (pp. 275-299). New York: Springer.

- Lacerda, G., Spirtes, P., Ramsey, J., & Hoyer, P.O. (2008). Discovering cyclic causal models by independent component analysis. *Proceedings of the 24th Conference on Uncertainty In Artificial Intelligence*, 366-374.
- Lauritzen, S.L. (1996). *Graphical Models*. Clarendon Press : Oxford.
- Lee, J.L. (2012). Correlation and causation in the study of personality. *European Journal of Personality*, 26, 372–390.
- Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. New York: Wiley.
- Lewis, D.K. (1973). Causation. *Journal of Philosophy*, 70, 556-567.
- Lindquist, M.A. & Sobel, M.E. (2011). Graphical models, potential outcomes and causal inference : Comment on Ramsey, Spirtes and Glymour. *Neuroimage*, 57, 334-336.
- Ling, R.F. (1982). Review of 'Correlation and Causation [sic].' *Journal of the American Statistical Association*, 77, 489-491.
- Little, D. (2011). Causal mechanisms in the social realm. In P. McKay Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 273-295). Oxford University Press.
- MacCallum, R.C., & Austin, J.T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51, 201-226.
- Mackie, J.L. (1980). *The cement of the universe: A study of causation*. Oxford University Press.
- Matsueda, R.L. (2012). Key advances in the history of structural equation modeling. In R.H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 17-42). New York : The Guilford Press.
- Morgan, S.L. (Ed.) (2013). *Handbook of causal analysis for social research*. New York: Springer.
- Morgan, S.L., & Winship, C. (2007). *Counterfactuals and causal inference : Methods and principles for social research*. Cambridge University Press.
- Mouchart, M., Russo, F., & Wunsch, G. (2010). Inferring causal relations by modeling structures. *Statistica*, LXX(4), 411-432.
- Mulaik, S.A. (1986). Toward a synthesis of deterministic and probabilistic formulations of causal relations by the functional relation concept. *Philosophy of Science*, 53, 313-332.
- Mulaik, S.A. (1993). Objectivity and multivariate statistics. *Multivariate Behavioral Research*, 28, 171-203.
- Mulaik, S.A. (2009). *Linear causal modeling with structural equations*. Chapman & Hall/CRC.
- Muthén, B.O. (1987). Response to Freedman's critique of path analysis: Improve credibility by better methodological training. *Journal of Educational Statistics*, 12, 178-184.
- Muthén, B.O. (2001). Second-generation SEM with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L.M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 291-322). Washington, D.C.: APA.
- Muthén, B.O. (2011). Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. Récupéré à <http://www.statmodel.com>
- Niles, H.E. (1922). Correlation causation and Wright's theory of path coefficients. *Genetics*, 7, 258-273.
- Niles, H.E. (1923). The method of path coefficients: An answer to Wright. *Genetics*, 8, 256-260.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann.
- Pearl, J. (1993). Comment: Graphical models, causality and intervention. *Statistical Science*, 8, 266-269.
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82, 669–710.

- Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods & Research*, 27, 226-284.
- Pearl, J. ([2000] 2009). *Causality: Models, reasoning, and inference* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press.
- Pearl, J. (2010a). The foundations of causal inference. *Sociological Methodology*, 40, 75-149.
- Pearl, J. (2010b). An introduction to causal inference. *The International Journal of Biostatistics*, 6(2), 1-59.
- Pearl, J. (2012). The causal foundations of structural equation modeling. In R.H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 68-91). New York : The Guilford Press.
- Reuchlin, M. (1978). Processus vicariants et différences individuelles. *Journal de Psychologie*, 2, 133-145.
- Reuchlin, M. (1980). Théories en psychologie : Explication et interprétation psychologiques. In M. Richelle & X. Seron (Dir.), *L'explication en psychologie* (pp. 237-260). Paris : PUF.
- Reuchlin, M. (1995). Totalités, éléments, structures en psychologie. Paris : PUF.
- Reuchlin, M. (1998). Signification statistique et signification scientifique. *Revue française et francophone de psychiatrie et de psychologie médicale*, 18, 98-104.
- Reuchlin, M. (1999). Evolution de la psychologie différentielle. Paris : PUF.
- Robins, J.M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modeling*, 7, 1393-1512.
- Rosenbaum, P.R. (2002). *Observational studies* (2<sup>nd</sup> edition). New York: Springer.
- Rothman K.J. (1976). Causes. *American Journal of Epidemiology*, 104, 587-592.
- Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.
- Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society*, 13, 1-26.
- Russo, F. (2009). Causality and causal modeling in the social sciences. Measuring variations. New York: Springer.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21, 157-170.
- Russo F., Wunsch G., & Mouchart M. (2011). Inferring causality through counterfactuals in observational studies. Some epistemological issues. *Bulletin de Méthodologie Sociologique*, 111, 43-64.
- Scheines, R., Spirtes, P., Glymour, C., Meek, C., & Richardson, T. (1998). The TETRAD Project: Constraint Based Aids to Causal Model Specification. *Multivariate Behavioral Research*, 33, 65-117.
- Shalizi, C.R. (in press). *Advanced data analysis from an elementary point of view*. Cambridge University Press. Récupéré à <http://www.stat.cmu.edu/~cshalizi/ADAfaEPOV/>.
- Shpitser, I., VanderWeele, T.J., & Robins, J.M. (2010). On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the 26th conference on Uncertainty and Artificial Intelligence* (pp. 527-536). Corvallis: AUAI Press.
- Simon, H. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association*, 49, 467-479.
- Sobel, M.E. (2009). Causal inference in randomized and non-randomized studies: The definition, identification and estimation of causal parameters. In R.E. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 3-22).
- Spirtes, P., Glymour, C., & Scheines, R. ([1993] 2000). *Causation, Prediction, and Search* (2<sup>nd</sup> ed.). Cambridge, MA.: The MIT Press.

- Suppes, P. (1970). *A probabilistic theory of causality*. North-Holland, Amsterdam.
- Williamson, J. (2009). Probabilistic theories. In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 185-212). Oxford University Press.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.
- Woodward, J. (2007). Interventionist theories of causation in psychological perspective. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy and computation* (pp. 19-36). New York: Oxford University Press.
- Woodward, J. & Hitchcock, C. (2003). Explanatory generalizations. Part I: A counterfactual Account, *Nous*, 37, 1–24.
- Wright, P.G. (1928). *The Tariff on Animal and Vegetable Oils*. New York: MacMillan.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics* 5, 161-215.