

Les cheminements de la causalité

Las vías de causalidad

The Progress of Causality

Serge CARLOS

Volume 2, numéro 2, novembre 1970

URI : <https://id.erudit.org/iderudit/001123ar>

DOI : <https://doi.org/10.7202/001123ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0038-030X (imprimé)

1492-1375 (numérique)

[Découvrir la revue](#)

Citer cet article

CARLOS, S. (1970). Les cheminements de la causalité. *Sociologie et sociétés*, 2(2), 189–202. <https://doi.org/10.7202/001123ar>

Résumé de l'article

Le thème de l'article est la convergence entre la réflexion sociologique et les développements des techniques d'analyse quantitative. L'auteur montre comment l'emploi des techniques basées sur le modèle général de régression amène le chercheur à pousser plus loin sa réflexion sur les phénomènes étudiés, chacune des techniques présentées impliquant des postulats quant à la dynamique des variables mesurées ou quant à l'influence de variables non mesurées. Partant de l'analyse écologique et de l'analyse contextuelle, l'auteur introduit l'analyse des cheminements de causalité en montrant la progression de la convergence des techniques et de la réflexion sociologique des premiers modèles au second.

Les cheminements de la causalité*



SERGE CARLOS

C'EST MAINTENANT DEVENU un lieu commun que de remarquer l'importance accordée aux développements méthodologiques et techniques en sociologie. Presque chaque nouveau numéro de chaque revue sociologique d'envergure contient un article sur ce sujet. Par tempérament et par formation, je me réjouis habituellement de cette richesse et de cette popularité. Cependant, cette effervescence n'est pas sans poser quelques problèmes. On a souvent l'impression que plus les méthodes statistiques d'analyse se développent, plus leur champ d'application se rétrécit. La plupart du temps, les modèles reposant sur les techniques de régression nécessitent des variables métriques alors que, plus souvent qu'autrement, les données sociologiques sont de type nominal ou ordinal. La sociologie a sûrement besoin d'une méthodologie plus précise et plus polyvalente pour l'analyse des relations simples et multiples de variables non métriques. Cette affirmation ne saurait être, malgré tout, la reconnaissance d'un fossé entre les données et les besoins analytiques du travail de recherche quotidien d'une part, et les modèles d'analyse statistique quelque peu complexes que véhiculent depuis une dizaine d'années les méthodologues d'autre part.

* Cet article est la version quelque peu modifiée d'un exposé présenté au Congrès annuel de la Société canadienne de sociologie et d'anthropologie (Winnipeg, mai 1970). Je tiens à remercier C. Michael Lanphier, Raymond N. Morris et David B. Nolle pour leurs commentaires au moment de l'exposé. Aucune de ces personnes n'est responsable du contenu de l'article. Une traduction anglaise est publiée simultanément dans *Interchange*, revue de l'Ontario Institute for Studies in Education.

Il y a plusieurs raisons qui permettent de soutenir l'inexistence d'un tel fossé. La première est qu'il y a maintenant des techniques d'analyse des relations entre variables qualitatives reposant sur des bases mathématiques aussi complexes que celles du modèle linéaire général employé pour les données quantitatives. On peut penser ici à la décomposition des probabilités conditionnelles proposée par Coleman¹, à l'analyse de l'incertitude de McGill² et au modèle logarithmique développé par Theil³. Une seconde raison serait l'attitude conservatrice de plusieurs chercheurs en sciences sociales telle que dépeinte par Tufte⁴. Bien souvent des chercheurs se privent de l'information supplémentaire contenue dans des données qualitatives pour éviter de discuter certains postulats concernant les distributions de leurs variables ou leur niveau réel de mesure. Le troisième argument qui doit nous pousser à ignorer le décalage entre le travail d'analyse de données empiriques du chercheur et les discussions parfois complexes des méthodologues est le plus important. Si l'on fait l'historique des développements méthodologiques, on s'apercevra qu'ils débouchent très souvent sur des considérations touchant au processus logique de la recherche empirique ou à la formulation plus précise de certains concepts et de certaines hypothèses sociologiques. L'histoire de la méthodologie statistique de la dernière décennie est un échange entre le méthodologue et le théoricien. Partant d'un problème technique d'analyse, on en vient rapidement à poser des problèmes plus substantiels. À preuve, cette longue confidence de Boudon où il explique comment il est arrivé à formuler le modèle d'analyse de dépendance :

Le principe une fois admis que l'interprétation des résultats d'une enquête consiste le plus souvent à analyser les propriétés des distributions de manière à déceler la présence d'une structure causale, comment étendre la méthodologie de Lazarsfeld de manière à traiter du cas général où non pas 3, mais n variables apparaissent comme interconnectées?... jusqu'à trois ou quatre variables, l'intuition suffit à conduire une analyse correcte... Au-delà, la lecture directe devient difficile, puis pratiquement impossible. Il faut alors inverser la procédure : c'est-à-dire faire l'hypothèse que les données recueillies peuvent être expliquées par une structure causale particulière et vérifier que cette structure : 1) est effectivement compatible avec les données, 2) en rend mieux compte que toute autre⁵.

Cet exposé doit traiter des modèles d'analyse des cheminements de causalité. Sous cette étiquette, j'englobe aussi bien le *path analysis* de Sewall Wright⁶ que la variante de Boudon, l'analyse de dépendance⁷. Je ne veux, en aucune façon, démontrer ici la structure mathématique de ces modèles, d'une part parce que ce

1. James S. Coleman, *Introduction to Mathematical Sociology*, Glencoe (Ill.), The Free Press, 1964.

2. William J. McGill, « Multivariate Information Transmission », *Psychometrika*, vol. 19, 1954, p. 97-116.

3. Henri Theil, « Logit Specifications for the Multivariate Analysis of Qualitative Data », *Report 6935 of the Center for Mathematical Studies in Business and Economics*, The University of Chicago.

4. Edward R. Tufte, « Improving Data Analysis in Political Science », *World Politics*, vol. 21, 1969, p. 641-654.

5. Raymond Boudon, « Les relations causales : problèmes de définition et de vérification », *Revue française de sociologie*, vol. 8, 1967, p. 391.

6. Sewall Wright, « The Method of Path Coefficients », *Annals of Mathematical Statistics*, vol. 5, 1934, p. 161-215; S. Wright, « Path Coefficients and Path Regressions : Alternative or Complementary Concepts », *Biometrics*, vol. 16, 1960, p. 189-202.

7. R. Boudon, « A Method of Linear Causal Analysis : Dependence Analysis », *American Sociological Review*, vol. 30, 1965, p. 365-374; R. Boudon, *l'Analyse mathématique des faits sociaux*, Paris, Plon, 1967, chap. III.

travail a été très bien réalisé dans un nombre suffisant de publications⁸ et, d'autre part, parce que cette structure n'a rien de compliqué si l'on est le moins familier avec le modèle linéaire général⁹ et avec l'algèbre des systèmes d'équations. Il est plus important, je crois, de discuter les modèles d'analyse des cheminements de causalité à la lumière de l'argument avancé plus haut. Ces nouveaux modèles ont des implications sur le plan de la théorie sociologique et de la logique du processus de recherche. Contentons-nous, pour l'instant, de noter qu'il est essentiel de comprendre la formalisation mathématique du *path analysis* si l'on veut s'en servir. Il n'y a pas, à ce jour (et j'espère qu'il n'y aura jamais), de programme d'ordinateur réalisant les calculs de l'analyse des cheminements de causalité comme il en existe pour l'analyse factorielle ou l'analyse de régression. L'application des dernières techniques peut toujours être automatique, celle de la première technique ne saurait l'être. Avant de produire des estimations des coefficients d'association entre les variables du modèle, le *path analysis* pose le modèle avec ses postulats logiques et théoriques ainsi que ses conséquences. Cette portion du travail ne peut être le résultat des manipulations de l'ordinateur, elle dépend essentiellement du chercheur. Vue sous cet angle, l'analyse des cheminements de causalité apporte à la méthodologie sociologique beaucoup plus qu'une nouvelle technique statistique. Si nous avions continué de citer Boudon nous aurions appris que son analyse de dépendance fait passer le processus de recherche empirique du mode déductif au mode inductif. Nous décrirons plus loin en détail comment se traduit ce passage. Auparavant, j'aimerais montrer que, conçue dans cet esprit, l'analyse des cheminements de causalité se situe en parfaite continuité avec les efforts de la méthodologie de la dernière décennie pour ouvrir un dialogue avec la sociologie empirique ou théorique. Ce phénomène m'apparaît important parce qu'il démystifie une partie des modèles d'analyse complexes et leur enlève leur saveur magique et aussi parce qu'il devrait fournir une motivation intéressante au sociologue non méthodologue pour entrer en contact avec les aspects rébarbatifs des nouvelles techniques.

PROPRIÉTÉS COLLECTIVES ET EFFETS DE COMPOSITION

Depuis que la sociologie possède une méthodologie qui lui est propre, soit depuis Durkheim, l'étude des variations concomitantes a porté plus souvent qu'autrement sur des caractéristiques définies sur le plan des individus. Pour sortir du déterminisme psychologique, les sociologues ont tenté, de tout temps, d'insérer

8. Pour le *path analysis*, le lecteur pourra se référer à : O. D. Duncan, « Path Analysis : Sociological Examples », *American Journal of Sociology*, vol. 72, 1966, p. 1-16; K. C. Land, « Principles of Path Analysis », dans E. F. Borgatta (édit.), *Sociological Methodology 1969*, San Francisco, Jossey-Bass, 1969, p. 3-37; C. E. Werts, « Path Analysis : Testimonial of a Proselyte », *American Journal of Sociology*, vol. 73, 1968, p. 509-512; D. R. Heise, « Problems in Path Analysis and Causal Inference », dans E. F. Borgatta (édit.), *op. cit.*, p. 38-73. Pour l'analyse de dépendance, en plus des articles de Boudon déjà cités, on pourra profiter de : F.-A. Isambert, « Traduction mathématique et vérification de quelques systèmes de relations causales », *Revue française de sociologie*, vol. 8, 1967, p. 369-384; A. Flavigny, « Notes sur l'analyse de dépendance », *Revue française de sociologie*, vol. 10, 1969, p. 333-351. De façon générale, l'article suivant intègre les deux approches : J. Feldman-Högaasen, « Analyse de dépendance, régression et corrélation », *Revue française de sociologie*, vol. 10, 1969, p. 318-332.

9. Pour la première fois dans la littérature sociologique, le modèle linéaire général est admirablement présenté par James Fennessey, « The General Linear Model : A New Perspective on Some Familiar Topics », *American Journal of Sociology*, vol. 74, 1968, p. 1-27.

dans leurs analyses des caractéristiques portant sur des « collectifs ». En 1950, cependant, Lazarsfeld et Kendall sentent le besoin de suggérer d'employer des caractéristiques à la fois individuelles et collectives dans l'analyse¹⁰. La présence de facteurs se rapportant à des groupes est un élément essentiel de toute théorie sociologique. La traduction de cette nécessité au niveau de la recherche empirique passe toutefois par des dédales méthodologiques sinueux. Jusqu'ici deux voies ont été surtout explorées : l'« analyse écologique » et l'« analyse des effets de composition ». Voies qu'on commence à peine à exprimer dans un même cadre d'analyse, et encore le rapprochement ne se fait-il que sur le plan des techniques statistiques¹¹.

L'« analyse écologique » consiste à mettre en relation des propriétés de collectifs, généralement d'agrégats géographiques, que ces propriétés soient analytiques, structurelles ou globales suivant les différenciations introduites par Lazarsfeld et Menzel¹². À l'usage, cette approche a dévié et est devenue un substitut. Certains chercheurs ont commencé à utiliser les statistiques administratives portant sur des aires géographiques pour étudier des propositions concernant le comportement des individus là où les données individuelles ne pouvaient être recueillies. L'« analyse écologique » de propriétés analytiques dégénéra en abus, entraînant une mise au point sans équivoque et un peu maladroite de Robinson¹³. La mise à nu et la condamnation de l'*ecological fallacy* furent si tranchantes que Menzel, dans un commentaire sur l'article de Robinson, se vit obligé de rappeler que l'« analyse écologique », même appliquée à des propriétés analytiques, avait un sens sociologique précis et indépendant des inférences sur le comportement individuel¹⁴. Le taux d'analphabétisme d'une division administrative est un indice valable d'une réalité sociale : il mesure d'une certaine façon la productivité d'un système scolaire. Une étude sociologique doit inclure de telles mesures si elle veut dépasser les interprétations psychologiques. Les études des structures urbaines et de la ségrégation résidentielle montrent bien que l'analyse écologique, à partir d'attributs analytiques, produit des connaissances sociologiques d'intérêt. Le fait que certains chercheurs emploient ces mesures mal à propos est donc un problème théorique tout autant que méthodologique. Menzel présente la nécessité de conserver des perspectives collectives dans l'analyse sociologique alors que Robinson dénonce implicitement une pratique répondant à des impératifs matériels de la recherche. La première intervention se situe sur le plan des préoccupations théoriques, la seconde débouche sur des progrès méthodologiques. Goodman et Boudon ont reconnu, quelques années plus tard, que l'*ecological fallacy* était un substitut intéressant à des enquêtes coûteuses au niveau des individus et ont stipulé clairement les conditions selon lesquelles l'emploi des techniques de

10. P. L. Kendall et P. L. Lazarsfeld, « Problems of Survey Analysis », dans R. K. Merton et P. L. Lazarsfeld (édit.), *Continuities in Social Research : Studies in the Scope and Method of « The American Soldier »*, Glencoe (Ill.), The Free Press, 1950, p. 195.

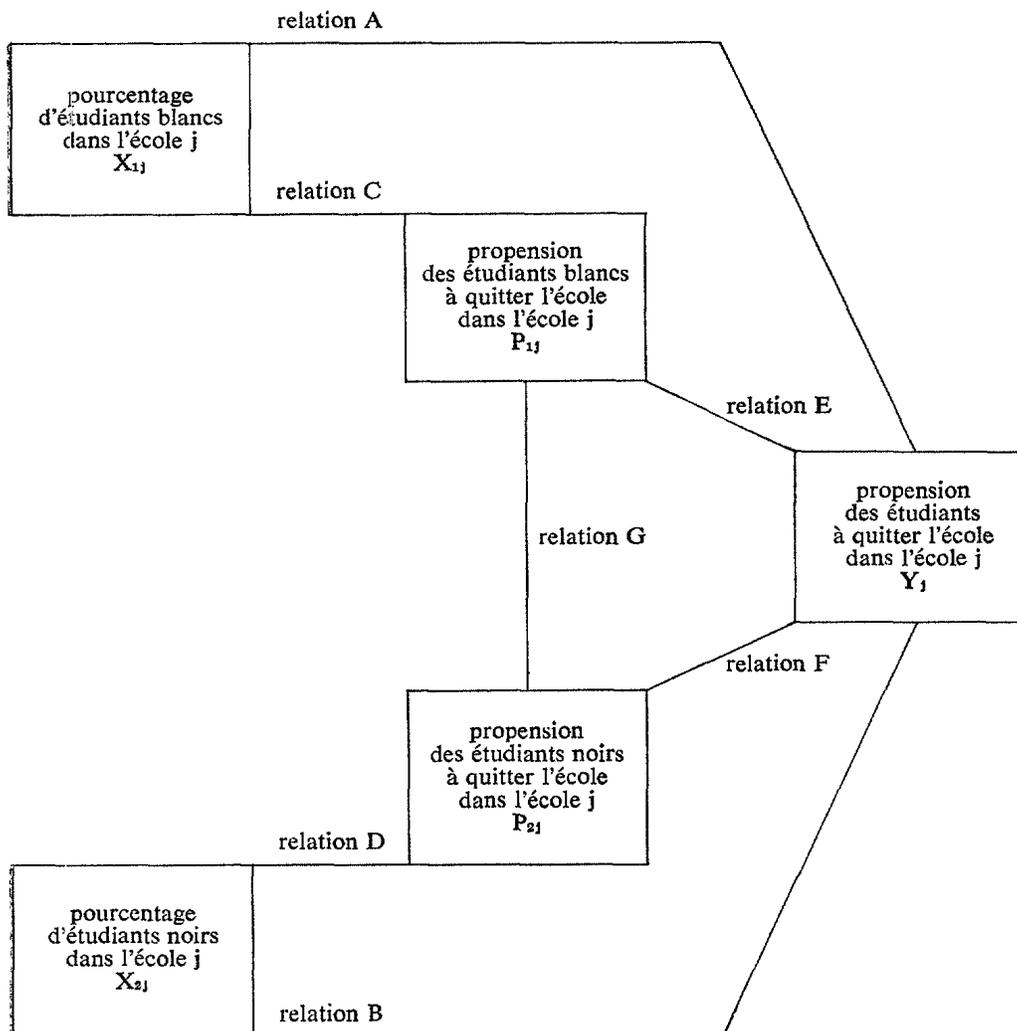
11. Voir K. Schuessler, « Covariance Analysis in Sociological Research », dans E. F. Borgatta (édit.), *op. cit.*, p. 219-244; et aussi J. Fenessey, « The General Linear Model : A New Perspective on Some Familiar Topics », *American Journal of Sociology*, vol. 74, 1968, p. 1-27.

12. P. L. Lazarsfeld et H. Menzel, « On the Relation between Individual and Collective Properties », dans A. Etzioni (édit.), *Complex Organizations*, New York, Holt, Rinehart and Winston, 1961, p. 422-440.

13. W. S. Robinson, « Ecological Correlations and the Behavior of Individuals », *American Sociological Review*, vol. 15, 1950, p. 351-357.

14. H. Menzel, « Comment on Robinson's Ecological Correlations and the Behavior of Individuals », *American Sociological Review*, vol. 15, 1950, p. 674.

GRAPHIQUE 1



corrélation « écologique » pouvait jouer adéquatement ce rôle de substitut¹⁵. Les travaux de Goodman demeurent très intéressants à un niveau de formalisation statistique mais ne font pas le lien avec de nouvelles trouées théoriques aussi systématiquement que ceux de Boudon. Le premier met adéquatement l'accent sur la condition fondamentale de la correspondance entre la corrélation écologique et la corrélation individuelle : le modèle de régression linéaire de variables collectives analytiques reproduit la relation qui existe pour les individus si les propensions des groupes comparés quant au comportement dépendant sont constantes d'un agrégat à un autre.

Pour clarifier ces constatations présentées rapidement, on peut se référer au graphique 1. La corrélation écologique y correspond à la relation A ou à

15. L. Goodman, « Some Alternatives to Ecological Correlation », *American Journal of Sociology*, vol. 64, 1959, p. 610-625; R. Boudon, « Propriétés individuelles et propriétés collectives : un problème d'analyse écologique », *Revue française de sociologie*, vol. 4, 1963, p. 275-299.

la relation B. La corrélation individuelle est représentée par la relation G dans chaque agrégat (chaque école dans le cas du graphique). La relation A ou B reproduit la relation individuelle pour toute la population si la relation G est constante d'un agrégat à l'autre. Le modèle linéaire au niveau des agrégats sera impuissant à reproduire la relation au niveau des individus si les relations C ou D, ou les relations E ou F ont un fondement dans la réalité étudiée. Ces conditions, que Boudon exprime dans des modèles non linéaires qu'il nomme « paraboliques » et « homographiques », ont des résonances théoriques. La première série de conditions (relations C et D) correspond à une variation des propensions spécifiques à chaque groupe étudié (étudiants noirs et blancs) d'un agrégat à l'autre en fonction de l'importance relative du groupe dans l'agrégat. C'est là un phénomène de normes de groupe. Les relations E et F renvoient à une dépendance de la propension de chaque groupe par rapport à la propension globale de l'agrégat. Boudon voit dans cette situation l'équivalent de l'« influence des normes du milieu ». Les notions de « normes de groupe » et de « normes du milieu » ont un contenu proprement théorique. Il est intéressant qu'elles trouvent ici leur expression dans la solution au problème méthodologique de l'emploi des corrélations écologiques pour évaluer des corrélations au niveau des individus. Le chercheur qui, après l'article de Boudon, veut tirer profit de données sur des « collectifs » pour vérifier des hypothèses sur les différences de comportement dues à des caractéristiques individuelles, ne peut plus calculer mécaniquement des taux de corrélation. Il doit envisager plusieurs avenues d'explication théorique et défendre son choix de l'une ou de l'autre. Ce qu'il importe de remarquer ici, c'est l'existence d'une technique d'insertion de facteurs plus « sociologiques » dans l'analyse, technique qui dévie de sa nature initiale pour rejoindre des phénomènes individuels. Dans l'effort pour corriger cette déviation et obtenir quand même des résultats au niveau individuel, les méthodologues posent les questions au niveau théorique et formulent des techniques d'appréhension des concepts concernant l'influence du contexte social. Cependant ces dernières techniques ne peuvent plus être appliquées automatiquement, elles ne peuvent plus être séparées de la discussion théorique du phénomène étudié. Le recours aux éléments théoriques du problème n'est pas rendu nécessaire par la seule absence des données au niveau individuel (les propensions de chaque groupe d'étudiants dans chaque école selon notre graphique); il se pose aussi au niveau de l'analyse des effets de composition.

La démarche d'analyse des effets de composition telle que systématisée par Davis consiste à étudier simultanément les relations C et D du graphique¹⁶. L'approche est cette fois directe : il ne s'agit pas de faire des hypothèses sur la source de variations des propensions spécifiques des groupes comparés. Les P_{1j} étant connus, il est possible de calculer les coefficients de relation entre ces propensions et les indices de composition (les X_{1j}). La technique introduite par Davis a été corrigée par Tannenbaum et Bachman¹⁷. Ces derniers auteurs mentionnent, bien que trop rapidement, le problème de la définition conceptuelle d'un effet de composition parmi les difficultés du modèle d'analyse discuté. Jusqu'à eux, tous les articles touchant l'analyse des effets de contexte avaient

16. James A. Davis, Joe S. Spaeth et Carolyn Huson, « A Technique for Analyzing the Effects of Group Composition », *American Sociological Review*, vol. 26, 1961, p. 215-225.

17. Arnold S. Tannenbaum et Jerald G. Bachman, « Structural versus Individual Effects », *American Journal of Sociology*, vol. 69, 1964, p. 585-595.

donné des définitions opérationnelles des effets présentés, qu'on les ait appelés effets de composition comme Davis ou effets structurels comme Blau¹⁸. Tannenbaum et Bachman font remarquer que même et surtout si la variable contextuelle est analytique (obtenue par une opération arithmétique sur des données mesurant des individus), son contenu théorique doit correspondre à un concept ayant une signification au niveau du contexte social. Quand il parle de normes de groupe, Boudon donne à la variable de composition une interprétation possible faisant référence à une caractéristique d'un contexte social. Il est donc essentiel de spécifier théoriquement ce qu'implique une variable de composition en termes de mécanismes ou de caractéristiques des agrégats sociaux avant de l'employer dans le cadre technique de Davis. Mais même ces mises en garde de Tannenbaum et Bachman ne sont pas suffisantes. Non seulement faut-il préciser théoriquement la signification de la mesure de composition avant de tirer des conclusions d'une analyse contextuelle, mais il faut aussi s'assurer que les interprétations renvoyant au jeu d'autres variables individuelles sont à exclure. Il est toujours possible que la variable de composition utilisée dans le modèle ne soit que le reflet plus ou moins parfait de toutes les autres variables individuelles influençant la fréquence d'apparition du comportement dépendant dans les agrégats. La découverte d'un effet de composition repose toujours sur le postulat que si ce sont des variables individuelles non mesurées qui déterminent véritablement la fréquence du comportement étudié dans les agrégats, celles-ci sont causées par la variable de composition.

Découvrant que la propension des étudiants blancs à quitter l'école est plus faible dans les milieux où les étudiants blancs sont en majorité, on doit se demander si ce n'est pas le niveau socio-économique plus élevé des étudiants de ces écoles qui entraîne leur plus grande persévérance scolaire. Si tel est le cas, l'effet de composition découvert sera réel si on peut défendre théoriquement que c'est la concentration des étudiants blancs dans une école qui entraîne la concentration d'étudiants de statut social plus élevé et non l'inverse. C'est là un problème de contrôle de variables supplémentaires. Si ces variables ne sont pas mesurées dans l'étude et ne peuvent être incluses dans l'analyse empirique, le chercheur doit discuter théoriquement leur impact sur le comportement dépendant par rapport à la mesure de composition. Donc, même si le format technique de l'analyse contextuelle est précisément formulé dans la littérature, ce type d'analyse nécessite une série d'interrogations théoriques de la part du chercheur.

Pour montrer que les développements méthodologiques renvoient toujours à une formulation nouvelle des questions théoriques auxquelles le modèle statistique le plus puissant ne saurait répondre, nous aurions pu parler des problèmes généraux d'interprétation des résultats de l'analyse de régression¹⁹. Nous avons choisi de présenter les cas de l'analyse « écologique » et de l'analyse « contextuelle » parce que ces approches sont plus spécifiques.

18. Peter M. Blau, « Formal Organizations : Dimensions of Analysis », *American Journal of Sociology*, vol. 63, 1957, p. 58-69.

19. Pour une discussion de ces problèmes, voir : Robert A. Gordon, « Issues in Multiple Regression », *American Journal of Sociology*, vol. 73, 1968, p. 592-616.

CHEMINEMENTS DE CAUSALITÉ ET RECHERCHE SCIENTIFIQUE

Nous avons déjà dit que les techniques d'analyse des cheminements de la causalité empruntent leurs fondements statistiques au modèle linéaire général tout comme l'analyse de régression et l'analyse factorielle. Comme ces dernières techniques, les premières sont soumises aux mêmes limitations. Les effets d'interaction demeurent difficiles à inclure dans l'analyse des cheminements de causalité. En principe, seules les relations linéaires peuvent être étudiées, les cas où la non-linéarité peut être attendue devant être transformés pour s'exprimer linéairement. De façon générale, il faut limiter l'analyse aux variables continues, bien que les variables qualitatives dichotomiques puissent être traitées si on les confine au rôle de facteurs indépendants dans le modèle ²⁰.

Ces restrictions posées, passons enfin à ce qui m'apparaît être la grande contribution des nouvelles techniques d'analyse de la causalité en sociologie. Dans les pages qui suivent, je vais essayer de montrer comment ces techniques sont particulièrement appropriées aux études sociologiques qui dépassent le stade exploratoire. L'emploi du *path analysis* est recommandable dans les cas où le chercheur peut discuter les implications d'un ensemble de relations causales en s'appuyant sur les connaissances déjà accumulées dans un secteur spécifique. À la fin de cet exposé, on devrait pouvoir conclure que le *path analysis* ne saurait avoir aucune utilité dans une entreprise de description d'un phénomène sociologique, mais qu'il est l'instrument le mieux adapté à la vérification d'un système théorique de causalité entre plusieurs variables.

Revenons à la question de l'évaluation des effets du milieu scolaire sur des variables de performance académique. Dans un article récent ²¹, Werts discute la méthode employée par Coleman ²² et par nombre de ses collègues en sociologie de l'éducation, pour distinguer l'influence des caractéristiques des étudiants de celle des caractéristiques du milieu scolaire sur le succès académique. Cette méthode se réalise en deux étapes d'analyse de régression. Werts montre comment cette approche sous-estime la portion d'explication due au milieu et repose sur l'hypothèse que ce sont les caractéristiques des étudiants qui déterminent les propriétés du milieu scolaire, alors que l'hypothèse inverse est tout aussi, sinon plus plausible. Il suggère plutôt de traiter le problème dans une seule analyse de régression et de faire une partition de la variance expliquée selon la technique proposée par McNemar ²³, en variance expliquée par a) les caractéristiques des étudiants; b) les caractéristiques du milieu; c) l'action conjointe des deux ensembles de facteurs. Employant cette technique, le problème de la détermination de la direction du lien causal entre les caractéristiques des étudiants et celles de l'école est évité.

Werts fonde la supériorité de la technique qu'il propose sur un ensemble de théorèmes statistiques démontrant que les coefficients de régression atteignent

20. Dans un récent article, Boyle discute les conditions de l'emploi de variables ordinales dans l'analyse des cheminements de causalité : R. P. Boyle, « Path Analysis and Ordinal Data », *American Journal of Sociology*, vol. 75, 1970, p. 461-480.

21. Charles F. Werts, « The Partitioning of Variance in School Effect Studies », *American Educational Research Journal*, vol. 5, 1968, p. 311-318.

22. James S. Coleman *et al.*, *Equality of Educational Opportunity*, Washington (D.C.), U.S. Government Printing Office, 1966.

23. Q. McNemar, *Psychological Statistics*, New York, John Wiley and Sons, 1962.

le statut de « meilleurs estimateurs de la relation » (*unbiased estimates*) quand les équations par lesquelles on les calcule représentent le plus fidèlement le jeu réel d'influence des variables considérées²⁴. Ainsi, on n'obtiendra pas les « meilleurs estimateurs de la relation » quand dans une équation de régression unique se trouvent des variables indépendantes qui sont en réalité les effets d'autres variables indépendantes de l'équation. Dès lors, on ne saurait dire que le problème de l'évaluation précise de l'influence de chaque ensemble de variables sur le phénomène dépendant est complètement résolu par la méthode suggérée par Werts. La méthode employée par Coleman peut reposer sur un postulat acceptable ou inacceptable en ce qui concerne la direction de la relation causale entre les caractéristiques des étudiants et celles du milieu scolaire, mais au moins, elle tient compte de l'ordre causal des variables indépendantes du modèle.

C'est aussi l'ordre causal des variables indépendantes qui sous-tend la controverse entre Sewell et Armer²⁵ d'une part et Turner, Boyle et Michael²⁶ d'autre part, au sujet de l'importance relative des caractéristiques du quartier de résidence dans l'explication des intentions des étudiants du secondaire de poursuivre des études au niveau collégial. Michael résume bien la critique principale que Turner, Boyle et lui-même formulent à l'endroit de l'approche de Sewell et Armer, quand il écrit :

Une différence majeure entre l'étude de Sewell et Armer et les études antérieures est que ces auteurs ne portent aucune attention à l'influence du quartier de résidence sur les aptitudes scolaires. Apparemment, Sewell et Armer ne s'intéressent pas à cette relation parce qu'ils ne visent qu'à prédire les études collégiales, alors que les études antérieures s'attaquaient à l'explication causale de ce phénomène.

Le chercheur qui veut tenir compte de la connaissance de la direction de la causalité entre variables originellement indépendantes, doit remplacer l'équation de régression unique par un système d'équations récursif (où chaque variable dépendante n'est expliquée que par des variables causalement antérieures). Dès qu'une étude porte sur plusieurs variables dans un domaine d'analyse où les premières explorations ont été réalisées, il est souvent possible de postuler un ordre causal entre les variables et de le représenter par un système d'équations récursif. C'est dans ce contexte que l'analyse des cheminements de causalité devient une technique intéressante, soit là où le chercheur peut postuler sur des bases intéressantes un ordre causal entre les variables. Le premier jalon de dialogue entre la théorie et la méthodologie que fournit le *path analysis* se situe à ce niveau, l'emploi de cette méthode d'analyse forçant le sociologue à déterminer un ordre entre ses variables et à le discuter. La réalisation des calculs qui produisent les coefficients de causalité ne peut en aucun cas dicter cet ordre et toute la responsabilité est entre les mains du chercheur. La formalisation mathématique de

24. H. Theil, « Specification Errors and the Estimation of Economic Relationships », *Review of the International Statistical Institute*, vol. 25, 1957, p. 41-51. Pour une présentation sociologique de problèmes statistiques connexes, voir : H. M. Blalock, Jr., « Evaluating the Relative Importance of Variables », *American Sociological Review*, vol. 26, 1961, p. 866-874.

25. W. H. Sewell et A. J. Armer, « Neighborhood Context and College Plans », *American Sociological Review*, vol. 31, 1966, p. 159-168.

26. R. H. Turner, « Communication on Neighborhood Context and College Plans (I) », *American Sociological Review*, vol. 31, 1966, p. 698-702; J. A. Michael, « Communication on Neighborhood Context and College Plans (II) », *American Sociological Review*, vol. 31, 1966, p. 702-706; R. P. Boyle, « Communication on Neighborhood Context and College Plans (III) », *American Sociological Review*, vol. 31, 1966, p. 706-707.

l'analyse des cheminements de causalité devient un instrument économique pour les systèmes complexes de relations entre plusieurs variables nécessitant un ensemble récursif d'équations, système récursif qui n'a de sens que si les variables sont ordonnées.

Un principe important de la méthode scientifique est celui de la parcimonie²⁷. Selon ce principe, même si le réseau de relations entre plusieurs phénomènes est complexe et comporte plusieurs paliers, il faut retrancher de chaque palier d'explication directe tous les facteurs qui ne sont pas essentiels. L'analyse des cheminements de causalité est une méthode qui devrait permettre de serrer de plus près les impératifs logiques de la parcimonie. Un système récursif devient facilement incompréhensible s'il contient plusieurs variables et si toutes les combinaisons de deux de ces variables doivent être considérées. Isambert a montré que l'analyse de dépendance appliquée à des systèmes récursifs complets (où toutes les relations de deux variables sont représentées par un lien causal direct) ne donne rien de plus que des coefficients de régression qui peuvent être obtenus sans la formalisation spéciale de Boudon²⁸. On peut donc avancer l'argument suivant : l'analyse des cheminements de causalité apporte une information supplémentaire qu'on ne peut obtenir par les méthodes classiques de régression multiple dans les cas où la structure causale étudiée ne comporte pas tous les liens possibles entre les variables. Encore ici, la technique d'analyse que nous discutons devrait amener le chercheur à réfléchir sur ses concepts, à faire un tri des relations théoriquement significatives, et à expliquer ses choix les uns par rapport aux autres. Dès lors, nous croyons que les systèmes récursifs incomplets sont plus appropriés à l'emploi du *path analysis* et posent plus clairement le lien nécessaire entre la discussion théorique et la formalisation méthodologique. Cependant, dans les cas où le chercheur s'intéresse au modèle récursif complet, la réflexion théorique reste nécessaire pour établir l'ordre causal des variables. L'analyse des cheminements de causalité devient alors une procédure de décomposition en causalité directe et causalité indirecte de la relation entre deux variables. Les résultats de cette décomposition dépendront de l'ordre causal postulé par le chercheur.

Vient ensuite l'évaluation empirique du modèle construit théoriquement. Pour ce faire, il convient de s'appuyer sur les impératifs statistiques de l'identification auxquels répond la formalisation mathématique du *path analysis*²⁹. Dans un système d'équations, les coefficients statistiques ne peuvent être estimés si quelque combinaison linéaire de deux équations en produit une troisième qui en contredit une quatrième déjà posée dans le système. Pour empêcher cette indétermination des coefficients, on pose des contraintes sur les éléments du système pour le rendre statistiquement identifiable. Les techniques d'analyse des cheminements de causalité sont originales par la réponse qu'elles apportent à ce problème de l'identification. Parmi les contraintes employées jusqu'ici, la plus commune est celle de la spécificité qui stipule que les variables implicites ne sont pas reliées entre elles, non plus qu'avec les variables explicites du système qui leur sont causalement

27. R. B. Braithwaite, *Scientific Explanation*, Cambridge, Cambridge University Press, 1953, p. 67 et s.

28. F.-A. Isambert, « Traduction mathématique et vérification de quelques systèmes de relations causales », *Revue française de sociologie*, vol. 8, 1967, p. 378-379.

29. Pour une discussion du problème de l'identification, voir : H. M. Blalock, Jr., « The Identification Problem and Theory Building : The Case of Status Inconsistency », *American Sociological Review*, vol. 30, 1965, p. 556-564.

antérieures³⁰. Ces hypothèses sur les liens statistiques portant sur les variables implicites sont nécessaires et on les présente habituellement comme des éléments purement statistiques de l'analyse des cheminements de causalité. Cependant, pour certains modèles (spécialement les modèles récurrents incomplets et les modèles impliquant plus d'une variable dépendante), il est possible de calculer les corrélations entre les variables implicites et entre celles-ci et quelques variables explicites, de sorte que le système de relations causales postulé théoriquement reproduise le mieux possible la matrice de corrélations des variables explicites obtenue empiriquement. La considération de ces corrélations impliquant des variables implicites permet au chercheur de discuter les erreurs de son modèle théorique par rapport aux réalités mesurées³¹. Cette façon de procéder est en parfaite continuité avec le rôle de guide dans la réflexion théorique que nous avons attribué jusqu'ici au modèle des cheminements de causalité.

Non seulement les nouvelles techniques d'étude de la causalité sont-elles comme les modèles « écologiques » et « contextuels » (introduits dans la première partie de cet exposé) des éléments qui interagissent avec les aspects théoriques de la recherche sociologique, mais elles permettent aussi de répondre à certaines questions laissées en suspens par les modèles cités. L'analyse écologique de Boudon fait des postulats sur les variations de données inconnues (les propensions spécifiques aux groupes comparés dans les agrégats), l'analyse contextuelle pose *a priori* la dépendance causale des variables individuelles non contrôlées dans le processus empirique par rapport aux variables de composition mesurées. Par son format mathématique, l'analyse des cheminements de causalité permet de transformer les postulats relatifs aux variables non mesurées en hypothèses dont les conséquences peuvent être considérées par rapport au système causal étudié. La réalisation mathématique du *path analysis* reposant sur la combinaison des coefficients de corrélations entre les variables explicites du modèle étudié, il n'est pas nécessaire que toutes ces corrélations soient issues d'une seule et même étude. Il sera parfois très rentable de pousser plus loin la discussion théorique en incluant dans le modèle causal des variables non mesurées dans le matériel analysé mais dont les corrélations avec les variables déjà mesurées sont connues et bien établies dans la littérature sociologique. Les corrélations ajoutées à celles calculées sur le matériel peuvent aussi être établies au sujet de variables qui n'ont jamais été mesurées dans la recherche sociologique antérieure, postulant certains niveaux de relation et discutant les conséquences de ces postulats pour l'ensemble du système causal³².

30. Les deux conditions de l'hypothèse de spécificité sont équivalentes dans certains cas. Cette équivalence est discutée par J. Feldman-Høgaasen, « Analyse de dépendance, régression et corrélation », *Revue française de sociologie*, vol. 10, 1969, p. 318-332. Isambert introduit une hypothèse moins rigide qu'il appelle dissociabilité dans les cas où l'équivalence mentionnée précédemment ne tient pas (soit pour les modèles récurrents incomplets).

31. Pour un exemple d'emploi de cette ressource d'analyse que sont les corrélations portant sur les variables implicites, voir R. W. Hodge et D. J. Treiman, « Social Participation and Social Status », *American Sociological Review*, vol. 33, 1968, p. 728 et s.

32. Cette pratique de recherche souhaitée par Blalock (« Making Causal Inferences for Unmeasured Variables from Correlations among Indicators », *American Journal of Sociology*, vol. 69, 1963, p. 62) est employée dans l'article déjà cité de Hodge et Treiman. D'autres exemples intéressants se trouvent dans P. M. Blau et O. D. Duncan, *The American Occupational Structure*, New York, John Wiley and Sons, 1967, p. 324-328, ainsi que dans l'article de R. M. Hauser, « Schools and the Stratification Process », *American Journal of Sociology*, vol. 74, 1969, p. 600 et s.

Pour résumer, tel que nous l'avons décrit jusqu'ici, l'emploi de l'analyse des cheminements de causalité permet de faire coïncider les processus de recherche empirique avec les caractéristiques formelles des modèles théoriques de causalité. Chapoulie définit quatre conditions nécessaires du modèle théorique de causalité : « 1) la donnée d'un ensemble de variables sociologiques, 2) la donnée d'un ordre strict sur ces variables, 3) la liaison entre cause et effet ... justifiée par une théorie spécifiant l'aspect sémantique de la relation », 4) des hypothèses statistiques sur l'explication des variables explicites par celles qui leur sont causalement antérieures³³. L'analyse des cheminements de causalité permet plus de souplesse par rapport à la première caractéristique. Les données sociologiques du modèle peuvent être autant des indicateurs réels que des concepts théoriques tel qu'expliqué au paragraphe précédent. Nous avons vu comment l'analyse des cheminements de causalité nécessite l'établissement d'un ordre causal des variables tel que stipulé par la seconde caractéristique. En suggérant une préférence pour les modèles récursifs incomplets, nous nous adressons à la troisième caractéristique. Quand la technique des cheminements de causalité force le chercheur à choisir les relations causales théoriquement significatives parmi l'ensemble des relations logiquement possibles, elle l'amène à préciser, dans sa discussion, la nature du processus de causalité soutenant son choix. La quatrième caractéristique est respectée par les hypothèses spéciales au sujet de l'indépendance des variables implicites et par la vérification de ces hypothèses pour certains modèles d'analyse. L'analyse des cheminements de causalité remplit complètement son rôle quand on l'emploie comme langage de formalisation d'un modèle théorique de causalité.

RÉSUMÉ

Le thème de l'article est la convergence entre la réflexion sociologique et les développements des techniques d'analyse quantitative. L'auteur montre comment l'emploi des techniques basées sur le modèle général de régression amène le chercheur à pousser plus loin sa réflexion sur les phénomènes étudiés, chacune des techniques présentées impliquant des postulats quant à la dynamique des variables mesurées ou quant à l'influence de variables non mesurées. Partant de l'analyse écologique et de l'analyse contextuelle, l'auteur introduit l'analyse des cheminements de causalité en montrant la progression de la convergence des techniques et de la réflexion sociologique des premiers modèles au second.

ABSTRACT

[*The Progress of Causality*] The theme of this article is the convergence between sociological thought and technical developments in quantitative analysis. The author shows how the use of techniques based upon the general model of regression leads researchers to carry further their reflexions on studied phenomena, each of the techniques presented implying postulates in regard to the dynamics of the measured variables or concerning the influence of variables which are not measured. Starting from ecological analysis and contextual analysis, the author introduces an analysis of the path followed by causality showing the progression, from preliminary to secondary models, in the convergence between techniques and sociological thought.

RESUMEN

[*Las vías de causalidad*] El artículo trata de la convergencia entre la reflexión sociológica y el desarrollo de las técnicas de análisis cuantitativo. El autor muestra de qué modo el uso de las técnicas basadas sobre el modelo general de regresión conduce al investigador a profundizar su reflexión sobre los fenómenos estudiados, en la medida en que cada una

33. J. M. Chapoulie, « Un type d'explication en sociologie : les systèmes de variables en relations causales », *Revue française de sociologie*, vol. 10, 1969, p. 352-362.

de las técnicas presentadas implican ciertos postulados en cuanto a la dinámica de las variables medidas o en cuanto a la influencia de las variables no medidas. A partir del análisis ecológico y del análisis contextual, el autor introduce el análisis de las vías de causalidad, mostrando la progresión en la convergencia de las técnicas y en la reflexión sociológica que implica el pasaje de los primeros modelos al segundo.