

La cumulativité en démographie

Daniel Courgeau

Ined

Si la démographie est née au XVII^e siècle avec l'œuvre de John Graunt (1662), elle était alors considérée comme une arithmétique politique, terme introduit par William Petty dans les années 1670 (Dupâquier et Dupâquier, 1985) et titre de son ouvrage publié en 1690. Le terme démographie est apparu au XIX^e siècle avec les *Eléments de statistique humaine, ou démographie comparée* de Guillard (1855). Elle est maintenant définie comme « une science ayant pour objet l'étude des populations humaines, et traitant de leur dimension, de leur structure, de leur évolution et de leurs caractères généraux envisagés principalement d'un point de vue quantitatif » (Henry, 1981). Cette science s'est développée sous diverses approches apparues tout au long de son histoire, entre autres, celle de l'analyse transversale, de l'analyse longitudinale, de l'analyse biographique et de l'analyse multiniveau (Courgeau, 2002).

Nous ne reprendrons pas ici la discussion de Berthelot (2001b) sur les programmes et paradigmes en sciences sociales, mais rappellerons seulement la définition de quelques termes, qui nous suffiront dans le cadre de cet article. Bien que Masterman (1970) ait relevé vingt et un sens différents du terme paradigme dans l'ouvrage de Kuhn (1962), ce dernier en a finalement distingué deux principaux :

d'une part, l'ensemble des convictions, des valeurs reconnues et des techniques qui sont communes aux membres d'une communauté scientifique donnée ; d'autre part, un élément isolé de cet ensemble, constitué de solutions concrètes de problèmes qui, employées comme modèles ou exemples, peuvent remplacer les règles explicites en tant que base pour résoudre les problèmes subsistant dans la science normale¹ (Kuhn, 1970).

Nous proposons ici une notion un peu différente de celles proposées par Kuhn qui répond en fait à la question suivante : comment passer des *phénomènes vécus* à l'*objet scientifique* au sens de G.-G. Granger (1994) ? Pour ce philosophe, « le vécu complexe saisi dans l'expérience des choses sensibles est devenu *objet* d'une mécanique et d'une physique, par exemple, lorsque l'on s'est avisé de le réduire à un modèle abstrait, où n'ont figuré tout d'abord que spatialité, temps et "résistance" au mouvement ». Et il reconnaît que cet objet ne bénéficie pas dès le départ d'une définition explicite et générale de son contenu. Ainsi les sciences comme la physique et la biologie font de leur objet des explicitations successives comme le montre le passage de la physique de Newton à la relativité générale d'Einstein. De même la démographie va expliciter son objet au travers de paradigmes successifs, chaque paradigme prescrivant une façon différente de rapporter les phénomènes observés à l'objet scientifique (Courgeau et Franck, 2007). Cette notion conviendra ici pour notre examen de la démographie, au travers des divers exemples que nous allons traiter. Nous verrons également

¹ On the one hand, it stands for the entire constellation of beliefs, values, techniques and so on shared by the members of a given community. On the other, it denotes one sort of element in that constellation, the concrete puzzle-solutions which, employed as models or as examples, can replace explicit rules as a basis for the solution of the remaining puzzles of normal science.

dans la conclusion qu'elle permet d'introduire une cumulativité, impossible avec les notions définies par Kuhn.

Nous examinerons si un paradigme unique a servi la démographie ou si, au contraire, celle-ci a rencontré des points de bifurcation entraînant des changements de paradigme. Un premier examen semble montrer, pour les sciences sociales,

l'élaboration progressive d'un corpus de procédures et d'outils constitutifs de noyaux disciplinaires d'un côté, l'inscription des opérations de connaissance menées sous l'autorité du rationalisme expérimental et de ses exigences de formalisation et de contrôle de l'autre, cette double orientation semble inviter à l'approfondissement, au renforcement, et à la constitution d'un savoir cumulatif (Berthelot, 2001a).

Mais l'examen plus approfondi des diverses sciences sociales, paraît lui montrer qu'en fait il n'en est rien :

tant sur le plan logique que dans leur évolution historique, le mouvement commun qui anime la sociologie, l'ethnologie, la démographie et la psychologie sociale n'est pas un mouvement de consolidation et de renforcement par épuration et rectification successives, mais un mouvement de dissémination, de constructions alternatives, de complexification et de ruptures.

Nous allons essayer de voir ici, plus en détail, si ce mouvement s'est réellement opéré en démographie au travers de *révolutions*, l'entraînant sur des voies et des procédures divergentes entre elles, ou au contraire au travers d'une *évolution* qui a permis un progrès du savoir scientifique (Courgeau, 2002). Cette évolution a-t-elle ouvert l'accès à des paradigmes différents, qui permettent simultanément l'accumulation des connaissances dans cette science humaine ? Pour ce faire nous allons examiner un certain nombre d'exemples qui nous montreront plus clairement ce qu'il en est.

Premier exemple : le taux de masculinité à la naissance

La naissance d'un plus grand nombre de garçons que de filles a intéressé les chercheurs depuis le XVIIe siècle : Graunt (1662) remarque déjà que l'on a baptisé à Londres « 139 782 garçons et seulement 130 866 filles, et que les comptes de la Province concordent assez bien avec ceux de Londres sur ce point² », soit un taux de masculinité de 51,65 %, et indique qu'il « voudrait demander aux voyageurs s'il en est de même dans d'autres pays³ ». Depuis lors de nombreux démographes se sont intéressés soit à ce taux, soit au rapport de masculinité à la naissance, rapport de l'effectif des garçons à celui des filles⁴.

Devant cette donnée simple à mesurer se dressent un certain nombre d'interprétations, qui peuvent conduire à des notions différentes. Nous présenterons ici certaines d'entre elles.

Au XVIIIe siècle les chercheurs sont frappés par l'observation d'une régularité extrême dans la répartition des naissances selon le sexe, de même que dans l'arrivée de nombreux autres phénomènes démographiques, tels que la naissance, la mort, la maladie, etc.

² That there have been also *Christened* within the same time, 139782 *Males*, and but 130866 *Females*, and that the Country Accompts are consonant enough to those of *London* upon this matter.

³ Onely we shall desire, that Travellers would enquire whether it be the same in other Countries.

⁴ Voir à ce sujet : Arbuthnott, 1710, Süßmilch, 1741, Laplace, 1781, Condorcet, 1793-1794, Poisson, 1830, Quetelet, 1835, Darwin, 1871, Halbwegs, 1933, 1936, Brian et Jaisson, 2007, etc.

Arbuthnot (1710), observant 82 années de naissances à Londres, de 1629 à 1710, montre que chaque année il y a plus de naissances masculines que féminines, que « cet excès se produit en proportions constantes, et que la différence reste dans des limites fixées⁵ ». Il est intéressant de voir là l'un des premiers tests statistiques publiés. Il pose en effet le problème sous la forme suivante : « l'hypothèse A est opposée à B, selon laquelle il y aurait plus de garçons nés que de filles : Trouver la chance de A ou la valeur de son espérance⁶ ». Il en conclut que ce ne peut être par hasard qu'il puisse y avoir plus de garçons nés que de filles, en proportion constante, et que cela prouve l'existence d'une *Providence Divine*. Cependant il est intéressant de noter que son premier argument selon lequel « la régularité constante observée dans la naissance des deux sexes⁷ » ne peut pas être une question de chance, n'est pas valide comme raisonnement statistique, comme l'a clairement montré Nicolas Bernoulli dans ses lettres à Montmort (1713).

Süssmilch (1741) a étendu cette observation à de nombreux pays et indique que dans ces naissances, il n'y a pas seulement toujours plus de garçons que de filles, mais encore ce qui est assurément digne d'étonnement, les garçons ont toujours avec les filles une certaine proportion ; et celle-ci n'est pas indéterminée ni ne vient une fois ainsi par hasard, mais est au contraire constante et presque toujours égale.

Il étend ensuite son observation à de nombreux phénomènes démographiques empiriques, qui devraient lui révéler l'existence et la nature de l'Ordre Divin : l'ordre du surplus des naissances sur les décès, l'ordre des degrés de peuplement de la surface terrestre, l'ordre des naissances, l'ordre des mariages, l'ordre de la mortalité, l'ordre des maladies, etc. Pour ce faire, l'instrument de la preuve est le recours aux grands nombres et à la notion de hasard. Cependant, même s'il montre que, pour la mortalité, « la différence entre villes et villages doit être recherchée dans la manière de se nourrir, dans les mœurs et dans le mode de vie » (1761), il conclut : « Si les mœurs et les modes de vie étaient partout semblables – comme l'est la nature -, la mortalité le serait également ». Il ajoute que, si les lois du hasard sont simplement descriptives et si des différences peuvent exister selon les modes de vie, les causes immuables sous-jacentes se trouvent dans l'Ordre Divin, qu'une société parfaite montrerait : la pensée mythologique vient alors suppléer l'absence d'explication des régularités observées, mais cependant cet ordre ne guide ni n'oriente la recherche.

C'est donc un courant de pensée tout à fait prépondérant au XVIIIe siècle que nous avons mis en évidence à partir de la lecture des œuvres de Arbuthnot et de Süssmilch, qui découvraient la régularité statistique des taux, et que l'on retrouve chez de nombreux penseurs de l'époque. Ainsi Pearson (1926) dans un article consacré à De Moivre écrivait :

L'idée de Newton d'une divinité omniprésente maintenant les valeurs statistiques moyennes fut à la base d'un développement des statistiques allant de Derham, Süssmilch, Niewentyt et Price jusqu'à Quetelet et Florence Nightingale [...]. De Moivre développa la théologie newtonienne et orienta la statistique dans une direction nouvelle où elle se maintint pendant presque un siècle.

Si l'on ne peut pas encore parler de paradigme, car il s'agit d'une pensée mythologique, nous pouvons cependant dire que cette approche est placée sous l'hypothèse

⁵ This excess shall happen in a constant Proportion, and the Difference lye within fix'd limits

⁶ A lays against B, that every year there shall be born more males than females : to find A's lot, or the value of his Expectation.

⁷ A very remarkable one to be observed in the exact Ballance that is maintained, between the Numbers of Men and Women

suivante : *la constance des lois de la nature, qui peut être prouvée par le recours aux grands nombres, devrait apparaître dans la régularité des taux démographiques observés dans toutes les populations humaines.* Ainsi le problème de la masculinité des naissances pourrait être résolu selon le même schème que tous les autres problèmes démographiques.

Cependant, même si cette hypothèse a pu se maintenir chez certains auteurs au cours du temps, l'importance des variations de ces taux selon les populations et les époques apparut comme beaucoup trop large pour être négligée et conduisit cette fois-ci à une véritable pensée scientifique et à un véritable paradigme.

Ainsi Laplace (1780) constate qu'il y a

plus de quatre cent mille à parier contre un, que les naissances des garçons sont plus faciles à Londres qu'à Paris ; ainsi on peut regarder comme une chose très-probable qu'il existe dans la première de ces villes, une cause de plus que dans la seconde, qui y facilite les naissances des garçons, & qui dépend soit du climat, soit de la nourriture & des mœurs.

Comme on le voit il ne s'agit plus de la constance d'un Ordre Divin mais de variations de l'observation, qui conduisent à faire des hypothèses sur les raisons des différences observées. Bien qu'il ne puisse montrer avec précision que les différences sont dues au climat, à la nourriture, etc., car il ne dispose pas des statistiques nécessaires, il ne se contente plus d'une soi-disant loi divine.

Condorcet (1772-1794) va encore plus loin en indiquant : « On peut demander encore quelle pourrait être sur l'espèce humaine [l'effet de] la découverte d'un moyen de produire un enfant mâle ou femelle suivant la volonté des parents, moyen que les observations faites avec soin doivent faire découvrir à la longue. » Non seulement il rejette la loi divine mais prévoit une action possible de la pratique humaine sur la masculinité des naissances.

Les recherches de Laplace sont poursuivies par Poisson (1829) qui montre que le taux de masculinité « dépend de l'état de la société, puisque le nombre de naissances hors mariage influe sensiblement sur la proportion des naissances masculines et féminines⁸ ». En revanche, « ce rapport est à peu près le même pour le midi de la France et pour la France entière, en sorte qu'il paraît indépendant de la variation du climat, au moins dans l'étendue de notre pays ». Il montre également que « la proportion des naissances des deux sexes n'est pas non plus la même à Paris et dans les départements ». De même Cournot (1843) examine de nouveaux résultats obtenus dans ces recherches. Il montre que les résultats obtenus pour Paris sont valables pour les principales villes de l'Europe. Il indique également que « la chance moyenne des naissances masculines est sujette à éprouver de fort notables variations d'un département à l'autre et d'une année à l'autre ». Mais les statistiques qui montrent ces différences lui semblent par ailleurs suspectes, et l'amènent à reporter à plus ample informé ces résultats. Enfin il donne les résultats obtenus par Hofacker (1828) et Sadler (1830), tout en critiquant leur source :

la cause de la prépondérance des naissances masculines résiderait uniquement dans la supériorité habituelle de l'âge du père sur l'âge de la mère. La chance pour chaque sexe ne dépendrait pas de l'âge absolu du père ou de la mère, mais seulement de la différence des âges. Malheureusement les nombres sur lesquels

⁸ Des indicateurs différents ont été utilisés au cours du temps : nombre de naissances vivantes de garçons pour cent filles, proportion des garçons sur cent naissances vivantes, etc. Il est facile de les relier les uns aux autres : d'un point de vue statistique la mesure la plus satisfaisante est la proportion de garçons pour cent naissances vivantes, qui est une probabilité.

s'appuient ces deux statisticiens sont insuffisants pour décider une question de cette importance.

On voit là l'importance du calcul des probabilités pour confirmer ou infirmer une hypothèse démographique.

Essayons maintenant de voir avec plus de précisions le paradigme qui soutient et permet une telle analyse.

Quittant l'intemporalité des travaux antérieurs, ces travaux se situent à un moment transversal : ils observent les résultats d'une année et cherchent à relier la masculinité des naissances à diverses caractéristiques de la société dans laquelle elle se produit. C'est donc bien dans le cadre d'une analyse du moment des groupes sociaux qu'elle se situe. La méthode des variations concomitantes, proposée par Durkheim (1895), est la plus pertinente pour effectuer cette analyse :

Pour qu'elle soit démonstrative, il n'est pas nécessaire que toutes les variations différentes de celles que l'on compare aient été rigoureusement exclues. Le simple parallélisme des valeurs par lesquelles passent les deux phénomènes, pourvu qu'il ait été établi dans un nombre suffisant de cas suffisamment variés, est la preuve qu'il existe entre eux une relation. Cette méthode doit ce privilège à ce qu'elle atteint le rapport causal⁹.

Il constate ainsi que le taux de suicide augmente régulièrement avec la proportion de protestants dans treize provinces de Prusse, montrant que « le suicide varie en raison inverse du degré d'intégration de la société religieuse » (Durkheim, 1897). Cela revient dans le vocabulaire d'aujourd'hui à réaliser une analyse de régression linéaire.

Cette approche peut être généralisée à tous les phénomènes démographiques et conduit vraiment à un paradigme, qui permet de réaliser une analyse du moment : *les faits sociaux s'expliquent par diverses caractéristiques économiques, politiques, religieuses, sociales, etc. de la société*. C'est la variation des taux qu'il faut maintenant chercher à expliquer par de nombreux facteurs sociaux (Russo, 2009). Cela définit une forme de causalité qui trouve son origine dans la société elle-même, et dont les effets se font sentir sur l'ensemble de la population. On voit dès lors combien ce nouveau paradigme est différent de l'hypothèse précédente, en faisant dépendre les phénomènes observés des caractéristiques de la société dans laquelle ils se produisent et non d'une puissance extérieure intemporelle. On peut cependant se demander si ce passage de la régularité des taux à leur variation introduit un mouvement de rupture, comme l'indiquait Berthelot, ou au contraire permet un enrichissement de la recherche démographique. En fait dans certains cas on pourra toujours vérifier leur régularité, alors que dans d'autres il ne sera plus possible de les considérer comme constants, comme les travaux de Poisson (1829) le montraient parfaitement. Ce passage permet donc de généraliser l'examen des taux démographiques, au lieu de les supposer constants, et de rechercher les raisons qui peuvent entraîner leur variation.

Nous n'irons pas plus loin dans l'examen de cette approche des taux de masculinité à la naissance, qui s'est poursuivie et améliorée jusqu'à aujourd'hui, et renvoyons le lecteur intéressé à de nombreux articles sur ce thème¹⁰, et surtout aux travaux récents de Brian et Jaisson (2007), qui montrent combien le taux de masculinité à la naissance peut s'éloigner des valeurs observées dans le passé : ainsi en Chine il atteignait 55,1% en 1999.

⁹ Cette dernière affirmation de Durkheim a été remise en question par divers chercheurs, comme Lazarsfeld et Boudon, qui ont montré par l'analyse multivariée que la méthode des variations concomitantes ne suffit pas pour établir le rapport causal, quoiqu'elle y contribue beaucoup. Mais là n'est pas le propos de cet article.

¹⁰ Entre autres : Halbwegs, 1933 ; Halbwegs et Sauvy, 1936 ; Coale et Bannister, 1994 ; Attané, 2005

Cet exemple nous montre très clairement que la construction des données est indépendante de la construction de la théorie (Suppe, 1989) : la mesure des données est restée pratiquement identique pendant plus de trois siècles, alors que la théorie a pour elle-même énormément changé.

Deuxième exemple : la génétique des comportements démographiques

Le second exemple que nous allons présenter ici est plus complexe car il cherche à combiner la part due à la nature et celle due à la culture dans les comportements sociaux, sous l'hypothèse qu'il est possible de décomposer la transmission des comportements en une part due aux gènes et une part due à l'environnement dans lequel vit l'individu. C'est donc de l'application d'un modèle qu'il s'agit ici.

Il est en premier lieu évident que tous les comportements humains sont influencés par nos gènes : il n'existe pas de trait observable qui ne soit pas fonction du patrimoine génétique de l'individu. Mais dès que l'on cherche à mesurer la magnitude de cette influence, cela revient à admettre qu'une mesure pourra être définie et calculée, permettant de caractériser les parts de la nature et de la culture dans la détermination du trait étudié. C'est cette mesure qui a permis le développement de la génétique du comportement, à partir des années 1920 et qui continue à justifier les nombreux travaux publiés dans ce domaine de nos jours, en particulier en démographie¹¹. Voyons plus en détail les hypothèses sur lesquelles elle est construite.

C'est Fisher (1918) qui le premier a posé les hypothèses utilisées après lui pour la mise en place de la génétique du comportement. A ce moment là, la structure des chromosomes n'était pas connue et l'hérédité mendélienne ne permettait que l'explication de la transmission de caractères qualitatifs, sous la dépendance d'un nombre limité de gènes. Les expériences de Mendel (1865), sur des graines de pois, lui permirent de mettre en évidence les lois de ségrégation, qui montraient le caractère discontinu du matériel héréditaire, et d'indépendance entre différents caractères.

Pour comprendre la transmission de caractères quantitatifs, Fisher (1918) fit l'hypothèse de polygènes soumis, indépendamment les uns des autres, à un processus de ségrégation. Il a également introduit la possibilité de dominance d'un gène sur un autre situé sur un même locus. Cela lui permit de calculer les corrélations génétiques entre individus reliés par divers liens de parenté (frère-sœur, parent-enfant, cousins, etc.) sous l'hypothèse d'unions aléatoires, et des formules semblables pour des unions basées sur l'apparentement. Il introduisit enfin l'effet de l'environnement, par l'intermédiaire d'un coefficient égal au rapport de la variance génétique à la variance phénotypique. Cet effet est alors considéré comme indépendant de celui des gènes. Ce travail, remarquable pour l'époque où il a été réalisé, posait clairement certaines hypothèses nécessaires à la génétique du comportement.

La notion d'héritabilité fournit ensuite le modèle nécessaire pour pouvoir séparer les diverses composantes génétiques et d'environnement qui entrent dans la génétique du comportement. Ce concept a été introduit dans les années 1920 par Sewall Wright (1921), pour décomposer la variance phénotypique d'un trait en une composante génétique et une composante due à l'environnement, indépendantes entre elles. Pour le définir, il est nécessaire de poser un certain nombre de conditions, qui ont été souvent oubliées par les utilisateurs de ce concept par la suite : les polygènes agissent additivement les uns par rapport aux autres, leur ségrégation est indépendante, l'effet de l'environnement est indépendant de celui des

¹¹ Voir à ce sujet les travaux de Kohler *et al.*, 1999 ; Rodgers *et al.*, 2000 ; Rodgers *et al.*, 2001 ; Morgan et King, 2001 ; Rodgers et Kohler, 2003, etc.

gènes et est aléatoire, il n'y a ni migrations, ni mutations ou sélection dans la population étudiée. Enfin pour simplifier les calculs on doit supposer que le nombre de polygènes est infini. Nous sommes bien ici en présence d'un modèle simple applicable à toute population vivante.

Sous toutes ces conditions, il est possible de définir avec précision les concepts d'héritabilité large et d'héritabilité étroite, qui sont directement utilisés par la génétique du comportement pour mesurer les parts respectives de la variance génotypique et de la variance environnementale. Voyons plus en détail le calcul de ces variances.

Essayons d'abord de séparer les deux séries de causes qui conduisent à un caractère phénotypique mesuré X , par exemple en démographie la fécondité d'un individu : l'information génétique transmise par ses parents et la contribution de l'environnement sur ses comportements. Il est alors tentant de décomposer la variance de ce même caractère, mesurée maintenant sur une population, en une part génétique et une part due à l'environnement :

$$\text{var}(X) = \text{var}_G(X) + \text{var}_E(X)$$

L'hypothèse sous-jacente qui est faite est alors que la covariance entre effet génétique et d'environnement est nulle, comme l'hypothèse d'indépendance entre ces effets le permet. On peut dans ce cas définir l'héritabilité au sens large comme la part des causes génétiques par rapport à la variabilité totale :

$$H^2 = \frac{\text{var}_G(X)}{\text{var}(X)}$$

Cependant les effets des polygènes ne peuvent pas être considérés comme rigoureusement additifs et il faut introduire dans l'effet génétique une part due à l'effet additif des gènes, une autre part due à l'effet d'interaction entre gènes d'un même locus (dominance) et une dernière due à l'interaction entre gènes situés à des locus différents (épistasie). Dans la mesure où l'épistasie est difficile à estimer sans une connaissance détaillée de la structure génétique, elle est à nouveau supposée négligeable par rapport aux précédents effets. On décompose dans ce cas la variance due aux causes génétiques en une part additive et une part de dominance :

$$\text{var}_G(X) = \text{var}_A(X) + \text{var}_D(X)$$

On peut alors définir l'héritabilité au sens étroit comme la part de l'effet additif des gènes par rapport à la variabilité totale :

$$h^2 = \frac{\text{var}_A(X)}{\text{var}(X)}$$

Ces différentes relations permettent, lorsque l'on dispose par exemple de données sur des vrais et faux jumeaux, d'estimer les divers types d'héritabilité et les composantes de la variance correspondant aux différents effets (voir par exemple Kohler *et al.*, 1999).

On peut dès lors dire que cette approche est placée sous le paradigme suivant : *il est possible de séparer mathématiquement l'effet des gènes et de l'environnement de façon additive sur tous les traits humains; certains traits sont essentiellement déterminés génétiquement, d'autres sont surtout influencés par l'environnement.* Ainsi Jensen (1969) a-t-il pu proclamer à l'aide de ses estimations de l'héritabilité du QI, que celui-ci ayant une très forte composante génétique et étant fortement corrélé aux résultats scolaires, les sommes

allouées au *Head Start Programme* pour les enfants noirs des Etats-Unis seraient dépensées en pure perte.

Il semblerait alors que les chercheurs aient réussi à réaliser le vieux rêve de l'humanité : pouvoir séparer ce qui est dû à la nature et ce qui est dû à la culture. Malheureusement cela n'est qu'une illusion, qu'une analyse plus poussée des résultats et des hypothèses à leur base vient réduire à néant. Nous ne détaillerons pas ici toutes les raisons qui vont conduire à rejeter cette approche (Vetta et Courgeau, 2003), mais indiquerons seulement les principales.

En premier lieu un certain nombre d'erreurs relevées dans les calculs faits par Fisher (1918) vient modérer l'utilisation incontrôlée de certains de ses résultats. Dès 1976, Vetta a montré que les corrélations calculées par Fisher entre apparentés, en particulier entre jumeaux n'étaient pas utilisables dans le cas d'unions consanguines, qui sont non négligeables entre humains. De même la méthode proposée par Jinks et Fulker (1970), en vue d'utiliser le modèle de Fisher pour analyser les comportements, qui est sans aucun doute l'article le plus cité par les généticiens du comportement, comporte des formules incorrectes qui la rendent inutilisable (Vetta, 1976).

Egalement, comme nous l'avons dit, Fisher a émis ses hypothèses sans rien savoir du génome. L'hypothèse de l'existence de polygènes implique un génome constitué de plusieurs centaines de milliers de gènes. Or le décryptage du génome humain conduit à un nombre de gènes compris entre 25 et 28 000 : cela rend cette hypothèse peu vraisemblable, car les caractéristiques humaines que la génétique du comportement pense pouvoir étudier sont innombrables (fécondité, nuptialité, longévité, intelligence, personnalité, homosexualité, alcoolisme, féminité, autisme, dépression maniaque, agression, bonheur, raisonnement spatial ou verbal, comportement criminel, obésité, etc.). Mais surtout, il ignorait que les gènes sont regroupés sur 23 paires de chromosomes : lors de la méiose, deux caractères subissent des ségrégations soit indépendantes, s'ils sont sous la dépendance de gènes situés sur deux paires de chromosomes distincts, soit totalement liés, s'ils sont sous la dépendance de gènes situés sur la même paire de chromosomes. En réalité, il peut y avoir des échanges entre deux chromatides et des recombinaisons de gènes sont possibles. On ne peut donc plus considérer que les polygènes soient soumis, indépendamment les uns des autres, à la ségrégation : la quantification de leur transmission devient impossible.

D'ailleurs, Fisher lui-même (1951) donne son opinion sur ce « coefficient d'héritabilité, que je considère comme l'un de ces raccourcis malheureux, qui apparaissent souvent en biométrie du fait d'un manque d'analyse plus approfondie des données¹² ». On peut regretter que ses craintes aient été parfaitement justifiées par l'utilisation qu'en font les généticiens du comportement.

Nous arrivons maintenant à l'argument le plus fort, à notre avis, de l'impossibilité de séparer les effets des gènes et de l'environnement. Gottlieb (2001) indique clairement qu'« il est maintenant reconnu que gènes et environnements sont simultanément impliqués dans tous les traits et qu'il est impossible de spécifier leur poids spécifique ou leur influence quantitative sur n'importe quel trait », et il ajoute que « Cela a été une bataille scientifique gagnée de haute lutte, qui n'a pas encore réussi à s'imposer auprès de l'ensemble de l'humanité¹³ ».

¹² Co-efficient of heritability, which I regard as one of those unfortunate short-cuts, which have often emerged in biometry for a lack of more thorough analysis of the data

¹³ It is now known that both genes and environments are involved in all traits and that it is not possible to specify their weighting or quantitative influence on any trait. This has been a hard-won scientific insight that had not yet percolated to the mass of humanity.

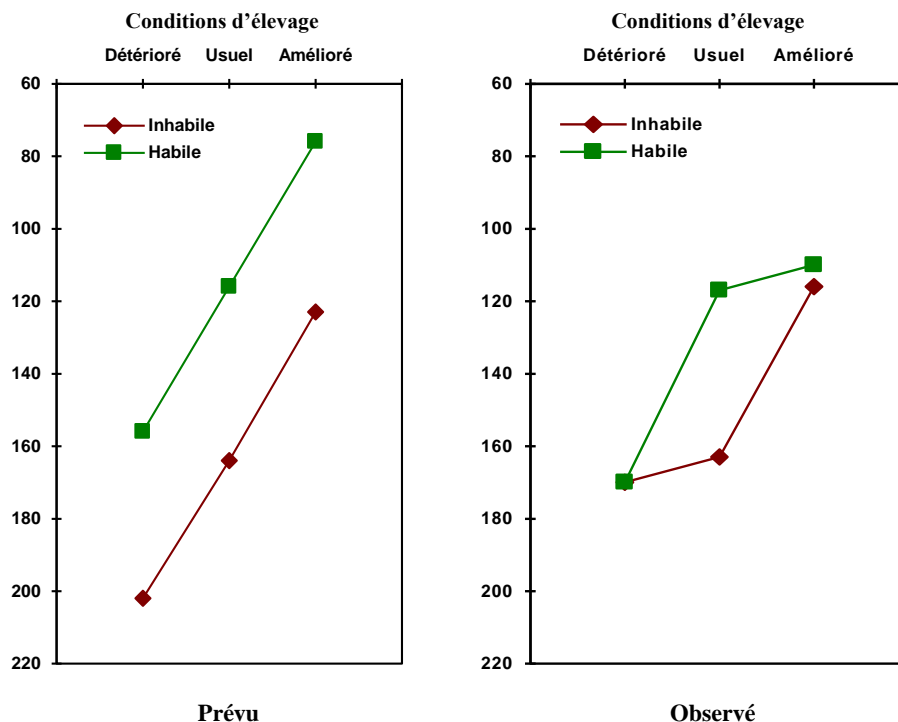
Pour démontrer ce résultat, il s'appuie sur une étude réalisée sur des rats car, pour des raisons d'éthique, elle n'est pas réalisable sur des hommes. Il est en premier lieu facile de sélectionner deux lignées de rats selon leur possibilité d'apprendre le meilleur cheminement à suivre dans un labyrinthe. On obtient ainsi une lignée « habile » et une autre lignée « inhabile » à trouver son chemin. Cette expérience est faite dans un environnement usuel.

Supposons maintenant que l'on change cet environnement, soit en l'améliorant, soit au contraire en le dégradant.

Sous les conditions d'un meilleur environnement, on peut penser que l'apprentissage des deux lignées va s'améliorer, mais on peut se demander si les différences entre elles resteront identiques. De même, sous les conditions d'un environnement peu favorable, leur apprentissage va se détériorer, mais on peut à nouveau se demander si les différences entre les deux lignées resteront identiques. Si l'on suppose que l'effet des gènes est indépendant de celui de l'environnement, alors les différences entre les deux lignées devraient rester identiques quel que soit l'environnement (voir figure 1, graphique de gauche). Sinon, cette expérience montrerait qu'il n'est pas possible de séparer un effet des gènes d'un effet de l'environnement.

L'expérience (voir figure 1, graphique de droite) nous apprend que, si l'on observe bien une amélioration ou une détérioration de l'apprentissage selon ces environnements, les différences entre les lignées vont disparaître entièrement : les rats inhabiles vont faire aussi peu d'erreurs que les rats habiles sous l'environnement amélioré et les rats habiles vont faire autant d'erreurs que les inhabiles sous l'environnement dégradé. Seuls ceux qui restent sous les conditions initiales gardent leurs différences.

Figure 1
Résultats prévus et observés sur deux lignées de rats élevés dans des environnements variables



Source : Gottlieb, 2001.

Ainsi, une sélection effectuée dans un environnement donné ne permet en rien de prédire le résultat d'une épreuve semblable dans un autre environnement. Le résultat de cette sélection est une propriété émergente, conduisant à des non-linéarités dans les effets observés.

Le rêve d'arriver à séparer la nature de la culture par un simple modèle additif n'est donc pas réalisable. Il faut bien voir que ces résultats incorrects ont souvent été utilisés à des fins politiques, pour soutenir une soi disant infériorité génétique de certains groupes ethniques, ou pour fournir une base pseudo-scientifique à des mesures politiques en vue d'améliorer l'espèce humaine. De nouvelles études en relation avec le décryptage du génome humain et traitant des prétendus liens entre gènes et comportements (Blum *et al.*, 1990 ; Rodgers *et al.*, 2001) ont été salués comme annonçant la naissance d'une « nouvelle génétique », mais leurs résultats sont toujours contestés par d'autres études (Sarkar, 1998). Le modèle simpliste sous-jacent à ces études est toujours incorrect.

Il est dès lors nécessaire de changer de paradigme. On peut maintenant dire que *les gènes et l'environnement sont simultanément impliqués dans tout comportement et qu'il est impossible de spécifier leur poids respectif ou leur influence quantitative sur quelque trait que ce soit. C'est cette interaction qu'il est maintenant nécessaire d'étudier*¹⁴. Nous ne poursuivrons pas plus avant ici dans l'étude de cette interaction, car ce n'est pas l'objet de cet article.

Si nous venons de montrer qu'il est impossible de mesurer de façon additive la part du génétique et la part de l'environnement dans la production d'un comportement donné, il nous faut voir plus en détail la façon dont la cumulativité des connaissances sur l'environnement d'une part, et des connaissances biologiques d'autre part, peut alors être envisagée.

Le premier exemple nous avait montré que le passage d'une recherche de la régularité des taux à celle de leur variation entraînait une vue plus complète de la démographie. De la même façon, nous passons ici d'un modèle additif à un modèle non additif, qui pourrait entraîner une vue plus complète des liens entre démographie et génétique des populations. En effet c'est le projet de Fisher de généraliser le modèle additif, parfaitement applicable aux caractères discrets sous la dépendance d'un petit nombre de gènes, aux caractères quantitatifs continus sous la dépendance d'un grand nombre de gènes, considéré comme infini, que nous avons critiqué. Mais on peut maintenant considérer le nouveau modèle non additif, comme une généralisation du modèle additif précédent, qui cette fois-ci peut être dans certains cas particuliers additif et dans la majorité des autres cas non additif. Ainsi nous pouvons dire que ce nouveau paradigme permet une vue plus complète des interactions entre démographie et génétique des populations.

Troisième exemple : l'introduction de l'analyse biographique

Le troisième exemple que nous allons traiter ici, consiste en l'apparition d'une nouvelle façon de voir les phénomènes démographiques qui va introduire un nouveau paradigme.

Depuis la fin de la Seconde Guerre mondiale, l'analyse longitudinale s'était peu à peu imposée comme plus apte à l'étude d'un phénomène démographique, que l'analyse transversale. En effet, la seconde approche étudiait les événements d'une année, ou d'une certaine période, en regroupant les résultats obtenus dans les diverses cohortes comme ceux

¹⁴ Voir par exemple Lewis et Lewontin, 1985, Gottlieb 1992, pour diverses approches de ce nouveau paradigme.

d'une cohorte fictive soumise de bout en bout aux conditions de l'année ou de la période considérée. Cette fiction soulevait de sérieuses objections en particulier pour l'interprétation d'indices synthétiques supérieurs à l'unité. L'analyse longitudinale évite ces objections et fournit des résultats faciles à interpréter (Courgeau, 2004). En France, Louis Henry en a été le principal théoricien en posant très clairement les bases sur lesquelles se fondait cette analyse (Henry, 1959). Dans les pays anglophones Welpton (1945) et Ryder (1951, 1965) en ont été les principaux instigateurs.

Si l'on veut étudier un événement démographique, tel que le mariage, cette observation, ici la proportion des adultes d'une certaine génération qui se sont mariés, est manifestement un mélange : une mort prématurée a empêché le mariage d'une fraction des adultes de la génération de même qu'une émigration hors de la zone d'étude, une immigration dans cette zone y amène des individus ayant peut-être un risque de se marier différent de celui des sédentaires, etc. Ces effets sont parasites dans une étude du phénomène et empêchent en particulier toute comparaison dans le temps ou l'espace de populations dans lesquelles ces phénomènes perturbateurs sont différents. Il faut donc s'efforcer de séparer l'effet des différents phénomènes et d'isoler le phénomène étudié, à l'état pur (Henry, 1959), exempt des perturbations provoquées par des événements liés à d'autres phénomènes.

Dès lors, le paradigme d'une analyse des phénomènes à l'état pur peut être approché par le postulat suivant : « le démographe ne peut étudier que l'arrivée d'un événement, et d'un seul, dans une sous-population qui conserve tous ses caractères et les mêmes caractères tant que le phénomène se manifeste » (Blayo, 1995). Pour que ce postulat soit vérifié il est nécessaire qu'il y ait d'abord *indépendance* entre le phénomène étudié et les phénomènes perturbateurs. Une seconde condition est également nécessaire : il faut que la cohorte étudiée soit *homogène*, c'est-à-dire que les probabilités de connaître les événements étudiés et perturbateurs soient les mêmes, à chaque âge, pour tous les membres de la cohorte. Les données utilisées pour effectuer une telle analyse sont essentiellement des données de l'état civil. L'ouvrage de Henry (1972) en constitue la parfaite exposition, où les différents phénomènes démographiques, sont traités chacun dans un chapitre séparé : nuptialité, fécondité, mortalité, déplacements et migrations.

Dès le départ Henry (1959) constatait que ces conditions n'étaient sans doute pas remplies :

Etant donné les différences de tous ordres entre les hommes, on peut être certain qu'aucun groupe humain n'est homogène ; l'observation courante et la réflexion nous incitent, d'autre part, à penser qu'il n'y a pas, le plus souvent, indépendance entre les risques.

Il constate cependant qu'en pratique les procédés courants sont à peu près acceptables pour l'étude d'un certain nombre de phénomènes démographiques. Il préconise cependant de :

pousser les recherches de démographie différentielle jusqu'aux caractéristiques individuelles, physiques et psychologiques, avec le souci d'étudier à la fois la dispersion et la corrélation des indices démographiques à l'intérieur des groupes, assez sommaires considérés jusqu'ici.

En dépit de ces recommandations l'analyse longitudinale a prévalu pendant une assez longue période, s'étendant de la fin de la Seconde Guerre mondiale à la fin des années 1970, et peu de chercheurs ont essayé d'approfondir ces études de démographie différentielle.

Il faut voir finalement que c'est une approche bien différente, l'analyse des biographies, et l'utilisation de données d'enquêtes qui ont permis de lever les conditions trop restrictives de cette analyse. En effet, ces données d'enquête permettaient de relier les divers

phénomènes démographiques que connaissait un individu, contrairement aux données d'état civil qui recueillaient séparément ces divers phénomènes. Nous avons ainsi pu montrer (Courgeau, 1976) l'existence de liens importants, tant entre mobilité et nuptialité, qu'entre mobilité et fécondité, rendant l'hypothèse d'indépendance entre phénomènes absolument non vérifiée.

L'analyse à proprement parler des biographies est dès lors née de la nécessité de donner une base théorique solide à l'étude des événements qui surviennent tout au long de la vie des individus. Une première base se trouve dans l'article de Cox (1972) et a été utilisée principalement dans des études médicales, sur des petits échantillons de malades, et dans des études industrielles de fiabilité. Ces méthodes ont commencé à être utilisées en démographie au début des années 1980 pour l'analyse d'un phénomène, la nuptialité (Menken et Trussell, 1981), puis pour l'analyse plus complète de biographies familiale, migratoire et professionnelle (Courgeau, 1982).

En dépit d'un nombre croissant de travaux utilisant ces méthodes, elles ont été longtemps considérées avec suspicion par les démographes français. Dans le numéro de *Population* destiné à célébrer le cinquantième anniversaire de la revue, un article (Blayo, 1995) résume les attaques contre ces méthodes en disant que la condition d'homogénéité doit toujours être remplie et que la condition d'indépendance doit toujours être posée. Elle conclut son article en disant que l'« analyste ne se contente pas d'appliquer une méthode, il réfléchit à ses conditions d'applications ; c'est ce qui le distingue du technicien ».

Après la lecture de cet article, le démographe Jan Hoem, éditeur du volume 15 *Demography* de l'*International Encyclopedia of the Social and Behavioral Sciences* (2001), nous écrivait dans un courrier personnel :

L'attaque inutile de Chantal Blayo n'a aucune base théorique en statistique. Il y a peu de nouveauté dans cet article et il n'aurait pas dû être publié dans un numéro spécial de la revue *Population* qui était censé nous rappeler la qualité de la recherche à l'INED. Je suppose que vous allez écrire une réponse et j'espère qu'en toute loyauté *Population* la publiera rapidement. Pour nous autres, les manières de penser que Mme Blayo présente dans la section IV de son article peuvent sans peine être ignorées. Il est triste de voir un démographe refuser d'accepter des développements méthodologiques et continuer de traiter de questions que les générations antérieures ont déjà affrontées et résolues, mais je doute que personne ne puisse empêcher cela¹⁵.

Nous avons en effet pu répondre à cette attaque dans *Population* (Courgeau et Lelièvre, 1996, 1997) en indiquant qu'il s'agit bien d'un changement de paradigme entraîné par cette analyse biographique. Nous allons en présenter ici les grandes lignes.

L'analyse longitudinale rencontrait dès la fin des années 1970 un certain nombre de difficultés, en particulier du fait de l'utilisation croissante de données d'enquêtes biographiques. Or cette analyse ne permet de suivre qu'une sous-population dans laquelle entrent des individus et de laquelle d'autres individus sortent. Un tel point de vue revient à nier toute spécificité de la vie individuelle, pour ne s'attacher qu'à l'arrivée d'un événement

¹⁵ Chantal Blayo's needless attack has no foundation in statistical theory. There is little novelty in the article and it should not have been published in a volume meant to remind us of the grandeur of research at INED. I suppose that you need to write a rejoinder and assume that in all fairness *Population* will publish it soon. For the rest of us, the attitudes that Mme Blayo presents in section IV of her paper may as well be ignored. It is sad when a demographer refuses to accept methodological developments and continues to address issues that earlier generations grappled with and solved, but I doubt that there is much one can do with it.

dans une sous-population qui demeure identique au cours du temps car constituée d'unités interchangeables.

De plus pour que cette sous-population garde son homogénéité on doit supposer que les nouveaux arrivants ont immédiatement tous ses caractères et vont se comporter de la même façon que les sédentaires. Cela revient à poser une hypothèse markovienne : l'étape actuelle n'est pas influencée par l'histoire passée des individus. Or des résultats d'enquêtes viennent montrer qu'il n'en est rien (Courgeau, 1979).

Le fait reconnu qu'une population n'est pas homogène, conduit à décomposer la population initialement étudiée en sous-populations plus homogènes. Ainsi, dans l'étude du mariage, il semble utile de distinguer la nuptialité des ruraux de celle de l'ensemble de la population car elle est très différente. Dans ce cas un nouveau phénomène perturbateur intervient : le départ du rural de célibataires. Ce phénomène touche un bien plus grand nombre d'individus, aux âges du mariage, que la mortalité et la migration internationale. Qui plus est ces individus risquent d'avoir un comportement très différent vis-à-vis du mariage que les sédentaires. Ensuite, les exploitants agricoles ont un comportement différent des ouvriers agricoles pour ce qui est de la nuptialité et ont également des risques différents de sortir de la population rurale. Cette sous-population rurale n'est donc pas homogène et il est nécessaire de la scinder plus avant.

Il conviendrait, dès lors, de scinder chaque sous-population en un nombre toujours croissant de sous-populations pour chercher à les rendre toutes homogènes. On va vite se trouver en présence de groupes trop restreints, dont le faible effectif va empêcher toute analyse sérieuse. Qui plus est on ne sera jamais certain d'avoir pris en compte tous les facteurs d'hétérogénéité de la population.

Pour ce qui est de la condition d'indépendance entre les divers phénomènes démographiques, la situation est aussi délicate avec les nouvelles données d'enquêtes. Il apparaît clairement que ces phénomènes sont intimement reliés entre eux (Courgeau, 1976) et qu'il n'est plus possible de les considérer indépendamment les uns des autres. De plus, comme le paradigme utilisé ne permet l'étude que d'un événement et d'un seul, on se trouve dans l'impossibilité d'étudier les sorties par événements compétitifs : les études de mortalité par cause, l'étude de la sortie du célibat par mariage ou par cohabitation, etc., ne sont pas permises.

Il en résulte que le paradigme longitudinal, du fait des conditions d'homogénéité de la population et d'indépendance entre les divers événements démographiques, empêche toute possibilité d'analyse plus poussée des données d'enquêtes : il nécessite des décompositions tellement fines de la population étudiée, qu'il invalide tout calcul ; il nécessite de telles contraintes sur les événements étudiés qu'il interdit tout un pan de l'analyse démographique (analyse d'événements compétitifs, analyse d'événements en interaction, analyse d'événements d'une population dans laquelle plusieurs événements permettent d'entrer ou de sortir). Il est dès lors nécessaire de remettre en cause ce paradigme et d'en poser un nouveau pour l'analyse biographique.

Le changement de point de vue opéré sera donc fondamental : l'intérêt ne va plus se porter sur l'étude de sous-populations homogènes, mais sur celle de trajectoires individuelles entre un nombre quelconque d'états. L'unité d'analyse ne sera plus l'événement, comme en longitudinal, mais la biographie individuelle, considérée comme un processus stochastique complexe. Le nouveau paradigme peut dans ce cas être approché par le point de vue suivant : *un individu parcourt, tout au long de sa vie, une trajectoire complexe, qui dépend, à un instant donné, de sa trajectoire antérieure et des informations qu'il a pu acquérir dans son*

passé. Ce changement d'optique va lever les conditions d'homogénéité de la population sur laquelle on travaille, et d'indépendance entre les phénomènes de l'analyse longitudinale.

En suivant au cours du temps les événements de la vie d'un individu, on voit que la principale manière pour lui d'échapper à l'observation sera de sortir de l'échantillon à la date de l'enquête ou à la date de l'étude, si l'on travaille sur des données de registre. Dans la mesure où ces dates n'ont aucune raison d'être liées à la vie d'un individu, la condition d'indépendance est, dans ce cas, parfaitement vérifiée : l'observation est dite non-informative et l'on sait tenir compte de ces sorties pour estimer les quotients. En revanche, il n'y a plus aucune raison que les divers événements rencontrés par l'individu soient indépendants les uns des autres : c'est la *dépendance* entre ces événements qui constituera l'essentiel de l'analyse des biographies.

Il importe de distinguer maintenant les événements en interaction, que l'analyse précédente appelait événements perturbateurs, des événements compétitifs ou concurrents. Un événement en interaction va modifier la probabilité de survenue de l'événement étudié : la probabilité de mariage d'un individu, qui a commencé sa vie comme agriculteur, va dépendre du fait qu'il soit sorti ou non de ce statut initial (Courgeau et Lelièvre, 1986). On peut, dans ce cas, tester cette modification du comportement en le comparant à celui des agriculteurs du même âge, ou à celui de la population qui n'a jamais connu le statut d'agriculteur. Cela permet de conclure sur une éventuelle sélection des individus ou, au contraire, une adaptation des comportements. Quand on parle de phénomènes compétitifs ou concurrents, il s'agit des diverses modalités d'un événement qui conduisent au même état final : mortalité par cause, mise en union par mariage ou cohabitation, etc.

L'analyse des interactions entre événements démographiques et celle des événements compétitifs est dès lors tout à fait réalisable en démographie (Courgeau et Lelièvre, 1989, 1992, 2001).

De même, lorsque l'on cherche à comprendre les comportements d'un individu, il faudra faire intervenir ses origines sociales et toute son histoire passée. Ces comportements ne sont pas innés, mais se modifient au cours de l'existence individuelle grâce aux expériences personnelles et aux acquis successifs. C'est donc bien une analyse de l'*hétérogénéité* des populations, que réalise l'analyse des biographies, vue sous forme dynamique et non plus statique. Elle permet en effet de faire une régression, non plus sur des données du moment, mais sur des processus temporels complexes. Cette généralisation a été l'œuvre de statisticiens (Cox, 1972, Aalen, 1978) et s'est révélée particulièrement adaptée à l'étude des phénomènes démographiques¹⁶. Elle conduit enfin à une combinaison, non analytique comme l'*équation fondamentale de la dynamique des populations* (Bourgeois-Pichat, 1994), mais tout autant valable, sous forme de modèles de micro-simulation (Billari et Prskawetz, 2003).

Comme on le voit facilement, ce changement de paradigme introduit une cumulativité des résultats en démographie, qui généralise celle d'une cumulativité monotone. Certains résultats peuvent rester approximativement les mêmes, lorsque les populations peuvent être considérées comme à peu près homogènes et les phénomènes comme à peu près indépendants, en analyse longitudinale et en analyse biographique. En revanche de nombreux résultats entièrement nouveaux peuvent être montrés en analyse biographique, qui ne pouvaient même pas être envisagés en analyse longitudinale, comme l'interaction entre phénomènes. On peut dire dans ce cas que l'analyse biographique, qui reconnaît parfaitement la possibilité d'indépendance entre phénomènes, cumule cette condition avec la possibilité simultanée que ces phénomènes puissent également être dépendants les uns des autres. De

¹⁶ Voir Menken et Trussell, 1981 ; Courgeau et Lelièvre, 1989, 1992, 2001.

même cette analyse reconnaît parfaitement la possibilité d'une population homogène, mais permet de généraliser l'analyse au cas d'une population hétérogène.

Quatrième exemple : synthèse multiniveau

Ce dernier exemple nous conduira à une synthèse de certaines des approches suivies dans les chapitres précédents. Nous les résumerons donc ici rapidement, pour montrer comment la démographie a évolué au cours du temps et l'étape, sans doute provisoire, à laquelle conduit la synthèse multiniveau (Courgeau, 2002, 2004, 2007).

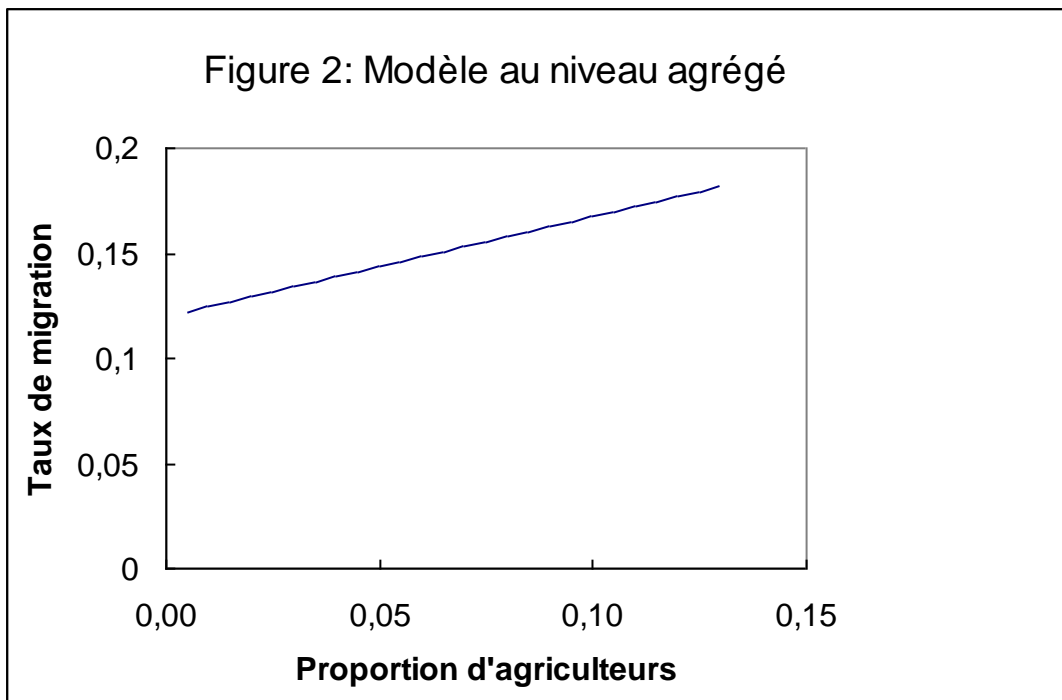
En effet l'un des problèmes majeurs de la démographie et de bien d'autres disciplines, voir la communication d'Alban Bouvier par exemple, est celui du passage d'un niveau d'agrégation à un autre, de l'individu aux divers groupes sociaux, politiques, géographiques, etc. Le changement de niveau exige un changement de regard sur les phénomènes étudiés et l'utilisation d'un nouveau paradigme, au sens que nous donnons à ce terme. Pour traiter ce problème le plus clairement possible, nous suivons simultanément le même exemple simple vu sous différents niveaux et divers paradigmes pour montrer ce qui les distingue les uns des autres et les diverses images qu'ils montrent d'une même réalité. Cet exemple considère les migrations entre les régions norvégiennes observées par le registre de population de ce pays¹⁷, effectuées par les hommes de la génération née en 1948, sur une courte période de deux ans suivant le recensement de 1970.

Comme nous l'avons indiqué dans le premier exemple, Durkheim (1895) proposait un paradigme transversal pour définir une science des faits sociaux. En particulier, il mettait en évidence la fonction d'un fait social en le liant à d'autres faits sociaux. Pour ce faire il proposait d'utiliser la méthode des variations concomitantes, appelée de nos jours méthode de régression linéaire. Ainsi, pour mettre en évidence les liens entre taux de suicide et pourcentage de protestants, il classait les provinces de Prusse observées entre 1883 et 1890 selon leurs pourcentages de protestants et montrait que « le suicide varie en raison inverse du degré d'intégration de la société religieuse » (Durkheim, 1897), l'Eglise protestante étant considérée comme « moins fortement intégrée que l'Eglise catholique ». Bien entendu son propos est bien plus général et complexe, mais il n'entre pas dans le sujet de cet article.

De la même façon il est possible de classer les 19 régions norvégiennes selon leur proportion d'agriculteurs et de réaliser une régression entre les taux de migration de chaque région et leurs proportions d'agriculteurs. La figure 2 porte les résultats de ce modèle.

Figure 2 : Modèle au niveau agrégé, indiquant la variation du pourcentage de migrants en Norvège, lorsque la proportion d'agriculteurs présents dans les régions augmente

¹⁷ Nous remercions ici les services statistiques norvégiens qui nous ont permis d'avoir accès à des fichiers créés à partir de ces données et des recensements de population par Kjetil Sørli et Øjsten Kravdal.



Les paramètres estimés pour cette régression permettent de calculer une probabilité de migrer égale à 0,597 pour les agriculteurs et 0,119 pour les autres professions, soit près de six fois plus forte pour les agriculteurs que pour les autres, si les hypothèses du modèle sont vérifiées. Ce résultat peut paraître surprenant étant donné le coût tant financier que personnel d'un changement de région pour cette profession.

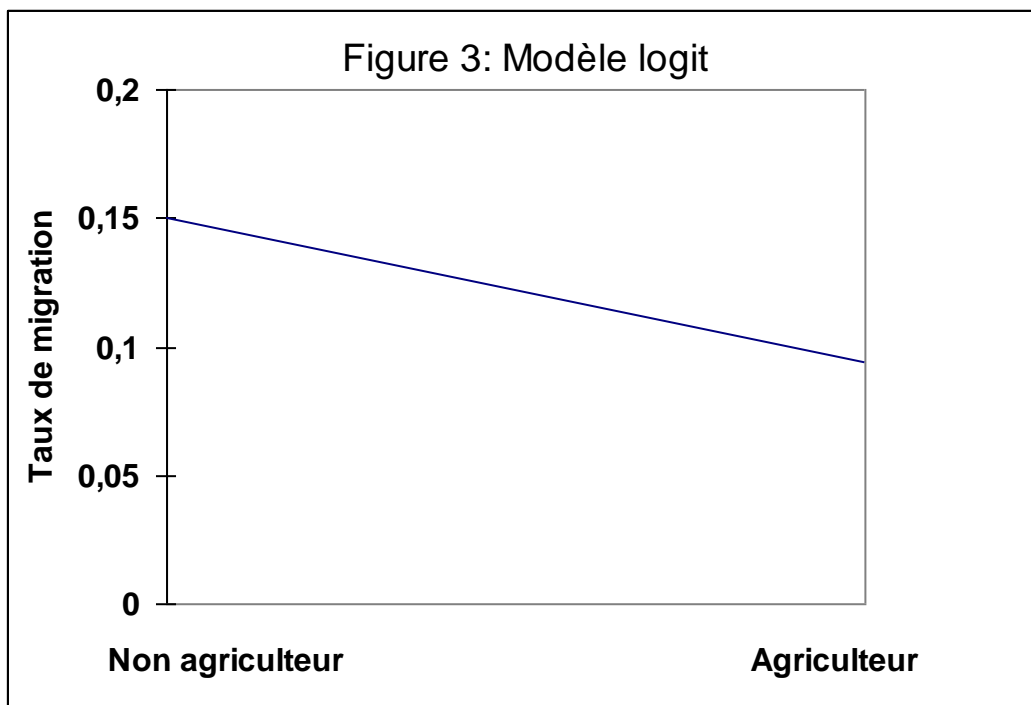
L'utilisation de ce paradigme conduit à un certain nombre de problèmes méthodologiques (Courgeau, 2004). Nous ne les détaillerons pas tous ici, mais insisterons sur l'*erreur écologique* à laquelle ce paradigme peut conduire. Un tel problème a été mis en évidence il y a plus de cinquante ans maintenant (Robinson, 1950). A cette époque de nombreux chercheurs travaillaient sur des données agrégées pour minimiser les problèmes de coût d'exploitation ou tout simplement parce que le plus souvent les données n'étaient disponibles que sous forme agrégée. Ils pensaient que les résultats obtenus étaient directement interprétables en termes de comportements individuels. Robinson a clairement montré que l'on ne pouvait utiliser une corrélation écologique calculée au niveau agrégé comme substitut d'une corrélation au niveau individuel.

Ainsi, la liaison positive entre le pourcentage d'agriculteurs et le pourcentage de migrants, trouvée plus haut, montre seulement que la plus forte probabilité de migrer est liée à une forte proportion d'agriculteurs sans que l'on puisse dire si ce sont les agriculteurs ou les non agriculteurs de ces régions qui migrent le plus. Seule l'hypothèse selon laquelle les faits sociaux ont une existence indépendante des individus qui les vivent vient appuyer l'interprétation du rôle prépondérant de la profession sur la migration : dans ce cas ce sont bien les agriculteurs des diverses régions qui ont une plus forte probabilité de migrer. Mais, il peut être tout à fait possible qu'un fort pourcentage d'agriculteurs, en limitant les offres d'emploi aux autres professions, contraignent les non agriculteurs à migrer pour trouver un meilleur emploi, alors que les agriculteurs eux-mêmes n'ont aucune raison de migrer plus hors de ces zones. Dans ce cas l'interprétation précédente ne vaut plus et il est nécessaire de chercher ailleurs une meilleure explication des comportements observés.

Le paradigme biographique, proposé plus tard, permet de vérifier au niveau individuel si les agriculteurs ont une plus forte probabilité de migrer que les non agriculteurs. En effet dans le cas précédent il nous suffisait de connaître les distributions marginales des événements étudiés pour pouvoir estimer, sous certaines hypothèses, la liaison recherchée. Maintenant, nous allons essayer d'estimer la distribution jointe de toutes les combinaisons de variables, au niveau individuel, et de relier ainsi de façon différente l'arrivée d'un événement aux caractéristiques d'un individu. Bien entendu cela nécessite de disposer de données individuelles et non plus agrégées, mais cela permet une estimation plus claire des effets sur les agriculteurs et les non agriculteurs. On arrive ainsi à une régression logistique qui va préfigurer une vraie analyse biographique.

L'estimation d'un tel modèle, en utilisant toujours les mêmes données norvégiennes, conduit à une probabilité de migrer de 0,095 pour les agriculteurs et de 0,150 pour les non agriculteurs. Ces résultats sont reportés sur la figure 3. Ce résultat au niveau individuel vient donc parfaitement contredire celui du modèle précédent au niveau agrégé. Il est plus proche de ce que l'on pouvait attendre des constatations de coût plus élevé d'une migration pour les agriculteurs. Nous verrons cependant plus loin comment comprendre ces résultats.

Figure 3 : Modèle logit, indiquant la probabilité de migrer des agriculteurs et des non agriculteurs norvégiens



Il est bien entendu possible de le généraliser à l'aide d'une régression temporelle, mais les données dont nous disposons ne le permettent pas et là n'est pas notre problème. Il se trouve dans la contradiction obtenue sous le paradigme transversal et sous le paradigme biographique. Il est dès lors utile de voir plus précisément les questions que pose la confrontation de ces deux approches pour montrer comment une analyse en niveaux permet de les surmonter.

A nouveau, nous ne présenterons pas le détail de ces questions, renvoyant le lecteur intéressé à Courgeau (2004, 2007). Nous nous intéresserons plus précisément ici au risque d'*erreur atomiste* auquel cette approche peut conduire. En effet l'approche biographique fait

essentiellement intervenir pour expliquer les comportements individuels, les caractéristiques de l'individu à la fois internes et externes à la démographie, aussi bien son âge et son sexe que sa carrière professionnelle. On ignore alors le contexte plus général dans lequel les conduites humaines se produisent. Ce contexte peut être défini comme le milieu familial dans lequel l'individu vit, ou plus généralement comme un entourage plus ou moins étendu autour de l'individu : son quartier, sa commune, etc. On peut penser que ce contexte peut jouer sur les comportements individuels et il paraît fallacieux d'isoler l'individu des contraintes imposées par la société et le milieu dans lequel il vit.

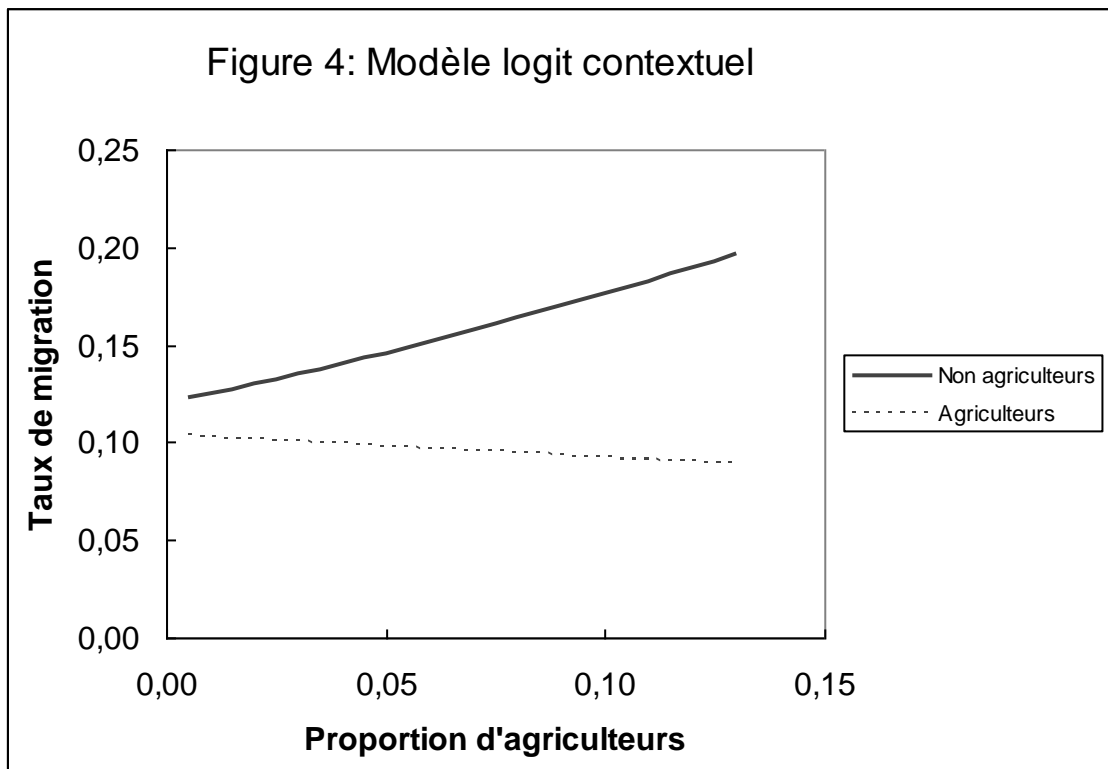
Ce risque d'erreur, que l'on peut opposer à l'erreur écologique, avait déjà été reconnu par des sociologues (Lazarsfeld et Menzel, 1961). Ils montraient la nécessité de définir avec suffisamment de précision divers types de groupes, de communautés, d'organisations et plus généralement d'ensembles quelconques d'individus, ayant un rôle social. Ceux-ci peuvent être composés de membres ayant une attitude d'ensemble vis-à-vis du comportement étudié et peuvent être décrits par un certain nombre de caractéristiques et de propriétés. Ce qui apparaît comme un groupe dans une étude peut être considéré comme un membre d'un groupement plus général dans une autre étude. Cette propriété est très importante car elle montre le caractère relatif de l'individu que l'approche biographique considère comme l'unité prépondérante. En considérant une pluralité de niveaux, nous pouvons quitter l'approche dualiste qui oppose l'individu à la société. Dans ce cas, « il n'y a plus de sens à choisir entre holisme et atomisme, et pour ce qui est des sciences sociales, entre holisme et individualisme » (Franck, 1995).

L'approche contextuelle (Coleman, 1986), complétée par une approche multiniveau plus générale (Goldstein, 2003 ; Courgeau, 2003), permet, comme on va le voir, de résoudre certains de ces problèmes.

En effet la première approche permet d'échapper à ces difficultés en cherchant à prédire un comportement individuel à l'aide à la fois de caractéristiques individuelles et de caractéristiques du groupe auquel ces individus font partie. Cela permet d'inclure différents types de mesure d'une caractéristique, considérée soit au niveau individuel, soit au niveau du groupe. Ainsi, par exemple, le comportement de chaque individu peut être associé au fait qu'il ait une caractéristique donnée (mesure individuelle) et au fait que les groupes auxquels il participe aient des caractéristiques données (mesures agrégées), qui peuvent être calculées à partir de celles de leurs membres. Non seulement cette approche contextuelle permet de mettre en évidence les sources et l'importance du biais écologique, mais elle permet de séparer les effets individuels, contextuels et écologiques qui jouent sur le phénomène étudié.

Reprenons l'exemple précédent du comportement migratoire des agriculteurs et des autres professions en Norvège. Nous avons maintenant la possibilité de faire intervenir non seulement le fait que l'individu soit agriculteur sur sa probabilité de migrer, mais également le pourcentage d'agriculteurs présents dans la région où il vit. Un tel modèle logit contextuel permet de mettre en évidence les probabilités de migrer non seulement des agriculteurs et des non agriculteurs, mais également comment ces probabilités vont varier lorsque le pourcentage d'agriculteurs change (figure 4)

Figure 4 : Modèle logit contextuel, donnant les taux de migration des agriculteurs et des non agriculteurs, selon le pourcentage d'agriculteurs présents dans chaque région.



On voit maintenant que le fait d'être agriculteur réduit toujours, comme dans le modèle biographique, sa probabilité de migrer, alors que le fait de vivre dans une région où le pourcentage d'agriculteurs est important va seulement augmenter les chances de migrer d'un non agriculteur sans modifier celle d'un agriculteur. On voit en quoi ce résultat est différent de ce que donnait le modèle agrégé, dans lequel il n'y avait pas la possibilité de distinguer les agriculteurs des non agriculteurs. Ce résultat vient parfaitement conforter l'hypothèse que nous avons précédemment posée : une forte densité d'agriculteurs dans une zone va accroître les probabilités de migrer seulement des autres professions, sans modifier celles des agriculteurs. Cela pourrait s'expliquer, dans ces régions, par un manque relatif d'autres emplois qu'agricoles, qui va pousser les autres professions à en émigrer plus que les agriculteurs lorsqu'ils recherchent un nouveau travail.

Dans ce cas, on voit apparaître un nouveau type de cumulativité des résultats : les résultats contextuels permettent de combiner les résultats agrégés avec les résultats biographiques. Cependant cette combinaison conduit à une interprétation différente du résultat agrégé. Le fait que la probabilité de migrer de l'ensemble de la population augmente avec le pourcentage d'agriculteurs n'est pas le fait d'une augmentation de la probabilité de migrer des agriculteurs, comme l'interprétait le paradigme de l'analyse du moment, mais au contraire d'une augmentation de la probabilité de migrer des autres professions, comme permet de l'interpréter le paradigme multiniveau. On voit donc apparaître une cumulativité qui maintient bien certains des résultats obtenus précédemment sous un autre paradigme, mais change leur signification. Il ne s'agit pas ici d'éviter de fausses corrélations au niveau agrégé comme lorsqu'on recourait aux histoires de vie (voir le chapitre d'Alban Bouvier dans cet ouvrage), mais maintenant de restituer à chaque niveau son effet sur le comportement étudié.

L'analyse multiniveau, en introduisant des effets plus complexes que précédemment, ne va pas bouleverser le cadre de raisonnement suivi dans l'analyse biographique, mais va

permettre de le compléter. Le nouveau paradigme¹⁸ qui en résulte va donc toujours considérer que *le comportement d'un individu dépend de son histoire passée, vue dans toute sa complexité*, mais il faut ajouter, *que ce comportement peut également dépendre de contraintes extérieures exercées sur l'individu, que celui-ci en soit conscient ou non*. Ainsi l'individu pourra être pleinement conscient des contraintes imposées par la société dans laquelle il vit et agir en fonction d'elles, aussi bien pour y résister, que pour les contourner ou les utiliser, si elles peuvent améliorer sa situation. Au contraire une pression de la société qui l'entoure peut influencer sur ses comportements sans qu'il en soit pleinement conscient : sa présence dans un milieu où, par exemple, le chômage est important peut l'amener plus facilement à une migration éloignée que s'il était dans une zone de plein emploi.

Le nombre de niveaux considérés est fonction à la fois du phénomène étudié et des effectifs observés pour mettre en évidence ces contraintes. Cependant cette analyse permet en utilisant des méthodes bayésiennes de considérer des groupes ayant très peu de membres, parfois de moins de 10 individus, tout en ayant des estimations fiables.

Nous n'allons pas plus loin dans l'examen d'une analyse pleinement multiniveau, qui n'apporte plus d'éléments nouveaux à notre discussion sur la cumulativité, mais conduit à une approche plus satisfaisante où les comportements d'un individu peuvent dépendre des comportements des individus du même groupe¹⁹.

Conclusion

L'examen des quatre exemples nous permet maintenant de voir plus clairement ce qu'il en est de la cumulativité en démographie.

Le premier exemple considérait une donnée simple caractérisant la population, le rapport de masculinité à la naissance, mesurée de façon à peu près identique²⁰ depuis plus de 300 ans, mais dont l'interprétation a varié au cours du temps. D'abord considéré comme une constante, que seules des différences entre les mœurs et les modes de vie viennent perturber, c'est bien cette régularité qui est primordiale : si les mœurs et les modes de vie étaient partout semblables, comme l'est la nature, le rapport de masculinité à la naissance le serait également. Les variations minimales observées alors selon les lieux s'estompent devant la grande régularité constatée. Par la suite, du fait que l'importance des variations réellement observées apparaît trop large, c'est au contraire sur les différences que les chercheurs vont se pencher. Il s'agit maintenant, non seulement d'expliquer ces différences, mais même de découvrir un moyen de produire un enfant mâle ou femelle selon la volonté des parents. Les travaux récents montrent l'importance de phénomènes de discrimination très précoces selon le sexe (Brian et Jaisson, 2007). Ainsi, bien que la mesure du phénomène reste parfaitement identique, avec une amélioration de la qualité des données et une observation plus générale, l'interprétation qu'en donnent différentes approches peut en être très différente.

Le deuxième exemple envisageait la possibilité de séparer la nature de la culture et de les traiter de façon distincte à l'aide d'un modèle simple. L'additivité serait alors possible

¹⁸ Nous pouvons dire qu'il s'agit bien d'un nouveau paradigme plus général car il conduit à un nouvel objet d'analyse, différent de celui de l'analyse biographique : il permet de prendre en compte le passage d'un niveau à un autre. Cette différence de point de vue avec Bouvier (2006) vient en partie de la définition que nous avons donnée au terme paradigme. Il s'agit cependant bien d'un des problèmes majeurs des sciences sociales.

¹⁹ Nous renvoyons le lecteur intéressé à Courgeau, 2003, 2004, 2007 et à Goldstein, 2003.

²⁰ Le nombre de garçons pour 100 filles, utilisé au XVIIIe siècle, s'en déduit par une fonction mathématique simple.

dans deux domaines que l'on pourrait alors considérer comme indépendants l'un de l'autre : la génétique de population, d'une part, les sciences sociales, d'autre part. Les hypothèses qui permettent une telle distinction, même si elles sont encore considérées par certains chercheurs comme valables, sont en fait parfaitement contredites par l'examen plus approfondi des bases de cette approche et par l'observation de populations sur lesquelles l'expérience est possible, contrairement aux populations humaines. Il est maintenant reconnu par les milieux scientifiques que, si les gènes et l'environnement sont simultanément impliqués dans tous les comportements humains, il est impossible de séparer leur influence quantitative sur tout comportement que ce soit. L'additivité des effets de la nature et de la culture, n'est pas vérifiée et il faudra faire intervenir simultanément ces effets de façon non additive pour raisonner correctement.

Le troisième exemple montrait en détail le passage d'un paradigme au suivant, les difficultés et les réticences des utilisateurs du paradigme antérieur à accepter la validité des résultats du second. L'analyse longitudinale posait des conditions restrictives à son utilisation : elle ne pouvait analyser que des populations homogènes et des phénomènes indépendants les uns des autres. Elle était soutenue par la pauvreté des données qu'elle utilisait. Si les statistiques de l'état civil encourageaient l'utilisation de ces hypothèses, dont elles ne permettaient pas de vérifier l'exactitude, dès que l'on utilise des données beaucoup plus riches d'enquêtes ces hypothèses paraissent bien trop restrictives. Les populations sur lesquelles on travaille sont en fait très souvent hétérogènes et les phénomènes que l'on étudie sont intimement dépendants les uns des autres. Dans ces conditions un nouveau type de cumulativité peut apparaître entre les résultats d'une analyse longitudinale et d'une analyse biographique : si les populations sont réellement homogènes vis-à-vis du phénomène étudié et si ce phénomène est indépendant des autres phénomènes qui se produisent en même temps que lui, le paradigme longitudinal est toujours vérifié. En revanche, lorsque l'une, l'autre ou les deux conditions ne sont pas vérifiées le nouveau paradigme permet d'intégrer l'étude de ce cas sans poser aucun problème.

Le dernier exemple montrait l'intégration de deux paradigmes antérieurs dans un nouveau paradigme plus général. Ce cas pourrait montrer une simple cumulativité si les résultats obtenus sous le paradigme plus général étaient la somme des résultats obtenus sous chaque paradigme précédent. A nouveau ce résultat est loin de ce que l'on observe en réalité. L'analyse transversale, en particulier celle utilisée par Durkheim pour montrer la liaison entre phénomènes agrégés du moment, et l'analyse biographique, pour montrer la liaison entre phénomènes individuels situés dans le temps, ne paraissaient guère pouvoir être cumulées additivement en une nouvelle forme d'analyse. La synthèse finalement proposée par l'analyse multiniveau constitue en fait une nouvelle forme de cumulativité. En distinguant différents niveaux d'agrégation elle va essayer de fournir une nouvelle méthode d'analyse qui permette de les traiter simultanément. Mais ce traitement simultané n'est en rien une additivité du traitement des données agrégées et individuelles. Il ne s'agit en fait que d'une nouvelle forme d'analyse, dans laquelle on retrouve des traces des formes antérieures, mais qui ne peut à aucun moment être considérée comme l'addition de ces deux analyses.

Comme l'indiquait déjà Berthelot (2001a), on constate que le mouvement généralisé qui anime les diverses sciences sociales, « où approfondissement et dispersion se mêlent, s'opère fondamentalement au niveau des programmes et des "paradigmes" ». Comme nous l'avons déjà indiqué, nous ne nous sommes placés ici qu'au niveau des paradigmes : c'est grâce à ce concept, défini dans l'introduction de ce chapitre, que nous avons examiné les données, les modèles ou les paradigmes eux-mêmes, et que nous examinerons la cumulativité des connaissances en démographie.

Voyons d'abord plus en détail comment se produit le passage du paradigme longitudinal au paradigme biographique, et essayons de comparer avec ce qui se produit lorsqu'on passe de la physique de Newton à la relativité généralisée d'Einstein. Comme nous l'avons dit, l'approche longitudinale en démographie se fondait sur la notion de population homogène et d'événements indépendants les uns des autres. De la même façon, la physique de Newton se fondait sur les notions d'espace homogène et isotrope dans ses dimensions. Les deux approches peuvent être considérées comme approximativement comparables si l'on considère que la physique s'intéresse à l'espace (nous laissons volontairement de côté le temps, qui complique la comparaison, sans en changer les conclusions) et la démographie aux populations (en laissant de nouveau de côté le temps) : les deux sont homogènes et ce que l'on peut considérer comme les dimensions de la démographie, les phénomènes étudiés, sont également isotropes, c'est-à-dire indépendants entre eux. Le passage au paradigme biographique, imposé par l'observation des données d'enquêtes plus détaillées que les données de registres, semble généraliser les résultats aux populations hétérogènes et aux événements dépendants les uns des autres. De même le passage à la relativité généralisée d'Einstein, imposé par d'importantes difficultés conceptuelles soulevées par la théorie de Newton, conduisit à un espace-temps courbe, déterminé par le contenu matériel de l'univers et en ce sens hétérogène, non isotrope, dans ses dimensions. La physique de Newton pourrait donc être considérée comme généralisée par celle d'Einstein : les axiomes correspondant aux deux vues sont cependant pour certains identiques et pour d'autres différents (Suppes, 2002).

Ainsi, pour nombre d'auteurs, cette généralisation n'est pas si évidente qu'il n'y paraît : Kuhn (1970) se demande : « la dynamique Newtonienne peut-elle réellement être dérivée de la dynamique relativiste ?²¹ » L'examen détaillé des concepts de dénomination identique chez Newton et chez Einstein, lui montre que les

références physiques de ces concepts d'Einstein ne sont en aucune façon identiques à celles des concepts de Newton qui portent le même nom. (La masse chez Newton est conservée ; la masse chez Einstein est convertible en énergie. Seulement les vitesses relatives basses peuvent être mesurées de façon identique, et elles ne peuvent même pas être conçues comme identiques)²².

Il en conclut que « l'on peut dès lors être assuré que les différences entre paradigmes successifs sont à la fois nécessaires et inconciliables²³ ».

De même, alors que ce sont les observations qui impliquent la théorie de Newton, ce sont au contraire les difficultés conceptuelles présentes dans cette théorie qui impliquent celle d'Einstein : ce n'est qu'après coup qu'il a pu vérifier sa théorie par des observations, telles que l'avance séculaire du périhélie de Mercure. Il ne peut dans ce cas y avoir aucune cumulativité, au sens où Kuhn conçoit les paradigmes.

Nous avons déjà pu, de façon semblable, montrer que l'approche biographique était très différente de l'approche longitudinale, et l'attaque de Chantal Blayo le montre clairement. En fait les deux approches ne définissent pas de façon semblable les populations sur lesquelles elles travaillent et étudient des liaisons très différentes, sinon incompatibles, entre les phénomènes. Pourrait-on donc en conclure qu'il n'y ait, dans ce cas, aucune cumulativité ?

²¹ Can Newtonian dynamics really be derived from relativistic dynamics?

²² The physical referents of these Einsteinian concepts are by no means identical with those of the Newtonian concepts that bear the same name. (Newtonian mass is conserved; Einsteinian is convertible with energy. Only at low relative velocities may the two be measured in the same way, and even then they must not be conceived to be the same.)

²³ Let us, therefore, now take it for granted that the differences between successive paradigms are both necessary and irreconcilable.

Il est nécessaire à ce point de la discussion de distinguer entre : (a) cumulativité des paradigmes, et (b) cumulativité des connaissances qui ont été acquises au sein de paradigmes différents. A priori, non-cumulativité des paradigmes n'exclut pas la cumulativité des connaissances obtenues au travers de paradigmes différents. Ainsi, comme nous l'avons déjà indiqué, lorsque Kuhn souligne l'hétérogénéité conceptuelle des paradigmes de Newton et Einstein, on ne peut en conclure que les connaissances de la mécanique classique sont annulées par la théorie de la relativité. Analysons d'abord plus en détail le second point avant de conclure sur le premier.

La cumulativité des connaissances apparaît comme évidente tout au long de l'histoire de la démographie et nos différentes présentations dans cet article le montrent parfaitement : passage enrichissant de la régularité des taux à leur variation ; passage d'un modèle considérant la génétique des populations comme indépendante de la démographie et agissant additivement avec celle-ci, à un modèle les considérant comme dépendant l'une de l'autre de façon non additive ; passage de l'indépendance entre phénomènes et l'homogénéité des populations, à la dépendance entre phénomènes et l'hétérogénéité des populations ; passage d'une dépendance à la société à une dépendance à l'individu, pour aboutir à une vue pleinement multiniveau. Chaque nouvelle étape intègre des éléments de la précédente et en rejette d'autres. C'est donc bien un progrès de la discipline que l'on observe, réalisé par l'introduction des paradigmes successifs. Chacun d'eux part de des insuffisances du précédent et propose une méthode permettant de combler ces insuffisances, tout en n'annulant pas la totalité des connaissances atteintes au travers du paradigme précédent. D'ailleurs pour certaines questions posées au démographe l'analyse transversale peut suffire, de même que pour la majorité des trajectoires de notre système planétaire l'analyse de Newton convient toujours parfaitement.

On peut donc dire que chaque nouveau paradigme vient compléter le précédent pour pouvoir traiter des cas sortant du champ de celui-ci, tout en conservant *en partie* certains des résultats obtenus avec celui-ci. Cependant cette conservation est loin d'être toujours assurée car le nouveau paradigme permet des raisonnements plus précis et détaillés que le précédent, pouvant conduire dans certains cas à des résultats très différents comme nous l'avons montré pour l'analyse biographique comparée à l'analyse longitudinale et surtout pour l'analyse multiniveau comparée à l'analyse transversale et l'analyse biographique.

Cela montre dès lors une certaine cumulativité des paradigmes qui est loin d'être linéaire. Cela vient du fait que les objets traités par la démographie sont sélectionnés par les différents paradigmes et sont donc propres à chacun d'entre eux. Ainsi l'analyse transversale ignore le temps vécu par un individu et met en relation les différents phénomènes démographiques observés à un moment donné avec les diverses caractéristiques des populations observées à ce même instant : ses objets sont donc ceux observés lors d'un recensement, par exemple. L'analyse longitudinale introduit le temps vécu par une génération ou une cohorte, mais n'arrive pas à intégrer les caractéristiques individuelles ni les liens entre les événements connus, car elle suppose la population homogène et les événements indépendants entre eux : ses objets sont ceux observés à l'aide des statistiques de l'état civil. L'analyse biographique permet enfin d'intégrer les diverses caractéristiques individuelles et les liens entre événements : ses objets sont ceux observés à l'aide d'enquêtes biographiques. Enfin l'analyse multiniveau permet une synthèse entre l'analyse transversale et l'analyse biographique en introduisant divers niveaux d'agrégation : ses objets sont ceux observés à l'aide d'enquêtes encore plus complexes que les enquêtes biographiques, faisant intervenir simultanément les caractéristiques observées à différents niveaux d'agrégation.

Chacun de ces paradigmes ne permet de raisonner que sur ses propres objets, mais il a été prouvé comme parfaitement cohérent par rapport à eux, comme différents ouvrages d'analyse démographique l'ont montré au cours du temps, depuis le XVIIe siècle. Mais en même temps la raison, pour laquelle il n'est pas possible de montrer une relation parfaitement linéaire entre eux, est que ces objets sont différents. En revanche une relation non-linéaire et une très forte continuité existent entre eux, et celles-ci peuvent être interprétées comme une cumulativité, pourvu que l'on n'oublie pas que chaque paradigme ne peut être vérifié que par ses propres objets (Agazzi, 1985). Ce dernier auteur, dépassant la position de Kuhn, montre très clairement, pour les sciences naturelles, mais cela pourrait être transposé pour les sciences humaines, que

le progrès scientifique ne consiste pas en une relation purement logique entre théories, et qu'en plus cette relation n'est pas linéaire. Cependant ce progrès existe et peut être même interprété comme cumulatif, dans la mesure où l'on n'oublie pas que chaque théorie scientifique est seulement vraie *pour* ses propres objets spécifiques²⁴.

Nous ne présenterons pas ici le détail de sa démonstration, mais en donnerons les éléments utiles pour notre discussion. Il constate en effet que le chercheur choisit dans la réalité un nombre restreint d'objets qu'il va inclure dans sa théorie. Une partie de ces objets dépend très fortement du contexte de la théorie (contextual part), une autre au contraire n'en dépend pas, constituant sa part référentielle (referential part), et permet de ce fait une comparaison entre théories différentes. Ainsi en démographie les divers événements traités ne dépendent pas du type d'analyse faite : la mortalité, la fécondité, la migration, etc., sont des objets indépendants de la théorie utilisée pour les traiter. En revanche les relations supposées entre ces objets en dépendent fortement : indépendance entre eux en analyse longitudinale, forte dépendance en analyse biographique. Il en résulte donc une possibilité de comparaison entre théories, qui ont la même part référentielle. Mais surtout que les résultats de l'une ne peuvent jamais être détruits par une autre dont les objets de recherche, qui en constituent la part contextuelle, sont différents. Ainsi, « les nouveaux résultats restent *simultanément présents avec* les anciens et les complètent²⁵ » : la mécanique classique est toujours vérifiée par ses objets, de même que l'analyse longitudinale. Mais la théorie de la relativité complémente la mécanique classique, de la même façon que l'analyse biographique complémente l'analyse longitudinale.

Comme le dit très justement Gilles-Gaston Granger (1994) : « Le fait humain ne peut certes être scientifiquement connu qu'à travers une pluralité de géométrals, mais à condition toutefois que soit découverte l'opération contrôlable qui le restitue à partir d'eux stéréoscopiquement ». Il nous semble que la pluralité des paradigmes observés en démographie corresponde bien à cette pluralité des géométrals et que les relations que nous avons pu montrer entre eux en permettent une restitution stéréoscopique fort prometteuse pour l'avenir. La notion de cumulativité qui en ressort est bien entendu loin de la simple additivité, mais permet une avancée, à notre avis, tout à fait pertinente pour la démographie et même pour l'ensemble des sciences humaines.

²⁴ Scientific progress does not consist in a purely logical relationship between theories, and moreover it is not linear. Yet it exists and may even be interpreted as an accumulation of truth, provided we do not forget that every scientific theory is true only *about* its own specific objects.

²⁵ The new truths remain *together with* the old ones and *complement* them

Remerciements :

Je remercie vivement ici Robert Franck pour les commentaires et avis si pertinents, qu'il m'a fournis tout au long de la rédaction de cet article. Je remercie également Jean-Claude Gardin, Alban Bouvier et Pierre Livet pour les commentaires écrits qu'ils m'ont fait parvenir, ainsi que tous les membres du groupe de travail pour les commentaires oraux faits lors de la présentation de deux versions antérieures de cet article : j'en ai tenu compte pour la rédaction de cette version finale. Bien entendu ce texte n'engage que la responsabilité de son auteur.

BIBLIOGRAPHIE

- Aalen Odd, « Nonparametric inference for a family of counting processes », *The Annals of Statistics*, no 6/4, juillet 1978, p.701-726.
- Agazzi Evandro, « Commensurability, incommensurability, and cumulativity in scientific knowledge », *Erkenntniss*, no 22/1-3, janvier 1985, p. 51-77.
- Arbuthnot John, « An argument for divine providence, taken from the constant regularity observ'd in the births of both sexes », *Philosophical Transactions of the Royal Society of London*, no 27, 1710, p. 186-190.
- Attané Isabelle, *Une Chine sans femmes*, Paris, Editions Perrin, 2005.
- Berthelot Jean-Michel, « Les nouveaux défis épistémologiques de la sociologie », *Sociologie et Sociétés*, no 30/1, printemps 1998, p. 23-29.
- Berthelot Jean-Michel, « Les sciences du social », in Jean-Michel Berthelot (ed.), in *Epistémologie des sciences sociales*, Paris, Presses Universitaires de France, 2001a, p. 203-265.
- Berthelot Jean-Michel, « Programmes, paradigmes, disciplines : pluralité et diversité des sciences sociales », in Jean-Michel Berthelot (ed.), *Epistémologie des sciences sociales*, Paris, Presses Universitaires de France, 2001b, p.457-519.
- Billari Francesco, Prskawetz Alexia, *Agent-based computational demography*, Heidelberg-New York, Physica-Verlag, 2003.
- Blayo Chantal, « La condition d'homogénéité en analyse démographique et en analyse statistique », *Population*, no 50/6, novembre-décembre 1995, p. 1501-1518.
- Blum Kenneth, Noble Ernest, Sheridan Peter, Montgomery Anne, Ritchie Terry, Jagadeeswaran Pudur, Nogami Harou, Briggs Arthur, Cohn Jay., « Allelic association of human dopamine D₂ receptor gene in alcoholism », *Journal of the American Medical Association*, no 263/15, 18 avril 1990, p. 2055-2060.
- Bourgeois-Pichat Jean, *La dynamique des populations*, Paris, Editions de l'Institut national d'études démographiques (coll. « Travaux et Documents »), 1994.
- Bouvier Alban, « L'architecture de la sociologie », *Revue du MAUSS*, no 28/2, 2^e semestre 2006, p. 391-402.
- Brian Eric, Jaisson Marie, *The descent of human sex-ratio at birth*, Dordrecht, Springer (coll. « Methodos Series »), 2007.
- Coale Ansley, Bannister Judith, « Five decades of missing females in China », *Demography*, no 31/3, août 1994, p. 459-479.
- Coleman, James, « Social theory, social research, and a theory of action », *American Journal of Sociology*, no 91/6, mai 1986, p. 1309-1335.
- Condorcet, *Tableau historique des progrès de l'esprit humain. Projets, esquisses, fragments et notes (1772-1794)*, Paris, Editions de l'Institut national d'études démographiques, 2004.
- Courgeau Daniel, « Mobilité géographique, nuptialité et fécondité », *Population*, no 31/4-5, juillet-octobre 1976, p. 901-915.

- Courgeau Daniel, « Migration and demographic phenomena in France », in J. White (ed.), *The Urban Impact of Internal Migration*, Chapel Hill, Institute for Research in Social Science, 1979, p.1-32.
- Courgeau Daniel, « Proposed analysis of the French Migration, family and occupation survey », in *Multistate Life History Analysis*, Laxenburg, IIASA, 1982, p. 1-14.
- Courgeau Daniel, « Evolution ou révolution dans la pensée démographique », *Mathématiques et Sciences Humaines*, no 40/160, hiver 2002, p. 49-76.
- Courgeau Daniel (ed.), *Methodology and Epistemology of Multilevel Analysis*, Dordrecht / Boston / Londres, Kluwer Academic Publishers (coll. « Methodos Series »), 2003.
- Courgeau Daniel, *Du groupe à l'individu: Synthèse multiniveau*, Paris, Editions de l'Ined, 2004.
- Courgeau Daniel, *Multilevel synthesis: from the group to the individual*, Dordrecht, Springer (coll. « Springer Series on demographic methods and population analysis »), 2007.
- Courgeau Daniel, Franck Robert, « La démographie, science constituée ou en voie de constitution ? Esquisse d'un programme », *Population-F*, no 62/1, juillet-octobre 2007, p.39-46 (« Demography – A fully formed science or a science in the making? An outline programme », *Population-E*, no 62/1, juillet-octobre 2007, p. 39-46).
- Courgeau Daniel, Lelièvre Eva, « Nuptialité et agriculture », *Population*, no 41/2, mars-avril 1986, p.303-326.
- Courgeau Daniel, Lelièvre Eva, *Analyse démographique des biographies*, Paris, Editions de l'Ined, 1989.
- Courgeau Daniel, Lelièvre Eva, *Event History Analysis in Demography*, Oxford, Clarendon Press, 1992.
- Courgeau Daniel, Lelièvre Eva, « Changement de paradigme en démographie », *Population*, no 51/3, mai-juin 1996, p. 645-654.
- Courgeau Daniel, Lelièvre Eva, « Changing Paradigm in Demography », *Population: An English Selection*, no 9, 1997, p. 1-10.
- Courgeau Daniel, Lelièvre Eva, *Análisis demográfico de las biografías*, México, El Colegio de México, 2001.
- Courgeau Daniel, Vetta Atam, « Nature, culture et génétique du comportement », in E. Vilquin (ed.), *Entre nature et culture : quelle(s) démographie(s)*, Chaire Quetelet 2002, Academia-Bruylant, Louvain-la-Neuve, p. 85-96.
- Cox David, « Regression models and life tables (with discussion) », *Journal of the Royal Statistical Society, Series B (Methodological)*, no 34/2, 1972, p. 187-220.
- Dupâquier Jacques, Dupâquier Michel, *L'histoire de la démographie*, Paris, Librairie Académique Perrin, 1985.
- Durkheim Emile, *Les règles de la méthode sociologique*, Paris, Alcan, 1895.
- Durkheim Emile, *Le suicide*, Paris, Alcan, 1897.
- Fisher Ronald Aymler, « The correlation between relatives on the supposition of Mendelian inheritance », *Transactions of the Royal Society of Edinburgh*, no 52, 1918, p.399-433.
- Fisher Ronald Aymler, « Limits to intensive production in animals », *British Agricultural Bulletin*, no 4, 1951, p. 217-218.
- Franck Robert, « Mosaiques, machines, organismes et sociétés. Examen métadisciplinaire du réductionnisme », *Revue Philosophique de Louvain*, no 93, 1995, p. 67-81.
- Goldstein Harvey, *Multilevel statistical models*, Londres, Arnold, 2003.
- Gottlieb Gilbert, *Individual Development and Evolution : The Genesis of Novel Behavior*, New York, Oxford University Press, 1992.
- Gottlieb Gilbert, « Genetics and Development », in *International Encyclopedia of the Social and Behavioral Sciences, Genetics, Behavior, and Society*, vol 9, N.J. Smelser, P.B. Baltes eds., Elsevier, 2001, p. 6121-6127.
- Granger Gilles-Gaston, *Formes, opérations, objets*, Paris, Librairie Philosophique J. Vrin, 1994.
- Guillard Achille, *Eléments de statistique humaine ou démographie comparée*, Paris, Guillaumin, 1855.
- Graunt John, *Natural and Political Observations Mentioned in a following Index, and made upon the Bills of Mortality*, Londres, Tho. Roycroft, 1662.

- Hacking Ian, *The emergence of probability*, Cambridge, Cambridge University Press, 1975.
- Halbwachs Maurice, « Recherches statistiques sur la détermination du sexe à la naissance », *Journal de la Société Statistique de Paris*, 74 /5, mai 1933, p. 164-191.
- Halbwachs Maurice, Sauvy Alfred (dir.), « Le point de vue du nombre », in *L'espèce humaine, Tome VII* (coll. « Encyclopédie Française »), Paris, 1936.
- Henry Louis, « D'un problème fondamental d'analyse démographique », *Population*, no 13/1, janvier-mars 1959, p. 9-32.
- Henry Louis, *Démographie. Analyse et modèles*, Paris, Larousse, 1972.
- Henry Louis, *Dictionnaire démographique multilingue*, Liège, Ordina Editions, 1981.
- Hoem Jan, *Demography*, vol 15, in *International Encyclopedia of the Social and Behavioral Sciences*, Neil Smelser, Paul Baltes (eds.), Oxford, Elsevier Science Ltd., 2001.
- Hofacker J.D., *Ueber die Eigenschaften*, Tübingen, Osiander, 1828.
- Jacob François, *La logique du vivant*, Paris, Gallimard, 1970.
- Jacquard Albert, « Heritability : one word three concepts », *Biometrics*, no 39/2, juin 1983, p. 465-477.
- Jensen Arthur, « How much can we boost IQ and scholastic achievement? », *Harvard Educational Review*, no 39/1, winter 1969, p. 1-123.
- Jinks John., Fulker David, « Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behaviour », *Psychological Bulletin*, no 73/5, mai 1970, p. 311-348.
- Kohler Hans-Peter, Rodgers Joseph Lee, Christensen Kaare, « Is fertility behaviour in our genes? Findings from a Danish twin study », *Population and Development Review*, no 25/2, juin 1999, p. 253-288.
- Kuhn Thomas, *The structure of scientific revolutions*, Chicago and London, The University of Chicago Press, 1962.
- Kuhn Thomas, « Postscript, in Kuhn Thomas », *The structure of scientific revolutions*, Chicago and Londres, The University of Chicago Press, 1970, p. 174-210.
- Laplace Pierre-Simon, « Mémoire sur les probabilités », *Mémoires de l'Académie Royale des Sciences*, 1778, p. 227-332.
- Lazarsfeld Paul-Felix, Menzel Herbert, « On the relation between individual and collective properties », in Etzioni Amitai (ed.), *Complex Organizations*, New York, Holt, Reinhart and Winston, 1961, p. 422-440.
- Levins Richard, Lewontin Richard, *The Dialectical Biologist*, Harvard, Harvard University Press, 1985.
- Masterman Margaret, « The nature of a paradigm », in Lakatos Imre, Musgrave Alan (eds), *Criticism and the Growth of Knowledge*, Cambridge, Cambridge University Press, 1970, p. 59-89.
- Menken Jane, Trussell James, « Proportional hazards life table models: An illustrative analysis of socio-demographic influences on marriages dissolution in the United States », *Demography*, no 18/2, mai 1981, p.181-200.
- Montmort Pierre Rémond de, *Essay d'analyse sur les jeux de hazard, Seconde édition revue et augmentée de plusieurs lettres*, Paris, Jacque Quillot, 1713.
- Nadeau Robert, *Vocabulaire technique et analytique de l'épistémologie*, Paris, Presses Universitaires de France, 1999.
- Pearson Karl, « Abraham De Moivre », *Nature*, no 2946, 1926, p. 551-552.
- Petty William, *Political Arithmetick*, London, Robert Clavel & Hen. Mortlock, 1690.
- Poisson Siméon Denis, « Mémoire sur la proportion des naissances des filles et des garçons », *Mémoires de l'Académie Royale des Sciences de l'Institut de France*, IX, 1830.
- Pressat Roland, *Dictionnaire de démographie*, Paris, Presses Universitaires de France, 1979.
- Robinson William., « Ecological correlations and the behavior of individuals », *American Sociological Review*, no 15/3, juin 1950, p. 351-357.

- Rodgers Joseph Lee, Hughes Kimberly, Kohler Hans-Peter, Christensen Kaare, Doughty Debby, Rowe David, Miller Warren, « Genetic influences helps explain variations in human fertility: Evidence from recent behavioural and molecular genetic studies », *Current directions in Psychological Science*, no 10/5, 2001, p. 184-188.
- Rodgers Joseph Lee, Kohler Hans-Peter (eds.), *The Biodemography of Human Reproduction and Fertility*, Dordrecht / Boston / London, Kluwer Academic Publishers, 2003.
- Rodgers J.L., Rowe D.C., Miller W.B., *Genetic influences on Human Fertility and Sexuality*, Dordrecht / Boston / London, Kluwer Academic Publishers, 2000.
- Rohrbasser Jean-Marc, « L'ordre divin: de l'arithmétique politique à la physico-théologie », in Süßmilch J.P., *L'ordre divin dans les changements de l'espèce humaine, démontré par la naissance, la mort et la propagation de celle-ci*, texte intégral de l'édition de 1741, Paris, Institut national d'études démographiques, 1998.
- Russo Federica, *Causality and causal modelling in the social sciences : measuring variations*, Dordrecht, Springer (coll. « Methodos Series »), 2009.
- Ryder Norman, *The cohort approach. Essays in the measurement of temporal variations in demographic behaviour*, Ph. D. Dissertation, New York, Princeton University (édition photographique de 1980, New York, Arno Press), 1951.
- Ryder Norman., « The cohort as a concept in the study of social change », *American Sociological Review*, no 30/5, décembre 1965, p. 843-861.
- Sadler Michael Thomas, *The Law of Population*, London, John Murray, 1830.
- Sarkar Sahotra, *Genetics and Reductionism*, Cambridge, Cambridge University Press, 1998.
- Suppe Frederick, *The Semantic Conception of Theories and Scientific Realism*, Urbana and Chicago, University of Illinois Press, 1989.
- Suppes Patrick, *Representation and Invariance of Scientific Structures*, Stanford, CSLI Publications, 2002.
- Süßmilch Johann Peter, *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, Tod, und Fortpflanzung desselben erwiesen*, Berlin, 1741.
- Vetta Atam, « Correction to Fisher's correlation between relatives and environmental effects », *Nature*, 263, 1976, p. 316-317.
- Vetta Atam et Courgeau Daniel, « Comportements démographiques et génétique du comportement », *Population*, no 58/4-5, juillet-octobre 2003, p. 457-488 (« Demographic behaviour and behaviour genetics », *Population, English Edition*, 58 /4-5, juillet-octobre 2003, p. 401-428).
- Whelpton Pascal, « Cohort analysis of fertility », *American Sociological Review*, no 14/6, décembre 1949, p. 735-749.