

Alain DESROSIÈRES

DU SINGULIER AU GÉNÉRAL

L'argument statistique entre la science et l'État

Les débats sur la pertinence et la portée des arguments statistiques se déroulent en général sur deux scènes distinctes, celle de la société et celle des sciences sociales. Dans le premier cas, une « information statistique » est produite par des instituts spécialisés. Puis elle circule, toujours inscrite, d'une manière ou d'une autre, dans un cercle plus vaste, sans début ni fin logique, qui la relie à une action ou à un jugement sur l'action. Dans le second cas, en revanche, la dimension cognitive de l'argument est seule mise en avant. La science se présente comme une fin en soi, une accumulation de savoir se justifiant elle-même. L'information statistique utilise des acquis de la science. En retour, elle lui fournit des matériaux et des outils. Mais elle s'en distingue, par ses modes de production et par ses usages. Les instituts de statistique ne sont pas des centres de recherche, même si leurs productions sont utilisées par ceux-ci. Pourtant, le rapprochement et la comparaison entre les controverses qui agitent ces deux scènes en apparence différentes sont instructifs. Certaines questions sont identiques. Quel statut de réalité faut-il attribuer aux objets statistiques ? Comment passer du singulier au général, puis opérer le mouvement inverse ? Comment l'argument statistique est-il articulé avec le langage ordinaire ?

Statistique et effets de réalité

La recherche historique a été transformée, dans la première moitié du siècle, par des méthodes statistiques. Les cycles longs des fluctuations économiques, les permanences de longue durée des relations entre les groupes sociaux prenaient alors le pas sur le foisonnement des histoires individuelles (fussent-

elles royales), des batailles et des traités. Puis, depuis les années 1970, par un retour du balancier, la micro-histoire a retrouvé les charmes du singulier et des trajectoires personnelles. Les grandes totalisations sont désormais dénoncées comme des artifices. Cette opposition cantonne l'argument statistique du côté du général et de l'universel, caractéristiques des sciences de la nature.

Ainsi, dans un article publié en 1980, l'historien italien Carlo Ginzburg oppose deux grandes façons d'explorer et d'interpréter les phénomènes sociaux. Il les nomme respectivement le paradigme *galiléen* et le paradigme de l'*indice*. Le premier est inspiré par les sciences de la nature, la physique et l'astronomie de Galilée. Il anime la science moderne, conquérante de l'universel, appuyée sur des formalisations mathématiques ou sur des totalisations statistiques. Le second paradigme, celui de l'indice, recherche au contraire le détail peu visible, la trace presque effacée, le *symptôme* révélateur d'une réalité cachée. Ce sont, par exemple, le lapsus pour Freud, l'empreinte digitale pour le détective, le pli obscur d'une robe pour l'expert en peinture. La statistique, avec ses additions et ses moyennes, pourrait sembler tout entière du côté du paradigme galiléen, celui de la généralité et des grands nombres. C'est bien ainsi que cet article a été lu le plus souvent, en confondant notamment la *mathématisation* des sciences sociales et leur *quantification*, qui constituent deux démarches historiquement distinctes, l'une hypothético-déductive, l'autre inductive.

Cette opposition apparente entre deux paradigmes, l'un statistique et l'autre indiciaire, est pourtant trompeuse. Comme le détective, le statisticien étudie des *indices*, à partir de ses enquêtes ou de ses fichiers, pour décrire l'inflation, la production ou le chômage. Il fait continuellement le va-et-vient entre des informations individuelles et des informations « agrégées », qui reflètent une réalité plus vaste. De même que le géographe passe du relevé cadastral élémentaire à la représentation d'un continent en modifiant l'échelle de sa carte, le statisticien passe du budget des recettes et dépenses d'une famille au tableau des consommations d'un pays, en choisissant le taux de sondage de son enquête. En manœuvrant la focale de leurs instruments d'observation, le géographe et le statisticien semblent pouvoir se promener entre le plus petit et le plus grand, entre le détail de Sherlock Holmes et le monde complet et sans faille de Galilée. De ce point de vue, le grand et le petit ne semblent plus opposer deux démarches radicalement différentes. Dans les deux cas, il s'agit de faire tenir une histoire plausible, mémorisable et réutilisable. L'échelle importe peu. L'argument « indiciaire » cherche à montrer l'invisible, à construire un récit, en rendant réels des personnages, qui peuvent être aussi bien le coupable d'un crime, pour le policier, qu'une variable latente (telle que la « propension au crime »), pour le sociologue statisticien.

L'étude du rôle de l'argument statistique, dans le débat social comme dans les sciences sociales, ne peut esquiver la question du *statut de réalité* de ses objets et de leurs mesures. Celui-ci oscille entre trois pôles, qui constituent trois façons de parler et de prendre appui sur un nombre. Un réalisme métrologique direct, inspiré des sciences de la nature, postule implicitement des objets sociaux aussi réels, définis et mesurables que le sont la circonférence de la terre ou l'intensité d'un courant électrique. Un réalisme symptomal indirect, plutôt issu des sciences de la vie et du corps humain, considère les mesures comme des signes, des reflets, des indices de réalités supposées mais inatteignables : la *propension* au crime ou au suicide de Quételet et Durkheim, l'*intelligence générale* de Spearman, les *variables latentes* des sociologues anglo-saxons sont typiquement utilisées dans des rhétoriques de ce type. Enfin, un conventionnalisme constructiviste, influencé par le droit ou les sciences politiques, met l'accent sur les conventions taxinomiques et sur les opérations de codage préalables aux mesures. Dans ce cas, la réalité des objets est liée à l'ampleur et à la solidité de l'investissement social dont résultent ces taxinomies et ces codages. Le fait d'utiliser telle ou telle de ces trois façons de dire, dans le feu de l'argumentation scientifique ou du débat social, résulte moins de choix épistémologiques que de contraintes de situation. Ceci vaut autant dans le débat social que dans le débat scientifique (Gollac, 1997).

Certains États latino-américains (le Brésil, le Mexique) réunissent dans un même institut la fabrication des statistiques et celle des cartes géographiques, comme si, en France, l'INSEE et l'IGN étaient fusionnés. Ce rapprochement incite à comparer le caractère réaliste de l'espace social du statisticien et de l'espace géographique du cartographe. Il est désormais courant de considérer le premier comme historique et construit. Les schémas de la représentation statistique d'un pays dépendent largement des structures politiques, sociales, juridiques, culturelles de l'État, de l'administration, et plus généralement, de toute la société, comme l'ont montré maints travaux d'histoire comparée des statistiques de différentes nations (Desrosières, 1993). Mais cette critique classique du réalisme photographique de la statistique est pertinente aussi pour les cartes, dont la texture et le contenu varient pour les mêmes raisons sociales. Par ailleurs, la structure démographique de la population, selon le sexe et l'âge, est souvent présentée comme une toile de fond de l'analyse sociale, comparable au « fond de carte » des fleuves et des montagnes du géographe. Cette vue un peu naïve ignore les difficultés pratiques des démographes à identifier et enregistrer certaines catégories de la population. Dans tous les cas, le rapprochement de ces deux constructions a l'intérêt de relancer la question de leur statut. Sont-elles des réalités, des fictions, ou des constructions sociales bien utiles, notamment pour l'action et le débat démocratique ?

Cette utilité n'est vérifiée que si ces constructions *tiennent bien*, si elles peuvent être transportées et combinées sans altérations trop fortes, et acceptées comme référence commune par la société dans son ensemble. Cette condition est souvent qualifiée de fiabilité. Elle doit être entendue non pas comme une ressemblance fidèle à la réalité, mais comme une consistance et une robustesse suffisantes pour supporter sans dommages des épreuves variées, transports, combinaisons, critiques des équivalences conventionnelles ou des frontières. Cette fiabilité est inextricablement technique et sociale. Elle suppose de longues chaînes d'enregistrements, de calculs et de mises en forme, qui, pris ensemble, suscitent ou non la confiance de la société. Dès lors que cette confiance est acquise, ces statistiques et ces cartes sont qualifiées d'informations, avec ce que ce mot passe-partout suppose de transparence et d'évidence, par suppression et oubli du long travail de leur construction. De ce point de vue, une information statistique est une production routinisée d'une chaîne d'enregistrement robuste, acceptée ou non comme point d'appui pour le débat et pour l'action. L'*indice des prix* et le *taux de chômage* constituent deux cas bien connus de telles informations. Leurs constructions et leurs usages sont bien documentés, dans la mesure où ces deux statistiques ont fait l'objet de maintes controverses, depuis une cinquantaine d'années. Dans les deux cas, peut être retrouvée toute la gamme des rhétoriques évoquées ci-dessus : réaliste directe (le « vrai chiffre du chômage »), réaliste indirecte (en termes d'« indicateur, parmi d'autres, d'un phénomène beaucoup plus complexe »), et enfin conventionnaliste constructiviste (en cas de controverse et de dénonciation).

Dans cette perspective, l'opposition de Ginzburg entre l'argument « galiléen » et l'argument « indiciaire » n'est plus aussi tranchée. Dans les deux cas il s'agit de *faire tenir*, de *mettre en forme* (en latin : *in-formare*) des éléments singuliers, des détails, des symptômes, *a priori* sans rapports entre eux. Le détective comme le statisticien rapprochent, confrontent, combinent, construisent des espaces d'équivalence et de comparabilité. Les enquêtes, les fichiers, les indices, les informations qu'ils fournissent au juge ou à l'acteur social contribuent à des buts analogues : faire tenir leurs arguments et leurs décisions, leur permettre de *clôre*, au moins provisoirement, la chaîne du doute et de l'incertitude. Cette façon de voir, qui gomme la différence entre les cas apparemment « petits » (ceux du détective) et « grands » (ceux du statisticien), n'est pas propre à la statistique. Elle est présente dans d'autres catégories d'arguments, notamment ceux qui visent à laisser des traces écrites. Goody (1979) a souligné, dans son analyse de la « raison graphique », l'importance du *tableau croisé*, permis par les deux dimensions de la feuille de papier écrite, mais beaucoup plus difficile à suggérer dans une culture orale. La forme tabulaire est bien sûr essentielle dans l'argument statistique.

Statistique et construction de l'État

Ce point de vue a le mérite d'attirer l'attention sur ce qui rapproche les manières d'argumenter et de prouver du juge, du savant ou du responsable politique. Il a été développé par la sociologie moderne des sciences et des techniques, de Bruno Latour (1989), Ian Hacking (1990) ou Donald McKenzie (1981). Très stimulant, il présente néanmoins un inconvénient, grave dans le cas de l'argument *statistique*. Il méconnaît la spécificité de la *construction de l'État*, à laquelle la statistique (comme d'ailleurs la cartographie) est étroitement associée. Le mot lui-même l'indique : la *Statistik* allemande du xv^e siècle était une description organisée de l'État, directement destinée au Prince. De nos jours encore, la « fiabilité », ou plutôt la légitimité des statistiques publiques, est fortement liée à *la fois* à deux autorités en général disjointes, celle de la science et celle de l'État. Combinaison tout à fait spécifique entre ces deux façons de faire tenir des choses, la statistique publique relève à la fois de la sociologie de la science et de celle des institutions et de l'État, c'est-à-dire de deux compétences et de deux formes de regard rarement réunies. Elle risque toujours de basculer d'un côté ou de l'autre, soit vers des débats sur le statut *scientifique* de cette statistique, soit vers des analyses de son rôle *institutionnel*, indépendamment de son contenu, et sans établir de lien entre ces deux dimensions.

Cette dichotomie classique peut être dépassée. L'histoire de la statistique peut être lue dans la perspective d'une « histoire concrète de l'abstraction », selon l'expression de l'historien Jean-Claude Perrot (1992). L'évolution conjointe du rôle de l'État et de ses technologies cognitives les plus matérielles offre un fil conducteur pour lire l'histoire de la statistique. On y retrouve par exemple une distinction centrale entre les activités de l'État visant à traiter des cas *singuliers* (par exemple les tribunaux), et celles qui organisent des politiques *générales*, valables pour la collectivité entière. La statistique a toujours été à la charnière entre ces deux volets de l'activité de l'État, en combinant continuellement le paradigme de l'indice et le paradigme galiléen, l'identification du cas singulier et la totalisation nationale. Le va-et-vient et la tension entre des traitements particuliers et des traitements généraux constituent le moteur, politique et cognitif, de la construction de la statistique, et ils en fournissent la matière première de base.

Au commencement était la *liste*. Symbole de l'activité de l'État, on la retrouve dans les plus anciennes traces écrites, chinoise, sumérienne ou égyptienne. L'*état* (avec un petit é) est l'outil de base de l'*État* (avec un grand É). La statistique ancienne est un comptage opéré sur ces listes. Les registres paroissiaux, ancêtres de l'*état civil*, sont des listes de baptêmes, de mariages et de décès, rendues obligatoires par des édits royaux, au xv^e siècle. Dès le

XVII^e siècle, les « arithméticiens politiques » anglais, Graunt et Petty, comptent les cas enregistrés. Ils transforment ces listes en nombres, en agrégats, utiles pour le Prince et pour les commerçants avisés. L'activité administrative d'enregistrement d'événements singuliers a été prolongée ou détournée vers la production d'une *information*, un *nombre résumant une liste*. La statistique était née. Avant de compter, il avait fallu édicter un règlement et des formes standardisées d'enregistrement obligatoire, puis identifier, transcrire et rassembler

Mais, à l'inverse, c'est l'évolution du rôle et des fonctions de l'État qui explique les transformations de la statistique publique. À chaque époque, des questions différentes ont été jugées sociales, et incorporées dans les responsabilités de l'État, suivant des modalités variables selon les pays. Pour les pays européens, on peut citer quelques exemples de telles questions. Jusqu'au XVIII^e siècle, la statistique royale était liée à la levée des armées et des impôts. Au XIX^e siècle, elle traite surtout de la pauvreté, des épidémies et de la santé publique. Puis, des années 1890 à 1930 : de l'organisation du travail salarié et de la protection des travailleurs. Des années 1940 à 1970 : du pilotage keynésien des politiques macroéconomiques. Enfin, depuis les années 1980 : du traitement des conséquences sociales de la crise, de la décentralisation des États, de l'unification européenne. À chacune de ces étapes correspondent des outils statistiques différents : listes exhaustives, calculs de moyennes, enquêtes par sondage, modèles économétriques, dont on peut retracer brièvement l'histoire.

Les recensements de la population du début du XIX^e siècle sont d'abord des listes des familles et des individus qui les composent, avant de fournir des tableaux de chiffres. Ces tableaux sont construits selon les découpages administratifs du territoire. Le recensement est une carte de la population. Dans les années 1830, l'Angleterre est quadrillée par des bureaux locaux, chargés à la fois d'enregistrer les actes de l'état civil, de gérer l'assistance aux pauvres, et de suivre, en cas d'épidémie, la morbidité et la mortalité, presque rue par rue. L'ensemble de ces bureaux est coiffé par le *General Register Office*, dirigé par un médecin, William Farr. La légitimité de cette statistique est enracinée dans une action *locale* de prévention et d'assistance. Le territoire est le support de l'agrégation statistique. Le danger de la contagion épidémique justifie l'intervention publique, mais celle-ci est d'abord locale.

En marge de la statistique administrative exhaustive issue des recensements et des activités des bureaux, une forme d'enquête directe se développe : la monographie locale. Les questions d'hygiène, de délinquance et de pauvreté sont souvent discutées et traitées par des initiatives privées de réformateurs sociaux, médecins, hommes de loi, ingénieurs ou professeurs. En France, Frédéric Le Play symbolise ce courant, catholique, plutôt hostile à l'État, en par-

ticulier à l'État républicain issu de la Révolution de 1789. La monographie porte sur des « cas typiques » jugés exemplaires. Son noyau est constitué par un relevé du budget des recettes et dépenses d'une famille ouvrière. Mais ces budgets ne sont pas agrégés, additionnés, comme ils le seront par exemple par l'économiste allemand Engel, auteur des célèbres lois de consommation. La monographie sert à illustrer un discours à la fois conservateur et social, hostile au salariat marchand capitaliste et au suffrage universel, mais favorable à la protection des conditions de vie des familles ouvrières contre les méfaits du capitalisme. Cette tradition de l'enquête sociologique de terrain, contrepoint du recensement exhaustif, se perpétuera, en dehors de la statistique officielle. Elle met en valeur la cohérence interne d'une constellation de traits enregistrés pour un cas singulier, traité comme *exemple* (Asad, 1994 ; Dodier & Baszanger, 1997). La statistique du XIX^e siècle, ignorant les sondages, accordait, au moins en principe, un poids comparable au recensement et à la monographie.

Entre 1875 et 1895, une crise économique grave frappe les pays industrialisés. Le salariat capitaliste s'est développé sans être codifié ni protégé par des lois spécifiques. La crise engendre le chômage, à la fois dans la réalité et comme catégorie d'analyse de celle-ci. Auparavant, la société assistait des *pauvres* et dressait des statistiques sur la pauvreté. Désormais, des institutions spécifiques vont aider les chômeurs, et, pour cela, les définir et les compter. Avant de chercher à mesurer le chômage en général, la statistique compte les chômeurs *secourus*, de même qu'auparavant, la pauvreté était mesurée, en Angleterre, par le nombre des personnes traitées par les bureaux de la loi sur les pauvres. Le lien entre la statistique et le traitement administratif local de cas individuels est encore direct. La statistique est alors une énumération exhaustive, calquée sur les procédures étatiques territoriales.

À partir du XX^e siècle, ce lien direct se transforme profondément. La forme de l'action de l'État et celle des procédures statistiques évoluent de façon parallèle. D'une part, la protection sociale est régie par des lois nationales, uniformes sur tout le territoire et fondées sur un *principe assurantiel de couverture statistique des risques*. D'autre part, la description statistique du monde social n'est plus issue seulement du comptage exhaustif de cas individuels. Elle recourt désormais à l'enquête par sondage sur un *échantillon aléatoire représentatif au niveau national*. Le lien entre ces évolutions, fondées toutes deux sur le calcul des probabilités, peut être vérifié. En 1895, le statisticien norvégien Kiaer présente une enquête par sondage empirique, dans le cadre d'un débat parlementaire sur de nouvelles lois de protection des travailleurs et des retraités. En 1906, l'anglais Bowley utilise le calcul des probabilités pour évaluer des intervalles de confiance, à l'occasion d'enquêtes préparatoires à la mise en place des lois sociales du *Welfare State* naissant.

La méthode des sondages aléatoires était connue depuis longtemps : le mathématicien Laplace l'avait utilisée en France dès les années 1780, pour estimer la population du Royaume. Elle avait pourtant été rejetée, au long du XIX^e siècle, par des statisticiens attachés à une pratique administrative exhaustive et territoriale. Pour eux la « fiabilité », c'est-à-dire la légitimité de la statistique, était liée à son exhaustivité. Le sondage était perçu comme une « acrobatie de géomètre ». Il n'avait pas place dans la statistique officielle.

Tout change au XX^e siècle, dès lors que les politiques sociales sont pensées comme un système cohérent et national, et non plus comme une somme de cas individuels traités localement. Dans le sondage, la singularité spécifique de l'individu disparaît, au profit du spécimen tiré au hasard dans une urne supposée uniforme. L'homologie des deux procédures, administrative et statistique, est frappante. Les lois sociales sont appliquées uniformément sur le territoire national. Les citoyens sont équivalents, de même que sont équivalentes les boules contenues dans l'urne probabiliste. La société, pensée comme un tout, de la sociologie solidariste de Durkheim est désormais dotée d'un équipement institutionnel avec les lois sociales, et d'un équipement cognitif avec la méthode des sondages. Les mises en place de ces deux technologies sont parallèles. Expérimentales et partielles entre 1900 et 1940, elles s'étendent massivement après la Seconde Guerre mondiale (INSEE, 1987 ; Anderson, 1988).

Un récit comparable peut être fait des développements conjoints, entre les années 1930 et 1960, des politiques macroéconomiques keynésiennes et des systèmes de *comptabilité nationale*. Celle-ci se place, dès son origine, au-dessus et en dehors des entreprises industrielles et commerciales, singulières, locales, identifiées par leurs noms et leurs activités spécifiques. Les agents économiques des *Tableaux économiques d'ensemble*, les branches des *Tableaux d'échanges interindustriels* de Léontieff, remplacent les énumérations et descriptions d'établissements de la statistique industrielle du XIX^e siècle. Pendant la période de croissance, des années 1950 à 1970, la politique économique est pensée du point de vue d'un acteur central régulant le jeu du marché par des interventions correctrices ou incitatives. La comptabilité nationale est bien adaptée à cette conception de l'action publique.

Trois modalités non exclusives de l'action à distance de l'État

Au XIX^e siècle, l'information statistique était fournie par la compilation de listes administratives, par des recensements, exhaustifs mais rares et portant sur peu de variables, et par des monographies locales, portant sur quelques individus mais sur beaucoup de variables. Au XX^e siècle, elle a été transformée par la mise en œuvre d'outils complètement nouveaux, comme les

enquêtes par sondage et la comptabilité nationale. De son côté, l'exploitation des fichiers administratifs a été bouleversée par le recours à la mécanographie, puis à l'informatique. Ces transformations peuvent être présentées comme des conséquences d'un progrès continu de la science et de la technique, ce qu'elles sont bien sûr. Mais, comme l'a montré le cas de la longue résistance contre la méthode des sondages aléatoires, elles sont aussi et surtout le reflet des transformations des fonctions de l'État et des formes de son action. L'État agit à *distance*, mais selon des modalités variables. On peut distinguer par exemple une action *administrative*, une action *statistique*, et une action *décentralisée*. Dans chaque cas, la « distance » est pensée et gérée de façon spécifique, et l'information statistique impliquée par ces formes de l'action de l'État est, elle aussi, spécifique. Ces trois formes coexistent et se mêlent bien sûr dans les États modernes. Elles ne constituent pas un découpage logique, selon un critère homogène, des formes possibles d'État.

L'État « *administratif* » édicte des lois générales, des règles, des codes. Ces règles sont appliquées, c'est-à-dire transformées en décisions singulières par des agents locaux. Ceux-ci ont un degré d'appréciation des circonstances particulières. Leurs décisions, dans des cas non prévus, peuvent faire jurisprudence et transformer ainsi la loi. L'incertitude porte sur la faculté d'interpréter la loi et de l'adapter aux circonstances. L'information statistique, fondée sur des récapitulatifs d'actes administratifs ne peut être qu'exhaustive ou monographique. La statistique est une forme de comptabilité des activités de l'État. Ainsi, les premières statistiques criminelles françaises, créées en 1825, portent longtemps le nom de « comptes généraux de la justice ». L'idée de mesurer la criminalité *dans la société* n'apparaît que beaucoup plus tard.

L'État « *statistique* » a implicitement intégré les idées de Quételet sur l'homme moyen, de Durkheim sur les faits sociaux distincts des faits individuels, et de Keynes sur la spécificité des dynamiques macroéconomiques. L'État-providence agit « en moyenne », couvre des risques, assure une péréquation statistique et assurantielle entre les individus. Il s'efforce de réduire des inégalités sociales, décrites par des statistiques issues d'enquêtes par sondage. Le ministre keynésien ajuste une offre globale et une demande globale, à l'aide de modèles macroéconométriques basés sur les tableaux de la comptabilité nationale. L'incertitude porte sur des risques, probabilisés ou non. Dans le premier cas, elle est maîtrisée par les propriétés de la « loi des grands nombres ». Celle-ci permet de surmonter la contingence et l'imprévisibilité des situations individuelles, et d'assurer une stabilité et une consistance à des agrégats statistiques, devenus ainsi pensables et gérables par l'État. Alors que la statistique de l'État « *administratif* » ne pouvait accepter l'idée de probabilité, celle-ci est en revanche essentielle pour l'État « *statistique* ».

Mais ces deux États ont en commun d'être dotés d'un *centre*. Le premier agit à travers des règles, leurs applications locales et leurs totalisations administratives. Le second prend appui sur la loi des grands nombres et agit à travers des moyennes statistiques. En revanche, les États *décentralisés*, les fédérations et les unions d'États souverains explorent de nouvelles voies. Celles-ci sont fondées sur les notions de subsidiarité, d'action procédurale, de négociation et de réseau. L'idée est de laisser le plus de liberté possible aux niveaux élémentaires de la société, en ne confiant aux niveaux supérieurs que les pouvoirs que ne peuvent assurer les premiers. L'action procédurale édicte des modalités de décision et de négociation et non pas des règles substantielles. Les lieux d'action, de décision, et par là de production et d'usage de l'information, sont de plus en plus nombreux et reliés entre eux de façons variées. Par ailleurs, les domaines considérés comme relevant d'une responsabilité collective sont multipliés, eux aussi : environnement, bioéthique, enfance maltraitée, toxicomanie, prévention du sida et d'autres nouvelles maladies, protection des minorités culturelles et linguistiques, égalité entre hommes et femmes, sécurité des équipements domestiques et industriels, normes de qualité des biens de consommation. Dans chaque cas sont élaborées et négociées *simultanément* des modalités de jugement et de mise en forme statistique de ces problèmes, des répartitions de responsabilités entre les divers acteurs, des façons d'évaluer *a posteriori* les actions publiques et de transformer celles-ci en fonction de ces évaluations. Des informations sont produites et utilisées dans tous les maillons de cette chaîne circulaire de la description, de l'action et de l'évaluation.

Une conséquence importante de ces formes nouvelles de l'action administrative est que la correspondance entre *statistique* et *territoire de l'État-nation* change de nature. Auparavant directe et évidente, et donc souvent oubliée, elle est désormais mise en question à partir de points de vue locaux ou supranationaux. La croissance des responsabilités des municipalités, des régions ou des États dans les pays fédéraux, implique un développement de l'information à ces niveaux. Mais celle-ci peut être pensée de deux façons. Elle peut résulter de la démultiplication d'une information déjà standardisée et produite nationalement. Dans ce cas, l'État central joue un rôle typique de producteur de normes et de standards. Mais l'information sur une localité ou une région peut aussi être conçue et produite localement. Dans ce cas, les informations produites en divers lieux du territoire national ne sont plus comparables et agrégeables. De ce point de vue, l'unification ou la dispersion de l'information produite dans un pays reflète son unification et sa dispersion politique et administrative.

Quand l'État central est fort et le territoire bien quadrillé par des administrations uniformes (cas français, depuis Napoléon), celles-ci produisent régu-

lièrement des informations organisées en termes de « variables » : à chaque bureau de ministère correspondent certaines variables d'action et d'évaluation. Des modèles économétriques peuvent relier les actions et leur résultats, à travers ces variables. En revanche, à un niveau local, l'information se présente moins sous forme de variables attachées à des actions spécifiques, que comme constellation d'attributs variés, plus proches de la monographie que de la statistique administrative. Le maire d'une ville veut tout savoir sur *sa ville*, tandis que le chef d'un bureau de ministère central veut tout savoir sur *sa variable*, pour tout le territoire. La tension entre ces deux points de vue constitue la richesse des systèmes statistiques des États décentralisés. En France, pays longtemps centralisé, elle n'est clairement ressentie que depuis les années 1970. L'Institut de statistique, l'INSEE, est doté de 21 directions régionales, mais celles-ci dépendent d'une direction générale parisienne. En revanche, en Allemagne ou en Suisse, États fédéraux, les *Länder* ou les cantons sont dotés de bureaux statistiques autonomes. Ces pays ont une longue tradition de statistiques municipales et régionales.

Les efforts d'harmonisation entre les statistiques de différents pays ont eux aussi pour effet de brouiller certaines évidences de l'information. C'est parce que la statistique est liée à l'organisation de l'État et aux structures d'une société que le rêve d'une standardisation complète n'est pas réalisable. Celle-ci impliquerait d'uniformiser tout ce qui est en amont de la production statistique : les systèmes fiscaux, la protection sociale, les définitions des emplois, les grilles salariales... Une solution intermédiaire, souvent retenue, est de définir en commun des grandeurs théoriques, chaque pays étant libre de les évaluer selon ses possibilités. Les agrégats de la comptabilité nationale et certains indicateurs sociaux sont ainsi construits. Ceci permet de publier des tableaux comparant les économies et les structures sociales de différents pays. Cette démarche est sans doute inévitable. Elle risque pourtant de rendre encore plus opaques les boîtes noires dans lesquelles sont rendus équivalents et sont combinés les éléments dispersés et hétérogènes du monde social. Elle devrait être accompagnée d'un travail parallèle et complémentaire de comparaison entre les systèmes statistiques et entre leurs modes d'insertion dans le tissu social et administratif de chaque pays. La construction d'une Union européenne et, plus généralement, la mondialisation de l'économie, ont rendu encore plus actuelles des questions que les statisticiens se posent en fait depuis 1853, année de leur premier congrès international, organisé par Quételet.

La tension entre deux espaces de références normatives

Cette extension mondiale de l'information statistique semble accentuer son caractère « galiléen », totalisateur et réducteur des singularités, individuelles

ou même nationales. Pourtant ses développements récents l'ont conduite aussi dans une direction apparemment opposée, celle d'une identification et d'une catégorisation fines des personnes, à des fins marchandes. De plus en plus souvent, les banquiers, les assureurs ou les chaînes de commerce ciblent leurs clients en utilisant des informations individuelles combinées par de subtils logiciels statistiques. L'intelligence artificielle et les systèmes experts permettent d'automatiser des choix, des décisions, des affectations d'individus à des classes de traitements variés. Le paradigme de l'indice, cher à Carlo Ginzburg, semble être lui aussi absorbé par la machine. Un diagnostic médical, une décision de justice, peuvent être aidés par des informations statistiques, grâce à des raisonnements probabilistes d'inspiration bayésienne : toute information supplémentaire sur un cas peut modifier les probabilités subjectives conduisant à une décision.

Mais cette automatisation a conduit les spécialistes des sciences cognitives et les sociologues des organisations à examiner de plus près les questions de contexte et les processus dynamiques de décision. Le décideur humain, médecin ou juge, tient compte d'informations purement locales, non codifiables. Ses actes ont un passé et un avenir. Ils sont inscrits dans un processus, une mémoire, une stratégie. Ils sont liés aux décisions d'autres acteurs, dans des réseaux coordonnés par des règles, des coutumes, des institutions, des objets ou des informations statistiques. Celles-ci paraissent de plus en plus comme des objets parmi d'autres, jouant un rôle de point d'appui pour argumenter, de signal et de référence commune pour négocier et se coordonner. L'objectivité de l'information quantitative est à la fois le produit d'un coûteux processus social d'objectivation, et une ressource précieuse pour permettre l'accord entre des acteurs dispersés et hétérogènes.

Cet accord suppose souvent des compromis entre des systèmes de justification et des biens communs radicalement hétérogènes, sinon incommensurables. Un exemple intéressant des négociations nécessaires pour surmonter de telles incompatibilités est fourni par le cas de l'information statistique elle-même. Les lieux du débat public où les objectifs et les modalités de celle-ci sont discutés sont de deux types. D'une part, des « conseils de la statistique » (en France, le Conseil national de l'information statistique, CNIS) réunissent des représentants des utilisateurs, administrations, syndicats d'employeurs et de salariés, associations... On y discute des besoins en information économique et des programmes des offices statistiques. D'autre part, des organes spéciaux ont été créés au moment de la mise en place des premiers grands fichiers informatiques. Ils veillent à la protection de la vie privée, au secret et à la confidentialité des informations fournies par les individus, non seulement aux statisticiens mais dans beaucoup d'autres circonstances. La Commission

nationale de l'informatique et des libertés (CNIL) vérifie soigneusement les questionnaires préparés par les statisticiens et les démographes. Les objectifs de ces deux lieux de débat démocratique, le CNIS et la CNIL, sont différents et peuvent entrer en conflit. Les discussions et négociations qui en résultent obligent chaque partie à expliciter clairement ses objectifs et les limites infranchissables selon elle, et à écouter les justifications de l'autre partie pour définir un bien commun.

Ainsi la construction même d'un système d'information fait apparaître une tension entre deux espaces de références normatives, aussi légitimes l'un que l'autre. Le besoin d'information économique et sociale est lié notamment (mais non uniquement) au projet d'organiser une « société plus juste ». Les actions sont évaluées selon leurs conséquences sur la collectivité. En revanche, la protection de la vie privée se réfère à l'idée que les individus, « libres et égaux en droit », selon la Déclaration des droits de l'homme de 1789, doivent être assurés que l'État garantit l'exercice de cette liberté et de cette égalité. La liberté implique que l'espace privé soit protégé. L'égalité porte sur les droits de citoyenneté, alors que, dans l'autre système de référence, elle porte sur les chances sociales et économiques.

La philosophie des *droits des individus*, longtemps réticente à l'investigation statistique, remonte aux révolutions anglaise, américaine et française du XVIII^e siècle. Celle de la *justice sociale* s'est développée au XX^e siècle, en prenant appui notamment sur l'analyse statistique des inégalités sociales, entre les classes, puis, plus tard, entre les sexes ou entre les groupes ethniques. Cet usage de l'information statistique est assez récent. Il remonte aux années 1930 aux USA, et aux années 1950 en Europe. La mesure statistique des inégalités entre groupes sociaux est à peu près absente des travaux de Durkheim, alors qu'elle occupe une place centrale dans la sociologie postérieure à 1945.

La Constitution française de 1958 est précédée de deux préambules. L'un est la Déclaration des droits de 1789. L'autre, datant de 1946, porte sur les droits *sociaux*. Les États de droit modernes combinent, chacun à leur manière, ces deux exigences. L'information statistique, plutôt bridée quand la première dominait, a changé de nature et de portée sociale quand la seconde est devenue aussi importante que la première.

La philosophie politique distingue les morales *déontologiques*, qui évaluent les actes en eux-mêmes et pour qui chaque personne humaine est irremplaçable et sacrée, et les morales *téléologiques*, qui évaluent les actes du point de vue de leurs conséquences, notamment sur la collectivité, ce qui suppose la définition d'un « bien commun » (Fagot-Largeault, 1991). Les sociétés modernes ne peuvent méconnaître ni les unes ni les autres de ces morales. Les débats sur le rôle de l'information statistique sont inscrits dans cette question

plus large. Les objectifs de justice sociale, d'efficacité économique et de protection des personnes doivent être pensés et discutés en même temps, dans des espaces publics où ces différentes logiques normatives peuvent se déployer et se confronter, sans que l'une d'entre elles domine les autres. La construction ou la reconstruction de l'État impliquent de tels espaces. L'information statistique est un des langages grâce auxquels les acteurs sociaux peuvent se coordonner.

L'argument statistique : entre la décision et la preuve

La recherche historique et sociologique sur la statistique se trouve souvent partagée entre deux domaines bien distincts : la sociologie politique et la sociologie des sciences. La première étudie la place de la statistique dans le développement des États modernes et de leurs bureaucraties impersonnelles. Elle fait souvent référence à Foucault et à son analyse de la « gouvernementalité ». Dans cette perspective, l'accent est mis sur la capacité des bureaux de statistique et des méthodes quantitatives à instaurer une forme puissante d'objectivité, socialement construite, et fondée sur la « confiance dans les nombres » (Porter, 1995). La statistique paraît alors surtout comme une composante essentielle d'une « sociologie de la modernité », tendue entre « liberté et discipline », selon l'expression de Peter Wagner (1996). La sociologie des sciences, en revanche, lit l'histoire de la statistique selon un autre point de vue. Elle suit l'évolution des formulations et des usages des *schémas probabilistes*, depuis leurs premières expressions par Huygens et Bernoulli (Gigerenzer *et al.*, 1989; Hacking, 1990; Fagot-Largeault, 1989). Cette histoire est relativement différente de la première, même si les points de rencontre existent, depuis Laplace et Quételet, et si l'État *statistique*, évoqué ci-dessus, prend appui sur les régularités macrosociales rendues visibles par l'interprétation dite « fréquentiste » du calcul des probabilités : les systèmes de protection sociale sont fondés sur de telles permanences statistiques.

L'argument statistique est toujours tendu entre une rhétorique (politique) de la décision et de la justification, et une rhétorique (scientifique) de la vérité et de la preuve. Les controverses qui, dans les années 1930, ont entouré la formulation et la diffusion des *tests statistiques*, offrent un bon exemple de cette tension (Gigerenzer *et al.*, 1989). Aujourd'hui massivement utilisés, ces outils ont été proposés sous deux formes différentes et concurrentes, d'une part, par Ronald Fisher, et, d'autre part, par Jerzy Neyman et Egon Pearson. Pour Fisher, le test permet de *vérifier* si deux valeurs observées sont ou non significativement différentes l'une de l'autre. Le but recherché est, dans une visée de *vérité*, la construction progressive d'un savoir scientifique, éventuellement falsifiable au sens poppérien. Pour Neyman et Pearson, en revanche,

le test a pour but de *décider* si deux valeurs vont être considérées comme significativement différentes, en envisageant successivement les *risques* (en terme de coûts) des deux choix erronés complémentaires : tenir une assertion pour vraie alors qu'elle est fausse, ou la tenir pour fausse alors qu'elle est vraie. Le but recherché est alors, dans une visée *économique*, de formuler une procédure objective et systématique (par exemple pour tester la qualité d'une fabrication en série) permettant d'orienter une *action*. Entre une visée de vérité et une visée de décision, le dialogue a été longtemps très difficile. Il n'est pas clos et offre un terrain de recherche illimitée pour une sociologie politique des sciences.

Bibliographie

- Anderson, M. J.
1988 *The American Census. A Social History*. New Haven, Yale University Press.
- Asad, T.
1994 « Ethnographic Representation, Statistics and Modern Power », *Social Research*, 61(1), p. 55-88.
- Desrosières, A.
1993 *La politique des grands nombres. Histoire de la raison statistique*. Paris, La Découverte.
- Dodier, N. & Baszanger, I.
1997 « Totalisation et altérité dans l'enquête ethnographique », *Revue française de sociologie*, XXXVIII (1), p. 37-66.
- Fagot-Largeault, A.
1989 *Les causes de la mort. Histoire naturelle et facteurs de risque*. Paris, Vrin.
1991 « Réflexions sur la notion de qualité de la vie », *Archives de philosophie du droit*, 36, p. 135-153.
- Gigerenzer, G. *et al.*
1989 *The Empire of Chance. How Probability Changed Everyday Life*. Cambridge, Cambridge University Press.
- Ginzburg, C.
1980 « Signes, traces, pistes. Racines d'un paradigme de l'indice », *Le Débat*, 6, p. 3-44.
- Gollac, M.
1997 « Des chiffres insensés ? Pourquoi et comment on donne un sens aux données statistiques », *Revue française de sociologie*, XXXVIII (1), p. 5-36.
- Goody, J.
1979 *La Raison graphique*. Paris, Minuit.

- Hacking, I.
1990 *The Taming of Chance*. Cambridge, Cambridge University Press.
- INSEE
1987 *Pour une histoire de la statistique*. Paris, INSEE-Economica, 2 vol.
- Latour, B.
1989 *La Science en action*. Paris, La Découverte.
- McKenzie, D.
1981 *Statistics in Britain, 1865-1930. The Social Construction of Scientific Knowledge*. Edimbourg, Edinburgh University Press.
- Perrot, J. -C.
1992 *Une histoire intellectuelle de l'économie politique, 17^e-18^e siècles*. Paris, Éd. de l'EHESS.
- Porter, T.
1995 *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*. Princeton, Princeton University Press.
- Wagner, P.
1996 *Liberté et discipline. Les deux crises de la modernité*. Paris, Métailié.

Geoffrey BOWKER et Susan Leigh STAR

PROBLÈMES DE CLASSIFICATION ET DE CODAGE DANS LA GESTION INTERNATIONALE DE L'INFORMATION¹

Marx a comparé la technologie à du « travail solidifié [« gelé »] », le travail et ses valeurs étant enchâssés [*embedded*] dans une forme transportable². Les nouvelles technologies de l'information incorporent pareillement du travail selon des modalités qui sont déterminantes pour les politiques tout en demeurant souvent difficiles à distinguer. Là où elles sont utilisées pour prendre des décisions ou représenter les étapes d'un processus décisionnel, ces technologies contribuent également à incorporer ces décisions. C'est dire qu'à l'intérieur d'une technologie, ou d'une représentation complexe, sont dissimulés des arguments, des décisions, des incertitudes ainsi que la nature procédurale de prises

1. Traduction de « Knowledge and Infrastructure in International Information Management. Problems of Classification and Coding », in L. Bud-Friedman (ed.), *Information Acumen. The Understanding of Knowledge in Modern Business*, Londres, Routledge, 1994, p. 187-213.
2. Nous aimerions exprimer notre reconnaissance à nos collègues Joan Fujimura et Alberto Cambrosio pour l'aide qu'ils nous ont apportée et les nombreuses heures de discussion stimulante concernant les questions présentées ici. Une partie de cette recherche a été réalisée grâce à une bourse de l'Université de Californie. Anne-Marie Mol, Lisa Bud-Friedman et Anne Fagot-Largeault ont apporté des commentaires étendus et très utiles sur une première version de cet article. Nous remercions aussi pour leurs commentaires utiles les différents discutants anonymes présents lors de la conférence de l'HICSS en 1991. Les Archives des Nations Unies et l'Organisation mondiale de la santé, située à Genève, nous ont donné très aimablement accès aux matériaux non publiés.