

# Les statistiques peuvent-elles donner une image réelle de la réalité sociale?

## Can Statistics Provide a Real Image of Social Reality?

Raymond BOUDON

Volume 8, numéro 2, octobre 1976

La mobilité sociale : Pour qui? Pour quoi?

URI : <https://id.erudit.org/iderudit/001080ar>

DOI : <https://doi.org/10.7202/001080ar>

[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

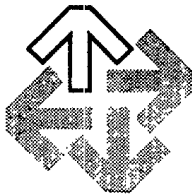
BOUDON, R. (1976). Les statistiques peuvent-elles donner une image réelle de la réalité sociale? *Sociologie et sociétés*, 8(2), 141-156.  
<https://doi.org/10.7202/001080ar>

Résumé de l'article

On peut déceler trois types fondamentaux d'homo sociologicus, conduisant à des sociologies plus ou moins accueillantes à l'égard de la statistique. Les deux premiers types introduisent une vision déterministe de l'homme, même si ce déterminisme est conçu comme probabiliste. Dans le premier cas, on recherche les explications des comportements sociaux dans l'environnement, le plus souvent à l'aide de données et de méthodes statistiques. Dans le second, ce sont plutôt les structures sociales globales qui sont invoquées comme causes; ceci conduit en général à une réticence par rapport aux statistiques. Enfin, un troisième type de modèle donne une place centrale aux notions d'action, de choix, de décision, tout en ne niant pas que leur exercice est soumis à des contraintes; dans ce contexte, les statistiques sont considérées comme un instrument utile de description, qui doit être dépassé par une analyse de l'agrégation des actions. L'auteur illustre ces trois types de modèles à l'aide de nombreux exemples, souvent tirés des recherches sur la mobilité sociale.

---

# Les statistiques peuvent-elles donner une image réelle de la réalité sociale?\*



RAYMOND BOUDON

---

Avant d'aborder mon sujet, je voudrais souligner que le titre donné à cette conférence, s'il est commode, comporte le danger d'une *métaphysique réaliste*. Je veux dire qu'il me paraît inutile et inopportun d'évoquer une réalité sociale « antérieure » à la connaissance que nous pouvons en avoir, indépendante d'elle et que la recherche des sciences sociales viserait à dévoiler. En fait mon propos, même s'il m'amènera à poser certaines questions philosophiques, est plus simple et plus modeste que le suggère son titre. Il vise à apporter quelques éléments d'explication aux difficultés qu'on observe dans le mariage fort ancien entre les sciences sociales d'une part — et en parlant de sciences sociales, je pense ici particulièrement à la sociologie et à la science politique — et la statistique de l'autre.

Si je parle d'un mariage ancien, c'est afin de dissiper la croyance répandue selon laquelle l'invasion de la statistique dans les sciences sociales serait récente et due à l'influence considérable des sciences sociales américaines entre 1955 et 1965. Il est vrai que si on observe l'évolution de la sociologie en France sur le court terme, je veux dire depuis la fin de la deuxième guerre mondiale, on remarque un développement spectaculaire de l'utilisation des statistiques par les politologues et les sociologues. Mais il suffit de prendre davantage de recul pour constater que l'utilisation des statistiques apparaît dès le 18<sup>e</sup> ou même le 17<sup>e</sup> siècle. Des mouvements de pensée comme

---

\*Conférence à l'École Polytechnique, Paris, 23 septembre 1975.

*l'arithmétique politique* dans l'Angleterre du 18<sup>e</sup> siècle, comme la *statistique morale* du 19<sup>e</sup> siècle sont inséparables de l'histoire des sciences sociales. Naturellement, ces mouvements de pensée sont liés à l'intensification des échanges économiques et commerciaux ainsi qu'à la complexité croissante des tâches de l'État. Sans m'étendre sur ces problèmes historiques qu'on trouve traités par Lazarsfeld<sup>1</sup>, je rappellerai seulement l'importance pour l'institutionnalisation de la sociologie d'une œuvre sur laquelle je reviendrai à plusieurs reprises: celle de Durkheim (c'est en effet à partir de Durkheim que la sociologie est reconnue de manière irréversible comme une discipline indépendante et légitime). Or Durkheim, et principalement le Durkheim des *Règles de la méthode sociologique* et du *Suicide*<sup>2</sup> aurait été impossible sans le mouvement de la statistique morale dont il constitue le prolongement et le couronnement. Sans ce mouvement, Durkheim n'aurait visiblement jamais eu l'idée d'analyser le suicide, ce phénomène individuel et humain par excellence, par la méthode des covariations ou des corrélations. Ce faisant, il prolongeait, quoiqu'avec une grande originalité, la longue lignée des chercheurs qui, au cours du 19<sup>e</sup> siècle, de Quételet à Tarde et aux positivistes italiens avaient soumis l'analyse du crime, du suicide, du divorce et de bien d'autres phénomènes à l'observation et à la méthodologie statistique.

Je reviendrai dans un instant sur les difficultés du mariage entre statistique et sciences sociales. Pour le moment, je voudrais simplement noter que celles-ci se devaient de conclure un mariage, sinon d'amour, du moins de raison avec la statistique. Sans doute a-t-on beaucoup discuté la méthodologie statistique appliquée par Durkheim et ses continuateurs à l'analyse du suicide. Peut-être est-elle insuffisante. Mais il est évident qu'un sociologue intéressé par le phénomène du suicide ne peut pas ne pas accueillir avec intérêt des données statistiques montrant par exemple que la fréquence du suicide se modifie régulièrement avec les saisons, avec les jours de la semaine, avec l'âge, qu'elle décroît en temps de guerre, qu'elle oscille avec les fluctuations économiques, ou encore que la fréquence du suicide masculin est stable depuis un demi-siècle en France et que la fréquence du suicide féminin augmente régulièrement<sup>3</sup>. Je laisse de côté pour le moment le problème de l'interprétation de telles données, pour me contenter d'énoncer que, sans elles, le sociologue serait privé de précieuses informations sur le phénomène du suicide. Naturellement, ce que j'avance à propos du suicide serait vrai de tous les ordres de phénomènes qui intéressent les sciences sociales. Ainsi la réflexion des sociologues sur l'inégalité des chances devant l'enseignement serait moins assurée si nous ne disposions pas de données permettant de mesurer, plus ou moins grossièrement, le phénomène, ainsi que ses variations dans le temps et selon les sociétés. J'ai conscience que ces réflexions sont banales. Mais elles me semblent nécessaires étant donné la moue que certains sociologues trouvent séant d'affecter devant les données statistiques. Cette moue est en fait celle d'enfants gâtés. Car on ne peut contester que le développement et le perfectionnement progressifs de l'observation statistique ait augmenté la connaissance des sociétés dans lesquelles nous vivons de manière à la fois considérable et diffuse. Pour le mesurer, il suffit de rappeler l'époque héroïque après tout pas si éloignée puisqu'il s'agit de la fin du 18<sup>e</sup> siècle où, faute de recensements et de données fiables sur la population, la fréquence des épidémies

1. P. Lazarsfeld, « Notes sur l'histoire de la quantification », in *Philosophie des sciences sociales*, Gallimard, Paris, 1970.

2. E. Durkheim, *Le suicide*, 1897, nouvelle édition P.U.F., Paris 1960.

3. Voir « Le suicide », *Population et sociétés*, n° 79, avril 1975, et Durkheim, *op. cit.*

et des disettes donnait à Montesquieu (ce thème apparaît dès les *Lettres Persanes* [Lettres CXII sq.] et se retrouve dans *L'esprit des lois* [Livre XXIII, chap. XXIV]) l'impression que la population du globe décroissait dangereusement, et suggérait à Jean-Jacques Rousseau de faire de la croissance démographique le critère par excellence à partir duquel juger de la qualité des institutions politiques et sociales, ou comme diraient certains de nos contemporains, du bonheur national brut:

« Toutes choses égales d'ailleurs, le gouvernement sous lequel... les citoyens peuplent et multiplient davantage est infailliblement le meilleur: celui sous lequel un peuple diminue et dépérit est le pire. Calculateurs, c'est maintenant votre affaire; comptez, mesurez, comparez. » (Du *Contrat Social*, livre III, chapitre IX).

Cet appel de Rousseau au « calculateur » ou, comme nous dirions, au statisticien apparaîtrait étrange à beaucoup de sociologues et politologues modernes qui dressent volontiers le procès de la méthodologie statistique en sciences sociales. Pourquoi ces réticences à l'égard de la statistique? C'est le problème que je voudrais maintenant tenter de traiter avec quelque détail, en procédant du plus simple au plus complexe, ou, si vous voulez, du problème traditionnel de l'exactitude des statistiques au problème aujourd'hui à la mode de la *pertinence* des relevés et méthodes statistiques pour les sciences sociales.

D'abord, il faut reconnaître que, par opposition aux phénomènes qu'on considère comme relevant de la démographie (nombre des citoyens, nombre des décès, des naissances, des divorces), les phénomènes relevant de la sociologie sont en général plus difficile à définir et, partant, à dénombrer. Je dis bien en général, car il n'est pas plus difficile de dénombrer les voix qui se sont portées sur un parti donné à l'occasion d'une consultation électorale que les naissances ou les décès. Mais il en va autrement si on s'intéresse par exemple au suicide ou au crime. Considérons le crime. On rencontre à propos des statistiques concernant ce phénomène deux difficultés considérables que, sous des modalités variables et avec des intensités différentes, on retrouve à peu près partout. La première est qu'on ne connaît jamais le nombre exact des crimes et délits, mais seulement ceux qui sont détectés par l'appareil répressif. C'est le fameux problème du chiffre noir de la criminalité. On doit d'ailleurs remarquer que les données recueillies ne sont pas affectées seulement par l'efficacité de l'appareil répressif. Toutes sortes d'autres facteurs viennent modifier la probabilité pour qu'un acte qu'on désirerait enregistrer soit effectivement enregistré. Ainsi, lorsque les vols de faible importance sont nombreux et que la police est incapable ou se considère incapable de les traiter, les victimes peuvent être tentées de considérer qu'il est inutile de porter plainte. De sorte qu'appareils d'observation et phénomènes observés se trouvent agir et réagir les uns sur les autres. La seconde difficulté est que la classification des crimes et délits à partir de laquelle s'effectue l'enregistrement statistique des actes délictueux et criminels peut apparaître inadéquate au sociologue, qui peut considérer par exemple comme hétérogène une catégorie de crimes ou délits telle qu'elle est définie par le code pénal. Mais on n'imagine mal que les classifications utilisées par la statistique criminelle puisse être indépendante des classifications juridiques. Avec une acuité et sous des formes différentes, le même problème se retrouve pratiquement à propos de tous les types de données statistiques: même si on peut chercher à tenir compte dans l'élaboration de l'appareil statistique des « besoins » scientifiques des sciences sociales, il reste que cet appareil doit satisfaire aux besoins administratifs et, naturellement, qu'il est soumis à des contraintes de coût.

Il est intéressant, dans ce contexte, de s'arrêter un instant sur la polémique instructive qu'a soulevée depuis plusieurs décennies l'utilisation par Durkheim de données statistiques dans l'analyse du suicide. Dès 1932, Halbwachs<sup>4</sup>, un élève de Durkheim, publiait un livre intitulé *Les causes du suicide*. Le livre était largement consacré à une critique de la méthodologie de Durkheim. Plus près de nous, en 1967, un sociologue américain, Jack Douglas<sup>5</sup> allait, dans *The Meanings of Suicide* beaucoup plus loin encore dans la critique. Tout récemment, en France, Jean Baechler<sup>6</sup> dans son livre important sur *Les suicides* reprenait et élargissait la critique de Douglas. Les objections adressées à Durkheim, que je me contente de résumer rapidement, concernent d'abord le problème de la définition du suicide: la population des suicides enregistrés par la statistique comporte des suicides qui n'en sont pas, des suicides accidentels, comme celui de cette jeune fille qui désire simuler un suicide pour exercer un chantage, avec le propos délibéré de se rater et qui, par accident, se tue. Faut-il compter ce décès comme un suicide? Réciproquement, des suicides authentiques ne sont pas dénombrés comme tels, les suicides lents par intoxication par exemple, ou les suicides dissimulés en accidents. Bref, on ne peut espérer que le nombre des suicides donné par la statistique corresponde à une définition homogène, précise et adéquate du suicide. Plus généralement, l'enregistrement des suicides n'est pas un processus socialement neutre. Ainsi, les appareils d'observation statistique n'enregistrent pas les tentatives de suicide, car celles-ci n'ont pas d'existence juridique. Mais l'étude scientifique du suicide peut difficilement se dispenser de considérer et en tout cas a tout intérêt à considérer simultanément les suicides réussis et les tentatives. D'un autre côté, l'émotion suscitée par un suicide et par conséquent la probabilité pour que celui-ci soit enregistré comme un accident varient par exemple avec le degré auquel, culturellement, l'acte suicidaire est considéré comme tabou. C'est en s'appuyant sur des remarques de ce genre que certains critiques de Durkheim ont contesté par exemple le résultat, qui tient une grande place dans son livre sur le *Suicide*, selon lequel les protestants se suicident davantage que les catholiques d'après les statistiques. Durkheim interprétait ce résultat qui lui paraissait congruent avec beaucoup d'autres en le mettant en relation avec le caractère plus autoritaire de la religion catholique qui, soumettant l'individu à des normes extérieures échappant à la critique de son libre-arbitre, le dissuaderait plus efficacement d'attenter à son existence. Les critiques de Durkheim rétorquent que la différence entre protestants et catholiques dans la fréquence des suicides résulte probablement simplement de ce que le suicide est plus sévèrement condamné par l'opinion dans les pays de tradition catholique, et par conséquent plus fréquemment dissimulé. Ainsi, les différences observées seraient un simple artefact résultant de différences dans les appareils d'enregistrement.

Ces critiques sont-elles fondées? En partie sans aucun doute. Ainsi Douglas a pu montrer que des variations dans le temps des taux de suicide observés par Durkheim dans tel ou tel pays étaient effectivement dues à des modifications survenues au niveau de l'appareil d'enregistrement, comme lorsque le soin de recueillir les données de suicide a été, dans la Prusse du 19<sup>e</sup> siècle, retiré au clergé pour être confié à l'administration. Mais la critique a généralement été excessive: à travers Durkheim, c'est en fait la méthodologie statistique dans son ensemble que des critiques comme Douglas voulaient

4. M. Halbwachs, *Les causes du suicide*, Alcan, Paris, 1930.

5. J. Douglas, *The Meanings of Suicide*, Princeton University Press, Princeton, 1967.

6. J. Baechler, *Les suicides*, Calmann-Lévy, Paris, 1975.

atteindre. Ainsi ont-ils eu tendance à jeter l'enfant avec l'eau du bain. Car si Durkheim a parfois sous-estimé l'interférence entre l'appareil d'enregistrement et les données enregistrées, un nombre important de résultats statistiques sur le suicide résistent aux types de critique que je viens de résumer: même si le nombre des suicides enregistrés par les appareils statistiques est toujours inexact, il est hors de doute par exemple que le suicide augmente avec l'âge, ou que le suicide réussi est plus fréquent chez les hommes, etc. etc. Bref, on peut montrer qu'un certain nombre de relations statistiques concernant le suicide sont valides.

Pour résumer cette discussion, je dirai que, dans de nombreux cas, l'observation statistique est plus difficile lorsqu'il s'agit de phénomènes relevant de la sociologie ou de la science politique que lorsqu'il s'agit de phénomènes démographiques ou, évidemment, de phénomènes relevant des sciences de la nature. Il est plus difficile de définir la notion de suicide que celle de naissance. D'un autre côté, l'appareil d'enregistrement, sauf cas limite, n'est pas « neutre »: le suicide n'est pas vécu de la même manière en milieu protestant et en milieu catholique. Les femmes n'utilisent pas les mêmes modes de suicide que les hommes. Ces circonstances et maints autres facteurs affectent la probabilité pour qu'un suicide soit enregistré comme tel. En bref, l'appareil d'enregistrement apparaît dans ce cas de suicide, comme dans la plupart des cas, comme une composante du système social dont il permet l'observation, mais qui se trouve aussi affectée par lui.

À toutes ces difficultés, il faut de surcroît ajouter le coefficient personnel qu'introduit, non pas cette fois l'appareil d'enregistrement, mais l'utilisateur, sociologue ou politologue, des données fournies par l'appareil d'enregistrement statistique. Pour illustrer ce point, prenons un exemple classique. Un des thèmes favoris des sociologues est celui de la mobilité sociale. Il s'agit de savoir, entre autres choses, dans quelle mesure les enfants tendent ou non à hériter de la position sociale de leurs parents. Pour répondre à ce type de questions, les sociologues utilisent ce qu'ils appellent des tableaux de mobilité entre générations, c'est-à-dire des tableaux à double entrée comportant verticalement et horizontalement une liste de types de positions sociales (par exemple: manœuvres, ouvriers qualifiés, employés, cadres moyens, etc.) et donnant les flux intergénérationnels entre les types de profession. Pour construire ces tableaux, ils utilisent en général des données recueillies sur de vastes échantillons nationaux. Naturellement le problème est de s'entendre sur la liste des types de positions sociales: il est évident, en effet, que selon les positions considérées comme distinctes sont plus ou moins fines, la mobilité apparaîtra comme plus ou moins grande. Or si on compare par exemple les études américaines de mobilité aux études françaises, on remarque aussitôt que les Américains<sup>7</sup> ont tendance à utiliser des listes fines. Ainsi, ils considèrent généralement deux catégories socio-professionnelles comme distinctes à partir du moment où elles correspondent à des différences dans les revenus moyens, dans le niveau d'instruction moyen ou dans le prestige tel qu'il est établi par des enquêtes d'opinion. En revanche, les chercheurs français ont tendance à adopter des catégories plus larges et, par exemple, à englober toutes les catégories de travailleurs manuels en un seul type. Naturellement, il en résulte que les premiers voient dans leurs résultats plus de mobilité que les seconds. Pourquoi cette différence *d'optique*? Il s'agit là d'une question compliquée. Mais on peut affirmer que cette différence provient, en partie au moins,

---

7. P. Blau et O. Duncan. *The American Occupational Structure*, Wiley, New York, 1967.

du fait que la notion de classe sociale et plus généralement la tradition marxiste a des racines beaucoup moins profondes chez les intellectuels américains que chez les intellectuels français. De façon curieuse, les phénomènes de stratification sociale sont traditionnellement perçus en France à partir de la conceptualisation marxienne et en particulier du concept de classe, tandis que la manière dont les Américains conçoivent les distinctions sociales apparaît comme traditionnellement influencée par la conceptualisation et les conceptions d'un sociologue français, A. de Tocqueville. Pour Tocqueville, la centralisation politique, l'individualisme, la sacralisation des valeurs égalitaires, l'importance pour le système capitaliste de la compétition, de la réussite, du *statut*, tous ces traits typiques des sociétés modernes étaient incompatibles avec des sentiments prononcés d'appartenance de classe<sup>8</sup>. Peut-être ces différences proviennent-elles aussi de ce que les symboles distinguant les classes sociales sont moins marqués aux États-Unis.

Il résulte de tout ceci que l'utilisation des statistiques dans les sciences sociales pose au chercheur des problèmes de « distanciation » ou, comme dirait Piaget, de « décentration » considérables. J'ajoute entre parenthèses que, paradoxalement, c'est peut-être lorsqu'il est le plus maître de ses données, lorsqu'il les recueille lui-même par enquête sur échantillon à partir de questions qu'il a formulées lui-même, que la distanciation est la plus difficile pour le chercheur. Je crois que ces difficultés de la distanciation expliquent pour une part le rejet de méthodes statistiques par certains sociologues et, plus généralement, l'illusion selon laquelle la connaissance des sociétés serait d'une nature différente de la connaissance de la nature et obéiraient à d'autres principes. En réalité, les difficultés que j'ai tentées d'illustrer et de résumer par la notion de *distanciation* peuvent généralement être contournées. Ainsi, on peut étudier la mobilité sociale en utilisant des classifications différentes des positions sociales et comparer les résultats. On peut acquérir la certitude que, même si les données du suicide sont toutes inexactes, la différence entre les fréquences de suicide caractérisant telle ou telle catégorie sociale a bien le signe donné par les statistiques. Grâce aux critiques de Durkheim nous savons distinguer ceux de ses résultats qui sont valides et ceux qui sont sujets à caution. Bref, même si les problèmes de définition, les problèmes posés par l'équation personnelle du chercheur, par le fait que les instruments d'observation ne sont pas socialement neutres dans la mesure où ils visent à répondre à des « besoins » de la gestion de sociétés complexes, même si tous ces problèmes sont considérables, ils ne sont pas en général insurmontables.

J'en viens maintenant à la critique de *pertinence* adressée par certains sociologues modernes à l'utilisation des données et méthodes statistiques dans les sciences sociales. Nous oublierons à ce point les problèmes posés par l'adéquation entre l'appareil d'observation et le phénomène qu'on désire étudier pour nous situer en aval, au niveau de l'analyse. À ce niveau, certains sociologues utilisent couramment les méthodes statistiques usuelles: analyse de régression, analyse de variance, etc. qui sont aussi utilisées par exemple en agronomie ou en économétrie. À première vue, cette pratique paraît légitime et naturelle. Elle est même si naturelle que Durkheim qui, au moment où il écrit le *Suicide*, dans les dernières années du 19<sup>e</sup> siècle, ignorait tout des notions statistiques élémentaires, les a appliquées sans les connaître<sup>9</sup>. Tout le *Suicide* repose en

8. Sur l'opposition entre « classes » et « statut » et plus généralement pour une introduction aux concepts fondamentaux de la sociologie, voir R. Nisbet, *The Sociological Tradition*, Basic Books, New York, 1966 et en français H. Mendras, *Éléments de sociologie*, 2<sup>e</sup> édition, Colin, Paris, 1975.

9. H. Selvin, « Aspects méthodologiques du suicide », in R. Boudon et P. Lazarsfeld (eds), *l'Analyse empirique de la causalité*, Mouton, Paris, 1966, pp. 276-291.

effet sur une analyse des variations des taux de suicide avec l'âge, avec le sexe, avec l'environnement (rural ou urbain) avec les cycles d'affaires, etc. Il s'est demandé si la relation entre la fréquence du suicide et l'âge était la même pour les deux sexes, si elle variait en fonction de l'environnement urbain ou rural. Bref, sans le dire et sans le savoir, Durkheim a utilisé dans son étude des formes frustes et intuitives de ce que nous appelons aujourd'hui l'analyse de variance ou l'analyse de régression. L'exploitation qu'il présente des matériaux statistiques dont il disposait montre même qu'il avait la notion de concepts statistiques complexes, comme ceux auxquels nous donnons aujourd'hui le nom d'effets d'interaction du premier, du deuxième ou du troisième ordre.

Ultérieurement, les sociologues manipulant des données statistiques se sont évidemment mis à employer couramment les méthodes mises au point par les statisticiens. Ainsi, si vous avez la curiosité de prendre connaissance des études de mobilité sociale conduites depuis une dizaine d'années dans de nombreux pays<sup>10</sup>, vous observerez que ces études suivent fréquemment le même patron. Elles consistent à utiliser des données (données de recensement, ou données recueillies sur de larges échantillons) permettant d'établir à propos d'une population un certain nombre d'informations concernant l'origine sociale de chaque individu (c'est-à-dire la position sociale de son père), le niveau d'instruction du père, le niveau d'instruction de l'individu considéré, sa position sociale, sa profession, son revenu, etc. en début de carrière, et/ou à une étape ultérieure de sa carrière. Cela fait, on applique à ces données les méthodes statistiques usuelles. Ainsi, on utilise souvent l'analyse de régression à équations multiples. On détermine ainsi le coefficient de régression partiel du niveau d'instruction du fils sur le niveau d'instruction et sur la position sociale du père, puis le coefficient de régression partiel de la position sociale du fils sur son niveau d'instruction, etc. On comprend l'intérêt de ce genre d'analyse. Il permet par exemple de mesurer l'influence des origines sociales sur le niveau d'instruction ou celle du niveau d'instruction sur la position sociale. Il permet de se demander dans quelle mesure ces diverses influences varient au cours du temps; dans quelle mesure elles diffèrent d'une société à l'autre. Bref, ce type d'analyse paraît instructif et légitime. Il est largement employé. Il a été employé par des pionniers de la sociologie, alors même que ces méthodes statistiques en étaient au début de leur développement. Malgré cela, il est contesté par de nombreux sociologues. Pourquoi? On peut donner à cette question, à mon sens, une réponse générale et des réponses particulières. Je veux dire par là que l'utilisation de ces méthodes statistiques en sociologie pose une question générale indépendante de la nature particulière des phénomènes étudiés et des questions particulières dans certains contextes d'application.

La question générale est la suivante. Les méthodes statistiques usuelles reviennent à tenter de rendre compte des données observées par un modèle de forme  $y = a_1 x_1 + a_2 x_2 + \dots + a_n x_n + e$  (je néglige le fait que ce modèle puisse contenir plusieurs équations de ce genre, ou que ces équations puissent être plus complexes). Ainsi le *Suicide* de Durkheim peut être mis sous la forme d'une série d'équations de ce genre. Le problème qui se pose alors est celui de l'interprétation d'une telle équation. Lorsque les variables  $x_1, x_2, \dots, x_n$  et  $y$  désignent des phénomènes naturels, l'interprétation est immédiate. Ainsi si  $x_1$  désigne la quantité de pluie,  $y$  la quantité de récoltes

10. W. Müller et K. U. Mayer (eds), *Social Stratification and Career Mobility*, Mouton, Paris/La Haye, 1973; A. P. M. Coxon et C. L. Jones, *Social Mobility*, Penguin Books, Londres, 1975; voir aussi Blau et Duncan, *op. cit.*



et que le coefficient  $a_1$  est positif, on conclut que la pluie exerce une influence sur la quantité des récoltes et la recherche peut s'arrêter, au moins provisoirement. En revanche, lorsque  $y$  désigne le suicide ou le crime ou le vote pour un parti et que  $x_1$  désigne par un exemple la position sociale ou la classe sociale, le fait que  $a_1$  soit petit ou grand, positif, négatif ou nul ne nous apprend pas grand'chose. Bref, il peut être intéressant d'observer par exemple que le suicide varie en fonction de l'âge ou de la classe sociale. Mais il y a intérêt à considérer cette observation comme un point de départ et non comme un point d'arrivée. La sociologie commence non pas lorsqu'on constate que le suicide varie avec l'âge, mais lorsqu'on est en mesure d'expliquer pourquoi il varie avec l'âge. Tandis que l'agronome peut se contenter d'observer que la pluie ou un certain type d'engrais influence la quantité de la récolte. De même, la sociologie commence, non pas lorsqu'on observe qu'il existe une relation entre origine sociale et niveau d'instruction, mais lorsqu'on est capable d'expliquer pourquoi il en est ainsi. Pour résumer mes remarques sur ce point, je dirais que les méthodes statistiques permettent d'établir des faits naturellement indispensables pour le progrès de la recherche, mais qu'elles ne fournissent pas l'explication, le *pourquoi* de ces faits. En bref, elles ont une portée plutôt *descriptive* qu'*explicative*. Par contraste, dans les applications de ces méthodes à des phénomènes relevant des sciences de la nature, description et explication *peuvent* coïncider, comme lorsqu'on constate que tel caractère génotypique a une influence sur un caractère phénotypique.

Derrière ces remarques se profile une discussion compliquée et pas toujours explicite qui sépare les sociologues en deux camps. Les uns pensent que les phénomènes sociaux peuvent être étudiés exactement comme des phénomènes naturels, qu'il suffit d'observer par exemple que l'origine sociale a une influence sur le niveau d'instruction, de mesurer cette influence aussi exactement que possible, et d'étudier les variations de cette influence en fonction d'autres variables de manière aussi complète et fine que possible. Ainsi ces sociologues considèrent que l'explication de la relation entre origines sociales et niveau scolaire peut être, sinon entièrement donnée, du moins facilitée si on est capable de déterminer les variables intermédiaires responsables de la relation, par exemple en mesurant le poids respectif sur la réussite scolaire des différences dans l'environnement culturel ou dans les conditions économiques qui résultent des différences dans les origines sociales. Les autres pensent que, même si on était capable d'analyser l'influence des origines sociales sur le niveau d'instruction en fonction de toutes les variables possibles, on n'en serait encore qu'au seuil de l'explication du phénomène. Pour cerner de façon peut-être plus claire l'opposition entre les deux écoles, on peut dire que pour les premiers, expliquer un phénomène c'est connaître de manière aussi complète que possible les variables qui influencent ce phénomène et la structure de ces influences, tandis que pour les seconds, une telle analyse ne fournit pas une *explication* du phénomène étudié pour la raison que la notion d'*influence* (ou de cause) est une notion en elle-même toujours obscure lorsqu'il s'agit de phénomènes humains et qui ne peut dans le meilleur des cas que représenter un dangereux raccourci. Comme vous le voyez, la discussion nous conduit tout droit à des questions philosophiques qu'il est impossible d'éviter. Pour les premiers, l'homme peut être assimilé à une entité sujette à des influences, dont les caractéristiques peuvent être expliquées par l'action sur lui de facteurs qui lui sont extérieurs. Pour les autres, on ne peut faire abstraction du fait que l'homme est libre, capable d'*actions*, de choix et de décisions et que, par conséquent, ses comportements ne sauraient être réduits à des réactions plus ou moins mécaniques à des influences extérieures que par une analogie forcée et par conséquent inadmissible.

Permettez-moi d'insister un instant sur ce point essentiel. Il me paraît incontestable qu'en sociologie c'est aujourd'hui la première *philosophie* qui est la plus représentée, et de loin. C'est-à-dire que la plupart des sociologues conçoivent l'homme comme le siège d'influences extérieures à lui et contribuant à déterminer ses comportements.

Mais à ce point, on doit introduire une nouvelle distinction d'une importance capitale: lorsque cette philosophie *déterministe* de l'homme s'accompagne d'une culture que pour des raisons de brièveté je qualifierai de *behavioriste*, comme c'est le cas par exemple aux États-Unis, le terrain est favorable à une application massive des techniques statistiques: le prototype de la recherche consiste en effet alors à identifier les différentes variables caractérisant l'environnement social de l'individu et à déterminer le poids de ces variables. Lorsque la culture behavioriste est absente ou traditionnellement contestée, comme c'est le cas généralement en France par exemple, et que les comportements humains sont considérés comme *déterminés* non pas par l'environnement, mais par les *structures sociales*, le terrain est évidemment immédiatement moins favorable à l'application des méthodes statistiques. D'abord parce que ces *structures sociales* ne sont pas toujours définies de manière assez claire pour donner lieu à une manipulation statistique. Ensuite parce que leur variance interindividuelle est souvent conçue comme nulle. C'est le cas par exemple des études qui présentent les structures de la société de consommation comme responsables d'effets identiques sur le comportement de tous. Il est intéressant à cet égard de considérer les critiques opposées à l'emploi des méthodes statistiques par les sociologues français. Très souvent on observe qu'elles se ramènent en fait à une critique de l'explication du comportement social par des variables d'environnement. Ainsi, les sociologues français de l'éducation<sup>11</sup> qui s'intéressent aux inégalités scolaires considèrent souvent d'un œil dubitatif les travaux de leurs collègues américains parce que ces derniers essaient de déterminer les facteurs qui, dans l'environnement familial, scolaire et social des individus, expliquent que la réussite scolaire ou les ambitions scolaires varient en fonction des origines sociales. Or pour beaucoup de sociologues français ces variations sont explicables par des *structures* qui ne se réduisent pas aux variables d'environnement des Américains, et qui ne donnent guère prise à l'emploi des méthodes statistiques. Cela explique que ces sociologues se contentent de prélever les données statistiques indiquant l'existence d'inégalités, mais quittent le terrain statistique au niveau de l'analyse.

Pour résumer on peut donc, en simplifiant, déceler trois types fondamentaux d'*homo sociologicus* conduisant à des sociologies plus ou moins accueillante à l'égard de la statistique. Les deux premiers types introduisent une vision déterministe de l'homme, même si ce déterminisme est conçu comme probabiliste. Mais il existe une différence considérable entre ces deux types. En effet, le premier localise le déterminisme au niveau de l'environnement, tandis que le second voit dans les comportements sociaux le produit de structures sociales globales, définies au niveau des sociétés dans leur ensemble. Le troisième type de modèle donne au contraire une place centrale à la notion d'action, de choix, de décision, ce qui ne veut pas dire évidemment que l'*homo sociologicus* soit alors conçu comme ayant la maîtrise ou le contrôle des conditions, situations et contraintes dans lesquelles s'exerce son action, ni que ses décisions prennent nécessairement les formes algorithmiques de la théorie mathématique de la décision.

---

11. P. Bourdieu et J. C. Passeron, *La Reproduction*, Minit, Paris, 1970. C. Baudelot et R. Establet, *L'école capitaliste en France*, Maspero, Paris, 1971.

Personnellement, j'ai tendance à trouver beaucoup plus fondamentales, beaucoup plus claires et beaucoup plus intéressantes les objections à l'emploi des méthodes statistiques de ceux qui considèrent que le sociologue ne peut faire abstraction de la capacité humaine d'action, de délibération, de choix et de décision pour expliquer les comportements humains. Pour ces sociologues, les méthodes statistiques ont une utilité descriptive, mais doivent être dépassées dans la mesure où une proposition du type «  $x$  a une influence sur  $y$  » n'est jamais claire et ne peut jamais être totalement clarifiée en multipliant les variables intermédiaires entre  $x$  et  $y$  pour la raison que, s'agissant de l'homme, la notion d'influence est elle-même obscure et incompatible avec les notions d'action, de décision et finalement de liberté.

Pour tenter de fixer les idées sur ces distinctions, permettez-moi de prendre un exemple beaucoup plus simple que celui du crime, du suicide ou des comportements scolaires, l'exemple des accidents d'automobile. Envisageons d'abord le cas du sociologue appartenant à la première des trois catégories que j'ai distinguées, je veux dire du sociologue estimant que le comportement humain doit être expliqué essentiellement par l'action de variables caractérisant l'environnement de l'individu. Très naturellement, ce sociologue fera un inventaire des variables qui lui paraissent susceptibles d'être statistiquement liées avec la probabilité d'occurrence d'un accident (âge du conducteur, type d'éducation reçu, densité de la circulation, etc. etc.) et analysera par les méthodes statistiques usuelles le poids respectif de ces variables. Considérons maintenant un sociologue appartenant à la deuxième de nos catégories, celui qui considère les comportements humains comme déterminés par les structures sociales: ce sociologue cherchera peut-être à expliquer les accidents de la route par le caractère sacré de l'automobile dans ce qu'on appelait il y a quelques années la société de consommation, par le fait qu'elle est devenue un moyen d'affirmer sa puissance et sa supériorité plutôt qu'un moyen de transport. Il est clair que ce type de théorie est difficilement justiciable de l'analyse statistique. Tournons-nous maintenant vers le troisième sociologue, celui qui conçoit l'homme non comme le produit de son environnement ou des structures sociales, mais comme un être *actif*, dont les comportements ne sont compréhensibles que si l'on connaît les objectifs auxquels il vise par ces comportements et la représentation qu'il se fait des moyens qui peuvent lui permettre d'atteindre ses objectifs. Notre troisième sociologue considérera sans doute avec intérêt les résultats statistiques obtenus par le premier, mais son analyse commencera là où celle du premier s'achève. Imaginons par exemple que, toutes choses égales d'ailleurs, les accidents apparaissent comme plus fréquents sur les routes à trois voies que sur les routes à deux voies. Le premier sociologue se contentera de cette observation et conclura que certains types de structures routières sont causes d'accident. Le troisième tentera d'aller au-delà et d'expliquer en termes d'action le *pourquoi* de la corrélation statistique. Ainsi, il suggérera peut-être que les routes à trois voies créent une structure d'interaction suggérant et en tout cas permettant aux automobilistes de jouer au célèbre jeu du poulet qui, avec le dilemme des prisonniers et quelques autres jeux moins connus, constitue un des cas de figure tétralogiques de la théorie des jeux: en effet, la route à trois voies incite chacun des deux automobilistes se faisant face qui se sont engagés dans la voie centrale à essayer de « gagner », c'est-à-dire à espérer que l'autre réintégrera le premier sa file d'origine, renonçant provisoirement au dépassement qu'il comptait effectuer.

Il ne serait pas difficile de montrer que les trois types de sociologie que je viens de distinguer apparaissent à propos de sujets considérés comme plus nobles scientifi-

quement que celui des accidents d'automobile. J'ai abondamment cité l'exemple du suicide. La sociologie de type I caractérise de nombreux travaux contemporains, comme l'analyse statistique présentée par Henry et Short<sup>12</sup> de la relation entre les cycles du suicide et les cycles d'affaires. Dans son récent livre sur les *Suicides*, J. Baechler représente de son côté le troisième type de sociologie: il s'efforce en effet de rendre compte des relations que la statistique observe à propos du suicide à partir d'une analyse des structures fondamentales d'interaction dans lesquelles le suicide peut apparaître comme un choix. Ainsi, les statistiques montrent que, bien que les suicides masculins soient plus nombreux que les suicides féminins, les tentatives de suicide sont beaucoup plus fréquentes chez les femmes: il cherche à expliquer ce fait en montrant que, pour des raisons complexes dues à la structure de la famille et à la division sexuelle du travail, les femmes se trouvent beaucoup plus fréquemment dans des situations où la tentative de suicide est un moyen stratégique efficace pour atteindre certains objectifs. Quant à Durkheim lui-même, il peut, selon la lecture qu'on en fait, être classé soit dans le type II, soit dans le type III.

La même classification pourrait être utilisée à propos d'un autre champ de recherche que j'ai plusieurs fois mentionné, celui des inégalités scolaires, de l'héritage social ou de la mobilité sociale.

Les sociologues de type I étudient l'influence des variables d'environnement sur la réussite scolaire, puis sociale. Les sociologues de type II s'efforcent de montrer comment, de manière plus ou moins complexe, les structures sociales et singulièrement la structure des classes sociales conduisent les individus ou leurs familles à adopter des comportements ayant la conséquence de perpétuer les inégalités et les phénomènes d'héritage social. D'autres enfin, que je classerai dans le type III, dans lequel j'inclurai si vous le permettez le travail que j'ai récemment consacré à ce sujet<sup>13</sup>, s'efforcent d'expliquer les relations que la statistique observe par exemple entre origines sociales et niveau scolaire, entre niveau scolaire et position sociale ou entre origine sociale et position sociale comme la conséquence de l'agrégation de décisions rationnelles effectuées par les individus en fonction de leur position sociale actuelle et dans le contexte de structures scolaires et sociales données (entendant par décisions « rationnelles » des décisions dont la logique peut être aisément saisie par un observateur plutôt que les décisions algorithmiques de type théorie des jeux).

Je crois que les trois types de sociologie que je viens de distinguer permettent de mieux comprendre ce que j'appelais au début de cet exposé les difficultés du mariage entre les sciences sociales et la statistique et, singulièrement, la contestation moderne de la « pertinence » des statistiques. Les sociologues de type I sont parfaitement ouverts à la statistique (je veux dire à la fois à l'idée que les données puissent prendre principalement la forme de données statistiques et que les méthodes statistiques soient l'instrument exclusif d'analyse de ces données). Les sociologues de type II sont beaucoup plus réticents à l'égard de la statistique dans la mesure où elle leur paraît liée à un environnementalisme qu'ils rejettent (la notion d'environnement s'opposant ici à celle de structure sociale et non, évidemment à celle d'hérédité). Les sociologues de type III tendent à considérer les méthodes statistiques comme des méthodes utiles d'un point de vue descriptif, mais refusent de voir dans le résultat d'une analyse

---

12. A. Henry et J. Short, *Suicide and Homicide*, Free Press, Glencoe, 1954.

13. R. Boudon, *L'inégalité des chances*, Colin, Paris, 1973.

de régression ou de variance, si complète soit-elle, le point final de la recherche: derrière une relation statistique, il leur importe de déceler une agrégation d'*actions* dont la logique puisse être mise en évidence.

Comme vous le voyez, le problème de la relation entre les méthodes statistiques et les sciences sociales est un problème complexe et qui renvoie en dernière analyse à la représentation que se donne le chercheur de l'*homo sociologicus*.

Avant d'abandonner ces questions, je voudrais encore introduire une remarque d'ordre historique: la critique de la « pertinence » des statistiques est quelquefois perçue comme le dernier cri de la méthodologie. Mais il faut voir qu'elle est aussi ancienne que la critique de l'exactitude des statistiques en matière de sciences sociales. On sait par exemple que, dans la première moitié du 19<sup>e</sup> siècle, Auguste Comte, polytechnicien et parrain de la sociologie, se livrait à de véhémentes attaques contre l'application par certains de ses contemporains du calcul des probabilités et de la statistique aux phénomènes sociaux. Or il n'est pas difficile de voir que cette attitude anti-statistique s'accompagne chez Comte d'un modèle d'*homo sociologicus* analogue à celui de ce que j'ai appelé les sociologues de type II: pour Comte, comme pour beaucoup de sociologues modernes, l'*homo sociologicus* est conçu comme un être passif, comme le produit de forces et de structures sociales situées au niveau des sociétés globales.

Je voudrais maintenant, pour finir, redescendre de ces hauteurs philosophiques pour souligner un dernier point qui me paraît important. Il me semble que les déceptions des sociologues ou des politologues devant les résultats de certaines analyses statistiques proviennent de l'usage mécanique qui tend à être fait des méthodes statistiques. Ainsi, ayant observé un certain nombre de variables explicatives  $x_1, x_2, \dots, x_n$  et une variable  $y$  qu'il cherche à expliquer, l'analyste a tendance, pour rendre compte de la dépendance de  $y$  par rapport à  $x_1, x_2, \dots, x_n$ , à recourir mécaniquement à des équations de régression, le plus souvent linéaires, quelquefois polynomiales. Or, dans certains cas — je quitte ici le terrain général sur lequel je me suis situé dans la partie précédente de mon exposé pour aborder des situations particulières — il se trouve qu'on possède des informations suffisantes sur la réalité étudiée pour pouvoir donner à la relation entre la variable à expliquer et les variables explicatives une forme plus utile et plus éclairante. Permettez-moi de me borner ici à un exemple emprunté de nouveau à l'analyse de la mobilité sociale.

Comme je l'ai indiqué plus haut, de nombreux chercheurs se consacrant à ce domaine utilisent un modèle de recherche qui consiste à analyser la régression de la position sociale acquise par un individu sur son niveau d'instruction, le niveau culturel de sa famille, le niveau social de sa famille, etc. Bref, il s'agit d'analyser un nombre de variables explicatives aussi grand que possible de manière à rendre compte d'une proportion aussi grande que possible de la variance de la variable finale, à savoir la position sociale des individus constituant la population observée. Or quand on se livre à ce genre d'analyse, dans n'importe quel contexte national, on ne parvient jamais à rendre compte que d'une proportion relativement modeste de la variance de la variable finale. Parmi les nombreuses études de ce type dont j'ai pu prendre connaissance, je n'en ai pas trouvé une seule qui parvienne à frôler les 50%. Autrement dit, le niveau culturel de la famille, le niveau socio-professionnel de la famille, le niveau d'instruction du sujet et toutes les autres variables qu'on peut imaginer en y incluant éventuelle-

ment des variables d'attitude, ne parviennent jamais, prises ensemble, à expliquer davantage qu'une proportion inférieure à 50% de la variance de la position sociale des individus observés. Pourquoi? L'interprétation statistique mécanique de ce résultat est que les plus de 50% inexplicés de la variance de la variable finale traduisent l'effet de ce que les statisticiens appellent des variables aléatoires. Une interprétation de ce genre a été donnée par certains sociologues. Ainsi, le sociologue américain Jencks<sup>14</sup> interprète les 60% de variance inexplicée qu'il observe à partir de ses données comme traduisant les effets de variables telles que la « chance », le piston, les accidents de santé... Mais il faut reconnaître que cette interprétation a laissé la plupart de ses collègues sceptiques. Il est peu vraisemblable que la « chance » exerce une influence plus grande dans la détermination du destin social individuel que les origines sociales, la réussite scolaire, le niveau d'instruction et les autres variables de ce type prises simultanément. En fait, la clé du mystère me paraît résider ailleurs: dans le fait que, malgré son apparence innocente, l'analyse de régression cherche dans ce cas à faire porter aux données un habit qui leur convient mal.

		Position sociale		
		+	-	
Niveau d'instruction	+	20	20	40
	-	0	60	60
		20	80	

Pour illustrer mon raisonnement, je supposerai que nous travaillons sur une société très simple où les individus ne pourraient accéder qu'à deux types (élevé/bas) de positions sociales et à deux types (élevé/bas) de niveaux d'instruction. Imaginons alors que notre société soit aussi méritocratique que possible, c'est-à-dire que les positions sociales de niveau élevé soient toujours attribuées dans la mesure du possible aux individus ayant le niveau d'instruction élevé. Pour fixer les idées, supposons alors qu'une cohorte de 100 jeunes gens arrivent sur le marché des positions sociales, 40 d'entre eux avec un niveau d'instruction élevé et 60 avec un niveau d'instruction bas, et que le marché des positions sociales leur offre 20 positions de niveau élevé et 80 de niveau bas. Dans ce cas, le coefficient de régression de la position sociale sur le niveau d'instruction ne peut-être, comme le montre le tableau ci-dessous, supérieur à 0,50. En effet, 20 individus au maximum peuvent alors recevoir une position élevée sur les 40 candidats. Le coefficient de régression maximum est donc égal à  $(20/40) - (0/60) = 0.50$ . De sorte que, dans ce cas fictif, le niveau d'instruction explique une proportion très inférieure à 50% de la variance de la variable à expliquer, à savoir la position sociale des individus. Cela signifie-t-il que le niveau social soit faiblement affecté par le niveau d'instruction et fortement par la « chance »? Non. La faiblesse du coefficient de régression est due au déséquilibre structurel entre la distribution des niveaux d'instruction et la distribution des positions sociales. Cet exemple illustre donc un cas où l'application mécanique aux données du modèle de régression classique de forme  $y = ax + e$ , bien que « tentante » et d'une certaine manière « naturelle », aboutit à des résultats confus, en particulier parce qu'elle conduit à mettre sur le compte d'erreurs

14. C. Jencks, *Inequality*, Basic Books, New York, 1972, et R. Boudon, « La sociologie des inégalités dans l'impassé? », in *Analyse et Prévision*, vol. 17, Janvier 1974, p. 83-95.

aléatoires ce qui relève en fait de discordances structurelles. Plus grave encore, elle engage le chercheur à tenter interminablement d'identifier les variables supplémentaires qui pourraient améliorer la proportion de la variance expliquée.

La proposition générale que je souhaitais illustrer par cet exemple, c'est qu'une autre raison de la réticence des sociologues devant les méthodes statistiques est probablement une conséquence de leur « facilité » c'est-à-dire de leur apparente généralité d'emploi. On peut toujours soumettre n'importe quel type de données de forme convenable à une analyse de régression, à une analyse de variance ou à une analyse factorielle. Mais le problème difficile est celui de la significativité des résultats. Dans certains cas, comme dans celui de l'exemple précédent l'analyse statistique peut contribuer à obscurcir plutôt qu'à l'éclairer.

Au fond la question à laquelle on aboutit est la question banale du bon usage des statistiques. Les méthodes statistiques peuvent conduire à des résultats obscurs qui ne font pas beaucoup avancer notre compréhension des phénomènes étudiés. Et ce qui est vrai des modèles statistiques est naturellement vrai des données de la statistique descriptive. Ce point est si banal que j'hésite à le mentionner. Mais le mauvais usage des statistiques descriptives est si répandu qu'il n'est peut-être pas inutile d'en donner un exemple: on constate couramment qu'on se contente pour apprécier l'évolution des inégalités de revenu dans le temps de comparer les coefficients de Gini caractéristiques de deux ou trois périodes ou, plus simplement, de comparer l'évolution de la proportion des revenus globaux détenus par les 5% (ou les 10%, ou les 20%, etc.) les plus favorisés et les 5% (ou les 10%, ou les 20%, etc.) les plus défavorisés. En fait, cette comparaison peut n'avoir aucun sens: s'il se trouve par exemple que, entre  $t_1$  et  $t_2$  la proportion des jeunes de 25 à 30 ans et des vieux de plus de 65 ans a augmenté par rapport à celle des personnes de 30 à 65 ans il peut en résulter une augmentation du coefficient d'inégalité de Gini, même si les différences de salaire entre les catégories d'emploi manifestent dans le même intervalle de temps une tendance à se rapprocher, cela résultant de ce qu'une proportion plus importante de personnes sont observées soit en début de carrière soit dans leur période de retraite. De même, il est clair que l'interprétation d'un coefficient de Gini varie avec l'intensité et le sens des différences de fécondité entre classes sociales: imaginons que les 10% les moins favorisés possèdent 3% de la masse globale des revenus, leur situation et le degré d'inégalité sont différents selon qu'ils doivent subvenir avec leurs revenus à la subsistance et à la formation de 10% ou de 20% de la population des enfants d'âge scolaire. Naturellement, pyramide des âges et fécondité ne sont pas les seuls facteurs à « surveiller » avant de passer à l'interprétation de l'évolution d'un coefficient de Gini. Il en est d'autres. Ce qui n'empêche pas les commentateurs de n'en considérer souvent aucun, notamment lorsque l'évolution observée est congruente avec leurs croyances.

Autre exemple: la littérature sur l'inégalité des chances insiste souvent sur des propositions telles que « à origines sociales plus élevées, niveau d'instruction plus élevé » ou « à niveau d'instruction plus élevé, revenus plus élevés ». Mais elle néglige très fréquemment la considération des dispersions pour ne retenir que celle des moyennes. Socialement et par conséquent sociologiquement, les dispersions sont cruciales, cependant. Si elles sont élevées, cela signifie que même si les moyennes de revenu croissent régulièrement avec le niveau d'instruction, pour un individu particulier, un

niveau d'instruction plus élevé peut être associé à une probabilité importante de revenu plus bas. Pour des raisons compréhensibles, ce point est généralement négligé.

Les statistiques peuvent-elles donner une image réelle de la réalité sociale? Il n'y a pas pour nous de réalité sociale, indépendamment de la connaissance que nous en avons. Il n'y a pas de connaissance de la réalité sociale sans instruments d'observation. Pas d'instrument d'observation qui puisse être tenu pour un miroir fidèle et non déformant. Comme tous les autres instruments d'observation, les statistiques donnent de la réalité sociale une image dont l'intérêt pour la compréhension des sociétés dépend de la manière dont elles sont conçues et utilisées. Rien ne me paraît plus absurde que d'opposer une sociologie quantitative et statistique à une sociologie qualitative et non statistique comme les ténèbres à la lumière (ou comme la lumière aux ténèbres); que d'opposer une prétendue méthode « ethnologique » qui atteindrait l'essence même des objets qu'elle étudie, aux abstractions et aux artifices de la statistique. Les difficultés de la distanciation ne sont pas moins grandes dans un cas que dans l'autre. Dans les deux cas, imagination et esprit critique sont les conditions nécessaires du succès. Dans les deux cas, le modèle que le chercheur se fait de l'*homo sociologicus* interfère avec ses observations et, en même temps, lui permet de donner une signification à ces observations qui, autrement, se réduiraient à des sensations désordonnées (je laisse ici de côté la question, qui supposerait à elle seule un débat, de savoir lequel de ces modèles d'*homo sociologicus* est préférable). La statistique n'est pas davantage, en soi et indépendamment de la manière dont données et méthodes statistiques sont utilisées, une garantie d'objectivité et de scientificité que les méthodes qualitatives sont une garantie d'accès direct à la « réalité » sociale. Mais il faut aussi ajouter que lorsqu'on traite de phénomènes situés à l'échelle de sociétés complexes, qu'on ne peut embrasser du regard, la méthode « ethnologique » est peu applicable, sauf à introduire le postulat douteux selon lequel les microcosmes accessibles à l'observation directe sont la réplique fidèle du macrosme auquel ils appartiennent. Dans ce cas, l'information statistique joue un rôle irremplaçable.

#### RÉSUMÉ

On peut déceler trois types fondamentaux d'*homo sociologicus*, conduisant à des sociologies plus ou moins accueillantes à l'égard de la statistique. Les deux premiers types introduisent une vision déterministe de l'homme, même si ce déterminisme est conçu comme probabiliste. Dans le premier cas, on recherche les explications des comportements sociaux dans l'environnement, le plus souvent à l'aide de données et de méthodes statistiques. Dans le second, ce sont plutôt les structures sociales globales qui sont invoquées comme causes; ceci conduit en général à une réticence par rapport aux statistiques. Enfin, un troisième type de modèle donne une place centrale aux notions d'action, de choix, de décision, tout en ne niant pas que leur exercice est soumis à des contraintes; dans ce contexte, les statistiques sont considérées comme un instrument utile de description, qui doit être dépassé par une analyse de l'agrégation des actions. L'auteur illustre ces trois types de modèles à l'aide de nombreux exemples, souvent tirés des recherches sur la mobilité sociale.

#### ABSTRACT

One can discern three fundamental types of *homo sociologicus*, leading to sociological practices more or less welcoming to the use of statistics. The first two types introduce a deterministic vision of man, even when this determinism is seen as probabilistic. In the first case, explanations of social behaviour are sought in the environment, mostly with the help of statistical data and methods. In the second case, it is rather



global social structures which are called upon as causes; this generally leads to a reticence with regard to the use of statistics. Finally, a third model holds as central the notions of action, choice, decision, while not ignoring that their exercise is subject to constraints; in this context, statistics are considered a useful instrument of description, which has to be supplemented by an analysis of the aggregation of actions. The author illustrates these three types of models with numerous examples, often drawn from research on social mobility.

#### RESUMEN

Se pueden descubrir tres tipos fundamentales de « homo sociologicus », que conducen a sociologías más o menos acogedoras con respecto a la estadística. Los dos primeros tipos intrducen una visión determinista del hombre, aún cuando ese determinismo sea concebido como probabilístico. En el primer caso, a menudo se buscan explicaciones de comportamientos sociales en el medio ambiente, con la ayuda de datos y de métodos estadísticos. En el segundo, son más bien las estructuras sociales globales que son invocadas como causas; esto conduce en general a una reticencia en relación a las estadísticas. En fin, un tercer tipo de modelo da un lugar central a las nociones de acción, de elección, de decisión, sin negar que su utilización está limitada; en este contexto, las estadísticas son consideradas como un instrumento útil de descripción, que debe ser depasado por un análisis de agregación de las acciones. El autor ilustra estos tres tipos de modelo con la ayuda de numerosos ejemplos sacados de investigaciones sobre movilidad social.