



HAL
open science

La simulation sur ordinateur des phénomènes sociaux

Jean-Paul Grémy

► **To cite this version:**

Jean-Paul Grémy. La simulation sur ordinateur des phénomènes sociaux. Sociologie. Université René Descartes - Paris V, 1971. Français. NNT: . tel-00654804

HAL Id: tel-00654804

<https://theses.hal.science/tel-00654804>

Submitted on 23 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

J.-P. GREMY

**LA SIMULATION SUR ORDINATEUR
DES PHENOMENES SOCIAUX**

Thèse pour le Doctorat de 3^e cycle
sous la direction de Monsieur
le Professeur DAVAL

J.-P. GREMY

LA SIMULATION SUR ORDINATEUR

DES PHENOMENES SOCIAUX

Je tiens tout d'abord à exprimer ma reconnaissance à Monsieur le Professeur Roger DAVAL, sans qui cette recherche n'aurait pu être menée à son terme, et qui, dans le contexte de l'Institut des Sciences Humaines Appliquées (ISHA) dont il est le Directeur, m'a accordé toute l'aide dont j'ai eu besoin pour ce travail.

Je remercie également Monsieur le Professeur Alain GIRARD qui m'a autorisé à réaliser une série d'expériences de simulation à partir de son étude sur Le choix du conjoint, et a eu l'obligeance de me laisser la libre disposition des données qui ont été recueillies pour cette étude.

Parmi les personnes dont les critiques ou les conseils m'ont été utiles à divers moments de cette recherche, et qui voudront bien trouver ici l'expression de mes remerciements, je citerai les Professeurs Raymond BOUDON et Marc BARBIT ; Messieurs Jean-Claude FAURE et Raphaël BENAYOUN (SEMA) ; Messieurs Jean SUTTER et Albert JACQUARD (INED) ; et mes anciens collègues de la Société d'Economie et de Mathématique Appliquées (SEMA), et de l'Institut des Sciences Humaines Appliquées (ISHA).

J'exprime enfin mes vifs remerciements à toutes celles qui, dans le cadre de l'ISHA, et de l'U.E.R. de Mathématiques, Logique Formelle et Informatique de la Sorbonne (Lettres), ont contribué à la présentation matérielle de ce travail ; et en premier lieu, à Mademoiselle TICHIT, qui a assuré avec beaucoup de dévouement les frappes successives du texte, et qui en a réalisé la mise en page définitive avec toute la compétence dont elle fait preuve pour le bulletin Informatique en Sciences Humaines ; ainsi qu'à Mademoiselle HAVAS qui, avec la collaboration de Mademoiselle CHAUVET, a dessiné la majeure partie des organigrammes, courbes, et histogrammes présentés dans le texte.

Jean-Paul GREMY

T A B L E

Introduction

	7
0.1. Objet de la recherche	7
0.2. Description de la méthode utilisée	10
0.3. Plan de l'exposé	11

Première partie : les techniques de simulation et leur utilisation en sociologie.	14
---	----

<u>1. QU'EST-CE QUE LA SIMULATION ?</u>	14
---	----

<u>1.1. Examen de quelques définitions proposées par les spécialistes.</u>	16
--	----

1.1.1. La définition proposée par D.G. KENDALL	16
1.1.2. La définition proposée par K.M. HOGLUND et P.L. PHIPPS	19
1.1.3. La définition élaborée par l'I.F.I.P. et l'I.C.C.	20
1.1.4. Les définitions proposées par R. BOUDON	22
1.1.5. La définition proposée par J. GUITTET	24
1.1.6. La définition proposée par R.E. DAWSON	26
1.1.7. La définition proposée par J. AGARD	27
1.1.8. La définition proposée par P. ROSSI	29

<u>1.2. Définition du domaine étudié.</u>	29
---	----

1.2.1. Démarche suivie pour aboutir à une définition	30
1.2.2. Proposition d'une définition de la simulation	35
1.2.3. Qu'est-ce qu'un système ?	35
1.2.4. Différents types de systèmes	38
1.2.5. Le concept de modèle	44
1.2.6. Les relations entre le système étudié et son modèle	47
1.2.7. Qu'entend-on par expérimentation ?	53
1.2.8. Champ d'application de notre définition	57

<u>2. LES DIFFERENTS TYPES DE SIMULATION.</u>	60
<u>2.1. Quelques jalons historiques</u>	60
2.1.1. L'Antiquité et le Moyen-Age	62
2.1.2. L'époque moderne	72
2.1.3. L'époque contemporaine	87
<u>2.2. Classification des exemples de simulation.</u>	93
2.2.1. Les objectifs de la simulation	95
2.2.2. Les caractéristiques de l'original	104
2.2.3. Les caractéristiques du modèle	109
2.2.4. La nature des relations liant le modèle à l'original	116
2.2.5. Remarques sur la possibilité de construire une typologie	124
 <u>3. LA SIMULATION EN SOCIOLOGIE.</u>	 127
<u>3.1. Place de la simulation dans les sciences sociales.</u>	127
3.1.1. Les simulations de processus individuels	130
3.1.2. Les simulations de processus de groupe	138
3.1.3. Les simulations électorales	143
3.1.4. Simulation du fonctionnement des organisations	145
3.1.5. Simulation de l'évolution d'un système	151
<u>3.2. Description détaillée de deux expériences de simulations en sociologie.</u>	155
3.2.1. La simulation du comportement de production dans un atelier (R. BRETON)	156
3.2.2. Le "projet Simulmatics" (POOL, ABELSON, POPKIN)	173

Deuxième partie : application de la simulation à l'étude expérimentale de l'homogamie.	198
--	-----

4. L'ELABORATION DU MODELE. 198

4.1. Le problème du choix du conjoint. 201

- 4.1.1. L'étude réalisée par Alain GIRARD 201
- 4.1.2. Les dimensions de l'homogamie 206
- 4.1.3. Quelques directions de recherche possibles 224

4.2. Un modèle pour l'étude expérimentale de l'homogamie. 227

- 4.2.1. Position du problème 228
- 4.2.2. Formalisation du modèle 235
- 4.2.3. La dynamique du modèle 259

5. LA PROGRAMMATION DU MODELE SUR ORDINATEUR. 264

5.1. Quelques choix préalables. 266

- 5.1.1. La présentation des données 266
- 5.1.2. La génération de "nombres au hasard" 278
- 5.1.3. Le choix du langage de programmation 289

5.2. La conception du programme. 292

- 5.2.1. Les mécanismes de constitution des couples 293
- 5.2.2. La structure générale du programme 301

5.3. L'écriture du programme. 305

- 5.3.1. Organigramme commenté 305
- 5.3.2. Texte du programme 319

6. PRINCIPAUX RESULTATS

324

6.1. L'influence du modèle de décision sur l'homogamie

325

6.1.1. Modèles de décision à un seul critère de choix

326

6.1.2. Modèles de décision à deux critères de choix

348

6.1.3. Comportement comparé des différents critères

358

6.2. Comparaison des données obtenues par simulation et des données réelles.

365

6.2.1. Peut-on reconstituer artificiellement les données
fournies par l'enquête ?

366

6.2.2. Description de quelques échantillons engendrés par
simulation

368

6.3. Bilan de l'expérience.

385

6.3.1. Evaluation des apports de la série de simulations

385

6.3.2. Examen des insuffisances et des limites de l'expérience

388

Conclusions

396

7.1. La simulation et l'ordinateur

396

7.2. Portée et limites des expériences de simulation

403

7.3. Intérêt de la simulation pour le sociologue

415

Références bibliographiques

433

INTRODUCTION

Depuis une vingtaine d'années environ, les techniques de simulation sur ordinateur ont connu un essor considérable, dans des domaines aussi variés que la physique appliquée, l'économie, l'organisation et la gestion de l'entreprise, la recherche opérationnelle. La sociologie elle-même a été touchée par cet essor, plus tardivement il est vrai, et avec semble-t-il un succès moindre. Malgré l'intérêt suscité par quelques expériences spectaculaires, en particulier en sociologie politique, il ne semble pas que l'apport à la sociologie des techniques de simulation soit actuellement important ; on peut même estimer cet apport disproportionné aux moyens mis en oeuvre (investissement intellectuel, coût du temps d'ordinateur). Nous nous sommes assigné comme objectif dans cette recherche l'évaluation de ce que les techniques de simulation peuvent apporter à la sociologie, tant au niveau des applications qu'à celui de la réflexion théorique.

0.1. Objet de la recherche.

Il s'agit par conséquent d'une recherche à *visée méthodologique*. Pour atteindre l'objectif que nous nous sommes fixé, il nous a été nécessaire de procéder à un examen critique des simulations

projetées ou déjà réalisées dans les sciences sociales, d'analyser les jugements portés sur l'intérêt de ces réalisations par les sociologues qui y ont eu recours, et d'effectuer nous-même une expérience de simulation à titre d'essai de la méthode. Ces diverses démarches devaient nous permettre de mettre en évidence les possibilités et les limites de l'usage de la simulation en sociologie, et d'illustrer sur un exemple simple, fabriqué pour la circonstance, les conditions requises pour la mise en oeuvre de cette approche, les qualités qu'elle exige du sociologue, et les précautions que requiert la réalisation d'une expérience simulée, et l'interprétation de ses résultats.

Nous aurions souhaité pouvoir aboutir ou bien à la rédaction d'une sorte de manuel destiné au sociologue désireux de recourir à la simulation, manuel édictant un ensemble de conseils pratiques et éventuellement de mises en garde sur les mauvais usages de cette approche ; ou bien à la description détaillée des tâtonnements, des essais, et des erreurs qui jalonnent la conception et la mise en oeuvre d'une expérience de simulation. La rédaction d'un *manuel de l'utilisateur* nous est apparue comme un objectif prématuré en l'état actuel des recherches sur la simulation. En effet, si nous nous limitons aux seuls problèmes *pratiques*, la grande variété des techniques utilisables nous interdisait d'énoncer des préceptes ayant un certain caractère d'universalité. Si par contre nous préférons nous situer au plan de la *méthode*, celle-ci n'ayant pas encore été codifiée, nous étions condamné à osciller entre des considérations empruntées à l'épistémologie, trop générales pour rendre service aux utilisateurs, et des recettes empiriques, fondées sur les tâtonnements des pionniers de la simulation, n'ayant pas encore reçu la consécration de la réflexion méthodologique. Quant à

retracer fidèlement la démarche réelle suivie par un sociologue utilisant une technique de simulation, c'eût été certainement entreprendre un travail très utile du point de vue de la philosophie des sciences. Car nous pensons que toute réflexion méthodologique sérieuse doit s'appuyer sur l'analyse détaillée du cheminement réel du chercheur, et non sur le compte-rendu académique et épuré qu'il donne de celui-ci lorsque sa recherche est achevée. Mais un tel travail aurait donné naissance à une sorte de *journal de recherche*, sans structure logique, renfermant de nombreux repentirs et redites, retraçant simultanément les péripéties de la mise en oeuvre sur ordinateur numérique de la méthode de Monte-Carlo, les réflexions inspirées par le compte-rendu d'une expérience de simulation, les avatars de notre projet de modèle, la recherche de données simulables, etc.. C'est dire qu'un texte de cette nature, tout en pouvant présenter un intérêt pour une recherche de méthodologie expérimentale, aurait été pour le sociologue extrêmement fastidieux à lire, et trop spécifique pour avoir une quelconque valeur exemplaire.

Nous avons donc adopté une sorte de compromis entre ces deux possibilités, en divisant notre recherche en deux branches parallèles : d'une part, une *réflexion théorique*, nourrie des réalisations et des recherches méthodologiques antérieures, et susceptible d'être présentée selon un cadre logique ; et d'autre part l'*expérience vécue* d'un sociologue utilisateur de la simulation, expérience qu'il était nécessaire de retracer selon son déroulement historique, mais en schématisant la démarche suivie et en l'élaguant de péripéties jugées *a posteriori* inutiles ou trop spécifiques, de manière à lui conférer un caractère d'illustration plus générale d'une procédure de simulation.

0.2. Description de la méthode utilisée.

Nos recherches sur les utilisations de la simulation dans les sciences sociales ont été dès le début entravées par le peu d'accord que nous avons constaté, entre les spécialistes de la simulation, sur ce qu'il convient de désigner de ce vocable ; le désaccord est d'ailleurs aggravé par l'usage que certains sociologues ont fait de ce mot, l'étendant, apparemment de façon abusive, au jeu de rôle et aux expériences de laboratoire en psychologie sociale. Aussi avons nous dû commencer par recenser et clarifier les notions de base ; ce faisant, nous avons été contraint de trancher entre les diverses acceptions de la simulation, et de définir nous-même le champ de notre étude. Pour cela, nous avons entrepris un travail de compilation et d'analyse qui a porté sur divers traités généraux, comptes-rendus de colloques, et manuels, concernant la cybernétique, l'informatique, la recherche opérationnelle, et, bien entendu, la simulation ; sur un nombre important de descriptions d'application de la simulation à la sociologie et d'exposés de réflexions sur ce type d'approche ; sur un nombre plus succinct de descriptions d'expériences de simulation réalisées en recherche opérationnelle, en gestion de l'entreprise, en économie, en démographie, en psychologie, en physiologie, en biologie ; enfin sur quelques ouvrages traitant de la méthode expérimentale, et de l'histoire des sciences et des techniques.

La réalisation, à titre de démonstration, d'un exemple concret de simulation, nous a imposé le choix d'un thème. Pour des raisons qui apparaîtront dans le cours de l'exposé, nous avons choisi de simuler un processus sociologique en utilisant au cours de la simulation des données empiriques tirées d'une enquête par

sondage. Plutôt que de laisser la détermination du thème au hasard de nos réflexions et de nos lectures, nous avons procédé de manière systématique. Nous avons utilisé une bibliographie établie par l'*Association pour la Diffusion de la Pensée Française*, et qui constitue une sorte d'anthologie des études réalisées en France, entre 1945 et 1965, en sociologie et psychologie sociale (207). Nous avons tout d'abord sélectionné dans cette bibliographie les études portant sur un processus d'ordre sociologique, et reposant sur les résultats d'une enquête par sondage. Pour les quelques titres satisfaisant à ces conditions, nous avons parcouru rapidement le compte-rendu de l'étude. En dernière analyse, nous avons retenu *Le choix du conjoint*, d'Alain GIRARD ; l'auteur a bien voulu mettre à notre disposition pour cette recherche les cartes perforées sur lesquelles avaient été reportés les résultats de l'enquête. A partir de quelques propositions tirées de l'analyse d'Alain GIRARD, nous avons élaboré un modèle simple utilisant certaines données de l'enquête ; nous avons écrit un programme de simulation correspondant au modèle, et exécuté ce programme sur l'ordinateur CDC 3600 du Centre de Calcul du C.N.R.S.

0.3. Plan de l'exposé.

L'exposé que nous donnons ci-après des résultats de notre recherche doit être considéré avant tout comme un *bilan provisoire* de ce que l'on peut dire sur la simulation en sociologie. Ceci pour diverses raisons. Quand nous avons entrepris cette recherche, nous pensions que le nombre des exemples d'application de la simulation à la sociologie était inférieur à trente ; nous avons constaté qu'à

elles seules les expériences ayant fait l'objet d'une publication dépassaient la centaine. Nous pensions que la notion de simulation était un concept clair et précis ; nous avons relevé presque autant de définitions qu'il y a d'auteurs. Nous pensions que les techniques de simulation étaient peu variées, et qu'il était possible de les résumer en une procédure-type ; il s'est avéré à l'examen que si la mise en oeuvre sur ordinateur numérique de la *méthode de Monte-Carlo* occupait une place privilégiée, elle n'était pas, et de loin, la seule procédure utilisée. Nous pensions enfin que la méthodologie de la simulation avait été élaborée et codifiée par les praticiens de la recherche opérationnelle, qui en font un usage quasi-quotidien ; nous n'avons trouvé dans ce domaine que des affirmations empruntées aux théoriciens de la science expérimentale, des réflexions plus originales mais parcellaires, des mises en garde prudentes, et des aveus d'ignorance. Aussi cet exposé n'est-il qu'une présentation, sans prétention à l'exhaustivité, de ce qui a été dit ou réalisé à propos de la simulation en sociologie, et des principales questions qui restent encore sans réponse.

Cet exposé est divisé en deux parties, chacune correspondant à l'une des branches de notre recherche.

La partie théorique compte trois chapitres :

- dans le premier, nous passons en revue les principales définitions de la simulation parmi toutes celles que nous avons relevées, et nous précisons le sens que nous donnons à cette notion.

- le second chapitre commence par un bref historique des événements qui ont concouru à favoriser l'apparition des techniques de simulation ; il propose ensuite un système de classification des diverses applications de la simulation telle que nous l'avons définie.

- le chapitre trois enfin a pour objet la place qu'occupent actuellement les expériences de simulation dans les sciences sociales ; il renferme la description détaillée de deux expériences qui ont eu un certain retentissement.

La partie descriptive relate notre propre simulation, et ne doit être considérée que comme une illustration des procédures évoquées dans la première partie. Elle compte également trois chapitres :

- le chapitre quatre expose le problème choisi comme exemple d'application, et décrit l'élaboration du modèle et la préparation des données.

- le cinquième chapitre traite des problèmes techniques de la traduction opératoire du modèle et de sa programmation.

- les résultats de cette expérience sont résumés dans le chapitre six, qui se clot par une évaluation de l'intérêt de l'expérience.

La conclusion esquisse un bilan des avantages et des limites de l'utilisation des techniques de simulation en sociologie, et tente de répondre à la question que nous nous sommes posée au début de cette recherche : "Que peuvent apporter à la sociologie les techniques de simulation ?".

PREMIERE PARTIE :

LES TECHNIQUES DE SIMULATION

ET LEUR UTILISATION EN SOCIOLOGIE

1. QU'EST-CE QUE LA SIMULATION ?

Le terme de simulation est un mot français d'usage assez répandu, dont l'ascendance latine est évidente (*simulatio*, de *simulare* : imiter, feindre). Dans le langage courant, selon LITTRE (140, T7, p. 180, articles *simulation*, *simuler*), *simulation* désigne l'acte de "faire paraître comme réel ce qui ne l'est point. (...) Simuler un combat, une maladie". *Simuler* se distingue à la fois de *feindre* et de *dissimuler* :

- de *feindre*, car son sens est un peu plus précis : "celui qui feint une maladie peut simplement se mettre au lit, et faire croire qu'il est malade. Celui qui simule une maladie, en détermine en soi quelques symptômes" (ibid.) ;

- de *dissimuler*, car "simuler une maladie, c'est feindre de l'avoir ; dissimuler une maladie, c'est feindre de ne l'avoir pas" (ibid.).

L'exemple de la maladie, choisi par LITTRE pour illustrer ses définitions, nous permet de pousser plus avant notre analyse du sens usuel du mot *simulation*. Pour qu'il y ait simulation, il est nécessaire qu'il y ait une situation réelle complexe (la maladie), que l'on ne peut intégralement reproduire pour des raisons pratiques (risque d'être soi-même malade !), mais dont on désire cependant reproduire certains traits (symptômes), jugés essentiels pour atteindre un but particulier que l'on poursuit (faire croire à autrui, et en particulier à un médecin, que l'on est malade). Dans un autre article (ibid., article *simulé*), LITTRE

apporte, sur le même exemple, deux précisions complémentaires. La première porte sur les moyens mis en oeuvre, qui sont des moyens artificiels ; la seconde est que le but visé par celui qui use de la simulation est de guider un choix, d'orienter une décision (provoquer une exemption de service par exemple) : "Maladies simulées, maladies dont on détermine en soi les symptômes par des moyens artificiels et qu'on paraît avoir, afin de s'exempter ainsi de remplir des devoirs imposés par la société ou par les lois".

A partir des indications fournies par LITTRE, on pourrait définir la simulation comme *la reproduction, par des moyens artificiels, de certains aspects d'un phénomène réel, en vue d'orienter une décision ou une action*. Le sens technique actuel du vocable de simulation est certes différent de la définition que nous proposons ici. Le terme utilisé en recherche opérationnelle est la transposition du mot anglo-américain, dont l'étymologie est la même, mais dont l'évolution sémantique a été légèrement différente de celle du mot français. Néanmoins, nous verrons que tous les traits que nous avons relevés dans la définition de LITTRE se retrouvent dans les acceptions techniques du terme de simulation ; ce qui peut justifier après coup son adoption par les chercheurs français.

Les définitions proposées par les spécialistes de la simulation varient d'un auteur à l'autre. Si la plupart sont, en fin de compte, complémentaires les unes des autres, il arrive que certaines s'avèrent incompatibles entre elles. Afin d'en mieux cerner les acceptions techniques, nous passerons en revue les principales définitions de la simulation que nous avons relevées. Nous nous efforcerons ensuite de préciser la notion de simulation, et de dégager les relations existant entre celle notion et les notions voisines.

1.1. EXAMEN DES DEFINITIONS PROPOSEES PAR LES SPECIALISTES.

Nous avons relevé une trentaine de définitions, dont un tiers sont proposées par des chercheurs en sciences humaines (en majorité sociologues ou psychologues), et deux tiers par des spécialistes en recherche opérationnelle. Nous analysons ci-après quelques unes des plus caractéristiques parmi les définitions proposées.

1.1.1. La définition proposée par D.G. KENDALL.

Selon H. WILLE (201, p. 319), D.G. KENDALL définit la simulation comme la "réalisation artificielle" (*artificial realisation*). A titre d'exemple, WILLE cite la technique de planification NPT (*Network Planning Technique*), dont le PERT constitue un cas particulier. Cette technique a pour but la "réalisation artificielle d'un projet situé dans l'avenir, et permet ainsi d'étudier dans le service de planification les conséquences d'évènements futurs possibles, avant que ceux-ci ne se soient produits".

Une telle définition est évidemment très pauvre, et par conséquent très large dans ses applications. Dans cette optique, le terme de simulation peut s'appliquer aux essais sur maquette (essais de maquette d'avion en soufflerie, de maquette de bateau en bassin, de modèle réduit de barrage, etc.), aux expériences réalisées sur des animaux pour en prévoir l'effet sur l'homme (greffes d'organes, brassage génétique, conditionnement et apprentissage, etc.), à la manipulation de groupes constitués artificiellement en laboratoire pour prévoir le comportement de groupes réels (psychologie sociale expérimentale), ou même aux procédés de synthèse chimique. Disons tout de suite qu'une

telle extension ne correspond pas à l'usage le plus répandu. Il faut néanmoins ajouter que cette définition a été citée dans le cadre d'un congrès sur "la simulation numérique en recherche opérationnelle" (Hambourg, septembre 1965), et qu'il pouvait sembler naturel aux participants de limiter l'acception du terme aux seules simulations sur ordinateur numérique.

Si l'on s'en tient à la lettre de la définition de D.G. KENDALL, celle-ci représente les caractères suivants :

- elle est assez proche du sens courant, en ce sens qu'elle s'applique à tout ce qui, ne pouvant être réalisé, se trouve reproduit artificiellement ;

- en conséquence, elle englobe des situations très diverses, ne présentant pas toutes un caractère de démarche scientifique (une pièce de théâtre est une réalisation artificielle d'une suite d'évènements) ;

- elle met l'accent sur l'aspect apparemment contradictoire de la simulation, en associant les notions antinomiques de réel et d'artificiel. Ce dernier point laisse entrevoir une difficulté méthodologique profonde, sur laquelle nous reviendrons (§7.2.c) : dans quelle mesure une simulation peut-elle reproduire la réalité ?

B. RENARD et J.P. RENAULT, sans citer D.G. KENDALL, s'élèvent contre une définition trop large de la simulation : "si l'on s'accorde pour donner au mot *simulation* le sens de *représentation de la réalité par des moyens artificiels*, qui d'entre nous ne passe pas son temps à *simuler* ?" (172, p. 201). Ainsi, le comptable simule des échanges de marchandises, l'architecte simule une construction future, l'ingénieur simule par le calcul le fonctionnement des mécanismes qu'il étudie. Du point de vue de l'informaticien, tout traitement sur

ordinateur deviendrait une simulation. "En effet, les calculateurs traitent des chiffres, les lisent, les transforment, etc. ; en aucun cas le calculateur ne traite la réalité elle-même" (ibid., p. 202). Le seul domaine du traitement de l'information qui pourrait à la rigueur échapper à une telle définition est l'ensemble des problèmes posés dans le cadre de la théorie des nombres : ces derniers étant eux-mêmes des abstractions, on peut considérer qu'ils ne peuvent être des objets de simulation.

Cette définition est par conséquent trop imprécise. Il est nécessaire soit d'augmenter sa compréhension, en lui ajoutant des termes qui la précisent, soit de restreindre son extension. C'est ce dernier parti qu'ont choisi B. RENARD et J.-P. RENAULT, qui énumèrent quatre cas pour lesquels on emploie en informatique le terme de *simulation* (simulation de modèles, programmés sur ordinateur numérique ; simulation d'un calculateur par un autre calculateur ; simulation simplifiée d'un calculateur selon le programme PRESTE ; simulateurs spécifiques, généralement hybrides). Une telle énumération présente le double inconvénient de ne pas constituer une définition, et d'être par nature incomplète. En effet, les auteurs ont été contraints de prévoir une cinquième catégorie ("autres"), dans laquelle ils font entrer les calculs "mathématiques" (calculs de tables numériques, recherches sur la théorie des nombres), les calculs "techniques" (calculs de trajectoires d'engins balistiques, étude de réseaux de distribution), et les applications nouvelles (économiques et biologiques). Remarquons que les applications sociologiques, démographiques, ou psychologiques, sont absentes de cette énumération ; il est vrai qu'elles constituent actuellement une partie infime des applications effectivement réalisées.

1.1.2. La définition proposée par K.M. HOGLUND et P.L. PHIPPS.

"La simulation est une procédure dans laquelle un dispositif ou objet est utilisé pour imiter un second dispositif ou objet" (114, p. 276). En proposant cette définition, HOGLUND et PHIPPS pensaient à l'utilisation d'un ordinateur numérique non spécialisé pour imiter l'utilisation, par l'armée de l'air, d'un système tactique de commande et de contrôle ayant recours à un ordinateur : le dispositif simulé comportait par conséquent un ordinateur fictif, des utilisateurs fictifs, et des données fictives. Dans l'esprit des auteurs, le dispositif simulé est par conséquent une machine, ou un système homme-machine.

Prise à la lettre, cette définition semble devoir s'appliquer :

- à toute forme d'imitation, quelle qu'elle soit ;
- à la simulation d'un objet quelconque, même inanimé ;
- aux seuls objets ou dispositifs physiques, à l'exclusion des êtres vivants.

Ainsi, un tableau représentant un portrait ou un paysage serait une simulation du personnage ou de la région représentée. En réalité, si l'on admet que toute simulation est une certaine forme d'imitation, il ne semble pas que la réciproque soit toujours vraie : la copie d'un objet n'est pas nécessairement une simulation. L'usage du mot, dans la littérature scientifique, ne s'étend pas à toute forme d'imitation, mais se restreint à certains types de relations entre l'original et sa copie.

D'autre part, on n'utilise guère le mot *simulation* qu'à propos de transformations, de processus évolutifs, ayant un certain

déroulement dans le temps ; c'est-à-dire à propos d'objets ou de dispositifs susceptibles de variations. L'original, l'objet simulé, doit par conséquent avoir un *comportement*, caractérisé par le fait de prendre successivement des états différents. L'exemple le plus élémentaire pouvant être celui d'une machine simple, composée de pièces inanimées, se mouvant les unes par rapport aux autres, et constituant de ce fait un dispositif animé. La définition de HOGLUND et PHIPPS, qui ne comporte pas de précision à ce sujet, pourrait s'appliquer à toute représentation matérielle d'un objet quelconque, comme par exemple le plan d'une ville, ou l'écriture du symbole chimique d'un corps.

Enfin, cette définition, qui nous semble par ailleurs pécher par manque de précision, ne paraît pas pouvoir s'appliquer à des êtres vivants, ou des collectivités d'êtres vivants. Selon les auteurs, la simulation s'applique à des objets (*objects*) ou des dispositifs (*devices*), c'est-à-dire à des machines. En fait, l'exemple choisi par eux est en contradiction avec cette énumération restrictive, puisque le système simulé tient compte du comportement des responsables du dispositif de défense.

1.1.3. La définition élaborée par l'I F I P et l'I C C.

Un comité de rédaction composé de membres de l'*International Federation for Information Processing* (I F I P) et de l'*International Computation Center* (I C C) a élaboré un vocabulaire de l'informatique (117), destiné à mettre de l'ordre dans un langage technique dont la croissance rapide a été particulièrement anarchique. Ce comité définit la simulation comme "la représentation de certains traits du

comportement d'un système au moyen des actions d'un autre système, par exemple la représentation de phénomènes physiques au moyen d'ordinateurs, ou même la représentation d'un ordinateur par un autre ordinateur" (op. cit., p. 6).

Cette définition apporte, par rapport à celles qui précèdent, deux précisions :

- il s'agit de reproduire le comportement (*behaviour*) d'un système, et non nécessairement le système lui-même ;
- ce n'est pas nécessairement l'intégralité de ce comportement que l'on s'efforce de reproduire : il est possible de se limiter à quelques aspects de celui-ci. Nous retrouvons ici cette notion importante déjà présente dans les analyses de LITTRE, qu'une simulation n'est pas une représentation exhaustive, mais qu'elle porte plutôt sur certains traits essentiels de la réalité.

En raison de sa concision, la définition de l'I F I P - I C C laisse subsister quelques imprécisions. En particulier, comment sont sélectionnés les traits du comportement original qui doivent figurer dans la reproduction simulée ? Quelles considérations déterminent le choix des critères de sélection ? Peut-on parler de simulation quels que soient les choix effectués ? D'autre part, de quelle nature est cette représentation ? Quels sont les liens qui doivent exister entre l'original et sa reproduction ? Sur ce dernier problème, une précision est apportée à l'article *simulateur* (ibid., p. 8), où il est mentionné qu'un simulateur est conçu de telle manière qu'il existe, entre celui-ci et le système qu'il reproduit, une *correspondance fonctionnelle*.

Une critique plus importante que nous adresserons au vocabulaire de l'I F I P - I C C est que certains termes essentiels, com-

me *comportement* ou *système*, ne sont nulle part définis. Certes, ces termes ont acquis une signification scientifique précise, en particulier chez les cybernéticiens ; mais un rappel de la définition d'un *système* présenterait l'avantage de définir les limites à l'extension du terme de *simulation*. Si nous admettons provisoirement qu'un système est défini par un ensemble de variables présentant pour l'observateur un certain caractère d'unicité, nous voyons que la définition de l'I F I P - I C C présente un caractère de grande généralité. Le système original, comme le système qui en reproduit le comportement, peuvent être indifféremment un ordinateur, une machine, un animal, un groupe social, etc.. Peuvent aussi être considérés comme des simulations le jeu de rôle, les essais d'avion en soufflerie, l'expérimentation sur les animaux, les manoeuvres militaires, etc.. Une telle généralité est à porter au crédit des informaticiens de l' I F I P - I C C, qui ont eu le souci de ne pas limiter l'acception du terme à son usage le plus répandu dans leur propre spécialité (simulation sur ordinateur).

1.1.4. Les définitions proposées par R. BOUDON.

Raymond BOUDON est, à notre connaissance, le seul sociologue de langue française qui ait tenté de définir la simulation. Nous avons relevé chez lui deux définitions complémentaires :

- "*Simuler un système* : construire un système S'(modèle) comme une suite d'évènements survenant aux éléments du système, selon des règles déterministes ou probabilistes". (31, p. 85).

- la simulation désigne "l'ensemble des procédures qui, pour analyser un système réel ou vérifier des hypothèses concernant

un système réel, ont recours à un autre système, plus simple ou plus facilement manipulable que le premier, et lié à lui par un rapport d'analogie". (60, p. 220).

Ces deux définitions nous apportent des précisions sur l'usage de la simulation, la nature du *modèle* qui reproduit le système réel, et les rapports qui existent entre l'original et son modèle simulé. Par contre, les procédures et les techniques utilisées sont quelconques, ce qui confère à la notion de simulation une extension presque aussi large que dans le cas de la définition de l'I F I P - I C C .

L'usage des techniques de simulation est bien défini : elles servent ou bien à analyser un système réel (en l'occurrence un système social), ou bien à vérifier des hypothèses sur ce système. Cependant, un tel objectif ne saurait être le caractère spécifique de la simulation. Des techniques statistiques telles que l'analyse de la variance ou l'analyse factorielle visent, elles aussi, à l'analyse des systèmes réels ; les tests statistiques reposant sur la notion de probabilité d'erreur dans l'estimation d'un paramètre ont, eux aussi, pour but la vérification d'hypothèses relatives à la réalité. De plus, le recours à un *modèle* pour analyser un système de variables ou vérifier un ensemble d'hypothèses n'est pas non plus un trait spécifique des techniques de simulation, puisque les techniques statistiques que nous avons citées se fondent toutes sur un modèle mathématique. Retenons cependant que la simulation est une des méthodes qui permettent d'analyser un système réel ou de valider des hypothèses en recourant à un *modèle* de ce système réel. Dans un autre texte, R. BOUDON précise d'ailleurs qu'en première approximation, "les modèles simulés peuvent être considérés comme une extension des

modèles mathématiques" (33, p. 381; 32, p. 3).

La nature du modèle est en partie différente de celle de l'original. Le modèle est également un système, mais un système *construit*, donc en quelque sorte artificiel, et moins complexe que le système réel qu'il représente. Le modèle évolue dans le temps, puisqu'il est conçu comme une suite d'évènements survenant au système original. L'apparition ou l'enchaînement de ces évènements obéit à des règles précises (déterministes ou stochastiques), fixées par le chercheur. Comment ces règles sont-elles formulées ? Ce problème de la construction du modèle est lié à celui des rapports entre le modèle et le phénomène original.

La relation du modèle avec le système réel original est une relation d'*analogie* ; c'est-à-dire qu'il existe entre eux une "ressemblance plus ou moins lointaine" (32, p. 51). Cette ressemblance est voulue par le chercheur, qui a construit le modèle "à l'image" de la réalité, tout en s'efforçant de le rendre plus simple et plus manipulable que celle-ci. Nous verrons au chapitre suivant (§ 2.2.4.d) quelles sont les procédures qui permettent au chercheur de construire le modèle ; retenons par contre la notion, suggérée par R. BOUDON dans sa seconde définition, de la manipulation du modèle. C'est là une idée de base de la simulation, qui rend possible ce que BOUDON appelle la "quasi-expérimentation" (32, p. 20). Nous retrouverons cette idée dans la définition suivante, de J. GUITTET.

1.1.5. La définition proposée par J. GUITTET.

"La connaissance d'un système quelconque S ne peut être acquise que par celle de l'ensemble R de ses réactions à un ensemble

A d'actions exercées sur ce système. On a : $R = S(A)$. Chaque fois que l'on aura réalisé un système Σ , dont les réactions à certaines actions, dont l'ensemble est une partie A' de A , sont voisines de celles de S , on dira que Σ simule le système S . Σ constitue le modèle de S , relativement à la partie A' ." (108, p. 191).

Cette définition semble, au premier examen, peu différente de celle de l'I F I P - I C C. Elle introduit cependant une notion essentielle : celle d'*expérimentation sur le modèle*, déjà suggérée dans l'une des définitions de R. BOUDON. GUITTET rappelle implicitement les liens qui unissent un système et son environnement : le comportement du système est l'ensemble des actions qu'il exerce sur son environnement. Pour connaître un système, il faut non seulement connaître les actions que celui-ci exerce sur l'environnement, mais également connaître les actions que l'environnement exerce sur le système ; il devient ainsi possible de relever des séquences de comportement permettant d'identifier les réactions du système à son environnement. C'est là une démarche de type expérimental.

En pratique, l'expérimentateur (qui est un élément de l'environnement du système) n'agit pas au hasard ; il s'efforce au contraire de sélectionner une partie des actions possibles sur le système (stimuli), afin d'en isoler les effets (réactions du système, ou réponses). Si l'on remplace le système étudié par un autre système, tel que ses réponses aux stimuli sélectionnés par l'expérimentateur soient semblables à celles du système étudié, le second système simule le système étudié relativement aux stimuli sélectionnés. Il s'ensuit que :

- en dehors du plan d'expérience défini préalablement, le second système ne peut a priori être considéré comme un modèle du

premier ;

- dans le cadre du plan d'expérience, les deux systèmes sont équivalents, puisque, sur chacun d'eux, des stimuli semblables donnent des réponses semblables.

Il est par conséquent possible de construire un modèle Σ plus simple que le système original S, puisqu'il suffit que Σ se comporte comme S relativement au sous-ensemble de stimuli A'. "Simuler un système physique S vis-à-vis d'un certain ensemble d'actions A', c'est obtenir artificiellement, à l'aide d'un système Σ plus simple que S, les réponses que donnerait S aux actions de A' " (ibid). Il devient dans ces conditions possible d'expérimenter sur Σ comme on expérimenterait sur S.

On peut reprocher à J. GUITTET de ne pas définir avec précision ce qu'il entend par "réactions *voisines* de celles du système réel". Il est vrai que ce point soulève le problème de l'ajustement du modèle à la réalité, problème qui sort un peu du cadre d'une définition. D'autre part, selon cette définition, le modèle pourrait, à la limite, se confondre avec le phénomène réel : plus le sous-ensemble A' des actions envisagées sur le système simulé est proche du système réel. Si A' est identique à A, peut-on dire que Σ est identique à S ? En d'autres termes, un "sosie" du système réel est-il encore un modèle, ou bien est-il lui aussi un système réel ?

1.1.6. La définition proposée par R.E. DAWSON.

Chercheur en sciences politiques, R.E. DAWSON s'est efforcé, après avoir passé en revue un certain nombre d'applications de la si-

mulation dans des domaines très divers des sciences sociales, de donner une définition générale à la simulation compatible avec ces applications. "En tant que technique de recherche des sciences sociales, la simulation consiste en l'élaboration et la manipulation d'un modèle *opératoire*, ce modèle étant une représentation physique ou symbolique de tous ou de certains aspects d'un processus sociologique ou psychologique. Pour le spécialiste en sciences sociales, la simulation est la construction d'un modèle opératoire d'un processus individuel ou collectif, et l'expérimentation sur cette maquette par la manipulation des variables et de leurs interrelations" (61, p. 3).

Cette définition rassemble la plupart des traits principaux de celles que nous avons examinées auparavant. Peut-être lui reprocherait on de n'envisager le cas que d'un processus. La notion de *modèle opératoire* est définie par ailleurs comme la représentation du comportement d'un système, visant à reproduire le déroulement du processus de ce système (62, p. 189) ; il s'agit donc de la représentation dynamique d'un phénomène dynamique. Enfin, il faut noter que le modèle peut être soit un dispositif physique, soit une représentation symbolique, comme un modèle mathématique.

1.1.7. La définition proposée par J. AGARD.

"Une simulation reproduit à vitesse accélérée l'évolution temporelle d'un système en tenant compte d'aléas pour donner sur le comportement du système des renseignements que d'autres méthodes ne fourniraient pas, si ce n'est à un prix de revient supérieur" (6, p. 2). Ajoutons que J. AGARD réserve explicitement l'usage du mot *simulation* aux modèles stochastiques de phénomènes stochastiques

(ibid). Ces conceptions sont reprises sans grand changement par E. TORCHET (192, p. 388).

Cette définition s'avère beaucoup plus spécifique que celles qui la précèdent. En particulier, le phénomène naturel doit présenter certaines caractéristiques aléatoires, que le modèle doit reproduire sans les soumettre à des lois déterministes. Mais le trait qui nous semble le plus important est la *contraction du paramètre temps*. De ce point de vue, une telle définition cadre parfaitement avec les divers types de simulation sur ordinateur, dans lesquels une campagne militaire ou la construction d'un grand ensemble, qui pourront durer plusieurs années sur le terrain, sont simulées en quelques heures.

Une telle définition écarte par conséquent toutes les simulations en temps réel, comme les jeux d'entreprise, ou l'utilisation des simulateurs de vol. Elle écarte également toutes les simulations en temps accéléré dans lesquelles le phénomène réel, ou bien le modèle, n'obéissent qu'à des lois déterministes. Elle écarte enfin, mais il s'agit de cas d'espèce, toutes les simulations en temps accéléré de phénomènes aléatoires par un modèle aléatoire, si le problème peut être traité, à précision égale, à un moindre coût par d'autres méthodes (analytiques par exemple). Il est pourtant intéressant de souligner que, malgré ces multiples contraintes, on peut imaginer des expériences en accord avec la définition de J. AGARD, et ne recourant pas à l'usage de l'ordinateur. Par exemple l'étude de la transmission des caractères héréditaires chez l'homme par analogie avec les résultats de croisements opérés sur la mouche *Drosophile* ou sur le champignon *Neurospora*.

1.1.8. La définition proposée par P. ROSSI.

"Une simulation est une expérience conduite sur un calculateur électronique à l'aide d'un modèle de simulation. (...) Un modèle de simulation est un triplet (Π, E, f) où Π est un processus stochastique (Ω, A, P) . E est l'ensemble du système et f une application de l'espace des trajectoires de Π dans l'espace F des évolutions du système". (179, p. 11).

P. ROSSI considère que le mot français *simulation* traduit l'expression anglaise *digital simulation*. L'instrument utilisé pour faire fonctionner le modèle est donc par définition l'ordinateur numérique. D'autre part, un modèle de simulation est nécessairement pour P. ROSSI le modèle mathématique d'un processus aléatoire (un modèle mathématique étant un ensemble d'objets et de relations). Bien que ne faisant pas intervenir la relation entre le temps réel et le temps simulé, la définition de P. ROSSI se range parmi les plus restrictives : seules méritent d'être appelées simulation les expériences étudiant le comportement un modèle stochastique à l'aide d'un ordinateur numérique.

1.2. DEFINITION DU DOMAINE ETUDIE.

A l'issue d'une revue qui est loin d'être exhaustive, nous ne pouvons que constater l'extrême hétérogénéité des définitions proposées pour la notion de simulation. Les définitions les plus larges englobent la plupart des démarches de toute recherche scientifique, quel qu'en soit le champ d'application (surtout si l'on admet, avec

Robert OPPENHEIMER, que "la science est analogie" (cité par 145, p. 288) ; les plus spécifiques se limitent à un exemple particulier d'utilisation de l'ordinateur numérique pour l'étude d'un phénomène physique (ou économique) de nature aléatoire. Il faut en outre, souligner le fait que cette hétérogénéité ne peut être expliquée ni par les différences de formation des chercheurs, ni par des différences de langue ou de culture : la même diversité se retrouve par exemple chez les sociologues ou les physiciens, et plus nettement encore, chez les chercheurs américains, anglais, ou français.

On comprend mieux pourquoi certains spécialistes de la simulation parmi les plus notables, comme K.D. TOCHER (*The Art of Simulation*), D.N. CHORAFAS (*La simulation mathématique et ses applications*), ou W.R. ASHBY (*Design for a Brain*), se sont gardés de proposer une définition d'une notion qui se trouve pourtant au centre de leurs travaux. Sans doute craignaient-ils, en ajoutant une définition de plus à la liste des définitions proposées, d'accroître la confusion de la terminologie, ou de paraître prendre trop vivement parti dans des querelles académiques.

1.2.1. Démarche suivie pour aboutir à une définition.

Pour poursuivre notre recherche, il est indispensable d'introduire plus de précision et de clarté dans les termes que nous utiliserons, fût-ce au prix d'un certain arbitraire. Il nous faut donc choisir parmi les définitions proposées, ou forger notre propre définition. Nous pouvons a priori écarter les définitions extensives (c'est-à-dire par énumération exhaustive des cas particuliers auxquels

s'appliquerait le vocable de simulation) : dans le cas d'une technique en évolution extrêmement rapide, et pouvant s'appliquer à une grande variété de domaines, ce type de définition risquerait d'être trop restrictif (comme nous l'avons signalé, au § 1.1.1., à propos de la définition proposée par B. RENARD et J.-P. RENAULT), et vraisemblablement périmé dès l'apparition de formes nouvelles de la simulation. Choisir une définition compréhensive est par conséquent la seule démarche qui nous permette de préciser l'objet de notre recherche sans l'assujettir aux seules réalisations passées.

Une telle option nous impose de trancher entre des caractéristiques dont certaines sont très techniques, et à propos desquelles les spécialistes ne sont pas d'accord entre eux : par exemple, devons-nous refuser d'utiliser le terme de *simulation* dans le cas d'un modèle non dynamique, comme le suggère R. CRUON (56, p. 386) ? ou lorsque le mécanisme interne du modèle n'est pas la copie fidèle du mécanisme de l'original, comme le fait R.F. SIMMONS (185, p. 363), en opposant *synthèse* à *simulation* ? ou encore chaque fois que le modèle simulé et le phénomène original ne sont pas tous deux de nature aléatoire (J. AGARD, 6, p. 2) ? Dans ces conditions, il nous faut préalablement définir des critères nous permettant de décider dans quels cas nous pourrions parler de simulation, et dans quels cas nous refuserons l'usage de ce terme.

Le choix de ces critères doit tenir compte de deux facteurs : le premier est que nous ne disposons pas de la qualification nécessaire pour prendre parti sur le fond de certains des problèmes techniques auxquels nous avons fait allusion ; le second est que le but de notre recherche est essentiellement l'étude des possibilités et des contraintes liées à l'utilisation d'une technique particulière, plus que l'ana-

lyse linguistique des diverses acceptions d'un mot. Aussi avons-nous adopté une démarche empirique, que l'on peut schématiser ainsi :

- recenser les exemples d'application que l'on a qualifiés de "simulation" ;

- parmi ceux-ci, sélectionner les cas auxquels le terme de simulation nous paraît pouvoir raisonnablement s'appliquer (*) ;

Nous réalisons ainsi une première partition (partition *a priori*) de l'ensemble des exemples d'application ;

- dégager les caractéristiques communes à la majorité des cas sélectionnés, et formuler à partir de celles-ci une définition provisoire ;

- réaliser une seconde partition (partition *a posteriori*) de l'ensemble des exemples d'application recensés, selon qu'ils correspondent ou non à la définition ;

- comparer les deux partitions et recenser les cas de désaccord (c'est-à-dire prendre la différence symétrique des deux sélections : exemples sélectionnés ne concordant pas avec la définition, et exemples conformes à la définition n'ayant pas été préalablement sélectionnés) ;

- analyser les caractéristiques des cas de désaccord de manière à réaliser une nouvelle partition empirique ;

- dégager une nouvelle définition provisoire, et poursuivre la démarche jusqu'à ce qu'il y ait recouvrement de la partition *a posteriori*, et qu'ainsi la définition adoptée englobe toutes les caractéristiques communes à tous les exemples retenus.

(*)

Les critiques que nous avons adressées aux définitions que nous avons examinées dans la première partie de ce chapitre, laissent entrevoir quels ont été nos critères implicites de sélection : nous n'avons retenu que les exemples paraissant, à première vue, compatibles avec la définition de J. GUITTET ou avec celle de R.E. DAWSON.

Avant d'aborder les critiques qu'il convient de faire de cette démarche, nous ferons deux remarques. La première remarque porte sur l'échantillon d'exemples de simulation que nous avons recensé ; le nombre d'exemples publiés en langue française et anglaise est certainement actuellement de l'ordre de plusieurs milliers. Il ne pouvait être question de les étudier tous. Nous avons limité notre recensement à tous les exemples de simulation en sociologie auxquels nous avons pu avoir accès, et à quelques exemples parmi les plus connus dans d'autres disciplines (en particulier en psychologie, démographie, recherche opérationnelle). La seconde remarque est relative à notre procédure de sélection *a priori* au sein de cet échantillon. Pour juger si le terme de *simulation* pouvait raisonnablement s'appliquer à un exemple donné, nous avons considéré la nature des techniques mises en oeuvre. Ainsi, pour réaliser la première sélection *a priori*, nous avons tenu le raisonnement suivant : quelle que soit la définition que l'on donne du terme de simulation, il est incontestable que cette technique a connu un essor considérable grâce à la recherche opérationnelle. Celle-ci est née aux environs de 1940 (cf. F.N. TREFETHEN, ch. 1 de 144). Au cours de notre première sélection, nous avons par conséquent éliminé tous les exemples mettant en oeuvre des techniques antérieures à cette date. Ce parti pris nous a conduit à éliminer les expériences de pionniers comme I. FISHER, ou de certains chercheurs "marginiaux" comme R.P. RICH ; nous les avons ensuite réintégrés au cours des phases ultérieures de notre démarche. Mais en dernière analyse, ce parti-pris s'est avéré nous avoir fourni un point de départ fécond.

La critique fondamentale que l'on peut adresser à notre démarche est qu'elle présente toutes les caractéristiques du cercle vicieux (*petitio principii*) : nous avons classé les expériences re-

censées selon une définition implicite, et nous avons formulé cette définition à partir de la sélection ainsi opérée ; nous n'avons par conséquent rien fait d'autre qu'explicitement notre critère de classification. Cette critique est en partie fondée. Nous avons en effet un double souci au départ : éliminer du domaine de la simulation certaines techniques de laboratoire fondées sur l'expérimentation sur des êtres vivants (psychophysiologie animale, techniques de groupe), et conserver toutes les expériences recourant à des dispositifs physiques pour simuler des phénomènes naturels. Ce souci était guidé moins par les définitions du mot simulation que par les usages qu'en étaient faits le plus fréquemment ; il était en outre renforcé par la crainte d'aboutir à une définition soit trop large, soit trop spécifique. Il faut remarquer néanmoins que notre démarche ne s'est pas traduite par la confirmation de la première sélection *a priori*, mais a comporté au contraire de nombreux tâtonnements, et des réajustements tantôt de l'extension, tantôt de la compréhension du concept analysé. En effet, d'une part notre définition implicite n'était pas parfaitement claire, et notre première sélection ne pouvait donc être exempte d'erreurs ou d'ambiguïtés ; d'autre part, la formulation d'une définition reposait sur les caractéristiques propres non pas à tous les exemples considérés, mais seulement à la majorité d'entre eux ; enfin, certaines de ces caractéristiques pouvaient être présentes à la fois chez certains exemples sélectionnés et chez certains exemples éliminés. Ainsi, s'il est vrai que la conception que nous avons *a priori* de la simulation a joué un rôle important dans les premières phases de notre démarche, son influence s'est trouvée atténuée au fur et à mesure que notre définition se précisait. D'autre part, l'intérêt de cette définition est surtout de nous fournir un cadre commode pour la suite de notre recherche, et, à ce titre, la formulation à laquelle nous avons abouti nous a paru suffisante.

1.2.2. Proposition d'une définition de la simulation.

La simulation est une procédure utilisée pour remplacer l'expérimentation sur un système réel, consistant à expérimenter sur un système quasi-isolé déterminé quelconque, constituant un modèle du système réel étudié.

Cette définition fait appel à plusieurs concepts qu'il convient de définir avec précision : *système, système quasi-isolé déterminé, modèle, expérimentation*. Après avoir précisé le sens que nous attribuons à chacun de ces concepts, nous pourrions délimiter l'extension que nous donnons au terme de simulation, et définir ainsi le domaine de notre étude.

1.2.3. Qu'est-ce qu'un système ?

La notion de *système* est l'un des concepts de base de la cybernétique ; afin de définir clairement cette notion, il nous faut la situer brièvement par rapport à la discipline qui l'utilise. "Science du contrôle et de la communication dans l'animal et dans la machine" (199), la cybernétique étudie le comportement des *mécanismes* (*), c'est-à-dire de "choses qui peuvent changer dans le temps" (12, p. 13). Ces mécanismes peuvent être fictifs ou réels, naturels (non construits par l'homme) ou artificiels (55, p. 25) ; (103, p. 26) ; ces distinctions importent peu au cybernéticien, dont le domaine d'étude

(*) Conformément à l'usage établi par G.T. GUILBAUD (107, p. 12) et L. COUFFIGNAL (55, p. 25-26), nous avons réservé le terme de *machines* à des "corps organisés construits par l'homme", c'est-à-dire à des "mécanismes finalisés". Nous utilisons donc le terme de *mécanisme* pour traduire le mot anglais "*machine*", contrairement à l'usage du traducteur français d'ASHBY.

est l'ensemble de tous les mécanismes possibles (13, p. 3).

Le changement est donc une propriété intrinsèque des mécanismes. Pour un terme donné (pour un aspect donné du mécanisme étudié), un changement est une transition d'un état à un autre, sous l'action d'un opérateur. Par exemple (13, Pp. 12-13), sous l'action de l'opérateur "soleil", le terme "peau" passe de l'état "blanche" à l'état "brune" ; on dira que "peau brune" est le *transformé* du terme "peau blanche". Lorsqu'un opérateur agit sur plusieurs termes à la fois, l'ensemble des transitions sur l'ensemble des termes constitue une *transformation* ; ainsi, l'opérateur "exposition au soleil", agissant sur le groupe de termes "sol", "pellicule photo-sensible", "pigment", produira le groupe de trois transitions (transformation) "sol froid → sol chaud", "pellicule vierge → pellicule exposée", pigment coloré → pigment décoloré".

Lorsque l'action d'un opérateur sur un groupe de termes ne crée aucun élément nouveau, c'est-à-dire qu'il existe pas de transformé qui ne soit présent dans le **groupe** des termes, le groupe des termes est invariant dans la transformation correspondante ; on dit alors que cette transformation est *close*. Lorsque chaque terme a un seul transformé, on dit que la transformation est *univoque*. Un mécanisme qui suit dans son comportement une transformation close et univoque est un mécanisme *déterminé*. Comme exemples de mécanismes déterminés, ASHBY cite (13, Pp. 31-33) une pendule en bon état et remontée, et le comportement des épinoches au moment de la ponte (d'après TINBERGEN).

La plupart des mécanismes naturels sont généralement complexes, et constitués d'une quantité de mécanismes plus simples et plus élémentaires, liés entre eux, et réagissant les uns sur les autres ; l'état d'un mécanisme complexe est constitué par les états de ses éléments

constitutifs (22, p. 19). Pour étudier le comportement d'un mécanisme complexe, il est nécessaire d'observer le comportement de chacune de ses parties ; on est ainsi conduit à identifier un certain nombre de *variables*. Par exemple (12, p. 14), pour étudier le fonctionnement d'une horloge à poids et balancier, on pourra prendre en compte, à chaque instant de l'observation, l'écart du balancier par rapport à la verticale, sa vitesse angulaire, la hauteur du poids, la position de l'aiguille des minutes, etc..

Tout mécanisme réel nécessiterait, pour être décrit complètement, la prise en compte d'une infinité de variables (12, Pp. 15-16) ; en pratique, un petit nombre seulement de variables est retenu par le chercheur. Cette sélection dépend évidemment du domaine d'intérêt du chercheur, de ses objectifs en effectuant la recherche, et de ses conceptions théoriques. C'est la liste des variables prises en compte par le chercheur qui constitue un *système* : "un système est défini comme un ensemble quelconque de variables que le chercheur sélectionne parmi celles qui sont disponibles dans le mécanisme réel" (ibid, p. 16).

La notion de *système* s'avère donc être une notion essentiellement *opératoire*. En effet, le système est une abstraction distincte du mécanisme réel, qu'il sert à décrire commodément ; il est donc possible de modifier cette abstraction selon les besoins de la démarche du chercheur. Il arrive souvent, au cours d'une recherche, que le système initialement défini se révèle impropre à rendre compte de manière satisfaisante du mécanisme observé. En effet, l'objectif de toute recherche scientifique est de découvrir certains *invariants*, c'est-à-dire d'identifier des transformations closes et univoques. ASHBY (13,

Pp. 48-51) cite l'exemple d'un expérimentateur étudiant le comportement d'un balancier ; son système est défini par l'écart angulaire du balancier par rapport à la verticale. Il enregistre la position du balancier à chaque quart de seconde, à partir d'un état initial correspondant à un écart de 30° ; il relève ainsi les positions 30° , 10° , -24° . Dans le cadre de la définition adoptée pour le système, la description de la transformation observée est :

$$\begin{array}{cc} & 30^\circ & 10^\circ \\ \downarrow & 10^\circ & -24^\circ \end{array}$$

Afin de vérifier la transition $10^\circ \rightarrow -24^\circ$, l'expérimentateur place le balancier à une position initiale de 10° ; il enregistre alors la transition $10^\circ \rightarrow -3^\circ$, qui est incompatible avec la précédente. La transformation n'est donc pas univoque. La solution pour le chercheur consiste à *redéfinir le système*, en modifiant le choix des variables qui le constituent. Dans l'exemple choisi, le système "écart angulaire - vitesse angulaire" permettrait d'aboutir à une transformation close et univoque.

1.2.4. Différents types de systèmes.

Dans notre définition de la simulation, nous avons adjoint au mot *système* trois qualificatifs : *réel*, *quasi-isolé*, *déterminé* ; nous précisons ci-après le sens que les cybernéticiens attribuent à ces mots.

a) Systeme réel

Nous avons signalé que le domaine de la cybernétique s'étendait à l'ensemble des mécanismes possibles, qu'ils soient réels ou fictifs. Par *mécanisme réel*, nous entendons tout mécanisme, naturel ou artificiel, existant actuellement ou ayant existé ; un mécanisme fictif

est un mécanisme possible n'ayant jamais existé. Par *système réel*, nous entendons tout système correspondant à un mécanisme réel, qu'il soit naturel ou artificiel. Nous avons vu que le nombre des variables nécessaires pour décrire complètement un mécanisme réel quelconque (cortex, fourmilière, système économique d'une société humaine, galaxie, machine) est infini. Pour un mécanisme réel donné, l'ensemble de systèmes possibles (c'est-à-dire l'ensemble des parties de l'ensemble des variables) est également infini ; or, chaque système appartenant à cet ensemble est un système réel.

b) Systeme quasi-isolé.

Il est possible de généraliser le concept de *transformation*, en considérant que les éléments d'une transformation peuvent être non seulement des états, mais également des transformations. Dans ces conditions, nous pouvons définir une transformation $U : T_1 \rightarrow T_2$, où T_1 et T_2 sont des transformations. A titre d'exemple, ASHBY (13, p. 53) suppose qu'un enfant dispose d'une machine fabriquée en méccano ; le fonctionnement de cette machine peut être décrit par la transformation T_1 . L'enfant désire construire une machine différente, dont le fonctionnement correspond à la transformation T_2 . Pour cela, il démonte la première machine, et utilise les éléments de celle-ci pour construire la seconde ; cette opération correspond à la transformation U , qui est la transition de T_1 à T_2 .

Si l'on pousse plus avant l'analyse de la notion de transformation généralisée, on s'aperçoit que cette notion est indispensable pour rendre compte de la plupart des mécanismes réels, et en particulier de la classe importante des machines (mécanismes artificiels). Prenons par exemple un tour automatique à métaux, et

limitons notre observation aux modifications que son fonctionnement fait subir à la pièce à usiner : les variables qui définissent notre système sont un petit nombre de dimensions de la pièce à usiner, et l'évolution du système peut être résumée par une seule transformation close et univoque, liant l'état initial de la pièce à son état final. Si nous considérons maintenant un tour-revolver automatique, chaque séquence d'opérations correspondant à une position de la tourelle porte-outils peut être décrite par une transformation close et univoque ; à l'outil numéro 1 correspond la transformation T_1 , à l'outil numéro i la transformation T_i . Si l'on veut décrire l'ensemble des opérations nécessaires pour l'usinage d'une pièce, on devra avoir recours à une transformation ayant pour éléments certaines transformations T_i . Cette transformation U correspond à une intervention extérieure au système lui-même : positionnement de la tourelle, mise en place d'un gabarit ; on parlerait plus justement d'une action sur le système de son environnement. On dit que l'indice i des transformations T_i est un *paramètre* du système ; ce paramètre constitue une *entrée* du système.

Notre analyse de la notion de transformation nous a conduit à introduire l'idée d'*environnement* d'un système. Par *environnement*, nous entendons "les variables du milieu extérieur dont les changements affectent le système, et celles qui sont modifiées par le comportement du système" (en nous inspirant des définitions figurant dans 12, p. 36, et 55, p. 15). A la notion d'*entrée* du système, il nous faut ajouter celle de *sortie*, correspondant à l'action exercée par le système sur la partie du milieu extérieur qui constitue son environnement.

D'un point de vue théorique, on peut distinguer deux clas-

ses importantes de systèmes, selon la nature de leurs interactions avec leur environnement (103, p. 1) :

- les *systèmes isolés*, qui ne sont pas influencés par le monde extérieur, et qui n'exercent aucune influence sur celui-ci ;

- les *systèmes quasi-isolés*, qui sont influencés par le monde extérieur, mais seulement selon certains processus déterminés (entrées), et qui exercent une influence sur celui-ci, mais seulement selon certains processus déterminés (sorties).

Sans que cela ait été démontré de manière satisfaisante (cf. 103, Pp. 6-7), on admet généralement que seuls existent dans la nature des systèmes quasi-isolés ; au point que pour H. GRENIEWSKY, l'objet de la cybernétique ne comprend que des systèmes quasi-isolés (103, p. 29). A nos yeux, l'intérêt majeur des systèmes quasi-isolés réside dans le fait qu'ils sont théoriquement manipulables ; il est donc possible (au moins théoriquement) d'expérimenter sur eux. Cette possibilité transparait d'ailleurs dans la terminologie adoptée par certains auteurs, qui utilisent indifféremment "stimulus" pour "état distinct d'une entrée", et "réponse" pour "état distinct d'une sortie" (103, p. 2), ou dans l'illustration de la notion de système par des schémas expérimentaux (204, p. 23).

c) Système déterminé

L'objectif de l'étude d'un système (par observation ou expérimentation) est de dégager des lois établissant des relations entre les entrées du système et ses sorties ; par exemple, l'étude de la "boîte noire" a pour objet de déterminer les connexions reliant les bornes de sortie aux bornes d'entrée (13, Pp. 105-114). Cet objectif est en fait l'étude de l'action de l'environnement sur le

ystème. Dans le cas particulier, mais important à la fois des points de vue théorique et pratique, où l'environnement est l'homme, et où le système est le milieu extérieur sur lequel s'exerce son action, on distingue trois types de comportement du système (55, Pp. 19-20) :

- comportement *passif* : les réponses du système sont celles que présume le programme d'action ;

- comportement *déterministe* : les réponses du système ne sont pas nécessairement celles que présume le programme, mais elles sont liées aux actions possibles des agents d'exécution par des lois connues ;

- comportement *aléatoire* : les réponses du système ne sont pas nécessairement celles que présume le programme, et l'on ne connaît pas de loi les reliant aux agents d'exécution.

Dans l'optique qui est la nôtre, il n'est pas nécessaire de distinguer entre comportement passif et comportement déterministe : dans chaque cas, les réponses du système dépendent nécessairement des actions que l'environnement exerce sur lui. Par contre, il nous paraît indispensable de distinguer deux types de comportement aléatoire :

- comportement à "*déterminisme statistique*" : les réponses du système sont liées aux actions possibles de l'environnement par des relations stochastiques. Le comportement du système présente des fluctuations autour d'un "comportement moyen" : il est possible d'énoncer des lois liant les états de sortie du système à ses états d'entrée, mais ces lois s'expriment non en termes de déterminisme pur, mais en termes de "liaisons les plus probables" (cf. 195, Pp. 165-171).

- comportement *indéterminé* : on ne peut déceler aucune relation de probabilité entre les actions de l'environnement et les réponses du système.

Nous distinguerons donc trois types de systèmes : les systèmes déterminés, les systèmes pseudo-déterminés, et les systèmes indéterminés. H. GRENIIEWSKY (103, Pp. 3-4) définit clairement les deux premiers types :

- un *système déterminé* est un système dont le répertoire de chaque entrée comporte au moins deux états distincts, et dont l'état distinct actuel de toute sortie est toujours déterminé de façon univoque par les états distincts, passés et actuels, de toutes les entrées du système ;

- un *système pseudo-déterminé* est un système dont le répertoire de chaque entrée comporte au moins deux états distincts, et dont l'état distinct de toute sortie est toujours déterminé par les états distincts, passés et actuels, de toutes les entrées du système, avec une probabilité constante donnée. GRENIIEWSKY impose à cette probabilité d'être toujours supérieure à 0,5.

Il faut souligner que les systèmes déterminés ne se limitent pas au cas particulier où l'état d'un système est déterminé entièrement par l'état immédiatement précédent. Ce cas ("state-determined system", cf. : 12, Pp. 26-28) est certainement le plus facile à étudier, probablement le seul que l'on sache bien traiter (l'Introduction à la Cybernétique d'ASHBY énonce les règles d'analyse de tels cas, mais d'eux seuls), et certainement le type de système que l'on cherche préférentiellement à identifier dans le monde naturel. Cependant, la définition générale du système déterminé comporte explicitement la possibilité de réponse différée, d'influence de séquences de stimuli sur la réponse, c'est-à-dire, en dernière analyse de mémoire.

1.2.5. Le concept de modèle.

Le recours à l'utilisation de *modèles* est l'un des traits dominants de la démarche scientifique dans les sciences expérimentales ; aussi retrouve-t-on cette notion dans les sciences les plus diverses. La diversité des usages de ce concept entraîne la diversité des acceptions : après avoir passé en revue sept définitions différentes de la notion de modèle, définitions empruntées à des domaines d'application aussi différents que la logique mathématique, la physique, la statistique mathématique, et les sciences sociales, P. SUPPES (188, Pp. 163-171) s'interroge sur la possibilité de regrouper les multiples usages du mot *modèle* sous un concept unique. Et certes il est permis de mettre en doute cette possibilité lorsque l'on confronte par exemple quelques unes des définitions proposées par les logiciens (143, Pp. 83-85 ; 87), des analyses effectuées par les statisticiens (88), les physiciens (104 ; 65), les psychologues (180, Pp. 4-8), et la description de certains modèles proposés par les économistes (102 ; 123, ch. IV et VI), les psychologues (17 ; 180) ou les sociologues (47 ; 11 ; 133).

C'est pourquoi L. APOSTEL (7) considère le problème de la définition du concept de *modèle* d'un point de vue plus pragmatique. Limitant son analyse aux seules sciences expérimentales, il décrit neuf cas d'utilisation de modèles :

- *formation d'une théorie*, en remplaçant l'étude d'un système inconnu par celle d'un système connu présentant avec le premier une certaine analogie ;
- *simplification* d'une théorie actuellement insoluble par le calcul ;
- *réduction* de deux théories distinctes à un modèle commun ;
- *extension* d'une théorie confirmée, mais incomplète, par l'étude d'un modèle qui la prolonge (par exemple en la quantifiant) ;

- *évaluation de l'adéquation* d'une théorie générale à des faits scientifiques nouvellement découverts ;

- *explication* d'un ensemble de faits pour lesquels on ne dispose que d'une théorie descriptive ;

- *concrétisation* d'un système trop grand, trop petit, trop éloigné, ou trop dangereux pour se prêter à l'investigation scientifique, sous la forme d'un modèle manipulable ou observable sans difficulté ;

- *globalisation* d'une théorie sous forme d'un modèle permettant à l'esprit de se la représenter comme un tout ;

- *relai* entre un niveau d'abstraction éloigné de l'expérience et le niveau de l'observation ou de l'action.

A la lumière de cette description, qui ne prétend nullement à l'exhaustivité, APOSTEL constate que le concept de modèle apparaît, dans les sciences expérimentales, comme un outil servant de médiateur entre deux facteurs : entre une théorie et une autre, entre un ensemble d'expériences et une théorie, entre deux ensembles d'expériences, ou entre des structures intellectuelles et les sujets qui les utilisent. A partir de cette analyse, APOSTEL tente de dégager des traits communs aux diverses fonctions que remplissent les modèles dans les sciences expérimentales, afin d'aboutir à une description unique du concept multiforme de modèle. Cette description est la suivante : "lorsqu'un sujet quelconque, pour obtenir de l'information sur un système B, utilise un système A qui n'a avec le système B aucune interaction ni directe, ni indirecte, (on dit que) ce sujet utilise A en tant que modèle de B" (7, p. 36).

Quelle peut être la nature d'un modèle correspondant à la description d'APOSTEL ? Dans la plupart des usages du mot *modèle*

(sans qualificatif) que nous avons relevés, le modèle est un système *physique*, généralement artificiel (machine). Par exemple :

- "Système artificiel, aussi peu compliqué que possible, suffisamment analogue par son fonctionnement à un organe ou à un système plus complexe (...) pour le représenter dans les limites de précision qu'on s'est fixées" (205, p. 202).

- "Représentation d'un objet réel ou d'un processus, construite généralement avec une forme, des dimensions, et/ou une matière différentes" (62, p. 189).

- l'expérimentation est réalisé par "l'étude de modèles dans des bassins de carène, des tunnels aérodynamiques, sur des maquettes d'ouvrages hydrauliques, ou à l'aide d'analogies électriques" (54, p. 183). D'autre part, certains auteurs ont jugé bon d'étendre la notion de modèle à "une phrase ou un système de phrases" (194, p. 178). Cette acception englobe toute description formalisée d'un système, et en particulier la description en langage mathématique ("modèles mathématiques"). Elle est résumée par la définition proposée par R.A. BRODY : "Un modèle est un ensemble d'assertions concernant une réalité passée, présente, ou prévue. C'est un ensemble de propositions qui prétend décrire la structure des relations existant entre les composants -unités et variables- de cette réalité" (36, p. 191).

En restant dans le cadre de la description d'APOSTEL, nous pouvons donc avec COUFFIGNAL (55, Pp. 107-109) distinguer deux types de modèles :

- les *modèles dialectiques* (*), qui sont "la description des phénomènes dans la langue ordinaire complétée par des termes techniques et des éléments de langages techniques spéciaux, notamment des mathématiques" (ibid., p. 107) ;

- les *modèles physiques*, "mécanismes artificiels comportant certaines analogies avec un mécanisme donné et ayant pour but de faire apparaître d'autres analogies" (ibid., p. 53).

L'étude des relations qui sont établies par le chercheur entre le système qu'il étudie et le modèle construit pour cette étude nous permettra de définir les rapports qui existent entre ces deux types de modèles.

1.2.6. Les relations entre le système étudié et son modèle.

Pour comprendre la nature des liens qui unissent le modèle au système original, il convient de rappeler la place du modèle dans la démarche du chercheur. Rappelons que l'on peut schématiser la

(*)

"Ce mot a reçu des acceptions si diverses qu'il ne peut être employé utilement qu'en indiquant avec précision en quel sens il est pris. Encore y a-t-il lieu de se défier, même sous cette réserve, des associations impropres qu'on risque d'éveiller ainsi" (132, p. 227). Nous conformant à cette recommandation d'André LALANDE, précisons que l'usage que les cybernéticiens font de ce terme se rattache à l'étymologie du mot plutôt qu'à son histoire : διάλεκτος signifie *langage articulé*, de διαλέγειν, *parler, expliquer*. Dans son *Dictionnaire Grec-Français*, BAILLY relève un sens de διαλεκτικός proche de celui des modèles dialectiques de simulation : "ἡ διαλεκτική (τέχνη), la discussion ou le calcul des probabilités, par *opposition à une démonstration positive*, ARISTOTE, *Rhétorique*, 1,1,1 ; *Topiques*, 1,1,2, etc."

démarche du chercheur en sciences sociales en la réduisant à trois étapes constituant un processus cyclique (123, ch. I) : l'induction, qui permet de passer des faits d'observation à la formulation d'une théorie ; la déduction, qui permet d'énoncer les conséquences possibles de la théorie en appliquant certaines règles de raisonnement ; la vérification, qui consiste à confronter les conséquences de la théorie avec les faits. Dans le cadre de cette description, le modèle constitue l'illustration de la théorie élaborée, (123, p. 4), ou plutôt il est "l'ensemble des hypothèses, d'une théorie" (19), traduit sous une forme opératoire.

En conséquence, le modèle doit présenter une certaine "similarité", une certaine *analogie* avec le système original, tel qu'il a été défini, analysé, et interprété par le chercheur ; un modèle "fonctionne de façon analogue à l'original" (103, p. 33). La notion d'*analogie* est une notion trop imprécise pour que nous puissions borner ici notre analyse. LALANDE (132, Pp. 51-52) définit l'analogie au sens courant comme une "ressemblance plus ou moins lointaine, particulièrement entre choses qui ne se ressemblent pas dans leur aspect général, et qui ne peuvent être subsumées sous un même concept". Il est vrai que cette notion avait à l'origine un sens beaucoup plus précis : "identité du rapport qui unit deux à deux les termes de deux ou plusieurs couples. Spécialement, et par excellence, proportion mathématique (appelée ἀναλογία dans EUCLIDE)" (ibid., p.51). C'est par exemple le sens auquel se réfère ARISTOTELE lorsqu'il parle des hommes "libres et égaux soit proportionnellement, soit arithmétiquement : ἑλευθέρων καὶ ἴσων ἢ κατ'ἀναλογίαν ἢ κατ'ἀριθμὸν" (10, p. 224). Bien que cette acception ne soit plus usitée de nos jours (la traduction en fait foi), il nous paraît que la signification du mot *analogie* dans le domaine qui nous intéresse est plus proche du sens primitif que du sens

courant. Nous disons en effet qu'il existe une relation d'analogie entre deux systèmes A et B lorsque par exemple la relation qui lie telle sortie de A à telle entrée de A est identique à celle qui lie la sortie correspondante de B à l'entrée correspondante de B. En d'autres termes, il y a analogie entre les deux systèmes lorsque deux éléments de A sont entre eux *dans le même rapport* (ἀνά τὸν αὐτὸν λόγον) que deux éléments correspondants de B. Cette conception apparaît dans es deux définitions suivantes, empruntées à des physiciens :

- "Etant donné un certain système S (...), l'analogue de ce système sera tout autre système S' régi par des équations identiques" (70, p. 3).

- "La notion d'analogie implique une correspondance entre les mesures de grandeurs physiques de natures différentes caractérisant deux sortes de phénomènes (...). (Dans les applications du calcul analogique), les analogies reposent sur une correspondance bien définie, dès que les conventions sont explicitées, entre les formes potentielle, cinétique et dissipative de l'énergie d'un système donné et les énergies électrostatique, électromagnétique et Joule d'un réseau électrique" (170, Pp. 17-18).

Nous allons préciser la nature de cette correspondance entre l'original et son modèle en nous appuyant d'une part sur les réflexions de G. PÓLYA et de J. ULLMO, et d'autre part sur les travaux de W. Ross ASHBY. Georges PÓLYA (160, Pp. 10-13) fait très nettement le départ entre l'analogie et la ressemblance : si l'analogie est une sorte de ressemblance, c'est une ressemblance exprimée à l'aide de concepts précis. Le poète comparant la femme à une fleur, perçoit quelque ressemblance ; mais il ne songe aucunement à une analogie. L'enfant qui trouve effrayants les squelettes d'animaux que lui pré-

sente un musée ne peut parler d'analogie, si c'est là l'unique ressemblance qu'il relève entre ceux-ci. Par contre, le naturaliste découvre bien une analogie entre divers organes, lorsqu'il note que des organes aussi dissemblables que la main de l'homme, la patte du chat, la jambe du cheval, l'aile de la chauve-souris, ou la nageoire d'une baleine, sont composées de parties semblables, *similairement reliées entre elles*. Aussi, "nous dirons que deux systèmes sont analogues si nous trouvons *les mêmes relations, bien définies, entre leurs éléments respectifs*" (ibid., p. 11). Jean ULLMO (195, Pp. 75-81) précise cette définition en la complétant par une mise en garde : citant deux exemples de raisonnement analogique, tirés de La structure des théories déductives de Louis ROUGIER, il montre que l'analogie n'est un procédé de raisonnement scientifique que lorsqu'elle porte sur les *lois* qui régissent les phénomènes, et non sur les *propriétés* des objets. Le concept aristotélicien, qui a un **contenu essentiel**, portant sur la "nature" des choses, ne peut être le support d'une relation d'analogie féconde. L'exemple cité est la notion de "phénomène ondulatoire", qui se définit par la *propriété* de présenter des interférences. L'"analogie" relevée entre les "phénomènes ondulatoires" acoustiques et lumineux a conduit FRESNEL à étendre abusivement toutes les propriétés des vibrations acoustiques aux ondes lumineuses. Par contre, l'identité des *formules* fondamentales qui rendent compte des phénomènes hydrodynamiques et des phénomènes électriques, a permis de transposer aux circuits électriques ce que l'on savait sur les canalisations d'eau. "L'exemple est valable parce qu'ici c'est *la loi qui est la même*, s'exprimant par la même formule" (ibid., p. 76).

Dans le langage de la cybernétique, W. Ross ASHBY exprime un point de vue voisin, à partir de la notion de *transformation univoque*. Un mécanisme déterminé peut être défini comme "un système dont

le comportement correspond à une application de M dans M'' (15, p. 458), M étant l'ensemble des états qui définissent le système. Deux systèmes sont *isomorphes* (13, Pp. 114-123) si une transformation bi-univoque des états d'entrée et de sortie de l'un des systèmes en ceux de l'autre permet de passer de l'application du premier à l'application de l'autre ; en d'autres termes, deux systèmes sont isomorphes si l'un peut être rendu identique à l'autre par un simple changement de variable. Dans un couple de systèmes isomorphes, l'un des systèmes peut constituer un modèle de l'autre. Ainsi, une carte est un modèle du terrain qu'elle représente. Dans la réalité, il n'est pas toujours possible de disposer d'un système isomorphe au système étudié ; et la construction d'un modèle peut n'aboutir qu'à la réalisation d'un système *homomorphe* (13, Pp. 124-133), c'est-à-dire d'un système plus simple que l'original. La simplification du système étudié est obtenue en regroupant des états que l'on distinguait auparavant, c'est-à-dire en appliquant une relation d'équivalence à l'ensemble M des états du système, de manière à obtenir un ensemble quotient M' plus petit que M . Mais cette simplification ne peut être opérée sans précaution. Supposons un système S défini par la transformation :

$$S \begin{array}{c} \downarrow \\ a \ b \ c \ d \ e \\ e \ b \ a \ b \ e \end{array}$$

On peut simplifier le système en considérant comme équivalents les états a et d d'une part, et les états b et e d'autre part. On obtient alors le système S' :

$$S' \begin{array}{c} \downarrow \\ f \ g \ h \\ h \ f \ h \end{array} \quad \text{où : } \begin{array}{l} f = a \text{ ou } e \\ g = c \\ h = b \text{ ou } e \end{array}$$

Cette transformation est close et univoque, et le système S' est déterminé. Supposons maintenant que l'on définisse à partir de S un système S'' en considérant comme équivalents les états a , b , et c d'une part, et les états d et e d'autre part :

$$S'' \downarrow \begin{matrix} k & l \\ k \text{ ou } l & k \text{ ou } l \end{matrix} \quad \text{où : } \begin{matrix} k = a \text{ ou } b \text{ ou } c \\ l = d \text{ ou } e \end{matrix}$$

La transformation n'est plus univoque ; S'' n'est pas déterminé, et ne peut être considéré comme un modèle de S . En conséquence, on posera qu'"un modèle doit être l'ensemble quotient d'une relation d'équivalence compatible avec l'application originale", (15, p. 460), où "compatible" signifie "conservant la propriété originale d'être un système déterminé".

Cette définition nous permet d'aborder le problème des rapports entre les modèles dialectiques et les modèles physiques. On peut, avec R. BOUDON, considérer les modèles simulés comme une extension des modèles dialectiques, et en particulier des modèles mathématiques : "étant donné un ensemble T de propositions verbales, associer à cet ensemble - qu'on peut, si on le désire appeler *théorie* - un modèle M , c'est, grosso modo, donner aux propositions de T , si elles ne l'ont pas déjà, une forme telle qu'il soit possible d'en déduire, par un calcul, de nouvelles propositions" (32, p. 3 ; 33, p. 381) ; le terme "calcul" étant pris au sens de "calcul logique ou mathématique" (c.f. 143, p. 16). Dans cette perspective, il est possible de classer les modèles simulés, ou plus exactement les systèmes physiques qui en sont la réalisation, en fonction de leur degré d'analogie avec le modèle mathématique ; R. BOUDON (32, Pp. 4-5 ; 33, Pp. 383-384) distingue ainsi les modèles simulés comme substitués des modèles mathématiques (en particulier lorsque ceux-ci ne sont pas actuellement solubles), et les modèles simulés pouvant "ne correspondre à aucune formalisation mathématique sauf très indirecte et simplifiée ou 'réductrice'". Une lecture hâtive du texte de R. BOUDON pourrait faire croire à l'existence de simulation sans modèle mathématique sous-jacent ; une telle croyance serait évidemment incompatible avec les analyses que nous avons conduites. En effet, pour un système réel donné

S, nous pouvons définir un certain nombre de systèmes quelconques M_1 , tous homomorphes de S, et isomorphes entre eux ; dans ces conditions, n'importe lequel des systèmes M_1 peut être pris comme modèle du système S. On pourra par exemple choisir indifféremment le modèle mathématique, électrique, ou mécanique du phénomène étudié, et seules des considérations pratiques guideront notre choix (13, p. 117). Nous n'insisterons pas sur les avantages du symbolisme mathématique (19), ni sur le fait que tout modèle autre que mathématique est réductible à un modèle mathématique ; la réciproque est d'ailleurs vraie (à des conditions de précision et de maniabilité près) pour tout système physique dans lequel nous disposons de sous-systèmes analogues aux opérateurs utilisés dans le modèle mathématique. Nous rappellerons simplement que tout modèle, s'il n'est pas mathématique, est "mathématisable" (ibid., p. 383).

1.2.7. Qu'entend-on par expérimentation ?

"Une expérience est le fait de provoquer, en partant de certaines conditions bien déterminées, une observation telle que le résultat de cette observation, qui ne peut être assigné d'avance, soit propre à faire connaître la nature ou la loi du phénomène étudié" (132, Pp. 322-323). Cette définition met l'accent sur deux traits essentiels de la notion d'expérimentation : son caractère *intentionnel* (l'expérimentateur désire vérifier une hypothèse), et le fait qu'il s'agit d'une *manipulation* (l'expérience est provoquée).

Cette notion est trop connue pour que nous nous y arrêtions. Nous ferons cependant à propos de l'expérimentation deux remarques concernant des acceptions particulières que ce mot a pris dans un passé récent. La première concerne l'*expérimentation artificielle*.

GOBLOT (94, Pp. 298-299) appelle ainsi ce qui, chez Claude BERNARD, porte le nom d'*expérience passive*, ou *non-intentionnelle* (24, Pp. 36-38) : il s'agit d'expériences que le chercheur ne pouvait pas réaliser lui-même (pour des raisons pratiques ou éthiques), et qui ont été réalisées par un accident fortuit, rendant ainsi l'observation possible. Il est intéressant de souligner l'usage de cette expression pour l'opposer à l'acception que lui donne R. FAURE dans un ouvrage à paraître portant ce titre : spécialiste de l'emploi des méthodes de simulation en recherche opérationnelle, R. FAURE désigne ainsi la manipulation du modèle physique d'un système sur lequel l'expérimentation directe n'est pas possible (BOUDON utilise le terme de "quasi expérimentation").

La seconde remarque porte sur la notion d'*expérimentation mentale*, définie par E. MACH, qui se trouve en quelque sorte préfigurer la simulation sur modèle dialectique : "en dehors de l'expérimentation physique, il en est encore une autre, l'expérimentation mentale, très abondamment employée au degré supérieur du développement intellectuel. Le faiseur de projets, l'homme qui bâtit des châteaux en Espagne, le romancier, l'inventeur d'utopies sociales ou techniques, font de l'expérimentation mentale. Mais le négociant à l'esprit solide, l'inventeur ou le chercheur sérieux procèdent de même. Tous se représentant des circonstances données, et rattachent à ces circonstances l'attente, la prévision de certaines conséquences ; ils expérimentent en pensée (...). Nos représentations, nous les avons sous la main d'une manière bien plus facile et plus commode que les faits physiques ; nous expérimentons sur nos pensées, si l'on peut ainsi parler, à moins de frais" (cité par 132, p. 324).

D'autre part, il convient de relever que les auteurs du 19e siècle qui ont contribué à codifier la méthode expérimentale, qui

ont étendu aux sciences naturelles le champ d'application de celle-ci, et qui lui ont donné la forme que nous lui connaissons aujourd'hui, donnent de l'expérimentation une description qui semble devoir en écarter la simulation. Ni John Stuart MILL (Systeme de logique rationnelle et inductive), ni Claude BERNARD (Introduction à l'étude de la médecine expérimentale), ni Thomas HUXLEY (Les sciences naturelles et les problèmes qu'elles font surgir) ne paraissent envisager la possibilité d'expérimenter sur un système pour en mieux connaître un autre. Pour chacun de ces auteurs, l'expérimentation est une observation provoquée artificiellement, afin d'en dégager les lois du phénomène naturel que l'on manipule ainsi. Le modèle invoqué par MILL est celui des chimistes, qui s'efforcent tout d'abord d'isoler une substance naturelle, de l'obtenir à l'état pur, pour ensuite y adjoindre une à une d'autres substances, afin d'observer quel sera le résultat de ces combinaisons. On cherche donc à faire varier, en les combinant, les circonstances qui peuvent influencer sur le phénomène, afin de mieux connaître le comportement de celui-ci. Et c'est bien dans cet esprit, même si leur démarche réelle semble plus complexe, que les chercheurs se réclamant de la méthode expérimentale ont utilisé l'expérimentation : Claude BERNARD ou Louis PASTEUR ont toujours cherché à agir directement sur les conditions du phénomène naturel étudié. Par contre, il est intéressant d'analyser les exemples choisis par MILL pour illustrer la fécondité de la méthode expérimentale : l'expérimentation nous apprend que, dans l'atmosphère, c'est l'oxygène et non l'azote qui est propre à entretenir la vie ; ou que les éclairs et le tonnerre sont des phénomènes électriques. Or, ce qui est remarquable dans ces exemples, c'est qu'ils constituent des ébauches de simulations plutôt que des expérimentations au sens strict. C'est dans des conditions de laboratoire, ne présentant que peu de ressemblance avec les conditions naturelles, à l'aide d'une "machine électrique", que l'on reconstitue pour les expliquer, les phénomènes de l'éclair ou du tonnerre. Cette

différence importante n'apparaît cependant pas à MILL, qui considère qu'en procédant ainsi, on n'a fait qu'isoler le phénomène étudié, au lieu d'en élaborer un équivalent. Les phénomènes créés dans des conditions de laboratoire semblent donc constituer, à ses yeux, un morceau de l'univers naturel, et par conséquent devoir *a priori* obéir aux mêmes lois. L'expérimentation, même *in vitro*, porte toujours sur un phénomène naturel, et non sur un simulacre de celui-ci.

Cette opposition, entre les thèmes associés à la notion d'expérimentation et le concept de simulation, ne doit pas arrêter notre réflexion. Au sens strict, l'expérimentation s'oppose à l'observation, tout en étant une extension de celle-ci : le "fait expérimental" est un fait produit artificiellement, dans des conditions que l'on peut ne jamais rencontrer dans la nature. De même, la simulation s'oppose à l'expérimentation, en ce que le phénomène simulé est sinon tout entier artificiel, du moins toujours distinct et différent du phénomène que l'on se propose d'étudier ; la notion de simulation n'en est pas moins une extension de celle d'expérimentation. Le point important qui distingue la simulation de l'expérimentation se situe au niveau de l'*intention du chercheur* : dans le cas d'une simulation sur un modèle physique, l'expérimentation sur ce modèle a bien pour effet d'apporter des informations nouvelles sur le comportement du modèle (sinon, il serait inutile d'expérimenter) ; mais l'intérêt majeur de cette expérimentation pour le chercheur réside ailleurs, dans la transposition possible de ces informations à un autre système, sur lequel le chercheur a renoncé à expérimenter. Nous verrons sur des exemples concrets, dans les chapitres suivants, comment peuvent se justifier la possibilité et l'intérêt de l'extrapolation à la réalité des inférences faites sur le modèle. Ajoutons cependant, en anticipant sur nos conclusions (§ 7.2.c), que la simulation doit être nettement distinguée de l'expérimentation au sens courant de ce terme : si l'on

considère les fonctions qu'elle remplit dans la démarche du chercheur plutôt que les techniques qu'elle utilise, alors la *simulation n'est pas une expérimentation*.

1.2.8. Champ d'application de notre définition.

Nous sommes maintenant en mesure de préciser les cas de "simulation" auxquels notre définition s'applique, et ceux que nous sommes contraint d'écarter. Comme nous avons opté pour une définition compréhensive, dans un domaine où les techniques évoluent rapidement, il est probable que l'extension qu'il convient de donner au concept de simulation, tel que nous l'avons défini, évoluera dans le temps. Il convient donc de considérer la description suivante comme essentiellement provisoire.

Actuellement, nous étendons l'usage du terme de simulation aux seuls cas présentant à la fois chacune des caractéristiques suivantes :

a) le *phénomène réel* étudié peut être quelconque, mais il doit être circonscrit sans imprécision ni ambiguïté et constituer, pour l'observateur, un *système rigoureusement défini* ; cette condition élimine les problèmes pseudo-scientifiques ou non formalisés.

b) le *modèle* peut être un système physique ou dialectique, mais il doit également constituer un *système bien défini*. S'il s'agit d'un système dialectique, les concepts de base, les axiomes, et les règles de déduction doivent être complètement explicités ; s'il s'agit d'un système physique, les variables considérées et leurs états doivent avoir été répertoriés d'une manière exhaustive, sans ambiguïté et avec le degré de

précision requis pour l'expérimentation.

c) le modèle doit être *quasi isolé*, c'est-à-dire que les relations qu'il entretient avec son environnement (auquel appartient l'expérimentateur) doivent être clairement identifiées et facilement contrôlables.

d) le modèle doit être *déterminé*, c'est-à-dire que le dispositif physique ou le système dialectique doivent être parfaitement connus de l'expérimentateur. S'il s'agit d'un modèle dialectique, les règles de déduction doivent satisfaire à la condition de cohérence. S'il s'agit d'un modèle physique, les relations entre les états d'entrée du système et les états de sortie doivent être univoques, et bien connues de l'expérimentateur. Cette condition élimine, en l'état actuel de nos connaissances, une grande partie des systèmes physiques naturels : en particulier, les "simulations" effectuées avec des sujets humains jouant un rôle (*all-man simulation*) sont à écarter, trop de facteurs incontrôlés intervenant entre les entrées et les sorties du système. Cette caractéristique par contre, n'élimine pas les modèles stochastiques (méthode de Monte-Carlo), car l'aléatoire intervient comme un paramètre d'entrée du système, et non en tant que facteur interne : le système est donc bien déterminé.

e) Le modèle doit être *homomorphe* au système étudié.

f) La procédure suivie doit être l'*expérimentation sur le modèle* ; c'est-à-dire que l'expérimentateur doit faire varier systématiquement certains paramètres d'entrée du modèle afin d'observer l'effet de ces manipulations sur les sorties du modèle. Cette condition écarte les opérations purement déductives et le calcul mathématique ; elle n'élimine pas par contre l'utilisation de la méthode de

Monte-Carlo pour la résolution de problèmes mathématiques actuellement non traitables par le calcul.

g) L'expérimentation sur le modèle doit avoir pour but de tirer des observations faites sur celui-ci *des propositions applicables au système original*.

Dans les chapitres suivants, nous désignerons par le vocable de "simulation" tous les exemples répondant aux conditions que nous venons d'énoncer. Nous réserverons aux exemples voisins, satisfaisant à certaines de ces conditions, mais non à toutes, le terme de "simulacre". Nous pouvons maintenant passer à l'examen des différents types de simulation.

2 - LES DIFFERENTS TYPES DE SIMULATION

Moins d'une vingtaine d'années après l'apparition des techniques modernes de simulation, les exemples d'application en sont innombrables. Ces techniques sont d'un emploi quotidien en recherche opérationnelle, et au moins fréquent dans les sciences appliquées déjà fortement mathématisées ; d'autre part, si les exemples d'utilisation en sont encore rares dans les sciences naturelles ou sociales, ils revêtent souvent un caractère original. Aussi, les cas concrets d'utilisation de la simulation présentent-ils une grande variété. Pour dresser un panorama même succinct des cas d'utilisation de la simulation, il est par conséquent nécessaire de "mettre de l'ordre" dans cette variété, et, à défaut de construire une typologie des exemples concrets que nous avons recensés, d'en établir les fondements en dégagant les principales dimensions permettant de classer ces exemples. Auparavant, nous esquissons un survol historique de l'emploi des techniques de simulation.

2.1. QUELQUES JALONS HISTORIQUES.

L'histoire de la simulation reste à écrire. Nous serions d'ailleurs tentés de parler plutôt de sa préhistoire, car l'histoire

proprement dite de la simulation ne se développe guère qu'après la seconde guerre mondiale. Mais nous verrons qu'un nombre non négligeable de chercheurs, tant dans l'antiquité que depuis la renaissance, ont "fait de la simulation sans le savoir". Il est probable qu'une recherche historique systématique apporterait bien des révélations (la découverte de l'"ordinateur" de l'île d'Antikythera, dont nous parlerons au § 2.1.1.d, en est un exemple). Une telle recherche n'a pas, à notre connaissance, été entreprise ; et il est hors de nos objectifs, et de notre compétence, de la tenter. C'est pourquoi nous nous bornerons, dans ce sous-chapitre, à indiquer, en nous limitant à la civilisation occidentale, les principaux courants qui ont permis à la simulation de naître et de se développer. Nous avons pris le parti de citer également les savants ou chercheurs qui nous paraissent être des précurseurs ou des pionniers de la simulation ; il convient à leur propos de garder présent à l'esprit le fait qu'il n'existe pas encore d'histoire de cette discipline, et que par conséquent les pionniers que nous citons doivent d'être passés à la postérité soit à d'autres travaux, soit à certains résultats qu'ils ont obtenus, et non aux techniques originales qu'ils ont mises en oeuvre. Aussi, il est possible que nous ayons attribué le mérite d'une innovation ou d'une découverte à un auteur, célèbre à d'autres titres, qui n'aura fait que copier un inventeur encore inconnu. C'est pourquoi nous pensons n'apporter ici que quelques jalons dans un domaine qui devrait être un champ d'investigation pour les historiens de la science et des techniques.

L'apparition de la simulation en tant que méthode spécifique est loin d'être un phénomène marginal par rapport à l'histoire des sciences et des techniques. Au contraire, on peut en détecter les signes précurseurs dans quatre courants majeurs du progrès scientifique, qui sont :

- la naissance et le développement de la pensée scientifique : effort de formalisation qui, dès l'antiquité grecque, porte sur l'élaboration et la définition des concepts, et sur la découverte de règles permettant de construire des théories et de les confronter à l'expérience ;

- les progrès du langage mathématique, constituant le premier système dialectique déterminé ;

- les progrès technologiques, depuis la construction des premiers appareils de mesure jusqu'aux ordinateurs actuels, en passant par les "automates" ;

- le développement et la systématisation des techniques expérimentales, parfois utilisées dans l'antiquité et au moyen-âge, mais n'ayant conquis leur place actuelle qu'à l'époque moderne.

2.1.1. L'Antiquité et le Moyen-Age.

Dans le monde méditerranéen, ces périodes sont caractérisées par la naissance et le développement de la pensée scientifique, mais aussi par l'absence de relations entre la science et les techniques. Aussi, la science antique est-elle surtout une science de l'observation, et non, malgré quelques brillantes exceptions rarement comprises et jamais suivies, une science de l'expérimentation. Il ne faut pas s'attendre, par conséquent, à découvrir dans ces périodes de nombreux précurseurs de la simulation ; tout au plus pourrons-nous signaler quelques *modèles dialectiques* ou quelques *simulacres* de phénomènes naturels. Nous verrons néanmoins que les plus grands parmi les savants de l'antiquité ont parfois suivi une démarche proche de la simulation.

a) Avant l'essor de la science grecque.

La science et les techniques des civilisations du bassin méditerranéen, autres que la civilisation gréco-romaine, sont mal connues, faute de documents en nombre suffisant. Aussi ne mentionnerons-nous pour mémoire que quelques jalons essentiels. Parmi les premiers *modèles dialectiques* du monde physique, nous ne citerons que le système astronomique des Babyloniens, qui atteint aux 8ème et 7ème siècles avant J.-C. un degré suffisant pour permettre de prévoir les éclipses de lune (38, Pp. 206-207) ; la géométrie des Egyptiens (ibid., p. 201), l'arithmétique des Egyptiens, dont les premiers éléments connus remonteraient au 19ème siècle avant J.-C. (ibid, Pp. 201-202) et des Babyloniens, dont on pense qu'elles permettaient de résoudre des problèmes algébriques complexes incluant des équations du 2ème degré (ibid., Pp. 204-205).

Les *modèles physiques* qui sont parvenus jusqu'à nous sont les plans cadastraux et les plans d'architectes des Babyloniens et des Egyptiens (203, T 14, p. 827) ; une carte du monde établie sur une tablette d'argile par les Babyloniens au 6ème siècle avant J.-C. (ibid.) ; et enfin le boulier, décrit par Hérodote, mais vraisemblablement plus ancien (ibid. T 1, Pp. 6-7). Harold BORKO voit à juste titre dans ces modèles matériels les lointains ancêtres de l'ordinateur (29, Pp. 24, et 32-34). Selon les historiens grecs, les Egyptiens auraient réalisé des statues animées de dieux et de déesses ; on a par ailleurs retrouvé dans les nécropoles divers jouets articulés (66, p. 10). Il s'agit là des premiers automates connus dans le bassin méditerranéen.

b) La naissance de la science dans la grèce antique.

C'est au 6ème siècle avant J.-C. que va naître la pensée scientifique, à partir du fonds de connaissances techniques dont disposent les savants grecs. Ceux-ci se dégagent peu à peu des préoccupations utilitaires qui étaient celles des savants égyptiens et babyloniens, et s'efforcent de dégager des *lois* et de les étayer par des *démonstrations*. En particulier, PYTHAGORE le premier, en développant ses recherches sur l'arithmétique et la géométrie, "remonta aux principes supérieurs, et rechercha les théorèmes abstraitement, et par l'intelligence pure" (EUDEME DE RHODES, cité par 38, p. 210). Il faut sans doute voir dans les travaux de l'école pythagoricienne le début de la "science pure" chez les Grecs ; la dernière manifestation de ce courant étant la création de l'algèbre par DIOPHANTE, au 3ème siècle après J.-C. (ibid., Pp. 292-294).

Les efforts des Grecs pour comprendre l'univers et le décrire en termes scientifiques ont abouti à la formulation de *modèles dialectiques* extrêmement nombreux. Le modèle le plus ancien semble être celui qui permit à THALES de prédire l'éclipse de soleil de 585 avant J.-C. (ibid., p. 208). Les *modèles dialectiques* élaborés par les Grecs ne se limitent pas au domaine de l'astronomie. A côté des grands systèmes de l'univers élaborés par EMPEDOCLE D'AGRIGENTE, ANAXAGORE DE CLAZOMENE, ou LEUCIPPE au 5ème siècle avant J.-C., HIPPARQUE ou PTOLEMEE au 2ème siècle avant J.-C., pour ne citer que les plus importants, il faut mentionner le système de l'homme élaboré au 5ème siècle par l'école hippocratique (38, passim).

c) Les premiers recours à l'expérimentation.

La méthodologie qui inspire la plupart des recherches empiriques de l'antiquité gréco-romaine est exprimée clairement par ARISTOTE et son disciple THEOPHRASTE. Ceux-ci posent, en réaction contre PLATON, que l'observation des phénomènes apporte une certitude plus grande que les considérations rationnelles, et qu'il convient d'accorder le premier rôle, dans la démarche du chercheur, à la recherche des concomitants (38, Pp. 228-229). Mais "si les savants grecs ont pris comme point de départ des faits réellement observés, ils n'ont pas cherché à renouveler ces faits, à varier les circonstances dans lesquelles ils se produisaient" (59, p. 20). S'il leur arrivait parfois de recourir à l'expérimentation, celle-ci ne leur apparaissait que comme un cas particulier de l'observation, à l'égard duquel il convenait de se méfier : en effet, "à leur avis, l'observation des faits produits au moyen de l'expérimentation ne pouvait donner des résultats probants, parce que ces phénomènes renferment quelque chose d'artificiel, susceptible de modifier ou d'altérer la marche naturelle des choses" (38, p. 230). Aussi, aucun savant de l'antiquité ne dispose-t-il d'un laboratoire, cette notion étant associée aux idées de magie et de charlatanisme (131, p. 25). De plus, on admet généralement que, dans la société esclavagiste antique, le mépris du travail manuel entraînait une indifférence des savants à l'égard des instruments, qu'il ne leur venait pas à l'esprit d'utiliser pour leurs investigations (59, p. 20).

De nombreux chercheurs de l'antiquité cependant ont utilisé l'expérimentation : depuis les pythagoriciens dans leurs recherches sur les cordes vibrantes, jusqu'à GALIEN dans ses vivisections d'animaux, ou PTOLEMEE dans son étude de la réfraction (38, passim).

Mais rares sont ceux qui ont eu conscience d'utiliser une méthodologie distincte de celle d'ARISTOTE. Dès le début du 3ème siècle avant J.-C., STRATON DE LAMPSAQUE introduit l'expérimentation comme procédé méthodologique normal, complétant l'observation ; mais son enseignement ne sera pas compris des savants de l'antiquité (ibid., Pp. 230-231). Il en sera de même pour ARCHIMEDE, chez qui certains auteurs découvrent les premières manifestations de l'esprit expérimental (59, p. 20). Dans ses Théorèmes mécaniques adressés à Eratostène, ARCHIMEDE décrit une méthode expérimentale permettant de découvrir, par la mécanique, des propositions, qu'il démontrera ensuite par la géométrie (cité par 38, Pp. 258-261). Il semble que l'on puisse presque, à ce propos, parler de simulation, les mécanismes construits et manipulés par ARCHIMEDE étant des *modèles physiques* de systèmes mathématiques abstraits. Il en est de même pour le fameux problème de la couronne de HIERON DE SYRACUSE, pour la solution duquel ARCHIMEDE construisit deux maquettes physiques de la couronne, ayant le même poids que celle-ci, mais étant l'une d'or, l'autre d'argent, ARCHIMEDE fit ainsi varier le paramètre "métal composant la couronne", observa les variations de volume qui en résultaient, et put ainsi évaluer la composition réelle de la couronne du roi (rapporté par VITRUVE, cité par 38, Pp. 255-256). Comme STRATON, ARCHIMEDE fut admiré, mais ne fut pas imité, en raison semble-t-il du fossé existant entre la science et les techniques. Il avait probablement appris l'expérimentation physique auprès des ingénieurs alexandrins, comme CTESIBIUS.

d) Les progrès de la technique dans l'antiquité gréco-romaine.

Les historiens de l'antiquité rapportent avec étonnement la construction d'automates divers par les mécaniciens grecs ou romains.

Ces automates, qui leur paraissaient imiter la vie, valurent à leurs constructeurs le nom de "faiseurs de miracles" (θαυματουργοί). L'un des plus célèbres fut la colombe de bois d'ARCHYTAS DE TARENTE (5ème siècle avant J.-C.), qui volait en utilisant l'air comprimé. HERON D'ALEXANDRIE (1er siècle après J.-C.) décrit minutieusement dans son traité Des automates les mécanismes qu'il utilisa pour réaliser la représentation entièrement automatique d'une pièce en 5 actes, la légende de NAUPLIUS. L'animation des personnages, les changements de décor, l'ouverture et la fermeture des portes, et même certains effets spéciaux (tonnerre et foudre), étaient obtenus à l'aide de cylindres à chevilles et de roues à rochets, mus par la chute d'un poids (182, Pp. 20-30). Les théâtres d'automates conçus selon ce modèle constituent les premières *machines à programme* ; encore ces programmes sont-ils partie intégrante de la machine, et ne peuvent-ils être modifiés. Les forces mises en oeuvre dans ces machines sont ingénieuses, mais peu puissantes : chute d'un poids freinée par du sable s'échappant lentement ; écoulement d'eau agissant sur les pales d'une roue ou comprimant de l'air dans un récipient ; jets d'eau, de sable, ou de mercure ; vapeurs dégagées par de l'eau (l'*éolipyle*) ou du mercure chauffés ; etc.. Aussi ces automates ne sont-ils pas à proprement parler des *machines*, mais des *simulacres* sans autre valeur que récréative. D'autre part, il ne pouvait s'agir de ce que nous appelons aujourd'hui *simulation*, puisqu'aucune recherche expérimentale n'inspirait ces travaux.

Par contre, les historiens ne parlent pratiquement jamais, en raison semble-t-il des préjugés déjà mentionnés, des instruments de mesure ou de calcul, dont certains pouvaient être très perfectionnés. Nous connaissons par une lettre d'ERATOSTHENE l'instrument qu'il inventa pour résoudre le problème de la duplication du cube, et qui

constitue le premier instrument de calcul connu avec certitude (38, p. 250) ; mais aucun exemplaire du *mésolable* n'a été retrouvé. Aussi faut-il accorder une place particulière à la découverte, dans un vaisseau coulé au 1er siècle avant J.-C., au large de l'île grecque d'Antikythéra (*), d'un instrument qui constitue un véritable modèle simulé du système solaire (166, p. 61). Bien que découvert au début du siècle, les éléments de ce mécanisme en bronze ont été longtemps pris pour des débris de statue. Étudié récemment par un chercheur américain, cet instrument s'est révélé être, avec ses vingt engrenages taillés dans une plaque de bronze, d'un niveau technique comparable à celui des horloges du 18ème siècle après J.-C. Manipulable à la main, il indique l'évolution des positions respectives des planètes ; l'auteur de sa découverte le compare à "une grande horloge astronomique sans échappement", ou à "un ordinateur analogique moderne qui utilise des éléments mécaniques, pour économiser des calculs fastidieux" (ibid., p. 66). Évidemment, une découverte de cette importance bouleverse les conceptions les plus généralement admises sur l'histoire des techniques, en révélant l'existence de simulations il y a plus de 2000 ans.

e) Le Moyen-Age.

Sur le plan technique, le Moyen-Age présente deux traits essentiels : par l'intermédiaire des Arabes, il hérite du vaste corps de connaissances mécaniques accumulées par les Grecs et non mentionnées par la littérature (166, p. 67) ; sous l'influence de la disparition de l'esclavage, les préjugés liés au travail manuel disparaissent, et les hommes de science ne répugnent plus à construire

(*) Petite île située "en face de Cythère" (ἀντί των Κύθηρων).

eux-mêmes les appareils dont ils ont besoin (59, p. 48). Cependant, les conséquences de ces facteurs ne se feront sentir nettement qu'à partir de la Renaissance.

Dans un bref historique de la simulation, nous nous bornerons à signaler le fait que les Arabes perpétuent l'acquis de la science et des techniques antiques : les savants appliquent avec succès la méthode expérimentale à l'étude de la réfraction (AL-RÂZÎ, 9-10ème siècles), des poids spécifiques (AL-BIRUNÎ, 11ème siècle), du rapport entre la force motrice des projectiles et leurs caractéristiques physiques (IBN-SÎNA, 11ème siècle) (38, Pp. 324-327) ; les mécaniciens construisent des calendriers astronomiques rappelant le mécanisme d'Antikythéra (166, p. 67).

Dans l'occident chrétien, il convient de faire une place à part à Roger BACON, qui, au 13ème siècle, ébauche une première construction de la méthode expérimentale, en liant l'expérience et l'induction, et en insistant sur l'importance des mathématiques dans les sciences de la nature, et les possibilités de démonstration qu'elles introduisent (38, Pp. 349-351). Précisons bien qu'il ne s'agit que d'une ébauche : s'il prône le recours à la méthode expérimentale pour fonder la connaissance scientifique, Roger BACON ne propose *aucune méthode précise* d'expérimentation, ni aucun principe permettant d'inférer des lois à partir de l'expérience. Bien plus, l'*experimentum* a moins pour objet le contrôle des faits naturels, que la découverte des forces occultes inconnues au reste des hommes (pierre philosophale, élixir de vie, pharmacopée quasi-miraculeuse) (34, Pp. 696-699). Cette conception pour nous ambiguë de la méthode expérimentale se trouve bien résumée par ce passage de BACON : "le raisonnement (...) ne peut distinguer le sophisme de la démonstration, à moins d'être vérifié

dans ses conclusions par les oeuvres certificatrices de l'expérience.(...
 Personne de nos jours n'a cure de cette méthode (...) ; c'est pourquoi
 tous les secrets ou peu s'en faut, et les plus grands de la science,
 sont ignorés de la foule de ceux qui s'adonnent au savoir" (cité par
 34, p. 697). Au siècle suivant, le courant nominaliste provoque un
 retour tardif à l'observation de la nature, en réaction contre le
 "réalisme" platonicien, qui s'était détourné de l'étude des phénomènes
 pour la contemplation des idées (38, Pp. 336-337).

Sur le plan technique, le Moyen-Age est l'époque de la
 domestication des forces naturelles (invention du collier d'attelage,
 du moulin à vent, du moulin à eau), et de la découverte empirique de
 l'effet de rétro-action (*feed-back*). Il semble que certains mécanismes
 de régulation aient été réalisées dès la fin du Moyen-Age : "baille-
 blé" ou "babillard", régulateur à boules, gouvernail permettant l'orien-
 tation des ailes du moulin (182, Pp. 15-17) ; mais ce n'est qu'à la
 Renaissance que ces inventions seront décrites et étudiées par les
 savants. D'autre part, la seconde moitié du 13ème siècle marque les
 débuts de l'horlogerie mécanique, avec l'invention du foliot
 (balancier), et de la roue de chaperon dont les cames constituent le
 programme des sonneries.

Enfin, au niveau des *modèles dialectiques*, il faut mention-
 ner l'oeuvre considérable du logicien Raymond LULLE (1235-1315),
 l'*Ars generalis* ou *Ars magna*. LULLE se propose de fournir à tous,
 pour la défense de la religion catholique, une méthode de penser
 universelle, constituant la clé de toutes les sciences et permettant
 à celui qui l'emploie de découvrir la vérité sur n'importe quel sujet (*)

(*) Titre complet d'une des nombreuses versions du *Grand Art* de LULLE :
*Divi Raymundi Lulli doctoris illustrati Ars Compendiosa inveniendi
 veritatem seu Ars Magna et Major quae est clavis et clausura
 omnium artium et scientiarum et omnium operum divi authoris* (141,
 T.1, p. 433).

On connaît la critique que DESCARTES adressera à l'art de LULLE dans la seconde partie du *Discours de la Méthode* (63, T.6, p. 17 ; cf. également la lettre à BEECKMAN du 29 avril 1619, *ibid.*, T.10, Pp. 164-166), critique qui atteste au moins l'importance que tenait encore cette oeuvre dans l'enseignement philosophique de l'époque. Le *Grand Art* se présente comme un complément de la logique d'ARISTOTE (cf. 34, T.1, Pp. 701-703). Il apporte une méthode de résolution (nous dirions un *algorithme*) pour deux problèmes techniques de l'utilisation du syllogisme : découverte des prémisses nécessaires pour démontrer une proposition donnée, et découverte du moyen terme qui unira deux termes extrêmes donnés. Cette méthode se fonde sur une classification de tous les prédicats possibles d'un sujet quelconque ; cette classification, étroitement liée aux catégories mentales du 13^{ème} siècle, pouvait être considérée comme caduque dès la Renaissance. Mais, à partir de cette classification (à laquelle on pourrait substituer n'importe quel autre système), LULLE indique le moyen de découvrir *automatiquement* le raisonnement recherché, en explorant systématiquement l'ensemble des combinaisons d'attributs possibles à l'aide de toutes les questions possibles. L'auteur fournit d'ailleurs de nombreuses abaques facilitant l'application de sa méthode de recherche (*). Le niveau technologique de l'époque ne permettait évidemment pas de construire des machines à raisonner mettant en oeuvre l'Art de LULLE. On trouve cependant des appareils rudimentaires, cercles de métal de diamètres différents, portant gravées les divisions correspondant aux concepts à combiner, et pivotant autour d'un même axe ; un appareil de ce type, la *figura universalis*, compte jusqu'à 14 cercles concentriques (90, p. 17). Outre que l'Art de LULLE constitue le premier exemple d'un dispositif de raisonnement automatique, on peut le considérer comme le premier *modèle dialectique universel*, simulant le fonctionnement de l'esprit humain.

(*)

Voir par exemple : 141, T.1, Pp 431, 481, 605, 665, 731.

2.1.2. L'époque moderne.

La période qui s'étend de la Renaissance au début du 20ème siècle est celle de la fusion de la science et des techniques, fusion déjà amorcée au cours du Moyen-Age. Elle est marquée par les progrès de la méthodologie expérimentale, et sa théorisation par François BACON ; par la construction des premières machines (horloges à balancier, pompes pneumatiques) ; par l'apparition de l'automatisme ; par l'introduction enfin, sous l'influence du courant mécaniste, des premières simulations systématiques sur machine. L'interdépendance de la science et des techniques à partir de la Renaissance fait que la préhistoire de la simulation apparaît comme un courant relativement homogène ; tel n'était pas le cas pour l'Antiquité, période pour laquelle également nos informations sur les progrès techniques sont quasi-nulles. Nous pouvons donc désormais suivre plus aisément l'ordre chronologique.

a) La Renaissance.

Les progrès des techniques de la Renaissance renouent avec la tradition antique, qui avait dans l'occident chrétien été méconnue au moins jusqu'au 12ème siècle ; c'est au 15ème siècle surtout que l'on redécouvre les connaissances des ingénieurs d'Alexandrie, et que l'on commence à utiliser, à des fins pratiques cette fois-ci, des machines assez complexes mues par le vent ou la force de l'eau (59, p. 49). C'est que, sous l'influence des voyages d'exploration et des grandes découvertes, qui remettent en cause l'enseignement des "savants" et les doctrines des Anciens, les habitudes intellectuelles vont se trouver profondément modifiées. On découvre avec étonnement que de simples artisans, ne sachant souvent ni lire ni écrire, connaissent mieux le monde que les philo-

sophes ; les témoignages en sont nombreux (cités par 178, Pp. 225-226). Par exemple, Jacques CARTIER écrivant (Bref récit et succinte narration, 1545) : "Et si quelques-uns voulaient dire le contraire de ceci, en alléguant le dire des sages philosophes du temps passé, qui ont écrit et fait la division de la terre en cinq zones (...), je dis pour ma réplique (...) que les simples marins d'à présent ont connu le contraire de cette opinion des philosophes par la vraie expérience" ; ou VIVES (De causis corruptarum artium, 1555), constatant que les paysans et les artisans connaissent mieux la nature que la plupart des philosophes ; ou enfin BELLEFOREST (Histoire des neuf rois Charles de France, 1568), notant ironiquement : "Les voyages de l'époque moderne ont montré que les Astrologues et les Géographes anciens ont eu peu de savoir et moins d'expérience, et qu'ils ont eu la hardiesse avec de froides et de frivoles raisons, de montrer que le monde était inhabitable outre l'équateur". Ces constatations commencent à discréditer les *docteurs* (au moins les plus dogmatiques d'entre eux), et à susciter un intérêt nouveau pour les techniques, qui sont en avance sur la science.

Ainsi, sans que l'idée de régularisation automatique soit encore maîtrisée, on en voit apparaître les premiers jalons non plus seulement chez les praticiens, mais aussi chez les auteurs scientifiques : régulateur à boules, dessiné par FRANCESCO DI GIORGIO vers 1475 ; description, par Agostino RAMELLI en 1588 du "baille-blé", dispositif réglant la quantité de blé admise sous la meule du moulin en fonction de la vitesse de rotation de celle-ci (92, Pp. 35-36). Par ailleurs, l'imagination des historiens du 15ème siècle a été frappée par les automates que construisit REGIOMONTANUS, comme sa mouche de fer capable dit-on de voleter autour d'une pièce et de revenir dans la main de celui qui l'a lâchée, ou son aigle qui

vola devant l'empereur Maximilien lors de son entrée à Nuremberg (203, T.19, p. 397).

Sur le plan des idées, François BACON, homme de transition entre la Renaissance et le siècle de DESCARTES, peut être considéré comme un des fondateurs de la science moderne, malgré l'archaïsme de ses conceptions scientifiques. Ayant "deviné que le monde réel est en grand ce que nos laboratoires sont en petit" (138, p. 439), il accorde le premier une place importante aux expériences de laboratoire, et élabore sa théorie de l'induction et de la preuve expérimentale (ibid. p. 450 ; 34, T.2, Pp. 39-41). D'autre part, on relève chez VINCI de très nombreuses suggestions d'expériences, par exemple sur la chute des corps ou l'écoulement de l'eau (131, Pp. 33-34), et des conceptions qui annoncent le mécanisme du 17^{ème} siècle : "un siècle avant la grande école mécaniste, il devine que connaître, c'est fabriquer" (138, n. 456).

b) Le 17^{ème} siècle.

Le 17^{ème} siècle est avant tout le siècle du *mécanisme* ; mais c'est également celui où l'on crée la physique mathématique et où l'on renonce à découvrir les "causes premières". La physique mathématique est née de la découverte d'un isomorphisme entre ce que l'on savait des phénomènes naturels, et certaines parties des mathématiques. Par exemple, on connaissait les valeurs de l'angle de réfraction à partir de l'angle d'incidence, et l'on disposait d'autre part d'éléments de trigonométrie ; mais l'idée n'était jamais venue de confronter ces deux ensembles de connaissances. L'idée, ébauchée par GALILEE, est redécouverte par DESCARTES : la nature est mathématique ; "pour le savant, plus d'intentions dans la nature, plus de causes

finales : l'univers, c'est l'arsenal de Venise en plus grand" (138, Pp. 487-488). Cette conception nouvelle de l'objet de l'investigation scientifique conduit les savants du 17^{ème} siècle à raisonner en termes de *rappports* entre les phénomènes, c'est-à-dire en termes de *lois*. Si l'on excepte DESCARTES, dont la science a également pour objet l'essence des choses, les savants du 17^{ème} siècle se détournent de la chose en soi pour ne plus s'intéresser qu'aux apparences (*φαινόμενα*), et aux lois qui les unissent.

Ayant conquis l'objectivité scientifique, les chercheurs du 17^{ème} siècle s'efforcent d'expliquer les phénomènes en prenant pour modèles des *systèmes physiques déterminés*, dont ils connaissaient déjà la théorie : les machines de l'époque. De là naît le courant mécaniste, qui, réduisant la nature à un jeu de forces mécaniques, tente d'expliquer les phénomènes à l'aide de ressorts, de leviers et de pompes. Ainsi, dans le *Traité de l'Homme*, DESCARTES établit-il une correspondance rigoureuse entre l'anatomie de l'homme et les éléments des machines que l'on peut voir dans les jardins du roi à Saint-Germain-en-Laye :

nerfs	tuyaux
muscles et tendons	engins et ressorts
esprits animaux	eau
organes de sens	commandes des mécanismes

Les esprits, venant des concavités du cerveau, et entrant dans les nerfs, changent la forme des muscles en qui ces nerfs sont insérés, et les font ainsi se mouvoir. De même, la seule force de l'eau meut diverses machines, et peut même "les y faire jouer de quelques instruments, ou prononcer quelques paroles, selon la diverse disposition des tuyaux qui la conduisent" (63, T.11, p. 130). Les

modèles mécaniques auxquels DESCARTES fait appel pour expliquer telle fonction particulière sont souvent relativement complexes et font penser aux machines actuelles. Ainsi, il distingue les circuits de commande et les circuits moteurs ; les "esprits animaux" parcourent ces deux types de circuit. Un circuit moteur relie des muscles antagonistes, et les fait se mouvoir ; un circuit de commande part du cerveau, et aboutit aux valvules qui règlent le circuit moteur (cf. par exemple l'explication des mouvements de l'oeil, *ibid.*, T.11, Pp. 135-137). Certains mécanismes même présentent un système de régulation automatique : c'est le cas du mécanisme de la respiration, qui fonctionne sans intervention du cerveau (*ibid.*, Pp. 138-140).

L'idéal des mécanistes n'est pas seulement d'expliquer, mais de reproduire ; c'est-à-dire de *simuler* les phénomènes qu'ils observent. MERSENNE s'extasie devant "la composition et le mouvement d'un mouche-ron, qui tout seul contient et renferme plus de merveilles que tout ce que l'art des hommes peut faire et représenter" et souhaite que l'on puisse "acheter la vue de tous les ressorts qui sont dans ce petit animal, ou bien apprendre l'art de faire des automates, et des machines qui eussent autant de mouvements" (cité par 137, p. 74). DESCARTES pense d'autre part que si nous parvenions à construire des automates "qui eussent des organes ou la figure d'un singe (...), nous n'aurions aucun moyen pour reconnaître qu'elles ne seraient pas en tout de même nature que ces animaux" (63, T.6, p. 56). Malheureusement, les techniques du 17^{ème} siècle ne permettent pas de construire des automates aussi complexes (comme nous le verrons à l'alinéa suivant), et il faudra attendre le 18^{ème} siècle pour que ces vœux se matérialisent. Aussi, à partir de l'école mécaniste, deux courants se dessineront. L'un, qui prévaudra chez les savants

français, cherchera pour expliquer les phénomènes des *modèles dialectiques* ("mettre le monde en équations") ; l'autre, illustré par HOBBS, NEWTON, MAXWELL, s'efforcera au contraire de construire des *modèles mécaniques* (138, p. 491).

Les techniques qui pourraient permettre de construire des modèles mécaniques ne montrent pas de progrès spectaculaires par rapport au siècle précédent. En particulier, les automates qu'a pu connaître DESCARTES, par exemple, ne semblent guère moins rudimentaires que ceux de l'antiquité (63, T.11, Pp.130-131). Une exception toutefois : réalisant le voeu exprimé par MERSENNE dans ses *Questions Théologiques* (cité par 137, p. 501, note 3), un médecin du Wurtemberg aurait "composé une statue (...) pour démontrer au doigt à et l'oeil la circulation du sang" ; "grâce aux principes hydrostatiques", on pouvait y voir tout ce qui se passe dans notre corps, "à la réserve des opérations de l'âme raisonnable" rapporte le *Journal des savants* de 1677 (cité par 68, p. 118). Il faut également porter à l'actif du 17ème siècle la mise au point par Salomon De CAUS, de la méthode de "programmation" sur cylindre à cames ou à taquets, pour commander les automates (182, Pp. 55-67) ; la construction d'un métier à tisser les rubans, dû à Willem Dierickzoon Van SONNEVELT (69, Pp. 216-218) ; un projet de métier à tisser dû à De GENNES, qui aurait par ailleurs inventé un canard automate capable de marcher, de manger, et de digérer (ibid., p. 218) ; enfin les machines à additionner de SCHICKARD (1623) et de PASCAL (1642), et la machine à multiplier et à diviser de LEIBNIZ (1663) (29, Pp. 35-36).

En l'état actuel de nos connaissances sur l'histoire de la simulation, une expérience mérite une place particulière dans les

recherches des savants du 17^{ème} siècle : il s'agit du *volcan de LEMERY* (*). Nicolas LEMERY (1645-1715), illustre professeur de chimie, recourait devant ses élèves à une véritable simulation pour démontrer ses hypothèses sur les mécanismes des éruptions volcaniques. Il mélangeait intimement 275 parties de limaille de fer, 30 parties de fleur de soufre, et 50 parties d'eau tiède. Avec la pâte ainsi constituée, il façonnait une maquette en forme de volcan, qu'il recouvrait de terre. Au bout de quelques heures, cette pâte s'échauffait jusqu'à l'incandescence, provoquant des jets de vapeur d'eau, et projetant en l'air la terre qui la recouvrait. Dans le mélange intime de soufre et de limaille de fer, l'eau tiède avait amorcé une réaction lente, exothermique, donnant naissance à du sulfure de fer ; la chaleur dégagée par cette réaction avait pour conséquence la vaporisation de l'eau. L'expérience de LEMERY est une bonne illustration de l'expérimentation sur une maquette, avec la confusion, que nous retrouverons jusque chez les expérimentalistes du 19^{ème} siècle, entre la maquette et la réalité (cf. § 1.2.7).

c) Le 18^{ème} siècle.

Du point de vue de l'histoire de la simulation, le 18^{ème} siècle pourrait être appelé le *siècle des automates* : les mécaniciens vont en effet parvenir à réaliser certains des modèles envisagés par les savants du siècle précédent, en particulier des modèles d'hommes ou d'animaux. L'ouvrage de LA METTRIE, L'homme-machine, publié en 1748, reflète de manière plus détaillée la conception de DESCARTES, et a

(*) Cette expérience nous a été décrite par le Professeur DAVAL. Les principales *Histoire des Sciences* n'en font pas mention, ni les *Traité élémentaire de chimie*. Certains détails de l'expérience de LEMERY nous ont été fournis par Monsieur J. BROSSARD, Ingénieur chimiste, à qui nous exprimons ici nos remerciements.

pu constituer un modèle pour les mécaniciens. Nous verrons que, sur le plan scientifique, l'apport des automates fut décevant, en raison de la rusticité des moyens mis en oeuvre, et du manque de rigueur (parfois même d'honnêteté) des mécaniciens. Sur le plan méthodologique cependant, ce siècle présente l'intérêt d'être celui où la simulation mécanique tend à se généraliser chez les savants ; il faudra semble-t-il attendre la seconde moitié du 20ème siècle pour retrouver un phénomène analogue.

Les progrès de la mécanique, et de l'horlogerie en particulier, ont permis aux techniciens de maîtriser l'automatisme. "C'est entre les mains des horlogers que l'automatisme a pris naissance sous la forme élémentaire des sonneries d'horloges, puis des automates à cycle de plus en plus complexe" (147, p. 296), les sonneries constituant nous l'avons vu le premier des mécanismes à programme. Au 18ème siècle, les constructeurs d'automates sont nombreux, mais la plupart ne songent encore qu'à réaliser des curiosités, plutôt qu'à faire oeuvre scientifique ; au nombre de ces curiosités, citons le *Petit carrosse qui va seul* de CAMUS (182, Pp. 30-41), et les nombreux automates androïdes comme le *Flûteur*, de VAUCANSON, la *Joueuse de tympanon*, de ROENTGEN et KINTZING, le *Joueur de flageolet* de Von KNAUSS, l'*Ecrivain* de Pierre JACQUET-DROZ et LESCHOT, le *Dessinateur* et la *Musicienne* de JACQUET-DROZ fils. (66, Pp. 17-27).

Cependant, les savants et le monde cultivé voient dans ces possibilités techniques nouvelles le moyen de vérifier les théories mécanistes du siècle précédent ; d'où la passion du public cultivé pour les "anatomies mouvantes" (*anatomia animata*). Certains savants, comme Nicolas LE CAT ou François QUESNAY, construisent des modèles mécaniques de l'homme ou de certains animaux dans un but didactique

ou de recherche. La simulation devient ainsi l'auxiliaire officiel de la science expérimentale. Ces travaux sont dominés par les réalisations de Jacques VAUCANSON, ainsi qu'en témoigne cet éloge de VOLTAIRE (*De la nature de l'homme*, cité par 68, p. 56) :

"Le hardi Vaucanson, rival de Prométhée,
Semblait, de la nature imitant les ressorts,
Prendre le feu des cieux pour animer les corps"

VAUCANSON mérite en effet l'attention du monde scientifique pour son *Canard digéreur*, construit pour démontrer expérimentalement la valeur de la théorie chimique de la digestion, selon laquelle les aliments sont digérés sous la seule action de la salive, du suc pancréatique, et de la bile ; cette théorie s'opposait à la théorie mécanique, qui faisait intervenir des meules, des pressoirs, et des scies (ibid. p. 127). Selon le prospectus distribué lors de sa présentation au public (Pâques 1739), le canard reproduit "le mécanisme des viscères destinées aux fonctions du boire, du manger, et de la digestion ; le jeu de toutes les parties nécessaires à ces actions y est exactement imité : il allonge son cou pour aller prendre le grain dans la main, il l'avale, le digère, et le rend par les voies ordinaires tout digéré (...). L'aliment y est digéré comme dans les vrais animaux, par dissolution et non par trituration comme le prétendent plusieurs physiciens (...). Je ne prétends qu'imiter la mécanique de cette action en trois choses, qui sont : d'avaler le grain, de le macérer, cuire, ou dissoudre, de le faire sortir dans un changement sensible" (cité par 68, Pp. 61 et 124-125). Dans une lettre à l'abbé D.F., VAUCANSON précise qu'il ne prétend pas "donner cette digestion pour une digestion parfaite, capable de faire du sang et des parties nourricières pour l'entretien de l'animal", mais que néanmoins les spectateurs "verront les expédients qu'on a

employés pour faire prendre le grain au canard artificiel, le lui faire aspirer jusque dans son estomac, et là, dans un petit espace, construire un laboratoire chimique, pour en décomposer les principales parties intégrantes, et le faire sortir à volonté" (cité par 182, p. 53). L'objectif visé est donc bien un objectif scientifique, malgré l'insuffisance du modèle qui ne tient pas compte de l'assimilation des matières nutritives par l'organisme.

Le canard de VAUCANSON, qui semble avoir illustré la thèse de la digestion chimique d'une manière satisfaisante pour les contemporains, est aujourd'hui perdu. Mais s'il faut en croire le témoignage d'un témoin oculaire, Friedrich NICOLAI, le mécanisme du canard digéreur ne serait qu'une supercherie : il n'y aurait pas de "laboratoire chimique" réalisant la digestion, mais simplement deux récipients, l'un pour emmagasiner le grain avalé, et l'autre contenant les matières à rejeter (ibid., p. 126). Ce témoignage est corroboré par l'illusionniste Robert HOUDIN, qui a réparé le canard digéreur en 1845, et qui écrit dans ses *Confidences et révélations* : "A mon grand étonnement, je vis que l'illustre maître n'avait pas dédaigné de recourir à un artifice (...). La digestion, si pompeusement annoncée dans le *Mémoire*, n'était qu'une mystification (...). L'eau et la graine ainsi aspirées tombaient dans une boîte placée sous le ventre de l'automate (...). L'évacuation étant chose préparée à l'avance, une espèce de bouillie, composée de mie de pain colorée de vert, était poussée par un corps de pompe" (cité par 182, p. 50).

Si l'on peut mettre en doute l'honnêteté scientifique de VAUCANSON au début de sa carrière, en acceptant comme circonstances atténuantes le fait qu'il ait commencé par construire des curiosités comme son *Joueur de flûte*, on peut porter à son crédit le projet qu'il soumet, en 1741 à l'Académie des Beaux-Arts de Lyon ; dans ce

projet, il fait montre d'une parfaite connaissance des ressources de la simulation, en proposant de "construire une figure automate qui imitera dans ses mouvements les opérations animales, la circulation du sang, la respiration, la digestion, le jeu des muscles, tendons, nerfs, etc.. L'auteur prétend que l'on pourra par le moyen de cet automate faire des expériences sur les fonctions animales et en tirer des inductions pour connaître les différents états de la santé des hommes afin de remédier à ces maux. Cette ingénieuse machine qui représentera un corps humain, pourra servir enfin à faire des démonstrations dans un cours d'anatomie" (cité par 68, p. 148). Ce projet ne fut bien entendu pas réalisé, malgré des recherches préliminaires ; en particulier un essai pour reconstituer la circulation du sang (ibid., p. 152), qui préfigure les expériences de MAREY.

Parmi les automates célèbres du 18ème siècle, nous citerons également le joueur d'échecs de Wolfgang Von KEMPELEN (182, T. 19, p. 397), qui appartient au siècle suivant au mécanicien MAELZEL, et dont la supercherie fut dévoilée par Edgar POE, dans la nouvelle intitulée Le joueur d'échecs de Maelzel.

En dehors des recherches sur les automates, quelques inventions méritent d'être mentionnées : celle du régulateur à boules appliqué à la machine à vapeur, de James WATT (1767) ; la programmation par cartons perforés, inspirée des orgues de barbarie du Moyen-Age, appliquée aux métiers à tisser par FALCON et par VAUCAISON ; et le perfectionnement de la programmation sur cylindre, par ENGRAMELLE (182, passim). Notons également la construction, par STANHOPE, de la première machine logique, le *démonstrateur*, qui

donne automatiquement la conclusion impliquée par les prémisses d'un syllogisme (90, Pp. 98-110).

d) Le 19ème siècle et le début du 20ème siècle.

Nous avons regroupé en une seule période les 140 années qui précèdent la seconde guerre mondiale. Celles-ci constituent en effet une période assez bien connue et relativement homogène séparant le siècle des automates de l'ère des ordinateurs. Cette période peut être considérée comme une simple continuation des siècles précédents, au cours de laquelle la théorie de l'expérimentation s'affine, les techniques mécaniques se perfectionnent, et l'usage de la simulation se répand, sans toutefois s'imposer comme il le fera après la seconde guerre mondiale.

La méthodologie expérimentale s'enrichit des travaux de John Stuart MILL et de Claude BERNARD, travaux trop connus pour qu'il soit utile de les rappeler ici (cf. § 1.2.7). Sur le plan théorique, le langage binaire, inventé par LEIBNITZ, est formalisé par BOOLE ; l'algèbre de BOOLE sera, au siècle suivant, le fondement de l'algèbre des circuits (SHANNON), utilisé pour la conception des machines à traiter l'information.

Parmi les innovations techniques, nous signalerons tout d'abord les machines logiques, dispositifs souvent assez complexes destinés à résoudre automatiquement les syllogismes. Les premières machines logiques sont purement mécaniques : *abaque logique* de JEVONS (1869), machines de MARQUAND (1882), PASTORE (1903), MACAULAY (1910) (90, Pp. 111-141) ; elles ont été suivies, selon le

projet de MARQUAND, par la machine électrique de BURACK (1936), qui préfigure l'ordinateur (ibid., Pp. 151-152). Dans une autre direction, nous mentionnerons l'*arithmomètre* de THOMAS, et surtout les machines à calculer de BABBAGE. La première, destinée au calcul de tables mathématiques, procédait au calcul de fonctions polynomiales selon la méthode des différences. Les difficultés techniques que soulevait sa réalisation provoqua l'arrêt des travaux, et seule une partie de cette machine fut effectivement construite. Il semble d'ailleurs que, si cette machine avait pu être entièrement construite, elle n'aurait pu fonctionner parfaitement, en raison de difficultés mécaniques (136, p. 105), dues au peu de "fiabilité" des éléments qui la constituaient (22, Pp. 14-15). La seconde machine, plus perfectionnée, est restée à l'état de projet ; elle peut être considérée comme un ordinateur mécanique, commandé par cartes perforées (29, Pp. 39-41). Mentionnons également le planimètre d'HERMANN, calculant automatiquement la valeur d'une intégrale (ibid., Pp. 27-28) ; la machine à prévoir les marées, de KELVIN, capable de calculer la hauteur des marées dans un port donné pour toute l'année à partir d'un petit nombre de paramètres (ibid., Pp. 27-28 ; 203, T. 16, p.873) ; l'analyseur différentiel de BUSH, entièrement mécanique (29, p. 29). Signalons enfin le perfectionnement par JACQUARD de la programmation des métiers à tisser ; la création des premiers équipements mécanographiques à cartes perforées par HOLLERITH, puis BULL ; l'invention par FARCOT du servo-moteur, qui prépare le couplage avec réaction de deux machines dont l'une contrôlera l'autre.

Les expériences de simulation mécanique sont plus variées qu'au siècle précédent. On trouve certes divers automates, très perfectionnés, qui sont plutôt des *Simulacres* que de véritables simulations. En particulier les automates de John Nevil MASKELINE, jouant aux cartes ou dessinant, ou le *Voder*, appareil à clavier reproduisant la voix

humaine (203, T. 19, p. 397). Plus remarquables sont le joueur d'échec de L. TORRES QUEVEDO, robot électromécanique capable dans des conditions données de gagner une fin de partie (136, p. 108), ou le *Componium* de WINKEL, orgue mécanique à cylindres construit en 1821, capable, par le jeu aléatoire du choix des cylindres, de simuler l'exécution de variations sur un thème donné, et ainsi de composer (182, p. 124). Cet orgue est, sur le plan musical, l'équivalent (aux problèmes technologiques près) de ce que sera sur le plan graphique la machine *Calliope*, d'Albert DUCROCQ (1953). D'autre part, d'authentiques expériences de simulation sont entreprises dans un but scientifique : celle de MAREY, montrant le rôle de l'élasticité des artères dans la circulation du sang ; celle de LILLIE, prouvant que la transmission de l'influx nerveux est un phénomène proprement physiologique, analogue à la transmission de l'onde électrique produite par l'excitation d'un fil conducteur ayant subi une certaine préparation (55, p. 108).

L'une de ces expériences mérite une mention particulière car touchant aux sciences humaines : il s'agit du modèle hydraulique du marché économique conçu en 1892 par Irving FISHER. Celui-ci commence par constater que "l'économiste emprunte à la mécanique une grande partie des mots de son vocabulaire", comme *équilibre, stabilité, contraction, force, pression, résistance, niveau*, etc. (82, p. 31). D'où l'idée d'aller plus loin, et de réaliser de véritables modèles mécaniques des phénomènes économiques en établissant un parallèle rigoureux entre les notions et les lois de l'économie, et celles de la mécanique (ibid., Pp. 132-133). Après avoir présenté des modèles simples (citermes communicantes) des mécanismes de la variation de l'utilité marginale en fonction de la quantité de biens disponibles, FISHER décrit (ibid., Pp. 53-63) un mécanisme complexe, représentant le marché économique à m biens et n

consommateurs. Ce mécanisme peut être grossièrement décrit comme un bassin, dans lequel se trouvent placées des citernes, contenant chacune la quantité d'un bien détenu par un consommateur ; il y a donc $m \times n$ citernes. Chaque citerne est séparée en deux compartiments, l'un de dimensions fixes, correspondant à la quantité de bien, l'autre de dimensions ajustables (en fonction des variations de prix), représentant la quantité de monnaie. Les citernes sont reliées entre elles suivant certaines lois par des tuyaux et des leviers articulés. A l'extérieur du bassin, des pompes permettent de faire varier la quantité de liquide contenu dans une série de réservoirs, faisant ainsi varier les revenus individuels ou la quantité d'un bien donné disponible sur le marché.

"Le mécanisme qui vient d'être décrit est l'analogue matériel du marché économique idéal. Il représente, chacun dans son rôle exact, les éléments qui contribuent à la détermination des prix et les livre aux investigations de l'oeil. Nous sommes ainsi à même (...) d'employer le mécanisme comme instrument d'investigation, et de suivre, grâce à lui, certaines variations compliquées qui, sans son aide, seraient difficilement étudiées avec succès" (ibid., p. 64). FISHER suggère diverses expériences, comme de faire varier les revenus individuels, les prix des produits, la valeur de l'étalon monétaire, ou la demande d'un produit (ibid., Pp. 64-76). Signalons enfin que *FISHER énonce toujours l'équivalent mathématique de ses modèles.*

Avant de quitter cette période pour la période contemporaine, mentionnons les simulations entreprises, un peu avant la deuxième guerre mondiale, pour étudier l'encombrement des réseaux de lignes téléphoniques. Les appels étant assimilés à des phénomènes aléatoires, les chercheurs utilisaient, pour imiter le hasard et décider si un

appel se produisait ou non pendant une unité de temps, une roulette analogue à celles que l'on trouve dans les casinos (85, p. 7). Sans doute faut-il voir dans cette anecdote l'origine du terme de *méthode de Monte-Carlo* (cf. §2.1.3.d).

2.1.3. L'époque contemporaine.

Cette période, que nous faisons commencer en 1940 et qui n'est pas encore achevée, est marquée par trois événements qui ont donné à la simulation son essor actuel :

- l'apparition d'un nouveau style de recherche pluridisciplinaire, la recherche opérationnelle ;
- l'invention des ordinateurs ;
- la création d'un courant de pensée nouveau, le courant cybernétique.

a) La recherche opérationnelle.

La recherche opérationnelle est le courant de recherche qui a donné à la simulation son essor actuel. Il est possible de faire remonter la recherche opérationnelle sinon à l'antiquité (ARCHIMEDE cherchant à déterminer l'utilisation optimum des moyens de défense de Syracuse assiégée par les Romains), du moins aux recherches de PASCAL, FERMAT et BERNOULLI sur l'application du calcul des probabilités à l'art de prendre des décisions (78, Pp. 5-6). Néanmoins, il ne s'agit là que de travaux de précurseurs ; l'histoire de la recherche opérationnelle commence en 1940, avec la création par

la défense anti-aérienne britannique, de l'équipe pluridisciplinaire appelée le "BLACKETT Circus", qui deviendra le groupe de recherche opérationnelle de l'armée de terre britannique. La recherche opérationnelle se développe ensuite rapidement en Grande-Bretagne, puis aux Etats-Unis dès leur entrée en guerre. Parmi les réussites de cette méthode de recherche, on peut citer l'analyse de l'efficacité des bombardements (ZUCKERMAN), la détermination de la taille optimum des raids de bombardiers afin de limiter les pertes, la mise au point de la tactique de bombardement la plus efficace (YOUDEFIN, SCOTT, et CLARKSON), la systématisation de la détection des navires, le perfectionnement de la lutte anti-sous-marine (WILLIAMS), la définition de la taille optimum que devaient avoir les convois de bateaux et l'élaboration de la tactique de défense anti-aérienne des navires (144, Pp. 8-14).

Depuis la fin de la guerre, la recherche opérationnelle s'est considérablement développée : des équipes se sont constituées dans la plupart des pays, qui ont étendu les applications de la recherche opérationnelle à de nombreux domaines de la vie civile (problèmes d'organisation, de gestion de stock ou de matériel, programmes économiques, prévision, etc.). Les "chercheurs opérationnels" préfèrent en général construire des modèles analytiques des phénomènes qu'ils étudient (78, p. 15) ; mais, en raison de la complexité de leur objet, les modèles analytiques correspondant sont souvent inutilisables pour le calcul. En effet, l'un des traits spécifiques de la recherche opérationnelle, à l'origine de la fécondité de cette méthode, est la pluridisciplinarité qui entraîne le *panachage des points de vue* (144, p. 10) ; les modèles auxquels aboutit une équipe disciplinaire doivent tenir compte d'un grand nombre de lois, empruntées à des sciences n'ayant pas le même degré de développement,

et qu'il est difficile d'unir en un modèle homogène. Lorsque cela est possible à force de schématisation, le modèle analytique est souvent trop complexe pour pouvoir être traité économiquement par le calcul. Aussi, la simulation est-elle l'une des méthodes les plus employées en recherche opérationnelle.

b) Les ordinateurs.

Les historiens font apparaître les ordinateurs à la fin de la première guerre mondiale. C'est oublier les pionniers que furent GERSCHGORINE, constructeur en 1929 de la première calculatrice soviétique (182, p. 7) ; KALIN et BURCKHARDT, qui réalisèrent en 1941 la première machine à calcul logique selon les principes de SHANNON (23, Pp. 163-190 ; 90, Pp. 155-157 ; 181, p. 38) ; ZUSE et SCHREYER, qui conçurent et commencèrent à construire, vers 1942 à Berlin, le Z3, calculateur utilisant 1.500 tubes à vide (76, Pp. 58-59). Le premier ordinateur moderne, le MARK 1, de Howard H. AIKEN, date de 1944 (23, Pp. 101-127). C'est un appareil électromécanique à bandes perforées, capable de réaliser les opérations arithmétiques. A titre d'exemple, la réalisation d'une addition nécessitait en moyenne 3/10 de seconde ; celle d'une multiplication, 4 secondes ; celle d'une division, 11 secondes (ibid., p. 123). Le premier ordinateur électronique, l'ENIAC, fut opérationnel en 1946 ; il utilisait 18.000 tubes à vide et pouvait effectuer 32.000 opérations élémentaires par seconde (ibid., Pp. 128-144). Il fut suivi de l'EDVAC, premier ordinateur binaire à programme enregistré, et de l'EDSAC, premier ordinateur à accumulateur et à compteur ordinal. Avec le WHIRWIND 1, apparaît enfin en 1953 la mémoire à tores de ferrite (29, Pp. 41-47 ; 136, Pp. 105-107). Les ordinateurs numériques passent ensuite avec

l'*UNIVAC 1*, de l'ère artisanale à l'ère industrielle et commencent à être utilisés dans les domaines les plus variés. Parallèlement, les ordinateurs analogiques sont conçus en vue d'applications particulières (conduite de machines, simulateurs).

La place que prend l'ordinateur dans les différents domaines de la recherche ou des applications est illustrée par le développement extrêmement rapide du marché de l'informatique : en dix ans, le volume du parc d'ordinateurs installés aux Etats-Unis est passé de 2.000 environ en 1959 à 68.500 environ en 1969 ; pour l'Europe des six et la Grande-Bretagne, les effectifs correspondants sont 250 (dont France : 20) en 1959, et 19.750 (dont France : 4.900) en 1969. Le taux de progression actuel est le 30 % pour les Etats-Unis, et 45 % environ pour l'Europe des six (206, Pp. 98-99). Les ordinateurs numériques représentent la grande majorité des ordinateurs installés. La part, en France, des ordinateurs analogiques ou hybrides représente moins de 4 % (ibid., p. 104) ; si l'on tient compte du fait que les ordinateurs numériques sont généralement plus gros, donc plus coûteux que les ordinateurs analogiques ou hybrides, on peut considérer que ces derniers représentent environ un dixième des appareils installés.

L'apparition de l'ordinateur constitue pour la simulation un événement de même importance que l'invention du microscope pour la biologie. Il ne s'agit en effet que du perfectionnement d'instruments existant déjà auparavant, mais ce perfectionnement multiplie dans des proportions considérables les possibilités du chercheur. L'intérêt de l'ordinateur numérique est sa très grande souplesse d'utilisation, puisque l'utilisateur peut écrire lui-même le programme du traitement qu'il désire voir effectuer. Ainsi que le souligne ASHBY (13, p. 117), "le grand calculateur numérique est remarquable, précisément parce qu'il

peut être programmé pour devenir isomorphe de tout système". Aussi l'ordinateur numérique joue-t-il un rôle primordial dans le développement actuel des techniques de simulation, et leurs applications.

c) La cybernétique.

Le début officiel de cette science nouvelle date de la publication, en 1948, du livre de Norbert WIENER, Cybernetics. Mais il est possible de trouver de nombreux précurseurs des cybernéticiens, comme WATT, MAXWELL (cf. 136, p. 110), ou J. LAFITTE (cf. 107, p. 12). "Science-carrefour" selon l'expression de G. Th. GUILBAUD, la cybernétique a donné naissance à de nombreux courants, parfois antinomiques, qui tous se réclament d'elle ; tous ces courants ont cependant en commun un intérêt pour la théorie de l'action finalisée et de l'adaptation (136, p. 101). Du point de vue d'un survol de l'histoire de la simulation, l'apport de la cybernétique est double : sur le plan théorique, la cybernétique a apporté une formalisation des "mécanismes" et de leur fonctionnement, utile pour l'étude des systèmes réels ; sur le plan pratique, les cybernéticiens ont réalisé des modèles physiques d'organismes adaptatifs qui sont autant de simulations.

Nous avons, au chapitre précédent (§ 1.2.3. et 1.2.4.), énoncé certains traits de la formalisation due aux cybernéticiens. Nous avons cependant négligé une notion fondamentale, celle de *feed-back* (information en retour, ou rétro-action). Il y a *feed-back* lorsque le système étudié agit sur l'environnement et l'environnement sur le système (12, p. 37), ou, ce qui revient au

même, lorsque deux systèmes sont couplés avec réaction (103, Pp.16-19). Le couplage peut être indirect, c'est-à-dire s'effectuer par l'intermédiaire d'un autre système faisant office de relai. Le feed-back peut être positif ou négatif (ibid., Pp. 19-22 ; 205, p. 200). Un feed-back *négatif* tend à ramener toujours une variable à une valeur constante ; un régulateur de WATT, un thermostat, fonctionnent selon ce principe. Un feed-back *positif* tend à écarter toujours la variable d'une position constante : le mécanisme de la "boule de neige" en est un exemple. Le feed-back est à la base même de tout mécanisme adaptatif ; aussi est-il d'un usage constant dans la plupart des exemples de simulation.

Les principaux modèles physiques construits par les cybernéticiens ont connu une grande notoriété. En particulier la tortue de W. Grey WALTER, mécanisme autonome muni d'organes sensoriels et susceptible d'être conditionné (55, p. 108) ; l'*homéostat* d'ASHBY, dispositif complexe recherchant automatiquement un état d'équilibre (12, Pp. 100-121) ; la souris de SHANNON, capable de sortir d'un labyrinthe et de se souvenir de la solution (136, p. 112) ; le mécanisme du Docteur SAUVAN, qui simule la création de l'instinct (55, Pp. 108-109). L'intérêt de tous ces mécanismes est de constituer des modèles physiques de comportements adaptatifs complexes, sur lesquels il est possible d'expérimenter.

d) La simulation proprement-dite.

Les travaux des cybernéticiens ne représentent qu'une partie modeste des applications de la simulation. Celle-ci, grâce aux progrès de l'électronique, peut rapidement prendre des formes multiples :

simulations stratégiques du type *war game* ou *business game*, simulateurs de postes de conduite d'engins (simulateurs de vol) ou de types d'environnements (simulateurs de plongée), "animaux" cybernétiques, modèles de prévision économique ou de marketing, etc.. S'agissant d'une période qui en est encore à sa phase d'expansion, il ne nous est pas possible de tenter d'en faire l'historique ; nous préférons donner un aperçu de sa diversité en décrivant quelques uns des exemples d'utilisation les mieux connus (dans la 2ème partie de ce chapitre). Signalons toutefois deux étapes importantes de cette période :

- la redécouverte, vers 1942, par Von NEUMANN et ULAM, d'une méthode de résolution expérimentale de problèmes complexes non solubles analytiquement, par tirage au hasard de points dans l'espace des solutions possibles. Cette méthode a été tenue secrète pendant la seconde guerre mondiale, et désignée sous le nom de *Méthode de Monte-Carlo*. Elle est actuellement utilisée également pour introduire l'aléatoire dans les modèles de simulation (cf. § 5.1.2.).

- les premières utilisations de la simulation en sociologie, vers 1960, avec en particulier les études de COLEMAN et WALDORF, et de Mc PHEE et SMITH, sur les mécanismes électoraux (cf. § 3.1.3.a).

2.2. CLASSIFICATION DES EXEMPLES DE SIMULATION.

Chez une partie des auteurs qui ont traité de la simulation, on trouve une esquisse de classification des cas concrets d'application de cette méthode. Dans la presque totalité des cas, ces classifications ne comptent qu'un petit nombre de types (de 2 à 5), chaque type regroupant des exemples encore relativement hétérogènes. D'autre

part, les typologies ainsi esquissées sont assez souvent de simples énumérations descriptives, sans effort de systématisation ; en conséquence, les types ne sont pas toujours exclusifs les uns des autres, et leur ensemble ne recouvre pas toujours la totalité du domaine visé. Les raisons de ces insuffisances nous semblent tenir principalement de deux facteurs. Le premier est la limitation imposée à la réflexion méthodologique par l'évolution très rapide des techniques et l'extension difficilement prévisible de leurs champs d'application ; tout effort de systématisation risque en effet de n'aboutir qu'à des propositions provisoires, qu'à des typologies très vite caduques. Le second facteur est le caractère pragmatique des classifications proposées. Aucun chercheur ne s'est fixé comme objectif la construction d'une typologie ; simplement, dans le cadre d'une recherche particulière, certains auteurs ont jugé nécessaire de distinguer plusieurs cas afin de mieux préciser quelle était l'originalité de leur démarche. Aussi, ces classifications sont-elles des typologies "de circonstance", se limitant le plus souvent à une seule dimension.

Nous ne saurions avoir l'ambition de construire une typologie "définitive". Mais il nous semble nécessaire, pour apporter plus de clarté dans le domaine de ce que nous avons appelé *simulation*, de réaliser une sorte de quadrillage systématique de ce domaine. Pour ce faire, nous reprenons, en les systématisant, certaines distinctions déjà proposées, et nous en introduisons d'autres, de manière à disposer d'un ensemble de dimensions définies sans ambiguïté. En nous inspirant de la formalisation d'APOSTEL (7, Pp. 4-5), nous pouvons représenter la relation de "modélisation" par $R(S, B, M, O)$, où S est le sujet qui entreprend la recherche, B les buts qu'il poursuit, M le modèle, et O le système original. Les "variables" qui jouent un rôle important dans la définition de la relation sont :

- l'objectif de la recherche, c'est-à-dire le couple (S, E) ;
- les caractéristiques de l'original O ;
- les caractéristiques du modèle M ;
- la relation qui lie le modèle à l'original L (M, O).

Après avoir étudié chacune de ces "variables", nous examinerons s'il est possible de combiner les dimensions retenues pour constituer une typologie des cas de simulation. Nous verrons que la complexité du domaine étudié interdit pratiquement l'élaboration d'une typologie universelle, et que toute typologie utilisable doit privilégier un petit nombre des dimensions que nous aurons dégagées.

2.2.1. Les objectifs de la simulation.

Une revue des objectifs visés par les auteurs des simulations que nous avons étudiées nous conduit à distinguer trois grandes catégories d'objectifs : la recherche en vue d'une meilleure compréhension de la réalité ; la résolution d'un problème concret, afin de guider un choix ; la formation ou l'entraînement (*training*). Ces objectifs ne sont pas exclusifs l'un de l'autre ; nous verrons en particulier au chapitre suivant (§ 3.2.2.) que le "projet Simulmatics" visait aux deux premières catégories citées. Cependant, même lorsqu'ils coexistent dans une même expérience, ils correspondent à des moments distincts et des démarches différentes de la part des chercheurs. Lorsque nous analysons un exemple de simulation visant plusieurs objectifs, nous nous trouvons en fait en face de plusieurs simulations juxtaposées, et portant sur le même matériel. Nous décrirons donc ces catégories d'objectifs comme des catégories s'excluant mutuellement.

a) La recherche en vue d'une meilleure compréhension de la réalité.

Les simulations conçues dans ce but se proposent de contribuer soit à la construction d'une théorie nouvelle, soit à l'enrichissement ou à la validation d'un corps de connaissances déjà constitué. En nous inspirant principalement des réflexions d'APOSTEL (7, Pp. 5-11) et d'ASHBY (15, Pp. 463-464), nous distinguerons quatre cas :

- les simulations ayant pour but la *construction d'une théorie*. Ce cas se présente lorsque nous disposons d'une masse de données d'observation, mais sans pouvoir intégrer l'ensemble de ces données dans une théorie globale. La difficulté pour élaborer une telle théorie est accrue lorsque les données dont nous disposons sont incomplètes, ou lorsque nous nous trouvons devant un phénomène que nous estimons *a priori* devoir être particulièrement complexe. La démarche du chercheur consiste alors à sélectionner une loi spécifique du domaine étudié, à construire un modèle satisfaisant à cette loi, et à expérimenter sur ce modèle. En l'enrichissant progressivement par l'introduction d'autres lois, il est possible de construire une théorie permettant de rendre compte de l'ensemble des faits observés avec une précision donnée.

- les simulations ayant pour but l'*achèvement d'une théorie*. Ce cas correspond aux simulations qui se proposent d'ajouter à une théorie existante quelques détails nouveaux afin d'accroître son adéquation aux faits observés. La démarche suivie consiste à construire un modèle correspondant à la théorie que l'on désire compléter, à modifier ce modèle en y intégrant les lois ou propositions nouvelles, et à ajuster expérimentalement les paramètres du modèle en confrontant aux données empiriques les résultats fournis

par la simulation. Cette démarche est d'un intérêt assez limité : elle ne peut à proprement parler apporter beaucoup d'informations nouvelles, puisqu'elle se borne à réajuster une théorie ancienne. De plus, l'enrichissement du modèle correspondant à une théorie ne peut se poursuivre indéfiniment ; lorsque le modèle est complet, c'est-à-dire cohérent et conforme aux observations, le processus s'arrête.

- les simulations ayant pour but la *restructuration d'une théorie*. On peut distinguer plusieurs problèmes correspondant à ce cas. Par exemple, on dispose d'un modèle bien ajusté aux observations, mais ne satisfaisant pas à certaines lois de la théorie. Le désaccord partiel existant entre la théorie et le modèle conduit à modifier la théorie. La nouvelle théorie peut suggérer une modification du modèle, qui conduira à une amélioration de la théorie, et ainsi de suite ; le processus peut aboutir à un développement continu de la recherche. Un autre problème de la même catégorie se pose lorsque l'on a construit plusieurs modèles partiels complémentaires, et que l'on se propose de fusionner ces modèles, ou d'en étendre le champ d'application : on est souvent conduit à reformuler la théorie correspondante. Un troisième problème est soulevé par la coexistence de plusieurs modèles distincts, en accord à la fois avec la théorie et les données empiriques, et entre lesquels la théorie ne permet pas de choisir. Le chercheur est alors placé devant la nécessité de rechercher un critère de choix, et par conséquent de modifier la théorie. Un quatrième problème enfin est dû aux insuffissances locales du modèle, qui ne rend pas compte avec une précision suffisante de certaines observations ; si le modèle est conforme à la théorie, c'est cette dernière qu'il convient de repenser.

- les simulations ayant pour but la *déduction des conséquences pratiques d'une théorie*. Elles consistent à construire un modèle correspondant à la théorie que l'on désire étudier, et à faire fonctionner ce modèle dans des conditions déterminées, afin d'en recueillir les résultats. Cet usage de la simulation correspond aux théories trop complexes pour que leurs implications concrètes puissent être déduites par simple calcul mathématique ou logique (en l'état actuel des procédés de déduction). C'est ce que l'on appelle l'*interprétation d'un modèle formel*. Cette catégorie d'objectifs correspond en gros à deux démarches. La première est l'*exploration de toutes les conséquences possibles d'une théorie*. Il est fréquent que la théorie élaborée pour rendre compte d'un ensemble de faits dépasse le cadre étroit dans lequel ces faits ont été recueillis ; dans le cas contraire, la fécondité de la théorie se bornerait à son extension à l'univers des faits de même nature, ce qui limiterait son intérêt pour le progrès de la science. La simulation permet de découvrir les implications de la théorie extérieures aux faits qui sont à l'origine de son élaboration, et de faire ainsi la preuve de sa fécondité. La seconde démarche est parfois appelée abusivement la "*validation*" d'une théorie. Elle consiste à vérifier que les conséquences de la théorie sont bien conformes aux données d'observation, en confrontant à celles-ci les résultats d'une simulation utilisant un modèle de cette théorie. Le rôle d'une démarche de ce type est analogue à celui d'un test statistique : elle permet de rejeter une théorie fautive, mais ne permet pas d'affirmer qu'une théorie est vraie. En effet, il est assez souvent possible d'ajuster les valeurs attribuées aux paramètres d'un système correspondant à une théorie, de telle sorte que les sorties de ce système soient conformes à certaines données ; on ne peut en déduire que le système est un modèle correct de la réalité. D'autre part, on peut, pour un même ensemble de faits, concevoir plusieurs théories distinctes susceptibles d'en

rendre compte. MERSENNE notait déjà que, lors même que nos expériences réussissent, elles ne nous apprennent pas comment Dieu a fabriqué les choses ; elles nous livrent seulement l'une des manières dont il a pu s'y prendre, car "il les a pu autrement faire" (137, p. 386).

b) La résolution d'un problème concret afin de guider un choix.

Par opposition aux simulations de la catégorie précédente, celles qui visent à la résolution d'un problème concret ont un objectif limité et très pragmatique : guider une décision. L'apport de cette catégorie de simulations consiste soit à fournir des éléments pour la décision, soit à choisir entre plusieurs décisions possibles compte tenu de certains critères de choix fixés a priori, soit même à prendre automatiquement la décision et à l'exécuter. Nous distinguerons quatre cas :

- les simulations ayant pour but de *fournir des éléments de décision*. Ces simulations ne jouent pas dans la décision un rôle déterminant ; elles se bornent à fournir aux responsables (*decision-makers*) des informations dont-ils ont besoin pour orienter leur action, sans que ces informations, indispensables par ailleurs, contribuent à simplifier le problème de la décision. L'exemple le plus fréquemment rencontré en est la prévision à court ou moyen terme de l'évolution d'une situation dont les facteurs sont connus, mais trop complexe pour donner lieu à une simple extrapolation. Cette catégorie de simulation s'applique particulièrement bien aux données économiques, démographiques, ou météorologiques, en raison à la fois de la complexité des phénomènes en cause et de l'intérêt

pratique que présente leur connaissance. Nous citerons également un exemple beaucoup plus rare actuellement, mais intéressant de cette catégorie de simulations : celui de la prévision "à rebours", correspondant en fait à une interpolation. Le cas particulier auquel nous pensons est le calcul de l'état de la population, pour une date intermédiaire entre deux recensements, en tenant compte des informations recueillies au cours des recensements, et des taux de nuptialité, mortalité, etc., de la population (156, Pp. 373-384).

- les simulations ayant pour but de *contribuer partiellement à la décision*. Cette catégorie de simulations regroupe les expériences incluant un modèle de décision incomplet, c'est-à-dire réalisant une partie de la décision, mais sans en conduire le déroulement jusqu'à son terme. Un exemple en est le programme *ELECTRE* (cf.39), qui simule une procédure de choix en présence de critères multiples, cette procédure pouvant aboutir à des pré-sélections différentes selon les valeurs attribuées à trois paramètres.

- les simulations ayant pour but de *choisir la meilleure décision possible*. Dans de telles simulations, le chercheur a préalablement défini de manière opératoire un critère de choix, qui peut être par exemple une grandeur que l'on cherche à maximiser : gain de voix en faveur d'un candidat ou d'un parti, taux d'augmentation des ventes ou du recrutement, proportion de produits finis acceptables au contrôle, productivité, rapidité de réponse ou d'exécution, etc.. Le chercheur construit un modèle du système dont il veut étudier le fonctionnement, fait varier les paramètres dont les valeurs dépendent de la décision qui sera prise, et observe les variations corrélatives de la grandeur choisie comme critère de choix. S'il n'y a pas remise en question du critère de choix adopté, la décision retenue sera auto-

matiquement celle pour laquelle le critère est maximum (ou minimum, ou optimum, selon la signification du critère pour les responsables). On peut donc dire que le choix de la décision est bien effectué par la simulation, même si ce choix est purement déductif et ne suppose aucune initiative (ne parle-t-on pas, dans les décisions humaines, de choix "dictés par les circonstances" ?). Remarquons que cette catégorie de simulations peut être utilisée non seulement pour des décisions simples (lancer ou ne pas lancer un nouveau produit), mais aussi pour des chaînes complexes de décisions agissant les unes sur les autres (définition d'une stratégie, restructuration d'une organisation).

- les simulations ayant pour but de *prendre et d'exécuter une décision*. Cette catégorie de simulations est à distinguer de l'asservissement des machines-outils. Le principe de l'asservissement se fonde sur l'utilisation de feed-backs négatifs, provoquant une décision de l'organe de contrôle de la machine ; mais cette décision est déclenchée par le processus de fabrication lui-même, et non par une simulation de ce processus. Dans une simulation orientée vers l'exécution d'une décision, nous sommes en présence d'un processus réel, sur le déroulement duquel l'organe de contrôle ne conçoit aucune information. Les décisions de l'organe de contrôle, qui agissent sur le déroulement du processus, ne peuvent être prises qu'à partir d'une simulation de ce processus. Il y a une forte analogie entre les simulations de cette catégorie, et un système de commande automatisée comme le triage à billes des wagons de marchandises. Dans ce système, la commande des aiguillages et du freinage sur voie est effectuée par un dispositif qui "simule" le déplacement de chaque wagon et le remplissage des voies à l'aide de billes d'acier allant se ranger dans des tubes verticaux (66, Pp. 79-80). Nous ne pouvons cependant

parler ici de simulation, car il n'y a pas expérimentation sur le modèle. Comme exemples de cette catégorie de simulation, nous citerons certains types de pilotage automatique où il est nécessaire d'anticiper les conséquences d'une manoeuvre, afin de choisir et d'exécuter la décision la plus opportune ; cela implique évidemment que les simulations se déroulent en temps accéléré, de sorte que la durée nécessaire à l'exploration des décisions possibles ne rendent pas caduque la prise de décision finale.

c) La formation ou l'entraînement.

L'objectif de cette catégorie de simulations est de familiariser des étudiants ou des professionnels avec une théorie ou une tâche nouvelle pour eux. Les simulations à caractère pédagogique, c'est-à-dire orientées vers la compréhension d'une idée ou l'entraînement à un travail, peuvent ne concerner qu'une partie de la théorie ou de la tâche, isolée de son contexte, ou au contraire porter sur la totalité de la théorie ou de la tâche. A partir des analyses d'ASHBY (15, Pp. 461-463, de KAUFMANN, FAURE, et LE GARFF (122, Pp. 84-91), et d'ADAMS (5, Pp. 33-37), nous considérerons quatre sous-catégories :

- les simulations visant à l'*assimilation d'un aspect isolé d'une théorie*. La réalisation d'un modèle correspondant à un point particulier permet de clarifier certains concepts, de mettre l'accent sur les plus importants d'entre eux, en matérialisant et en "donnant vie" aux propositions abstraites. La modélisation rend les propositions scientifiques plus convaincantes et plus stimulantes pour l'esprit. On peut distinguer plusieurs niveaux dans cette catégorie, selon le degré de complexité du sous-système abstrait qui est l'objet de la simulation. Au niveau le plus élémentaire, le modèle est consti-

tué de quelques règles simples ; ASHBY donne comme exemple de ce niveau la tortue de W. Grey WALTER, qui simule un comportement adaptatif dans la poursuite d'un objectif. On peut également illustrer ce niveau de complexité par le modèle de Van der POL, simulant les battements du coeur (161, Pp. 141-143). A un niveau plus complexe, le modèle doit permettre une investigation dans la logique interne du mécanisme ; pour cela, on peut concevoir un mécanisme discret, dont les entrées sont contrôlées par l'expérimentateur. L'exemple proposé par ASHBY est l'étude du fonctionnement d'un réseau de neurones identiques (cf. 57). Au niveau le plus élevé, le modèle peut être un mécanisme analogique très complexe, comme l'*homéostat* d'ASHBY (cf. 12, Pp. 100-121).

- les simulations visant à l'*assimilation d'un corps de connaissances dans son ensemble*. Ces simulations, encore rares, ont pour objectif la constitution d'un corpus dynamique, jouant auprès de l'étudiant le rôle d'un manuel. Mais tandis que le manuel présente les lois scientifiques de manière isolée, en consacrant par exemple un chapitre à l'action du système végétatif sur le cortex, un autre à l'action des glandes endocrines sur le système végétatif, etc., la simulation permet de représenter toutes les interactions entre les différents systèmes. L'étudiant peut ainsi consulter le corpus en expérimentant sur le modèle qui résume l'ensemble des connaissances actuelles sur le problème dans toute leur complexité.

- les simulations visant à l'*apprentissage d'une tâche partielle*. Dans de telles simulations, le sujet qui apprend se trouve plongé dans une situation aussi analogue que possible aux conditions réelles d'exécution de la tâche. Le dispositif utilisé porte généralement le nom de *simulateur* (simulateur de vol, de plongée, etc.). Un

simulateur est "un ordinateur (souvent un ordinateur analogique) ou un autre dispositif, ou bien un programme, capable de réaliser une simulation parce que conçu de telle sorte qu'il existe une correspondance fonctionnelle entre le système physique étudié, et le dispositif utilisé pour l'étude" (117, p. 8). Un simulateur de tâche partielle est conçu de manière à permettre l'apprentissage d'une phase particulièrement délicate de la tâche globale.

- les simulations visant à l'*apprentissage d'une tâche globale*. Elles utilisent également des simulateurs, mais ceux-ci sont plus complexes, et par conséquent plus coûteux, que les simulateurs de tâches partielles. Actuellement, de telles simulations servent à faire l'économie d'un apprentissage sur le tât particulièrement coûteux ou dangereux : pilotage d'engins rapides, gestion d'entreprises en régime concurrentiel, manoeuvres militaires, etc.. En termes de jeux d'entreprises ou de *war games*, on leur réserve parfois le nom de *simulations stratégiques*, par opposition aux *simulations tactiques* qui désignent les simulations de tâches partielles.

2.2.2. Les caractéristiques de l'original.

Le champ des applications possibles des techniques de simulations est très vaste, et en grande partie inexploré. Nous nous contenterons donc d'esquisser ce qui pourrait être un système de classification des originaux ; cette esquisse devrait être complétée et précisée au fur et à mesure que l'usage des techniques de simulation s'étendra. Nous ne considérerons que cinq caractéristiques de

l'original : le domaine scientifique dont il dépend ; son degré de théorisation ; le degré de certitude des facteurs qu'il fait intervenir ; le caractère périodique ou non du phénomène qui le caractérise ; le caractère stationnaire ou non de ce phénomène. Contrairement aux divisions que nous avons établies au paragraphe précédent entre les objectifs de la simulation, et qui ne constituaient qu'une seule dimension, les caractéristiques de l'original que nous venons d'énumérer doivent être considérées comme autant de dimensions indépendantes.

a) Domaine scientifique.

Nous avons choisi le terme de *domaine scientifique* afin d'éviter d'aborder le problème difficilement soluble, en l'état actuel du développement scientifique, de la classification des sciences. C'est aujourd'hui un lieu commun d'affirmer qu'aucune des classifications proposées par les grands auteurs ne permet de rendre compte du foisonnement de la science contemporaine, et que les frontières sont bouleversées par l'apparition de disciplines hybrides, qui remettent en question les divisions classiques du champ scientifique. C'est pourquoi nous ne tenterons pas une division systématique des domaines auxquels l'original d'une simulation peut se rattacher. Nous esquisserons une énumération sommaire des principaux "mots-clés", "principaux" signifiant seulement les plus fréquemment rencontrés actuellement lors d'un recensement des exemples de simulation. Ce sont : les mathématiques, la physique, la mécanique et ses applications, la chimie, la recherche opérationnelle, l'art militaire, l'économie, la gestion de l'entreprise, la biologie, la physiologie, la psychologie, la démographie, l'histoire, la sociologie.

b) Degré de théorisation.

Le système constituant l'original de la simulation doit être défini avec rigueur ; mais cette rigueur n'implique pas nécessairement que le système soit entièrement formalisé. En particulier, les relations qui lient les variables entre elles peuvent n'être pas parfaitement connues. Dans certains cas, les lacunes de la théorie peuvent aboutir à l'impossibilité de réaliser une simulation correcte, si elles ne sont pas compensées par des données empiriques contenant des informations sur les relations internes au système. Nous considérerons quatre niveaux de théorisation :

- l'original est *complètement formalisé*, et les relations entre les variables sont parfaitement connues. A ce niveau, il est possible de construire un modèle analytique du système que l'on étudie, même si ce modèle n'est pas "calculable".

- l'original est *composé de sous-systèmes entièrement formalisés*, mais les relations entre ces sous-systèmes sont mal connues. C'est ce qui se produit par exemple lorsque l'on désire intégrer, dans une expérience de simulation, un ensemble de lois parfaitement définies, les effets de chacune étant bien connus, mais les effets de l'action simultanée de ces lois étant ignorés. C'est ce que COLEMAN appelle la "synthèse de propositions simples pour répondre à des questions concernant le système" (49, Pp. 92-96). La simulation consistera alors à réaliser des essais de couplage des sous-systèmes composant l'original.

- l'original n'est que *partiellement formalisé*. Certaines lois ont été formulées à son sujet, mais elles ne concernent pas

l'ensemble des variables prises en compte. Dans ces conditions, l'alternative qui se pose au chercheur est ou bien construire un modèle incluant des propositions *a priori* sur les relations inconnues, ou bien construire un modèle partiel correspondant aux aspects formalisés de l'original, et adjoindre au modèle, au cours de la simulation, des données empiriques contenant des informations sur les relations non encore formalisées. Pour reprendre la terminologie utilisée au § 1.2.3., nous dirons que les données empiriques constituent la description d'un nombre fini d'états du système étudié. Il est par conséquent possible de tirer de ces données des propositions concernant les relations liant les variables sur lesquelles elles portent ; seulement, la validité de ces propositions dépend évidemment du caractère "représentatif" des états sélectionnés.

- l'original n'est *pas formalisé*. Il constitue ce que les cybernéticiens appellent une *boîte noire*, c'est-à-dire un système sur la nature duquel on ne peut *a priori* faire aucune supposition (cf. 13, Pp. 105-114). Pour qu'une simulation soit possible, il faut que le chercheur dispose de données empiriques constituant un échantillon des états du système, et de plusieurs modèles *a priori* qu'il compte éprouver au cours d'une série de simulations:

c) Degré de certitude des facteurs.

Même lorsque l'original est complètement formalisé, il est possible que les relations entre les variables ne soient pas connues avec une certitude absolue. Ainsi, LICHNEROWICZ établit une distinction entre deux familles de théories physiques : les théories mécanistes (comme la théorie de NEWTON sur le mouvement des astres), qui ne font

appel qu'à des fonctions certaines, et certaines théories contemporaines (SCHRODINGER, DIRAC), qui font appel au calcul des probabilités (139, Pp. 481-483). Nous distinguerons donc trois degrés :

- *tous les facteurs* de l'original sont connus de façon certaine. Il est par conséquent possible d'en formuler un modèle mathématique au sens strict, c'est-à-dire déterministe.

- *une partie seulement des facteurs* de l'original est connue avec certitude. Toute formalisation de l'original comportera donc, à côté des fonctions déterministes, certaines lois de probabilités.

- *aucun facteur* de l'original n'est connu avec certitude. Toute formalisation de l'original sera donc entièrement stochastique.

d) Caractère périodique du phénomène.

Nous éliminerons tout d'abord les phénomènes constants, qui prennent un état unique invariant dans le temps, et qui ne peuvent par conséquent faire l'objet d'une simulation. Nous distinguerons trois cas (78, Pp. 15-17) :

- les phénomènes *périodiques non constants*, appelés aussi phénomènes *permanents*. Ce sont des phénomènes cycliques, c'est-à-dire se reproduisant d'une manière identique à l'issue d'une période de longueur donnée, appelée phase.

- les phénomènes *non périodiques*, pour lesquels on n'a pu identifier de phase.

- les phénomènes *transitoires*, qui constituent une étape non périodique entre deux phases périodiques. Tels sont les phénomènes de réglage ou de mise au point d'une machine, d'apprentissage par essais

et erreurs, et, selon certains, les conflits militaires (179, p. 16, intervention de DALKEY).

e) Caractère stationnaire du phénomène.

L'allure d'un phénomène peut dépendre ou non du moment auquel commence l'observation. On peut de ce point de vue (78, p. 16) classer les phénomènes en :

- phénomènes *stationnaires*, c'est-à-dire ne dépendant pas de l'instant initial d'observation.

- phénomènes *non-stationnaires*, c'est-à-dire dont l'allure varie selon l'instant initial d'observation.

2.2.3. Les caractéristiques du modèle.

Parmi les caractéristiques qui permettent de décrire un modèle, nous considérerons tout d'abord la distinction entre modèles *prédictifs* et modèles *descriptifs*, faite par KEMENY et SNELL (123, Pp. 5-6). Un modèle *prédictif* permet d'étudier un processus, comme l'évolution d'un marché, la diffusion d'une espèce animale dans une zone géographique, la circulation de la monnaie, etc.. Un modèle *descriptif* permet d'affecter une mesure à un état complexe, comme le statut sociométrique ; c'est en quelque sorte un procédé de calcul. Il est clair qu'un modèle *descriptif* au sens que KEMENY et SNELL donnent à ce qualificatif, ne peut absolument pas faire l'objet d'une simulation ; le calcul d'un indice sociométrique, ou le traitement des

réponses à une question comportant le classement de trois objets, ne permet pas en soi l'expérimentation. Les modèles dont nous parlerons sont donc tous des modèles *prédictifs*.

Les dimensions que nous considèrerons pour classer les modèles de simulation concernent quatre aspects distincts de ces modèles : leur structure logique, leur caractère statique ou dynamique, leur degré de mathématisation et leur nature.

a) Structure logique du modèle.

ZELDITCH et EVAN distinguent deux catégories de modèles : les modèles déductifs, ou tautologiques, et les modèles inductifs, ou non-tautologiques (202, Pp. 49-51).

Les *modèles tautologiques* sont constitués par un ensemble d'hypothèses, et une longue chaîne d'opérations déductives permettant de découvrir les conséquences lointaines de cet ensemble d'hypothèses, là où le calcul mathématique ou logique seul ne pourrait aboutir. Un exemple simple en est fournie par la recherche opérationnelle appliquée à la direction des entreprises (27 ; 161, Pp. 174-181). Lorsqu'un chef d'entreprise désire étudier les répercussions prévisibles d'un changement de politique de gestion, il doit prendre en compte un très grand nombre de facteurs, internes à l'entreprise ou relatifs aux marchés du travail, des matières premières qu'il utilise, et du produit qu'il fabrique. S'il est éventuellement possible d'élaborer un modèle analytique incluant l'ensemble de ces facteurs et leurs relations, ce modèle est rarement calculable (ou bien alors, il est schématique ou incomplet). Il est relativement plus aisé de programmer les opérations

relatives à chaque élément du modèle, et de disposer ainsi d'autant de sous-programmes, ou de *modules*, qu'il y a de systèmes dynamiques entrant dans le modèle. Pour un modèle de production industrielle, on aura par exemple (27, Pp. 52-70) un module pour les approvisionnements en matières premières, un pour la fabrication, un pour le stockage des produits finis, et un pour la distribution de ces produits. Ces modules seront intégrés dans un programme général exprimant les actions exercées par un modèle sur les autres. L'intérêt d'un modèle tautologique dans ce domaine d'application réside dans la complexité des phénomènes étudiés, et dans l'étendue du champ des situations possibles.

Les *modèles non-tautologiques* ont pour objectif de reproduire *in vitro* des processus que leurs conditions d'apparition *in vivo* ne permettent pas d'étudier d'une manière satisfaisante ; ces processus peuvent ainsi être mesurés avec précision et manipulés au gré de l'expérimentateur au cours de la simulation. De tels modèles sont utilisés en particulier pour reproduire avec une fréquence suffisante des états qui se présentent rarement dans la réalité ; pour fournir des copies d'un processus unique ; pour éliminer les effets parasites de variables d'environnement masquant la relation que l'on désire étudier ; pour isoler un sous-système de l'original qui ne peut être isolé dans le monde réel.

b) Caractère statique ou dynamique.

Ce caractère dépend en partie du rôle que joue le paramètre *temps* dans le modèle. Si le temps ne joue aucun rôle, le modèle est statique ; mais il est possible que le temps soit une variable explicite

du modèle, sans que celui-ci soit nécessairement dynamique. Nous distinguerons trois cas (134, Pp. 18-19) :

- modèles *statiques*, dans lesquels le temps ne joue aucun rôle. Les modèles de cette catégorie sont conçus pour représenter une coupe instantanée dans un processus ; ils sont donc "structurellement statiques" (ibid., p. 19). Au nombre de ces modèles, on peut citer le modèle d'équilibre préférentiel de ROY et LE BOULANGER, le modèle utilisé par LEVEQUE pour prévoir la libération et la demande de la terre à partir des attitudes des agriculteurs (ibid.), et bon nombre de modèles économiques (par exemple, les modèles de maximisation de l'*utilité* par le consommateur, dans la théorie marginaliste).

- les modèles *faussement dynamiques*. Ce sont des modèles dans lesquels le paramètre *temps* joue un rôle explicite, mais qui ne visent qu'à reproduire un état d'équilibre par une procédure itérative, en simulant le processus que l'on pense devoir aboutir à cet état d'équilibre. Par exemple, pour reconstituer l'ensemble statique des lecteurs individuels de la presse périodique, en tenant compte de la périodicité de la lecture, et des duplications, on simule ce que l'on estime être le comportement réel d'un panel d'individus fictifs pendant un laps de temps donné (modèle *SCAL*).

- les modèles *dynamiques*. Dans ceux-ci, non seulement le temps joue un rôle explicite, mais le fonctionnement du modèle dans le temps n'est pas limité *a priori* à l'aboutissement à un état d'équilibre. Tels sont en particulier les modèles d'apprentissage, et les modèles markoviens. Les modèles de ce type permettent l'analyse de la phase transitoire d'un phénomène, et l'estimation des délais nécessaires pour atteindre un état d'équilibre donné.

c) Degré de mathématisation.

Tous les modèles ne sont pas des modèles mathématiques, mais tous les modèles sont "mathématisables", ainsi que nous l'avons noté au chapitre précédent (§ 1.2.6). Cependant, les limites des possibilités de calcul imposent des bornes à la "mathématisation" des modèles ; pour des raisons opératoires, la plupart des modèles utilisés en simulation ne sont pas des modèles mathématiques, mais des combinaisons de sous-modèles mathématiques. A partir des distinctions de GREEN (95, Pp. 196-198) et de BOUDON (32, Pp. 4-5 ; 33, Pp. 383-384), nous considérerons quatre cas :

- les modèles *mathématiques*, c'est-à-dire susceptibles d'être traités en n'utilisant que les seules opérations du calcul mathématique. Cette catégorie, qui englobe les modèles déterministes ou probabilistes formulés de manière analytique, s'oppose pour GREEN à toute une famille de modèles qu'il appelle *modèles de traitement de l'information (information processing models)*. Ces modèles présentent sur les modèles mathématiques l'avantage de permettre la formalisation de processus impliquant des interactions plus complexes, et des feed-backs ou itérations. On peut les représenter commodément, dit GREEN, sous la forme d'un organigramme de programmation.

- les modèles de traitement de l'information utilisés comme *substituts de modèles mathématiques*. Cette catégorie correspond aux modèles utilisés lorsque le modèle mathématique correspondant à la réalité ne peut être traité de manière analytique (il n'existe pas encore d'algorithme permettant de le résoudre). On peut les considérer comme la complication ou la modification d'un modèle mathématique existant, ou encore comme la réalisation physique d'un modèle mathé-

matique. Exemple : la simulation de la diffusion d'innovations par HAGERSTRAND (cf. § 3.1.5.a).

- les modèles constitués par une *juxtaposition de propositions empiriques* concernant le système étudié ("modèles réalistes"). La différence principale entre cette catégorie et la précédente réside dans la plus grande richesse des hypothèses, et la structuration plus lâche du modèle. L'intérêt de tels modèles est d'éviter une trop grande simplification du phénomène étudié, en rendant possible l'incorporation de toutes les propositions connues sur un segment de réalité donné. Exemple : la simulation par R. BRETON du comportement de production dans un atelier (cf. § 3.2.1).

- les modèles constitués par la *combinaison d'un petit nombre de propositions théoriques* ("modèles simulés sans équivalent mathématique" (*)). Comme les modèles mathématiques, ces modèles obéissent à l'impératif de la parcimonie des hypothèses. Mais ils ne sont pas traitables de manière analytique, et ne peuvent être considérés comme la complication ou la modification d'un modèle mathématique. Exemple : la simulation des mécanismes de l'abandon des poursuites judiciaires, par BOUDON et DAVIDOVITCH (cf. § 3.1.4.a).

d) Nature du modèle.

Par *nature du modèle*, nous entendons le dispositif matériel que

(*)

Il faut comprendre non pas que ces modèles sont sans équivalent mathématique *possible*, ce qui serait évidemment en contradiction avec l'affirmation selon laquelle tout modèle est "mathématisable". Mais plutôt que la formalisation mathématique n'a pas été tentée, en raison des difficultés qu'elle présentait, et du peu d'avantages supplémentaires qu'elle paraissait offrir au chercheur. Aussi, R. BOUDON précise-t-il ailleurs qu'il s'agit de modèles "sans équivalent mathématique *direct*", ou "sans équivalent mathématique *apparent*" (33, p. 391).

son fonctionnement nécessite. Nous distinguerons quatre sous-catégories :

- les modèles *dialectiques*, ne nécessitant aucun dispositif physique autre que les supports des informations. Par exemple, dans une simulation utilisant un modèle dialectique stochastique, on pourra utiliser une table de nombres au hasard, et diverses feuilles de papier pour "mémoriser" les résultats intermédiaires ou les informations constituant les sorties du modèle (on en trouvera un exemple simple dans 173).

- les modèles physiques *constitués exclusivement par des "machines"*. Dans cette sous-catégorie, il est possible de distinguer un grand nombre de cas : machines hydrauliques, comme celle conçue par FISHER (décrite au § 2.1.2.d) ; automates mécaniques, électromécaniques, ou électroniques, comme ceux construits par les cybernéticiens ; ordinateur numérique, dont le caractère "universel" rend possible des simulations très diverses ; ordinateur analogique, moins souple que le précédent, mais travaillant sur des grandeurs continues ; ordinateur hybride ; etc..

- les modèles physiques *utilisant des dispositifs biologiques*. On peut concevoir une simulation d'un phénomène de diffusion utilisant comme modèle physique une culture de bactéries se développant dans un espace non isotrope. La plupart des tentatives de simulation utilisant comme modèle un dispositif biologique fait appel à des organismes d'un très haut niveau de complexité : le rat, le singe, l'homme, voire le groupe social. En l'état actuel de nos connaissances, de tels organismes ne sauraient qu'exceptionnellement être considérés comme des systèmes déterminés, ainsi que nous l'avons souligné dans le chapitre précédent (§ 1.2.8.d).

- les modèles physiques *hybrides*, constitués par un ou plusieurs systèmes biologiques couplés avec une machine.

2.2.4. La nature des relations liant le modèle à l'original.

Le problème des relations entre l'original et le modèle qui servira à la simuler est un problème fondamental du point de vue méthodologique. C'est en effet de l'exacte appréciation de ces relations que dépend la valeur des conclusions que l'on peut tirer des résultats d'une simulation. Les principaux problèmes qu'il convient d'examiner à ce propos, et qui correspondent chacun à une dimension de notre classification, sont les suivants :

- le modèle appartient-il à la même classe de systèmes que l'original ?
- le modèle est-il une copie de l'original, ou d'une théorie simplifiée sur l'original ?
- le modèle reproduit-il la totalité de l'original ?
- par quels procédés a-t-on construit le modèle ?
- le système unitaire dans le modèle est-il le même que dans l'original ?
- le temps du modèle est-il le temps réel ?
- dans quelle mesure les propriétés mises en oeuvre dans le modèle sont-elles identiques à celles de l'original ?
- le caractère déterministe ou stochastique de l'original est-il conservé dans le modèle ?

a) L'appartenance à une même classe de systèmes.

APOSTEL (7, p. 4) distingue deux cas :

- *cas homogène* : le modèle et l'original appartiennent tous deux à la même classe d'entités. Par exemple, tous deux sont des formalismes, des langages, des systèmes physiques.

- *cas hétérogène* : le modèle et l'original ne sont pas de même type. Par exemple, l'original est un système biologique, et le modèle un langage.

b) L'existence d'un modèle intermédiaire.

Deux cas sont à considérer :

- le modèle est une *réplique opératoire du système original*, tel qu'il est perçu et formalisé par le chercheur. Il n'y a pas de relai entre eux.

- le modèle est une *réplique opératoire d'une formalisation intermédiaire*, qui traduit certains aspects du système original tel qu'il est perçu et formalisé par le chercheur. Cette formalisation intermédiaire doit, pour ne pas altérer la valeur de la simulation, obéir à certaines conditions. Soient T 1 la formalisation de l'original, T 2 le modèle utilisé dans la simulation, et M le modèle intermédiaire : "un modèle est un intermédiaire entre deux formalismes s'il est un modèle des deux, et si les propriétés du modèle qui le lient à T 1 sont, dans M, reliées aux autres propriétés qui le lient à T 2" (7, p. 11).

c) Le degré de complétude du modèle.

Le modèle élaboré pour reproduire le système original apporte une formalisation qui peut correspondre à l'ensemble des processus connus ou supposés dans le système, ou seulement à certains d'entre eux. Cette dimension est *a priori* indépendante du degré de théorisation de l'original (§ 2.2.2.b). En particulier, il arrive que l'on construise un modèle complet d'un système non formalisé, ce modèle étant un moyen d'avancer dans la connaissance théorique de l'original en confrontant les résultats de la simulation avec les données d'observation sur l'original. Il peut arriver également que, pour des raisons essentiellement pratiques (facilité de réalisation sur ordinateur), on réalise un modèle incomplet d'un système complètement formalisé. Nous distinguerons :

- les modèles *complets*, qui visent à reproduire les effets de l'ensemble des processus de l'original.

- les modèles *incomplets*, qui ne sont qu'une reproduction partielle du système original. Pour simuler un système à l'aide d'un modèle incomplet, il est nécessaire d'apporter les informations marquantes sur l'original sous la forme de données empiriques. Ces données sont des constantes de la simulation, tandis que les paramètres sur lesquels portent l'expérimentation sont des variables. Un modèle complet ne nécessite aucune donnée empirique ; *moins un modèle est complet, plus la simulation requiert un volume important de données empiriques, et réciproquement.*

d) Les procédures de construction du modèle.

L'élaboration du modèle à partir de l'original fait appel à trois types de procédures, pouvant être utilisées conjointement (32, Pp. 3-5 ; 33, Pp. 381-384) :

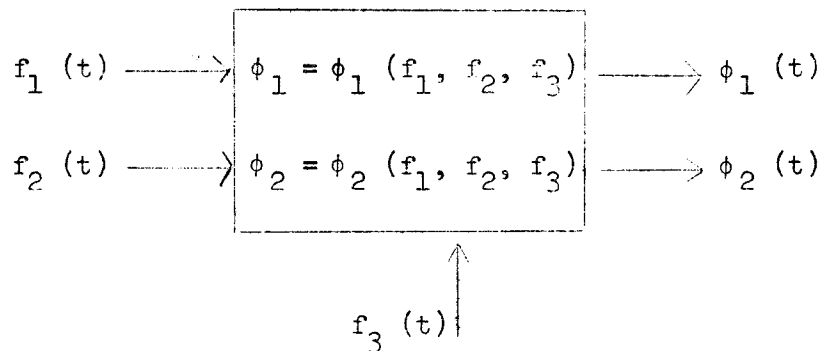
- la *spécification*, qui consiste à redéfinir certaines notions ou certaines relations de l'original, de manière opératoire.

- la *simplification*, qui consiste à ignorer certaines variables d'entrée du système original afin de faciliter le fonctionnement du modèle. Ainsi, pour simuler un phénomène de diffusion, on peut ignorer les obstacles naturels séparant les individus ; cela revient à introduire l'hypothèse simplificatrice d'un espace amorphe.

- la *réduction*, qui consiste à associer le modèle à seulement certaines variables de sortie du système original. La simulation ne s'applique par conséquent qu'à certaines conséquences du comportement de l'original, et ne peut servir à éprouver la validité de la formalisation de l'original.

e) L'utilisation des mêmes unités.

Cette dimension soulève le problème du *niveau de définition des unités élémentaires* dans l'original de la simulation. Par *unité élémentaire*, nous entendons ce que PAPERTE représente par un diagramme cybernétique atomique ; "un diagramme cybernétique atomique est un rectangle avec un ensemble de flèches orientées vers lui (nommées *entrées*) et de flèches qui en sortent (nommées *sorties*). A chaque flèche est associée une



fonction d'une variable t . Associée au diagramme, c'est une fonction qui exprime les fonctions associées aux sorties (*output*) comme fonction des fonctions (*input*) associées aux entrées (...). Un diagramme cybernétique est un ensemble de diagrammes atomiques liés de façon connexe et avec au moins une entrée libre" (159, Pp. 851-852). Dans la terminologie de PAPERT, le diagramme cybernétique est la représentation graphique d'un système, décomposé en sous-systèmes élémentaires couplés.

La notion de *niveau* est voisine de la notion d'*échelle*, ou d'*ordre de grandeur* des phénomènes observés : "selon les méthodes opératoires par lesquelles nous l'étudions, un gaz dans un récipient, d'abord fluide continu doué de pression, nous apparaîtra ensuite comme un ensemble de corpuscules comparables à des billes qui se choquent, puis ses molécules manifesteront leur structure complexe constituée par des arrangements d'atomes, puis ces atomes seront atteints dans la richesse de leur structure électronique, puis les objets constituant des noyaux atomiques (...) seront reconnus à leur tour (...)" (195, p. 74). Or, les êtres scientifiques atteints à un niveau donné d'observation "obéissent" à des lois différentes de celles des autres niveaux, même si, comme le postule BOHR (*principe de correspondance*), une loi d'un niveau donné doit pouvoir être encore vérifiée avec une approximation suffisante dans le domaine-frontière où il y a recouvrement avec le niveau immédiatement au-dessus ou au-dessous (ibid.,

Pp. 74-75). L'idée de lois scientifiques différentes selon l'échelle ou le niveau peut être illustrée par un exemple emprunté à la physique : si l'original est constitué de corps dont les dimensions se mesurent en centimètres, les lois que l'on énoncera à son propos sont celles de la mécanique classique ; si l'original appartient à un domaine dans lequel les dimensions sont de l'ordre de 10^{-8} centimètres, les lois correspondant à ce niveau de la nature sont celles de la théorie des *quanta*, "domaine décrit par une nouvelle mécanique, qui tient compte de caractères plus fins, plus complexes, de la réalité. Par extension, nous admettrons de même que si on entre dans le domaine des dimensions de 10^{-13} centimètres, il faudra édifier une nouvelle mécanique, et ainsi de suite " (197, p. 11). De même, si l'on étudie le comportement économique des consommateurs, les lois énoncées ne seront pas identiques selon que l'on partira, pour reconstituer le résultat global, de formulations relatives à l'offre et à la demande, au niveau de la population entière, ou bien relatives aux comportements individuels (134, Pp. 16-18).

Ces réflexions nous conduisent à distinguer deux familles de modèles :

- les modèles *macroscopiques*, dans lesquels l'unité élémentaire se situe au même niveau que le phénomène original. Les variables prises en compte dans le modèle sont par conséquent des variables manifestes dans l'original.

- les modèles *microscopiques*, dans lesquels l'unité élémentaire se situe à un niveau plus profond (plus fin, plus complexe, plus élémentaire) que le phénomène original. Dans ce cas, le modèle prend également en compte des variables sous-jacentes dans l'original.

f) Temps du modèle et temps réel.

Le processus original faisant l'objet d'une simulation occupe une certaine durée ; le fonctionnement du modèle occupe également un certain laps de temps. Ces durées sont-elles identiques ? Nous examinerons quatre cas :

- la simulation a lieu en *temps réel*. La mise en oeuvre du modèle dure exactement autant que le déroulement du processus original. C'est le cas par exemple pour les simulations de trajectoire d'engins, de traitement des produits dans une raffinerie de pétrole, ou pour les simulations de vol.

- la simulation a lieu *au ralenti*. C'est le cas dans certaines simulations pédagogiques mettant en évidence le déroulement d'un processus trop rapide pour être étudié sur le terrain.

- la simulation a lieu *en temps accéléré continu*. Ce cas est le plus fréquent, et certains auteurs restreignent à ce seul cas l'usage du mot simulation (cf. § 1.1.7).

- la simulation a lieu *en temps accéléré discontinu*. Elle se déroule en plusieurs phases, entre lesquelles les expérimentateurs ont à prendre des décisions concernant la phase suivante au vu des résultats précédents. On peut citer comme exemple les jeux d'entreprises.

g) Identité des propriétés du modèle.

Nous avons déjà abordé (§ 1.2.6) le problème de la correspondance entre les propriétés du modèle et celles de l'original. Nous avons en particulier indiqué qu'un modèle pouvait n'être qu'homomorphe (et non isomorphe) au système original. ASHBY fait remarquer qu'une souris en fer blanc peut être un modèle satisfaisant d'une souris vivante,

si la nature mécanique de l'une et protéinique de l'autre est sans importance (13, p. 132) ; ou qu'un physicien peut considérer une masse de gélatine moulée en forme d'hémisphères cérébraux comme un modèle correct du cerveau, s'il étudie la propagation des ondes de choc à travers celui-ci, tandis qu'un neurologue étudiant le cheminement de l'influx nerveux préférera utiliser comme modèle un réseau de fils électriques (15, p. 460). Dans chacun de ces exemples, l'analogie qui compte n'est pas la conformité des apparences, mais la "similarité génotypique" (202, p. 53).

Avec ZELDITCH et EVAN (ibid., Pp. 52-53), nous considérerons quatre degrés de ressemblance entre une propriété de l'original et la propriété correspondante du modèle :

- la propriété du modèle est *identique* à la propriété correspondante de l'original. Il s'agit du même phénomène, de la même loi, du même sous-système incorporé tel quel dans le modèle (propriétés "réelles").

- la propriété du modèle est *identique à l'échelle près* à celle de l'original. Un sous-système de l'original a été "miniaturisé" pour être incorporé au modèle (propriétés "iconiques").

- la propriété du modèle est *différente* de celle de l'original, mais *elle se comporte de la même manière* que cette dernière : il y a par conséquent toujours isomorphisme (propriétés "analogues").

- la propriété du modèle est *différente* de celle de l'original, et *la ressemblance entre leurs comportements est superficielle* (propriétés "homologues").

h) Degré de certitude des facteurs de l'original et du modèle.

Nous avons distingué (au § 2.2.2.c) trois degrés dans la certitude des facteurs de l'original. Pour la commodité de l'exposé, nous allons la réduire en deux, en considérant que tout original qui n'est pas entièrement déterministe est stochastique. Nous pouvons de même classer les modèles en deux catégories, selon qu'ils font ou non appel à des distributions de probabilité. Nous considérerons quatre cas (cf. 42, Pp. 374-375 ; 78, Pp. 104-113 ; 6, p. 2) :

- un modèle déterministe d'un original déterministe. Exemples : établissement d'horaires, simulation du vol d'un engin.

- modèle déterministe d'un original stochastique. Exemples : applications des probabilités à l'étude de phénomènes dont les lois sont mal connues.

- modèle stochastique d'un original déterministe. Exemple : calcul d'intégrales définies multiples par la méthode de Monte-Carlo.

- modèle stochastique d'un original stochastique. Exemples : problèmes de transport, de circulation, de files d'attente.

2.2.5. Remarques sur la possibilité de construire une typologie.

L'effort de classification que nous avons entrepris devrait permettre d'aboutir à une typologie assez complète des exemples de simulation. Quelques problèmes se posent encore cependant ; les principaux sont dus :

- au manque de clarté de certaines dimensions, en particulier celle du domaine scientifique auquel se rattache l'original ;

- au caractère empirique ou insuffisamment systématique de certaines distinctions, comme celle que nous avons proposée entre les différents objectifs possibles de la simulation ;

- à la précision insuffisante pour des sous-catégories correspondant à la majorité des cas actuellement réalisés, en particulier pour la nature des modèles physiques constitués exclusivement par des machines ;

- à la redondance de certaines informations, traduisant le fait que toutes les dimensions ne sont pas indépendantes : par exemple, les simulations ayant pour but de prendre et d'exécuter une décision sont nécessairement en temps accéléré.

Il est possible de pallier certains de ces défauts, en négligeant quelques unes des dimensions retenues. Entre autres, deux dimensions concernant l'original : le degré de certitude des facteurs dans l'original (§ 2.2.2.c), partiellement redondant avec le degré de certitude des facteurs de l'original et du modèle (§ 2.2.4.h) ; et le domaine scientifique dont dépend l'original (§ 2.2.2.a). D'autre part, le degré de précision dans certaines sous-catégories peut être augmenté. A titre d'exemples, il est possible de distinguer quatre types de problèmes pour les simulations ayant pour but la restructuration d'une théorie (2.2.1.a) ; trois niveaux de complexité pour les simulations visant à l'assimilation d'un aspect isolé d'une théorie (§ 2.2.1.c) ; quatre modes d'utilisation des modèles non-tautologiques (§ 2.2.3.a). Il serait de même intéressant de relever les caractéristiques exactes des machines utilisées (§ 2.2.3.d), et plus particulièrement, dans le cas de simulations sur ordinateur numérique, le type de celui-ci, sa capacité de mémoire centrale, et les périphériques qui lui sont connectés. Ces caractéristiques se trouvent souvent être d'une grande importance pratique, et peuvent obliger le chercheur à modifier son modèle afin

de le rendre exécutable sur la machine dont il dispose.

Cependant, si l'on combine les dimensions retenues en se limitant à leur précision minima, on aboutit à un nombre de types de l'ordre de 10^8 . Une telle typologie serait évidemment inutilisable. Il nous a cependant paru utile de "baliser", avec autant de précision qu'il était possible dans le cadre d'un travail limité, le domaine de la simulation tel que nous l'avions défini dans le chapitre précédent. Ce faisant, nous espérons avoir réussi à donner une idée de l'étendue et de la complexité d'un champ de recherche que nous avons pourtant étroitement circonscrit au départ. Nous avons voulu également aborder, à propos de certaines distinctions, quelques-uns des problèmes méthodologiques qui se posent au chercheur désireux de recourir aux techniques de simulation. Le "balisage" que nous avons entrepris constitue en fait une sorte de *questionnaire* devant permettre d'établir rapidement le profil d'une simulation particulière, et d'identifier les points sur lesquels elle diffère d'une simulation voisine. Si l'on désire avoir une vue d'ensemble des simulations réalisées dans un domaine particulier, les dimensions d'une typologie peuvent être sélectionnées parmi celles que nous avons décrites, en tenant compte de l'objectif poursuivi. Pour notre part, nous jugeons particulièrement importantes les dimensions concernant les objectifs de la simulation, la structure logique du modèle, et les relations entre les propriétés du modèle et celles de l'original. Si nous nous limitons par exemple aux seules simulations orientées vers la recherche, nous pouvons disposer d'une typologie encore maniable, malgré un nombre théorique de types dépassant la trentaine. Dans le chapitre suivant, nous allons étudier plus concrètement quelques exemples de simulation en sociologie.

3 - LA SIMULATION EN SOCIOLOGIE

Dans ce chapitre, nous nous proposons de donner tout d'abord un aperçu de la diversité des applications de la simulation déjà réalisées dans les sciences sociales, en décrivant succinctement une trentaine d'exemples. Nous analyserons ensuite de manière plus détaillée deux expériences de simulation, l'une en microsociologie, l'autre en macrosociologie.

3.1. PLACE DE LA SIMULATION DANS LES SCIENCES SOCIALES.

En permettant d'intégrer en un tout opérationnel des propositions empruntées à des domaines scientifiques divers, la simulation est, par nature, une démarche pluridisciplinaire. Aussi serait-il vain de vouloir s'efforcer, à partir d'une définition stricte de la sociologie, d'écarter toutes les simulations ne concernant pas directement la sociologie telle que nous l'aurons définie. Nous avons préféré considérer un échantillon des applications de la simulation pouvant intéresser les sociologues, c'est-à-dire dont le modèle avait pour sortie un fait social, ou un ensemble de faits sociaux. Cela nous a conduit à prendre en compte certaines simulations réalisées par des économistes, des démographes, des spécialistes de

l'organisation. D'autre part, il arrive que certains modèles, relatifs à des faits sociaux, aient pour unité de base l'individu (modèles microscopiques) ; dans ce cas, ce sont des processus individuels qui sont mis en oeuvre pour reproduire un phénomène se situant au niveau sociologique. C'est pourquoi nous décrivons également quelques simulations réalisées par des physiologistes ou des psychologues, et concernant des processus tels que l'apprentissage, le comportement de choix, la formation des habitudes. De tels modèles peuvent en effet être repris ultérieurement dans une simulation sociologique d'un niveau de complexité plus élevé que celui des simulations réalisées actuellement.

Dans son ensemble, cet échantillon est suffisamment diversifié pour illustrer la place que les techniques de simulation occupent aujourd'hui dans les sciences sociales. D'un point de vue quantitatif, cette place paraît extrêmement modeste, en particulier lorsqu'on la compare au rôle que joue la simulation dans les sciences physico-mathématiques. Sur le plan qualitatif, les modèles utilisés semblent souvent trop schématiques, les processus simulés trop artificiels. Parfois même, le phénomène original que l'on veut reproduire est lui-même un processus artificiel, ou une expérience de laboratoire. Il arrive enfin que les propositions incorporées au modèle s'avèrent purement tautologiques, et sans intérêt scientifique. Ces caractéristiques nous paraissent être celles qui peuvent s'appliquer à toute approche scientifique n'ayant pas dépassé le stade de la mise au point. Le chercheur en sciences sociales est peu familiarisé avec l'expérimentation ; son domaine de recherches s'y prête peu. Il connaît mal les procédés modernes de traitement de l'information et les possibilités qu'ils lui offrent ; il a trop peu de contacts avec les mathématiciens et les informaticiens et,

lorsqu'il fait appel aux services de ces spécialistes, la communication n'est pas toujours facile à établir. Aussi les expériences de simulation déjà réalisées dans les sciences sociales doivent elles être considérées comme les premières phases d'un apprentissage devant conduire ultérieurement à la maîtrise d'un nouveau mode d'investigation scientifique.

Par rapport à la classification que nous avons élaborée au chapitre précédent, les exemples que nous avons recensés se localisent tous dans une zone relativement étroite de l'espace typologique. L'objectif principal de la simulation est soit la recherche, soit la résolution d'un cas concret en vue de fournir des éléments de décision (bien que certains modèles aient également servi pour la formation) ; la technologie utilisée est toujours l'ordinateur, le plus souvent l'ordinateur numérique ; la simulation est effectuée en temps accéléré sans interruption. Cette spécificité des exemples de notre échantillon nous paraît refléter la spécificité des applications actuelles de la simulation dans les sciences sociales.

Nous examinerons successivement les modèles se situant au niveau individuel ; puis ceux visant à reproduire les phénomènes observables dans les petits groupes ; enfin ceux se situant à un niveau macrosociologique : simulations électorales, simulations du fonctionnement des organisations, simulations de l'évolution d'un système.

3.1.1. Les simulations de processus individuels.

Nous avons classé les exemples que nous avons choisis en cinq catégories : simulations concernant le fonctionnement du système nerveux, les mécanismes de la pensée, l'apprentissage, la formation des attitudes, la prise de décision.

a) Simulation du fonctionnement du système nerveux.

Le fonctionnement du système nerveux a fait principalement l'objet de deux types de projets de simulations : simulation d'un réseau de neurones, et tentatives pour reproduire le fonctionnement du cerveau.

La recherche conduite par J.T. CULBERTSON (57) appartient au premier type. L'auteur se propose de montrer que les comportements complexes des êtres vivants, décrits en termes de stimulus-réponse, peuvent être reproduits par des automates constitués d'éléments non différenciés, mais dont les connexions forment un réseau complexe ; ce sont les propriétés des réseaux, et non celles des éléments, qui expliquent le comportement (on trouvera une approche théorique de ce problème chez KLEENE, 126). A partir de trois types seulement de neurones (récepteurs, neurones centraux, effecteurs), l'auteur construit des réseaux permettant de rendre compte de la perception et de la reconnaissance des formes, et laisse pressentir la possibilité de reproduire, par le même procédé, le phénomène de conscience. Une recherche de ce type débouche évidemment sur la conception d'automates neuronaux.

W.R. ASHBY (15) étudie la possibilité de simuler le fonctionnement du cerveau dans sa totalité, en construisant un automate hautement sélectif, extrêmement sensible, extrêmement actif, et possédant des connexions internes nombreuses et complexes. Ces qualités lui sont indispensables pour pouvoir atteindre un but fixé extérieurement. Le dispositif analogique construit par ASHBY sous le nom d'*Homéostat* est une première étape dans la voie de la réalisation d'un "cerveau artificiel" ; ce dispositif élémentaire, possédant les quatre propriétés énoncées par l'auteur, est en particulier capable d'adaptation (réorganisation de l'organe à la suite d'une lésion) et d'apprentissage, mais présente certaines imperfections (difficultés d'adaptation au milieu, délai de réponse trop long), que l'on retrouve il est vrai dans les organismes vivants (12, Pp. 100-121).

b) Simulation des mécanismes de la pensée.

Cette catégorie se distingue de la précédente en ce qu'elle se situe à un niveau différent : les simulations du système nerveux concernent principalement le fonctionnement d'un organe déterminé, tandis que les simulations de la pensée se placent d'emblée au niveau de l'individu dans sa totalité. Les tentatives et les réalisations dans cette catégorie peuvent être classées, selon la complexité des fonctions simulées, en trois sous-catégories : simulation d'activités purement logiques, simulation d'une ou plusieurs fonctions de la pensée, simulation de la pensée sous toutes ses formes.

La simulation d'*activités purement logiques* est la voie dans laquelle s'engagent dans leur majorité les chercheurs de formation mathématique. Nous citerons, à titre d'exemples et sans les développer, les automates capables de jouer aux dames ou aux échecs (au moins des fins de partie), ou de démontrer des théorèmes mathématiques (*). Les modèles développés à ce propos nous paraissent présenter actuellement un plus grand intérêt pour la psychologie ou l'épistémologie que pour la sociologie, en raison du caractère très artificiel des activités simulées.

Parmi les activités de la pensée, la *résolution de problèmes* a retenu l'attention de nombreux chercheurs. L'exemple le plus connu est celui du *General Problem Solver*, conçu par NEWELL et SIMON (153). Ce programme a pour but de reproduire, à partir de trois types d'opérations élémentaires, les démarches faites par un sujet pour résoudre un problème de calcul logique ; le modèle étant fourni par les "protocoles", c'est-à-dire les commentaires faits à haute voix par les sujets pendant leur réflexion. Dans un texte ultérieur (151), NEWELL a montré la difficulté d'aller plus avant et de simuler un protocole donné.

Un autre type de problème qui a fait l'objet de nombreuses recherches est celui de la *perception*. Un travail de pionnier a été entrepris dans ce domaine par ROSENBLATT, avec son projet de *Perceptron* ; le problème fondamental auquel se heurtent actuellement les chercheurs est celui de la reconnaissance des formes, qui met en jeu des aspects particulièrement complexes de la pensée (148, Pp. 411-425). Dans le cadre d'une recherche sur les processus cognitifs, FELDMAN (80 :81) a tenté de simuler les anticipations d'un

(*) Voir par exemple FEIGENBAUM et FELDMAN, *Computers and Thought* (Mc Graw-Hill, 1963), 1ère partie, sections 2 et 3.

sujet en présence d'évènements présentés sous forme d'alternatives ; les résultats de la simulation paraissent assez décevants. Dans un autre domaine, celui de la formation des concepts, nous citerons à titre d'illustration le modèle de HUNT et HOVLAND (116 ; cf. également 115). Pour reproduire ce processus, les auteurs ont utilisé un matériel analogue à celui des expériences psychologiques sur la formation des concepts, afin de confronter les résultats de la simulation à ceux des tests effectués sur des étudiants. Le modèle inclut des procédures de reconnaissance des stimuli, de mémorisation et d'oubli, et d'élaboration de la réponse ; cette dernière, particulièrement complexe, repose essentiellement sur la comparaison des stimuli, et la réalisation d'opérations sur les ensembles.

La simulation de la pensée *sous toutes ses formes* n'est encore qu'à l'état de projets. Parmi ceux-ci, nous citerons une recherche visant à construire un automate capable d'observer ce qui se passe dans son environnement, et d'en faire une description verbale (40 ; 41). L'automate est conçu comme une vaste machine hybride, c'est-à-dire comportant des organes connectés avec un ordinateur numérique. Le schéma général de la machine est le suivant : les organes sensitifs explorant l'environnement sont connectés à des dispositifs de différenciation et de fragmentation des informations recueillies, réalisant un codage binaire de ces informations. Selon les connaissances antérieures de l'automate et ses attentes, les sensations donnent lieu, par l'intermédiaire de l'ordinateur, à une perception. Cette perception est mémorisée, et associée à des "souvenirs" sémantiquement proches. La proximité de ces "souvenirs" varie selon l'"attitude" de la machine ; l'auteur prévoit six "attitudes" parmi lesquelles l'opérateur peut

choisir : attitudes de description objective, d'explication, de sympathie, d'antipathie, de jugement esthétique, de jugement moral. La description de l'évènement, obtenu par association d'idées, est ensuite imprimée par l'unité d'expression verbale.

c) Simulation de l'apprentissage.

La modèle le plus simple que nous connaissons dans ce domaine est celui de conditionnement inventé en 1952 par GRENIEWSKI (103, Pp. 37-45). L'auteur montre qu'un automate très rudimentaire, composé de six éléments réalisant chacun soit une opération logique (conjonction, disjonction), soit une fonction élémentaire (duplication des signaux, retard), et disposant de deux entrées et une sortie, peut imiter le mécanisme du conditionnement. La particularité essentielle de cet automate est de ne pas utiliser le *feed-back*. Son intérêt est surtout pédagogique, car le conditionnement ainsi établi est vite éteint, et n'est pas susceptible de généralisation.

Un modèle plus complet de conditionnement est analysé par UTTLEY (196) ; il est basé sur deux notions mathématiques : la relation d'inclusion, et les probabilités conditionnelles. Pour deux ensembles d'entrée, disjoints ou non, des connexions s'établissent dans la machine entre ces deux ensembles et l'ensemble formé par leur réunion. Ces connexions sont établies lorsque la fréquence d'occurrence simultanée des deux ensembles d'entrée atteint un certain seuil. Lorsque les deux ensembles deviennent indépendants, la réponse conditionnelle disparaît, mais plus lentement qu'elle n'est apparue (conformément à ce qui se passe dans le conditionnement animal).

FEIGENBAUM a réalisé un modèle élémentaire d'apprentissage verbal (79), dont le programme d'exécution sur ordinateur (écrit en IPL-V) a été baptisé EPAM (*Elementary Perceiver and Memorizer*). Ce modèle vise à reproduire le processus d'acquisition de syllabes sans signification étudié en psychologie expérimentale. Il comporte un sous-programme chargé de trouver l'image associée au stimulus présenté, puis de rechercher l'orientation de réponse associée à l'image, et enfin de choisir la réponse correspondante. L'apprentissage consiste à accroître et différencier l'arborescence ("questionnaire" au sens de PICARD) servant à la discrimination des images.

d) Simulation de la formation des opinions et attitudes.

Nous avons pris comme exemple de cette catégorie de simulations l'expérience d'ABELSON et CAROLL (3), ayant pour but de simuler le système de croyances en politique étrangère d'un individu. Les mécanismes mis en oeuvre dans ce modèle sont de quatre types : mémorisation, test de crédibilité, rationalisation, et système exécutif. La *mémorisation* du système de croyances individuel fonctionne selon le processus suivant. Les croyances sont stockées sous forme de phrases simples (sujet, verbe, complément direct), constituant la *structure horizontale* de la mémoire. La *structure verticale* repose sur une hiérarchie de niveaux d'abstraction : un concept donné est un *exemple* pour un concept plus abstrait, une *qualité* pour un concept moins abstrait. A cette structure objective se superpose la coloration affective que le sujet attribue à ces concepts. Le *test de crédibilité* est un sous-programme évaluant le degré de vraisemblance, pour le sujet, de nouvelles propositions entrées dans le

programme ; il consiste à explorer la structure verticale de la mémoire pour découvrir des exemples confirmant ou infirmant toute nouvelle proposition. Le *processus de rationalisation* est également un sous-programme qui intervient lorsqu'une proposition nouvelle attribue à un concept valorisé un prédicat dévalorisé dans le système de valeurs de l'individu, ou lorsqu'à l'inverse cette proposition tend à valoriser un concept dévalorisé. Trois modalités de rationalisation sont alors possibles. Le *système exécutif* supervise les activités des trois mécanismes que nous venons de décrire. Son rôle est de déterminer, pour toute nouvelle proposition, si c'est la proposition elle-même, ou bien la source d'où elle provient, dont il convient d'évaluer le degré de crédibilité. Ce système définit également la réponse qui sera donnée à la proposition examinée. La simulation soumet le modèle à quatre types d'actions : création de nouveaux exemples à intégrer dans la mémoire verticale, développement d'un style de rationalisation pour des contenus donnés, stockage de nouvelles croyances acceptées, changements graduels dans les évaluations.

Une recherche à rapprocher de la simulation d'ABELSON et CAROLL est celle de STEFFLRE (187), qui porte sur les attitudes et les comportements à l'égard d'objets ou événements nouveaux. L'auteur projette de construire un modèle permettant de prédire les comportements et les croyances concernant un item absolument nouveau. L'hypothèse de base de cette recherche est qu'en face d'un objet ou un événement nouveau, un individu aura un comportement analogue à celui qu'il aurait en face d'un objet ou d'un événement jugé par lui similaire (c'est-à-dire codé de la même manière).

e) Simulation de prise de décision.

Malgré l'importance de ce type de simulation pour la compréhension des comportements collectifs, nous n'avons trouvé que peu d'exemples de modèles de décision individuels. Nous commencerons par la description d'un projet de modèle décisionnel conçu pour simuler les choix économiques d'un individu (18). Le modèle repose sur trois postulats fondamentaux. Le premier postulat est que tout agent économique prend ses décisions *dans un certain ordre*, représentant leur degré d'importance l'une par rapport à l'autre. Pour une décision donnée, le choix entre les options possibles est le résultat d'un arbitrage entre les variables jouant un rôle prépondérant dans cette décision. L'arbitrage est déterminé par les décisions antérieures du sujet. Selon le second postulat, le comportement général de l'agent est caractérisé par des *options fondamentales*. A titre d'exemple, la répartition d'un budget entre les différents postes, ou l'emploi du temps de loisir, sont déterminés par les goûts et le mode de vie de l'agent. Enfin, le dernier postulat concerne l'action, sur une décision, des décisions antérieures et des options fondamentales de l'agent : celles-ci *limitent le champ des possibilités de choix* (marge) pour les décisions suivantes. Le processus de choix dépend des courbes de demande individuelles exprimant la probabilité d'achat selon un critère donné (prix, avantages présentés, etc.). Il se déroule en trois étapes : élimination des options refusées *a priori*, élimination des options ne satisfaisant pas aux critères (se situant en dessous du seuil d'acceptation dans les courbes de demande), et sélection finale parmi les possibilités restantes, en fonction de motivations extra-économiques. Ce modèle peut être appliqué soit à un individu isolé, soit à une sous-population sup-

posée homogène ; un modèle général aggrège ensuite les modèles élémentaires pour représenter le phénomène économique au niveau de la population dans son ensemble.

CLARKSON a réalisé une simulation de prise de décision individuelle devant le problème du choix des investissements des trusts (45). L'auteur fait en effet dépendre ce type de décision d'un seul agent, le représentant du trust. Le processus de choix de celui-ci repose sur une *mémoire* dans laquelle les compagnies sont classées par secteur industriel ; des *procédures* de recherche et de sélection utilisant les informations stockées en mémoire concernant les compagnies et les industries ; un *ensemble de règles* propres à un individu donné. La comparaison du résultat des simulations avec les décisions réelles d'un représentant d'un trust a permis de conclure en faveur de la validité du modèle.

3.1.2. Les simulations de processus de groupe.

Les exemples que nous avons retenus peuvent être classés sous trois rubriques : simulations de comportements sociaux élémentaires, d'interactions dans un groupe, de travail en groupe.

a) Simulation de comportements sociaux élémentaires.

Les exemples de cette catégorie concernent les interactions entre deux individus. L'exemple le plus simple de ce type de modèles est présenté par GRENIEWSKY (103, Pp. 45-49) ; il est constitué de deux exemplaires du modèle de conditionnement déjà

mentionné (§ 3.1.1.c), couplés l'un à l'autre après une légère modification de l'un deux. Ce modèle constitue un modèle "d'instruction", c'est-à-dire qu'un conditionnement acquis par le système instructeur est ensuite enseigné par lui au système instruit.

Le modèle présenté par John et Jeanne GULLAHORN sous le nom d'HOMUNCULUS (109 ; 110) est plus complexe. Ce modèle reproduit le comportement social d'un individu (*George*), qui demande à un autre individu (*Ted*) de l'aider à effectuer une tâche ; le déroulement des interactions entre les deux individus à partir de cette situation de départ est reproduit en s'inspirant des théories de HOMANS, expliquant le comportement social en termes de punitions et récompenses. Le programme (écrit en IPL-V) traduit de manière opératoire les cinq propositions suivantes :

- si la réponse à un stimulus a été récemment récompensée, un stimulus analogue déclanchera une réponse analogue (généralisation de la réponse) ;
- plus un individu récompense fréquemment un type d'activité de son interlocuteur, plus celui-ci tend à avoir ce type d'activité (influence de la fréquence sur le renforcement) ;
- plus un individu attache de valeur aux réactions de son interlocuteur, plus il s'efforcera de provoquer ces réactions par son comportement (influence de la valeur de la récompense attendue) ;
- plus un individu a reçu fréquemment un type de récompense, moins ce type de récompense a de valeur pour lui (utilité marginale de la récompense) ;
- plus un individu est victime d'injustice par rapport aux normes sociales, plus il tend à réagir par la colère.

Ce programme "est une tentative pour expliquer, d'une manière qui ne serait pas réalisable verbalement, l'aptitude d'un individu engagé dans une interaction sociale normale à évaluer le contexte du comportement, à rassembler l'information nécessaire à l'élaboration de plans d'action, et (...) à sélectionner les conditions dans lesquelles il émettra une réponse plutôt qu'une autre" (97, p. 386).

Un autre modèle d'interaction entre deux individus mérite d'être décrit : c'est la première ébauche du modèle *CRISISCOM*, qui a pour objectif d'expliquer les crises politiques à partir de facteurs psychologiques. Ce modèle a été testé sur un fait historique passé : le déclenchement de la première guerre mondiale (164). Il représente l'interaction de deux Chefs d'Etat (ici, le Kaiser et le Tsar), et la manière dont ils perçoivent chacun une information identique et y répondent en émettant d'autres informations. Selon les auteurs, ce modèle, appliqué aux derniers jours de juillet 1914, période sujette à de grands changements politiques, a donné satisfaction. Après l'avoir perfectionné, les auteurs se proposent de l'appliquer à l'étude des crises internationales contemporaines.

b) Simulation des interactions dans un groupe.

Nous citerons deux exemples, empruntés tous deux à COLEMAN, dont les recherches portent surtout sur l'étude dynamique des structures sociales.

Le premier exemple concerne l'influence du groupe de référence sur ceux qui le prennent comme modèle (46, Pp. 64-66). Les données sont obtenues par enquête auprès d'élèves du second degré (*high*

schools). On pose à chacun deux questions : auquel de ses camarades aimerait-il le plus ressembler ? et aimerait-il mieux laisser le souvenir d'un brillant élève, d'un sportif accompli, ou d'un bon camarade ? A partir de la première question, on obtient un graphe des préférences, pouvant comporter de nombreux chemins. On fait ensuite l'hypothèse que, si l'individu A, qui désire être un sportif accompli, admire B, qui lui, voudrait laisser le souvenir d'un brillant élève, il y a une certaine probabilité pour que A adopte l'idéal de B. Le problème est de déterminer dans quel sens évolueront les attitudes du groupe dans une telle hypothèse. La détermination des probabilités de transition d'une attitude à une autre est arbitraire ; ces paramètres peuvent par conséquent être manipulés afin de mesurer leur effet sur l'évolution. Selon COLEMAN, les valeurs attribuées à ces paramètres ne modifient pas le sens de l'évolution du groupe.

Le second exemple porte sur l'évolution des interactions dans un groupe de trois personnes (46, Pp. 66-69 ; 48, Pp. 1053-1054 ; 49, Pp. 92-94), étudiée par KIRK et COLEMAN. Le problème posé est le suivant : lorsque trois personnes sont en interaction, comment se fait-il que le groupe, à l'issue de son évolution, se décompose en une paire et un isolé ? Le modèle repose sur l'hypothèse qu'on interagit plus avec les personnes qu'on préfère, et qu'on aime plus les personnes avec lesquelles on interagit plus. La simulation montre que le groupe *peut* évoluer vers le système *paire + isolé*, mais qu'il peut également rester équilibré (avec de faibles oscillations), ou présenter un *leader* dominant les deux autres individus faiblement liés entre eux.

c) Simulation du travail en groupe.

Kullervo RAINIO a réalisé une série de simulations reproduisant une expérience de laboratoire sur la résolution de problèmes en groupe (168 ; 167, Pp. 76-88 ; 169, Pp. 289-290). Les comportements des individus en présence sont formalisés à partir du modèle d'apprentissage à deux opérateurs de BUSH et MOSTELLER ; l'interaction sociale est considérée par RAINIO comme un processus séquentiel de choix individuel, chaque contact étant systématiquement sanctionné (récompensé ou puni). La sanction de chaque choix dépend de l'ensemble des choix simultanés des individus du groupe. La tâche à accomplir consiste à parcourir un "labyrinthe verbal", chaque membre du groupe fournissant à chaque essai un élément du choix. RAINIO a entrepris parallèlement des expériences de laboratoire et des simulations portant sur des groupes de trois sujets ; la comparaison entre ces deux types d'investigation révèle un accord entre les résultats d'expérience et les résultats simulés pour certains types de groupes expérimentaux seulement. Le modèle semble donc n'être applicable que dans certaines conditions.

SCHELLHORN a mis au point un modèle d'équipe au travail intégrant les facteurs personnels (rythme d'activité individuel, interactions entre membres de l'équipe, satisfaction retirée du travail) souvent négligés par les simulations réalisées auparavant (183). L'équipe est composée de huit exécutants déjà formés aux travaux qui leur incombent ; ils ont à assumer six types de fonctions. Le programme de travail est figuré par un graphe orienté (de type *PERT*), indiquant le temps de travail nécessaire à chaque tâche et leur ordre de succession. Les événements qui surviennent au cours de l'accomplissement du programme de travail sont rythmés par le calendrier des membres de l'équipe (alternance travail/repos) :

un évènement arrivant pendant une période de travail est ressenti par tous les exécutants d'une même tâche, tandis qu'un évènement se produisant en dehors du temps de travail n'affecte qu'un individu. Les réponses émotionnelles des individus à ces évènements se traduisent par des freinages stationnaires ("peine") devant un échec, des activations transitoires ("colère") en présence d'un obstacle, des activations stationnaires ("joie") à la suite d'un succès. La succession de deux réponses émotionnelles de types différents suscite soit des sentiments durables tels que l'affliction ("frayeur" suivie de "peine") ou la satisfaction ("colère" suivie de "joie"), soit des conflits. Le comportement de chaque individu, c'est-à-dire son niveau d'activité dans l'équipe, est déterminé par son état affectif.

3.1.3. Les simulations électorales.

Les simulations électorales, assez répandues en raison de leur intérêt pratique, sont de deux types : les simulations par agrégation de modèles de décision microsociologiques, et les simulations macrosociologiques.

a) Les simulations par agrégation de modèles microsociologiques.

A ce type de simulations appartiennent les travaux des pionniers de la simulation électorale, COLEMAN et Mc PHEE. L'objectif principal de ces expériences est de reproduire les processus de formation et de modification des opinions dans une microsociété.

Le modèle de COLEMAN et WALDORF (cf. 49, Pp. 97-98 ; 48, Pp. 1055-1056), élaboré à l'occasion des élections présidentielles de 1960, a été construit pour utiliser des données d'enquête recueillies spécialement. Les données, concernant 800 personnes de Baltimore, portaient sur leurs attitudes envers divers problèmes politiques, l'importance à leurs yeux de ces problèmes, leur exposition aux différents moyens d'information de masse, leurs attitudes envers les candidats au début de la campagne, et les relations existant entre les interviewés (structure sociale). Les entrées de la simulation furent constituées par les événements de la campagne électorale, et les sorties par les attitudes en faveur de chaque candidat ; le modèle visait à reproduire l'influence de la campagne sur les intentions de vote, à la fois directement par l'intermédiaire du système de valeurs des électeurs, et indirectement par l'intermédiaire de la structure sociale (réseau de communications). Au niveau de l'ensemble de l'échantillon étudié, les changements d'attitudes au cours de la campagne semblent avoir reflété correctement les changements réels.

La simulation de Mc PHEE et SMITH a été réalisée à la même époque pour rendre compte des élections primaires dans le Wisconsin (cf. 32, Pp. 12-13 ; 49, p. 97 note 5) ; les données ont été fournies par l'*Institut Roper*. Le modèle de Mc PHEE avait été conçu à l'origine pour expliquer les fluctuations politiques que l'on observait entre les élections présidentielles (163, p. 7) ; il faisait intervenir trois processus : processus de *stimulation* (effets sur l'individu des événements de la campagne), de *discussion* (modification des opinions sous l'influence des échanges de vues avec l'entourage), et d'*apprentissage* (perméabilité aux influences selon la régularité et l'homogénéité des comportements antérieurs).

b) Les simulations macrosociologiques.

L'exemple le plus connu de ce type de simulations est probablement le "projet SIMULMATICS", que nous analysons dans la seconde partie de ce chapitre (§ 3.2.2.). Howard ROSENTHAL a examiné la possibilité d'appliquer la démarche suivie par l'équipe *Simulmatics* à la vie politique française (177). D'autre part, diverses expériences de simulation ont été réalisées en France, à l'occasion des élections législatives des 5 et 12 mars 1967 (en particulier pour l'I.F.O.P., et la S.E.M.A./S.O.F.R.E.S.) ; malheureusement, le résultat de ces expériences reste confidentiel, et n'a pas fait l'objet de publications (*).

3.1.4. Simulation du fonctionnement des organisations.

Nous avons classé les exemples de ce type en deux catégories : les simulations du comportement interne d'une organisation, et les

(*)

Il ne faut pas confondre ces simulations avec les expériences de prévision faites à l'occasion de ces élections. Les stations de radio périphériques ont diffusé des essais de prévision des résultats définitifs du scrutin à partir des dépouillements effectués dès 20 heures dans quelques bureaux de vote judicieusement choisis. De telles opérations reposent sur l'évaluation des "fourchettes", c'est-à-dire des marges d'erreurs dans l'estimation d'un paramètre à partir d'un échantillon de la population totale (cf. 71). Une autre expérience intéressante a consisté à expliquer a posteriori les transferts de voix entre le premier et le second tour (cf. 72 ; 73). Ces tentatives se distinguent de la simulation en ce qu'aucune expérimentation (manipulation de paramètres) n'a été réalisée, les modèles sous-jacents étant des modèles statistiques très simples. Les simulations réalisées après le premier tour de ces élections ont consisté à reproduire expérimentalement les effets, sur les transferts de voix, de diverses stratégies de désistement des candidats en vue du second tour. Les résultats de ces simulations demeurent la propriété des organisations qui les ont financées.

simulations des interactions entre plusieurs organisations.

a) Simulation du comportement d'une organisation.

Le premier exemple de cette catégorie est la simulation du fonctionnement d'une entreprise par SPROWLS (186). Le postulat de base du modèle est qu'une firme a pour tâches essentielles de provoquer des transactions de différentes sortes, de conserver de l'information concernant ces transactions, et de prendre des décisions fondées sur le traitement de cette information. Les sous-systèmes les plus importants pour la simulation sont les ouvriers de la production, les machines, les acheteurs du produit fabriqué, les fournisseurs du matériel brut, et les sources de financement. Le modèle de chaque sous-système inclut des unités individuelles (ouvrier, machine, vendeur, etc.). Le modèle général comporte, en plus des cinq sous-systèmes, diverses procédures telles que tenue des comptes, inventaire des stocks, etc.. L'entreprise fictive choisie comme objet de la simulation est une usine métallurgique de transformation, employant 400 salariés, et ayant un chiffre d'affaires de 4 millions de dollars. Sous la forme décrite par l'auteur en 1962, le modèle était incomplet, et simulait seulement le processus de production.

Le second exemple est un aspect de la recherche connu sous le nom de "projet LEVIATHAN" (176). Cette simulation reproduit également le fonctionnement d'une firme, mais en mettant l'accent sur les processus de direction à l'intérieur de la firme. La firme, dont l'objectif est de déchiffrer les messages codés comportant des erreurs, est composée de deux sous-systèmes : le sous-système

technologique (production), et le sous-système de gouvernement (décisions). Ces deux sous-systèmes sont mutuellement adaptatifs. L'ensemble du système intègre les trois types de structures définies par les études d'organisation : structure formelle verticale (*ligne*), structure formelle latérale (*staff*), et structure informelle (réseau des relations personnelles). Le sous-système le plus important est celui qui prend des décisions ; c'est également le sous-système le plus complexe. Les décisions sont de deux types : décisions de contrôle, destinées à réaliser les objectifs fixés, et décisions de commande, fixant les objectifs. Ce second type de décision suppose la définition d'une politique, et une anticipation de l'avenir. Les responsables des décisions sont considérés comme les noeuds d'un réseau d'information. Chaque responsable est au départ affecté de traits de caractère et d'attitudes (par exemple : forte tendance à choisir des décisions en faveur de l'expansion) ; des facteurs de modification (liés à la productivité de la main d'oeuvre, aux taux de production, au capital, etc.) sont ensuite introduits. Le processus individuel de prise de décision est lui-même un processus assez complexe, comportant huit phases ; la décision finale est la combinaison des décisions individuelles. Elle peut concerner la charge de travail dans le sous-système de production, la répartition de la main d'oeuvre, et la définition des priorités internes.

Une expérience d'un type différent est la simulation réalisée par BOUDON et DAVIDOVITCH pour expliquer les abandons de poursuite judiciaire (60). L'objectif de cette simulation est de rendre compte des variations du taux de *classement* ou de *non-lieu* d'une année à l'autre. L'hypothèse de base du modèle est que l'augmentation du taux de criminalité, en exerçant une pression croissante sur l'ap-

pareil judiciaire, entraîne un dépassement des capacités d'absorption des institutions ; tout se passe alors comme si les magistrats s'efforçaient d'optimiser la répression, en tenant compte d'une estimation de la "productivité" des organismes répressifs. Les critères guidant le choix des magistrats sont essentiellement la probabilité d'aboutir à une condamnation si l'affaire est portée à l'audience, et le préjudice social causé par le délit. Ces critères sont d'autant plus stables, pour un type de délit donné, que celui-ci est plus grave ou plus fréquent. Le modèle lui-même est construit à partir de quelques propositions simples. La première est que, pour chaque type d'infraction, on dispose d'une distribution de la mesure globale de l'aptitude des infractions de ce type à être portées à l'audience. La seconde proposition concerne l'appréciation, par les juges, de la capacité d'absorption de l'appareil judiciaire : le seuil *a priori* des poursuites pour un type d'infraction donné est calculé à partir du taux de poursuite de l'année précédente. Le modèle (programmé en Fortran) fonctionne de la manière suivante : pour une année donnée, le corps des magistrats sélectionne un certain nombre d'affaires parmi celles qui leur ont été soumises. Si le total des affaires retenues excède la capacité de l'appareil judiciaire, les affaires les moins aptes à aboutir seront éliminées (classement ou non-lieu). Cette élimination suppose que l'on calcule la distance de chaque affaire retenue au seuil de poursuite des infractions du même type ; ces distances sont pondérées par un coefficient variant en sens inverse de la gravité et de la fréquence de ces infractions. Les affaires qui, après pondération, sont les plus proches du seuil, sont éliminées. Si le nombre d'affaires retenues excède encore la capacité de l'appareil, le processus recommence ; si le nombre d'affaires retenues est trop faible, la démarche s'effectue alors en sens inverse. La simulation, pour quinze catégories d'infraction,

sur une période de plus de cinquante années, donne des résultats par type d'infraction proche des résultats observés.

Les auteurs de simulation de systèmes économiques ont, à la suite de FISHER (cf. § 2.1.2. d), marqué une certaine préférence pour les dispositifs analogiques. Signalons d'ailleurs à ce propos que GRENIEWSKY a réalisé un modèle simple d'économie planifiée par un organe central (103, Pp. 125-127). Deux expériences intéressantes de simulation ont été réalisées sur le problème complexe des crises de surproduction (8, Pp. 201-206). La première a été la construction, par Otto J.M. SMITH, d'un dispositif électronique reproduisant le fonctionnement d'une économie de marché. La circulation des capitaux était représentée par celle d'un courant électrique d'une fréquence donnée, un phénomène réel durant une année étant simulé par un processus de 0,2 millième de seconde. Les principaux résultats de la série de simulations entreprise à l'aide de ce modèle concernent l'inévitabilité et la périodicité des crises de surproduction, et l'effet de certains palliatifs tels que la réglementation de la production, la limitation des investissements, les subventions gouvernementales, ou le crédit à bon marché. Arnold TUSTIN (103) a expérimenté sur des dispositifs électroniques reproduisant divers modèles de type "keynésien". Il a montré que, contrairement aux théories de KEYNES, la réglementation de l'économie nationale par l'Etat ne permet pas d'éviter les crises de surproduction, et que celles-ci ne sont pas dues à l'insuffisance des investissements.

Une tentative pour simuler un système social dans son ensemble a été faite par C. DOMINGO et A. VARSAVSKY (67). Ces auteurs ont voulu étudier expérimentalement la stabilité sociale dans l'*Utopie* de Thomas MORUS. Les éléments du modèle sont les caractéristiques de l'Etat (11 variables qualitatives, et 2

variables "mesurables"), et les acteurs, définis par la catégorie sociale à laquelle ils appartiennent (12 groupes), leurs opinions relatives à chacune des caractéristiques de l'Etat, et leurs traits de caractères (10 variables "mesurables"). Diverses lois définissent les variations de chacune des caractéristiques individuelles selon leurs valeurs antérieures. Les premières simulations ont en particulier permis d'étudier le mécanisme de formation de coalitions.

b) Simulation des interactions entre plusieurs organisations.

Les simulations des interactions entre organisations reproduisent toutes des situations de concurrence entre organisations : concurrence commerciale entre firmes, ou concurrence politique entre nations.

HOGGATT a simulé un marché fictif, dans lequel les firmes subissent les effets de la concurrence (32, Pp. 14-15). Sur ce marché, le prix est déterminé de manière à réaliser un équilibre entre l'offre et la demande ; celle-ci dépend de l'évolution des prix. Les firmes sont de dimensions diverses, et leur productivité est fonction de leur taille. Il y a deux types de firmes, selon la manière dont elles déterminent leur production en faisant des pronostics sur l'offre : celles qui anticipent l'offre à partir des prix de la période écoulée, et celles qui l'anticipent à partir de l'agrégation des fonctions de coût particulières. Les séries de simulation entreprises par HOGGATT ont montré que les profits étaient réalisés essentiellement par les firmes de la seconde catégorie, celles de la première catégorie quittant progressivement le marché.

HOGGATT et BALDERSTON ont construit un modèle reproduisant non plus un marché fictif, mais le marché du bois dans la région nord-ouest du Pacifique (48, p. 1058). Cette simulation, utilisant des données empiriques, a permis de mieux comprendre le fonctionnement de ce système économique, et a révélé une tendance à la segmentation du marché.

Un certain nombre d'expériences de laboratoire sur les relations internationales ont été réalisées sous la direction de Harold GUETZKOW (cf. 105 ; 106) ; mais la quasi-totalité de ces expériences ressortissent plus au jeu de rôle qu'à la simulation telle que nous l'avons définie. Ces recherches ont néanmoins inspiré quelques essais de simulation sur ordinateur. BENSON (21) a simulé les interrelations entre dix-huit Etats (neuf Etats principaux, et neuf Etats satellites). Les Etats principaux étaient définis par leur potentiel militaire (en termes de population, produit national brut et par tête, production d'énergie, taux d'alphabétisation, etc.), leur "propension à agir" (c'est-à-dire leur agressivité), leurs bases à l'étranger, et leur puissance atomique. De plus, on disposait d'informations complémentaires sur l'ensemble des dix-huit Etats : localisation géographique, relations commerciales, etc.. Le modèle reproduit la formation des alliances et des coalitions, les conflits, et leurs conséquences.

3.1.5. Simulation de l'évolution d'un système.

Les expériences concernant l'évolution des systèmes portent soit sur des phénomènes de diffusion à l'intérieur d'un réseau de communication donné, soit sur la croissance et le développement d'une population.

a) Simulation de phénomènes de diffusion.

Les simulations de phénomènes de diffusion peuvent concerner des populations d'individus indifférenciés, sans relations entre eux. Dans de tels modèles, chaque individu a une probabilité égale d'être en contact avec n'importe quel autre individu de la population. L'intérêt de ces modèles est d'être plus aisément manipulables que des modèles plus réalistes. On les rencontre principalement en épidémiologie (cf. p.ex : 20). Dans les sciences sociales, au contraire, les recherches sur la diffusion s'appuient sur la *structure sociale*, c'est-à-dire sur les réseaux de communication qui lient entre eux les individus et les groupes. Un exemple intéressant d'une telle approche, antérieur à l'utilisation de la simulation, est fourni par COLEMAN dans son étude sur la diffusion de produits nouveaux auprès des médecins (50).

Par ailleurs, COLEMAN a réalisé une expérience de simulation sur la diffusion de l'habitude de fumer chez les lycéens (49, Pp. 98-101). Les données empiriques recueillies préalablement sur les individus étudiés concernaient le caractère occasionnel ou régulier de l'habitude de fumer, les fréquentations et les relations d'amitié entre lycéens. La simulation permet d'estimer l'évolution du taux de fumeurs en fonction de la fréquence des contacts des non-fumeurs avec des amis fumant déjà auparavant.

Torsten HÄGERSTRAND a reproduit, par simulation, la diffusion d'une technique nouvelle auprès des fermiers d'une petite région de Suède au relief assez accidenté (111). Il a tout d'abord recueilli des informations sur le réseau des relations entre fermiers, en utilisant comme indicateurs le volume des communications

téléphoniques, les cessions de terrains, les migrations d'ouvriers, et les mariages. Une première série de simulations a utilisé un espace isotrope, dans lequel seule la distance entre deux fermes régissait la probabilité de contact. Les séries ultérieures de simulation ont tenu compte des accidents de terrain (fjords, montagnes) faisant obstacle aux communications ; ces expériences, utilisant un espace non-isotrope reproduisant la configuration de la région, ont fourni des résultats convergeant avec les données empiriques.

b) Simulation du développement d'une population.

Cette catégorie de simulation aborde des problèmes intéressant principalement la génétique des populations et la démographie.

En s'inspirant des travaux du généticien BARRICELLI, deux chercheurs de l'équipe *Simulmatics* ont conçu le projet de simuler un processus évolutif susceptible d'aboutir à l'apparition d'un organisme artificiel capable de résoudre des problèmes (25). Les organismes seraient représentés en mémoire par leur patrimoine génétique, incluant l'aptitude à résoudre certains problèmes logiques et arithmétiques. Ces organismes auraient la faculté de se reproduire. Ils seraient sélectionnés dans une "jungle", en fonction de leur "intelligence" (aptitude à résoudre au moins partiellement les problèmes) ; les organismes sélectionnés seraient ensuite conservés et protégés dans une "arche de Noé". Le jeu simultané des sélections et des mutations devraient permettre d'aboutir à la création d'organismes "intelligents".

Albert JACQUARD a simulé l'effet, sur le taux de reproduction humaine, de diverses techniques contraceptives (119). La population étudiée est une cohorte de 200 femmes de 15 à 45 ans. Au cours de cette période (correspondant à la période de fertilité), chaque femme est exposée à divers "risques" : mariage, conception, veuvage, remariage, décès ; ces risques sont mesurés par des distributions de probabilités (voisines des fréquences observées dans les pays occidentaux). Le paramètre variant d'une simulation à l'autre concerne l'attitude des femmes à l'égard des procédés contraceptifs féminins. Techniquement, les procédés eux-mêmes sont efficaces à 98 % ; mais la femme peut les utiliser d'une manière relâchée si son information est mal faite ou si sa motivation n'est pas forte. Son attitude à cet égard peut être soit le désir de retarder les naissances afin d'obtenir un échelonnement jugé optimal, soit la volonté d'empêcher la venue d'un enfant supplémentaire. L'auteur a réalisé sur ce modèle 37 simulations, en faisant varier la taille maximum désirée de la famille, l'espacement optimum des naissances, et l'efficacité des procédés contraceptifs. Les conclusions principales qui se dégagent de cette recherche sont que, *sur le plan familial* (planification des naissances), seule une attitude contraceptive très stricte est efficace, la contraception de retardement étant d'ailleurs plus efficace que la contraception d'arrêt ; *sur le plan collectif* (limitation des naissances), même une attitude peu efficace au niveau individuel peut abaisser sensiblement le taux d'accroissement de la population.

Un modèle démographique destiné à reproduire artificiellement l'évolution de la population des Etats-Unis a été élaboré par ORCUTT et ses collaborateurs, dans le cadre d'une recherche

économique (156, 2ème et 4ème parties ; cf. également 154 et 155). Ce modèle est constitué d'unités de deux types : les *individus* (sous-systèmes ayant pour sorties la mort et la mariage) et les *couples* (sous-systèmes ayant pour sorties la conception et le divorce). La simulation de l'évolution démographique a pour unité le mois. C'est un processus markovien, dans lequel les probabilités des différentes sorties sont liées à certaines caractéristiques de l'unité. Ainsi, le taux de mortalité dépend de l'âge, du sexe, et de la race de l'individu ; le taux de mariage, de l'âge, du sexe, et de la situation de famille (célibataire, veuf ou divorcé) de l'individu ; le taux de divorce, de la durée du mariage ; le taux de conception, de l'âge de la femme, du nombre d'enfants qu'elle a déjà eus, et de l'intervalle de temps qui s'est écoulé depuis la dernière naissance. Un facteur saisonnier corrige ces taux. Les probabilités sont tirées des informations fournies par les recensements. Pour la simulation des mariages, l'appariement des conjoints fait l'objet d'une procédure particulière, fondée sur l'âge et la situation de famille des individus désignés. Les premières simulations réalisées ont montré une précision satisfaisante dans les résultats ; le modèle est prévu pour s'intégrer ultérieurement à un modèle de prévision économique.

3.2. DESCRIPTION DETAILLEE DE DEUX EXPERIENCES DE SIMULATION EN SOCIOLOGIE.

Nous avons décrit succinctement un certain nombre d'applications des techniques de simulation dans les sciences sociales, afin de fournir un aperçu de la variété des problèmes traités. Ce survol ne permet pas de comprendre le détail de l'élaboration des modèles, et de suivre les étapes de la démarche des chercheurs. C'est pourquoi

nous exposons dans ce sous-chapitre deux expériences assez différentes, bien que réalisées toutes deux sur ordinateur numérique. La première expérience est une simulation stochastique effectuée pour vérifier une théorie microsociologique ; elle repose sur un modèle mathématique relativement élaboré, et n'utilise aucune donnée empirique. Les expériences de ce type étant assez nombreuses, nous avons choisi de décrire une simulation souvent citée (entre autres par BOUDON et COLEMAN), mais peu accessible en France, car le texte en a été simplement ronéotypé par une Université américaine. La seconde expérience est le premier essai de simulation du mécanisme des élections présidentielles américaines, qui, à ce titre constitue une simulation "historique" dont la presse et de nombreux ouvrages se sont fait l'écho. L'écrivain Eugene BURDICK, co-auteur de *The Ugly American*, a même écrit un livre d'anticipation, *The 480*, inspiré de cette expérience. Contrairement à la simulation précédente, celle-ci repose sur un modèle mathématique déterministe très simple, mais utilise une masse considérable de données empiriques.

3.2.1. La simulation du comportement de production dans un atelier.

Cette recherche a été effectuée par Raymond BRETON (35), sous la direction de James COLEMAN (*). En simulant le comportement de production d'une équipe au travail, l'auteur a voulu montrer qu'il était possible de provoquer des comportements identiques chez les travailleurs, sans nécessairement faire appel à l'intériorisation ou la transmission d'une norme implicite que chacun s'ef-

(*)

Nous tenons à remercier le Professeur Raymond BRETON de l'Université Johns Hopkins (Baltimore, Maryland, Etats-Unis), qui nous a aimablement communiqué le texte de cette étude.

forcerait de respecter afin de maintenir la cohésion de l'équipe.

a) Objectifs de la recherche.

Depuis les travaux d'Elton MAYO (cf. 37, ch. 3), les spécialistes de la sociologie industrielle ont généralement admis que la production d'un atelier dépend non seulement des instructions données par la maîtrise et des avantages accordés aux travailleurs, mais également de l'action de processus sociaux informels à l'intérieur de l'atelier. MAYO note par exemple que, dans certains ateliers, la production est freinée par les travailleurs malgré toutes les incitations pécuniaires à produire plus, et que ce phénomène semble dû aux pressions du groupe sur chaque travailleur. La thèse généralement admise est qu'il existe une *norme de production* dans l'atelier à laquelle les travailleurs sont conduits à se conformer sous l'influence d'un système informel de sanctions morales. Cette norme jouerait un rôle essentiel dans le maintien de la structure sociale de l'atelier, et de la cohésion de l'équipe.

Dans la littérature traitant de ce problème, on ne traite pas séparément les facteurs qui sont causes de l'apparition d'une uniformité de comportement, et ceux qui donnent à cette uniformité un caractère *normatif*. De même, on distingue rarement les facteurs qui déterminent le *contenu* de la norme (accroître ou restreindre la production) de ceux qui expliquent l'*efficacité* de cette norme sur la tendance à l'uniformité. Or il faut souligner qu'une tendance à l'uniformité peut se manifester dans un groupe sans devenir une norme, à moins que les membres du groupe ne s'accordent à considérer ce comportement uniforme comme une manière correcte de se

comporter. Ainsi, ce qui apparaît aux observateurs comme une norme de groupe peut n'être en réalité qu'une uniformité de comportement, résultant de la recherche, par chaque membre du groupe, de son propre avantage individuel. Certes, si l'observateur fait prendre conscience aux membres du groupe de l'existence de cette uniformité de comportement, ceux-ci peuvent découvrir l'intérêt de ce comportement homogène, et lui conférer, par une décision commune, une valeur normative ; mais il s'agit alors d'un processus distinct du précédent. D'autre part, le sens dans lequel la norme agit, et les facteurs qui déterminent cette orientation, sont d'une importance primordiale pour le problème étudié. Selon HOMANS, l'apparition d'un comportement uniforme, et la nature de ce comportement, sont déterminées uniquement de façon indirecte par l'environnement du groupe. Les paramètres de l'environnement expliquent les processus de groupe. Ces processus impliquent certaines formes d'interaction entre les membres du groupe ; les interactions ont pour conséquence la création de sentiments (positifs ou négatifs) de chaque membre du groupe à l'égard de chacun des autres membres. L'apparition d'un comportement uniforme peut alors être la conséquence de ces sentiments réciproques.

Le but de cette recherche est de simuler certains aspects du comportement d'une équipe au travail, afin de déterminer les conditions de l'apparition d'une uniformité dans les comportements de production. Le modèle utilisé reproduit les processus qui se déroulent à l'intérieur du groupe, tels qu'ils ont été décrits par les "sociologues industriels" au cours de recherches empiriques. Les entrées du modèle sont : diverses règles portant sur l'organisation de la production ; le système de rémunération ; et le profil psychologique des travailleurs avant leur intégration à l'équipe. Les sorties sont le niveau global de la production par équipe, et le

degré d'uniformité de la production des membres de l'équipe. Les relations entre équipes et l'éventualité d'une uniformisation de la production au niveau des ateliers n'ont pas été étudiées.

b) Les facteurs d'entrée et leur influence sur la structure de l'équipe.

Les règles portant sur l'organisation de la production sont édictées par la Direction ; elles définissent les relations fonctionnelles qui doivent s'établir entre les ouvriers pour effectuer la tâche qui incombe à l'équipe. Ces règles créent une interdépendance entre les ouvriers (exemple du travail à la chaîne), qui peut prendre la forme d'une coopération (exemple de tâches collectives imposant une coordination des efforts de chaque membre de l'équipe). Le modèle de BRETON suppose que le travail de chaque individu est totalement indépendant de celui des autres (mais non nécessairement identique à celui-ci). En conséquence, la Direction n'impose aucune relation fonctionnelle entre les ouvriers ; si des interactions s'établissent, elles ne peuvent qu'être la conséquence d'autres facteurs impliqués par la situation de travail.

Le système de rémunération peut être défini à partir de deux dimensions : la nature de l'unité payée, et le mode de paiement. L'unité payée peut être le temps de travail, la pièce réalisée, le niveau de qualification, ou une combinaison de ces facteurs. Le mode de paiement peut être individuel ou collectif ; l'ouvrier est payé dans le premier cas en fonction de sa propre production, dans le second cas en fonction de la production moyenne de l'équipe. Le mode de rémunération peut être *fixe* (c'est-à-dire être indépendant des

variations dans les conditions de travail) ou *ajusté* ; un mode de rémunération ajusté peut être fonction du nombre de pièces produites, ou peut être fondé sur la quantité minimum du travail par unité de temps. Dans le modèle de BRETON, le système de rémunération est un système ajusté de rémunération individuelle aux pièces ; le salaire d'un ouvrier est fonction linéaire de sa production. On suppose également que le niveau de qualification est le même pour chaque ouvrier. Le réajustement du taux de rémunération peut avoir lieu en fonction des variations des conditions de travail ou de la situation économique.

Le *profil psychologique* de chaque travailleur influe sur son comportement dans l'équipe en ce sens qu'il doit, dans son travail, prendre la décision d'agir ou non, de produire ou non, et que ces décisions dépendent de ses motivations et de ses intérêts. Selon l'auteur, les motivations et les intérêts d'un travailleur américain dépendent principalement de sa formation, de ses croyances religieuses, des valeurs qu'il attache à la famille et à la collectivité, de son histoire professionnelle, de ses besoins économiques, etc. ; ils sont donc déterminés par des facteurs extérieurs à la situation de travail. Dans le modèle, on suppose que chaque ouvrier a un niveau donné de motivation pour maximiser ses gains ; ce degré reste constant dans le temps. On suppose de plus que les ouvriers sont tous également soucieux de leur propre intérêt, et tous également capables de développer des sentiments positifs ou négatifs les uns envers les autres, et à l'égard de la Direction. Enfin, dans certaines conditions, les ouvriers peuvent avoir tendance à exercer certaines contraintes sur leur propre comportement, ou sur celui des autres membres de l'équipe.

Ces facteurs ont pour effet de structurer l'équipe de travail. La Direction et les ouvriers recherchent chacun leur propre intérêt. La Direction a établi les conditions de rémunération, et a défini des règles de réajustement des salaires lorsque les conditions de travail changent : lorsque la productivité moyenne de l'équipe croît, le salaire aux pièces décroît, et réciproquement. Dans l'exécution de leur tâche, si les ouvriers recherchent la satisfaction de leur propre intérêt, cet intérêt ne coïncide pas nécessairement avec l'idée que la Direction s'en fait. Ainsi, ils ne cherchent pas systématiquement à produire plus que la moyenne, ni à maximiser leurs gains, ni à les empêcher de descendre en dessous d'un certain seuil. De même, ils ne font pas nécessairement coïncider leurs intérêts propres avec ceux de la Direction : ils ont tendance à exercer certaines contraintes lorsque leurs intérêts sont menacés. Or, le fait que le calcul des taux de rémunération soit basé sur la production moyenne implique que *chaque ouvrier peut avoir une action sur le taux de rémunération*. Il peut accroître ou diminuer sa propre production ; il peut même exercer une action plus importante s'il réussit à influencer sur le taux de production d'autres membres de l'équipe. S'il y a un changement dans le taux de rémunération, ou simplement si les ouvriers redoutent un tel changement, chaque ouvrier s'efforcera d'exercer une contrainte sur les autres membres de l'équipe dès qu'il percevra la possibilité d'influer ainsi sur le réajustement ; *en agissant ainsi, il ne fera que défendre son intérêt propre*.

c) Le déroulement du processus dans l'équipe de travail.

Il y a deux types d'unités de décision dans le modèle : les ouvriers, et la Direction. Chaque ouvrier peut prendre deux types

de décisions : travailler ou non, et exercer ou non des contraintes sur les autres ouvriers. La Direction décide du taux de rémunération par pièce. Le déroulement du processus peut entraîner l'apparition de sentiments nouveaux. Nous allons examiner chacun de ces points en détail.

La décision de travailler ou de ne pas travailler pendant une unité de temps est une approximation discrète d'un phénomène continu. En réalité, l'ouvrier qui désire faire varier son taux de production agit sur la vitesse de travail ; dans ce modèle, le même résultat est obtenu pour un cycle de production (décomposé en vingt unités) en considérant que l'ouvrier, dont la productivité est supposée constante, travaille pendant certaines unités de temps et ne travaille pas pendant d'autres unités. Cette décision dépend de trois facteurs externes et trois facteurs internes. Les facteurs externes sont la difficulté du travail (qui, pour un même travail, peut varier d'un ouvrier à l'autre), le taux de rémunération, et les contraintes subies de la part des autres travailleurs. Les facteurs internes sont la force de sa motivation pour le salaire, sa tendance à exercer une contrainte (à la fois sur les autres et sur lui-même), et l'efficacité des contraintes qu'il subit. Cette efficacité dépend de la nature et de la force des sentiments qu'il éprouve à l'égard de ceux qui exercent les contraintes. La prise de décision, par un individu i , de travailler pendant une unité de temps t , a été programmée de la manière suivante : la propension à travailler P_i est tout d'abord calculée, comme étant le rapport des forces PT_i poussant l'individu vers le travail, à l'ensemble des forces $(PT_i + P\bar{T}_i)$ agissant sur l'individu.

$$P_i = \frac{PT_i}{PT_i + \overline{PT}_i}$$

$$\text{avec : } PT_i = a_t k + \sum_{j=1}^{j=n-1} (d_{ji}) (D_{ij})$$

$$\text{et : } \overline{PT}_i = \beta_i + \sum_{j=1}^{j=n-1} (d_{ji}) (L_{ij}) + C_i$$

où : a_t = rémunération par pièce au temps t ,

k = constante,

n = nombre d'ouvriers dans l'équipe,

d_{ji} = contrainte de l'individu j sur l'individu i ,

D_{ij} = sentiment négatif de i à l'égard de j ,

β_i = coefficient de difficulté de travail pour l'individu i ,

L_{ij} = sentiment positif de i à l'égard de j ,

C_i = tendance de i à exercer une contrainte.

La valeur calculée de P_i est ensuite comparée à un nombre pseudo-aléatoire : si P_i lui est supérieur, l'individu i travaille ; sinon, il ne travaille pas.

La décision d'exercer une contrainte sur les autres ouvriers dépend de l'éventualité d'un changement dans le taux de rémunération, et du degré d'aliénation du travailleur à l'entreprise. L'intensité de la réaction à un changement du taux de rémunération varie d'un ouvrier à l'autre ; de plus, cette intensité s'affaiblit avec le temps, selon des modalités variables d'un individu à l'autre. Le degré d'aliénation d'un ouvrier à l'entreprise (c'est-à-dire d'iden-

tification à la Direction) peut varier d'un ouvrier à l'autre, mais reste constante dans le temps. Dans le modèle, l'influence Pc_i des facteurs qui poussent l'ouvrier i à exercer une contrainte est évaluée selon la formule :

$$Pc_i = 1 - e^{-[\eta (a_0 - a_t) + r_i]}$$

où :

η = constante.

a_0 = taux de rémunération initial.

a_t = taux de rémunération au temps t .

r_i = degré d'aliénation à l'entreprise.

Si Pc_i est supérieure à un nombre aléatoire, la tendance de l'individu i à exercer une contrainte est positive ; sinon, elle est nulle. La contrainte s'exerce à l'égard des seuls ouvriers dont la production est supérieure à celle de l'ouvrier i .

La décision de la Direction de modifier le taux de rémunération dépend de la production de la période précédente, de l'idée que le directeur a d'une rémunération "raisonnable", de ce qu'il estime pouvoir convertir en salaires, et du nombre minimum de pièces qu'il doit fournir à ses clients. Le processus de décision se déroule en plusieurs phases. Le directeur calcule d'abord la production moyenne et le salaire moyen par ouvrier ; il compare ce salaire moyen avec le salaire qu'il avait l'intention de verser. Il compare également le nombre de pièces fabriquées avec le nombre demandé par ses clients. Si la masse salariale dépasse la somme qu'il était disposé à payer, le directeur calcule un nouveau taux

de rémunération a :

$$a = \frac{Rd}{Aw}$$

$$\text{avec : } Aw = \frac{1}{T} \left[\sum_{t=1}^{t=T} \left(\frac{\sum_{i=1}^{i=n} w_i}{n} \right)^t \right]$$

où :

Rd = taux de rémunération souhaité par la Direction,

T = nombre de cycles depuis le dernier changement de taux de rémunération,

n = nombre d'ouvriers dans l'équipe,

w_i = production de l'individu i .

Si la production globale est insuffisante, le taux de rémunération est multiplié par le taux d'accroissement de la production nécessaire pour atteindre le niveau de production souhaité.

L'apparition de sentiments nouveaux entre les ouvriers dépend de leur productivité et des contraintes qu'ils exercent ou subissent. Le sentiment positif que deux ouvriers ressentent l'un envers l'autre croît fortement s'ils ont la même productivité, et si tous deux ont tendance à exercer une contrainte. Le sentiment positif entre deux ouvriers croît faiblement s'ils ont seulement la même productivité ; le sentiment positif de l'ouvrier i à l'égard de l'ouvrier j croît également faiblement si i a tendance à exercer une contrainte, et si la productivité de j est inférieure à la sienne. Enfin, le sentiment négatif de i à l'égard de j croît lorsque j a une productivité plus élevée que i , ou lorsque j exerce une contrainte sur i .

Le modèle a été programmé en FORTRAN et exécuté sur l'IBM 7090 de l'Université Johns Hopkins. L'organigramme simplifié du programme est représenté par la figure 1 page 167.

d) L'expérimentation.

Le modèle a été utilisé au cours de cinq séries de simulations, correspondant chacune à une situation différente, c'est-à-dire à un ensemble différent de valeurs des paramètres d'entrée. Les valeurs de ces paramètres n'ont pas été estimées à partir d'études empiriques, mais ont été définies arbitrairement. Ces valeurs ont été néanmoins choisies de telle sorte qu'elles aboutissent à un comportement vraisemblable, tant de la part du directeur que des ouvriers, et que ces valeurs présentent un certain intérêt pour l'expérimentation. Ainsi, celle-ci n'aurait eu aucun sens si les ouvriers avaient tous travaillé au maximum de leur capacité de production, ou si le directeur avait augmenté le taux de rémunération de manière à atteindre finalement un niveau utopique. Au cours des cinq séries de simulations, la taille de l'équipe a été de quatre ouvriers, et les caractéristiques psychologiques de ceux-ci sont demeurées constantes, sauf la motivation pour les gains et le salaire minimum acceptable. En particulier, le sentiment d'aliénation envers la Direction était uniformément faible chez tous les ouvriers au cours de toutes les expériences. D'autre part, le taux de rémunération initial a été fixé à 1,50 dollar par pièce dans chaque simulation. Les valeurs des paramètres d'entrée qui ont varié d'une expérimentation à l'autre sont les suivantes :

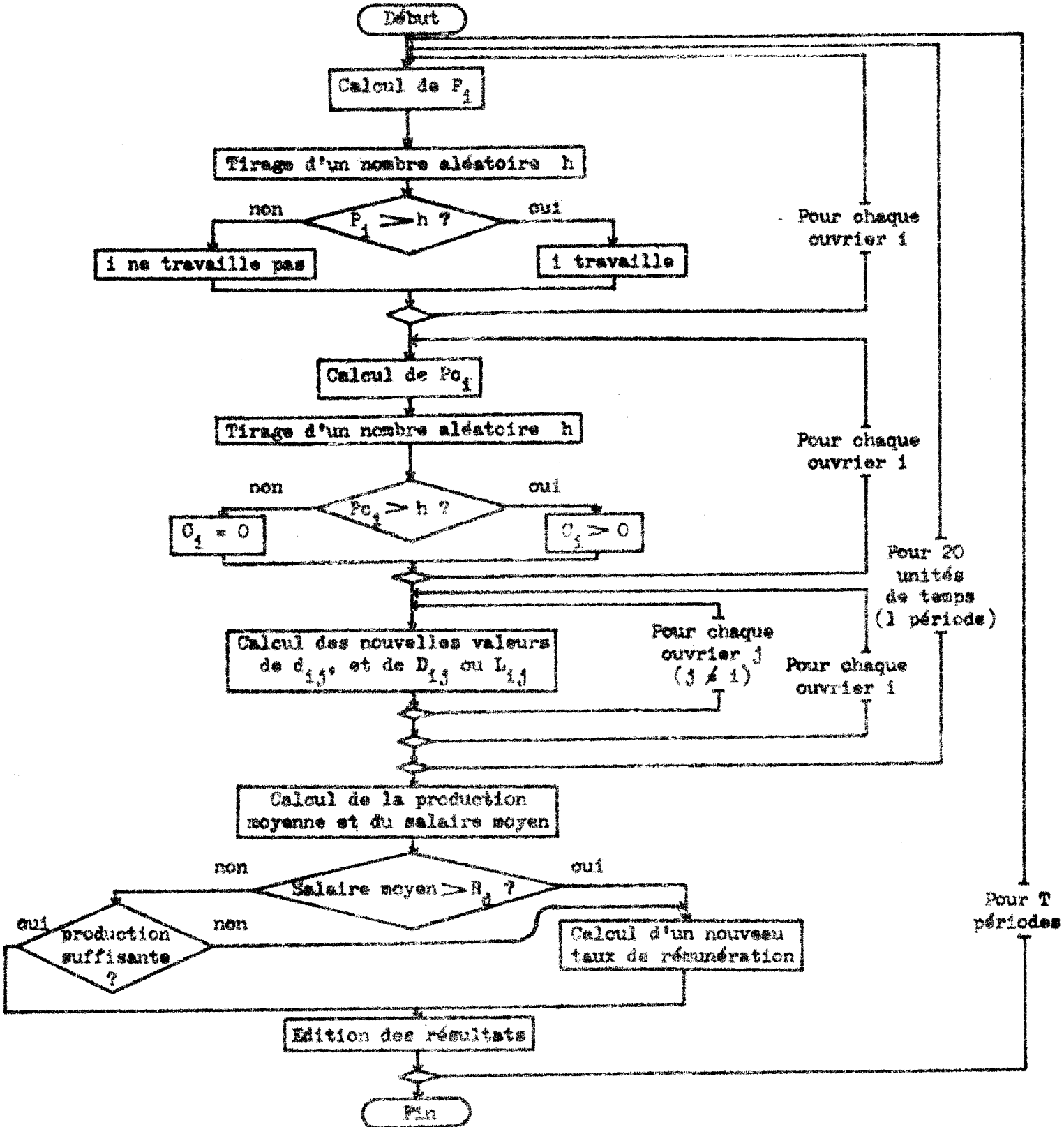


Fig. 1 Organigramme simplifié de la simulation du comportement de production dans un atelier (d'après Raymond BRETON)

Numéro de l'expérience	Degré de motivation pour les gains	Salaire minimum acceptable pour une période
1	uniformément faible	0 dollar
2	uniformément élevé	0 dollar
3	faible pour 2 ouvriers, élevé pour les 2 autres	0 dollar
4	uniformément faible	12 dollars
5	uniformément élevé	12 dollars

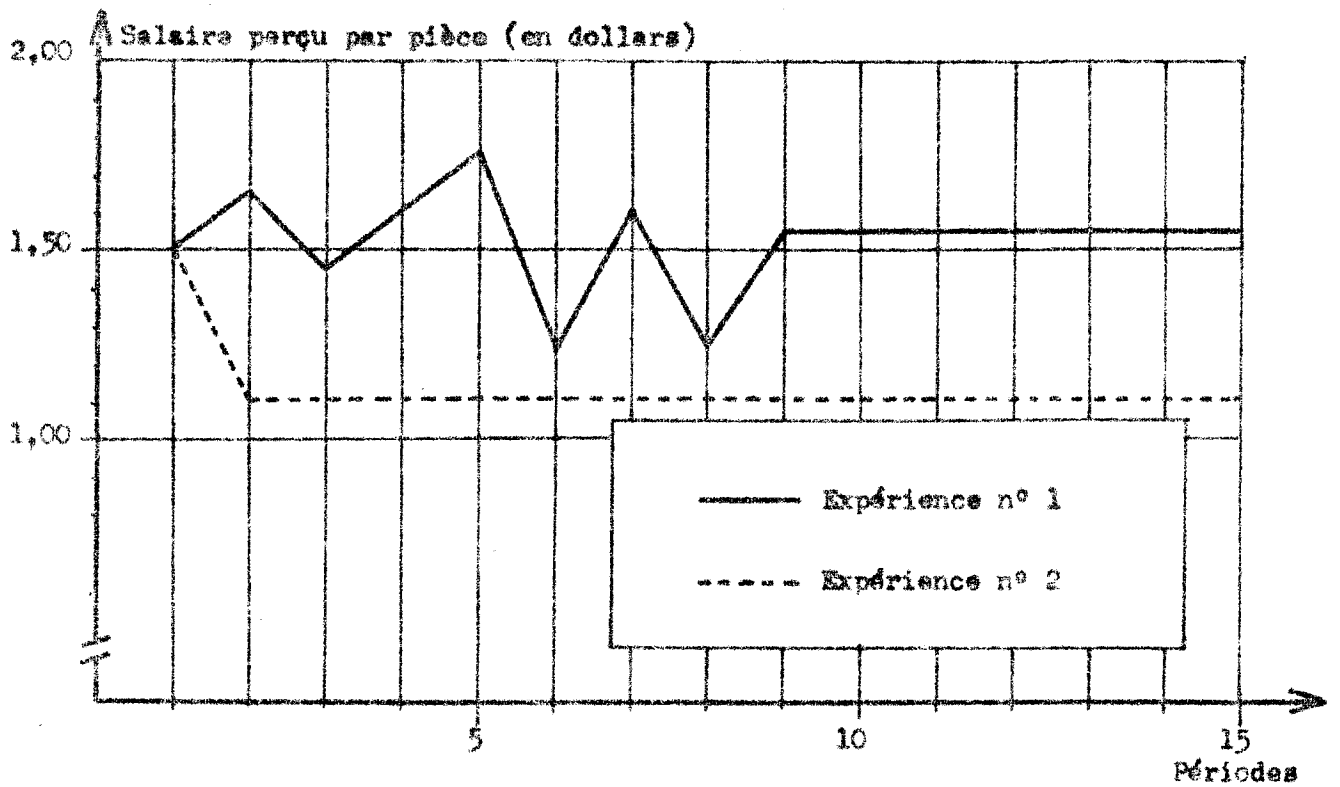
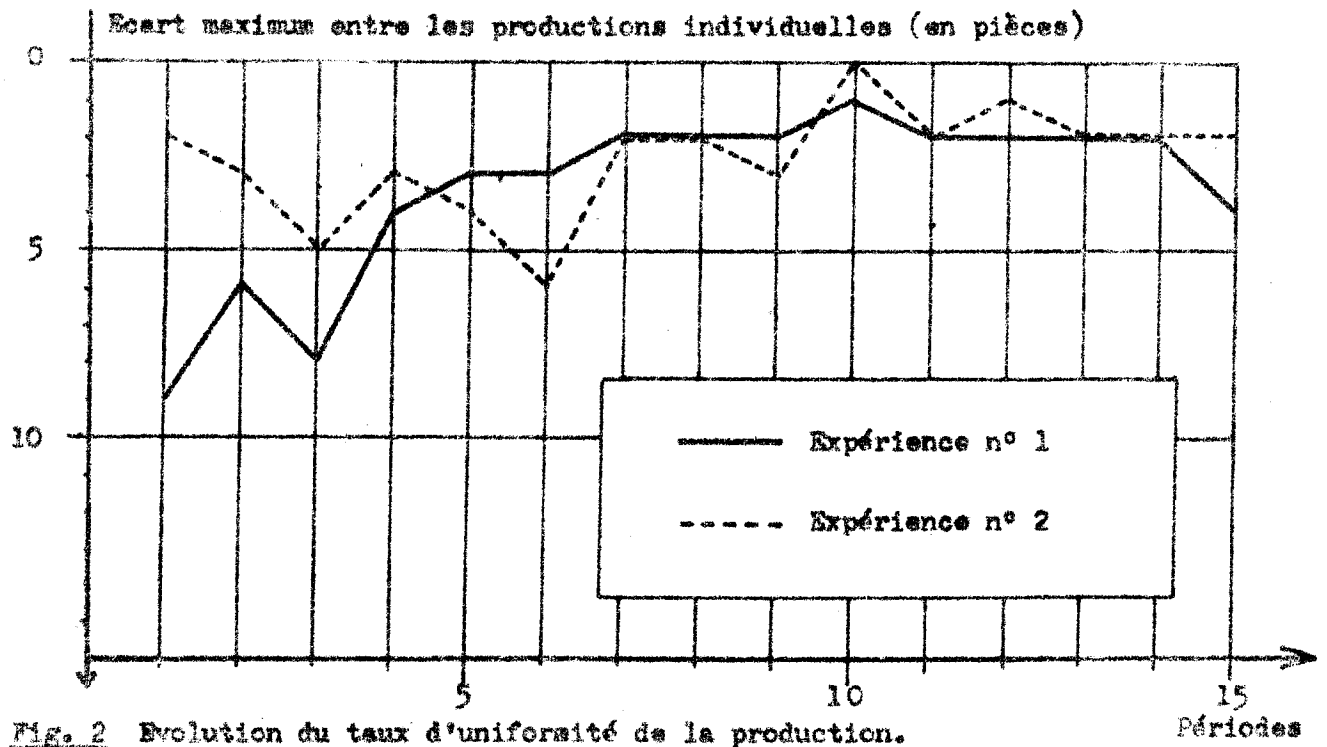
Tableau n° 1 : Conditions d'expérimentation.

Dans la première expérience, on note dès le début une forte tendance vers l'uniformisation de la production ; dès la 7ème période simulée, l'écart entre l'ouvrier qui produit le plus et celui qui produit le moins se stabilise à 2 pièces. Le taux de rémunération se stabilise également dès la neuvième période à 1,55 dollar par pièce. La pression exercée par les ouvriers est assez forte lorsque le taux de rémunération a diminué ; elle disparaît après la stabilisation du taux de rémunération, et la productivité moyenne des ouvriers demeure à peu près constante (autour de 14 pièces fabriquées par période de 20 cycles).

Dans la seconde expérience, l'influence de la force des motivations se traduit par des résultats très différents. La tendance à l'uniformité, forte au départ (écart maximum = 2 pièces) décroît ensuite et reste faible jusqu'à la 6ème période (écart

maximum = 6 pièces) ; elle remonte rapidement ensuite, pour osciller autour d'une valeur moyenne (écart maximum = 2 pièces). Le taux de rémunération s'est trouvé fixé dès la seconde période à 1,10 dollar par pièce, et ne varie plus ensuite. La pression exercée par les ouvriers, très forte au moment de la baisse du taux de rémunération, baisse fortement lorsque l'uniformité s'est établie. La productivité moyenne des ouvriers est nettement plus élevée au départ que dans la première expérience ; elle se stabilise dès la 7ème période (autour de 16 pièces par période de 20 cycles).

Les résultats des deux premières expériences (cf. figures 2 à 5, pages 170 et 171) pourraient être interprétés en faisant appel à une norme implicite de production ; en effet, l'intensification des contraintes exercées par les ouvriers entre eux suit de très près l'abaissement du taux de rémunération, et l'intensification des contraintes a coïncidé avec l'élévation du taux d'uniformité. Une analyse plus fine montre les faiblesses de cette explication. Dans l'expérience n° 1 (motivation faible), le taux d'uniformité croît ou se maintient dans les périodes où la pression est faible (périodes 4, 5, 7 et 9) ; ces expériences sont celles où le taux de rémunération augmente. Dans l'expérience n° 2, on remarque que, de la 2ème à la 6ème période, alors que les contraintes sont très fortes, le taux d'uniformité décroît (le taux de rémunération restant constant). D'autre part, quand les motivations sont fortes, le degré d'uniformité est très élevé au début, alors qu'aucune norme n'a pu s'établir. On peut rendre compte de ces résultats en disant que lorsque les ouvriers sont fortement motivés, leur productivité tend vers le maximum technologique, et leur comportement est peu différencié ; lorsqu'ensuite des contraintes s'exercent (à la suite de la baisse du taux de rémunération), elles réduisent inégalement l'influence des motivations, et les comportements se diversifient ; enfin, lorsque



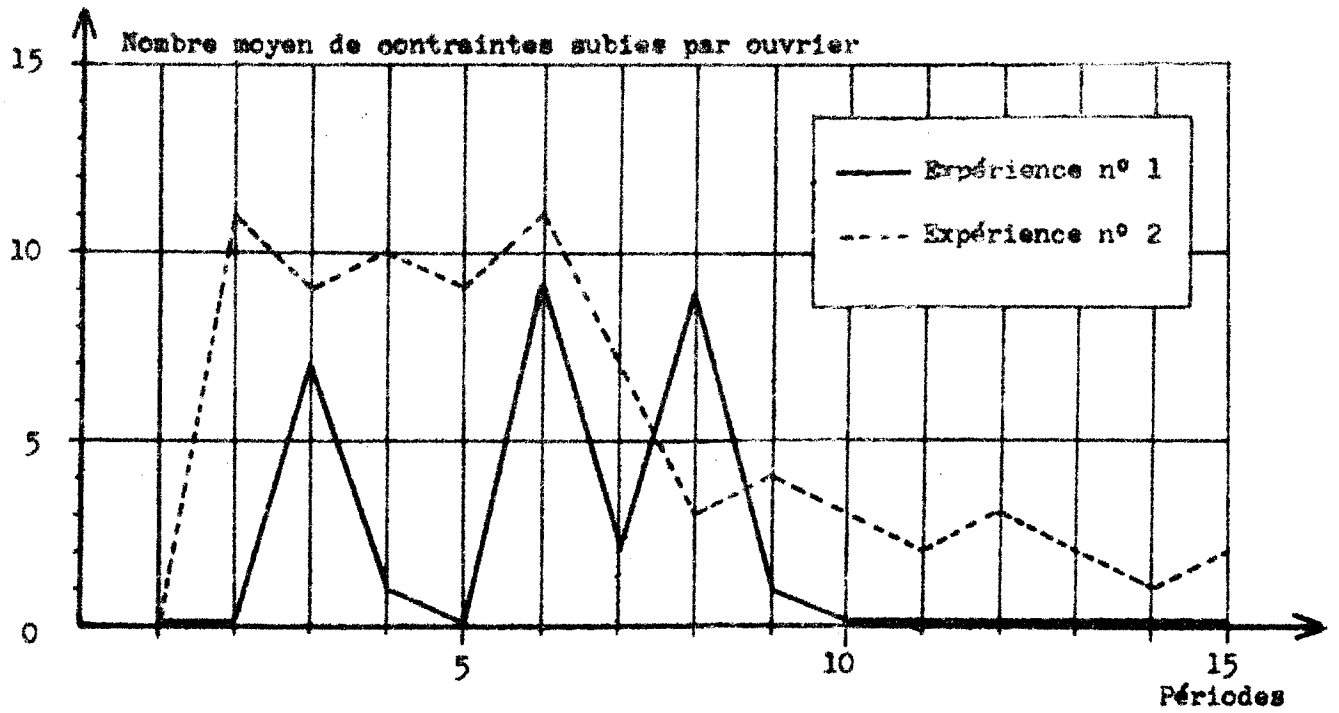


Fig. 4 Evolution des pressions de groupe.

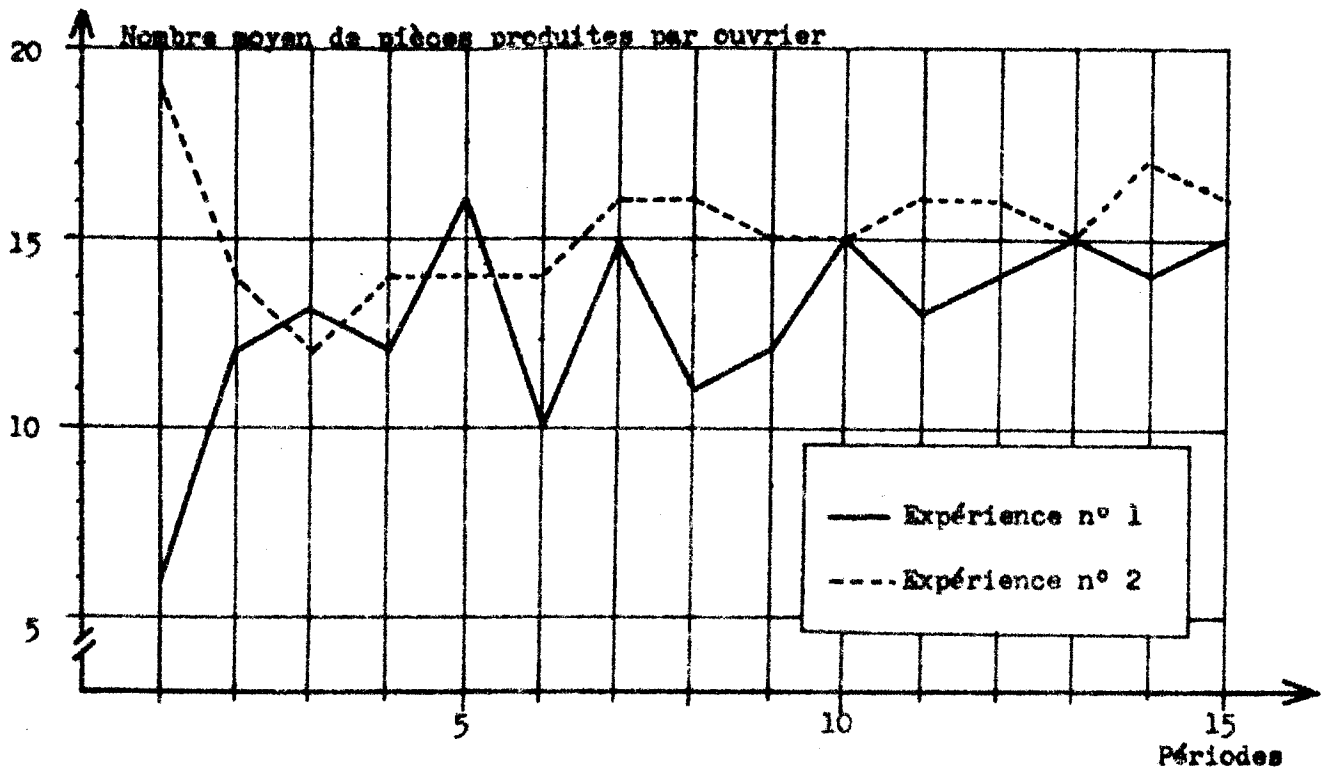


Fig. 5 Evolution de la productivité.

l'intensité des contraintes diminue, l'uniformité réapparaît sous l'influence des fortes motivations. Lorsque les ouvriers sont peu motivés, leur productivité est très inférieure au maximum, et les différences individuelles jouent largement ; les variations dans la productivité sont ensuite réduites, sous l'influence soit de l'augmentation du taux de rémunération, soit des pressions exercées par les membres de l'équipe.

L'expérience n° 3 permet de comparer les comportements, au sein d'une même équipe, d'ouvriers fortement motivés et d'ouvriers faiblement motivés. La tendance vers l'uniformité est moins marquée que dans les expériences précédentes ; elle résulte de la tendance des ouvriers fortement motivés à diminuer leur productivité, et de la tendance des ouvriers faiblement motivés à accroître la leur, sous l'effet des contraintes exercées. L'efficacité de ces contraintes augmente en raison de la croissance des sentiments positifs éprouvés par les forts producteurs envers ceux qui produisent peu. Néanmoins, les 15 périodes simulées ne suffisent pas à stabiliser le système, dont les oscillations, quoique d'amplitude décroissante, restent fortes.

On peut faire l'hypothèse que la stabilité dépend en partie du degré de motivation des ouvriers. Ainsi, dans l'expérience n° 1, le comportement de la Direction (taux de rémunération) ne se stabilise qu'à la 8ème période, alors qu'il atteint un niveau constant dès la seconde période de l'expérience n° 2 ; il en est en gros de même pour le comportement des ouvriers (productivité). Mais la stabilité du système dépend également de la rigidité des besoins minimum des ouvriers, comme le montrent les expériences n° 4 et 5. Dans celles-ci, le salaire minimum de chaque ouvrier ne peut être inférieur à 12 dollars par période, alors qu'il pou-

vait être nul dans les expériences n° 1 et 2. Dans l'expérience n° 4 (motivation faible), le taux de rémunération reste fixé à 1,50 dollar, et la productivité varie peu. Dans l'expérience n° 5 (motivation forte), le taux de rémunération tombe à 1,10 dollar dès la seconde période, et n'est plus modifié ensuite ; le système est également très stable.

e) Conclusion.

"Il est possible d'expliquer le degré d'uniformité du comportement dans une équipe, et le niveau de la production, sans postuler l'existence d'une norme. Le comportement de production de l'équipe est une conséquence de la structure de la situation de travail, du système de rémunération, et des interactions entre ouvriers. Il arrive que les ouvriers agissent par référence à une norme informelle. Mais il est également possible qu'ils agissent comme s'ils obéissaient à une norme informelle, alors qu'en fait ils réagissent aux caractéristiques de leur environnement" (35, p. 33).

3.2.2. Le projet SIMULMATICS.

Cette simulation politique (162 ; 163) a été entreprise dans le but d'aider John Fitzgerald KENNEDY à définir sa stratégie au cours de la campagne qui l'opposait à Richard NIXON pour les élections de novembre 1960 à la Présidence des Etats-Unis. Outre son intérêt pour l'historien, cette simulation présente aux yeux du sociologue plusieurs sources d'intérêt : c'est un travail de pionnier, puisque l'une des premières expériences de simulation en

sociologie politique ; elle visait la résolution d'un problème concret, et devait par conséquent être confrontée aux événements afin d'être validée ou infirmée ; sa mise en oeuvre a soulevé des problèmes méthodologiques complexes et relativement neufs, tels que celui de la fusion de données d'enquêtes ; elle a permis de clarifier certains points théoriques, concernant par exemple les mécanismes de l'abstention et du transfert des voix ; elle a enfin connu des prolongements tant sur le plan pratique (simulation de la campagne électorale de 1964) que sur le plan méthodologique (validation *a posteriori* du modèle) et théorique (analyse des courants politiques de 1950 à 1960).

a) Objectifs de la recherche.

Au cours de la préparation de la campagne pour les élections présidentielles américaines en 1960, un certain nombre de chercheurs en sciences sociales, ayant des affinités pour le parti Démocrate, s'étaient groupés autour d'un homme d'affaires new-yorkais supporter de ce parti. Ils avaient constitué une association, la *Simulmatics Corporation*, dont l'unique objectif était de réaliser une simulation politique à l'échelon de l'ensemble des Etats-Unis afin de guider la tactique électorale de J.F. KENNEDY. Depuis, cette association est devenue une entreprise commerciale s'occupant de l'utilisation de l'ordinateur dans la recherche en sciences sociales (*). Le projet soumis par l'association au parti Démocrate proposait la constitution à partir des résultats d'enquêtes politiques archivées dans une banque de données (le *Roper Public Opinion Research Center* à Williamstown), d'une sorte

(*) Sur l'éventail des activités de cette firme, cf. *Social Research with the Computer*, n° spécial de *The American Behavioral Scientist*, vol. 8, n° 9 (mai 1965). On trouvera la liste des principaux membres de la Simulmatics en page 44 de cette brochure.

de panel simulé représentatif du corps électoral américain. Une fois ce panel constitué, il devenait possible de simuler les diverses péripéties de la campagne, et ainsi d'orienter le choix des thèmes développés par KENNEDY pendant la campagne.

Les données nécessaires à la constitution du panel étaient mises, par le *Roper Public Opinion Research Center*, à la disposition du parti Démocrate à condition que les tableaux de dépouillement des données soient accessibles au parti Républicain, et que les résultats des exploitations de ces données soient publiés dans un but scientifique après les élections. Les propositions permettant de simuler la dynamique de la campagne pouvaient être tirées des recherches empiriques déjà réalisées en sociologie politique, et principalement des travaux de LAZARSEFELD et de ses collaborateurs. En particulier, l'influence, sur le vote des électeurs, de leur milieu social, de leur affiliation politique, et des thèmes de la campagne, est bien connue. On sait que les intentions de vote peuvent être modifiées par l'influence de l'entourage, la distorsion perceptive, et les contacts avec des individus appartenant à d'autres milieux sociaux. Le comportement des minorités soumises à des pressions contradictoires a été décrit par LAZARSEFELD, BERELSON et Mc PHEE ; il peut prendre la forme du vote pour le candidat de l'autre parti, de la fidélité à son parti au prix d'une distorsion perceptive, ou de l'abstention. Enfin, au moment où l'équipe *Simulmatics* préparait son projet, il existait déjà deux exemples de simulation politique, dûs à COLEMAN et à Mc PHEE (cf. § 3.1.3. a) ; ces exemples, quoique non publiés, étaient connus des chercheurs de la *Simulmatics*.

Lorsque le panel fut prêt, en août 1960, après environ une année de travail, le problème posé aux chercheurs fut de déterminer

l'attitude que le parti Démocrate devait adopter sur la question religieuse. Deux autres problèmes furent également traités au cours de la campagne ; celui du programme relatif aux droits civiques, et celui de l'importance qu'il convenait d'accorder aux questions de politique étrangère.

b) Le choix et l'organisation des données.

Les données utilisées pour constituer le panel ont été fournies par des sondages réalisés sur l'ensemble des Etats-Unis, à l'occasion des élections des représentants au Congrès, ou de la désignation des mandats pour l'élection à la Présidence, en 1952, 1954, 1956 et 1958. Les données sélectionnées devaient comporter les caractéristiques socio-économiques des interviewés, et leur intention de vote ou leurs attitudes à l'égard de divers thèmes politiques. Les auteurs de la recherche ont retenu 100.000 interviews (correspondant à 50 sondages différents), comportant des réponses à au moins quatre ou cinq thèmes jugés fondamentaux. Les données de ces 100.000 questionnaires ont été condensées à partir d'une typologie sommaire des électeurs. Cette condensation des informations était nécessité par les limites de la capacité de mémoire et la relative lenteur d'exécution de l'ordinateur utilisé (IBM 704). Les auteurs désiraient ne pas dépasser 500 types, ce qui donnait une moyenne de 200 questionnaires par type. Aussi se sont-ils limités aux variables les plus importantes pour prédire le comportement électoral, en minimisant le nombre d'états de ces variables. L'âge et le niveau d'instruction, qui pourtant exercent une influence sur le comportement de vote, ont dû être négligés au niveau de la typologie. Les variables retenues sont

la race, la religion, l'affiliation politique (*), le taux d'urbanisation, la région, le statut socio-économique, et le sexe. Les types ont été obtenus par combinaison de ces variables ; lorsqu'un type théorique avait un effectif trop faible, il pouvait être regroupé avec un type voisin. On aboutit ainsi à 480 types d'électeurs.

Les informations concernant les attitudes politiques pouvant présenter un intérêt pour la campagne ont été sélectionnées. La première difficulté était d'identifier les questions se rapportant à une attitude donnée ; la seconde était de déterminer si deux questions se rapportant à une même attitude correspondent ou non au même degré dans cette attitude. Pour résoudre la première difficulté, les auteurs ont considéré que lorsqu'une sous-population répond *oui* à une question dans un sondage donné, et *oui* à une autre question au cours d'un autre sondage, il y a une corrélation entre les deux questions. Il faut souligner que cette corrélation, observée au niveau d'une sous-population, ne se retrouve pas nécessairement au niveau des individus. Par exemple, l'analphabétisme est lié à un statut socio-économique bas ; la lecture de romans-photos également. Ce ne sont évidemment pas les mêmes individus qui sont

(*)

Aux Etats-Unis l'affiliation politique est différente de ce qu'est l'adhésion à un parti en France. En particulier, il n'y a ni carte d'inscription, ni cotisation. Il s'agit plutôt d'une identification à un parti qu'une affiliation véritable. C'est pourquoi il est fréquent qu'un électeur "affilié" par exemple au parti Démocrate, vote pour le candidat Républicain. En 1960, les taux d'affiliation étaient :

- Démocrates : 45 %
- Républicains : 31 %
- Indépendants : 24 % (177, P.23, note 5).

Le système politique américain étant bipartite, sont classés "Indépendants" ceux qui ne sont ni Républicains, ni Démocrates.

analphabètes, et qui lisent les romans-photos. Cependant, au niveau des sous-populations, ces deux variables sont liées entre elles. Pour calculer les corrélations, il a été nécessaire d'abandonner le niveau des types, niveau auquel les effectifs étaient trop faibles, pour travailler au niveau de 15 macro-types seulement. Pour établir l'équivalence entre deux questions, les auteurs ont exigé que ces deux questions paraissent intuitivement se rapporter à la même attitude latente, et que leurs distributions de réponses soient identiques pour chaque macro-type. Pour résoudre la seconde difficulté, c'est-à-dire définir le point de séquence de l'attitude latente, les auteurs ont construit des échelles de type GUTTMAN en tenant compte du taux de réponses affirmatives à chaque question. Aux groupes d'items (*clusters*) concernant les attitudes ont été ajoutées les informations qui n'avaient pu être conservées comme dimensions de la typologie : l'âge, le niveau d'instruction, et les intentions de vote. On dispose ainsi, pour chaque type d'électeurs, d'informations concernant 52 "*clusters*" ou items. D'autre part, les auteurs conservaient la possibilité d'ajouter à ces données les informations apportées par des sondages effectuées au cours de la campagne.

La procédure de l'élection présidentielle aux Etats-Unis est à deux niveaux. Les différents Etats disposent chacun d'un certain nombre de voix pour l'élection du Président. Il y a en tout 538 voix. Chaque Etat dispose automatiquement de 2 voix ; l'Etat de Washington dispose de 3 voix supplémentaires ; les 435 voix restantes sont réparties entre les Etats proportionnellement au nombre d'habitants. Lors des élections, le dépouillement des votes des électeurs est fait par Etat ; le parti qui obtient la majorité dans un Etat donné dispose de toutes les voix attribuées à l'Etat pour l'élection présidentielle. A titre d'illustration, l'Etat de

New-York dispose de 45 mandats, ceux de Pennsylvanie et de Californie de chacun 32, alors que le Wyoming, le Nevada, le Delaware et le Vermont n'ont chacun que 3 mandats. Il est donc nécessaire de tenir compte de la division par Etats dans l'organisation des données. Etant donné le volume important des interviews utilisés, on dispose en moyenne de 2.000 questionnaires par Etat ; le nombre minimum est compris entre 300 et 400 questionnaires. Compte tenu du fait que la région intervient dans la définition des types, on ne compte au maximum que 108 types différents pour un Etat. A partir des données de recensement, il a été possible de définir la proportion de chaque type d'électeurs par Etat. Pour simplifier la manipulation des données, les auteurs ont décidé de négliger les facteurs politiques locaux, et de faire comme si les différences des votes par Etat provenaient uniquement de différences dans la proportion de chaque type d'électeurs par Etat. Seuls ont été considérés dans la simulation les quarante-huit Etats continentaux.

c) La simulation sur le problème religieux.

Dès le début de la campagne, les Démocrates craignirent les effets du sectarisme religieux des électeurs. Un slogan exprimait clairement ce sectarisme : "pas de catholique à la Maison Blanche" (autre version : "n'élisez pas le Pape à la Maison Blanche"). Dans deux Etats (Wisconsin, et Virginie Occidentale), KENNEDY s'était attaqué à ce préjugé avec succès. Mais brusquement, il y eut une recrudescence des activités anti-catholiques, et les sondages révélèrent que NIXON était en tête. Le problème posé aux chercheurs de la Simulmatics fut alors :

qu'advient-il si le problème religieux devient le thème central de la campagne électorale ?

Les sociologues de la *Simulmatics* avaient annoncé qu'ils refusaient de faire une prévision de l'issue des élections, et qu'ils se limiteraient à évaluer les gains et les pertes de voix qu'entraînerait l'adoption d'une stratégie donnée. La première étape de la résolution du problème posé fut l'élaboration du modèle représentant ce que les chercheurs pensaient être la dynamique des élections. Ils considérèrent que, dans l'hypothèse où le problème religieux serait au centre des préoccupations des électeurs, deux facteurs auraient un rôle prépondérant : les intentions de vote passées, et les attitudes à l'égard de la possibilité qu'il y ait un catholique à la Maison Blanche. Pour évaluer les taux d'abstentions, les auteurs considérèrent que les électeurs soumis à des pressions contradictoires (comme les Démocrates protestants ou les Républicains catholiques) devaient s'abstenir plus fortement que les autres électeurs. Les formules de calcul des proportions de vote en faveur de chacun des candidats ont été établies pour 9 macro-types, obtenus en combinant une variable composite basée sur la race et la religion (protestants blancs, catholiques blancs, autres) et l'affiliation politique des électeurs (Démocrates, Républicains, Indépendants).

Les données servant de base aux calculs varient d'un macro-type à l'autre. Les paramètres utilisés sont :

D_c = pourcentage des intentions de vote Démocrate aux élections pour le Congrès ;

R_c = pourcentage des intentions de vote Républicain aux élections pour le Congrès ;

D_p = pourcentage des intentions de vote pour STEVENSON aux élections présidentielles de 1956 ;

R_p = pourcentage des intentions de vote pour EISENHOWER
aux élections présidentielles de 1956 ;

$\bar{C}Dc$ = estimation du pourcentage d'électeurs anti-catholiques
parmi ceux ayant déclaré une intention de vote Démocrate
aux élections pour le Congrès ;

$\bar{C}Dm$ = estimation du pourcentage moyen d'électeurs anti-
catholiques parmi ceux ayant déclaré une intention de
vote pour STEVENSON ou une intention de vote Démocrate
aux élections pour le Congrès ;

A = taux d'abstentions.

Le principal problème rencontré au cours de l'estimation de ces paramètres concerne le taux d'électeurs anti-catholiques ayant déclaré une intention de voter Démocrate. En effet, les auteurs n'avaient à leur disposition qu'un petit nombre de sondages ayant porté à la fois sur les intentions de vote et l'attitude anti-catholique (*). Il n'était en conséquence pas possible d'estimer, pour chaque macro-type par Etat le taux d'électeurs anti-catholiques ayant déclaré voter Démocrate. La démarche suivie par les sociologues de la *Simulmatics* a été d'estimer le degré de corrélation entre l'intention de vote et l'attitude anti-catholique pour l'ensemble de l'échantillon. Cette corrélation a été supposée constante pour chaque macro-type ; ce qui a permis de calculer un taux d'électeurs anti-catholiques ayant déclaré voter Démocrate.

Le modèle lui-même est assez simple : d'autant plus que les mêmes opérations ont parfois été utilisées pour des macro-types différents (avec évidemment les données propres à chaque macro-type). Pour un Etat donné, le mode de calcul par macro-type des taux de voix

(*) La question-type correspondant à ce "cluster" était : "Si votre parti désignait comme candidat un homme compétent, mais qui se trouverait être catholique, est-ce que vous voteriez pour lui ?".

pour NIXON (V_N) et pour KENNEDY (V_K) est le suivant :

Macro-type	Pourcentage pour Kennedy	Pourcentage pour Nixon
Républicains protestants blancs	$V_K = D_P (1-A)$	$V_N = R_P (1-A)$
Démocrates protestants blancs Indépendants protestants blancs	$V_K = (Dc - \bar{C}Dc) (1-2A)$	$V_N = (Rc + \bar{C}Dc) (1-2A)$
Démocrates catholiques blancs Indépendants catholiques blancs	$V_K = (Dc + \frac{Rc}{3}) (1-A)$	$V_N = \frac{2Rc}{3} (1-A)$
Républicains catholiques blancs Républicains "autres" *	$V_K = (Dc + \frac{Rc}{3}) (1-2A)$	$V_N = \frac{2Rc}{3} (1-3A)$
Démocrates "autres" ** Indépendants "autres" **	$V_K = \left(\frac{Dc + Dp - \bar{C}Dm}{2} \right) (1-2A)$	$V_N = \left(\frac{Rc + Rp}{2} + \bar{C}Dm \right) (1-2A)$
* Cette catégorie "autres" groupe essentiellement les noirs de toutes obédiences religieuses et les juifs.		

Tableau n° 2 - Mode de calcul des taux de voix pour chaque candidat selon les macro-types.

Le score final par Etat était obtenu en faisant la somme des voix obtenues par chaque candidat, l'effectif des voix pour un candidat dans chaque macro-type étant égal au produit du pourcentage de voix pour ce candidat par l'effectif du macro-type dans l'Etat considéré (*). La formulation de ces équations traduit un certain nombre de décisions de la part des auteurs ; elles portent principalement sur le choix des bases de calcul, l'évaluation des glissements de

(*) Un exemple fictif du procédé de calcul est donné par 177, Pp. 28-30.

voix dûs au problème religieux, et l'évaluation des taux d'abstentions dûs aux effets des pressions contradictoires. Il y a trois types différents de bases de calcul ; pour les Républicains protestants blancs, les intentions concernant l'élection présidentielle de 1956 sont considérées comme de bons prédicteurs ; pour les autres macro-types, l'influence personnelle d'EISENHOWER semble avoir joué trop fortement, et les prédicteurs choisis sont les intentions de vote pour le Congrès, sauf pour les Démocrates et les Indépendants noirs et juifs ("autres"), pour qui les auteurs ont préféré prendre la moyenne entre le Congrès de 1958 et les élections présidentielles de 1956. L'évaluation des glissements de voix a été fait de manière assez grossière ; l'apport à KENNEDY des voix des catholiques ne votant pas Démocrate a été évalué au tiers de l'ensemble des catholiques, pondéré par le taux d'abstentions ; les auteurs ont estimé par ailleurs qu'aucun anti-catholique ne voterait pour Kennedy. Pour le calcul des taux d'abstentions, la théorie des pressions contradictoires (*cross-pressures*) a conduit les auteurs à considérer que les Républicains non-protestants s'abstiendraient fortement de voter NIXON s'ils n'acceptaient pas de voter KENNEDY ; et que seuls les Républicains protestants ou les Démocrates ou Indépendants catholiques, n'étant pas soumis à des pressions contradictoires, conserveraient leur taux habituel d'abstention. Les coefficients choisis arbitrairement pour pondérer le taux habituel d'abstention sont 3 pour un fort abstentionnisme, 2 pour un abstentionnisme assez fort, et 1 pour l'abstentionnisme habituel.

Pour diverses raisons politiques (*), la simulation n'a pas été appliquée aux Etats du sud. Elle a permis d'ordonner les 32 Etats

(*)

Dans certains Etats du sud, les organisations locales du parti Démocrate risquaient de se trouver en désaccord avec les instances nationales du parti ; ainsi, voter Démocrate pouvait ne pas signifier nécessairement voter Kennedy.

étudiés, de ceux où la stratégie proposée est bénéfique pour KENNEDY, à ceux où elle lui est néfaste. Le conseil donné à KENNEDY a été d'attaquer de front le sectarisme religieux. Cette stratégie devait avoir pour effet de lui faire perdre des votes au niveau global, mais lui permettre de l'emporter dans des Etats disposant d'un nombre élevé de voix pour l'élection du Président, et ainsi de gagner des voix (cf. § 3.2.2.e). Bien que les auteurs n'aient pas cherché à réaliser une prévision de l'issue du scrutin, ils ont confronté après les élections les résultats de celles-ci et ceux de leur simulation. Ils ont constaté que les résultats de la simulation avaient une corrélation (r de BRAVAIS-PEARSON) de 0,82 avec les élections, alors que les données de sondage ayant servi pour la simulation n'avaient avec les élections qu'une corrélation de 0,53. Le modèle semble donc avoir représenté assez correctement l'effet des changements politiques survenus entre 1958 et 1960.

d) La mesure de la sensibilité du modèle.

Il est possible qu'un modèle complexe ait permis une prévision globale exacte, tout en étant faux dans le détail. En possession des résultats des élections, les auteurs se sont efforcés de déterminer les éléments du modèle qui ont contribué à son efficacité, et ceux qui au contraire l'ont compromise. Pour cela, ils ont recommencé leur simulation en faisant varier leurs hypothèses, c'est-à-dire la forme des équations du modèle. Les résultats de cette série de recherches portent sur l'estimation correcte des paramètres, et l'influence des variables qui ont été négligées au cours de la simulation de 1960.

Parmi les paramètres d'entrée du modèle, les uns étaient des données empiriques, et les autres des coefficients arbitraires

intervenant dans les équations. Une première série de simulations *a posteriori* a porté sur les données empiriques. En utilisant les résultats des sondages qui ont immédiatement précédé les élections, on améliore encore la prévision. Cet effet semble dû surtout à la taille de l'échantillon ainsi obtenu (130.000 interviews, provenant de 65 sondages), qui confère aux données de sondages plus de stabilité et de précision ; si l'on utilise les derniers sondages seuls, ou n'importe quel sous-ensemble des sondages utilisables, la prévision est moins bonne. Une autre série de simulations a porté sur les variations des coefficients arbitraires. On a constaté qu'en ne tenant pas compte du facteur religieux, on diminue fortement la corrélation de la simulation avec les élections. Si l'on fait varier l'apport à KENNEDY des Républicains catholiques de 10 % à 100 %, la corrélation varie peu, mais la dispersion est minimum pour un apport de 70 %. Si l'on fait varier simultanément les taux des défections dues à une attitude anti-catholique, et des ralliements de Républicains catholiques, la corrélation obtenue reste inférieure à celle de la simulation de 1960 (cf. figures 6 et 7, page 186). Le meilleur ajustement du coefficient multiplicatif du taux d'abstentions consiste à l'égaliser à 1, c'est-à-dire à le négliger. La théorie selon laquelle l'abstention est un effet des pressions contradictoires n'est donc pas vérifiée. Il semble que le comportement d'abstention soit différent selon que les électeurs sont placés devant un conflit "approche-approche" (Républicains catholiques), ou "évitement-évitement" (Démocrates anti-catholiques).

Au nombre des variables qui ont été négligées dans la simulation réalisée pour KENNEDY figurent les votes antérieurs, et les attitudes à l'égard de la politique étrangère et des droits civiques. Les prévisions fondées sur l'hypothèse d'une continuité des attitudes à l'égard des partis ne donnent pas d'aussi bons résultats que la

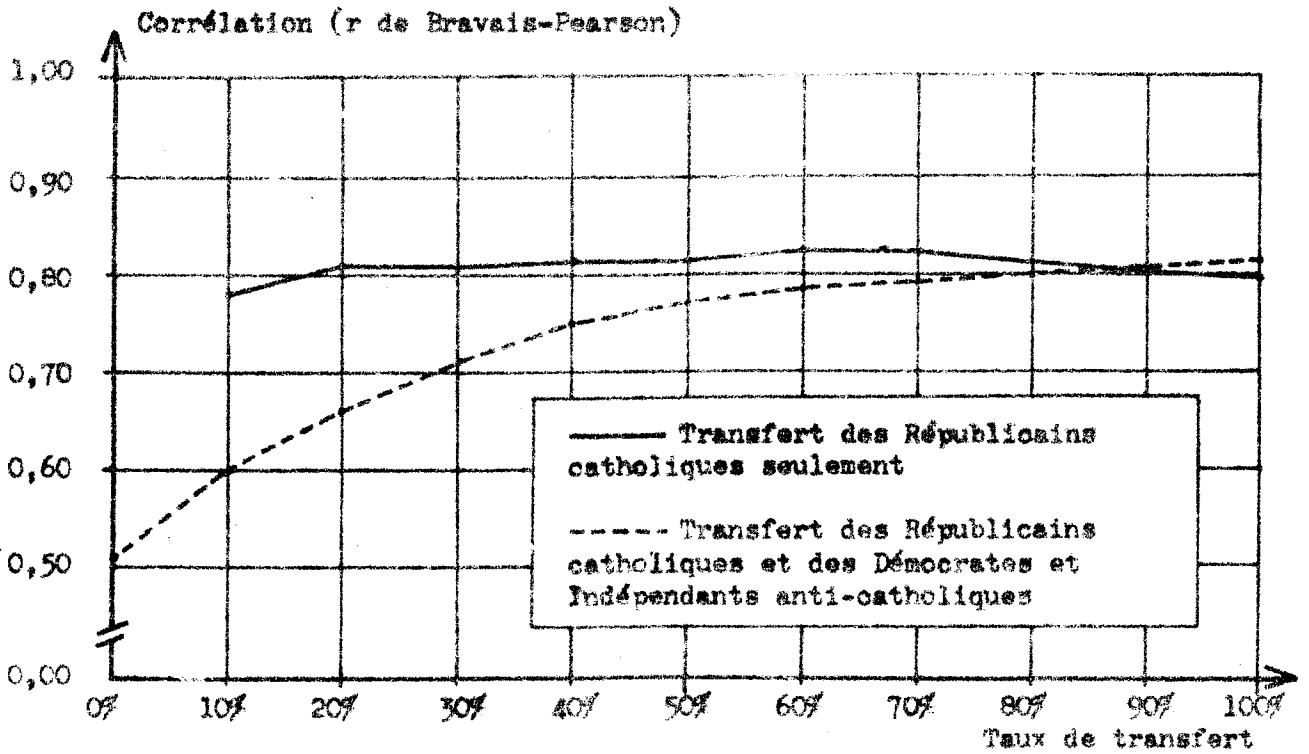


Fig. 6 Influence du taux de transfert de voix dû aux attitudes religieuses sur la corrélation des résultats de la simulation avec le scrutin par Etat.

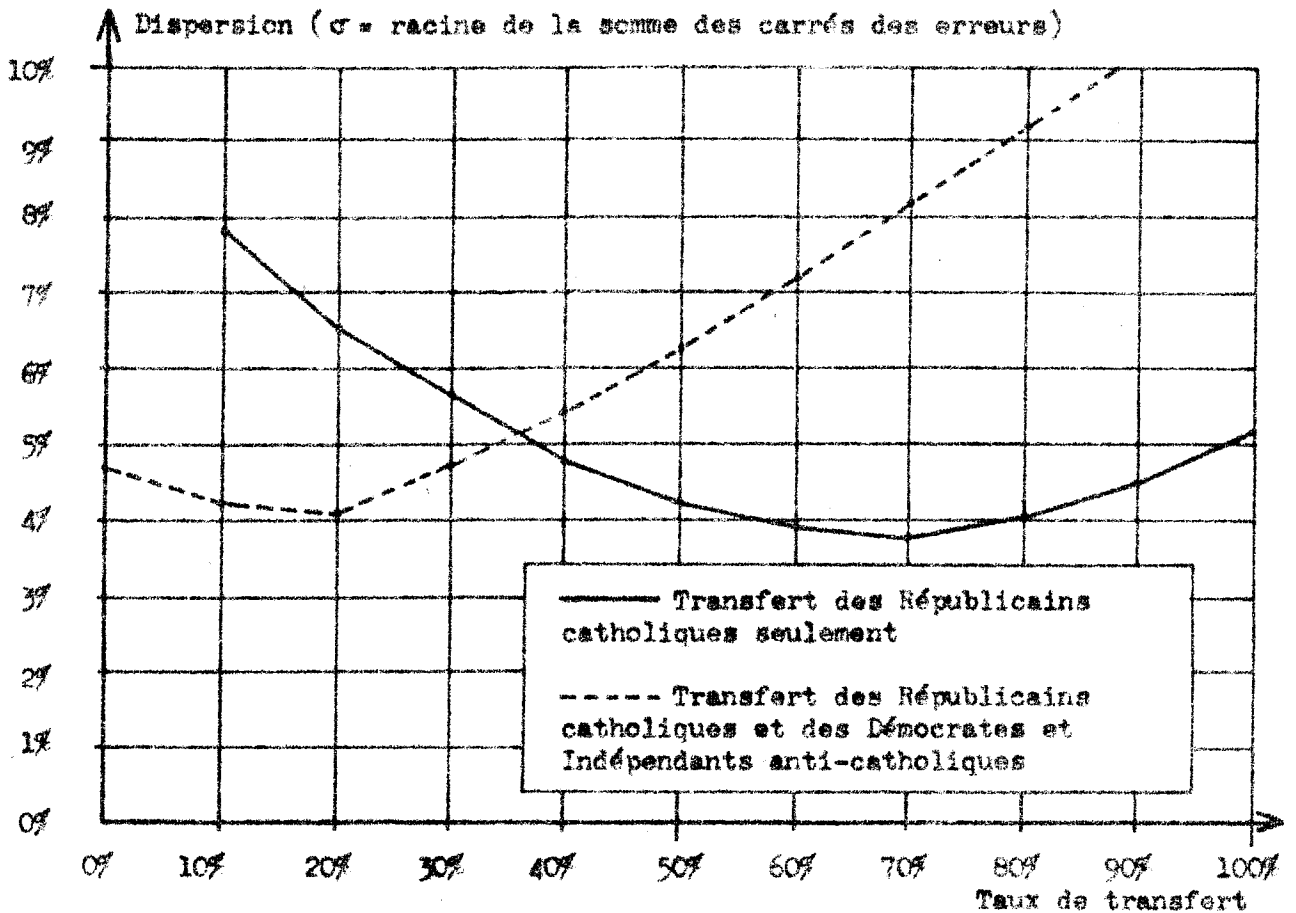


Fig. 7 Influence du taux de transfert de voix dû aux attitudes religieuses sur la dispersion des écarts entre les résultats de la simulation et le scrutin par Etat.

simulation. D'ailleurs, pour les 32 Etats ayant fait l'objet de la simulation, les corrélations (r de BRAVAIS-PEARSON) sont de 0,24 entre les sondages de 1956 et les votes de 1960 ; 0,43 entre les votes de 1956 et ceux de 1960 ; 0,37 entre les votes de 1958 et ceux de 1960 ; et enfin de 0,53 entre l'ensemble des votes pour le Congrès de 1952 à 1958, et les votes de 1960. Cependant, on remarque que l'on parvient à une prévision satisfaisante (corrélation avec les élections : 0,85) si l'on ajoute dans chaque Etat, au nombre de votes en faveur de STEVENSON, un apport de 35 % des voix catholiques. Il faut peut être voir dans ce résultat plus qu'une coïncidence numérique. On peut faire l'hypothèse que la dynamique des élections de 1960 présente certaines analogies avec celle des élections de 1956 ; et qu'en particulier les Démocrates qui en 1956 ont subi l'ascendant de la personnalité d'EISENHOWER ont les mêmes caractéristiques sociales que ceux qui, en 1960, ont refusé de voter pour un catholique.

Deux simulations ont été réalisées dans l'hypothèse où le thème central de la campagne aurait été la politique étrangère, ou le problème des droits civiques. Dans le premier cas, si les candidats avaient développé leur point de vue sur la négociation avec l'U.R.S.S., ou l'aide aux insurrections dans les démocraties populaires, que ce serait-il passé ? En 1960, l'opinion était préoccupée du problème des relations avec les pays socialistes ; et si une nouvelle crise à Berlin, ou un accord sur l'interdiction des essais nucléaires, étaient survenus, l'opinion aurait été fortement divisée. L'équation de base choisie pour cette simulation est :

$$V_K = \frac{Dp + Dc}{2} + \frac{Neg}{3} - \frac{\overline{Neg}}{3}$$

où :

Neg = pourcentage d'électeurs partisans de la négociation avec l'U.R.S.S.

$\overline{\text{Neg}}$ = pourcentage d'électeurs hostiles à une négociation avec l'U.R.S.S.

V_K , D_p , et D_c ont le même sens qu'au § 3.2.2.c.

Les résultats de cette simulation n'ont avec les résultats des élections qu'une corrélation de 0,43, ce qui montre l'importance de ce facteur. NIXON serait élu, avec 54 % des suffrages. Cette hypothèse est confirmée par les résultats d'un sondage téléphonique réalisé pendant la campagne, où NIXON apparaissait comme "l'homme le plus capable de traiter avec les Russes" (43 % contre 24 % à KENNEDY), en raison de son passé aux affaires étrangères et de l'inexpérience de KENNEDY. Dans l'hypothèse d'une campagne centrée sur le problème de l'égalité des droits civiques, les glissements de voix n'auraient concerné qu'une minorité d'électeurs, principalement les noirs et les ségrégationnistes. Mais dans 8 Etats du nord, un léger apport de voix de la part des noirs aurait pu être décisif. Les résultats de la simulation sur les droits civiques n'ont qu'une corrélation de 0,47 avec les élections. Pourtant, l'attitude de KENNEDY envers Martin Luther KING au cours de la campagne lui a valu un apport non négligeable de suffrages parmi les noirs.

Un problème proche de celui de la mesure de la sensibilité du modèle a été soulevé par les auteurs : c'est celui de la possibilité de simplifier le modèle utilisé pour la simulation de 1960. Ainsi que nous l'avons noté au chapitre précédent (§ 2.2.4.c), moins un modèle est complet, plus il nécessite de données empiriques, et réciproquement. C'est parce que les auteurs ne disposaient pas d'une théorie simple pour prédire les attitudes de chacun des types d'électeurs à l'égard de chaque item qu'ils ont dû utiliser 72.000

variables empiriques (480 types \times 50 items ou "clusters" d'attitudes \times 3 modalités de réponse). Est-il possible, rétrospectivement, de réaliser un modèle plus élaboré, requérant moins de données ? BERELSON, LAZARSELD et GAUDET ont construit un indice de la prédisposition à voter Démocrate ; chaque modalité des indicateurs est affectée d'un poids (ex. : catholique = 2 ; urbain = 1 ; etc.), et l'indice est obtenu par la sommation de ces poids. Pour qu'un tel indice soit valide, il faut qu'il n'y ait pas d'interactions entre les variables servant d'indicateurs. Les sociologues de la *Simulmatics* ont réalisé une série d'analyses de la variance portant sur 8 "clusters" d'attitudes. Pour chaque attitude, les facteurs étudiés étaient la religion, l'affiliation politique, la région, le statut socio-économique et le sexe. L'analyse a été conduite pour 336 types pour lesquels les données étaient facilement utilisables. Pour les items où le parti politique joue un rôle (image du parti, attitudes à l'égard des hommes politiques, etc.), il y a une interaction systématique entre l'affiliation politique et les autres facteurs ; encore ces interactions sont-elles de faible amplitude comparées aux effets principaux. Pour les autres "clusters" (politique étrangère, droits civiques, anti-catholicisme, etc.), il n'y a pas d'interaction. En pratique donc, il y a *additivité des facteurs* liés aux prédispositions, et il devrait être possible d'utiliser cette propriété pour simplifier les modèles ultérieurs. Ainsi, s'il y a parmi les Démocrates 45 % d'électeurs favorables à STEVENSON de plus qu'en moyenne, et s'il y a parmi les habitants des villes 1 % d'électeurs favorables à STEVENSON de plus qu'en moyenne, alors le taux d'électeurs Démocrates habitant en ville et favorables à STEVENSON doit être de $45 \% + 1 \% = 46 \%$ de plus que la moyenne.

e) Les apports du projet Simulmatics à la sociologie politique.

Indépendamment de son intérêt pratique, pour les choix stratégiques du parti Démocrate au cours de la campagne présidentielle de 1960, le projet *Simulmatics* a fait progresser la compréhension des mécanismes de la vie politique américaine. Tout d'abord, une analyse secondaire des données préparées pour la simulation a permis de retracer les grands courants politiques de 1950 à 1960. Cette analyse a porté sur les 336 types pour lesquels on disposait des données les plus complètes. Chaque type était affecté du même poids, et non plus d'un poids proportionnel à son effectif. Pour chacun des cinq "clusters" étudiés, deux indices étaient pris en considération : le degré d'implication de la question (taux de réponses favorables aux Démocrates + taux de réponses favorables aux Républicains), et le degré d'attraction de la position Démocrate (taux de réponses favorables aux Démocrates - taux de réponses favorables aux Républicains). L'étude des séries chronologiques pour chacun de ces indices a révélé une profonde stabilité dans les attitudes politiques des Américains : si, en 1952, EISENHOWER a attiré les suffrages de certains électeurs Démocrates et Indépendants en raison de son programme de paix, en 1956 on observe un retour aux normes des partis. Cependant, dans les Etats du sud, la tendance politique à voter Républicain dans les milieux urbains pauvres et Démocrate dans les milieux ruraux aisés est en train de se renverser ; le sud tend donc à s'aligner sur le reste du pays.

La simulation de 1960 aide à comprendre la dynamique des élections. Pour parfaire cette recherche, les auteurs ont réalisé

le meilleur modèle possible, en ajustant les valeurs des paramètres du premier modèle, et en introduisant de nouvelles variables. (cf. figure 8, page 192). Dans ce nouveau modèle, les auteurs ont supprimé la correction du taux d'abstention ; utilisé les données des sondages réalisés pendant la campagne ; évalué à 40 % l'apport de voix à KENNEDY des catholiques non-Démocrates ; introduit, dans les Etats du sud, un glissement de 10 % des noirs non-Démocrates en faveur de KENNEDY, et de 10 % des blancs Démocrates en faveur de NIXON. La validation du modèle a été réalisée non seulement d'après les résultats des élections, qui ne sont pas ventilés par les caractéristiques des électeurs, mais surtout d'après un sondage GALLUP effectué à la fin de 1960. Les principaux enseignements de la simulation de ce modèle concernent ;

- le *rôle des habitudes politiques*. Le facteur personnel du candidat à la Présidence est nettement plus important que celui du candidat aux élections pour le Congrès. Cependant, l'affiliation politique reste un facteur de base pour les élections présidentielles : ce facteur reste le meilleur prédicteur des votes pour l'élection de 1960.

- les *attitudes anti-catholiques*. La proportion de Démocrates ou d'Indépendants non-catholiques ayant refusé de voter pour KENNEDY doit être d'environ 20 %.

- le *soutien des catholiques*. On peut évaluer à environ 1/3 la proportion de catholiques qui, ayant voté Républicain en 1958 (élections au Congrès), ont voté pour KENNEDY. Le bilan de l'effet du facteur religieux peut être estimé ainsi (si le modèle est correct) :

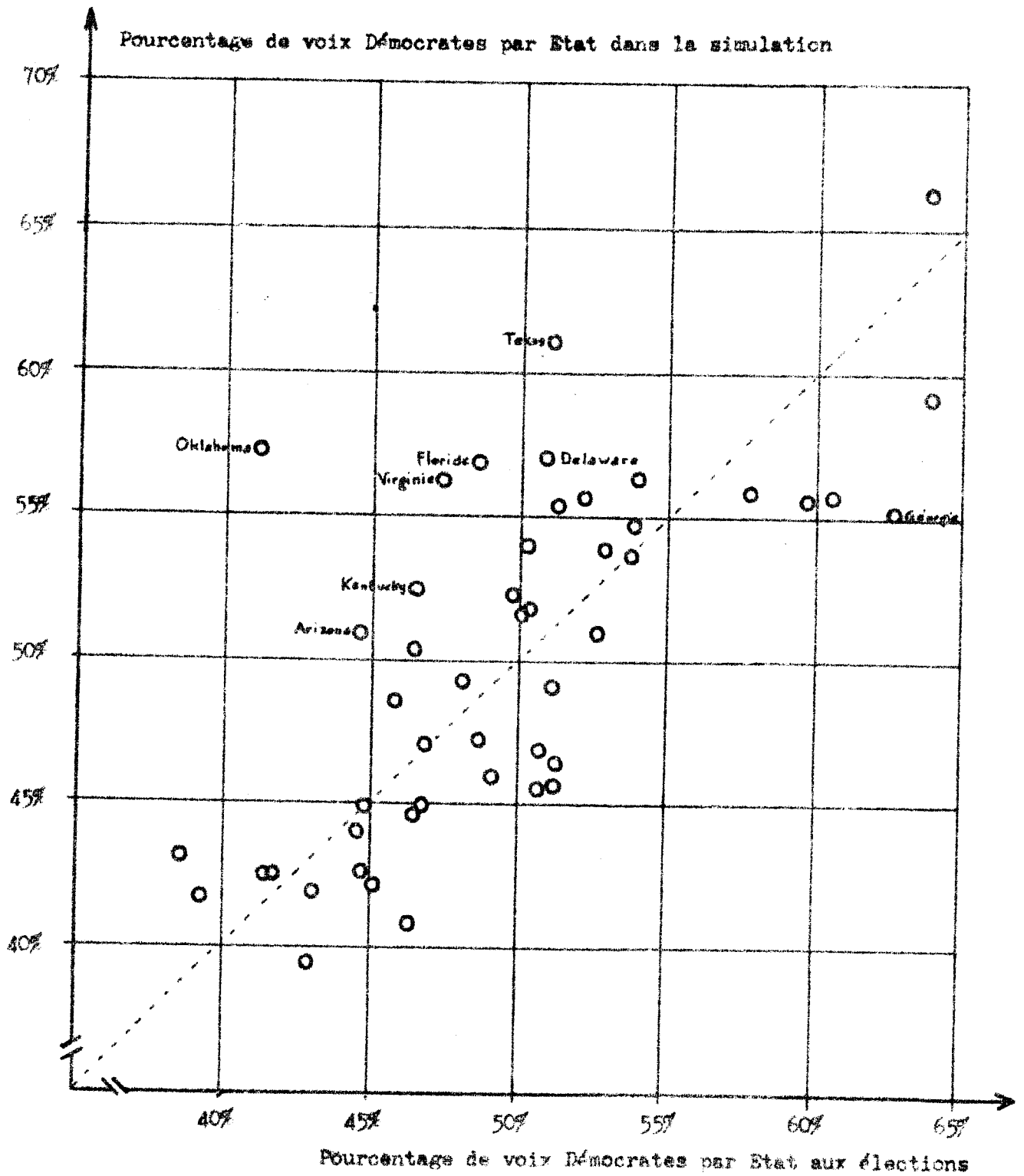


Fig. 8 Comparaison des résultats du modèle le mieux ajusté et du scrutin par Etat.

	Gains	Pertes	Bilan
<u>En voix</u> (électeurs)	2.800.000	4.300.000	-1.500.000
soit en pourcentages	4,1 %	6,3 %	-2,2 %
<u>En mandats</u> (états)	122	132	+ 10

Tableau n° 3 : Bilan de l'effet du facteur religieux.

- l'influence des *attitudes à l'égard des droits civiques*. Ce facteur a joué un rôle important dans les Etats du sud.
- l'influence des *facteurs locaux*, qui expliquent les variations résiduelles.

f) L'application du modèle de 1960 aux élections de 1964.

En 1964, les sociologues de la *Simulmatics* ont voulu rééditer leur expérience de 1960 avec cette fois-ci un objectif plus ambitieux : prédire les résultats du scrutin. Mais ils n'ont bénéficié d'aucune aide financière, et n'ont pu en conséquence améliorer leur modèle comme ils le souhaitaient. Ils ont donc utilisé les données de 1960 auxquelles ils ont ajouté les réponses à 20 sondages réalisés depuis, et ils ont adopté le modèle utilisé en 1960 en tenant compte des caractères propres à la campagne de 1964. Selon les auteurs, les élections de 1964 sont de celles qui risquent de remodeler la physionomie politique des Etats-Unis. En effet, dans les élections précédentes, la stratégie des candidats de chacun des deux partis consistait à convaincre le "Marais" de

leur modération. La candidature GOLDWATER a bouleversé cette tradition en proposant un choix politique aux électeurs : en conséquence, des millions d'électeurs ont rompu avec leur parti traditionnel, en refusant l'option GOLDWATER. Si ce phénomène était prévisible, il était difficile à chiffrer. Les auteurs ont successivement utilisé 3 modèles, le modèle de base faisant appel à 5 facteurs, les autres modèles reposant respectivement sur 6 et 7 facteurs.

Les facteurs du modèle de base sont l'affiliation politique, les habitudes de vote, et les attitudes à l'égard des droits civiques, de la sécurité sociale, et de la politique atomique. La logique du modèle repose sur les propositions suivantes :

- tous les Démocrates votent pour JOHNSON, excepté ceux qui sont fortement opposés aux droits civiques.

- tous les Républicains votent pour GOLDWATER, excepté ceux qui sont d'accord avec les Démocrates sur au moins deux des trois thèmes retenus (droits civiques, sécurité sociale, politique atomique).

- tous les Indépendants qui ont une attitude nette sur le problème des droits civiques votent en conséquence ; les autres votent selon leur attitude sur les deux autres thèmes.

Les équations permettant d'évaluer les suffrages en faveur de JOHNSON (V_J) sont les suivantes :

- pour les types d'électeurs *Démocrates* :

$$V_J = H (1 - \overline{DC} \cdot DC^+)$$

- pour les types d'électeurs *Républicains* :

$$V_J = H (DC \cdot CN + DC \cdot SS(1 - CN) + CN \cdot SS (1 - DC))$$

- pour les types d'électeurs *indépendants* :

$$V_J = H \left(\frac{X}{X+Y} \right)$$

avec : $X = DC + (1-DC-\overline{DC}) CN.SS + (1-DC-\overline{DC}) (1-CN-\overline{CN}).SS$
 $+ (1-DC-\overline{DC}) CN (1-SS-\overline{SS})$

$Y = \overline{DC} + (1-DC-\overline{DC}) \overline{CN}.SS$

où :

H = pourcentage d'électeurs ayant des habitudes de vote établies,

DC = taux d'électeurs favorables aux droits civiques,

\overline{DC} = taux d'électeurs hostiles aux droits civiques,

DC^+ = taux d'électeurs fortement concernés par les droits civiques,

CN = taux d'électeurs favorables au contrôle nucléaire,

\overline{CN} = taux d'électeurs hostiles au contrôle nucléaire,

SS = taux d'électeurs favorables à la sécurité sociale,

\overline{SS} = taux d'électeurs hostiles à la sécurité sociale.

Remarquons que ce modèle postule l'indépendance des attitudes envers les trois thèmes de la campagne à l'intérieur d'un type d'électeurs. Il faut ajouter également que certains pourcentages de réponses brutes sur les thèmes de la campagne ont dû être pondérés. En effet, en 1960 il était possible de considérer les valeurs brutes même si elles affichaient un biais systématique, puisque l'objectif était d'ordonner les Etats, et qu'un biais systématique n'affectait pas l'ordre obtenu. Pour la prévision par contre il a été nécessaire de réajuster certaines valeurs : on a considéré que 10 % des adversaires de la sécurité sociale avaient évolué en faveur de cette institution, et qu'un tiers des adversaires du contrôle nucléaire avaient été ébranlés par le traité d'octobre 1963 sur l'interdiction des essais nucléaires.

Ces ajustements ont fixé la distribution moyenne des attitudes sur chaque cluster à la valeur moyenne de 60 % en faveur de la thèse Démocrate. La simulation a prédit 62 % des voix pour JOHNSON, qui en a obtenu en fait 61,4 %. Les prévisions par Etat sont moins satisfaisantes : la corrélation avec les résultats n'est que de 0,52 avec une erreur médiane de 6,1 % (*). La faiblesse du modèle de base est due au fait qu'il ne tient pas compte des particularismes locaux. A partir de l'estimation de l'erreur faite en 1960, les auteurs ont corrigé de 10 % les votes des plaines de l'Ouest (modèle à 6 facteurs), obtenant une corrélation de 0,63 et une erreur médiane de 3,6 %. Après les élections, les auteurs ont amélioré le modèle à 6 facteurs en tenant compte de la décision des organisations locales du Mississippi et de l'Alabama de voter contre JOHNSON (modèle à 7 facteurs) ; la corrélation avec les élections s'est élevée à 0,90, avec une erreur médiane de 3,4 %. Les différences relevées par Etat entre les prévisions du modèle à 7 facteurs et les résultats des élections s'expliquent par le facteur "enfant du pays" pour l'Arizona (GOLDWATER) et le Texas (JOHNSON), par l'effet de majorité (*bandwagon effect*) contre les droits civiques pour la Californie, et par la sous-estimation des voix Démocrates dans les quatre Etats de la Nouvelle Angleterre (Rhode Island, Massachusetts, Vermont, Maine).

(*)

Dans le cas d'une prévision des résultats des élections, le coefficient de corrélation n'est pas un indicateur suffisant de la valeur de la simulation, car il ne tient compte que des rapports entre les intervalles des scores prédits, et les intervalles des scores réels. Si $y = ax + b$, la corrélation entre x et y est de 1, quels que soient a et b . Ainsi, une simulation qui ferait une erreur systématique correspondant exactement à 0,8 fois le score réel - 5 % des voix, aurait une corrélation de 1, avec une erreur variant de 13,2 % pour les scores les plus faibles à 21,2 % pour les plus élevés (en se basant sur les résultats par Etat des élections de 1964).

La dynamique des élections de 1964 est claire : GOLDWATER, en obligeant les électeurs à s'exprimer sur un programme politique, a perdu les partisans des droits civiques, du contrôle nucléaire, et de la sécurité sociale ; en contrepartie, il n'a gagné que les ségrégationnistes. Alors qu'en 1960, on pouvait prédire les comportements électoraux à partir de l'affiliation politique et du facteur religieux, en 1964 ce sont les thèmes politiques qui ont été déterminants. En 1960, 10 % environ des électeurs ont voté contre leur parti traditionnel pour des raisons religieuses ; en 1964, environ 20 % ont fait de même pour des raisons politiques (et parmi ceux-ci plus du tiers des Républicains, qui ont voté pour JOHNSON).

DEUXIEME PARTIE :

APPLICATION DE LA SIMULATION

A L'ETUDE EXPERIMENTALE DE

L'HOMOGAMIE

4 - L'ELABORATION DU MODELE

Pour recenser et illustrer les problèmes méthodologiques ou techniques que le recours à la simulation pose au sociologue, nous avons voulu réaliser une expérience de simulation assez simple pour avoir une valeur pédagogique, et assez complète pour nous permettre de traiter tous les problèmes que le sociologue rencontre effectivement. Nous avons dû prendre deux séries de décisions : l'une à propos des *moyens* techniques que nous envisagions d'utiliser, l'autre concernant le *thème* sociologique et les données qui devaient servir de base à l'expérience.

Les décisions techniques ont été prises assez aisément. Il nous a paru naturel de recourir à la simulation *sur ordinateur numérique*, car cette technique est actuellement de très loin la plus utilisée dans les sciences sociales ; et ni l'ordinateur analogique, ni *a fortiori* les machines ou dispositifs conçus spécialement pour une expérience ne semblent devoir le supplanter dans l'immédiat, en raison des difficultés pratiques que soulève leur mise en oeuvre. De plus, l'ordinateur numérique est, comme nous l'avons souligné, une machine universelle, dont les utilisations ne se limitent pas à la simulation, et qui est appelée à jouer le rôle d'auxiliaire général de la pensée ; il est possible, et en tout cas souhaitable, que la pratique de l'ordinateur soit aussi familière pour les prochaines générations de sociologues que l'est, pour le citoyen des pays

industriels, l'utilisation domestique de l'énergie électrique. Enfin, nous avons écarté les modèles purement dialectiques parce que difficiles à utiliser dans le cas de phénomènes un peu complexes : dans de tels modèles en effet, l'homme ne s'efforce-t-il pas de simuler le fonctionnement de l'ordinateur, alors qu'il ne dispose ni de la même capacité de mémoire en accès direct, ni de la même vitesse d'exécution ? D'autre part, nous souhaitons mettre en oeuvre un modèle *stochastique*, afin de pouvoir aborder le problème du tirage au hasard. En effet, la solution de ce problème est quasi-immédiate si l'on utilise une machine conçue pour l'expérience (dispositif physique analogue à une roulette ; cf. 191, Pp. 67-71) ou un ordinateur analogique (générateur de bruits ; cf. op.cit, Pp. 66-67) ; sur un ordinateur numérique par contre elle soulève quelques difficultés (cf. § 5.1.2.).

Le choix du thème sociologique a été plus délicat. En analysant, dans le chapitre précédent, les apports de la simulation aux sciences humaines, nous avons pu constater la nette différence existant entre les modèles micro-sociologiques et les modèles macro-sociologiques (cf. 99). Les premiers sont en général *complets*, c'est-à-dire (§ 2.2.4.c.) que l'ensemble des variables impliquées dans la simulation et de leurs interactions est contenu dans le modèle lui-même, qui peut même assez souvent être formulé de manière analytique. En cela, ils sont assez proches d'ailleurs des modèles de simulation élaborés par les psychologues ou les psychophysiologistes. Les modèles macro-sociologiques par contre n'expriment qu'une partie de la réalité que l'on se propose de simuler, et il est nécessaire de compenser les lacunes de la formalisation en introduisant, au cours de la simulation, un volume de données empiriques d'autant plus grand que le modèle est plus incomplet. Sur le plan technique, les simulations micro-sociologiques semblent relativement

faciles à réaliser, à partir de l'énoncé du modèle ; dans de nombreux cas, il serait possible de les traiter "à la main". Elles présentent par conséquent un intérêt pédagogique moindre que les simulations macro-sociologiques. Ces dernières font généralement appel à une quantité de données tellement importante qu'il n'est pas possible d'envisager leur réalisation autrement que sur ordinateur. Du point de vue méthodologique, il semble que de nombreux cas de simulation en micro-sociologie tiennent lieu d'expériences de laboratoire ; leur intérêt est donc moins grand que s'il s'agissait, comme en macro-sociologie de phénomènes *pour lesquels aucune expérimentation n'est possible*. D'un point de vue proprement sociologique enfin, les simulations de phénomènes micro-sociologiques nous paraissent rester encore en deçà du domaine propre de la sociologie, puisqu'actuellement aucune simulation n'a eu pour objet d'intégrer divers modèles micro-sociologiques pour déboucher au niveau des faits sociaux ; il est certain cependant que cette direction est celle dans laquelle les sociologues devront engager leurs recherches dès que les moyens techniques (en particulier la capacité de mémoire des ordinateurs) le permettront. En attendant, les simulations macro-sociologiques nous semblent les seules susceptibles d'appréhender réellement les phénomènes sociaux. (Nous développerons ce point dans nos conclusions, § 7.3.c).

Compte-tenu de l'ensemble des considérations qui précèdent, nous avons recherché un thème relatif à un phénomène macro-sociologique facilement isolable, assez bien formalisé, et pour lequel une expérimentation artificielle pourrait présenter un intérêt théorique. Pour des raisons pratiques, nous souhaitons en outre que les données empiriques concernant ce thème portent sur un échantillon

de population représentatif de l'ensemble, mais de dimensions moyennes, et qu'il nous soit possible de disposer de ces données sous une forme utilisable sur ordinateur. Ainsi que nous l'avons expliqué dans l'introduction, ces multiples conditions se sont trouvées remplies par l'étude d'Alain GIRARD (*) sur *le choix du conjoint* (93).

4.1. LE PROBLEME DU CHOIX DU CONJOINT.

4.1.1. L'étude réalisée par Alain GIRARD

Cette étude, publiée en 1964 dans les cahiers de *Travaux et Documents* de l'Institut National d'Etudes Démographiques, sous le titre *Le choix du conjoint*, est trop connue pour qu'il soit utile de l'exposer ici. Nous nous bornerons à rappeler les éléments de cette étude dont la connaissance est nécessaire pour comprendre la position du problème choisi pour servir d'objet de la simulation.

a) Objectifs de l'étude.

L'étude réalisée par A. GIRARD a pour objectifs la description et l'analyse de certains aspects sociologiques relatifs au choix du conjoint dans la société française contemporaine. Nous savons bien que, dans notre société, le mariage est non seulement un évènement important pour la vie biologique et la vie sociale de l'individu, mais aussi une institution sur laquelle l'influence de

(*)

Nous renouvelons nos remerciements au Professeur Alain GIRARD, qui a bien voulu mettre à notre disposition l'ensemble des documents concernant cette étude.

la société se fait particulièrement sentir. Le souci de l'auteur est d'éprouver la thèse selon laquelle le rite du mariage aurait été profondément remodelé sous l'influence d'une part de l'accroissement démographique, d'autre part du progrès technique et de ses corollaires, le développement des moyens de communication, la concentration urbaine, l'augmentation de la mobilité géographique. Une telle hypothèse conduirait à penser que l'univers de choix du conjoint (*) tend progressivement à se confondre avec la population globale toute entière, de sorte que peu à peu l'influence des barrières sociales et de l'autorité des parents disparaîtrait pour faire place au libre choix. Ainsi, dans la société contemporaine, on pourrait observer un effacement progressif des idéaux propres à chaque groupe et des valeurs sociales, dont l'influence limite l'initiative individuelle dans le choix du partenaire. A la limite, et sans tenir compte des facteurs psychologiques d'attrait mutuel, il n'y aurait donc plus à proprement parler de "choix du conjoint", mais simplement appariement selon les lois du hasard, c'est-à-dire panmixie.

Pour évaluer le degré d'exactitude d'une telle hypothèse, A. GIRARD a recouru à une enquête psychosociologique menée auprès d'un échantillon représentatif de l'ensemble des couples actuellement constitués. Cette enquête a permis non seulement d'évaluer la *distance* (sociale, géographique, culturelle, religieuse) entre les conjoints, mais aussi d'analyser les *circonstances* ayant précédé le mariage, et de dégager le *modèle collectif* auquel les conjoints tendaient à se conformer.

(*) C'est-à-dire l'ensemble des conjoints "possibles", compte-tenu de l'éloignement géographique et de la distance sociale, au sein duquel s'effectue le choix du conjoint.

b) L'enquête.

L'enquête a été réalisée entre juin et novembre 1959. L'unité d'observation était le couple. 1646 couples ont répondu au questionnaire ; la personne qui a répondu est tantôt la femme (826 fois), tantôt le mari (820 fois). La définition des couples à interroger obéissait aux conditions suivantes :

- les deux conjoints sont vivants au moment de l'enquête ;
- ils vivent ensemble ;
- le mari n'a pas plus de 65 ans, la femme pas plus de 62 ans ;
- avant leur mariage, ils étaient tous deux célibataires.

Ces conditions ont eu pour effet d'éliminer de l'enquête environ 30 % des personnes mariées : les veufs, les divorcés ou séparés, les remariés, et les couples qui se sont mariés avant 1914. La sélection des couples à interroger, parmi les couples correspondant aux conditions énoncées, a été effectuée selon la méthode des *quotas*, à partir de ceux critères : habitat (région x catégorie d'agglomération) et catégorie socio-professionnelle (profession du mari).

Le questionnaire comportait cinq séries de questions :

- renseignements concernant *le mari* (11 questions) : nationalité, date et lieu de naissance, position dans la famille, scolarité et diplômes, religion, localités de résidence successives, profession avant le mariage et profession actuelle ; nationalité, région d'origine et profession du père et de la mère.

- renseignements concernant *la femme* : mêmes questions que pour le mari, plus s'il y a lieu la date et les raisons de cessation du travail.

- *historique des relations* entre les futurs époux (9 questions) : première rencontre, fréquentation, attitude des familles, fiançailles.

- *circonstances du mariage* et situation actuelle du ménage (8 questions) : date et lieu du mariage, mariage religieux, contrat, dot ; nombre d'enfants, standing.

- *opinions et attitudes* à l'égard du mariage (13 questions) : importance dans le choix du conjoint du milieu social et des facteurs individuels ; influence de l'autorité parentale ; tolérance sexuelle avant le mariage ; âge optimum du mariage ; travail de la femme ; nombre d'enfants souhaités.

L'exploitation des résultats de l'enquête a été réalisée mécanographiquement. Le report sur cartes de l'ensemble des réponses a nécessité deux cartes par questionnaire. Le type de codage des réponses est lié à l'exploitation sur trieuse : utilisation des zones hors-texte, création d'indices ou combinaison de réponses avant la perforation.

c) Principaux résultats.

L'enquête a montré que le degré d'*homogamie* (*) des couples est assez élevé. Ainsi, la moitié environ des futurs conjoints habitaient la même commune, et près des trois quarts une commune d'importance comparable ; près des trois quarts sont nés dans la même région (en comptant 24 régions) ; 45 % appartiennent au même milieu social ;

(*) *Homogamie* : ressemblance entre les conjoints ; *hétérogamie* : dissemblance. Bien que ces notions soient des notions très générales, en pratique, les critères d'homogamie ou d'hétérogamie utilisés le plus souvent sont les caractéristiques socio-démographiques des conjoints.

66 % ont le même niveau d'instruction, malgré la disparité des niveaux selon le sexe ; 2 % seulement se réclament de religions différentes. On voit par conséquent que les effets de la mobilité sociale, comme de la mobilité géographique, sont relativement faibles.

L'analyse des circonstances de la première rencontre apporte un élément d'explication : bien que ces circonstances soient très variées, elles mettent le plus souvent en présence des individus de situations ou de rangs sociaux très proches. La période de fréquentation qui précède le mariage dure en moyenne deux ans. Prévenue en moyenne au bout d'un an, la famille se montre favorable dans 8 cas sur 10 ; la raison d'opposition invoquée le plus souvent est la différence des milieux sociaux. Les traditions sont très suivies (bague de fiançaille dans 70 % des cas ; mariage religieux dans 90 % des cas) ; par contre, le contrat de mariage et la dot sont rares (respectivement 20 % et 29 %).

Les *modèles collectifs* du mariage sont assez homogènes. Il y a en particulier accord sur l'âge idéal au mariage (homme : 23-25 ans ; femme : 20-22 ans) ; sur l'importance pour les futurs époux d'être du même milieu social ; sur la taille supposée de l'univers de choix. Quand on relève des différences sensibles, elles mettent en évidence une évolution des couples âgés aux plus jeunes, des milieux traditionnels aux milieux plus favorables à l'émancipation de la femme : cette évolution se manifeste surtout à propos de la liberté sexuelle de la jeune fille, de la limitation des naissances, et du travail féminin.

En résumé, on peut déceler une lente évolution de la société

vers une relative hétérogamie. Mais actuellement, l'homogamie géographique reste forte, en dépit de la mobilité géographique croissante, car l'évolution vers l'hétérogamie se trouve freinée par la tendance des émigrants à se réunir entre eux (reconstitution de la "petite patrie"). De même, l'homogamie socio-culturelle est élevée, surtout pour les catégories situées aux extrêmes de la hiérarchie sociale ; les normes sociales freinent l'hétérogamie sociale et culturelle en restreignant l'univers de choix (institutionnalisation des circonstances de la rencontre ; nécessité de l'approbation de la famille). Ces normes collectives non seulement régissent la constitution du couple, mais tracent la voie à la famille future. La famille apparaît ainsi "comme l'institution sociale qui contrôle et distribue les chances d'élévation sociale, freine les mouvements de régression, et permet aux différents milieux sociaux de se perpétuer d'une génération à l'autre" (93, p. 200). En conclusion, "la mobilité croissante tend à alléger les contraintes, et les choix peuvent être d'autant plus individualisés que ces contraintes sont moins fortes. Mais si les mariages ne sont plus "arrangés" dans la société contemporaine, ils continuent à subir toutes sortes de pressions extérieures" (ibid., p. 201).

4.1.2. Les dimensions de l'homogamie.

La notion d'*homogamie*, empruntée à la génétique des populations, est au centre de l'étude d'A. GIRARD. L'homogamie, et son contraire l'hétérogamie, sont les degrés extrêmes d'une échelle permettant de mesurer la ressemblance ou la dissemblance entre les conjoints. Cette ressemblance peut traduire les effets d'un comportement orienté vers la recherche d'un partenaire semblable (compor-

tement endogamique) ou au contraire dissemblable (comportement exogamique). Lorsque le comportement d'appariement semble dû au hasard, on parlera de *panmixie*. Il nous a paru intéressant d'approfondir ces diverses notions, et de définir les relations qui les lient.

a) *Panmixie et distance sociale.*

L'hypothèse de panmixie correspond au cas où il n'existe aucune tendance à restreindre l'univers de choix du conjoint : "tout le monde peut épouser tout le monde", et les couples se forment selon les lois du hasard. Selon le généticien BINDER, "une population est panmictique lorsque chacun des individus qui la composent a des chances égales de se reproduire avec n'importe quel autre individu de sexe opposé" (28, p. 20). Comme nous l'avons vu, l'un des objectifs de l'étude d'Alain GIRARD est d'évaluer dans quelle mesure la population française contemporaine s'écarte de l'hypothèse de panmixie. Si l'on compare, pour l'ensemble des couples étudiés, une caractéristique du mari à la caractéristique correspondante de la femme, la réalité se situera probablement entre les deux cas extrêmes suivants : celui où les caractéristiques des conjoints sont toujours identiques (*homoqamie absolue*), et celui où elles sont toujours différentes (*hétéroqamie absolue*). Nous illustrerons ces extrêmes par un exemple schématique. Supposons que l'on définisse l'origine sociale des personnes interrogées à partir de la profession du père, et que cette variable ne prenne que trois états : agriculteurs, ouvriers, cols blancs. Si notre échantillon renferme 200 couples, et si, pour chacun des conjoints, nous avons les mêmes distributions marginales (70 fils et 70 filles d'agriculteurs, 60 fils et 60 filles d'ouvriers, 70 fils et 70 filles

de cols blancs), nous pouvons construire les tableaux croisés d'effectifs correspondant aux hypothèses d'homogamie ou d'hétérogamie absolues selon l'origine sociale. Dans le tableau représentant l'hypothèse d'homogamie absolue (tableau n° 4), toutes les cases seront vides, à l'exception de la diagonale ; dans celui correspondant à l'hétérogamie absolue (tableau n° 5), seules seront vides les cases de la diagonale, les autres cases renfermant des effectifs quelconques compatibles avec les effectifs marginaux (on peut donc concevoir plusieurs tableaux représentant l'hétérogamie).

Profession du père du mari	Profession du père de la femme			Total
	Agriculteur	Ouvrier	Col blanc	
Agriculteur	70	-	-	70
Ouvrier	-	60	-	60
Col blanc	-	-	70	70
Total	70	60	70	200

Tableau n° 4 Homogamie absolue selon l'origine sociale.

Profession du père du mari	Profession du père de la femme			Total
	Agriculteur	Ouvrier	Col blanc	
Agriculteur	-	30	40	70
Ouvrier	30	-	30	60
Col blanc	40	30	-	70
Total	70	60	70	200

Tableau n° 5 - Exemple d'hétérogamie absolue selon l'origine sociale.

Pour représenter de la même manière la répartition au hasard d'une caractéristique chez les conjoints, nous calculerons pour chaque case un effectif proportionnel aux effectifs marginaux correspondant à cette case (*). Nous admettrons en effet que, dans

(*) L'effectif théorique a'_{ij} correspondant à la case a_{ij} du tableau sera :

$$a'_{ij} = \frac{\sum_1 a_{.j} \cdot \sum_j a_{i.}}{\sum_1 \sum_j a_{..}}$$

l'hypothèse d'une répartition aléatoire de l'origine sociale (par exemple), le nombre de filles d'agriculteurs épousant un fils de col blanc est proportionnel à l'effectif des fils de cols blancs dans l'ensemble de la population. Ce qui correspond au tableau croisé suivant :

Profession du père du mari	Profession du père de la femme			Total
	Agriculteur	Ouvrier	Col blanc	
Agriculteur	24,5	21	24,5	70
Ouvrier	21	18	21	60
Col blanc	24,5	21	24,5	70
Total	70	60	70	200

Tableau n° 6 - Répartition au hasard selon l'origine sociale.

Si nous observons une répartition de ce type pour une seule des variables étudiées, nous pourrions seulement en déduire que cette caractéristique n'intervient pas dans le processus de choix du conjoint, et qu'elle ne semble pas liée à une variable servant de critère de choix. Dans l'hypothèse où nous observerions une telle répartition aléatoire pour l'ensemble des variables étudiées, si nous sommes certain de ne pas avoir omis une variable pouvant jouer un rôle important dans le processus de choix du conjoint, nous pourrions affirmer que notre population est une population *panmictique*. Un comportement panmictique se traduit donc par une indépendance totale par rapport à l'*ensemble des caractéristiques*, qu'elles soient géographiques, chronologiques, ou socio-culturelles.

Pour mesurer à propos de l'origine sociale la différence entre la population française contemporaine et une population panmictique, A. GIRARD a construit un indice simple qui fait intervenir deux hypothèses implicites : la première est que l'on peut ordonner les professions en tenant compte de la "hiérarchie sociale", les ouvriers agricoles occupant le bas de la hiérarchie, et les cadres supérieurs et professions libérales en occupant le sommet ; la seconde hypothèse est que les intervalles séparant deux catégories professionnelles contiguës quelconques sont égaux. Il devient alors possible de mesurer d'une manière simple la *distance sociale* entre les conjoints, et de recenser les effectifs des couples caractérisés par une distance donnée. On peut également calculer les effectifs théoriques de couples pour chaque degré de distance sociale, dans l'hypothèse d'une population panmictique. A titre d'illustration, les effectifs théoriques correspondant aux tableaux que nous avons constitués ci-dessus seraient :

Distance sociale	Homogamie absolue	Panmixie	Hétérogamie absolue (exemple)
0	200 = 100 %	67 = 33,5 %	0 = 0 %
1	0 = 0 %	84 = 42 %	120 = 60 %
2	0 = 0 %	49 = 24,5 %	80 = 40 %

Tableau n° 7 : Variations de la distance sociale selon les hypothèses considérées.

La comparaison de la distribution théorique des distances sociales dans l'hypothèse de panmixie et des effectifs réellement observés dans l'enquête sur le choix du conjoint (93, Pp. 74-75), montre que l'homogamie est deux fois plus forte que celle qui serait due au hasard (45,1 % au lieu de 21,4 %). La différence entre les deux distributions est d'ailleurs hautement significative (*).

En suivant un raisonnement de même nature pour les principales variables pouvant exprimer une distance entre les conjoints, A. GIRARD a élaboré un indice de similitude globale. Cet indice (93, Pp. 93-94) est construit à partir de douze caractéristiques

(*) Le chi-carré calculé sur les effectifs ne serait pas inférieur à $\chi^2 = 462$ pour $v = 12$, ce qui est significatif à $P < 0,001$.

géographiques et socio-culturelles ; pour chaque caractéristique, on attribue à chaque couple une "note" comprise entre 1 (distance maximum) et 7 (similitude totale). Cet indice global ayant été construit *a priori*, pour des raisons techniques inhérentes au mode de dépouillement de l'enquête (*), il n'était pas possible de calculer les coefficients de pondération et éventuellement de correction, qui auraient tenu compte de l'importance relative de chaque caractéristique dans le phénomène de distance entre les conjoints, et de l'effet des interactions entre ces caractéristiques. L'indice global, qui est en réalité un indice de similitude plutôt que de distance, est la somme équipondérée des "notes" attribuées pour chaque caractéristique ; il varie par conséquent de 12 (distance maximum) à 84 (similitude totale). La valeur correspondant à la panmixie est la moyenne théorique de l'indice : 48.

L'examen de la distribution des valeurs prises par cet indice (93, Pp. 87-92) confirme le degré élevé d'homogamie de la population française. La moyenne et la médiane observées sont de 69,7, et le mode de 73. Seuls 1,2 % des couples étudiés ont un indice inférieur ou égal à 48 (panmixie). La comparaison de l'indice global moyen selon la profession du mari montre que la distance entre les conjoints est d'autant plus grande que l'on s'élève plus dans

(*) Le dépouillement de l'enquête a été réalisé sur trieuse. Les procédés de traitement mécanographique n'auraient pas permis de recourir, d'une manière techniquement simple, à un indice construit *a posteriori*, au vu des premiers résultats ; à plus forte raison, ils n'auraient pas permis de modifier le mode de calcul de cet indice au gré des chercheurs (ce qui peut être réalisé très simplement sur ordinateur). Il était donc nécessaire de définir la forme de l'indice et de calculer sa valeur avant d'avoir procédé au report sur cartes mécanographiques des réponses au questionnaire, afin de pouvoir faire figurer cette valeur sur les cartes au même titre que les autres variables.

la hiérarchie sociale (indice moyen de similitude variant de 74,9 pour les cultivateurs à 65,6 pour les cadres supérieurs et professions libérales). Si l'on rapproche ce résultat de l'observation (citée au paragraphe précédent) selon laquelle l'homogamie socio-culturelle est plus élevée pour les catégories extrêmes de la hiérarchie sociale, on peut faire l'hypothèse que, dans les catégories de rang social élevé, la mobilité géographique joue un rôle important dans le choix du conjoint.

b) Comportements endogamiques et exogamiques.

En génétique des populations, si l'on utilise l'hypothèse de panmixie comme une hypothèse commode pour les calculs, on n'ignore pas que de nombreux facteurs interviennent pour fausser cette hypothèse : fécondité différentielle, consanguinité, mutations, migrations, effectif et limites de la population, comportement de choix du partenaire (189, p. 320). Le choix du conjoint peut prendre deux formes antagonistes : la recherche chez le partenaire de caractéristiques identiques aux siennes propres (choix positif), ou au contraire la recherche d'un partenaire ayant des caractéristiques différentes, accompagnée du refus de partenaire identique (choix négatif). Le choix positif, dans lequel les individus de sexe opposé porteurs des mêmes caractéristiques se recherchent, est le mieux connu et le plus facilement observable, puisqu'il se traduit par un certain degré d'homogamie. Dans certaines espèces animales, on a relevé un appariement préférentiel entre individus de même couleur (couleur du manteau chez les mollusques ; couleur du pelage chez un rongeur) (28, Pp. 85-87). Chez l'homme, les facteurs d'homogamie les mieux connus des généticiens (189, p. 318 ; 190, p. 54) sont l'âge, la taille, et les aptitudes intellectuelles (mesurées par le Q.I.) ; certaines

anomalies comme la surdi-mutité ou les névroses, et certaines caractéristiques purement sociales comme le veuvage, jouent également un rôle. Le généticien SUTTER fait d'ailleurs remarquer que cet appariement phénotypique est fortement contrôlé par des considérations sociales, comme les normes implicites sur la taille des époux (crainte de constituer un "couple à la Dubout") (189, p. 319). Les exemples de choix négatifs sont plus rares ou moins bien connus ; SUTTER (ibid.) cite le cas des personnes à cheveux roux, ou de caractère violent, qui recherchaient des partenaires de caractéristiques différentes. Dans tous les cas, le comportement de choix systématique des individus a pour effet d'éloigner la distribution des caractéristiques des couples de la distribution théorique correspondant à la panmixie : il y a une forte ressemblance (tendance homogamique) dans le cas du choix positif, et plus forte dissemblance (tendance hétérogamique) dans le cas du choix négatif.

Si nous quittons le niveau descriptif pour considérer les facteurs explicatifs du choix, nous distinguerons avec les généticiens le *génotype*, qui est l'ensemble des caractères héréditaires propres à l'espèce, du *phénotype*, ensemble des caractères visibles chez l'individu, qui constituent les manifestations observables du génotype. Lorsque l'on étudie non plus un caractère phénotypique pris isolément, mais un ensemble de caractères que l'on considère comme des indicateurs du génotype, on aborde le problème de l'homogamie sous l'angle du problème de l'*hybridation interspécifique*, c'est-à-dire du croisement des espèces. En transposant ces notions dans le langage de la sociologie, nous dirons que lorsque nous considérons un ensemble de caractéristiques comme des indicateurs de l'appartenance à un groupe social, nous abordons le problème de l'homogamie non plus sous l'angle du choix du conjoint, mais sous celui du brassage social. Si nous

examinons par exemple comment se font les mariages chez les habitants des banlieues industrielles, nous verrons que ces habitants se marient beaucoup plus fréquemment à l'intérieur de cette population qu'à l'extérieur ; nous en concluons que cette population a un degré élevé d'*endogamie*. Lorsque le degré d'*endogamie* est maximum, la population considérée constitue un *isolat*. Dans le cas opposé, s'il arrive que les membres d'une population contractent autant que possible des unions à l'extérieur, soit par une sorte d'extension de la prohibition de l'inceste, soit sous l'influence de migrations ou de dispersions, nous dirons que cette population est fortement *exogamique*.

Si l'on considère un instant le problème de l'*endogamie* du point de vue du choix individuel, tout se passerait comme si, de prime abord, un individu appartenant à une population donnée évitait tout contact à implications matrimoniales avec tout individu de sexe opposé n'appartenant pas à la même population. Si un comportement *endogamique* était systématique et valable pour l'ensemble des individus de la population, il aboutirait à un isolement complet, et la population constituerait un *isolat*. En milieu humain, une telle situation est très rare, et suppose sinon un isolement physique total des populations, du moins un esprit de caste ou un racisme très fort ; dans la plupart des cas, et quel que soit le critère de définition des populations, la réalité observée se situe entre la panmixie et l'*endogamie*. Ainsi, il nous faut modifier notre hypothèse sur le comportement des membres d'une population à tendance *endogamique*, en supposant ou bien que seuls certains individus ont un comportement *endogamique* systématique, ou bien que chaque individu a une tendance *endogamique* de degré variable d'un individu à l'autre.

Si, à l'intérieur d'une population globale, on définit

diverses sous-populations à partir par exemple de l'origine sociale des individus, on comprend aisément qu'un comportement fortement endogamique au sein de chaque sous-population se traduira dans l'ensemble de la population par un degré élevé d'homogamie selon l'origine sociale. On peut également supposer qu'un tel comportement se traduira également par un degré élevé d'homogamie pour les caractéristiques fortement liées au critère de définition des sous-populations, c'est-à-dire l'origine sociale. Cette hypothèse est conforme aux relations supposées entre le génotype qui définit l'espèce, et les caractères phénotypiques, qui permettent de la décrire. Par exemple, SUTTER fait remarquer qu'aux Etats-Unis, où les conjoints sont habituellement de la même origine sociale, de la même religion, du même niveau socio-économique, on observe un degré d'homogamie élevé pour des traits biologiques tels que la taille, la couleur des cheveux et des yeux, et pour l'intelligence. On comprend en effet que "si les juifs se marient entre eux, comme ils ont plus fréquemment les yeux bruns et les cheveux foncés, il y a naturellement une corrélation entre les époux" (189, p. 319). On tiendrait un raisonnement analogue à propos des rapports entre l'hétérogamie et le comportement exogamique.

c) Endogamie et panmixie.

La notion de panmixie, telle que les généticiens la définissent, ne semble pas pouvoir s'appliquer sans modification à une population humaine. Deux contraintes en particulier font obstacle, dans les sociétés modernes, à une véritable panmixie ; une contrainte spatiale, et une contrainte temporelle. *Contrainte spatiale* car, malgré les progrès des moyens de communication, les probabilités de rencontre entre individus géographiquement éloignés sont sensiblement moindres qu'entre individus proches. Ce fait se traduit non seule-

ment sur le plan purement spatial (distance géographique, ou différences dans la région habitée avant le mariage), mais également sur le plan de la taille de l'agglomération de résidence. Ainsi, dans l'enquête d'Alain GIRARD, moins de 25 % des couples sont formés d'individus dont les communes de résidence n'étaient pas de même importance ; et dans 8,8 % des couples seulement, les futurs conjoints habitaient dans deux régions différentes. "La possibilité du choix implique la rencontre. La liberté de l'individu est limitée d'abord par l'étendue physique du milieu où il vit et se meut" (93, p. 19). C'est cette constatation qui a conduit les généticiens à définir la notion d'*isolat*, c'est-à-dire "la population à l'intérieur de laquelle un individu a la possibilité de trouver son conjoint" (190, p. 55). La *contrainte temporelle* est de même nature : la proximité chronologique joue aussi puissamment que la proximité spatiale. Dans l'échantillon ayant servi de base à l'étude sur le choix du conjoint, plus de la moitié des femmes avaient entre 20 et 24 ans au moment de leur mariage et moins de 10 % avaient plus de 27 ans ; de même, plus de la moitié des hommes avaient entre 23 et 27 ans, et moins de 10 % dépassaient la trentaine (*) A partir de ces chiffres, on voit que le degré d'homogamie selon l'âge doit être élevé dans la population française.

Doit-on, en raison de ces contraintes, abandonner totalement l'hypothèse de panmixie ? Une telle hypothèse est commode, pour permettre l'évaluation de l'intensité d'un comportement endogamique ou exogamique. Mais si cette hypothèse est complètement irréaliste, son

(*) Rappelons d'une part qu'il s'agit de mariages entre célibataires, et d'autre part que ces proportions ne peuvent donner une idée exacte du phénomène, puisqu'elles sont affectées par la mortalité.

intérêt, même théorique, s'en trouve évidemment réduit. Si le sociologue désire évaluer la force d'un comportement endogamique déterminé par des facteurs sociaux, il devra mesurer la différence entre le comportement observé et un comportement sur lequel les facteurs sociaux sont sans influence. C'est-à-dire qu'il devra *faire abstraction de l'endogamie due aux facteurs géographiques ou à l'âge*, et partir d'une hypothèse théorique de *panmixie restreinte*, en quelque sorte à âge au mariage et facteur géographique constants.

Ce raisonnement nous ramène à la notion d'*isolat*, en étendant cependant légèrement la signification que lui donnent les généticiens. Si nous parvenons à définir diverses sous-populations, homogènes du point de vue des facteurs géographiques et de l'âge des individus qui la composent, nous aurons constitué des *isolats artificiels*, au sein desquels nous pourrions faire l'hypothèse de la panmixie. A titre d'exemple, nous reprendrions les données fictives que nous avons utilisées au § 4.1.2.a, en supposant pour simplifier que la population est homogène en ce qui concerne l'âge et la distance géographique, et que le comportement endogamique se manifeste seulement à propos de la taille des agglomérations. Les effectifs des hommes selon leur origine sociale et la taille de leur agglomération de résidence sont les suivants :

Profession du père du mari	Catégorie d'agglomération		Total
	Ville	Campagne	
Agriculteur	10	60	70
Ouvrier	40	20	60
Col blanc	50	20	70
Total	100	100	200

Tableau n° 8 : Répartition des hommes selon l'origine sociale et l'habitat.

On admettra que la répartition des femmes selon les mêmes critères est identique à celle des hommes. Si nous supposons que les deux catégories d'agglomération constituent des isolats, à l'intérieur desquels il y a panmixie, nous avons pour chaque catégorie d'agglomération les répartitions suivantes :

Profession du père du mari	Profession du père de la femme			Total
	Agriculteur	Ouvrier	Col blanc	
Agriculteur	1	4	5	10
Ouvrier	4	16	20	40
Col blanc	5	20	25	50
Total	10	40	50	100

Tableau n° 9a : Panmixie dans les villes.

Profession du père du mari	Profession du père de la femme			Total
	Agriculteur	Ouvrier	Col blanc	
Agriculteur	36	12	12	60
Ouvrier	12	4	4	20
Col blanc	12	4	4	20
Total	60	20	20	100

Tableau n° 9b : Panmixie dans les campagnes.

En additionnant les tableaux 9a et 9b, nous obtenons le tableau théorique correspondant à l'hypothèse de panmixie restreinte sur notre population fictive :

Profession du père du mari	Profession du père de la femme			Total
	Agriculteur	Ouvrier	Col blanc	
Agriculteur	37	16	17	70
Ouvrier	16	20	24	60
Col blanc	17	24	29	70
Total	70	60	70	200

Tableau n° 10 : Panmixie restreinte.

Il est intéressant de comparer ce tableau à celui correspondant à l'hypothèse de panmixie absolue. On constate en particulier que l'hypothèse de panmixie restreinte telle que nous l'avons définie pourrait donner l'impression au niveau de l'ensemble de la population, de l'existence d'une tendance à l'homogamie sociale. Cette impression est d'ailleurs confirmée au niveau de l'indice de distance sociale :

Distance	Homogamie absolue	Panmixie restreinte	Panmixie absolue
0	200 = 100 %	86 = 43 %	67 = 33,5 %
1	0 = 0 %	80 = 40 %	84 = 42 %
2	0 = 0 %	34 = 17 %	49 = 24,5 %

Tableau n° 11 : Variations de la distance sociale selon les hypothèses considérées.

Or, il est clair que cette apparente homogamie est en réalité un *artefact*, du à l'endogamie géographique.

Il peut sembler trop schématique de réduire l'univers de choix de plusieurs individus à une sorte de *pool*, dans lequel chacun devrait trouver son partenaire. Les contraintes d'âge et les contraintes géographiques jouent en fait d'une manière beaucoup plus souple, et d'autre part il n'est pas légitime de "mettre en commun" les univers de choix de plusieurs individus, alors qu'il est peu probable qu'ils se confondent dans la réalité. Il conviendrait donc, pour construire un modèle de panmixie restreinte un peu plus réaliste, de passer du "tout ou rien" à une variable continue, et de substituer à la notion d'isolat celle de *vicinité*, qui exprime un degré variable de proximité entre les individus. Le raisonnement que nous avons tenu à partir de la notion d'isolat peut en gros s'appliquer à la notion de *vicinité* ; le seul problème théorique serait la définition d'un indice global de *vicinité*, tenant compte à la fois de la différence d'âge et des disparités géographiques.

4.1.3. Quelques directions de recherche possibles.

L'intérêt de la simulation pour le sociologue est, rappelons-le, de lui permettre de réaliser des expériences artificielles sur des phénomènes qu'il ne lui est pas possible de manipuler dans la réalité. A partir des données de l'étude d'Alain GIRARD sur le choix du conjoint, nous avons d'abord envisagé trois directions principales de recherche :

- réaliser l'hypothèse de panmixie restreinte ;
- étudier l'influence sur le brassage social de lois endogamiques ou exogamiques.
- étudier les mécanismes de choix du conjoint.

a) Réalisation de l'hypothèse de panmixie restreinte.

Nous avons montré sur un exemple simple l'intérêt qu'il peut y avoir, pour une étude sociologique, de maintenir constants certains facteurs comme les caractéristiques géographiques ou l'âge. Maintenir constants ces facteurs, c'est réaliser ce que nous avons appelé l'*hypothèse de panmixie restreinte*. Mais la réalisation concrète de cette hypothèse soulève de nombreux problèmes. Dans le cas où l'on se limite à la forme la plus fruste de panmixie restreinte, on suppose qu'il y a panmixie absolue à l'intérieur d'*isolats* déterminés en fonction de l'âge, de la distance géographique, et de l'âge. La dimension de ces isolats peut être définie de nombreuses manières ; et, pour une dimension donnée des isolats, leur composition peut être déterminée différemment, selon que l'on accorde plus d'importance à la partition de l'ensemble des individus selon l'âge, selon la région, ou selon la catégorie d'agglomérations. Si l'on introduit la notion de

vicinité, on retrouve les mêmes problèmes sous une autre forme : quelle dimension donner à l'univers de choix du conjoint, et comment calculer la distance spatio-temporelle séparant deux individus ?

Les réponses à cet ensemble de questions pourraient être fournies empiriquement par une série de simulations au cours de laquelle on réaliserait l'hypothèse de panmixie restreinte pour diverses valeurs des paramètres dont on désire mesurer l'influence. Ces paramètres sont :

- le type de modèle choisi (isolat ou *vicinité*) ;
- la taille moyenne de l'univers de choix ;
- les variations admises de la taille de l'univers de choix autour de cette valeur moyenne (dispersion) ;
- les poids respectifs attribués à l'âge et aux facteurs géographiques pour la composition des isolats ou le calcul de la *vicinité*.

b) *Etude du brassage social*.

Le *brassage social* est la résultante de deux types de forces : les unes qui ont pour effet d'augmenter les probabilités de contact entre deux individus quelconques de sexe opposé, les autres qui limitent les possibilités de mariage au sein de l'ensemble des individus de sexe opposé avec lesquels un individu donné est en contact. Parmi les forces du premier type, nous citerons le développement des moyens de transport et de communication, l'accroissement des migrations professionnelles, l'augmentation de la tolérance à l'égard des rencontres entre jeunes gens et jeunes filles. Les forces

du second type appartiennent pour la plupart à l'ensemble des facteurs qui tendent à perpétuer les groupes sociaux et à maintenir leur cohésion au cours du temps (facteurs endogamiques).

Dans une expérience de simulation, il serait possible de prendre comme paramètres l'intensité de ces différentes forces, afin d'étudier sur la population décrite par A. GIRARD l'effet des variations d'intensité. L'expérience consisterait à simuler le mariage des 3.292 individus recensés, autant de fois qu'il y a de combinaisons des valeurs choisies pour les paramètres, et de relever les variations de certains indices descriptifs, comme l'indice de distance globale moyen, ou le taux d'homogamie au niveau de l'ensemble de la population.

c) Etude du mécanisme de choix.

L'étude du brassage social se situe au niveau global de la population, et repose sur les comportements endogamiques ou exogamiques moyens, et non sur les particularités des comportements individuels. Dans l'enquête sur le choix du conjoint, de nombreuses informations ont été recueillies sur les croyances et les attitudes de la personne interviewée à l'égard du mariage. Il serait par conséquent possible de construire un modèle général de décision, qui permettrait, pour un ensemble de croyances et d'attitudes donné, de définir les caractéristiques de l'univers de choix, ou même le profil exact du conjoint recherché (qui n'est pas nécessairement le conjoint réel). En agrégeant dans une expérience de simulation les comportements individuels de choix ainsi définis, on obtiendrait un ensemble de couples dont les caractéristiques globales pourraient être comparées

à celles de la population de l'enquête. S'il y a désaccord, la simulation pourrait être recommencée après modification du modèle de choix, jusqu'à ce que l'on aboutisse à un ajustement satisfaisant du modèle. Cette série de simulations permettrait à la fois de mieux comprendre les mécanismes du choix du conjoint, et de voir ce qui se passerait si certaines attitudes venaient à être modifiées.

4.2. UN MODELE POUR L'ETUDE EXPERIMENTALE DE L'HOMOGAMIE.

Le choix de notre direction de recherche s'est porté sur le brassage social. Nous avons écarté *a priori* l'étude du mécanisme de choix du conjoint, en raison de la trop grande complexité du modèle qu'il faudrait envisager. En effet, à notre connaissance, aucune simulation n'a encore été réalisée qui réussisse à intégrer un grand nombre de modèles de décision individuels pour l'étude d'un phénomène macro-sociologique. La difficulté est à la fois théorique et pratique : théorique, car il n'existe pas encore, en l'état actuel des recherches sociologiques, de schémas explicatifs permettant le passage du niveau individuel au niveau proprement sociologique ; pratique, car les limites de la capacité de mémoire centrale des ordinateurs, plus encore que celles de la rapidité de calcul et d'accès à la mémoire centrale, sont un obstacle à la simulation de modèles très complexes. Aussi les expériences de simulation doivent-elles tenir compte du dilemme suivant : ou bien atteindre une complexité satisfaisante au niveau individuel, mais restreindre son étude aux petits groupes, ou bien se placer d'emblée au niveau des lois sociologiques, mais accepter en contre-partie de considérer les individus comme des entités assez schématiques (ce qui est d'ailleurs également le cas dans

les modèles macro-économiques). Nous avons d'autre part estimé que l'intérêt du problème de la panmixie restreinte était trop réduit pour justifier à lui seul une expérience de simulation ; nous pouvons d'ailleurs considérer ce problème comme un aspect particulier de celui du brassage social.

Le problème du brassage social nous a paru pouvoir présenter un certain intérêt pour la recherche sociologique, être assez complexe pour que les principaux problèmes méthodologiques y soient abordés, et être assez simple cependant pour ne pas demander un travail de programmation trop lourd. Afin de simplifier encore le problème, nous avons mis l'accent sur l'étude des variations du taux d'homogamie pour l'ensemble de la population.

4.2.1. Position du problème .

L'objectif de la recherche que nous avons entreprise pour décrire les étapes de la mise en oeuvre des techniques de simulation, est de déterminer *dans quelle mesure les caractéristiques des couples sont affectées par les comportements à tendance endogamique ou exogamique* des individus qui doivent les constituer. Pour cela, il nous faudra tout d'abord dissocier les couples recensés au cours de l'étude, pour constituer deux populations disjointes (les hommes et les femmes) que nous apparierons selon des règles édictées et contrôlées par nous au cours de l'expérience ; nous ferons ensuite varier globalement la nature et l'intensité des attitudes endogamiques ou exogamiques qui déterminent les comportements de choix, et nous relèverons les variations correspondantes du taux d'homogamie au niveau de l'ensemble de la population. Nous pourrons ainsi étudier expérimentalement le bras-

sage social à partir d'une population réelle, et même éventuellement comparer les résultats des simulations avec les données de l'enquête.

Nous allons examiner quelles sont les caractéristiques des individus ou des couples qui constituent les paramètres de la simulation, comment ces caractéristiques interviennent dans le processus de choix du conjoint, et selon quelles règles il peut être opportun de les manipuler pour atteindre l'objectif fixé.

a) Facteurs de l'endogamie.

On peut supposer que les facteurs qui déterminent le comportement endogamique que l'on observe dans la population française, sinon coïncident, du moins sont liés avec les variables pour lesquelles on relève un degré élevé d'homogamie. D'après l'étude d'Alain GIRARD, ces variables sont la nationalité (97,2 % d'homogamie), la religion (92 %), la région de résidence (91,2 %), la taille de la commune de résidence (75,5 %), la région de naissance (72,2 %), le niveau d'instruction (66 %), la taille de la commune de naissance (52 %), et la profession du père (45,1 %). Nous pouvons *a priori* classer ces variables en deux groupes : celles qui influent sur les probabilités de contact (caractéristiques de résidence), et celles qui ont une probabilité d'orienter le choix. Les premières, auxquelles il convient d'ajouter l'âge, constituent des contraintes physiques plutôt que des facteurs d'endogamie. Toutes les autres variables peuvent être considérées comme des paramètres du comportement d'endogamie ou d'exogamie ; nous écarterons cependant la nationalité, qui joue peu, puisqu'il s'agit d'une population résidant en France (l'échantillon ne comprend que 0,5 % de couples entièrement étrangers).

Les variables que nous avons isolées comme étant liées au comportement d'endogamie peuvent difficilement être considérées *a priori* comme indépendantes les unes des autres. Nous avons d'ailleurs vérifié, pour quatre d'entre elles, que leurs liaisons étaient toutes significatives, tant chez les hommes que chez les femmes (cf. § 6.1.1.b). Cependant, pour la commodité de la recherche, nous pourrions agir sélectivement sur l'un de ces facteurs, en réalisant par exemple un comportement endogamique relatif au seul niveau d'instruction, afin d'étudier les variations des taux d'homogamie spécifiques de chacun des autres facteurs. Nous pourrions ainsi constater l'effet, sur le taux d'homogamie, des liaisons entre variables. En première analyse donc, nous retiendrons 5 facteurs qui nous paraissent pouvoir servir de variables d'entrée dans une série de simulations sur l'endogamie : la région d'origine, la taille de la commune de naissance, la profession du père, le niveau d'instruction, et l'appartenance religieuse.

b) Processus d'appariement.

Ce sont, nous l'avons vu, les caractéristiques spatio-temporelles des individus qui déterminent leurs probabilités de contact entre eux. Si l'on considère le mariage qui en est l'aboutissement, on peut affirmer par exemple que, dans notre échantillon, une femme a 5,7 chances sur 100 de se marier dans sa 26ème année ; ou que, si elle est née en France dans une ville de 2.000 à 5.000 habitants, elle n'a que 1,7 chances sur 100 environ d'épouser un parisien (*). Ces caractéristiques (âge, région d'origine, taille de la commune de naissance) doivent nous permettre de définir l'*univers*

(*) Le premier chiffre est tiré du tableau V de la page 52 de l'étude d'Alain GIRARD, le second est calculé à partir du tableau VII de la page 63.

de choix du conjoint. En supposant cet univers de choix défini pour chaque individu, nous devons déterminer la dynamique du processus d'appariement. Puisque nous avons renoncé à construire des modèles psychologiques vraisemblables, nous réduirons les individus à un schéma unique de décision, les différences interindividuelles de comportement étant considérées comme réductibles aux différences dans les caractéristiques propres des individus. Ce schéma de décision devra comporter deux paramètres d'entrée : le *contenu* de l'attitude endogamique ou exogamique (c'est-à-dire la ou les dimensions sur lesquelles elle porte), et l'*intensité* de cette attitude (c'est-à-dire la probabilité d'acceptation ou de rejet de partenaires ne correspondant pas aux critères de sélection). Il comportera un paramètre de sortie : l'identification du partenaire sélectionné.

Chaque individu va donc devoir sélectionner, au sein de son propre univers de choix, un partenaire compatible avec les composantes de son attitude endogamique (ou exogamique). On ne peut évidemment envisager la possibilité que tous les univers de choix soient disjoints. Quelle que soit la définition de ces univers (isolat, ou vicinité), un même individu appartiendra aux univers de choix de plusieurs individus de sexe opposé. Au cours du processus d'appariement, il est donc inévitable qu'un individu ne trouve plus, dans son univers de choix, de partenaire acceptable. Comme d'autre part notre objectif est de reconstituer autant de couples qu'il y en avait dans l'échantillon d'Alain GIRARD, nous devons définir une procédure réalisant certains mariages forcés, afin qu'à l'issue d'une simulation, il ne reste plus aucun célibataire. L'inconvénient d'une procédure de mariage forcé est qu'elle introduira un biais dans l'homogamie observée par rapport à l'homogamie escomptée pour une intensité donnée de l'attitude endogamique. En effet, les mariages forcés feront bais-

ser le taux de l'homogamie observée pour les variables sur lesquelles porte l'attitude endogamique, le taux théoriquement prévisible se calculant sur des univers de choix disjoints. Afin d'atténuer autant que possible les artéfacts provoqués par le chevauchement des univers de choix, nous devons donc nous efforcer, lorsque nous procéderons à la formalisation du modèle, de lier la procédure de mariage forcé à l'intensité de l'attitude endogamique.

Lorsque tous les individus de l'échantillon auront été appariés, nous pourrons procéder à l'analyse des caractéristiques des couples, et au calcul des taux d'homogamie. Il faut dès maintenant signaler qu'un dépouillement concernant les variables spatio-temporelles ayant servi à définir les univers de choix n'apporterait aucune information nouvelle : par définition, ces variables ne peuvent être manipulées au cours des simulations, et par conséquent les taux d'homogamie selon l'âge et les caractéristiques du lieu de résidence avant le mariage, doivent conserver la valeur qu'ils ont dans l'étude d'Alain GIRARD.

c) Déroulement de l'expérience.

Nous envisageons d'étudier l'influence, sur le taux d'homogamie selon quelques variables servant à décrire les couples, d'attitudes endogamiques (ou exogamiques) de contenu et d'intensité variables. Pour cela, nous procéderons à plusieurs séries de simulations, au cours desquelles nous ferons varier le contenu et l'intensité de ces attitudes. Au cours d'une simulation, il est possible d'affecter à tous les individus le même contenu et la même intensité d'attitude ; il est possible également de leur affecter des contenus ou des intensités différents, ou bien au hasard, ou bien selon certaines

caractéristiques individuelles. Il est probable que l'affectation d'attitudes différentes est moins éloignée de la réalité que l'uniformisation des attitudes. Cependant, au début d'une recherche, il est prudent de n'envisager que des cas simples, dont il soit facile pour l'expérimentateur de conserver le contrôle ; c'est pourquoi nous opterons, dans les premières simulations sur le thème du brassage social, pour l'uniformisation des attitudes.

Le *contenu* des attitudes peut être déterminé par une variable (religion, origine sociale, etc.) ; cela signifiera que l'appariement est déterminé par la concordance ou la proximité des individus relativement à cette variable (même religion, origines sociales voisines). Il peut être également déterminé par une combinaison de variables. Dans ce cas, nous devons distinguer les modèles conjonctifs et les modèles disjonctifs. Dans les modèles *conjonctifs*, l'appariement est déterminé par la concordance des individus relativement à l'ensemble des variables considérées (même religion et même origine sociale) ; dans les modèles *disjonctifs*, il est déterminé par la concordance relativement à l'une quelconque des variables considérées (même religion ou même origine sociale). Par définition, les modèles conjonctifs sont les plus restrictifs, et les modèles disjonctifs les plus tolérants : les modèles ne mettant en jeu qu'une variable unique se situent évidemment entre ces deux types extrêmes. On peut également concevoir des modèles *mixtes*, à la fois conjonctifs et disjonctifs, si l'on dispose d'au moins trois variables ; ou des modèles plus complexes encore, reposant sur une attitude endogamique pour certaines variables, exogamique pour d'autres, exprimés à l'aide de relations conjonctives et disjonctives. Dans cette voie, l'imagination risque de se laisser entraîner par le vertige de la combinatoire, en perdant de vue le

nombre énorme des combinaisons qu'elle suggère (*). Il convient, dans une première étape, de se limiter aux modèles simples ; par exemple, en n'explorant que les modèles à une seule variable, et les modèles conjonctifs et disjonctifs à deux variables (**).

L'*intensité* des attitudes doit permettre de classer des attitudes de même contenu en fonction de leur degré d'intransigeance. Ainsi, si une attitude endogamique concernant la religion aboutit à une proportion moindre de mariages mixtes qu'une autre, on dira que la première est plus intense que la seconde. Par convention d'écriture, on pourra affecter la valeur 0 à l'intensité d'une attitude se traduisant par un phénomène de panmixie restreinte, et utiliser les nombres entiers positifs pour ordonner les intensités correspondant à des attitudes d'autant plus endogamiques que la valeur sera plus élevée. De même, par souci de symétrie, on pourra utiliser des nombres entiers

(*)

On démontrerait que, pour n variables, le nombre total M_n de modèles à une variable, et de modèles conjonctifs ou disjonctifs purs est de :

$$M_n = n + 2 \sum_{p=2}^n \frac{n!}{p!(n-p)!}$$

ou, plus simplement :

$$M_n = n + 2 M_{n-1} \quad \text{avec } M_1 = 1$$

Soit 11 modèles pour 3 variables, 26 pour 4 variables, et 57 pour 5 variables.

(**)

On démontrerait de même que, pour n variables, le nombre de modèles possibles dans ce cas est égal à

$$M_n = n^2$$

Soit 9 pour 3 variables, 16 pour 4 variables, et 25 pour 5 variables.

négatifs pour les attitudes exogamiques, la valeur absolue des nombres étant d'autant plus élevée que l'attitude est d'intensité plus forte. Il conviendra, lors de la formalisation du modèle, de définir si possible des relations plus précises entre ces valeurs numériques. Nous pouvons dès maintenant prévoir pour chaque contenu d'attitude envisagé de faire varier l'intensité de celle-ci d'autant de degrés vers l'exogamie que vers l'endogamie, le point zéro servant d'axe de symétrie.

4.2.2. Formalisation du modèle.

Nous venons de dresser une sorte de cahier des charges de ce que ce modèle doit nous permettre de faire. Il reste à la construire. Cette construction nous obligera à choisir entre les diverses possibilités que nous avons décrites, et à introduire certaines précisions là où notre définition du problème a laissé dans l'ombre des détails importants. Ainsi que nous l'avons vu (§ 2.2.4.d), le passage d'une théorie à un modèle opératoire peut se faire selon trois procédures non exclusives : la spécification, la simplification, et la réduction (32, Pp. 3-5 ; 33, Pp. 381-384). D'après la description que nous avons donnée des composantes et des objectifs de ce modèle, nous voyons immédiatement que nous allons devoir utiliser la *spécification* pour rendre opératoire la notion d'univers de choix du conjoint, ou celle d'attitude endogamique ; la *simplification*, pour définir le contenu des attitudes à partir de la combinaison de quelques variables seulement, liées de manière rigide par des opérateurs logiques ; la *réduction*, en limitant notre recherche à l'observation du seul taux global d'homogamie selon quelques variables, sans nous intéresser aux autres conséquences des comportements endogamiques. Ces

opérations ont pour but soit de rendre le "calcul" possible (au sens de *calcul logique* ; cf. #3, p. 16), soit simplement de le rendre plus aisé. Dans le second cas, il est certain que les recherches ultérieures devraient tendre à "recomplexifier" le modèle de manière à se rapprocher de la complexité du phénomène réel.

Les principaux points sur lesquels il est nécessaire de faire un effort de formalisation, fût-ce provisoirement au prix d'un appauvrissement des concepts utilisés, sont soulevés par les questions suivantes :

- comment définir l'univers de choix ?
- quelles variables de l'enquête d'A. GIRARD retenir dans ce modèle, et pour remplir quelles fonctions ?
- comment l'intensité des attitudes agira-t-elle sur le mécanisme de sélection du partenaire ?
- qui doit choisir son partenaire : l'homme, ou la femme ?
- quel degré de précision dans les résultats doit-on exiger ?

Nous allons préciser chacun de ces points.

a) Comment définir l'univers de choix de chaque individu ?

Par *univers de choix*, nous entendons l'ensemble des partenaires de sexe opposé avec lesquels un individu donné peut envisager l'éventualité d'une union. Cela implique que ces partenaires ne soient pas trop éloignés de lui géographiquement, et que leur âge ne soit pas trop différent du sien. Pour définir l'univers de choix des individus, nous devons soit recourir à la notion de *vicinité* (distance spatio-temporelle entre les individus), soit regrou-

per les individus dans des *isolats*. Nous allons tout d'abord examiner la possibilité de traduire de manière opératoire la notion de *vicinité*, notion plus conforme à ce que nous savons du processus réel de formation des couples. Nous verrons que la traduction opératoire de cette notion soulève d'importantes difficultés, et nous serons alors conduit à opter pour la constitution d'*isolats*, solution plus artificielle, mais plus facile à intégrer dans un modèle simple.

Pour définir, à partir de la notion de *vicinité*, l'univers de choix des individus impliqués dans la simulation, il nous faut mesurer entre eux une *distance* spatio-temporelle. Celle-ci peut être évaluée à partir de trois variables : l'âge, la région de résidence, et la taille de la commune de résidence de chaque individu avant son mariage. Il y a deux procédures possibles pour calculer une telle distance : l'élaboration d'un indice empirique analogue à celui utilisé par Alain GIRARD pour évaluer la *distance globale entre les conjoints* (93, Pp. 87-94), ou le calcul de la *probabilité* pour qu'un individu de caractéristiques spatio-temporelles données soit en contact avec un individu de sexe opposé défini à partir des mêmes caractéristiques. L'élaboration par Alain GIRARD d'un indice de distance calculé *a priori* se justifie surtout par les conditions techniques dans lesquelles le dépouillement de l'enquête a été effectué : l'emploi de procédés mécanographiques rend en effet très délicate, et dans certains cas pratiquement impossible, l'utilisation d'indices calculés à partir des résultats d'un premier dépouillement (cf. § 5.1.1.a). Ce problème technique ne se posant pas pour nous, il nous est facile de critiquer le recours à un coefficient de distance spatio-temporelle calculé *a priori*. Un tel indice se présenterait par exemple comme la somme ou le produit de trois indices relatifs chacun à l'une des variables servant à l'évaluation de la distance. Outre le fait qu'un tel indice ne tiendrait pas compte des liaisons qui existent nécessairement entre

les trois variables (*), il présenterait l'inconvénient d'introduire des coefficients de pondération arbitraires : nous n'avons *a priori* aucune raison de donner le même poids aux trois variables. Il serait possible il est vrai de calculer des coefficients de distance *a posteriori*, par exemple en donnant à chaque variable un poids d'autant plus fort que la distribution correspondante est moins dispersée (c'est-à-dire que le taux d'homogamie γ est plus élevé) ; mais même dans ce cas, le choix de la fonction qui liera le coefficient de pondération à la dispersion est nécessairement arbitraire, et il n'est pas certain que l'ordre obtenu sur les distances, à partir d'indices utilisant des coefficients de pondération différents, soit invariant. Dans ces conditions, nous ignorons si nous pourrions aboutir à une famille d'indices équivalents, qui ne risquent pas de définir des univers de choix dissemblables pour un individu donné.

Dans la mesure où nous disposons d'un ordinateur, rien ne s'oppose à ce que nous calculions un indice de distance à partir de la probabilité effective qu'ont de se rencontrer un homme et une femme de caractéristiques spatio-temporelles données. Les éléments nécessaires au calcul d'un tel indice se trouvent dans l'enquête d'Alain GIRARD : certains de ces éléments figurent même dans le compte-rendu des résultats de l'enquête. Ainsi, par exemple, un tableau de fréquences croisées met en correspondance l'importance de la localité de naissance des conjoints (93, p. 58). A partir de ce tableau, il est facile de calculer quelle était, dans cet échantillon, la probabilité moyenne pour qu'un homme résidant dans une commune d'une taille donnée épouse une femme résidant dans une commune de même taille, ou d'une taille

(*) L'existence de ces liaisons est confirmée par le tableau IV page 61 de l'étude d'A. GIRARD (93), qui indique une mobilité géographique plus forte chez les jeunes (moitié inférieure du tableau).

différente. On constate par exemple qu'un homme vivant dans une commune rurale avait en moyenne une probabilité de 0,84 d'épouser une femme vivant également dans une commune rurale, contre une probabilité inférieure à 0,01 d'épouser une femme résidant dans la région parisienne. Il est possible, à partir de telles informations, de définir un indice de distance qui soit une fonction monotone décroissante de cette probabilité : on admet aisément qu'une femme vivant dans une commune rurale soit plus proche d'un homme vivant également dans une commune rurale qu'une femme de la région parisienne, et que l'indice mesurant la proximité soit une fonction directe de la probabilité de contact. L'indice de distance peut simplement être le complément à 1 ou l'inverse de l'indice de proximité. Rien ne prouve il est vrai que la probabilité de *mariage* et la probabilité de *contact* soient égales dans tous les cas, et qu'estimer celle-ci à partir de celle-là soit une démarche correcte. Mais il ne faut pas perdre de vue que, *dans le contexte de notre simulation*, les seuls contacts que nous désirons prendre en compte sont ceux qui peuvent effectivement aboutir à un mariage. Du point de vue opératoire qui est le nôtre, les deux notions se confondent.

Une mesure de la distance spatio-temporelle devrait satisfaire à au moins trois conditions : correspondre à l'idée intuitive que l'on a de la distance spatio-temporelle, être sensible à l'influence des variables impliquées dans la simulation, et être calculable avec les données dont nous disposons. La première condition se ramène essentiellement à l'idée d'*invariance* de l'ordre sur les distances, quel que soit le mode de calcul choisi, et à celle de *symétrie* de l'indice. Si la condition d'invariance est une exigence que l'on peut admettre sans discussion, il n'en est pas de même pour la condition de symétrie. Certes, il est d'expérience courante que la

distance en mètres entre le point A et le point B est égale, quels que soient ces points, à la distance entre le point B et le point A ; mais si A est sur un sommet et B dans une vallée, et que l'on mesure la distance en heures de marche, $AB \neq BA$. De même, la distance spatio-temporelle mesurée en probabilité de contact n'est pas symétrique. Alors qu'une femme résidant dans une commune de 2.000 à 5.000 habitants a 1,7 chances sur 100 d'épouser un Parisien, un Parisien a 2,3 chances sur 100 environ d'épouser une habitante d'une commune de 2.000 à 5.000 habitants (cf. 93, p. 58, tableau II). On comprend d'ailleurs bien que la probabilité qu'une femme donnée appartienne à l'univers de choix d'un homme déterminé n'est pas nécessairement égale à la probabilité que cet homme appartienne à l'univers de choix de ladite femme.

La seconde condition exprime l'idée que, si la probabilité de contact entre individus dépend non seulement de leurs caractéristiques spatio-temporelles, mais aussi des variables socio-culturelles prises en compte dans la simulation, l'indice de distance doit tenir compte de ces variables. Nous savons par exemple que l'âge moyen au mariage n'est pas le même selon la profession ou le niveau d'instruction, que la mobilité géographique varie selon la profession (93, Pp. 60-61), etc.. Si nous considérons deux hommes h_1 et h_2 dont les caractéristiques spatio-temporelles sont identiques mais les caractéristiques socio-culturelles différentes, et deux femmes, f_1 et f_2 , répondant aux mêmes conditions, alors la mesure de la distance spatio-temporelle entre h_1 et f_1 doit différer (aux coïncidences numériques près) de la mesure de la distance spatio-temporelle entre h_2 et f_2 . En conséquence, le mode de calcul de l'indice de distance doit tenir compte à la fois des caractéristiques spatio-temporelles et des caractéristiques socio-culturelles des individus considérés. En

outre, il doit tenir compte également de l'évolution, au cours des 45 années environ sur lesquelles s'étendent les mariages recensés, de l'influence de ces variables socio-culturelles sur les probabilités de rencontre entre les individus de l'échantillon.

La condition de calculabilité est celle qui se heurte au plus grand nombre de difficultés pratiques. Supposons tout d'abord, pour simplifier le problème, que nous désirions simplement évaluer la probabilité moyenne qu'avaient de se rencontrer les conjoints recensés par l'enquête, en négligeant l'influence des variables socio-culturelles, mais en tenant compte des liaisons statistiques existant entre les trois variables spatio-temporelles retenues. Pour évaluer cette probabilité, il faudrait partir d'un tableau croisé à 6 dimensions donnant l'effectif des couples correspondant à chaque combinaison des états pris par les variables spatio-temporelles. En raison du nombre élevé de ces combinaisons, et de la taille relativement réduite de l'échantillon, un tel tableau serait d'ailleurs ininterprétable, le nombre de cases vides étant nécessairement plus élevé que celui des cases où figurerait ne fût-ce qu'un seul couple (*). Dans ces conditions, il serait utopique de vouloir tenir compte de l'influence des variables socio-culturelles, et de leur évolution.

(*) Rappelons que l'enquête a porté sur 1646 couples. Si nous conservons la précision fournie par l'enquête sur la taille de la localité de résidence (6 catégories), si nous regroupons les 24 régions utilisées par A. GIRARD (op. cit. p. 66) en 6 zones (op. cit., p. 46), et enfin si nous ne distinguons que 5 classes d'âge (op. cit., p. 48), le tableau à six dimensions correspondant au croisement de ces trois variables pour chaque couple aura $6^2 \times 5^2 \times 5^2 = 35.400$ cases. C'est-à-dire que la probabilité qu'une case quelconque ne soit pas vide est inférieure à 0,05. Pour qu'un tableau soit interprétable, il faut que la présence de cases vides ne puisse pas être imputée à la faiblesse de l'échantillon, mais à l'absence, dans la population étudiée, des caractéristiques correspondantes.

Admettons cependant que, compensant l'insuffisance des données empiriques par un effort de théorisation (cf. § 2.2.4.c), nous ayons mis au point une formule de calcul de l'indice cherché satisfaisant aux deux premières conditions. Le problème ici n'est plus de calculer une distance à l'intérieur des couples *constitués* (comme dans l'étude d'Alain GIRARD), mais à l'intérieur de chacun des couples *possibles*. C'est-à-dire qu'il nous faut évaluer la distance spatio-temporelle existant entre un individu donné, et chacun des individus de sexe opposé se trouvant dans l'échantillon. Pour l'ensemble de l'échantillon, nous devrions évaluer plus de 5 millions de distances (*). Pour alléger le calcul, il serait possible d'adopter un indice symétrique, réduisant à 2,5 millions environ le nombre de distances à évaluer. On pourrait également envisager de mettre au point un algorithme de tri destiné à diminuer le nombre de distances à calculer, en évitant de calculer les distances manifestement trop grandes pour devoir être prises en compte dans la constitution des univers de choix ; par exemple, en négligeant tous les couples d'individus dont la différence d'âge est supérieure à un seuil donné.

Supposons ce nouveau problème résolu ; nous disposons alors d'une mesure de distance satisfaisant aux trois conditions énoncées ci-dessus. Après le calcul de toutes les distances spatio-temporelles nécessaires, nous pourrions, pour constituer les univers de choix, choisir entre deux procédures : ou bien on fixera un seuil de distance spatio-temporelle, et tous les individus étant en deçà du seuil seront retenus ; ou bien on imposera un effectif maximum à l'univers de choix, et l'on retiendra le nombre fixé d'individus en commençant par les plus proches. La première procédure semble la plus simple à réaliser, et aussi la plus réaliste. En effet, les dimensions des univers de choix sont

(*)

Exactement $1646^2 \times 2 = 5.418.632$ distances.

variables d'un individu à l'autre, en fonction des caractéristiques spatio-temporelles de chacun ; ces caractéristiques déterminent les coordonnées des partenaires possibles, et aussi leur nombre. Quel que soit le processus de sélection adopté, sa mise en oeuvre sera délicate, car les univers de choix ainsi constitués comporteront nécessairement de nombreuses intersections ; après chaque choix, il sera nécessaire de "mettre à jour" les univers de choix des individus encore célibataires, en retirant de ces univers les partenaires éventuels déjà choisis par d'autres. Il arrivera nécessairement que certains univers de choix se trouveront ainsi vidés de leur contenu, avant que leur possesseur ait eu l'occasion de choisir un partenaire ; il arrivera de même que certains individus n'aient pas été sélectionnés au sein de tous les univers de choix dont-ils dépendent. L'utilisation de la *vicinité* pour définir les univers de choix risque donc de nous conduire à faire un certain nombre de célibataires au cours de chaque simulation. Or, étant donnée la nature même de l'enquête qui sert de base à cette recherche, si nous dissociions les couples, c'est afin de reconstituer autant de couples qu'il y en avait dans les données empiriques. Pour remplir coûte que coûte une telle condition, il nous faudrait alors soit procéder à des mariages "contre nature", en appariant les célibataires au mépris de leur distance spatio-temporelle (au mieux, en les appariant de manière à minimiser la somme de ces distances), soit élargir certains univers de choix, en abaissant arbitrairement le seuil de proximité requis pour y être admis. Dans les deux cas, cette procédure compliquée aurait pour effet de modifier *de façon non prévisible* les conditions de l'expérience, et donc d'*en fausser les résultats*.

Si maintenant nous procédons à l'examen critique de la solution qui consiste à constituer des *isolats*, nous voyons qu'une

telle solution présente des avantages pratiques certains. Tout d'abord, les isolats sont assez faciles à réaliser : il suffit de diviser l'ensemble des individus en sous-ensembles disjoints relativement homogènes des points de vue de l'âge, et du lieu de résidence au moment du mariage. Pour calquer le degré d'homogénéité des isolats sur les relations réelles relevées entre les caractéristiques spatio-temporelles des individus mariés, on pourrait tenir compte de la dispersion de ces caractéristiques autour de la tendance centrale, c'est-à-dire en particulier du degré d'homogamie observé pour l'âge et les coordonnées du lieu de résidence. Un second avantage majeur est que la division de la population en isolats simplifie le problème de la "mise à jour" des univers de choix au cours du processus d'appariement. A chaque mariage en effet, un homme disparaît de l'univers de choix des femmes de l'isolat, et une femme disparaît de l'univers de choix des hommes. Ces univers de choix étant communs pour un sexe donné, et de même effectif pour les deux sexes, il ne peut y avoir de "laissés pour compte" à l'issue du processus. En outre, les "mariages forcés", en fin de processus, se font toujours entre partenaires qui étaient *a priori* appariables, puisque telle est la définition de l'isolat.

Ces facilités sur le plan technique sont la contrepartie de difficultés théoriques importantes. On peut difficilement admettre qu'un découpage aussi sommaire de la population corresponde à un aspect de la réalité. Il est vrai que l'on a pu étudier de véritables isolats dans des populations humaines, mais l'existence d'isolats est en général due à une économie de type agraire, et une absence de moyens de communication que l'on ne retrouve plus en France à notre époque, même en région de montagnes (cf. 93, Pp. 19-21). D'autre part, même chez les populations animales isolées géographiquement, les isolats ne constituent pas des systèmes totalement clos, en raison des migrations occasionnelles, ou des propagations de gènes à travers les espaces peu peuplés

(28, Pp. 20-21). Aussi devons-nous considérer l'utilisation de la notion d'isolat comme un *artifice opératoire* destiné à faciliter le calcul, et nous efforcer de compenser les biais théoriques que cet artifice introduit.

Si nous définissons les isolats à partir de l'âge des conjoints, la base que nous devons prendre est évidemment l'âge des conjoints au moment de l'enquête, et non au moment de leur mariage : sinon, nous risquerions d'apparier des personnes appartenant à des générations différentes. Afin d'éviter que puissent apparaître de telles incohérences, il nous faut regrouper dans les isolats des individus nés à la même époque. Nous savons exactement quelle est, pour l'ensemble de l'échantillon, la dispersion autour de l'âge moyen au mariage. Il nous est donc possible de constituer des échantillons tenant compte de cette dispersion ; par exemple, un isolat comptant 100 personnes des deux sexes et correspondant approximativement à un mariage célébré en 1936 devrait comporter d'une part environ 12 hommes nés entre 1895 et 1908, 13 entre 1909 et 1911, 14 entre 1912 et 1913, 11 entre 1914 et 1920, et d'autre part 13 femmes nées entre 1899 et 1911, 15 entre 1912 et 1914, 12 entre 1915 et 1916, et 10 entre 1917 et 1921. L'inconvénient de cette procédure est qu'elle se fonde sur l'*âge moyen* et la *dispersion moyenne*, sans tenir compte de l'évolution de ces paramètres dans le temps ; il est possible cependant de corriger ce biais en calculant et en prenant pour bases la moyenne et la dispersion de l'âge au mariage par générations (*). Un

(*) L'auteur fait remarquer (93, p. 52-53) que ces paramètres sont calculés sur des échantillons faibles, et de surcroît biaisés sous l'effet de la mortalité. Ces points limitent l'intérêt d'une extrapolation des valeurs prises par ces paramètres à l'ensemble de la population française. Par contre, ils décrivent parfaitement l'échantillon tel que nous l'utilisons dans la simulation, et, à ce titre, peuvent être utilisés sans aucune restriction dans la constitution des isolats.

second inconvénient est que, si nous constituons chaque isolat en sélectionnant hommes et femmes à partir du seul critère de l'âge, nous ne tenons pas compte de l'influence, sur l'âge au mariage, des facteurs socio-culturels ; il est vrai que le modèle collectif semble relativement homogène, et que, dans l'ensemble, les comportements semblent l'expression de la norme admise (cf. 93, Pp. 143-148). Il est cependant aisé de tenir compte de l'ensemble des facteurs socio-culturels en faisant porter la sélection non sur les individus dissociés préalablement, mais sur les couples, que l'on ne dissociera qu'au sein d'un isolat constitué. La procédure de constitution des isolats en est d'ailleurs considérablement simplifiée. Il suffit en effet de sélectionner les couples en fonction de la *date de leur mariage* ; on dispose ainsi d'un sous-échantillon dont la dispersion de l'âge au mariage des individus reflète les variations dues à l'époque, et à l'influence des facteurs socio-culturels.

Les isolats doivent non seulement tenir compte de l'âge des individus au mariage, mais aussi de leur lieu de résidence à la même époque. Nous avons caractérisé le lieu de résidence à partir de deux variables : la région, et la taille de la localité. Pour ces variables, le problème se pose dans les mêmes termes que pour l'âge : il est indispensable que, dans chaque isolat, les distributions correspondant aux coordonnées géographiques des hommes et des femmes tiennent compte de la répartition des différences de région et de taille des agglomérations. La solution calquée sur celle que nous avons adoptée pour l'âge aurait consisté à sélectionner les couples selon le lieu où le mariage a été célébré. Les données dont nous disposons n'ayant pas été prévues pour cette recherche (puisque celle-ci constitue un cas particulier d'analyse secondaire), le codage des réponses à la question sur le lieu du mariage ne nous aurait pas

permis d'effectuer une telle sélection (*). Nous pouvons tourner cette difficulté en sélectionnant les couples à partir de la localité de résidence de l'un des conjoints, le mari par exemple. Dans ces conditions, chaque isolat regroupera une sous-population d'hommes homogène du point de vue du lieu de résidence, et une population de femmes présentant une certaine hétérogénéité de ce point de vue. Dans un isolat composé uniquement de Parisiens et de (futurs) femmes de Parisiens, la probabilité moyenne pour qu'un Parisien épouse une femme résidant dans une ville de 2.000 à 5.000 habitants sera toujours égale à 2,3 %, tandis que la probabilité qu'une femme quelconque de l'isolat épouse un Parisien sera évidemment de 100 %. Au niveau de l'ensemble des isolats cependant, ce biais se trouvera nécessairement corrigé, de telle sorte que par exemple sur le total des femmes résidant dans une ville de 2.000 à 5.000 habitants, il y en ait 1,7 % dans les isolats composés de Parisiens (et près de 60 % dans les isolats composés d'hommes résidant dans des villes de 2.000 à 5.000 habitants).

Nous constituerons donc nos isolats en sélectionnant des couples selon leur date de mariage, et la localité de résidence du mari

(*)

Ce codage résulte en effet de la comparaison de la réponse à la question *lieu du mariage* avec les réponses relatives à la localité de résidence de chacun des conjoints avant le mariage. Ces modalités prévues sont :

- localité de résidence du mari et de la femme,
- localité de résidence du mari,
- localité de résidence de la femme,
- autre localité en France,
- hors de France.

Dans 8 % des cas, le mariage a eu lieu dans une autre localité, soit en France, soit hors de France (93, p. 133), et nous ne possédons pas d'autres informations sur le lieu du mariage (on pourrait il est vrai supposer par exemple que le mariage a eu lieu conformément à la tradition chez les parents de la femme si celle-ci ne résidait pas avec eux).

avant le mariage. Il reste encore un problème à aborder : celui de la taille des isolats. La constitution des isolats peut se faire soit en découpant la population à partir d'intervalles définis sur les variables spatio-temporelles retenues, soit en regroupant dans des unités d'effectif constant des couples proches selon ces variables. La seconde solution présente de nombreux inconvénients que nous avons déjà soulignés à propos de la notion de *vicinité*, et ne présente aucun avantage décisif ; dans la réalité, les univers de choix ne sont pas nécessairement de mêmes dimensions. Aussi nous rallierons-nous à la première solution, en nous imposant cependant une limite raisonnable pour la dimension maximum des isolats ; sans chercher à justifier cette décision par des considérations théoriques, nous nous interdirons de constituer des isolats égaux à 10 % de l'ensemble de la population totale (*).

(*) Les considérations qui nous ont conduit à choisir 10 % comme limite supérieure à la taille des isolats reposent sur des considérations d'ordre essentiellement pratique. Sur le plan théorique en effet, les études concernant l'isolat aboutissent généralement à des propositions portant sur la distance géographique entre le postulant au mariage et ses partenaires possibles, et non sur le nombre de partenaires constituant l'univers de choix. Dans l'enquête elle-même, une question concerne ce problème : "Vous-même, avant votre mariage, combien de personnes estimez-vous avoir rencontrées, avec qui un mariage aurait été possible ?" Les réponses à cette question donnent une moyenne de 1,7 personnes (93, p. 156). Il s'agit évidemment de l'univers *perçu*, compte-tenu des restrictions socio-culturelles, et non de l'ensemble des personnes de sexe opposé et de même génération *avec qui l'on s'est trouvé en contact*. Sur le plan pratique, il nous a fallu tenir compte de la dispersion relative à une variable géographique, la taille d'agglomération. Dans tous les recensements et sondages, les enquêteurs ont renoncé à relever avec précision la dimension des communes rurales (moins de 2.000 habitants) ; aussi cette catégorie est-elle toujours pléthorique, quel que soit le phénomène étudié. Ainsi, dans l'étude sur le choix du conjoint, les communes rurales représentent-elles 37,7 % des communes de résidence des futurs conjoints (93, p. 58), et 45,8 % des localités de naissance (pour les conjoints nés en France ; *ibid.*, p. 63). Compte-tenu des variations de ces proportions dans les diverses classes d'âge, il nous a paru difficile d'abaisser la dimension maximum des isolats à moins de 8 % de la population, sous peine d'avoir un nombre important d'isolats très petits (moins de 2 % de la population).

b) Quelles variables retenir ?

Les variables dont nous avons explicitement besoin dans notre modèle sont de trois types : les variables spatio-temporelles définissant les isolats ; les variables expérimentales, relatives au contenu des attitudes endogamiques ou exogamiques (variables d'entrée du modèle) ; et les variables sur lesquelles nous mesurerons le taux d'homogamie (variables de sortie du modèle).

Nous pouvons définir les isolats à partir de la date du mariage, la taille et la région de la localité de résidence du mari. La date du mariage varie de 1914 environ à 1959, et peut être recodifiée en un nombre de catégories supérieur ou égal à 4 (93, Pp. 48-49) ; la dimension de la localité de résidence comporte 6 catégories, que l'on pourrait regrouper en 3 ou 4 compte-tenu de leurs effectifs ; la région de résidence est codée en 24 états (ibid., p.66), recodifiables en 6 zones (ibid., p. 46). En combinant ces trois variables recodifiées, nous constituerions au minimum 72 isolats, ce qui, pour plus de 3200 individus de deux sexes, constitue un nombre raisonnable (dimension moyenne de l'isolat : 22 personnes environ). Nous avons préféré cependant accroître la précision sur l'*époque du mariage*, et sur la *taille de la localité de résidence*, en estimant que ces deux variables ont une grande influence sur la plupart des variables pour lesquelles on relève un degré élevé d'homogamie dans l'enquête (cf. § 4.2.1.a). On sait en particulier que la mobilité géographique, le niveau d'instruction, la qualification professionnelle, ont évolué avec le temps ; on sait également que la proportion des diverses couches sociales varie fortement selon la catégorie d'agglomération. Par contre, la région de résidence nous a semblé d'un intérêt moindre. Non certes que nous ignorions l'importance des disparités

régionales ; mais il nous a paru que l'incidence de ces disparités sur des variables telles que le niveau d'instruction, la religion, ou l'origine sociale des conjoints était assez faible, relativement à celles des variables précitées. C'est pourquoi nous avons choisi de définir les isolats à partir de la date du mariage regroupée en 9 époques (avant 1920, puis de 5 ans en 5 ans jusqu'à 1959), et de la taille de la localité de résidence du mari en 6 catégories. Nous avons ainsi 54 isolats, ayant un effectif d'environ 30 personnes en moyenne.

Parmi les variables d'entrée du système, il ne pouvait être question de choisir la région de résidence des conjoints, ni la région de naissance, qui lui est assez fortement liée : notre définition de l'isolat nous interdit en effet toute analyse au niveau des phénomènes régionaux, puisqu'elle suppose une *mobilité interrégionale absolue*, en contradiction avec l'expérience (27,8 % des couples sont formés d'individus nés dans des régions différentes, et seulement 8,8 % de personnes qui résidaient dans des régions différentes avant leur mariage). Les variables d'entrée doivent donner un contenu aux attitudes endogamiques ou exogamiques ; pour être opératoires, elles doivent correspondre à des caractères effectivement mesurés sur les individus étudiés. Une attitude endogamique consciente chez les personnes interrogées a trait au milieu social : 68 % d'entre elles estiment qu'il vaut mieux que les époux appartiennent au même milieu social (contre 24 % qui déclarent que c'est sans importance) (93, p. 150). Nous ne disposons pas d'informations sur d'autres attitudes endogamiques portant sur les caractéristiques mesurées. Mais, compte-tenu du fait que les attitudes ne sont pas nécessairement conscientes, nous avons retenu toutes les variables pour lesquelles on relève un taux élevé d'homogamie, hormis la nationalité. Nous avons donc choisi

d'étudier les incidences sur l'homogamie de comportements simulés relatifs à l'*origine sociale* des partenaires (définie par la profession du père), à leur *origine rurale ou urbaine* (définie par l'importance de leur localité de naissance), à leur *appartenance religieuse*, et à leur *niveau d'instruction*.

Nous avons également choisi de retenir ces quatre variables comme variables de sortie.

c) Le mécanisme de sélection du conjoint.

Pour mettre au point ce mécanisme, et le relier à l'intensité des attitudes, qui constitue un des paramètres d'entrée du modèle, nous nous limiterons tout d'abord au problème de l'*endogamie selon une seule variable*. Notre objectif était d'élaborer un algorithme simple qui fasse intervenir le hasard au niveau individuel, et qui, au niveau de l'isolat, aboutisse à un taux d'homogamie selon la variable considérée qui soit une fonction croissante monotone du degré d'intensité. Pour atteindre ce but, nous pouvons décomposer le phénomène du mariage en deux processus successifs : la prise de contact, et la décision. La *prise de contact* peut être le résultat d'un processus purement aléatoire : le postulant "tire", au sein de l'isolat, un partenaire de sexe opposé. Si le processus complet s'arrêtait là, il y aurait panmixie à l'intérieur de l'isolat, et panmixie restreinte au niveau de l'ensemble de la population. La *prise de décision* comporte un jugement sur l'acceptabilité du partenaire : ou bien celui-ci correspond au critère de sélection exprimé par le contenu de l'attitude endogamique, et le mariage se fait ; ou bien il ne correspond pas, et nous sommes ramené au point de départ. Le postulant

doit avoir la possibilité de "tirer" de nouveau un partenaire. Si l'univers de choix était sinon illimité, du moins très grand, chaque postulant finirait probablement par sélectionner un partenaire qui lui convienne ; nous serions d'ailleurs en présence d'un cas d'endogamie absolue. Mais les univers de choix sont restreints, et nous désirons établir des degrés dans l'endogamie. Une procédure simple à mettre en oeuvre consiste à limiter la possibilité qu'a le postulant de recommencer à choisir au hasard dans l'isolat un autre partenaire s'il n'est pas satisfait. Ainsi, si le postulant a le droit de tenter à nouveau 3 essais, cela impliquera qu'au 4e tirage, il devra se marier, qu'il soit ou non satisfait de son partenaire. L'avantage de ce système, outre sa simplicité, est qu'il donne une signification concrète aux degrés d'intensité de l'attitude endogamique : le degré *zéro* correspond au cas où l'on n'a pas le droit de choisir à nouveau (*panmixie*), le degré *un* au cas où l'on peut tenter *un* nouvel essai, etc..

Il nous faut définir les modalités du jugement qui se traduira par l'acceptation ou le rejet d'un partenaire possible. Ce jugement peut être plus ou moins complexe : pour accepter un partenaire, on peut par exemple exiger qu'il soit rigoureusement de même origine sociale, ou d'une origine sociale voisine, ou encore, si l'on considère une hiérarchie sociale (transformation d'une variable *nominale* en variable *ordinaire*), exiger qu'il soit d'une condition supérieure à celle du postulant. Pour cette première série de simulations, nous avons choisi la solution la plus simple, qui est également applicable à chacune des quatre variables retenues : elle consiste à fonder l'acceptation du partenaire sur la relation d'*équivalence stricte*. Dans le cas d'attitudes dont le contenu est une combinaison de deux variables (cf. § 4.2.1.c), si le modèle est

conjonctif, la condition d'acceptation du partenaire sera qu'il remplisse à la fois les deux conditions d'équivalence ; si le modèle est *disjonctif*, il suffira qu'il en remplisse une seule.

L'extension de ce mécanisme de sélection à l'exogamie pose quelques problèmes pratiques relatifs tant au degré qu'au contenu de l'attitude. Nous avons convenu (§ 4.2.1.c) d'instaurer une relation de symétrie entre les degrés d'endogamie et les degrés d'exogamie, de telle sorte que des degrés d'intensité comparable ne se distinguent entre eux que par leur signe (négatif dans le cas d'attitude exogamique). En pratique, un nombre négatif d'autorisations de choisir à nouveau un partenaire n'aurait pas grande signification. Nous conviendrons donc que, si le degré d'intensité d'une attitude est négatif, tout se passe comme si le contenu de l'attitude était remplacé par sa négation, et son degré d'intensité par sa valeur absolue. Par exemple, une attitude ayant pour contenu l'*origine sociale*, avec un degré d'intensité égal à -3, correspondra à une décision d'acceptation si l'origine sociale du partenaire est différente de celle du postulant, et, dans le cas contraire, à une décision de rejet, avec possibilité de "tirer" trois fois successives (*). Le cas est légèrement plus complexe pour un modèle de décision à deux variables. Si nous sommes en présence d'une attitude ayant pour contenu un modèle disjonctif portant sur l'origine sociale et le niveau d'instruction, et de degré d'inten-

(*) Si nous appelons X la variable qui sert de critère de décision (contenu de l'attitude), X_a la valeur de cette variable chez l'agent qui décide (le postulant a au mariage), et X_p la valeur de la même variable chez le sujet "tiré" au hasard p (partenaire "passif") :

- dans le cas d'une attitude endogamique, nous aurons mariage si $X_a \equiv X_p$;

- dans le cas d'une attitude exogamique, nous aurons mariage si $\neg (X_a \equiv X_p)$,

c'est-à-dire si $X_a \neq X_p$.

sité égal à -3, il y aura acceptation si l'origine sociale et le niveau d'instruction du partenaire sont différents de ceux du postulant, et rejet dans le cas contraire, avec possibilité de "tirer" trois fois successives (*). En règle générale, la négation d'un modèle disjonctif est la conjonction de la négation des variables, et réciproquement (formules de DE MORGAN) :

$$\begin{aligned} \neg (a \vee b) &\equiv \neg a \wedge \neg b \\ \neg (a \wedge b) &\equiv \neg a \vee \neg b \end{aligned}$$

Le déroulement du mécanisme de sélection du conjoint au sein d'un isolat a pour conséquence de restreindre progressivement l'univers de choix des individus qui n'ont pas encore choisi leur partenaire. Ainsi, à l'intérieur d'un même isolat, l'individu qui choisit a un univers de choix plus restreint d'une personne que l'univers de choix de celui qui a choisi avant lui. Le même processus devant se répéter autant de fois qu'il y a d'isolats dans la population, l'ordre dans lequel les individus choisissent leur partenaire n'est pas indifférent. En effet, il serait normal que, d'un isolat à l'autre,

(*) Si nous appelons X et Y les variables qui servent de critères de décision, si nous conservons les mêmes symboles que dans la note précédente, et si nous considérons le cas d'un modèle *disjonctif* :

- dans le cas d'une attitude *endogamique*, nous aurons mariage si $(X_a \equiv X_p) \vee (Y_a \equiv Y_p)$;

- dans le cas d'une attitude *exogamique*, nous aurons mariage si $\neg ((X_a \equiv X_p) \vee (Y_a \equiv Y_p))$,

soit $\neg (X_a \equiv X_p) \wedge \neg (Y_a \equiv Y_p)$,

soit $(X_a \neq X_p) \wedge (Y_a \neq Y_p)$.

Si nous considérons le cas d'un modèle *conjonctif* :

- dans le cas d'une attitude *endogamique*, nous aurons mariage si $(X_a \equiv X_p) \wedge (Y_a \equiv Y_p)$;

- dans le cas d'une attitude *exogamique*, nous aurons mariage si $(X_a \equiv X_p) \vee (Y_a \neq Y_p)$.

ce ne soient pas toujours les mêmes individus (par exemple ceux qui se trouvent rangés les premiers dans l'isolat) qui bénéficient de la plus grande latitude de choix ; une telle régularité risquerait d'introduire des biais systématiques dans les résultats. Ces biais se retrouveraient également d'une simulation à l'autre. Nous allons voir que cette difficulté peut se trouver résolue par la solution que nous apporterons au problème suivant : qui doit choisir, de l'homme et de la femme ?

d) Qui doit choisir : l'homme, ou la femme ?

Ce problème a déjà fait l'objet des réflexions d'ORCUTT et de ses collaborateurs, lorsqu'ils ont mis au point la partie démographique de leur modèle économique (156, Pp. 75-76). Ces auteurs ont envisagé trois procédures différentes pour simuler le processus de mariage sur une population. Dans la première, seul l'homme choisit sa partenaire ; il détermine, par un processus aléatoire tenant compte de ses propres caractéristiques, les caractéristiques de sa future épouse, et la sélectionne ensuite dans la population féminine. Il est évidemment possible d'inverser les rôles, de sorte que ce soit toujours la femme qui choisisse son mari. Dans la seconde procédure, hommes et femmes choisissent à tour de rôle : un homme sélectionne son épouse, puis une femme son époux, et ainsi de suite. Dans la troisième procédure, les hommes et les femmes à marier sont sélectionnés en bloc pour constituer une sorte de "marché du mariage" (*marriage market*), dans lequel les unions se réalisent par ajustement de l'offre et de la demande. La difficulté de recourir à la troisième procédure réside dans la nécessité d'élaborer un algorithme d'ajustement de l'offre et de la demande suffisamment plausible sans être trop complexe ; un

tel algorithme devrait vraisemblablement faire appel à un indice de distance entre les partenaires, indice dont nous avons montré les inconvénients à propos de la notion de *vicinité*. Les auteurs ont préféré utiliser la seconde procédure, dans laquelle hommes et femmes choisissent alternativement leur conjoint.

Les objectifs du modèle démographique d'ORCUTT diffèrent sensiblement de ceux de notre modèle. ORCUTT ne vise pas à expérimenter sur une population close, mais étudie au contraire l'évolution dans le temps d'une population réelle, avec ses naissances, ses décès, ses mariages, ses séparations (cf. § 3.1.5.b). Aussi le *temps* joue-t-il dans le modèle d'ORCUTT un rôle qu'il n'a pas dans le nôtre, ce qui conduit les auteurs à proposer une procédure générale de formation des couples, pour une unité de temps donnée, qui est sensiblement plus coûteuse que celle à laquelle nous nous sommes arrêté (*). Néanmoins, les

(*) L'ensemble de la population avec les caractéristiques des individus est stockée dans un fichier sur bande magnétique. Lors de la lecture de la bande, pour le premier célibataire lu, on détermine par tirage au hasard s'il se mariera au cours du mois. Si le résultat est *oui*, un second tirage définit les caractéristiques de l'épouse ; le célibataire est alors rangé dans un fichier d'attente, et la lecture de la bande continue. Le célibataire restera en attente jusqu'à ce qu'on ait trouvé une femme célibataire correspondant aux caractéristiques recherchées. Si l'unité suivante est une femme célibataire, on vérifie d'abord si ses caractéristiques correspondent à celles que l'on recherche. Si oui, elle est unie à l'homme en attente ; sinon, on détermine par tirage si elle se mariera "de sa propre initiative" au cours du mois. Si le résultat est *oui*, un nouveau tirage déterminera les caractéristiques de l'époux recherché, et la postulante est rangée dans le fichier d'attente. Cette procédure est utilisable en raison de l'étendue de l'univers de sélection (toute la population), comparée au petit nombre de candidats au mariage chaque mois. Mais les manipulations de bandes qu'elle nécessite la rendent assez coûteuse (Il pourrait être intéressant par contre d'étudier le coût du même processus appliqué à des données stockées sur disque).

mécanismes de choix qu'il analyse méritent notre attention. Nous écarterons la troisième procédure en raison des difficultés de sa mise en oeuvre, et nous analyserons les deux premières qui, formellement, sont équivalentes. La première procédure revient à prendre le premier homme de l'isolat, lui faire choisir une épouse, et passer au suivant jusqu'à épuisement de l'isolat ; mais si nous agissons ainsi, nous donnons systématiquement au premier une plus grande latitude de choix, et nous infligeons l'inévitable "mariage forcé" systématiquement au dernier. On peut remédier à cela en tirant au hasard l'individu dont ce sera le tour de se marier ; on obtient ainsi une version satisfaisante de cette procédure. La seconde procédure est un peu plus complexe. Lorsque le premier homme de l'isolat a choisi son épouse, la première femme restante choisit son époux ; puis le premier homme restant choisit son épouse, etc. jusqu'à épuisement de l'isolat. Si le dernier individu ayant choisi est un homme, dans l'isolat suivant c'est la première femme qui choisira la première son époux. Cette seconde procédure ne supprime pas l'inconvénient que nous avons souligné à propos de la première procédure, mais l'atténue sensiblement. En effet, au fur et à mesure que la procédure d'appariement se déroule, la probabilité des individus restants d'être désignés au prochain tour pour se marier, tendent à être égales pour chaque individu. Au départ, dans un isolat donné, seuls les hommes et les femmes occupant les premiers rangs avant la médiane (les 50 % classés en tête) ont une chance de pouvoir choisir leur conjoint. Mais lorsque c'est par exemple à un homme de choisir, il peut parfaitement prendre pour épouse une femme située dans la première moitié de l'isolat, retirant à celle-ci la possibilité de choisir son époux, et donnant par contre cette possibilité à la première femme située au delà de la médiane. Ce processus, en se répétant, et ce aussi bien pour les hommes que pour les femmes, tend ainsi à égaliser les chances. Néanmoins, si l'on veut que les probabilités de choisir soient

réellement égales pour tous les individus de l'isolat, il nous faut tirer au hasard également l'individu qui choisit. Après cette modification, les deux procédures deviennent équivalentes sur le plan pratique, la seconde étant légèrement plus complexe. Nous avons cependant opté pour la seconde procédure, qui a le mérite d'avoir déjà été testée par ORCUTT, et nous l'avons modifiée en y ajoutant le tirage au sort de l'individu "actif".

e) La précision des résultats.

Les simulations stochastiques font appel à la méthode dite de Monte-Carlo (cf. 161, Pp. 178-181), qui repose sur une série de tirages "au hasard". Aussi, si l'on répète la même simulation, dans des conditions identiques, les résultats seront légèrement différents d'un passage à l'autre. Le meilleur moyen d'obtenir des résultats stables est de répéter l'expérience, de manière à pouvoir évaluer l'espérance mathématique des paramètres mesurés. En effet, la complexité de la formulation sous une forme analytique des modèles de simulation interdit pratiquement toujours le calcul *a priori* des limites de variation des paramètres de sortie. Il sera par conséquent nécessaire de procéder à quelques essais répétés pour une valeur donnée des paramètres d'entrée, afin de mesurer la dispersion des valeurs trouvées expérimentalement pour le taux d'homogamie. Ces essais nous permettront de décider s'il convient de répéter systématiquement toutes les expériences prévues : dans l'affirmative, un problème de coût se poserait, qui risquerait de nous obliger à réduire notre plan d'expérience.

4.2.3. La dynamique du modèle.

Le fonctionnement de ce modèle, tel que nous l'avons décrit, correspond à l'organigramme de la figure 9. Ce fonctionnement peut être décrit en 21 phases principales successives correspondant chacune à une case de l'organigramme. Rappelons au préalable que les variables prises en compte par ce modèle sont, au niveau des couples, la *date du mariage* et la *localité de résidence du futur mari*, et au niveau individuel l'*origine sociale*, le *niveau d'instruction*, la *religion*, et l'*importance de la localité de naissance*.

1^{ère} phase : *constitution des isolats* par tri de l'ensemble des couples selon la date du mariage et le lieu de résidence du mari. Les effectifs des isolats sont vérifiés ; s'ils étaient égaux ou supérieurs à 10 % de l'effectif total, il y aurait lieu de reconsidérer le degré de précision sur les variables, pour obtenir une partition plus fine. Comme la variable *localité de résidence du mari* est utilisée avec son degré de finesse maximum, il faudrait agir sur la date du mariage, en divisant une tranche de 5 ans en deux. Après vérification de la dimension des isolats, les couples sont dissociés afin de constituer dans chacun deux sous-populations sans lien entre elles.

2^{ème} phase : *positionnement de l'"inverseur"*. Cet "inverseur" est destiné à régler l'alternance *homme/femme* pour le choix du conjoint ; après chaque mariage, il sera placé sur la position opposée à celle qu'il avait précédemment (cf. *phase 14*). Nous supposons que nous le plaçons au départ de la position *homme*.

3^{ème} phase : *positionnement du premier modèle de décision*, correspondant au contenu de l'attitude endogamique. Ce sera par exemple un modèle à une variable, la *localité de naissance*.

4^{ème} phase : "initialisation" du degré d'intensité de l'attitude. Ce degré est exprimé par une valeur entière, dont le signe se réfère au contenu de l'attitude (négatif si exogamique), et la valeur absolue au nombre d'"essais" autorisés au postulant. La première valeur pourra correspondre à l'intensité la plus élevée d'exogamie, par exemple -10.

5^{ème} phase : contrôle du signe du degré d'intensité. Si le signe est positif, nous sommes dans le cas de l'endogamie, et nous sautons à la phase 8. Sinon, nous passons aux phases 6 et 7.

6^{ème} phase : calcul du modèle exogamique par négation du modèle endogamique.

7^{ème} phase : calcul du nombre d'essais autorisés au postulant. Il suffit pour cela de prendre la valeur absolue du degré d'intensité.

8^{ème} phase : détermination de l'individu qui choisit. La position de l'"inverseur" indique s'il doit être recherché dans la sous-population des hommes ou dans celle des femmes. A l'intérieur d'une sous-population, les individus sont rangés dans un ordre arbitraire, mais constant ; chaque individu possède ainsi un numéro d'ordre. Un tirage au hasard permet de désigner l'individu dont c'est le tour de choisir un conjoint.

9^{ème} phase : tirage au hasard d'un conjoint. Si l'individu qui choisit est un homme, il tire un numéro correspondant au numéro d'ordre de la femme qu'il peut épouser. Si c'est une femme qui choisit, elle tire au hasard un numéro d'ordre correspondant à un homme.

10^{ème} phase : choix ou rejet du futur conjoint selon le modèle de décision. La comparaison des caractéristiques des futurs époux permet de décider si le mariage est conforme aux critères de

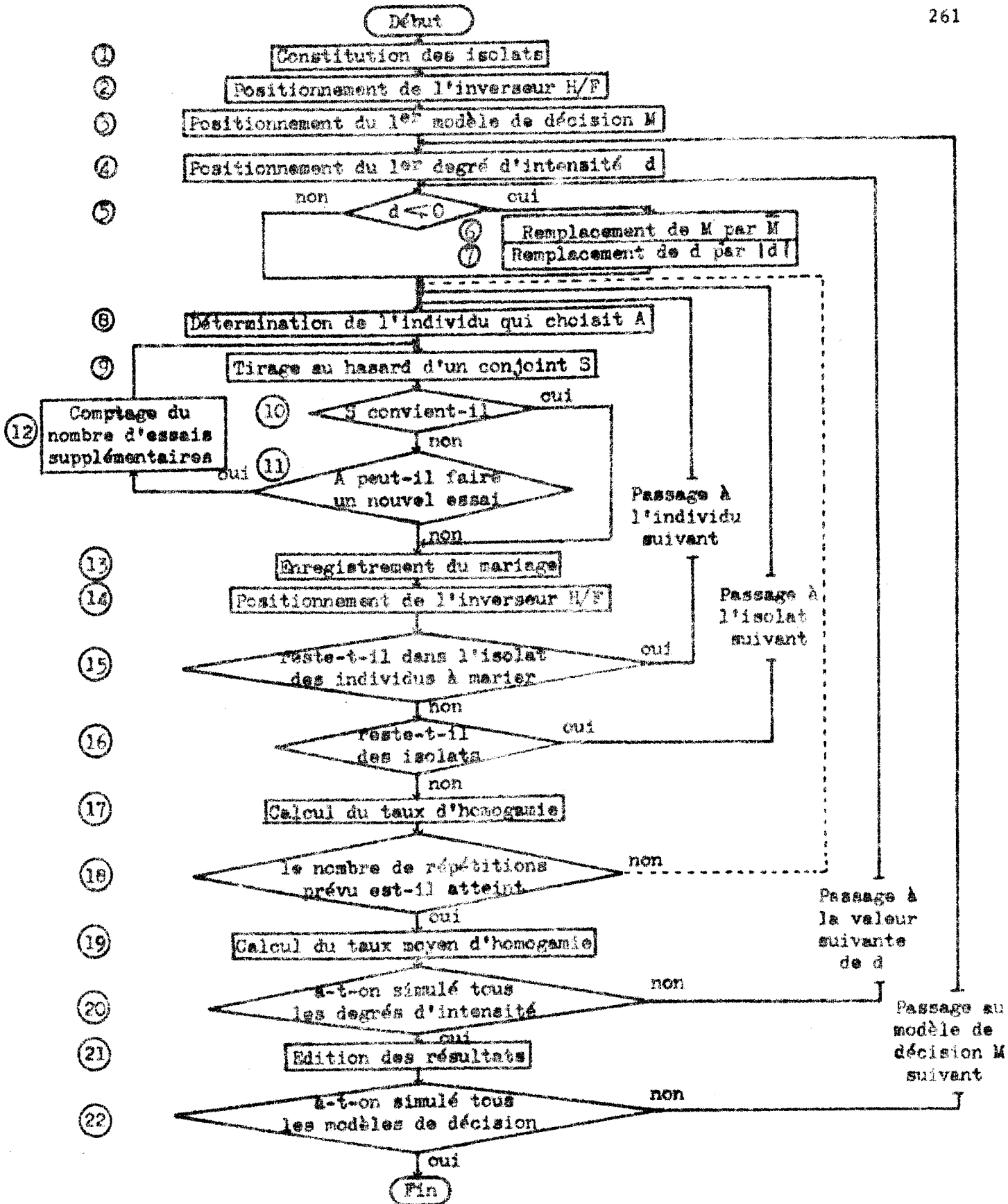


Fig. 9 Organigramme simplifié de la simulation de l'influence sur l'homogamie de divers comportements endogamiques ou exogamiques.

choix, ou non. S'il est conforme, le mariage est prononcé, et nous sautons à la phase 13. S'il ne l'est pas, le couple est mis en attente.

11^{ème} phase : *vérification du droit de l'individu qui choisit à tenter un nouvel essai*. Selon le degré d'intensité de l'attitude, le postulant au mariage a droit, s'il n'est pas satisfait, à tenter un certain nombre d'essais (correspondant à la valeur absolue du degré d'intensité). La comparaison de ce nombre maximum, avec le nombre réel d'essais qu'il a tentés en plus du premier choix, indique s'il peut encore tenter sa chance. S'il n'a plus droit à une nouvelle tentative, le couple en attente est marié "de force", et nous sautons à la phase 13. S'il reste encore un "crédit" d'essais, le couple est dissout, et le partenaire choisi retourne dans l'univers de choix constitué par la sous-population dont il provient.

12^{ème} phase : *comptage du nombre d'essais supplémentaires octroyés au postulant*. Ce comptage est destiné à permettre le contrôle du nombre d'essais (phase 11). Nous revenons ensuite à la phase 9 (tirage d'un nouveau partenaire).

13^{ème} phase : *enregistrement du mariage* pour permettre les calculs ultérieurs (caractéristiques croisées des conjoints), et élimination des deux individus mariés dans l'isolat.

14^{ème} phase : *positionnement de l'"inverseur"* à la position *femme* si c'est l'homme qui a choisi lors du dernier mariage, à la position *homme* dans le cas contraire.

15^{ème} phase : *contrôle du nombre d'individus restant à marier*. S'il reste encore des individus à marier dans l'isolat, nous revenons à la phase 8 pour l'individu suivant.

16^{ème} phase : *contrôle du nombre d'isolats au sein desquels les mariages ont été simulés*. S'il reste encore des isolats, nous reve-

nous à la phase 8 pour le premier individu de l'isolat suivant.

17^{ème} phase : calcul du taux d'homogamie pour chaque variable pour l'ensemble de la population.

18^{ème} phase : contrôle du nombre de répétitions de l'expérience nécessaires pour atteindre le degré de précision souhaité dans les résultats. Si le nombre prévu n'est pas atteint, nous revenons à la phase 8, pour le premier individu du premier isolat.

19^{ème} phase : calcul des taux moyens d'homogamie et de la dispersion des valeurs obtenues au cours de l'ensemble des répétitions. Les phases 18 et 19 pourront être supprimées si les premiers essais montrent que les résultats ont une stabilité satisfaisante.

20^{ème} phase : contrôle des degrés d'intensité simulés. Pour un modèle de décision donné, nous avons envisagé de simuler divers degrés d'intensité. Si tous les degrés n'ont pas été simulés, nous passons au degré suivant, et nous revenons à la phase 5.

21^{ème} phase : édition des résultats pour un modèle de décision. Si tous les degrés d'intensité ont été simulés pour un modèle donné, il est possible d'éditer un tableau des variations des taux d'homogamie (pour chacune des quatre variables) selon le degré d'intensité de l'attitude endogamique ou exogamique.

22^{ème} phase : contrôle des modèles de décision simulés. S'il reste des modèles de décision à simuler, nous passons au modèle suivant et retournons à la phase 4. Sinon, la série de simulations est terminée.

Dans le chapitre suivant, nous examinerons les procédures techniques permettant la réalisation, sur ordinateur, des opérations décrites dans cet organigramme.

5 - LA PROGRAMMATION
DU MODELE SUR ORDINATEUR

L'élaboration du modèle, telle que nous l'avons présentée au chapitre 4, est un travail qui incombe à l'*analyste* ; le rôle de celui-ci est en effet de formuler le problème posé sous une forme opératoire, avant de le soumettre au *programmeur d'études*. Ce dernier procède alors à une étude poussée du modèle, met au point les procédés techniques qui permettront son exécution sur ordinateur, et établit un schéma détaillé des opérations que la machine devra effectuer. A partir de cet organigramme détaillé, le *programmeur d'application* écrira le programme qui sera ensuite donné à la machine. Une telle division du travail est possible dans les grands centres de calcul, qui disposent de ressources en personnel qualifié et de moyens financiers importants. Le chercheur isolé par contre se trouve souvent dans l'obligation de prendre en charge lui-même le travail de programmation. C'est pourquoi nous décrivons dans ce chapitre les problèmes techniques ou même matériels que nous avons dû résoudre : ces problèmes sont ceux que rencontre actuellement tout sociologue voulant réaliser une expérience de simulation sur ordinateur. De plus, même si dans l'avenir le sociologue se trouve intégré au sein d'équipes pluridisciplinaires incluant des informaticiens, il lui faudra avoir une connaissance au moins sommaire des problèmes techniques posés par une simulation stochastique à partir de données empiriques, et des solutions qui peuvent leur être apportées.

Ce travail de programmeur ne peut être mené à bien que si l'on sait avec précision sur quel type de machine le programme sera exécuté. Pour réaliser notre expérience de simulation, nous avons eu accès au centre de calcul du C.N.R.S. (*). Ce centre dispose d'un *Control Data 3.600*, et d'un *IBM 360/75* couplé avec un *IBM 360/50*. Par rapport aux problèmes que le centre traite habituellement, notre simulation représente un travail relativement modeste. Aussi, pouvait-elle être réalisée indifféremment sur l'une ou l'autre de ces machines, malgré les différences importantes de celles-ci (**). Notre choix s'est porté sur le *Control Data 3.600*, en raison de l'expérience que nous

(*)

Centre de Calcul de l'Institut Blaise Pascal, devenu en 1968 le *Centre Inter-Discipline Régional de Calcul Electronique (CIRCE)*, à ORSAY.

(**)

Les principales différences sont des différences de *génération*, car ces machines comptent parmi les plus grosses de la 2e génération pour le *Control Data 3.600*, et de la 3e génération pour l'*IBM 360/75*. Les caractéristiques essentielles de ces ordinateurs sont :

<i>Caractéristiques</i>	<u><i>Control Data 3.600</i></u>	<i>IBM 360/75</i>
<i>capacité de mémoire centrale</i>	64 K mots de 48 bits soit plus de 3 millions de positions binaires	1024 K octets (8 bits) soit plus de 8 millions de positions binaires
<i>cycle de base</i>	1,5 micro-seconde	0,75 micro-seconde
<i>périphériques</i>	3 dérouleurs de bandes	8 dérouleurs de bandes 8 unités de disques 2 tambours
<i>Mode d'utilisation</i>	traitements par lots (<i>batch processing</i>)	multi-traitement

avons acquise depuis novembre 1966 sur un ordinateur du même modèle.

L'élaboration de notre programme a comporté les trois phases suivantes :

- une phase de recherches préalables, touchant au déroulement général du modèle, et aux langages de programmation ;
- l'étude détaillée du modèle, afin d'utiliser au mieux les possibilités offertes par le type de machine choisi ;
- l'écriture du programme.

5.1. QUELQUES CHOIX PREALABLES.

Avant de pouvoir procéder à l'examen détaillé du modèle, nous avons dû prendre des décisions sur les problèmes suivants :

- comment présenter les données ?
- comment réaliser sur un ordinateur numérique un dispositif de "tirage au hasard" ?
- quel langage de programmation choisir ?

Nous décrivons ci-après la procédure que nous avons suivie pour résoudre ces questions.

5.1.1. La présentation des données.

Les réponses au questionnaire mis au point par Alain GIRARD au cours de ses recherches sur le choix du conjoint ont été reportées sur cartes perforées en vue de leur exploitation mécanographique. Le fichier auquel nous avons eu accès comportait par conséquent des cartes

codées en fonction d'un traitement sur trieuse et tabulatrice, et non sur ordinateur, et classées en vue d'un dépouillement d'enquête, et non d'une simulation. Nous avons dû procéder à la recodification des données (ce qui nous a conduit à éliminer certains couples), et au reclassement des cartes.

a) Recodification des données.

Avant l'introduction du dépouillement d'enquêtes sur ordinateur, il était usuel de condenser au maximum l'information codée, de manière à réduire le nombre de cartes nécessaires par questionnaire ; l'idéal, consistant à faire tenir toutes les réponses à un questionnaire sur une seule carte. Pour atteindre ce but, il était courant d'utiliser les deux lignes de perforations hors-texte de la carte, portant ainsi à douze perforations la capacité d'une colonne ; il arrivait également que l'on code dans une même colonne plusieurs réponses, soit qu'il s'agisse d'une question à réponses multiples, soit que l'on ait codé à la suite les unes des autres les réponses à plusieurs questions. En utilisant ces divers artifices, on parvenait à accroître dans des proportions considérables la quantité d'information codable sur une seule carte : la capacité théorique maximum passait de 80 codes décimaux à 960 codes binaires, ce qui revient à tripler environ celle-ci (*). En pratique il est vrai, la quantité d'information contenue dans une seule carte restait très inférieure à ce maximum. La raison de ces efforts

(*)

En utilisant un code décimal, la capacité d'une colonne, c'est-à-dire la quantité d'information qu'elle peut contenir, est de $\log_2 10 = 3.322$ bits. Si l'on recourt à un code binaire, en utilisant les 12 zones de perforation possibles, la capacité d'une colonne s'élève par définition à 12 bits ; on multiplie ainsi la capacité d'une colonne par 3,6. Si l'on raisonne au niveau de la carte toute entière, celle-ci peut contenir soit un nombre décimal de 80 chiffres, soit un nombre binaire de $12 \times 80 = 960$ chiffres ; le codage binaire fait passer la capacité globale maximum d'une carte de $80 \log_2 10 = 265,754$ bits à 960 bits.

pour condenser l'information était d'ordre technique. Sur une trieuse, il est possible de compter ou de sélectionner les cartes selon une perforation donnée dans une colonne donnée, c'est-à-dire de travailler au niveau du *bit*. En procédant à plusieurs passages successifs, on peut compter ou sélectionner les cartes à partir d'une combinaison de perforations : cela permet de compter ou sélectionner les individus en fonction de leurs réponses à plusieurs questions considérées simultanément, et de constituer des tableaux croisés d'effectifs. Mais si deux informations relatives à un même individu ont été codées sur deux cartes différentes en raison de la longueur du questionnaire, le dépouillement croisé des effectifs de réponses aux deux questions correspondantes pose des problèmes délicats (*) ; c'est pourquoi on s'efforce de condenser au maximum les informations, si possible sur une seule carte par individu.

L'introduction de l'ordinateur a modifié sensiblement les données du problème. En principe, un ordinateur travaille au niveau du *bit*, et est par conséquent capable de lire n'importe quelle configuration de perforations : les informations lues sont mémorisées

(*)

La solution la plus fréquemment adoptée pour réaliser de tels croisements consiste à reporter sur la 2e carte l'information nécessaire contenue sur la première carte. Dans ces conditions, l'établissement du plan de cartes suppose que l'on ait établi un plan de dépouillement définitif, et que l'on s'interdise pratiquement tout dépouillement ultérieur (à moins de créer de nouvelles cartes). Afin d'atténuer cet inconvénient, il est prudent de réserver sur chaque carte une série de colonnes constituant une "zone de travail", qui restent vierges lors de la perforation des cartes, et qui permettent d'ajouter ultérieurement les informations nécessaires (simples reports, recodification de variables, indices, etc.).

et traitées sous forme binaire. L'existence d'une fonction de mémorisation supprime les contraintes imposées par la mécanographie classique : si deux informations relatives à un même individu sont codées sur deux cartes différentes, elles seront lues successivement par la machine, mais seront ensuite présentes simultanément dans la mémoire centrale de l'ordinateur, autorisant ainsi toutes les comparaisons et tous les traitements désirés. Il est possible d'utiliser autant de cartes par questionnaire qu'il est nécessaire pour coder les informations recueillies avec toute la précision souhaitée. En pratique, la lecture des perforations et leur interprétation au niveau du bit n'est possible que dans deux cas : ou bien le programme a été écrit dans le langage propre à l'ordinateur sur lequel on travaille (c'est-à-dire en *langage d'assemblage*, qui est un fidèle reflet du langage interne), et l'on peut utiliser toutes les ressources de la machine ; ou bien il a été écrit dans un langage évolué conçu spécialement pour traiter ce type de problèmes (*). Si l'on ne se trouve dans aucun de ces deux cas, on est obligé d'écrire le programme dans un langage évolué conçu généralement pour le calcul scientifique ou la gestion. Ce type de langage ne permet pas de lire n'importe quelle configuration de perforations ; aussi est-il prudent de coder les informations suivant le code décimal admis par tous les langages, c'est-à-dire en s'interdisant l'usage des perforations multiples et des perforations hors-texte.

(*) Il existe deux langages évolués conçus spécialement pour le dépouillement d'enquêtes : *NLT*, et *DAPHNE*. Il existe quelques programmes de dépouillement standards écrits en assembleur, tels que *PVS* ou *PRALINE* (cf. 98). Mais il ne peut être question, pour dépouiller une enquête, d'écrire soi-même un programme en *langage d'assemblage*, en raison du temps nécessaire pour écrire un programme, même simple, dans ce type de langage. Or, tous les programmes standards écrits en assembleur et tous les langages de dépouillement d'enquête sont la propriété de centres de calcul privés ; ils ne peuvent généralement être utilisés que dans ces centres de calcul, et à titre onéreux (en plus du coût du temps de traitement).

Dans le cas de l'analyse secondaire sur ordinateur d'une enquête dont les données ont été codées en vue du dépouillement sur trieuse, il est par conséquent nécessaire de procéder à la *recodification* des informations utilisées. Les réponses au questionnaire sur le choix du conjoint ont été codées sur deux cartes ; la première (grise) correspond aux renseignements factuels relatifs à chacun des conjoints ; la seconde (rose) est plutôt consacrée aux circonstances du mariage et aux opinions et attitudes. Les perforations hors-texte sont fréquemment utilisées ; les perforations multiples par contre sont rares. Toutes les informations que nous avons retenues dans notre modèle sont codées sur la carte n° 1. Nous avons limité notre recodification aux six variables retenues dans le modèle ; cela nous a obligé à recodifier les réponses à onze questions pour chaque couple : *commune de naissance* du mari (colonne 8) et de la femme (col. 40), *niveau d'instruction* du mari (col. 16) et de la femme (col. 47), *religion* du mari (col. 19) et de la femme (col. 50), *localité de résidence* du mari (col. 20) et de la femme (col. 51), *profession du père* du mari (col. 32) et de la femme (col. 63), *date du mariage* (col. 71-72). Deux problèmes se sont alors posés : transformer les codes hors-texte (Y,X) en codes numériques, et regrouper les modalités de réponses qui ont été distinguées au niveau de l'enquête, mais confondues dans l'analyse des résultats. Par exemple, pour toutes les variables relatives à l'*importance d'une agglomération* (de naissance, de résidence), nous avons codé 0 à la fois les communes indéterminées (X), étrangères (Y), ou des territoires d'outre-mer (codés 7,8 ou 9). Pour la *religion*, nous avons regroupé catholiques non pratiquants (2) et catholiques indéterminés (3) ; protestants, juifs et autres religions (4 à 9 et Y) ; sans religion (0) et indéterminés (X). Nous nous sommes en cela conformé à l'usage de l'auteur lui-même dans *Le choix du conjoint*.

Il aurait été possible d'écrire en langage d'assemblage un programme de recodification, qui aurait été exécuté par l'ordinateur. Pour des raisons d'économie, nous avons préféré procéder à la recodification de manière plus artisanale, en utilisant le matériel mécano-graphique en libre-service du centre de calcul. Nous avons utilisé la trieuse et la reproductrice pour mener à bien cette opération. Nous avons d'abord reproduit l'ensemble des premières cartes du fichier initial, en laissant cependant une zone vierge sur les cartes ainsi obtenues, de manière à disposer d'une zone de travail pour les recodifications. Après avoir ensuite procédé au tri de ce nouveau fichier selon la première variable à recodifier, nous avons regroupé les cartes devant porter le même code pour cette variable, puis nous avons perforé le nouveau code dans une colonne de la zone de travail. Cette perforation pouvait-être exécutée en série sur la reproductrice en utilisant une "carte maîtresse". Nous avons répété l'opération pour chacune des variables retenues. Nous avons en outre procédé à une recodification supplémentaire, en affectant un numéro d'ordre à chaque couple selon la date de son mariage ; le numéro *un* correspondant aux mariages d'avant 1920, le numéro *deux* aux mariages célébrés entre 1920 et 1924, et ainsi de suite pour chaque tranche de cinq ans jusqu'en 1959. Par mesure de sécurité, nous avons enfin vérifié, à l'aide d'un programme standard (écrit en langage d'assemblage), qu'aucune des variables recodifiées ne comportait de perforations multiples.

b) Sélection des couples.

En procédant à cette série de recodifications, nous avons constaté que certaines personnes interrogées n'avaient pas répondu à toutes les questions qui nous intéressaient ; après réflexion, nous avons décidé d'éliminer de notre échantillon tous les couples pour

lesquels nous ne disposions pas de toutes les informations requises. Nous avons pensé qu'une étude expérimentale sur l'homogamie serait faussée si l'on acceptait l'éventualité d'unions entre par exemple un homme d'origine sociale connue avec une femme dont l'origine sociale ne serait pas précisée, ou entre deux personnes d'origine sociale indéterminée : de tels mariages sont-ils homogamiques, ou hétérogamiques ? Il nous était évidemment impossible de décider, et opter pour l'une ou l'autre réponse revenait à accepter de surestimer ou sous-estimer systématiquement l'homogamie "vraie". D'autre part, nous avons dû éliminer également les couples dans lesquels l'un au moins des conjoints résidait hors de France avant son mariage. En effet, dans de tels cas, la dimension de l'agglomération de résidence n'a pas été mentionnée, et les mariages entre par exemple une personne résidant à l'étranger et une personne résidant en France dans une commune rurale auraient soulevé les mêmes problèmes que les réponses incomplètes. En sélectionnant les seuls couples ayant fourni des réponses "utiles" à toutes les variables retenues, nous avons ramené l'échantillon de 1646 couples à 1602.

Il y aurait lieu de s'interroger sur le bien fondé d'une telle amputation. En supprimant de notre échantillon près de 3 % de couples, on introduit un biais qui altère son caractère représentatif. Si notre expérience avait eu pour but de reconstituer artificiellement l'échantillon d'Alain GIRARD, afin de tirer de cette reconstitution des indications sur la manière dont les mariages se réalisent effectivement, et d'étendre nos conclusions sur le mécanisme du mariage à l'ensemble de la population française, ce biais aurait risqué de limiter la portée de nos conclusions. Encore pourrions-nous montrer que les incidences du biais introduit sont en fait atténuées. Nous pouvons en effet faire l'hypothèse que les éliminations dues aux réponses indéterminées,

d'ailleurs peu nombreuses, se sont produites de manière aléatoire, sans affecter la représentativité de l'échantillon. En ce qui concerne l'élimination des couples dont l'un au moins des conjoints résidait hors de France, elle porte sur 37 couples (cf. 93, Pp. 58-59, tableaux II et III) ; son influence sur la composition de l'échantillon est donc particulièrement sensible, tant en raison de son importance numérique que de son caractère systématique. Mais ce biais systématique nous conduirait simplement à définir d'une autre manière la population étudiée. Nous avons énuméré (§ 4.1.1.b) les restrictions qui définissent la population touchée par l'enquête ; à celles-ci s'ajoute d'ailleurs la condition implicite que le couple résidait en France au moment de l'enquête (1959), ce qui inclut dans l'échantillon les couples étrangers (0,5 %) ou mixtes (2,8 %) résidant en France, et exclut les couples français ou mixtes résidant à l'étranger. Notre sélection a pour effet d'éliminer de l'échantillon, et par conséquent de la population étudiée, tous les couples dont l'un des conjoints résidait hors de France avant le mariage. Si nous considérons la population définie par les conditions citées par Alain GIRARD, auxquelles nous ajouterons l'obligation pour les conjoints d'avoir résidé tous deux en France immédiatement avant leur mariage et d'y résider en 1959, nous pouvons dire que notre échantillon est représentatif de cette population (*).

La série de simulations que nous entreprenons constitue en quelque sorte une expérience de laboratoire, et la sélection de la

(*)

En réalité, l'échantillon d'Alain GIRARD ayant été sélectionné par quotas, et non purement au hasard, il y aurait lieu de vérifier que l'élimination des couples dont au moins l'un des conjoints résidait hors de France n'a pas introduit un biais systématique au niveau des quotas (dans l'hypothèse par exemple où tous les étrangers recensés seraient des manoeuvres, ou des ouvriers agricoles peu qualifiés). Si ce biais existait, et avait une valeur significative, il faudrait alors redresser l'échantillon.

population étudiée doit obéir aux règles d'un plan d'expérience, plutôt qu'à celles d'un plan d'échantillonnage. En effet, notre objectif principal n'est pas l'étude d'une population, ni la description d'un fait social, mais l'*analyse expérimentale de certains aspects d'un processus dynamique*. Ce que nous recherchons n'est pas en soi un ensemble de relations entre des individus, mais un ensemble de *relations entre des variables*. C'est pourquoi nous pouvons limiter notre recherche à un petit nombre de variables sélectionnées *a priori*, et non prendre en compte l'ensemble des variables ayant une influence sur le mariage. Si le modèle de simulation que nous nous proposons d'utiliser constituait un modèle complet du phénomène étudié, nous n'aurions besoin d'aucune donnée empirique, et nous pourrions nous en tenir à un plan d'expérience strict, puisque toutes les variables et leur interactions seraient contrôlées. Dans un modèle incomplet du type de celui que nous utilisons, nous pouvons définir de manière rigoureuse les variables et les états que celles-ci peuvent prendre, mais nous ignorons la nature des relations entre ces variables. Les données empiriques constituent en quelque sorte la *description brute* (ni condensée, ni formalisée) *des relations* qui unissent les variables entre elles, compte-tenu des interactions de tout ordre qui peuvent exister. Pour que ces données remplissent le rôle qui leur est assigné, il est nécessaire qu'elles renferment un échantillon représentatif des *relations* entre les variables. En éliminant les couples dont un des conjoints résidait hors de France, nous avons donné une définition plus restrictive à notre étude, et supprimé une variable parasite apparentée à la région, et non à la taille d'agglomération. Il eut peut-être suffi, dans les cas d'ailleurs minoritaires (cf. 93, p. 57, tableau I) où l'un des conjoints seulement résidait hors de France, d'éliminer de l'échantillon ce seul conjoint. Mais outre qu'une telle procédure risquait d'aboutir à la constitution d'isolats ne contenant pas un nombre égal

d'hommes et de femmes, elle influait sur la nature des relations entre les variables définissant les couples. Par contre, l'élimination au niveau des couples de la variable "pays de résidence" n'altère pas l'échantillon de *relations entre les variables retenues*, pour la population d'individus résidant en France et correspondant aux conditions énoncées par Alain GIRARD.

c) Regroupement des individus.

En procédant à la recodification des variables, nous avons modifié l'ordre dans lequel les cartes étaient rangées. Si quelques programmes standard de dépouillement d'enquête (*OSIRIS ; programme de tri croisé de la Maison des Sciences de l'Homme*) imposent avant dépouillement le classement des cartes selon le numéro de questionnaire, ceci à des fins de contrôle, tel n'est pas le cas dans les dépouillements sur trieuse, pour lesquels il n'existe aucune contrainte de ce type. Puisque nous écrivons nous-même notre programme de simulation, c'est à nous qu'il incombe de définir s'il y a lieu de telles contraintes, soit afin de permettre un contrôle, soit pour faciliter l'écriture ou l'exécution du programme. Nous avons décidé de regrouper les couples appartenant à un même isolat, de manière à faciliter l'exécution du programme. Pour cela, nous avons procédé à deux tris successifs, le premier sur la taille de la localité de résidence du mari, le second sur le numéro d'ordre correspondant à l'époque du mariage. Puis nous avons reproduit le fichier ainsi classé, par mesure de sécurité. En effet, si les machines mécanographiques lentes (reproductrice, tabulatrice, etc.) acceptent des cartes légèrement détériorées, les machines rapides (trieuses) et les lecteurs de cartes, qui traitent plus de 1000 cartes à la minute, ont une tolérance très réduite (*) ;

(*) Rappelons que les dimensions des cartes mécanographiques sont définies avec une grande précision : longueur = 187,3 mm \pm 0,1 ; largeur = 82,6 mm \pm 0,1 ; épaisseur = 0,17 mm \pm 0,01, à 20° centigrades et à 65° d'hygrométrie. Un changement important du degré d'humidité ou un léger froissement dû aux passages répétés en machine suffisent à altérer la forme des cartes, et à les rendre impropres au passage sur un lecteur de cartes rapide.

il n'est donc pas rare que des "bourrages" se produisent, arrêtant la machine et détruisant quelques cartes (souvent une dizaine). D'autre part, plutôt que de reproduire intégralement les cartes que nous avons utilisées pour la recodification, nous avons préféré ne reproduire que les colonnes correspondant aux variables retenues, plus le numéro de questionnaire et le numéro de période de mariage. Afin de faciliter un éventuel contrôle visuel, nous avons réparti ces informations sur les cartes d'une manière plus aérée, en laissant au moins quatre colonnes vierges entre chaque colonne utilisée.

Le reclassement des couples par isolat préparait la première phase de l'organigramme général (fig. 9). Il nous a semblé qu'il était préférable, plutôt que de confier cette tâche au programme, de la mener complètement à bien avant le passage sur ordinateur. En effet, la première phase de l'organigramme constitue une phase préalable de la simulation, qui n'est ensuite modifiée à aucun moment de l'exécution du programme. Pour achever cette phase manuellement, il nous suffisait de vérifier si les dimensions des isolats étaient convenables, et d'introduire une marque de séparation entre chacun d'eux. L'effectif des isolats correspond au tableau croisé du "numéro de période du mariage" selon la "localité de résidence du mari" (tableau n° 12). L'isolat ayant le plus gros effectif est celui qui correspond aux couples mariés entre 1945 et 1949, dont le mari résidait dans une commune rurale ; cet effectif (117 couples) est acceptable, puisqu'il ne représente que 7,3 % de l'effectif total de l'échantillon (1602 couples). Par contre, pour la première période (mariages célébrés avant 1920), les effectifs des isolats sont trop faibles pour laisser aux futurs conjoints une latitude de choix suffisante ; nous avons donc décidé de procéder à un regroupement. Comme nous estimions que l'influence de la taille de la commune de résidence sur

Période du mariage	Importance de la localité de résidence du mari						Total
	Moins de 2.000 habitants	De 2.000 à 5.000 hab.	De 5.000 à 20.000 hab.	De 20.000 à 100.000 hab.	Plus de 100.000 hab.	Paris	
1 - avant 1920	24	6	3	6	1	6	46
2 - de 1920 à 1924	55	13	19	19	19	14	139
3 - de 1925 à 1929	56	10	26	18	17	11	138
4 - de 1930 à 1934	65	22	22	19	14	15	157
5 - de 1935 à 1939	69	18	32	31	24	17	191
6 - de 1940 à 1944	69	19	25	21	22	9	165
7 - de 1945 à 1949	117	39	50	49	54	22	331
8 - de 1950 à 1954	76	27	37	33	41	19	233
9 - de 1955 à 1959	67	26	26	30	26	27	202
Total	598	180	220	226	218	140	1602

Tableau n° 12 : Date du mariage selon l'importance de la localité de naissance du mari.

les facteurs étudiés était plus importante que celle de l'époque du mariage, nous avons conservé les six catégories de communes, et regroupé les deux premières périodes en une seule. La taille maximum des isolats ainsi formés ($24 + 55 = 79$ couples) reste inférieure à celle de l'isolat déjà signalé comme ayant le plus gros effectif.

Pour séparer les 48 isolats ainsi constitués, nous avons intercalé entre les paquets de cartes correspondant à un isolat une carte vierge, portant un 5 en colonne 80. Pour signaler la fin du dernier isolat, nous avons ajouté au fichier une carte vierge portant un 9 en colonne 80. Sur les cartes correspondant aux couples, la colonne 80 était vierge de toute perforation. Nous verrons au § 5.2.2.a l'usage de ces cartes de séparation.

5.1.2. La génération de "nombres au hasard".

La simulation du processus de mariage demande au minimum deux tirages au hasard par couple : le premier pour sélectionner le prochain candidat au mariage, le second pour choisir son conjoint. Les techniques d'échantillonnage nous ont familiarisé avec l'utilisation de nombres au hasard pour sélectionner des individus au sein d'une population. Le problème qui est posé est : comment obtenir des nombres au hasard sur un ordinateur numérique ? En théorie, trois solutions sont possibles : utiliser une source de "bruits", lire une bande sur laquelle une table de nombres au hasard a été enregistrée, obtenir par calcul des nombres pseudo-aléatoires. Une source de bruits est par exemple une source radio-active, ou un dispositif électronique dont la sortie est un signal analogique dont

l'amplitude varie de manière imprévisible ; pour être utilisé par l'ordinateur numérique, le signal doit être transformé par un convertisseur analogique-numérique. Le centre de calcul du C.N.R.S. ne possédant pas de tels dispositifs, cette possibilité était exclue *a priori*. L'utilisation d'une liste de nombres au hasard, tirée des tables de TIPPETT ou de FISHER et YATES, et enregistrée sur bande magnétique, ne présente par contre aucune difficulté. Son unique inconvénient est de nécessiter au moins deux opérations de lecture par couple ; les opérations de lecture étant très lentes, relativement à la vitesse des opérations de traitement d'information, cette procédure risque d'allonger très sensiblement la durée d'exécution du programme, et par conséquent le coût de la simulation. L'obtention par calcul de *nombres pseudo-aléatoires* est une solution bien adaptée à la physiologie de l'ordinateur ; la difficulté est de disposer d'algorithmes engendrant des nombres qui puissent effectivement être considérés comme des nombres "au hasard".

C'est à cette dernière solution que nous nous sommes arrêté, à la fois en raison du faible coût de la procédure, et de son élégance. Nous avons passé en revue un certain nombre d'algorithmes proposés pour la génération de nombres pseudo-aléatoires, et les critiques qui leur étaient adressées. Après avoir décrit les principales propriétés que doit présenter une série de nombres "au hasard", nous présenterons trois groupes d'algorithmes, et nous indiquerons les raisons de notre choix.

a) Propriétés des nombres "au hasard".

Vouloir obtenir par le calcul des séries de nombres aléatoires est évidemment une absurdité. Une série de nombres aléa-

toires est une série ordonnée de nombres tels que, connaissant les $n-1$ premiers nombres de la série, je ne puisse pas prévoir le $n^{\text{ième}}$. Or, tout procédé de calcul étant déterminé, son résultat l'est également, et aucun algorithme ne peut produire des nombres "au hasard". Force nous est donc de limiter notre objectif à l'obtention, par le calcul, de séries de nombres qui puissent être considérées d'un point de vue opératoire comme des nombres aléatoires, bien qu'elles soient en réalité parfaitement déterminées. Pour cela, nous devons énoncer un ensemble de conditions auxquelles ces séries de *nombres pseudo-aléatoires* doivent satisfaire.

Les conditions minimum imposées aux nombres pseudo-aléatoires sont (191, Pp. 72-73) :

- *condition 1.* Etant donné l'intervalle défini par les bornes à l'intérieur desquelles doivent se trouver les valeurs prises par les nombres de la série, chacune des valeurs possibles à l'intérieur des bornes doit apparaître avec la même fréquence (distribution rectangulaire).

- *condition 2.* Les corrélations séquentielles entre les nombres (au moins les corrélations d'ordres les moins élevés) doivent être nulles.

Ces conditions sont nécessaires pour garantir que les nombres obtenus peuvent être considérés comme des nombres au hasard. Elles ne sont cependant pas suffisantes, comme le montre l'exemple suivant, dans lequel l'intervalle de variation est $[1,3]$:

1 2 2 3 3 1 3 2 1 1 2 2 3 3 1 3 2 1

Dans cette série, chaque valeur apparaît 6 fois (condition 1), et les corrélations séquentielles d'ordre 2 sont nulles (condition 2), puisque chaque couple de nombres possible apparaît avec la même fréquence. Cependant, la série est évidemment cyclique (ibid., p. 73).

Toute série de nombres pseudo-aléatoires de longueur finie est nécessairement cyclique. S'il était possible de respecter la seconde condition de manière absolue, c'est-à-dire de veiller à ce que les corrélations séquentielles *de tous ordres* soient nulles, il n'y aurait pas de cycles. Mais l'exemple de la série ci-dessus nous montre qu'une telle condition est irréalisable pour toute série de longueur finie. Dans cette série, le cycle est de longueur 9 ; c'est-à-dire que la séquence des 9 premiers chiffres se répète régulièrement. Cette séquence elle-même pourrait être considérée comme aléatoire, puisque chaque chiffre y apparaît avec la même fréquence, et que chaque couple de chiffres y apparaît également avec la même fréquence. Si l'on considère les corrélations séquentielles d'ordre trois, on peut voir sans calcul qu'il n'est pas certain que celles-ci soient nulles : en effet, il y a 3^3 triplets possibles par arrangement avec répétition des trois chiffres {1,2,3} pris 3 à 3. Dans une séquence de longueur 9, qui se répète ensuite, 9 triplets seulement apparaissent réellement ; par conséquent, 9 arrangements possibles sur 27 sont observés, ce qui introduit une contrainte séquentielle (cf. 97, p. 467). On ferait le même raisonnement pour les quadruplets (9 observés pour 81 possibles), etc.. Il est donc inévitable que des contraintes séquentielles existent dans une suite de longueur finie ; c'est-à-dire que l'on peut toujours (au moins en théorie) deviner, avec une probabilité plus grande que le hasard, la $n^{\text{ième}}$ valeur d'une série connaissant les $(n-1)^{\text{ièmes}}$ valeurs qui la précèdent (*). Pour qu'une série soit réellement aléatoire, il faudrait l'augmenter de telle sorte que les contraintes d'ordre élevé elles-mêmes disparaissent : on aboutit ainsi

(*)

A condition évidemment que n soit assez grand. Si des contraintes séquentielles non nulles apparaissent pour tout ensemble d'au moins i valeurs consécutives, il faut pour avoir des chances supérieures à la moyenne de deviner la $n^{\text{ième}}$ valeur que n soit au moins égal à i .

à une suite infinie, qu'il est impossible d'engendrer par calcul (nécessairement déterministe).

S'il est inévitable qu'il y ait des cycles dans une suite de nombres pseudo-aléatoires, il faut si possible supprimer les inconvénients pratiques de ces cycles. Dans le cours d'une expérience de simulation, il est fréquent que l'on procède à plusieurs milliers, voire dizaines ou centaines de milliers, de tirages au hasard (*). Pour que les contraintes séquentielles risquent le moins possible de perturber le déroulement de la simulation en y introduisant des biais systématiques, il faut que le cycle soit le plus long possible. C'est pourquoi on ajoute généralement aux premières conditions énoncées les deux suivantes (156, Pp. 357-358) :

- *condition 3*. La longueur d'une série de nombres pseudo-aléatoires doit être au moins égale à 10^6 nombres.

- *condition 4*. Si l'on considère un intervalle de dimension supérieure à une dimension minima d (de l'ordre de 10^{-6}), la connaissance de l'intervalle dans lequel se trouve le nombre n_{i-1} ne doit pas permettre de prévoir dans quel intervalle tombera le nombre n_i .

A ces considérations s'ajoutent deux impératifs, d'ordre pratique : le procédé de calcul doit être *récuratif*, et le temps de calcul d'un nombre doit être très sensiblement plus court que le temps de lecture d'un nombre sur bande magnétique. Nous passerons en revue maintenant quelques algorithmes devant en principe satisfaire à ces conditions.

(*) Dans une série de simulation de notre modèle, répétant les phases 3 à 17 de l'organigramme (fig. 17) pour 21 degrés d'endogamie ou d'exogamie (de -10 à + 10), il y a en moyenne un peu plus de 130.000 tirages au hasard, dont un peu plus de 100.000 (mais ce nombre varie selon le contenu de l'attitude) pour la procédure de choix du conjoint proprement dite.

b) Génération de nombres pseudo-aléatoires par la technique du "milieu du carré".

Cette technique (*mid-square technique*, ou *middle-of-the-square method*) proposée par Von NEUMAN et METROPOLIS (156, p. 358, notes 4 et 5), est historiquement la première technique utilisée pour engendrer des nombres pseudo-aléatoires. Elle est adaptée aux caractéristiques des premiers calculateurs électroniques, dont la capacité de mémoire était très réduite, et le temps de calcul très long ; elle est facile à programmer comme à exécuter, et utilise peu de place en mémoire (158, p. 55). Son principe est le suivant (191, Pp. 73-74) : à partir d'un nombre initial x_0 , de n chiffres, on calcule son carré x_0^2 , qui a $2n$ chiffres, puis on prend comme valeur de x_1 les n chiffres du milieu de $x_0^2 = 1849$, et $x_1 = 84$. L'inconvénient de la technique du "milieu du carré" est que les séries engendrées se satisfont pas à la condition 1 (distribution rectangulaire des nombres). De plus, selon la valeur initiale x_0 choisie, le cycle est plus ou moins long, et l'on risque dans certains cas d'aboutir à une valeur nulle, qui clôt la série (*). Enfin, TOCHER (191, p. 74) démontre que la troncature de x_i^2 pour donner x_{i+1} introduit un biais qui favorise les valeurs basses de x_{i+1} ; ce biais se répétant à chaque

(*) On s'en convaincra aisément sur l'exemple que nous avons cité :

$$x_0 = 43, \quad x_0^2 = 1849, \quad x_1 = 84, \quad x_1^2 = 7056, \quad x_2 = 05.$$

$$x_2^2 = 0025, \quad x_3 = 02, \quad x_3^2 = 0004, \quad x_4 = 00.$$

METROPOLIS a étudié diverses séquences selon la valeur initiale x_0 , la base de numération, et le nombre de chiffres. Ces séquences peuvent être représentées sous forme de graphe, illustrant le déroulement de la série. Ainsi (158, p. 56), en utilisant des nombres de 2 chiffres écrits en base 7, on obtient selon le nombre x_0 choisi, des chaînes de longueur variée, qui toutes aboutissent soit à un nombre qui se répète indéfiniment (0, 7, 30), soit à une boucle (alternance 14-28).

itération, il sera plus grand pour x_{i+2} que pour x_{i+1} , provoquant ainsi une dégradation continue de la séquence.

Pour tenter de remédier à ces inconvénients, on a proposé la technique du "milieu du produit" (*mid-product method*), qui est une extension de la technique précédente (191, Pp. 74-75). Nous avons deux valeurs initiales, x_1 et x_2 , chacune ayant n chiffres ; la valeur x_3 est constituée par les n chiffres du milieu du produit $x_1 x_2$. Par exemple, pour $x_1 = 25$ et $x_2 = 61$, $x_1 x_2 = 1525$ et $x_3 = 52$; par la suite, nous aurons $x_4 = 17$, $x_5 = 88$, etc.. Cette technique introduit elle aussi un biais, mais moindre que la technique précédente. On démontre d'ailleurs que toutes les transformations que l'on peut utiliser après la troncature conduisent à un biais.

c) Génération de nombres pseudo-aléatoires par l'opération de congruence sur une somme.

Cette technique (*additive congruential method*) recherche l'obtention d'un cycle de longueur maximum. La formule générale est $x_i = x_{i-1} + x_{i-2} \pmod{m}$. Par exemple, en prenant comme valeurs initiales $x_0 = 14$ et $x_1 = 17$, pour $m = 20$, nous engendrons la série : {14, 17, 11, 8, 19, 7, 6, 13, 19, 12, ...}. La difficulté est de choisir correctement les valeurs x_0 , x_1 , et m , de manière à se conformer aux conditions énoncées, et obtenir le cycle de longueur maximum. Parmi les expériences réalisées sur des séries engendrées selon cette technique, nous citerons une recherche de DAVID et RABINOWITZ (décrite par 191, p. 84), qui a consisté tout d'abord à tester la valeur d'une série obtenue avec $x_0 = 0$, $x_1 = 1$, et $m = 2^{44}$. Cette série n'ayant pas donné

satisfaction, les auteurs ont pris comme valeurs des paramètres $x_0 = 2^{42}$, $x_1 = 5^{17}$, $m = 2^{44}$. Ils ont de plus introduit une procédure de sélection dans la série ainsi engendrée, en sautant 1, 2, ou 3 nombres selon que le dernier chiffre du dernier nombre retenu, écrit en base 16, était 1, 2, ou 3 (mod. 3). Cette seconde procédure aurait donné satisfaction.

D'autres perfectionnements ont été apportés à cette technique. Par exemple, au lieu d'additionner x_{i-1} et x_{i-2} , on peut, si l'on dispose d'une série initiale de n nombres, additionner x_{i-1} et x_{i-n} (95, Pp. 164-167). On peut également (95, p. 359) faire la somme des n nombres compris entre x_{i-1} et x_{i-n} . Ces modifications présentent l'inconvénient d'allonger considérablement le temps de calcul, sans présenter d'avantages décisifs sur la technique décrite précédemment.

d) Génération de nombres pseudo-aléatoires par l'opération de congruence sur un produit.

Cette technique (*multiplicative congruential method*) est due pour l'essentiel à LEHMER, qui eut l'idée d'utiliser la théorie des nombres pour découvrir des procédures permettant d'engendrer un cycle de longueur maximum. La formule générale est $x_n = kx_{n-1} \pmod{m}$, où k et m sont des nombres entiers ; on utilise parfois pour le calcul la formule équivalente $x_n = k^n x_0 \pmod{m}$. A titre d'exemple, avec les valeurs $x_0 = 7$, $k = 3$, et $m = 17$, nous engendrons la série : $\{7, 4, 12, 2, 6, 1, 3, 9, 10, 13, 5, 15, 11, 16, \dots\}$; cette série ne satisfait d'ailleurs qu'à la première condition.

Pour obtenir un cycle de longueur maximum, il importe de bien définir les valeurs de k et m . TOCHER (191, Pp. 75-81) et GREENBERGER (156, Pp. 359-370) se livrent à un examen approfondi du problème. Dans le cas de la génération de nombres pseudo-aléatoires sur ordinateur, la longueur N du cycle (correspondant à la première apparition de $x_N = x_0$) est nécessairement inférieure ou égale au plus grand nombre autorisé par la dimension du mot machine. D'autre part, N dépend directement des valeurs attribuées à k et m . La valeur de m doit être choisie de manière à permettre une division rapide (cette opération étant généralement la plus longue opération de base à effectuer). Si le système de numération interne de l'ordinateur est à base b , on montre que m doit être la forme b^p ou b^{p-1} , pour $p = \text{entier}$. La plupart des ordinateurs utilisant un langage binaire, m vaudra 2^p ou $2^p - 1$. Dans le premier cas ($m = 2^p$), la longueur maximum du cycle est $N = 2^{p-2}$, et l'on doit avoir $k = 3 \pmod{8}$ et $x_0 = \text{impair}$. Dans le second cas ($m = 2^p - 1$), le cycle le plus long est obtenu pour $2^p - 1 = \text{nombre premier}$ (ce qui suppose que p soit un nombre dit de MERSENNE), et $N = 2^{p-2}$ également. La valeur de k ne peut être entièrement définie de manière analytique, et doit être sélectionnée par essais et erreurs.

La littérature consacrée à la *Méthode de Monte-Carlo* fournit une grande variété de valeurs du couple (k, m) , susceptibles d'engendrer des séries de nombres satisfaisant aux conditions requises pour être utilisées comme des séries aléatoires (*). En pratique, le problème

(*)

Pour les machines binaires, les valeurs les plus couramment utilisées sont :

$$(k, m) = (5^{13}, 2^{36}), (5^{13}, 2^{39}), (5^{17}, 2^{40}), (5^{17}, 2^{42}),$$

$$(5^{17}, 2^{43}), (2^{18} + 3, 2^{35}), (13^{13}, 2^{31} - 1), (7^{11}, 2^{29} + 1), (23, 2^{35} + 1), \text{etc}$$

Pour les machines décimales :

$$(k, m) = (7, 2^{11} \cdot 5^9), (7^5, 2^{12} \cdot 5^{10}), (7^9, 2^{12} \cdot 5^{10}),$$

$$(7^{13}, 2^{12} \cdot 5^{10}), (7^{17}, 2^{12} \cdot 5^{10}), (3^{19}, 2^{21} \cdot 5^{19}), (23, 10^8 + 1),$$

$$(5 \pmod{10^{12}}, 10^{12}), \text{etc..}$$

consiste à choisir parmi ces couples celui pour lequel le temps de calcul est minimum. Ce problème admet généralement plusieurs solutions, et il faut ensuite choisir, au sein du sous-ensemble sélectionné, un sous-ensemble plus petit en tenant compte des caractéristiques de l'ordinateur utilisé (dimension du mot, dimension minimum de l'intervalle défini dans la condition 4, au chap. 5.1.2.a). Le choix final dépend des tests effectués sur les séries engendrées par les valeurs retenues.

Diverses variantes ont été introduites dans la technique de congruence sur un produit. En particulier :

$$\boxed{x_{n+1} = k x_n + c \pmod{m}} \quad (95, \text{ p. } 167 ; 158, \text{ p. } 56)$$

$$\boxed{x_{n+1} = (4k + 1) x_n + 3k \pmod{m}} \quad (191, \text{ p. } 79)$$

$$\boxed{x_{n+2} = k x_{n+1} + 1 x_n \pmod{m}} \quad (158, \text{ p. } 58)$$

Il ne semble pas que ces variantes aient pour effet d'allonger le cycle dans des proportions importantes.

e) Choix d'une technique.

La technique de génération de nombres pseudo-aléatoires par congruence sur un produit est actuellement la technique la plus éprouvée, et celle qui permet d'obtenir le cycle le plus long. C'est donc cette technique que nous avons envisagé d'utiliser dans notre série de simulations. Le programme correspondant est assez aisé à écrire ; la seule difficulté est l'ajustement des paramètres, qui suppose une série d'essais assez importante, et de nombreux tests

statistiques. Un laboratoire américain, le *Argonne National Laboratory*, utilisant un *Control Data 3.600*, a effectué une recherche sur cet ajustement (9). Le résultat de cette recherche s'est traduit par une fonction qui a été incorporée aux fonctions de bibliothèque du système *Control Data 3600* (52, p. 89), sous le nom de *RANF* (*RAN*dOm-*N*umbers *F*unction). La longueur du mot étant de 48 bits, dont 1 bit de signe, la valeur choisie pour m est $m = 2^{47}$; ainsi, l'opération de congruence s'effectue automatiquement par perte des chiffres de poids fort du nombre engendré. Les autres paramètres sont $k = 5^{15} = 30.517.578.125 = -3 \pmod{8}$, et $x_0 = 2.020.703.135$; ces valeurs répondent bien aux conditions énoncées lorsque $m = 2^p$. Les nombres engendrés sont des nombres entiers répartis uniformément de 1 à 2^{47} ; une option permet de les transformer en nombres compris entre 0 et 1. La longueur du cycle est $N = 2^{45}$ (soit environ 10^{14}). Le programme a été écrit dans le langage d'assemblage de la machine (*COMPASS*), ce qui permet de minimiser son temps d'exécution : celui-ci est de 30 micro-secondes pour un calcul en nombres entiers, et de 38μ s avec la conversion en nombres réels.

Cette procédure a été validée à partir de 45 échantillons de nombres générés, de dimensions comprises entre 10.240 et 1.024.000 nombres (9, p. 5). Sur chaque échantillon, les chercheurs du *Argonne National Laboratory* ont utilisé un test statistique pour vérifier que chaque nombre apparaissait avec une fréquence égale (χ^2 avec $\nu = 1023$). et onze tests sur l'absence de liaisons séquentielles. Tous les résultats ont été satisfaisants.

5.1.3. Le choix du langage de programmation.

Le software du *Control Data 3.600* est assez riche pour nous permettre d'écrire notre programme soit en langage d'assemblage (*COMPASS*), soit en langage scientifique (*FORTRAN*, *ALGOL*), soit dans un langage de simulation (*SIMSCRIPT*). Nous avons d'emblée écarté l'assembleur : ce type de langage se justifie essentiellement pour un programme d'usage courant, pour lequel il est important de minimiser le temps d'exécution. La contrepartie de cet avantage est le temps nécessaire pour écrire et mettre au point le programme. Le problème se ramène donc au choix entre un langage scientifique, ou un langage de simulation. Avant d'entreprendre l'écriture du programme, nous avons dû procéder à une étude comparative des langages de simulation et des langages scientifiques. Si, en fin de compte, nous avons opté pour le *FORTRAN* plutôt que pour le *SIMSCRIPT*, c'est surtout en raison de l'extrême simplicité de notre modèle. Mais il n'est pas certain que, pour une simulation plus complexe, notre choix eut été identique.

Les *langages évolués* de programmation sont des langages qui sont en principe indépendants du type d'ordinateur sur lequel sont exécutés les programmes qu'ils permettent d'écrire. Le prototype des langages évolués à usage scientifique est le *FORTRAN*. La version la plus complète de ce langage, le *FORTRAN IV*, lit et traite des variables numériques et logiques, isolées ou sous forme de vecteurs ou de matrices ; les opérations qu'il permet d'effectuer sont les principales opérations booléennes et arithmétiques, auxquelles il convient d'ajouter un certain nombre de transformations usuelles existant sous forme de fonctions de bibliothèque (congruence, fonctions trigonométriques, racines, etc). L'*ALGOL*, conçu à l'origine pour servir de langage de communication entre mathématiciens se

préoccupant de recherche d'algorithmes, est devenu un langage de programmation dont le domaine d'application est sensiblement le même que celui du *FORTRAN IV* ; son avantage sur ce dernier réside principalement dans sa plus grande rigueur syntaxique, et dans l'existence de procédures récursives. Il existe d'autres langages à usage scientifique. A part le *PL 1*, langage "universel" qui ajoute aux facilités de l'*ALGOL* les possibilités des langages de gestion de type *COBOL*, et quelques possibilités empruntées aux langages de traitement de listes, les autres langages scientifiques sont peu répandus. Le *PL 1* n'est actuellement disponible que sur la gamme *IBM 360*.

Les langages de simulation sont paradoxalement plus nombreux que les langages scientifiques, alors qu'ils sont d'un usage moins courant. En parcourant la littérature spécialisée, nous en avons relevé 17 ; et leur nombre réel est certainement plus élevé encore (*). Ces langages ont pour avantages principaux (172, p. 205) de fournir une structuration standard de l'information, de prendre en charge l'enchaînement des activités, de faciliter le traitement de certaines opérations non numériques (ensembles, files d'attente), et de faciliter l'édition des résultats. Nous avons surtout étudié le *SIMSCRIPT* (53), langage de simulation basé sur le *FORTRAN*. Ce langage appartient à la famille des langages de *micro-programmation*, qui permettent de définir un modèle de manière très détaillée (par opposition aux langages de *macro-programmation*, comme le *SIMPAC*, ou le *GPSS*, qui n'utilisent qu'un nombre restreint de macro-instructions très puissantes). Les notions de base du *SIMSCRIPT* sont les *attributs*, qui servent à décrire l'état du système,

(*) Un panorama de ces langages est fourni par les articles de RENARD et RENAULT (172), et KRASNOW (129). Les langages que nous avons recensés sont : *CLP*, *CSL* (*Control and Simulation Language*), *ESP*, *FCRSIM*, *GASP*, *GPSS* (*General Purpose Systems Simulata*), *GSP* (*General Simulation Program*), *MILITRAN*, *MONTE-CODE*, *QUICK-SCRIPT*, *SIMSCRIPT*, *SIMON*, *SIMPAC*, *SIMTRAN*, *SIMULA*, *SOL*, *WASP*. La plupart de ces langages sont soit basés sur le *FORTRAN* ou l'*ALGOL*, soit dérivés d'un autre langage de simulation.

les *entités*, définies par un groupe d'attributs de même type, et les *événements*, qui introduisent des changements dans les attributs et modifient ainsi l'état du système. Les opérations de base sont la création et la destruction d'entités temporaires, l'apparition des événements, et certains types de calcul arithmétique ou de sélection.

Compte-tenu des possibilités offertes par la machine, et de nos propres connaissances, l'alternative était d'opter pour le *FORTRAN*, ou pour le *SIMSCRIPT*. Ces deux langages ont une syntaxe analogue, et même en partie un vocabulaire commun ; le *SIMSCRIPT* par contre offre plus de possibilités que le *FORTRAN*. Il est certain que les facilités relatives à l'édition des résultats sont négligeables dans notre cas, puisque nous n'envisageons d'étudier que les variations des taux d'homogamie pour quatre variables. D'autre part, nous ne faisons pas à proprement parler évoluer un système dans le temps ; ce paramètre n'intervient dans notre modèle que pour la définition des isolats, qui ne fait pas partie du processus de simulation lui-même. Aussi l'enchaînement de l'unique événement (mariage) de notre modèle ne présente-t-il aucune difficulté. Enfin, les entités que nous utilisons (individus et couples) sont définies par un petit nombre d'attributs (origine sociale, religion, etc.) faciles à manipuler, et nous n'effectuons aucune opération sur les ensembles d'entités. Les avantages du *SIMSCRIPT* sur le *FORTRAN* sont par conséquent négligeables pour notre modèle. Le *SIMSCRIPT* présente d'autre part quelques inconvénients : lenteur dans l'écriture des instructions (une instruction nécessite deux cartes au lieu d'une), et compilation plus longue, donc plus coûteuse. C'est pourquoi nous avons en dernière analyse opté pour la solution la plus simple, consistant à utiliser le *FORTRAN*.

5.2. LA CONCEPTION DU PROGRAMME.

Il est rare que l'on écrive en une seule fois un programme satisfaisant. Dans la plupart des cas, la version initiale du programme comporte quelques erreurs que l'on ne détecte qu'empiriquement, en procédant à son essai : c'est la phase de mise au point (*debugging*). De plus, un programme correct peut très souvent être amélioré, de manière à optimiser l'occupation de la mémoire centrale, ou réduire le temps d'exécution. Ce dernier point est particulièrement important pour un programme devant être utilisé à plusieurs reprises. Dans notre cas, ce problème était relativement secondaire, puisqu'une économie de quelques minutes sur le temps d'exécution constituait le gain maximum que nous pouvions escompter pour l'ensemble de nos simulations. C'est pourquoi nous n'avons pas cherché à réaliser à tout prix le programme optimum, mais à obtenir un programme correct, correspondant exactement à nos objectifs. Ainsi, nous n'avons pas eu besoin de tester à plusieurs reprises l'ensemble du programme, ce qui aurait nécessité plusieurs heures de temps d'ordinateur ; il nous a suffi de mettre au point (et au besoin de perfectionner) séparément chaque élément important, et de procéder à un bref contrôle de l'ensemble, sur un nombre réduit d'itérations, avant de procéder à l'exécution. La version que nous présentons ci-après ne doit donc pas être considérée comme un exemple, mais comme un outil de travail commode, quoique d'usage restreint ; nous aurions certainement pu perfectionner cet outil, à condition d'y consacrer un temps qui aurait été soustrait au temps disponible pour notre recherche proprement dite.

Notre effort a porté dans deux directions principales : la mise au point des mécanismes de constitution des couples, et la conception de la structure générale du programme.

5.2.1. Les mécanismes de constitution des couples.

Les mécanismes du choix du conjoint constituent évidemment les rouages essentiels de notre simulation. Pour mettre au point et perfectionner ces mécanismes, nous avons procédé à plusieurs essais sur un échantillon réduit de vingt couples d'individus, regroupés en deux isolats de dimensions inégales. Chaque essai a duré environ 40 secondes. Les principaux problèmes que nous avons eu à résoudre ont trait au tirage au hasard dans l'isolat, à l'appariement des individus, au reclassement des individus non sélectionnés, et à l'alternance homme/femme de la personne qui choisit.

a) Le tirage au hasard dans l'isolat.

La fonction de bibliothèque *RANF*, que nous avons mentionné au § 5.1.2.d, nous fournit une série de nombres pseudo-aléatoires compris entre 0 et 1. Chaque nombre doit nous permettre de désigner soit le candidat au mariage (phase 8 de l'organigramme de la fig. 9), soit son futur conjoint (phase 9 de l'organigramme). Si notre isolat comporte n couples, nous supposerons que les hommes et les femmes sont séparés, et qu'ils sont numérotés de 1 à n à l'intérieur de chacune des deux sous-populations ainsi constituées. Pour désigner au hasard un individu au sein d'une sous-population, il nous suffira de transformer notre nombre pseudo-aléatoire a , compris dans l'intervalle $]0, 1[$, en un nombre $f(a)$ entier, compris dans l'intervalle $[1, n]$. On voit que cette fonction doit être de la forme $f(a) = \text{partie entière de } an + 1$. En effet, supposons par exemple que l'isolat considéré comporte 4 hommes et 4 femmes, numérotés de 1 à 4 ; par conséquent, $n = 4$. Selon que le nombre a tombera dans l'un des intervalles $]0, 1/n[$, $[1/n, 2/n[$, ..., $[n-1/n, 1[$, la fonction $f(a)$ prendra les valeurs suivantes :

a	$f(a)$
$]0, 0,25[$	1
$[0,25, 0,50[$	2
$[0,50, 0,75[$	3
$[0,75, 1[$	4

En pratique, le tirage au hasard ne se limite pas à la désignation de l'individu dont le numéro a a été tiré. Pour permettre la comparaison des caractéristiques du candidat au mariage à celles du futur conjoint tiré au sort, il faut en quelque sorte extraire ceux-ci de l'isolat. Cela implique qu'à l'intérieur de la sous-population à laquelle il appartient (hommes ou femmes), chaque individu soit représenté par son numéro d'ordre dans l'isolat, et que chacune de ses caractéristiques soit codée. Nous exposerons plus loin (§ 5.2.2.a) la manière dont ces informations sont stockées en mémoire centrale ; il suffit de dire ici que ce problème a été résolu en constituant pour chaque sous-population une matrice rectangulaire, de n lignes (où n = nombre de couples dans l'isolat), et 4 colonnes. Chaque ligne correspond à un individu, celui-ci étant désigné par son numéro de ligne. Chaque colonne correspond à l'une des 4 caractéristiques retenues comme variable de sortie, dans l'ordre suivant : localité de naissance, niveau d'instruction, religion, origine sociale. Ainsi, la case $h_{12,3}$ de la matrice H correspondant aux hommes de l'isolat contient l'information relative à la *religion* (3^e caractéristique) du 12^{ème} homme de l'isolat.

Supposons que ce soit l'homme qui choisisse son épouse. Un premier tirage désignera par exemple le 12^{ème} homme de l'isolat ; nous extrairons alors ses caractéristiques (12^e ligne de la matrice des hommes) que nous conserverons dans une zone particulière de la mémoire, pour procéder ensuite à la comparaison avec celles de la future épouse.

Un second tirage désignera par exemple la 3^e femme de l'isolat : nous extrairons ses caractéristiques (3^e ligne de la matrice des femmes), que nous conserverons dans une zone réservée au partenaire "passif".

b) L'appariement des individus.

Lorsqu'un individu "actif" et un individu "passif" ont été extraits de l'isolat, il convient d'examiner si leurs caractéristiques satisfont au critère de choix ; si oui, le couple est constitué. Sinon, on procède au comptage du nombre d'essais restant à tenter à l'individu "actif", ce qui l'autorise ou non à extraire un nouveau partenaire de l'isolat (phases 10 à 12 de l'organigramme de la fig. 9). Cette procédure pose deux problèmes : celui de la comparaison des caractéristiques des individus en présence en fonction du modèle de décision, et celui du contrôle du nombre d'essais tentés.

Le modèle de décision est exprimé par une identité logique (§ 4.2.2.c), c'est-à-dire par une variable pouvant prendre deux états : *vrai*, ou *faux*. Par exemple, si le critère de choix est le niveau d'instruction, et si le modèle est endogamique, l'identité sera : *niveau d'instruction du partenaire "passif" \equiv niveau d'instruction du partenaire "actif"*. Le mariage sera prononcé si et seulement si l'identité est vraie ; il suffit donc de prévoir dans le programme un aiguillage basé sur cette identité.

Nous avons prévu d'autre part de donner à chaque candidat au mariage la possibilité de tenter un certain nombre de nouveaux essais si le mariage n'est pas prononcé, et d'imposer un mariage forcé avec le dernier partenaire tiré au sort si le nombre d'essais alloué est atteint. Il nous faut donc prévoir dans le programme un "compteur"

qui enregistrera chaque essai du partenaire "actif". Après chaque essai, nous vérifierons que le nombre d'essais tentés est inférieur au nombre d'essais alloués. Si oui, il y aura nouvel essai ; sinon, il y aura mariage forcé. Ce nouvel aiguillage se combinera au précédent, de sorte que l'on ait mariage si et seulement si l'expression $\left[(\text{identité exprimant le modèle de décision} \equiv \text{vraie}) \vee (\text{nombre d'essais tentés} < \text{nombre d'essais alloués}) \right]$ est vraie.

c) Le reclassement des individus non sélectionnés.

Lorsqu'un mariage a été prononcé, il convient d'éliminer de l'isolat les individus sélectionnés. Cette opération ne s'effectue pas automatiquement : en extrayant de l'isolat les deux partenaires possibles, on ne fait que recopier, dans une autre zone de la mémoire, leurs caractéristiques, qui restent stockées dans les matrices décrites au § 5.2.1.a. Rappelons qu'un transfert d'information à l'intérieur de la mémoire centrale ne détruit pas l'information contenue dans la zone de départ, mais réalise seulement un *duplicatum* de cette information, qui est rangé dans la zone d'arrivée du transfert. Ceci présente un avantage certain lorsque le mariage n'est pas prononcé : il n'est alors pas nécessaire de remettre les caractéristiques du partenaire "passif" dans la matrice d'où elles ont été tirées ; il suffit de détruire la copie qui en a été faite, puisque l'original est toujours rangé dans sa matrice d'origine. Par contre, lorsque le mariage est prononcé, il faut détruire le "dossier" correspondant à chacun des époux, afin que ceux-ci ne risquent pas d'être choisis à nouveau. Cette opération doit se faire dans les matrices contenant les caractéristiques des hommes et des femmes constituant l'isolat.

La principale difficulté réside dans l'identité du numéro de ligne de la matrice et du numéro d'ordre dans l'isolat de l'individu correspondant. Supposons en effet que, pour détruire le dossier du $i^{\text{ème}}$ homme de l'isolat, nous remplissons de zéros la $i^{\text{ème}}$ ligne de la matrice des hommes : nous avons bien détruit une partie du "dossier" de cet homme, mais son numéro d'ordre subsiste, et il peut encore par conséquent être tiré au sort.

Pour résoudre ce problème, nous avons envisagé successivement trois solutions. La première consisterait à conserver en mémoire le numéro d'ordre des individus mariés, et à ne pas modifier le contenu des matrices décrivant les membres de l'isolat. A chaque tirage au sort, on vérifierait que le numéro tiré est différent des numéros déjà tirés précédemment. Dans l'affirmative, le processus d'appariement se déroulerait normalement. Dans le cas contraire, on procéderait à un nouveau tirage, mais sans décompter le tirage précédent du crédit alloué au candidat au mariage. L'inconvénient majeur de cette procédure est qu'elle deviendrait de plus en plus lourde au fur et à mesure que le nombre de couples augmenterait. De plus, la comparaison du numéro tiré avec ceux tirés précédemment demanderait un temps relativement long. Nous avons donc éliminé cette procédure sans en faire l'essai.

La seconde solution est fondée sur le raisonnement suivant. Si nous nous représentons concrètement les "dossiers" correspondant aux hommes et aux femmes de l'isolat, nous avons deux piles de dossiers, celle des hommes et celle des femmes. La mise en oeuvre du processus d'appariement consiste tout d'abord à tirer au sort le numéro d'ordre du partenaire "actif", par exemple un homme, et à extraire son dossier

de la pile correspondante. Si le numéro tiré est par exemple 8, on extraira le 8^e dossier de la pile ; or, en procédant à cette extraction, nous avons transformé les numéros d'ordre des dossiers suivants : le 9^e dossier est devenu le 8^e, et le $n^{\text{ième}}$ le $(n-1)^{\text{ième}}$. Pour réaliser la même opération dans la matrice des hommes, il nous suffit de transférer le contenu de la ligne 9 dans la ligne 8, et ainsi de suite jusqu'au transfert du contenu de la ligne n dans la ligne $n-1$; puis, au tirage au sort suivant, nous engendrerons un nombre pseudo-aléatoire compris dans l'intervalle $[1, n-1]$; au $i^{\text{ème}}$ tirage au sort, le nombre pseudo-aléatoire sera compris dans l'intervalle $[1, n-i + 1]$. On procéderait de même pour le partenaire "passif", en prenant soin évidemment de n'extraire le dossier de celui-ci qu'après que le mariage soit prononcé. Nous avons expérimenté cette procédure sur un échantillon restreint, et elle a fonctionné d'une manière satisfaisante ; il nous a cependant paru possible de l'améliorer, de manière à réduire son temps d'exécution.

La troisième solution, que nous avons finalement adoptée, est par conséquent le perfectionnement de la procédure précédente. Le transfert du contenu de la $i^{\text{ème}}$ ligne d'une matrice dans la $(i-1)^{\text{ième}}$ ligne est une opération longue sur ordinateur, comme toutes les opérations de classement. Le temps d'exécution est d'autant plus grand que le nombre de colonnes est plus élevé. Nous avons donc cherché à minimiser le nombre de colonnes de la matrice à reclasser ; nous sommes parvenu à le réduire à un, en utilisant la technique de l'*adressage indirect*. Cette procédure fait appel, pour chaque sous-population de l'isolat, à deux tableaux. Prenons comme exemple la sous-population masculine. Une matrice rectangulaire H de n lignes et 4 colonnes contient la description des hommes de l'isolat ; un tableau TH de n lignes et 1 seule colonne (*matrice-colonne*) contient les numéros d'ordre des mêmes hommes : au début de la simulation, ce

numéro coïncide avec l'indice de ligne de ce tableau. Au cours de la simulation, seul ce second tableau sera reclassé, la matrice rectangulaire demeurant inchangée.

Le déroulement du processus passe par les phases suivantes :

- un nombre pseudo-aléatoire a , compris dans l'intervalle $[1, n]$, est donné ; il désigne la $a^{\text{ième}}$ ligne du tableau HT .
- le nombre i contenu dans la $a^{\text{ième}}$ ligne du tableau HT est le numéro d'ordre de l'homme à extraire du tableau T . On extrait donc la $i^{\text{ème}}$ ligne du tableau T .
- le reclassement s'effectue dans le tableau HT , de la $a^{\text{ième}}$ ligne du tableau à la $(n-1)^{\text{ième}}$. Le contenu de la $a^{\text{ième}}$ ligne (numéro d'ordre de l'individu extrait) est remplacé par celui de la $(a + 1)^{\text{ième}}$ ligne (numéro d'ordre de l'individu suivant), jusqu'au contenu de la $(n-1)^{\text{ième}}$ ligne qui est remplacé par celui de la $n^{\text{ième}}$ ligne. Le contenu de la $n^{\text{ième}}$ ligne reste inchangé.
- au tirage suivant, le nombre pseudo-aléatoire sera compris dans l'intervalle $[1, n-1]$. Ainsi, la $n^{\text{ième}}$ ligne de HT ne pourra plus être lue. Etc..

La figure 10 illustre une partie du déroulement de ce processus pour la sous-population masculine d'un isolat limité à 7 individus. Le mécanisme est évidemment identique pour la sous-population féminine. Cette procédure, un peu plus difficile à programmer que la précédente, permet un gain de temps de plus de 40 % de la durée totale de la simulation.

Origine sociale
Religion
Niveau d'instruction
Localité de naissance

	1	2	3	4
1	1	3	1	8
2	2	5	4	3
3	1	1	2	4
4	7	2	2	1
5	2	1	2	3
6	1	1	2	6
7	3	1	4	4

Situation initiale

Tableau H

1	1
2	2
3	3
4	4
5	5
6	6
7	7

Tableau HT

1er tirage au sort

Intervalle : [1, 7]
Nombre tiré : 4

Individu sélectionné :
HT(4) = 4
H(4) =

7	2	2	1
---	---	---	---

Reclassement de HT :

1	1
2	2
3	3
4	5
5	6
6	7
7	7

Intervalle suivant

2ème tirage au sort

Intervalle : [1, 6]
Nombre tiré : 5

Individu sélectionné :
HT(5) = 6
H(6) =

1	1	2	6
---	---	---	---

Reclassement de HT :

1	1
2	2
3	3
4	5
5	7
6	7
7	7

Intervalle suivant

3ème tirage au sort

Intervalle : [1, 5]
Nombre tiré : 1

Individu sélectionné :
HT(1) = 1
H(1) =

1	3	1	8
---	---	---	---

Reclassement de HT :

1	2
2	3
3	5
4	7
5	7
6	7
7	7

Intervalle suivant

.....

FIG. 10 - Illustration de la procédure de reclassement des individus sélectionnés.

d) L'alternance homme/femme.

Nous avons décidé que ce serait tantôt un homme, tantôt une femme, qui jouerait le rôle du partenaire "actif". La réalisation de cette convention complique légèrement l'écriture du programme. Il nous faut tout d'abord prévoir un "inverseur", variable prenant alternativement l'un des deux états, *homme* ou *femme*. Nous avons utilisé pour cela une variable logique, en convenant que la valeur *vrai* correspondrait à l'état *homme*, et la valeur *faux* à l'état *femme*. La valeur de cette variable, "initialisée" à *vrai* au début du programme (phase 2 de l'organigramme fig. 9), est remplacée par sa négation après chaque mariage (phase 14 de l'organigramme) ; l'alternance est par conséquent réalisée. Cet inverseur nous servira à trois reprises :

- lors de l'extraction du partenaire "actif", qui devra être pris dans la matrice *hommes* si l'inverseur est en position *vrai*, et dans la matrice *femme* dans le cas contraire.

- lors du tirage au sort des partenaires "passifs" jusqu'à conclusion du mariage ; ce tirage au sort devant être fait dans la matrice *femmes*, si l'inverseur est en position *vrai*, et dans la matrice *hommes* sinon.

- lors du reclassement des tableaux d'adressage indirect, afin d'éviter que le reclassement du tableau *femmes* ne soit effectué à partir de la ligne correspondant au rang de l'homme sélectionné, et réciproquement.

5.2.2. La structure générale du programme.

Le programme s'organise autour des mécanismes de choix que nous venons de décrire. Les problèmes qu'il nous reste à examiner sont

relatifs à la lecture des données, aux itérations, et à l'édition des résultats.

a) La lecture des données.

Nous avons tout d'abord envisagé de procéder au report sur bande magnétique de l'ensemble des données concernant les 3204 individus impliqués dans cette série de simulations. Pour chaque combinaison d'un modèle de décision et d'un degré d'intensité de l'attitude endogamique, la bande aurait été lue, puis rembobinée pour être relue à la combinaison suivante. Une telle procédure aurait occasionné une perte de temps considérable, en raison de la lenteur relative des opérations de lecture. Nous avons décidé d'utiliser la capacité de mémoire élevée du *Control Data 3600*, en stockant l'ensemble de ces informations en mémoire avant le commencement de la phase de simulation.

Les informations à stocker étaient les codes correspondant aux quatre variables d'entrée de notre modèle, soit huit valeurs par couple, plus l'identification de l'isolat auquel le couple appartient. Cette dernière information correspond, dans les données, aux cartes de séparation des isolats. Plutôt que de reproduire fidèlement en mémoire centrale l'image du fichier tel que nous l'avions constitué, nous avons préféré accoler à chaque couple le numéro d'ordre de l'isolat dont il dépend. Nous avons ainsi constitué une matrice de 1602 lignes et 9 colonnes, chaque ligne correspondant à un couple, et les colonnes contenant les caractéristiques de l'homme (col. 1 à 4), de la femme (col. 5 à 8), et le numéro de l'isolat (col. 9). Cette dernière colonne n'a pas été lue, mais remplie automatiquement par l'ordinateur : il a suffi de donner l'instruction de mettre au début le nombre 1 dans la 9^o colonne

pour tous les couples lus, et de passer au nombre suivant à chaque signe de séparation des isolats (carte portant un 5 en colonne 80), jusqu'à la fin du fichier (carte portant un 9 en colonne 80).

L'utilisation des données ainsi rangées en mémoire a été facilitée par la procédure d'*adressage indirect* décrite au § 5.2.1.c. Il est possible en effet de ranger, dans le tableau d'adressage indirect correspondant à l'isolat traité, les numéros d'ordre des individus non plus dans l'isolat, mais dans l'ensemble de la population. Pour un isolat donné, le tableau d'adressage peut contenir les numéros de ligne de la matrice des données ; la procédure adoptée ne modifiant que le tableau d'adressage, la matrice des données a pu être utilisée pour l'ensemble des simulations, sans nécessiter ni transfert d'information particulier, ni relecture des données.

b) Les itérations.

Nous avons prévu trois boucles imbriquées : la première permettant de passer d'un modèle de décision au suivant (phases 3 et 22 de l'organigramme de la fig. 9) ; la seconde faisant varier le degré d'intensité de l'attitude (phases 4 et 20 de l'organigramme) ; la troisième réitérant la même simulation un nombre de fois suffisant pour obtenir la précision désirée dans les résultats (phase 18 de l'organigramme).

La boucle permettant de simuler les divers modèles de décision retenus a été réalisée en créant un indice variant de 1 à 16, et en attribuant un numéro de 1 à 16 à chacune des expressions logiques correspondant à un modèle. Ce système nous a obligé à écrire *in extenso*

chaque expression, faute d'avoir trouvé une procédure simple permettant soit de lire ces expressions comme des paramètres externes soit de les engendrer automatiquement. Notre programme pourrait certainement être amélioré sur ce point.

La boucle faisant varier le degré d'intensité des attitudes porte sur un indice numérique ; nous avons choisi, après chaque essai, de faire prendre à celui-ci toutes les valeurs entières de l'intervalle $[-10, 10]$. Afin de nous réserver la possibilité d'attribuer d'autres valeurs à l'indice, nous avons décidé de considérer ces valeurs comme des paramètres externes du programme, et de les lire au même titre que les données. Ces valeurs sont rangées en mémoire dans un tableau, de manière à pouvoir être désignées par la valeur d'un indice variant de 1 à 21.

La boucle relative aux répétitions a été supprimée après essais. Nous avons constaté en effet que les variations aléatoires du taux d'homogamie étaient relativement faibles, pour une série de dix simulations identiques (cf. § 6.1, tableau n° 13). Il nous a semblé peu justifié de multiplier par dix le temps d'exécution de notre programme, pour un gain de précision qui n'améliorait pas notre compréhension des relations entre les variables étudiées.

c) L'édition des résultats.

Au cours des essais préliminaires, nous avons constitué et édité les tableaux donnant, pour l'ensemble de la population, la répartition des couples selon les caractéristiques croisées des conjoints :

localité de naissance, niveau d'instruction, religion, et origine sociale du mari et de la femme. Si nous avions conservé ce principe pour l'ensemble de la recherche, le temps d'exécution du programme aurait été considérablement allongé ; de plus, nous aurions disposé de plus de 1200 tableaux à dépouiller, ce qui aurait dépassé largement le cadre de notre recherche. Aussi nous sommes-nous borné à relever, pour chaque combinaison d'un modèle de décision et d'un degré d'intensité, le *pourcentage d'homogamie* pour chacune des quatre variables retenues. Nous avons jugé intéressant de relever également le *nombre moyen d'essais* tentés par le partenaire "actif" pour conclure un mariage. L'ensemble de ces résultats, pour une série de 336 simulations, tient en 16 pages de résultats, que nous analysons au chapitre 6.

5.3. L'ECRITURE DU PROGRAMME.

5.3.1. Organigramme commenté.

Le programme que nous avons écrit comporte un court programme principal, un sous-programme plus important correspondant au mécanisme de choix, et une fonction fournissant les nombres "au hasard". Nous donnons ci-après (fig. 11 à 12c) l'organigramme détaillé de ce programme. Les chiffres figurant dans des cercles correspondent aux étiquettes numériques du programme ; ils permettent de se reporter plus aisément au texte de celui-ci (§ 5.3.2.).

a) Le programme principal.

Le programme principal (fig. 11) prend en charge la lecture des données, l'exécution des itérations relatives aux modèles de déci-

sion et aux degrés d'intensité, et l'édition des résultats. Parmi les données à lire, figurent dans l'ordre les valeurs prises par le degré d'intensité des attitudes, les caractéristiques des couples sur lesquels porte la simulation, et enfin le libellé des titres correspondant à chaque modèle, et devant figurer sur les feuilles de résultats. Les résultats correspondant à un modèle donné sont édités sous la forme d'un tableau de 21 lignes et 6 colonnes. Chaque ligne correspond à un degré d'intensité de l'attitude. La première colonne indique le degré d'intensité, la seconde le nombre moyen d'essais tentés pour aboutir au mariage, et les suivantes les taux d'homogamie pour les quatre variables.

L'"inverseur" *INV*, qui commande l'alternance *homme/femme*, est "initialisé" à la valeur *vrai*. Les premières données lues sont les 21 valeurs du degré d'intensité. Celles-ci sont rangées directement dans la première colonne du tableau *RSLT* ("résultats"), de 6 colonnes et 21 lignes, destiné à contenir les résultats pour l'édition ; ces valeurs font en effet partie des données à éditer. Ainsi, dans le cours du programme, la $n^{\text{ième}}$ valeur du degré d'intensité sera recherchée dans la $n^{\text{ième}}$ ligne et la $1^{\text{ère}}$ colonne du tableau *RSLT*. Par exemple, la $12^{\text{ème}}$ valeur du degré d'intensité sera appelée *RSLT (1, 12) (*)*.

La phase suivante réalise la lecture des caractéristiques des couples, et leur rangement dans le tableau *KE*, qui comporte 9 colonnes et 1610 lignes. La variable *IS* ("isolat") est le numéro de l'isolat, que l'on enregistre automatiquement en colonne 9 de *KE* : $KE (9, L) = IS$.

(*)

Rappelons qu'en *FORTRAN*, le 1^{er} indice correspond, dans l'implantation du tableau en mémoire, à l'indice de *colonne* ; le 2^{e} à l'indice de *ligne* ; le 3^{e} , à l'indice de *bloc*. Nous nous en sommes tenu à cette particularité bien qu'en pratique elle puisse être ignorée (à condition d'explicitier toutes les boucles en entrée/sortie).

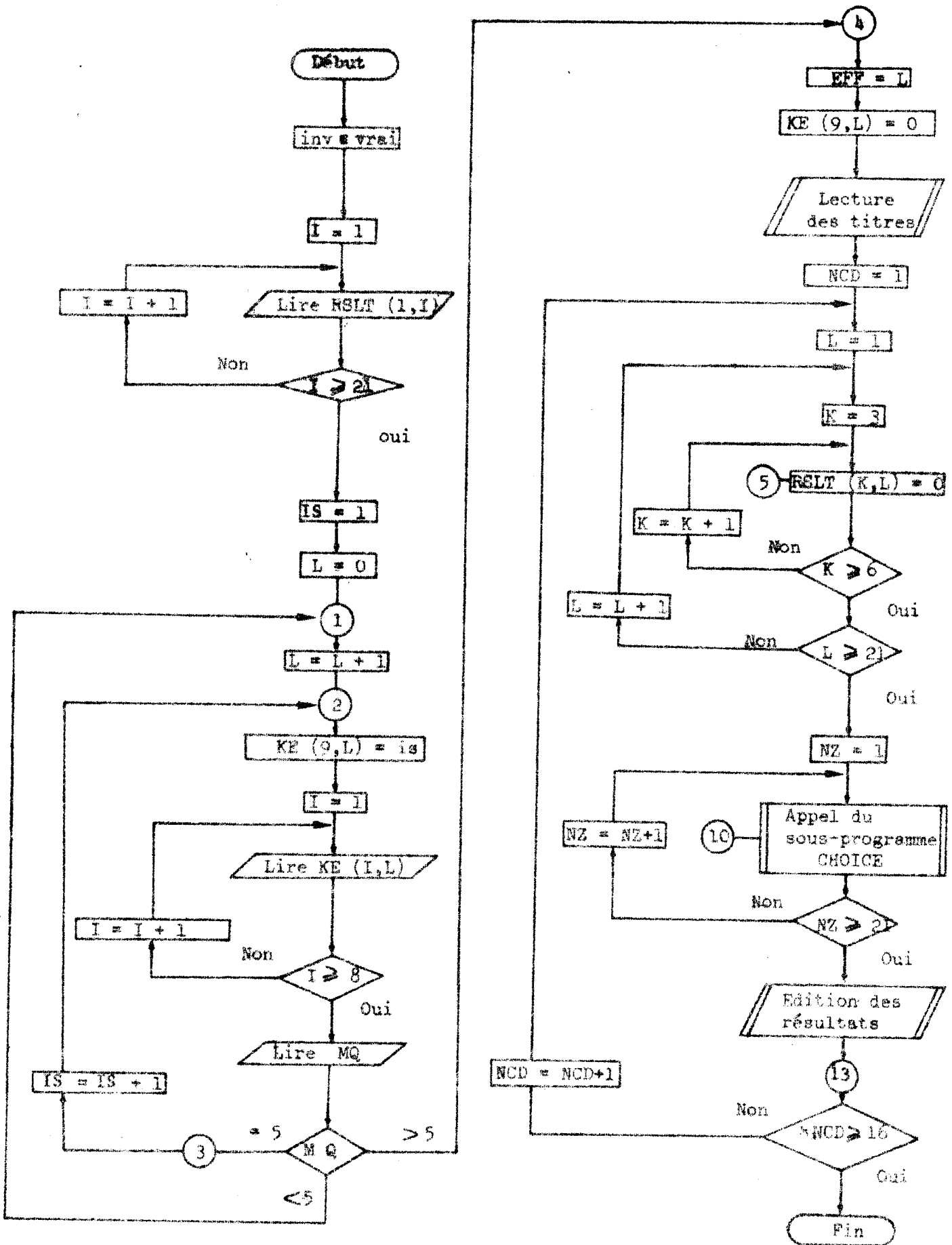


Figure 11 - Organigramme du programme principal.

Au départ, $IS = 1$. L est l'indice de ligne du tableau KE ; cet indice augmente d'une unité à chaque carte lue, sauf lorsque la carte est une carte de séparation des isolats. I est l'indice de colonne de KE , servant à ranger les valeurs codées des 8 caractéristiques du couple (4 pour l'homme, 4 pour la femme). En même temps que ces valeurs sont lues, on vérifie s'il existe une perforation en colonne 80 de la carte, en lisant la variable MQ ("marque"). S'il s'agit d'une carte correspondant à un couple, la colonne 80 est vierge, et $MQ = 0$ (*) ; s'il s'agit d'une carte séparant deux isolats, $MQ = 5$; s'il s'agit de la carte de fin de fichier, $MQ = 9$ (cf. § 5.1.1.c). Si $MQ = 0$ (dans le test, $MQ < 5$), on passe à la carte suivante après avoir augmenté L d'une unité. Si $MQ = 5$, la carte lue était une carte de séparation ; on augmente donc IS d'une unité (passage à l'isolat suivant), et on enregistre la carte suivante dans la même ligne du tableau KE . Si $MQ = 9$ (dans le test, $MQ > 5$), la lecture des données relatives aux individus est terminée. On conserve en mémoire l'effectif EFF des couples, qui est égal à la dernière valeur de l'indice de ligne L . Puis, pour faciliter l'exécution du sous-programme, on place un zéro en colonne 9 de la ligne suivante du tableau KE ; ce zéro joue le rôle de la marque de fin de fichier.

Les libellés des titres sont lus ensuite. Chaque libellé occupe une carte, soit 10 mots de 8 caractères. Nous n'avons pas fait figurer dans l'organigramme le détail de cette lecture, qui ne présente pas d'intérêt particulier ; on pourra en prendre connaissance

(*)

Ceci n'est vrai que parce que MQ est définie dans le programme comme étant une variable numérique. Pour ces variables en effet, le *blanc* est conventionnellement identifié à zéro.

dans le texte du programme.

La dernière partie du programme principal comporte une boucle extérieure, sur l'indice *NCD* ("numéro du critère de décision"), qui correspond au numéro de chacun des 16 modèles de décision ; l'usage de cet indice sera précisé dans le sous-programme. A l'intérieur de cette boucle, les colonnes 3 à 6 du tableau *RSLT* sont remises à zéro ; ces colonnes sont en effet destinées à servir au déroulement des calculs des taux d'homogamie. Puis on fait varier de 1 à 21 l'indice *NZ*, qui correspond au numéro d'ordre des valeurs des degrés d'intensité de l'attitude ; à l'intérieur des boucles imbriquées sur *NCD* (modèles de décision) et *NZ* (degrés d'attitude) se trouve l'ordre d'appel au sous-programme de choix du conjoint, dont l'exécution constitue une simulation complète. Après l'exécution des 21 itérations sur *NZ*, mais à l'intérieur de la boucle sur *NCD*, se situent les ordres d'édition des résultats calculés par le sous-programme. Comme pour les instructions de lecture des titres, nous renvoyons pour le détail de celles-ci au texte du programme.

b) Le sous-programme de choix du conjoint.

Ce sous-programme effectue à lui seul une simulation complète, pour un type de modèle de décision donné et une valeur du degré d'intensité de l'attitude endogamique. Il correspond aux phases 5 à 17 de l'organigramme de la figure 9, et comporte neuf parties :

- constitution des tableaux d'adressage indirect pour chaque isolat ;
- calcul du nombre maximum de choix autorisé ;

- tirage du partenaire "actif" ;
- tirage du partenaire "passif" ;
- sélection du modèle de décision ;
- comparaison des caractéristiques des partenaires, et constitution des couples ;
- procédure de reclassement ;
- procédure de mariage forcé en fin d'isolat ;
- calcul des pourcentages sur les résultats.

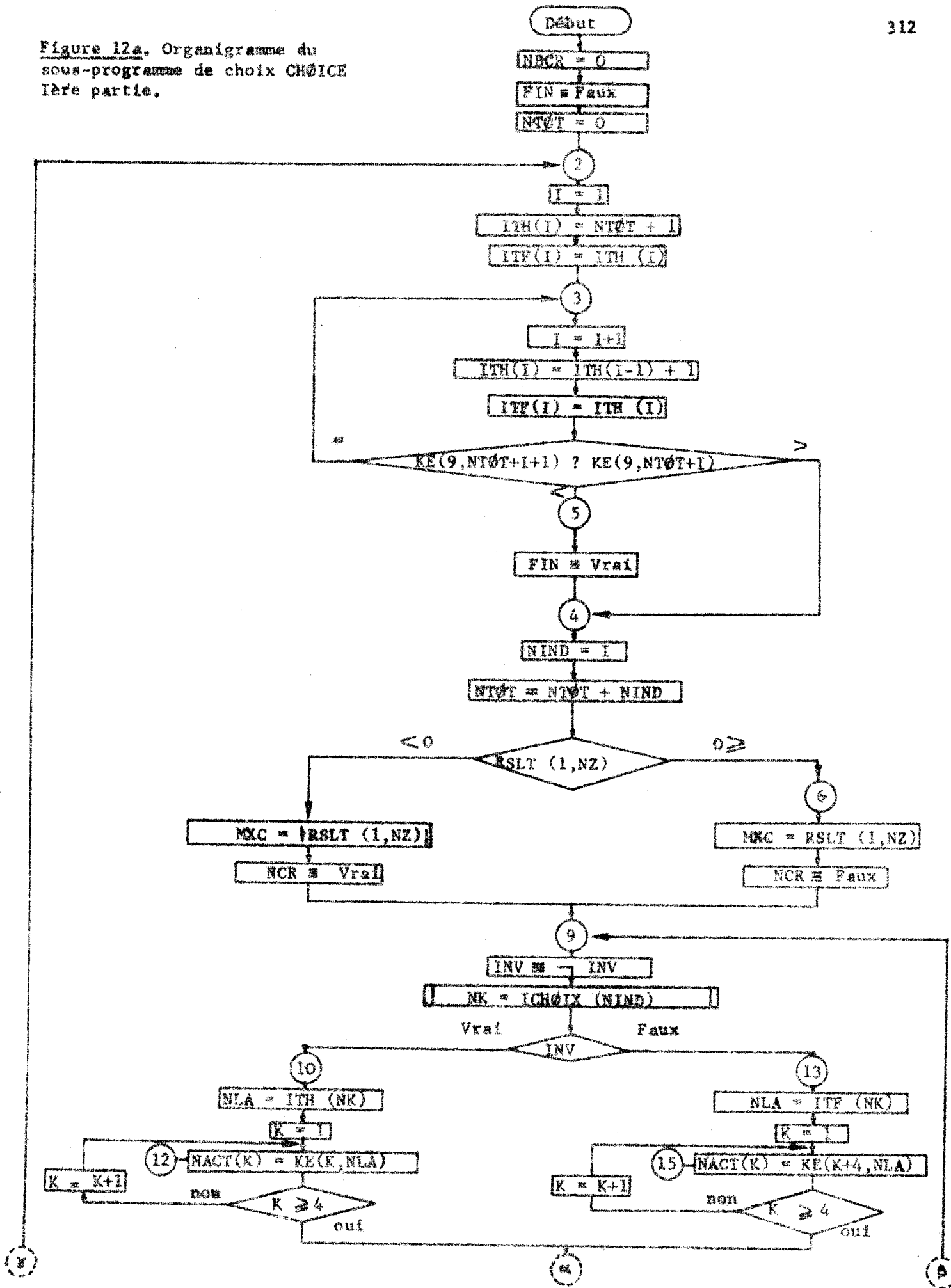
Au début du sous-programme, (Fig. 12a), un "compteur" *NBCR* ("nombre de choix réalisés"), est "initialisé" à zéro. Ce "compteur" est destiné à enregistrer le nombre total de tirages au sort d'un conjoint nécessaires pour appairier l'ensemble des individus de notre population au cours d'une simulation. Divisé par le nombre de couples, ce total fournira le nombre moyen de choix par simulation. D'autre part, une variable logique, *FIN*, est "initialisé" à la valeur *faux*. Elle deviendra *vraie* après l'enregistrement du dernier isolat.

Les *tableaux d'adressage indirect* sont au nombre de deux. Le premier, *ITH*, sert à désigner les hommes dans l'isolat ; le second, *ITF*, les femmes. Pour un isolat donné, le tableau *ITH* doit contenir les numéros d'ordre, dans le tableau *KE*, des hommes de l'isolat. Par exemple, pour le premier isolat, nous aurons $ITH(1) = 1$, $ITH(2) = 2$, etc. jusqu'à $ITH(79) = 79$, puisque l'effectif du premier isolat est 79. Pour le second, nous aurons $ITH(1) = 80$, $ITH(2) = 81$, etc., jusqu'à $ITH(19) = 98$, puisque les 19 hommes du second isolat correspondent aux lignes 80 à 98 du tableau *KE*. Pour le quarante-huitième isolat, nous aurons $ITH(1) = 1576$, jusqu'à $ITH(27) = 1602$. Il en sera exactement de même pour le tableau *ITF*.

Pour obtenir automatiquement ce résultat sans faire intervenir l'effectif de chaque isolat, nous utilisons un "compteur", $NTOT$, qui indique à quelle ligne du tableau KE nous sommes arrêté à l'étape précédente ; au début de la simulation, nous avons par conséquent $NTOT = 0$. Nous commençons par remplir la première ligne des tableaux ITH et ITF , en donnant à l'indice I la valeur 1, et à $ITH(I) = ITF(I)$ la valeur $NTOT + 1$. La ligne suivante des tableaux ITH et ITF , de coordonnées $ITH(I + 1)$ et $ITF(I + 1)$, est remplie en ajoutant 1 au contenu de la ligne précédente. On vérifie ensuite si les individus de la ligne suivante du tableau KE (dont le numéro est égal à $NTOT + I + 1$) ont le même numéro d'isolat (en colonne 9 du tableau KE) que les individus précédemment enregistrés (correspondant à la ligne de KE dont le numéro est égal à $NTOT + I$). Ce numéro d'isolat peut être égal, supérieur, ou inférieur au numéro des individus précédemment enregistrés. S'il est égal, nous sommes toujours dans le même isolat, et nous continuons l'enregistrement dans ITH et ITF en passant à la ligne suivante de ces tableaux. S'il est supérieur, nous sommes arrivé à la fin d'un isolat autre que le dernier ; on relève alors l'effectif de l'isolat, $NIND$ ("nombre d'individus"), qui est égal à la valeur de l'indice I , et on ajoute cette valeur $NIND$ à $NTOT$, pour indiquer le nombre total d'individus enregistrés. S'il est inférieur, c'est que nous venons de rencontrer la marque de fin de fichier (un zéro en colonne 9 du tableau KE) ; nous effectuerons alors les mêmes opérations que dans le cas précédent, mais auparavant, nous aurons donné la valeur *vrai* à la variable logique FIN .

Le calcul du *nombre maximum de choix* MXC est simple : il est égal au degré d'intensité de l'attitude (rangé en colonne 1 du tableau $RSLT$) si celui-ci est positif ou nul, et à sa valeur absolue s'il est

Figure 12a. Organigramme du sous-programme de choix CH@ICE ière partie.



négatif. Nous profitons du test sur la valeur du degré d'intensité pour positionner une variable logique, *NCR* ("négation du critère"), qui nous servira ultérieurement à déterminer s'il y a lieu de remplacer le modèle de décision par sa négation. Il est en effet plus économique de faire un test sur une variable logique simple que sur une valeur arithmétique qu'il faut rechercher dans un tableau à deux dimensions (*RSLT*).

Le tirage du partenaire "actif", c'est-à-dire la désignation de l'individu qui choisit, commence par le "basculement" de l'inverseur *INV* à sa position contraire (pour réaliser automatiquement l'alternance *homme/femme*). Il y a ensuite tirage au sort d'un nombre *NK*, compris entre 1 et *NIND*; ce tirage est réalisé par la fonction *ICHOIX*, dont le texte figure à la fin du programme. Le partenaire "actif" devra ensuite être recherché parmi les hommes si *INV* est en position *vrai*, et parmi les femmes sinon. Supposons que *INV* \equiv *vrai* ; le nombre *NK* indique le numéro de la ligne du tableau *ITH* dans laquelle doit être recherchée l'adresse du partenaire "actif", c'est-à-dire son numéro de ligne *NLA* dans le tableau *KE*. On enregistre ensuite les caractéristiques de celui-ci (figurant dans les 4 premières colonnes de la ligne *NLA* de *KE*) dans un tableau à 4 colonnes *NACT*. Si *INV* \equiv "*faux*", l'adresse *NLA* du partenaire "actif" figure à la ligne *NK* du tableau *ITF* ; les caractéristiques de ce partenaire sont à rechercher ensuite à la ligne *NLA* de *KE*, mais dans les colonnes 5 à 8 puisqu'il s'agit d'une femme.

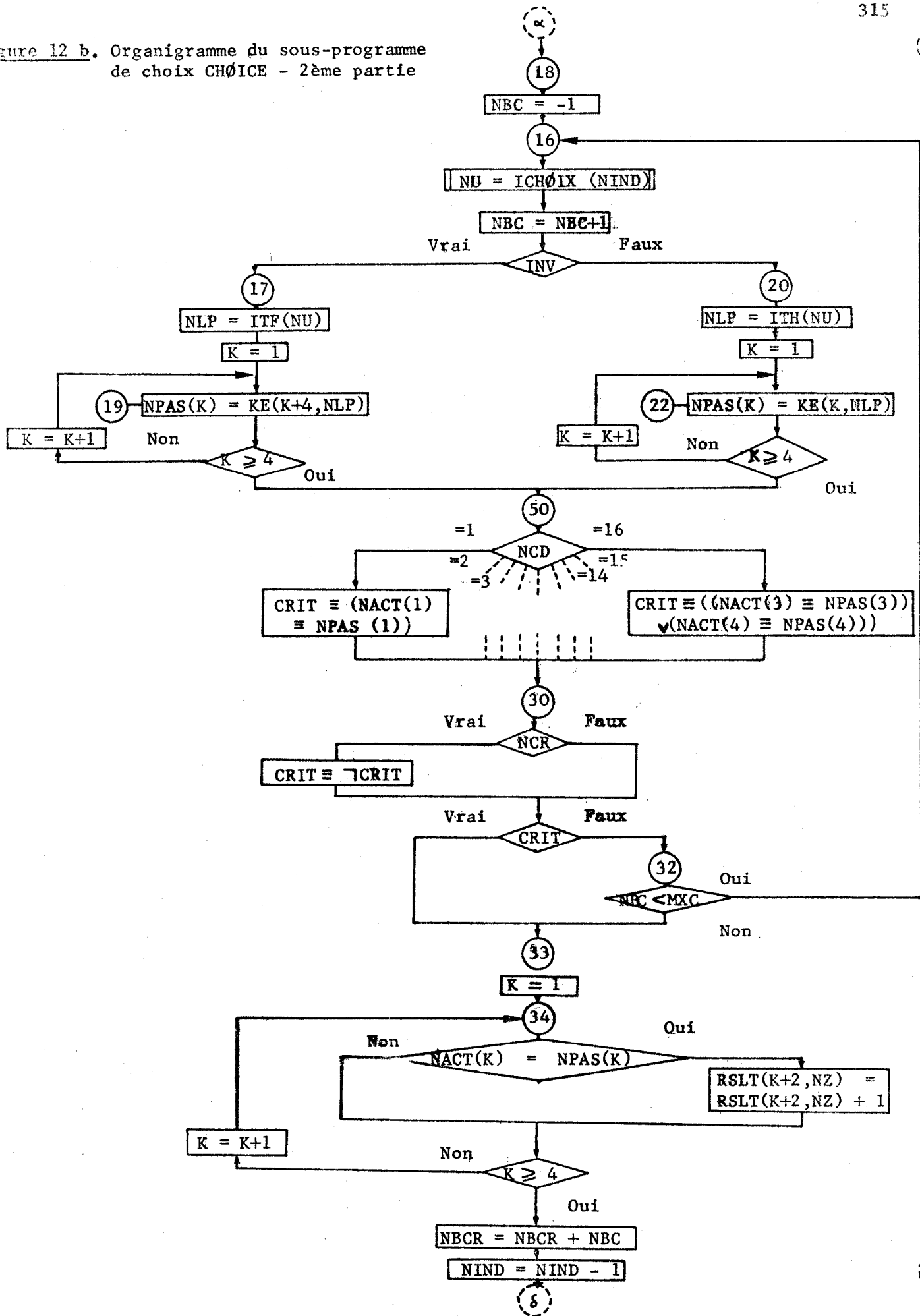
Le tirage du partenaire "passif" (Fig. 12b) est réalisé selon les mêmes principes que le tirage du partenaire "actif". Un nombre aléatoire *NU*, compris entre 1 et *NIND*, indique le numéro de ligne du tableau d'adressage indirect contenant l'adresse *NLP* de ce

partenaire. Cette adresse figure dans le tableau *ITF* si *INV* \equiv *vrai*, et *ITH* sinon. Le tirage du partenaire "passif" peut être réitéré pour un même partenaire "actif". Il convient par conséquent de relever le nombre de choix effectués par un individu donné : le "compteur" *NBC* ("nombre de choix") sert à cet usage. Il est "initialisé" à $NBC = -1$, afin de ne pas compter pour un choix le premier tirage au sort, qui constitue en quelque sorte un tirage obligé ; dans le cas de panmixie en effet, il y a un tirage au sort, alors que le nombre de choix autorisé *MXC* est nul. Les choix véritables commencent après le premier tirage au hasard.

La *sélection du modèle de décision* dépend de la valeur de l'indice *NCD* ("numéro du critère de décision"). Un test sur *NCD* renvoie aux 16 modèles, écrits intégralement sous la forme d'une expression du type "le critère est vrai si le niveau d'instruction des deux partenaires est égal" (modèle n° 2), ou "le critère est vrai si les deux partenaires ont le même niveau d'instruction et la même religion" (modèle n° 11). Si la variable *NCR* (négation du critère) est *vraie*, nous sommes dans une situation d'exogamie, et il convient de remplacer la valeur du critère, *CRIT*, par sa négation.

La *comparaison des caractéristiques des partenaires* selon le modèle de décision se limite à la constatation de la valeur prise par la variable logique *CRIT* à l'issue de la phase précédente. Si *CRIT* \equiv *vrai*, le mariage est prononcé. Sinon, on vérifie si le partenaire actif peut encore choisir, c'est-à-dire si $NBC < MXC$. Dans l'affirmative, on procède au tirage d'un autre partenaire "passif" ; sinon, le mariage est prononcé malgré tout. L'enregistrement du mariage se limite à la comptabilisation des cas d'homogamie. Pour cela, chaque caractéristique du partenaire "actif" est comparée à la même

Figure 12 b. Organigramme du sous-programme de choix CHØICE - 2ème partie



caractéristique du partenaire "passif" ; s'il y a identité, on enregistre un cas d'homogamie de plus pour la variable correspondante. Ces enregistrements se font directement dans les colonnes 3 à 6 du tableau *RSLT*, à la ligne *NZ* correspondant au degré de l'attitude sur lequel porte la simulation.

A l'issue de l'enregistrement du mariage, il convient de procéder à trois opérations. La première consiste à enregistrer au "compteur" *NBCR* le nombre de choix effectués par le partenaire "actif". La seconde est de diminuer d'une unité le nombre de couples d'individus restant dans l'isolat ($NIND = NIND - 1$), afin que la borne supérieure du nombre pseudo-aléatoire corresponde bien au nombre d'hommes (ou de femmes) à marier. La troisième opération est le reclassement des individus dans les tableaux d'adressage.

La *procédure de reclassement* (Fig. 12c) des tableaux *ITH* et *ITF* consiste à partir de la ligne d'où le conjoint a été tiré, et à reclasser le contenu des lignes suivantes ainsi que nous l'avons exposé au § 5.2.1.c. Si le partenaire "actif" était un homme ($INV \equiv vrai$), le reclassement de *ITH* se fait à partir de la ligne *NK*, et celui de *ITF* à partir de la ligne *NU* ; si le partenaire "actif" était une femme ($INV \equiv faux$), c'est évidemment le contraire. La dernière ligne à reclasser est celle dont le numéro correspond à l'effectif de couples d'individus restant à marier, soit *NIND*. A l'issue de la procédure de reclassement, s'il reste plus d'un couple dans l'isolat, le processus est repris à partir du tirage du partenaire "actif". S'il ne reste plus qu'un homme et une femme, ils sont mariés automatiquement.

La *procédure de mariage forcé* en fin d'isolat réalise ce

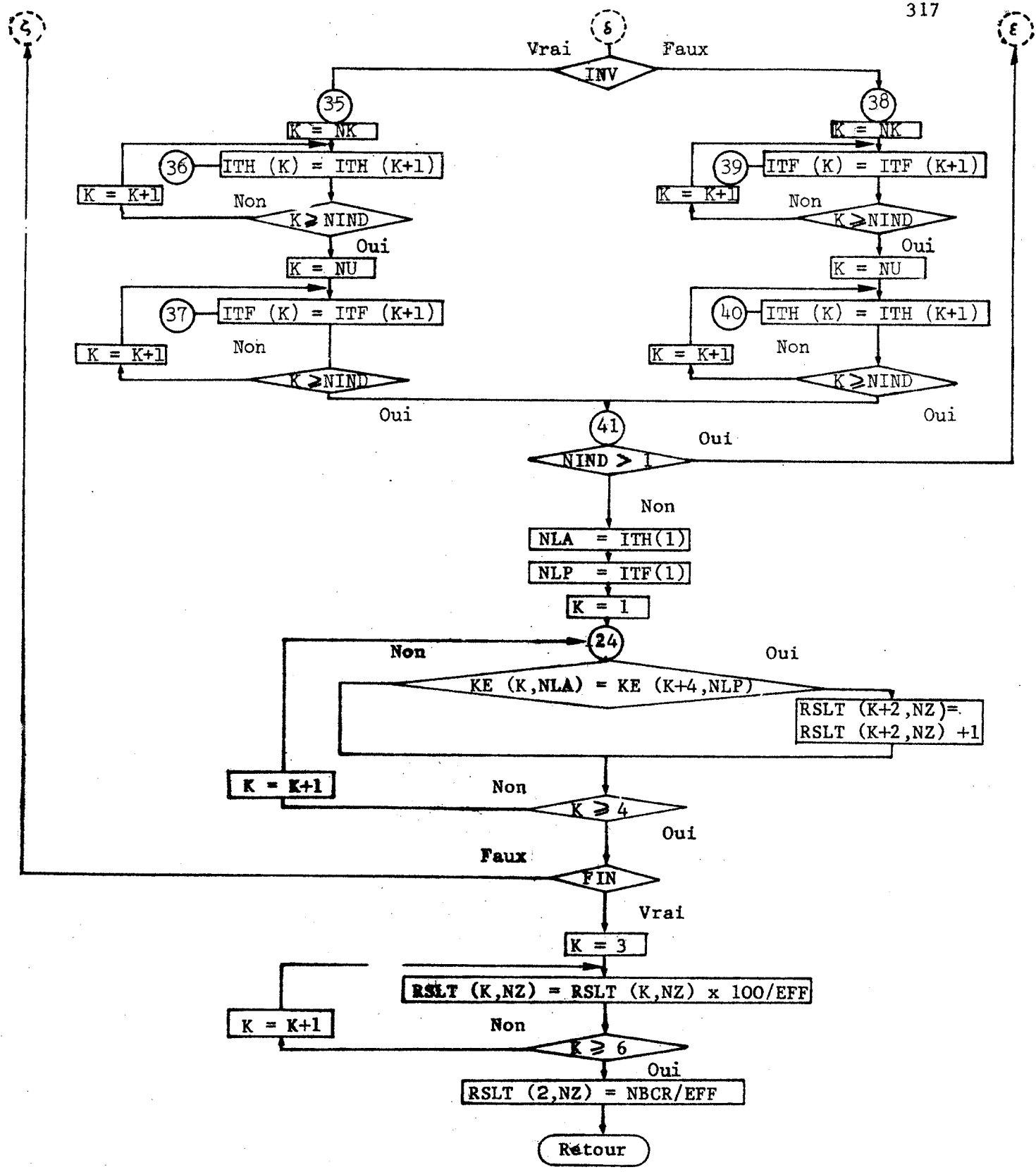


Figure 12 c. Organigramme du sous-programme de choix CHØICE- 3ème partie

mariage automatique. L'adresse de l'homme se trouve dans la première ligne du tableau *ITH*, celle de la femme dans la première ligne du tableau *ITF*. Leurs caractéristiques, figurant dans le tableau *KE*, sont comparées directement, et les cas d'homogamie enregistrés dans le tableau *RSLT*. A l'issue de cette procédure, s'il ne s'agissait pas du dernier isolat (*FIN* \equiv *faux*), le processus est repris à partir de la constitution des tableaux d'adressage indirect pour l'isolat suivant. S'il s'agissait du dernier isolat (*FIN* \equiv *vrai*), on exécute les calculs finaux sur les résultats.

Le calcul des pourcentages sur les résultats consiste à transformer les nombres de cas d'homogamie pour chaque variable (colonnes 3 à 6 du tableau *RSLT*) en pourcentages sur l'effectif total des couples *EFF*. En même temps est calculé le taux moyen de choix par mariage, en divisant le nombre total de choix *NBCR* par *EFF*.

c) La fonction engendrant des nombres pseudo-aléatoires.

Cette fonction, qui utilise la fonction de bibliothèque *RANF* est très simple ; aussi n'avons-nous fait figurer que son texte *FORTTRAN* à la fin du programme. Pour un nombre entier *NEL* ("nombre d'éléments") correspondant à la valeur supérieure de l'intervalle $[1, NEL]$, elle fournit un nombre pseudo-aléatoire tiré d'une distribution rectangulaire dans cet intervalle.

5.3.2. Texte du programme.

Le programme a été écrit en *FORTRAN 3600* (51). Les seules différences entre le langage que nous utilisons et le *FORTRAN IV* défini par les normes internationales concernent :

- les *ordres d'entrée-sortie*, écrits sous la forme (*) :

```
READ < numéro de format > , < liste >
```

```
PRINT < numéro de format > , < liste >
```

au lieu de :

```
READ ( < n° d'unité logique > , < n° de format > ) < liste >
```

```
WRITE( < n° d'unité logique > , < n° de format > ) < liste >
```

- le *IF logique à deux adresses*, qui s'écrit IF (< expression logique >) < étiquette 1 > , < étiquette 2 > . Si l'expression logique est vraie, il y a saut vers l'instruction repérée par < étiquette 1 > ; si l'expression est fausse, le saut se fait vers l'instruction repérée par < étiquette 2 > .

(*)

En utilisant la notation syntaxique de J.W. BACKUS.

PROGRAM HOMOGAMY

```

C
C*** PROGRAMME PRINCIPAL ***
C
  DIMENSION TITLF(16,10)
  COMMON KE(9,1610),INV,NZ,RSLT(6,21),NCD,EFF
  LOGICAL INV
  INV=.TRUE.
  READ 72,(RSLT(I,I),I=1,21)
  IS=1
  L=0
1  L=L+1
2  KE(9,L)=IS
  READ 70,(KE(I,L),I=1,8),MQ
  IF(MQ-5)1,3,4
3  IS=IS+1
  GO TO 2
4  EFF=FLOAT(L)
  KE(9,L+1)=0
  READ 71,((TITLE(I,J),J=1,10),I=1,16)
  DO 13 NCD=1,16
  DO 5 L=1,21
  DO 5 K=3,6
5  RSLT(K,L)=0.
  DO 10 NZ=1,21
  CALL CHOICE
10 CONTINUE
  PRINT 80,NCD,(TITLE(NCD,K),K=1,10)
  DO 11 L=1,10
11 PRINT 81,(RSLT(K,L),K=1,6)
  PRINT 82,(RSLT(K,11),K=1,6)
  PRINT 82,(RSLT(K,12),K=1,6)
  DO 12 L=13,21
12 PRINT 81,(RSLT(K,L),K=1,6)
13 CONTINUE
70 FORMAT (10X,I1,2(4X,I1),9X,I1,19X,I1,2(4X,I1),9X,I1,8X,I1)
71 FORMAT (10A8)
72 FORMAT (12X,14F4.0)
80 FORMAT (10HIJ-P GREMY,110X,7HPAGE NO,I3//33X,
  A68HETUDE DE L'INFLUENCE SUR L'HOMOGAMIE DE L'ENDOGAMIE ET DE L'EXO
  RGAMIE/33X,68(1H-),4(/),14X,104(1H*)/14X,1H*,102X,1H*/14X,1H*,2X,
  C18HCRITERE DE CHOIX =,10A8,2X,1H*/14X,1H*,102X,1H*/14X,104(1H*),
  D4(/),11X,15HNOMBRE DE CHOIX,6X,21HLOCALITE DE NAISSANCE,3X,
  E20HNIVEAU D'INSTRUCTION,10X,8HRELIGION,12X,15HORIGINE SOCIALE/
  FBX,7HMAXIMUM,7X,5HMOYEN,3X,4(24H TAUX D'HOMOGAMIE ),4(/))
81 FORMAT (10X,F3.0,6X,F8.3,9X,4(F8.3,16X))
82 FORMAT (1H0,9X,F3.0,6X,F8.3,9X,4(F8.3,16X))
  STOP
  END

```

SUBROUTINE CHOICE

```

C
C*** SOUS-PROGRAMME DE CHOIX DU CONJOINT ***
C
COMMON KE(9,1610),INV,NZ,RSLT(6,21),NCD,EFF
DIMENSION ITH(120),ITF(120),NACT(4),NPAS(4)
LOGICAL INV,CRIT,FIN,NCR
NBCR=0
FIN=.FALSE.

C
CONSTITUTION DES TABLEAUX D'ADRESSAGE INDIRECT DE L'ISOLAT
C
NTOT=0
2 I=1
ITH(I)=NTOT+1
ITF(I)=ITH(I)
3 I=I+1
ITH(I)=ITH(I-1)+1
ITF(I)=ITH(I)
IF(KE(9,NTOT+I+1)-KE(9,NTOT+I))5,3,4
5 FIN=.TRUE.
4 NIND=I
NTOT=NTOT+NIND

C
CALCUL DU NOMBRE MAXIMUM AUTORISE DE CHOIX
C
IF(RSLT(1,NZ).GE..0)GOTO 6
MXC=ABS(RSLT(1,NZ))
NCR=.TRUE.
GOTO 9
6 MXC=RSLT(1,NZ)
NCR=.FALSE.

C
C*** TIRAGE DU PARTENAIRE 'ACTIF' ***
C
9 INV=.NOT.INV
NK=ICHOIX(NIND)
IF(INV)10,13
10 NLA=ITH(NK)
DO 12 K=1,4
12 NACT(K)=KE(K,NLA)
GOTO 18
13 NLA=ITF(NK)
DO 15 K=1,4
15 NACT(K)=KE(K+4,NLA)

C
C*** TIRAGE DU PARTENAIRE 'PASSIF' ***
C
18 NBC=-1
16 NU=ICHOIX(NIND)
NBC=NBC+1
IF(INV)17,20
17 NLP=ITF(NU)
DO 19 K=1,4
19 NPAS(K)=KE(K+4,NLP)
GOTO 50
20 NLP=ITH(NU)
DO 22 K=1,4
22 NPAS(K)=KE(K,NLP)

```

```

C
C*** SELECTION DU MODELE DE DECISION ***
C
50 GOTO (81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96)NCD
81 CRIT=NACT(1).EQ.NPAS(1)
   GOTO 30
82 CRIT=NACT(2).EQ.NPAS(2)
   GOTO 30
83 CRIT=NACT(3).EQ.NPAS(3)
   GOTO 30
84 CRIT=NACT(4).EQ.NPAS(4)
   GOTO 30
85 CRIT=NACT(1).EQ.NPAS(1).AND.NACT(2).EQ.NPAS(2)
   GOTO 30
86 CRIT=NACT(1).EQ.NPAS(1).OR. NACT(2).EQ.NPAS(2)
   GOTO 30
87 CRIT=NACT(1).EQ.NPAS(1).AND.NACT(3).EQ.NPAS(3)
   GOTO 30
88 CRIT=NACT(1).EQ.NPAS(1).OR. NACT(3).EQ.NPAS(3)
   GOTO 30
89 CRIT=NACT(1).EQ.NPAS(1).AND.NACT(4).EQ.NPAS(4)
   GOTO 30
90 CRIT=NACT(1).EQ.NPAS(1).OR. NACT(4).EQ.NPAS(4)
   GOTO 30
91 CRIT=NACT(2).EQ.NPAS(2).AND.NACT(3).EQ.NPAS(3)
   GOTO 30
92 CRIT=NACT(2).EQ.NPAS(2).OR. NACT(3).EQ.NPAS(3)
   GOTO 30
93 CRIT=NACT(2).EQ.NPAS(2).AND.NACT(4).EQ.NPAS(4)
   GOTO 30
94 CRIT=NACT(2).EQ.NPAS(2).OR. NACT(4).EQ.NPAS(4)
   GOTO 30
95 CRIT=NACT(3).EQ.NPAS(3).AND.NACT(4).EQ.NPAS(4)
   GOTO 30
96 CRIT=NACT(3).EQ.NPAS(3).OR. NACT(4).EQ.NPAS(4)
30 IF(NCR)CRIT=.NOT.CRIT
C
COMPARAISON DES CARACTERISTIQUES DES PARTENAIRES
C
   IF(CRIT)33,32
32 IF(NBC.LT.MXC)GOTO 16
33 DO 34 K=1,4
34 IF(NACT(K).EQ.NPAS(K)) RSLT(K+2,NZ)=RSLT(K+2,NZ)+1.
   NBCR=NBCR+NBC
   NIND=NIND-1

```

```

C
C*** PROCEDURE DE RECLASSEMENT ***
C
      IF(INV) 35,38
35 DO 36 K=NK,NIND
36 ITH(K)=ITH(K+1)
      DO 37 K=NU,NIND
37 ITF(K)=ITF(K+1)
      GOTO 41
38 DO 39 K=NK,NIND
39 ITF(K)=ITF(K+1)
      DO 40 K=NU,NIND
40 ITH(K)=ITH(K+1)
41 IF(NIND.GT.1)GOTO 9
C
C*** PROCEDURE DE MARIAGE FORCE (FIN D'ISOLAT) ***
C
      NLA=ITH(1)
      NLP=ITF(1)
      DO 24 K=1,4
24 IF(KE(K,NLA).EQ.KE(K+4,NLP))RSLT(K+2,NZ)=RSLT(K+2,NZ)+1.
      IF(.NOT.FIN)GOTO 2
C
CALCUL DES POURCENTAGES SUR LES RESULTATS BRUTS
C
      DO 44 K=3,6
44 RSLT(K,NZ)=RSLT(K,NZ)*100./EFF
      RSLT(2,NZ)=FLOAT(NBCR)/EFF
      RETURN
      END
C
C
C
C
C
C
      FUNCTION ICHOIX(NEL)
C
C*** FONCTION ENGENDRANT DES NOMBRES ENTIERS ALFATOIRES
COMPRIS ENTRE 1 ET NEL (NOMBRE D'ELEMENTS) ***
C
      CHX=RANF(-1)
      PROP=FLOAT(NEL)*CHX
      ICHOIX=XINTF(PROP)+1
      RETURN
      END

```

6 -- PRINCIPAUX RESULTATS

Les informations apportées par la série de simulations que nous avons entreprises constituent, par rapport à l'objet de notre recherche qui est avant tout méthodologique, un point relativement secondaire. Il nous a cependant paru utile de présenter un échantillon des résultats de cette expérience simulée, pour pouvoir dresser ensuite un bilan de l'expérience. Rappelons toutefois auparavant que les résultats présentés concernent une *expérimentation sur un modèle théorique* relatif au brassage social, et non une recherche sur les mécanismes de choix du conjoint.

Nous exposerons d'abord les relations qui sont apparues entre les variables d'entrée et de sortie de notre modèle, et en particulier l'influence du type de décision simulé sur les taux d'homogamie ; l'analyse de ces relations était d'ailleurs le but de notre série de simulations. Ensuite, nous confronterons avec les données de l'enquête quelques échantillons des populations que nous avons artificiellement reconstituées. Nous terminerons ce chapitre sur un bilan de cette expérience.

6.1. L'INFLUENCE DU MODELE DE DECISION SUR L'HOMOGAMIE.

L'étude de cette influence était l'objectif principal visé par notre expérience. Afin d'éviter la lecture, toujours fastidieuse, des tableaux de résultats, nous avons préféré présenter ceux-ci sous une forme graphique. La plupart des figures qui suivent sont des diagrammes dans lesquels nous avons porté, en abscisses, le degré d'intensité de l'attitude exogamique ou endogamique, et en ordonnées le pourcentage d'homogamie. Rappelons que le degré d'intensité exprime à la fois le sens de l'attitude, et un nombre de choix maximum autorisé. Ainsi, une valeur pour ce degré de -7 correspond à une attitude exogamique, avec 7 choix autorisés ; une valeur de 5, à une attitude endogamique avec 5 choix autorisés. La valeur *zéro* correspond à la panmixie.

Ainsi que nous l'avons signalé (§ 5.2.2.b), nous n'avons pas répété chaque observation comme nous en avons eu préalablement l'intention (cf. § 4.2.2.e). L'objet de ces répétitions de l'expérience aurait été de donner plus de précision aux résultats enregistrés. Nous avons effectué quelques essais, en réalisant 10 fois de suite la même simulation : l'écart-type des taux d'homogamie relevés a presque toujours été de l'ordre de 1 %. Compte-tenu des variations importantes de certains des taux d'homogamie d'un modèle de décision à l'autre, il nous a semblé que la précision obtenue pour une observation isolée était suffisante. Si les phénomènes étudiés avaient consisté en des variations plus fines, nous aurions dû nous en tenir à notre intention première de répéter chaque observation. Dans ces conditions, notre expérience aurait vite atteint un coût prohibitif : elle aurait nécessité vraisemblablement un peu plus de 6 heures d'ordinateur CDC 3.600 (*). L'in-

(*) La durée de notre série de simulations (16 modèles de décision, 21 degrés d'intensité, soit 336 simulations) a été de 38 minutes sur le *Control Data 3600* ; soit une durée moyenne pour une simulation de l'ordre de 6 secondes et demie. Si nous avions réitéré chaque simulation, la durée totale de l'expérience aurait été d'environ 6 heures 10 minutes. Pour estimer les durées correspondantes sur l'*IBM 360/75* (en temps d'unité centrale), il convient de multiplier ces temps par 0,5.

térêt de cette recherche étant essentiellement d'ordre méthodologique, cette application ne nous a pas paru justifier une telle dépense. Les résultats définitifs relatifs à la panmixie confirment *a posteriori* le bien-fondé de cette décision. En effet, pour chacun des 16 modèles étudiés, la situation de panmixie a été nécessairement identique. Nous disposons donc de 16 exemples de panmixie, semblables aux fluctuations aléatoires près. Pour chacune des quatre variables étudiées, la dispersion des résultats reste acceptable, comme il ressort du tableau ci-dessous :

<i>Variable</i>	<i>Moyenne</i>	<i>Ecart-type</i>
Localité de naissance	41,922 %	0,495 %
Niveau d'instruction	47,183 %	0,628 %
Religion	43,110 %	1,020 %
Origine sociale	25,744 %	1,022 %

Tableau n° 13 : Tendances centrale et dispersion des taux d'homogénéité dans l'hypothèse de panmixie (16 observations).

Nous commençons par étudier les résultats correspondant aux modèles à un seul critère de choix. Nous abordons ensuite les modèles à deux critères. Nous examinons enfin les différences enregistrées entre les critères.

6.1.1. Modèles de décision à un seul critère de choix.

L'étude des quatre simulations relatives à ces modèles nous apporte deux ensembles principaux d'informations qui seront extrapolables

aux autres modèles :

- sur la forme générale de la fonction d'homogamie,
- sur le comportement des variables non retenues comme critères.

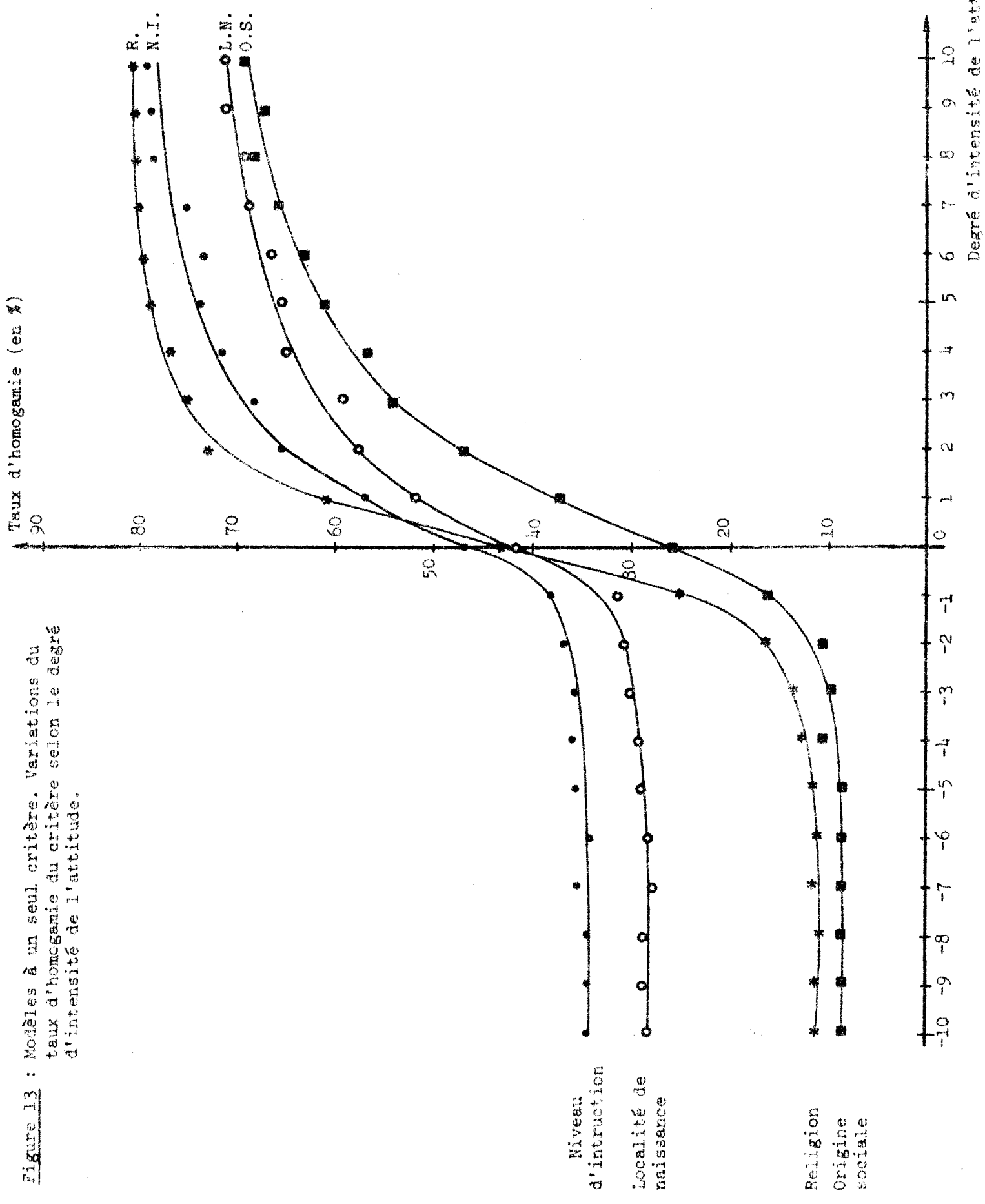
a) La forme générale de la fonction d'homogamie.

Si nous ne considérons que les seules variables ayant joué le rôle de critère dans le choix du conjoint (fig. 13), nous constatons que les courbes qui expriment le taux d'homogamie pour un critère de choix en fonction du degré d'intensité de l'attitude exogamique ou endogamique présentent des caractéristiques semblables. Ces caractéristiques sont les suivantes :

- le taux d'homogamie est une *fonction monotone croissante* du degré d'intensité tel que nous l'avons défini ;
- cette fonction a une *valeur minima*, supérieure au minimum théorique (0 %), et variable d'un critère à l'autre ;
- cette fonction a une *valeur maxima*, inférieure au maximum théorique (100 %), et variable d'un critère à l'autre ;
- la *pente* la plus forte de la courbe se situe au point d'abscisse 0, correspondant à la panmixie ;
- l'écart entre l'ordonnée de la panmixie et celle de la valeur maxima est plus grand que l'écart entre l'ordonnée de la valeur minima et celle de la panmixie.

La première caractéristique de ces courbes ne constitue pas une surprise ; on pouvait s'attendre, en construisant le modèle général de la simulation, à voir croître l'homogamie au fur et à mesure de l'élévation du degré de sévérité dans l'endogamie, et réciproquement à voir augmenter l'hétérogamie avec l'élévation du degré de sévérité dans l'exogamie.

Figure 13 : Modèles à un seul critère. Variations du taux d'homogamie du critère selon le degré d'intensité de l'attitude.



L'existence d'asymptotes à la courbe qui ne se confondent pas avec les limites théoriques (0 % et 100 %) est moins évidente. Elle nous paraît être la conséquence de deux facteurs : les différences existant au niveau de l'isolat entre la répartition des hommes et celle des femmes pour les quatre variables considérées ; et la faible dispersion relevée pour certaines de ces variables. Un exemple simple illustrera la manière dont ces facteurs peuvent jouer au niveau de l'isolat. Soit un isolat de 100 hommes et 100 femmes, dont le niveau d'instruction, avant dissociation des couples, se répartit comme dans le tableau n° 14. Les distributions marginales correspondant respectivement au niveau d'instruction des hommes et à celui des femmes sont relativement dissemblables. La somme des valeurs absolues des différences est égale à 24 % ; ces différences, évaluées

<i>Niveau d'instruction de la femme</i>	<i>Niveau d'instruction du mari</i>				<i>TOTAL</i>
	Primaire	Technique	Secondaire	Supérieur	
Primaire	51	6	9	-	66
Technique	2	3	2	1	8
Secondaire	5	4	10	6	25
Supérieur	-	-	-	1	1
<i>TOTAL</i>	58	13	21	8	100

Tableau n° 14 : Niveau d'instruction des conjoints dans un isolat (données empiriques).

selon le test du χ^2 , sont significatives au seuil de $p = 0,025$ (*). D'autre part, les dispersions de ces distributions sont assez faibles, en raison du nombre élevé d'individus ne possédant pas un niveau d'instruction supérieur au certificat d'études primaires. A partir de ces distributions, nous pouvons construire les tableaux croisés correspondant aux hypothèses d'endogamie ou d'exogamie extrêmes. Supposons tout d'abord que les couples dissociés aient été reconstitués en utilisant un modèle de décision correspondant à un degré très élevé d'attitude endogamique relative au niveau d'instruction. L'effectif maximum de couples ayant le même niveau d'instruction est égal au plus faible des deux effectifs marginaux correspondants. Dans notre exemple, où 58 hommes et 66 femmes ont un niveau d'instruction inférieur ou égal au C.E.P., il n'y aura au maximum que 58 couples pour lesquels le niveau d'instruction des deux conjoints sera inférieur ou égal au C.E.P. On voit dans le tableau n° 15 que le degré le plus élevé possible d'homogamie est de 88 %, en raison des disparités existant entre les distributions marginales.

Niveau d'instruction de la femme	Niveau d'instruction du mari				TOTAL
	Primaire	Technique	Secondaire	Supérieur	
Primaire	58	5	-	3	66
Technique	-	8	-	-	8
Secondaire	-	-	21	4	25
Supérieur	-	-	-	1	1
TOTAL	58	13	21	8	100

Tableau n° 15 : Niveau d'instruction des conjoints dans l'hypothèse d'homogamie maximum.

(*) $\chi^2 = 7,498$, avec $v = 3$.

cf. p.ex. : Comparaison simultanée de plusieurs distributions de fréquences, in 149, Pp. 245-246. Notons d'ailleurs que la meilleure mesure de la disparité entre la distribution des hommes selon le niveau d'instruction et celle des femmes est un coefficient mesurant le degré de liaison statistique entre le sexe et le niveau d'instruction. Le coefficient de contingence de KENDALL vaudrait ici $C = 0,19$.

Dans l'hypothèse d'une attitude exogamique très forte quant au niveau d'instruction, la faible dispersion des distributions marginales impose un taux minimum d'homogamie sensiblement supérieur à 0 %. En effet, les 58 hommes n'ayant qu'une instruction primaire pouvant épouser 1 femme ayant le niveau de l'enseignement supérieur, 25 femmes ayant le niveau du secondaire, et 8 ayant une formation technique, il restera 24 hommes qui devront nécessairement choisir leur épouse parmi les femmes de niveau primaire, soit un taux minimum d'homogamie de 24 %.

Niveau d'instruction de la femme	Niveau d'instruction du mari				TOTAL
	Primaire	Technique	Secondaire	Supérieur	
Primaire	24	13	21	8	66
Technique	8	-	-	-	8
Secondaire	25	-	-	-	25
Supérieur	1	-	-	-	1
TOTAL	58	13	21	8	100

Tableau n° 16 : Niveau d'instruction des conjoints dans l'hypothèse d'hétérogamie maximum.

Le fait que la pente la plus forte de la courbe d'homogamie se situe au voisinage du point correspondant à la panmixie peut être expliqué de manière intuitive. Considérons, dans l'exemple que nous avons uti-

lisé précédemment, les 58 hommes n'ayant qu'une instruction primaire et supposons, pour simplifier, que ce soit toujours l'homme qui choisisse son épouse. Dans l'hypothèse de panmixie, la probabilité moyenne que chacun d'eux épouse une femme de même niveau est par définition égale à 0,66 ; l'espérance mathématique du nombre de couples primaires est en gros de 38 couples. Il reste donc 20 hommes et 28 femmes de niveau primaire qui, lors du premier tirage au sort, n'ont pas rencontré un partenaire de même niveau. On calculerait de même que, sur les 200 personnes de la situation de départ, 112 (soit 56 hommes et 56 femmes) n'ont pas rencontré un partenaire de même niveau. Si ceux-ci ont la possibilité de procéder à un second tirage au sort, c'est-à-dire à une première tentative de choix, la probabilité moyenne pour qu'un célibataire de niveau primaire épouse une femme de même niveau est maintenant égale à $\frac{28}{56} = 0,50$; l'espérance mathématique du nombre supplémentaire de couples primaires est de $20 \times 0,50 = 10$ couples. En poursuivant le même raisonnement, on montrerait qu'au troisième tirage au sort, le nombre supplémentaire de couples primaires serait de 5 environ ; au quatrième tirage, de 2, etc.. L'effectif total de couples primaires passe successivement de 38 (panmixie) à 48, puis 53, puis 55 ; le taux d'homogamie a donc une accélération négative. On ferait le même raisonnement pour un comportement exogamique, en montrant que le taux d'hétérogamie est négativement accéléré à partir de la panmixie.

La dernière caractéristique de la courbe est que l'écart entre la valeur maxima d'homogamie et celle correspondant à la panmixie est plus grand que l'écart entre celle-ci et la valeur minima. On pourrait penser, au premier abord, que cette particularité est due au fait qu'en aucun cas, le taux d'homogamie correspondant à la panmixie n'est au moins égal à 50 % ; ceci pouvant introduire une dyssymétrie purement accidentelle, dépendant non pas de l'équation générale des fonctions,

mais des valeurs particulières prises par les paramètres pour un critère donné. Pour éprouver la validité d'un tel raisonnement, il est nécessaire de pousser un peu plus loin l'analyse de la relation liant le taux d'homogamie du critère au degré d'intensité de l'attitude simulée. Compte-tenu de la relative imprécision des résultats, nous nous sommes limité à une analyse intuitive, fondée sur l'examen des fonctions de probabilité observées. Dans le cas de simulations répétées de manière à réduire l'imprécision sur les observations, il conviendrait évidemment de mesurer les coefficients d'ajustement aux fonctions théoriques envisagées.

Notre première hypothèse a été que la courbe de variation du taux d'homogamie, était une ogive de GALTON. Pour en éprouver empiriquement le bien-fondé, nous avons reporté les courbes observées sur du papier gaussien-arithmétique. Le résultat n'a pas été une famille de droites, mais de nouvelles courbes en S. L'ogive de Galton étant formée par les fréquences cumulées d'une distribution de LAPLACE-GAUSS, nous avons jugé utile, pour étudier les distributions que nous avons obtenues expérimentalement, de construire les courbes correspondant à l'accroissement du taux d'homogamie en fonction du degré d'intensité de l'attitude. Pour atténuer l'effet des fluctuations aléatoires, nous avons regroupé les degrés pour lesquels le taux d'accroissement variait peu. Les histogrammes ainsi obtenus pour les quatre critères (fig. 14a et 14b, pages 334 et 335) montrent tous une forte asymétrie : la moyenne est située systématiquement à droite de la valeur modale.

Une seconde hypothèse serait qu'il existe bien une fonction de répartition des taux d'accroissement de l'homogamie obéissant à la loi de LAPLACE-GAUSS, mais que cette relation est masquée par la distorsion existant entre les comportements et les attitudes, c'est-à-dire entre le

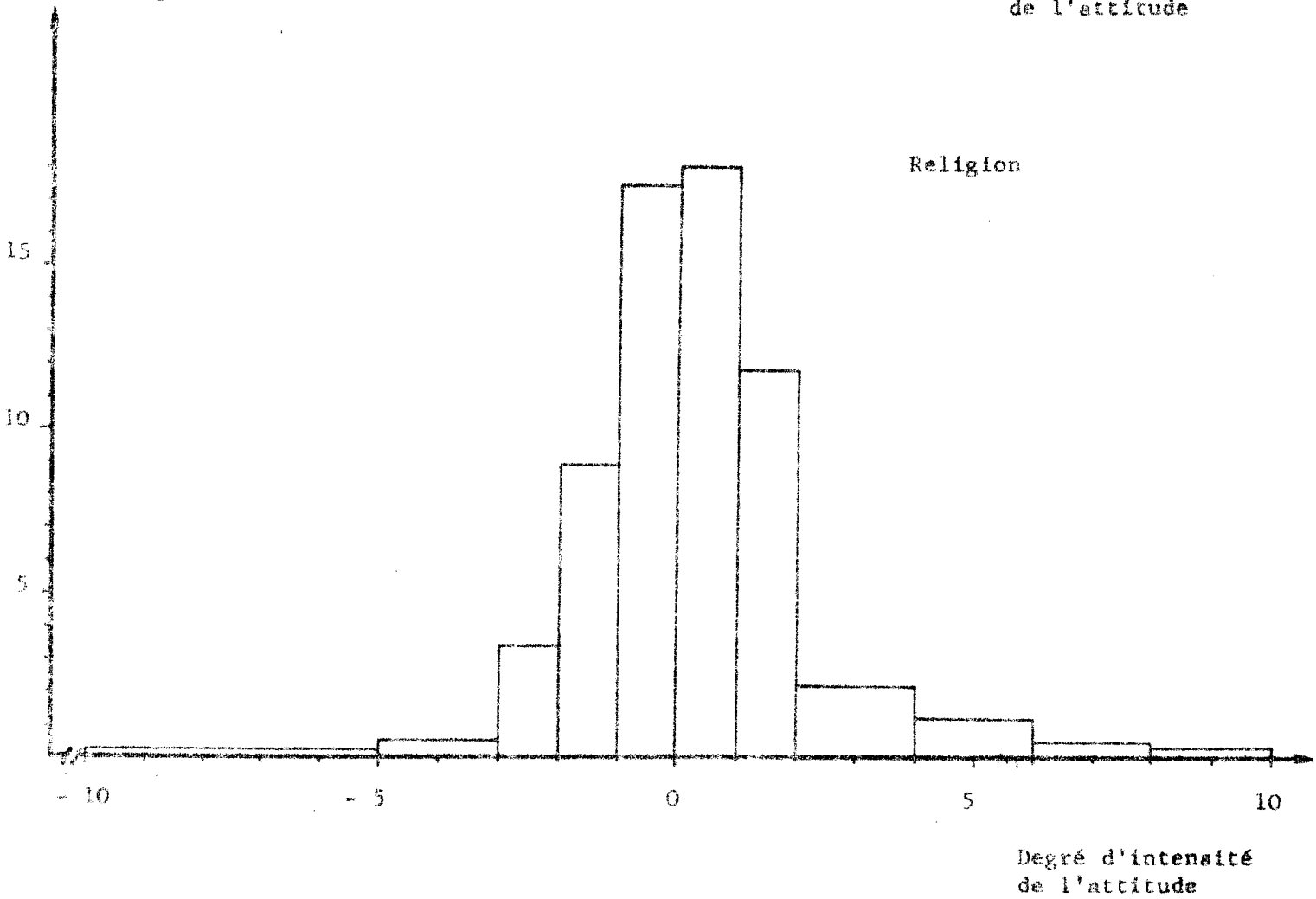
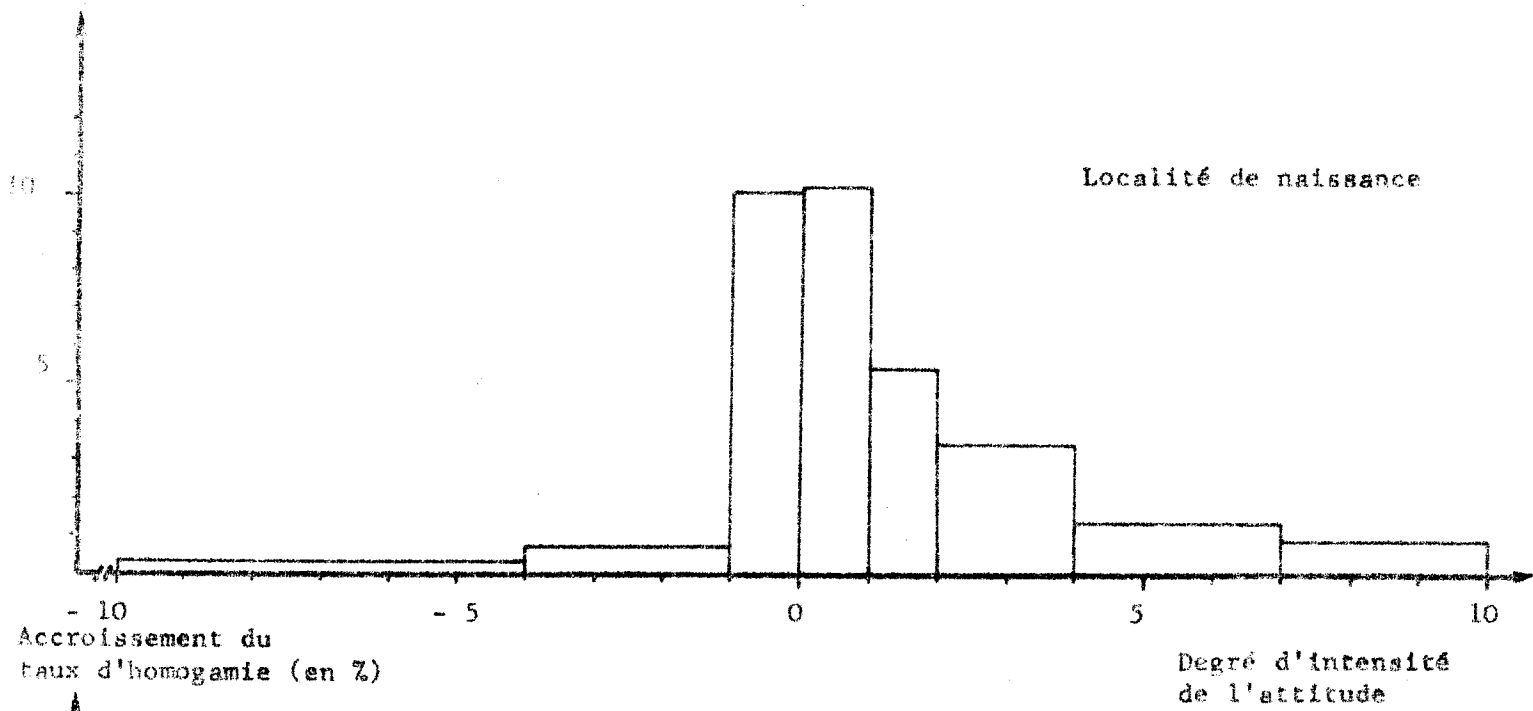
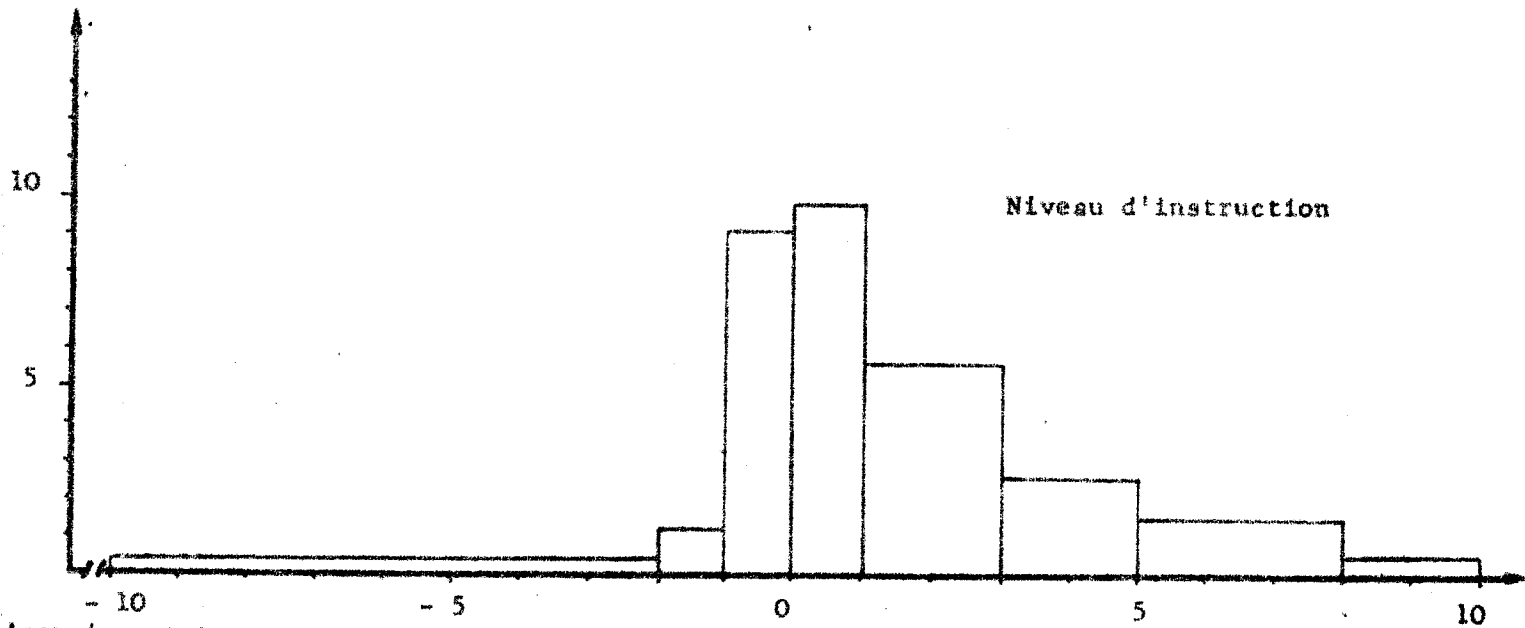


Figure 14 a - Pourcentage d'accroissement du taux d'homogamie selon le degré d'intensité de l'attitude (endogamique ou exogamique)

Accroissement
du taux d'homogamie (en %)



Accroissement
du taux d'homogamie (en %)

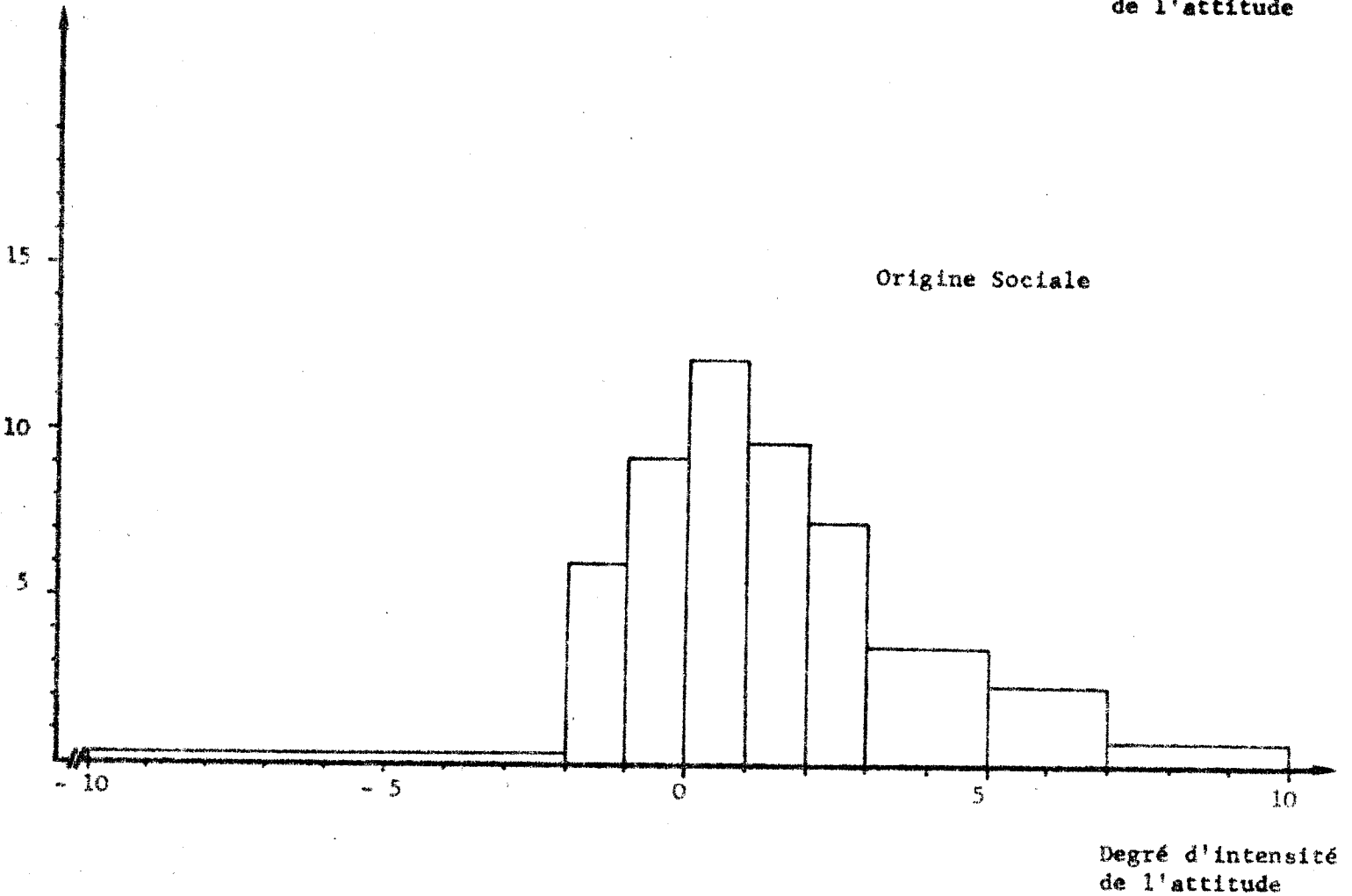


Figure 14 b. - Pourcentage d'accroissement du taux d'homogamie selon le degré d'intensité de l'attitude (endogamique ou exogamique)

nombre de choix réellement effectués, et le degré d'intensité qui exprime le nombre de choix théoriquement possibles. Si l'on compare le nombre de choix réellement effectués au degré d'intensité de l'attitude, on remarque que la relation entre ces deux grandeurs n'est pas linéaire (fig. 15, page 337) ; il n'est donc pas impossible que la fonction liant le taux d'homogamie au nombre moyen de choix soit du type ogive de GALTON (fig. 16, page 338). Si l'on construit les histogrammes correspondant au taux d'accroissement de l'homogamie selon le nombre moyen des choix réels (fig. 17a et 17b, pages 339 et 340), on constate que la dyssymétrie de la fonction de répartition est plus accentuée que dans le cas précédent (fig. 14a et 14b) en raison du déplacement du mode vers la gauche ; il ne saurait donc s'agir d'une courbe en cloche de LAPLACE-GAUSS.

Nous avons signalé que l'imprécision de nos mesures, qui sont relatives à une seule simulation sans répétition des observations, rend malaisé tout essai d'ajustement d'une fonction théorique. Il serait cependant intéressant, au cours d'une recherche ultérieure, de déterminer l'équation générale des courbes d'accroissement du taux d'homogamie. Nous savons déjà que cette équation doit être recherchée dans une extension de la loi binomiale, qui peut être par exemple une distribution dérivée de la loi hypergéométrique, ou de la loi de LAPLACE-GAUSS (*). *A priori*, deux types de courbes nous paraissent de-

(*) Nous savons en effet qu'au niveau de l'individu, la procédure de choix que nous avons simulée peut être assimilée formellement à une série de tirages avec remise dans une population (l'univers de choix) présentant un caractère dichotomique (conjoint acceptable ou non) ; cette procédure répétée aboutirait à une distribution binomiale si le tirage ne se trouvait interrompu par la procédure d'appariement. Au niveau de l'isolat, les appariements successifs constituent une série de tirages au sort sans remise et exhaustifs ; les caractéristiques des couples devraient donc obéir à une loi dérivée de la loi hypergéométrique. Au niveau de l'ensemble de la population, la superposition des distributions correspondant à chaque isolat pourrait avoir comme limite, à une transformation près, la loi de LAPLACE-GAUSS.

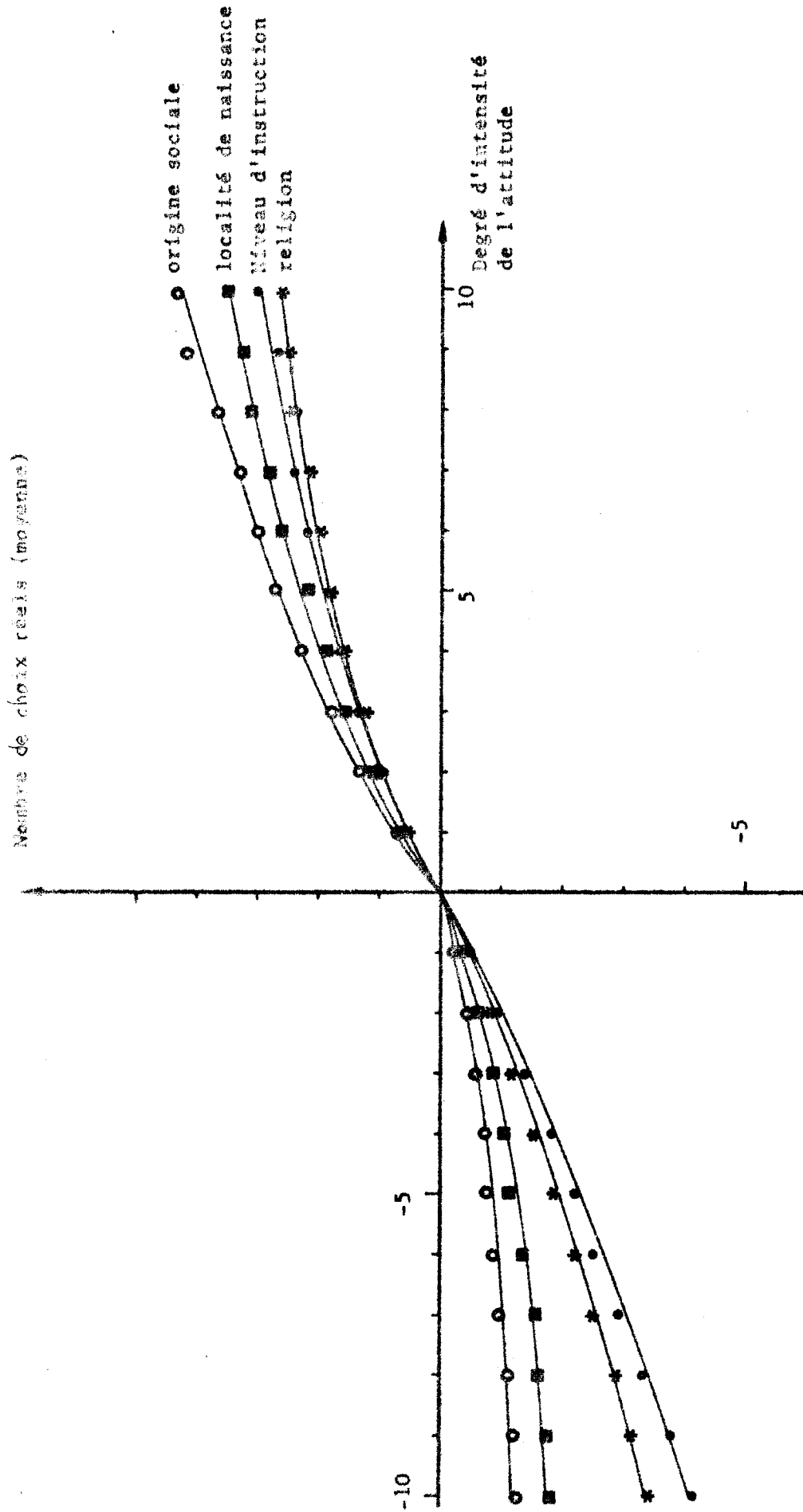


Figure 15 - Relation entre le nombre moyen de choix réellement effectués et le degré d'intensité de l'attitude (modèles à un seul critère.)

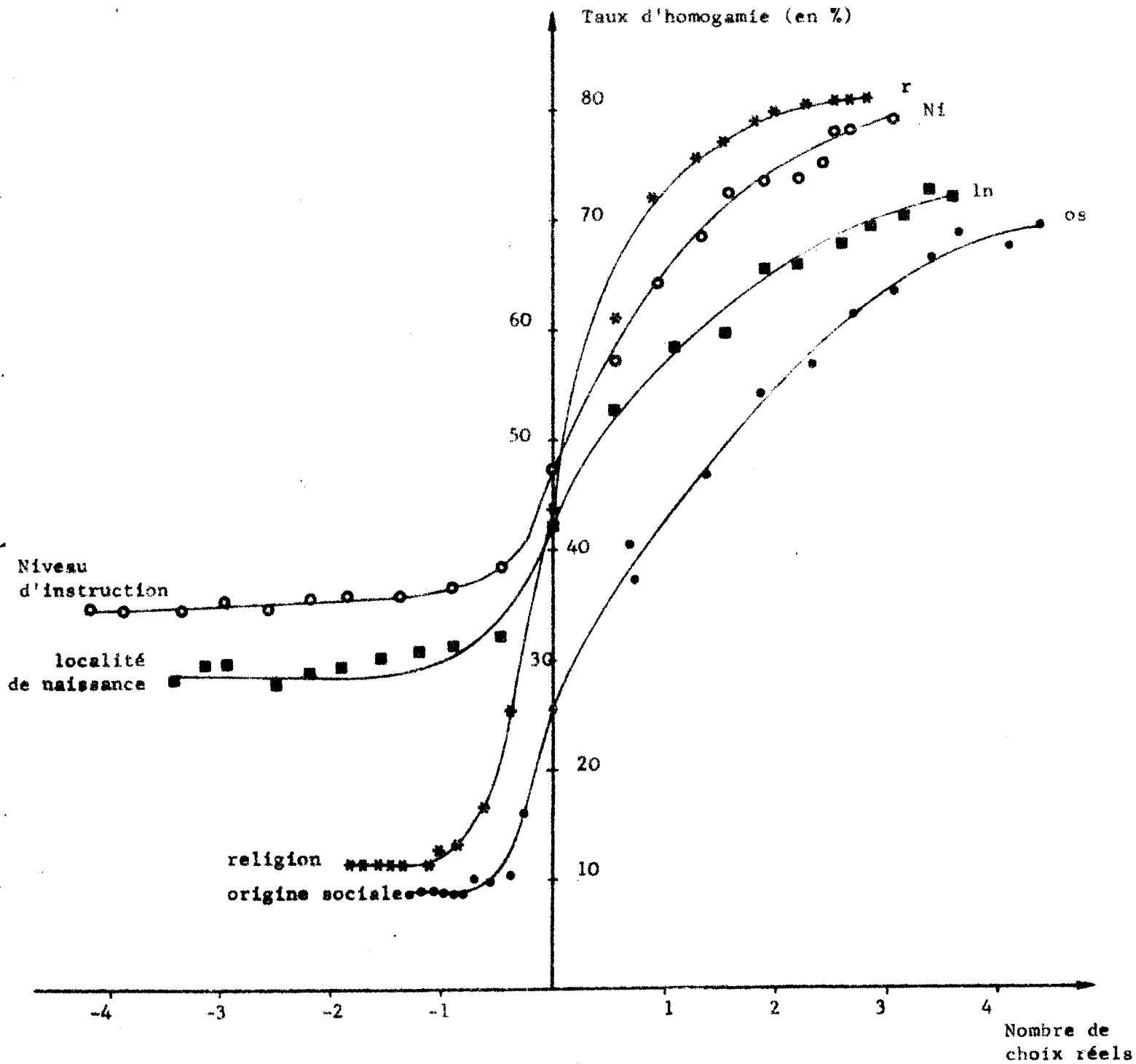
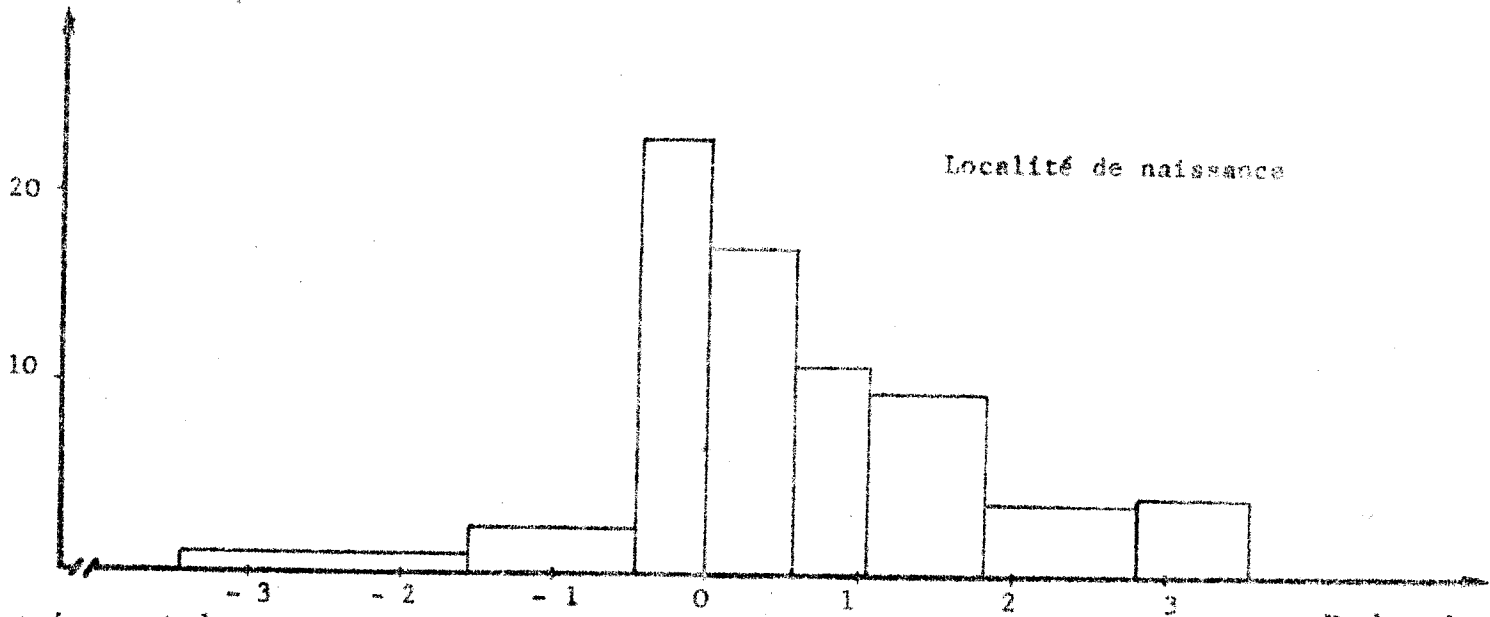
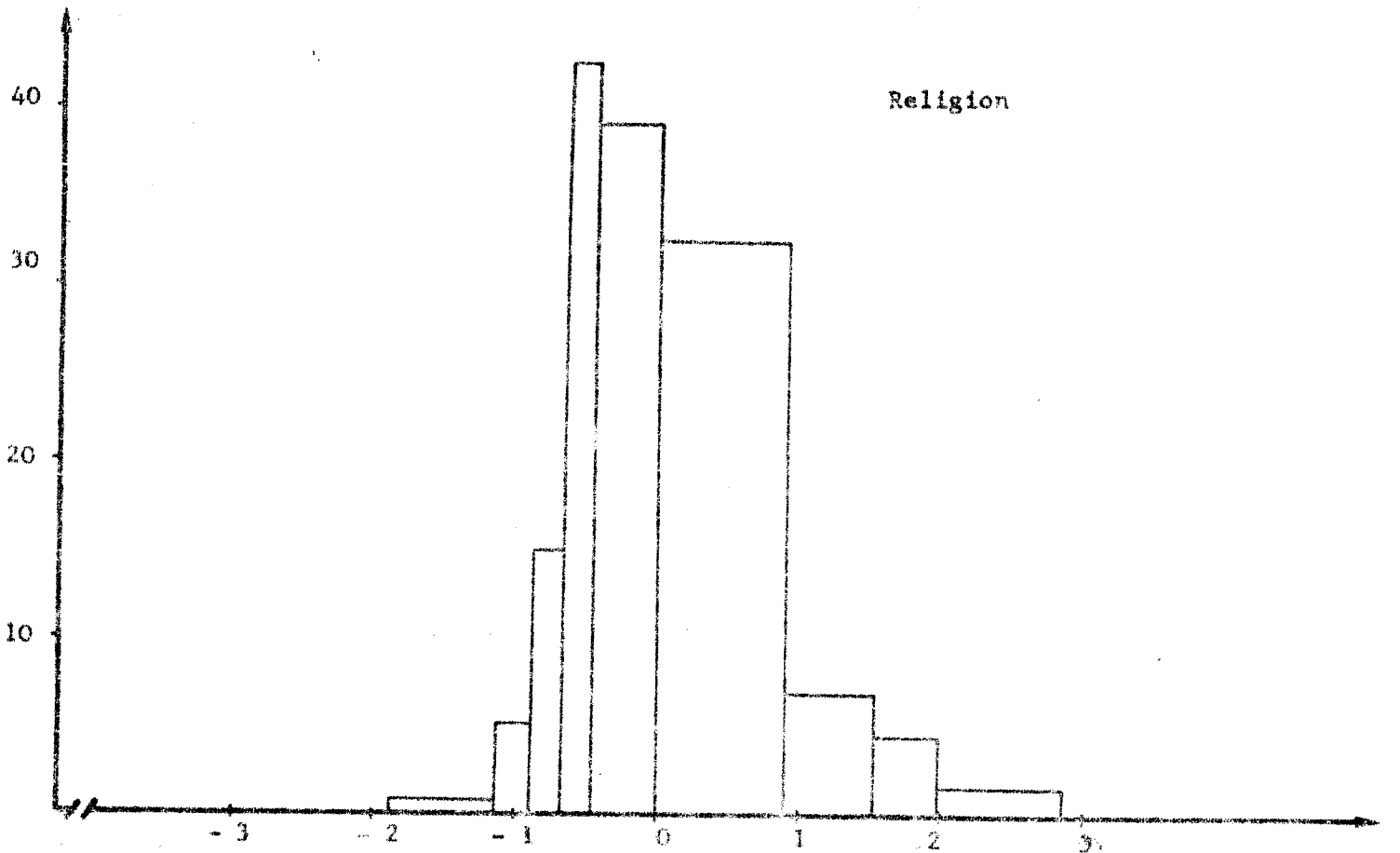


Figure 16 - Variations du taux d'homogamie selon le nombre moyen de choix réellement effectués (modèles à un seul critère)

Accroissement du
taux d'homogamie (en %)



Accroissement du
taux d'homogamie (en %)



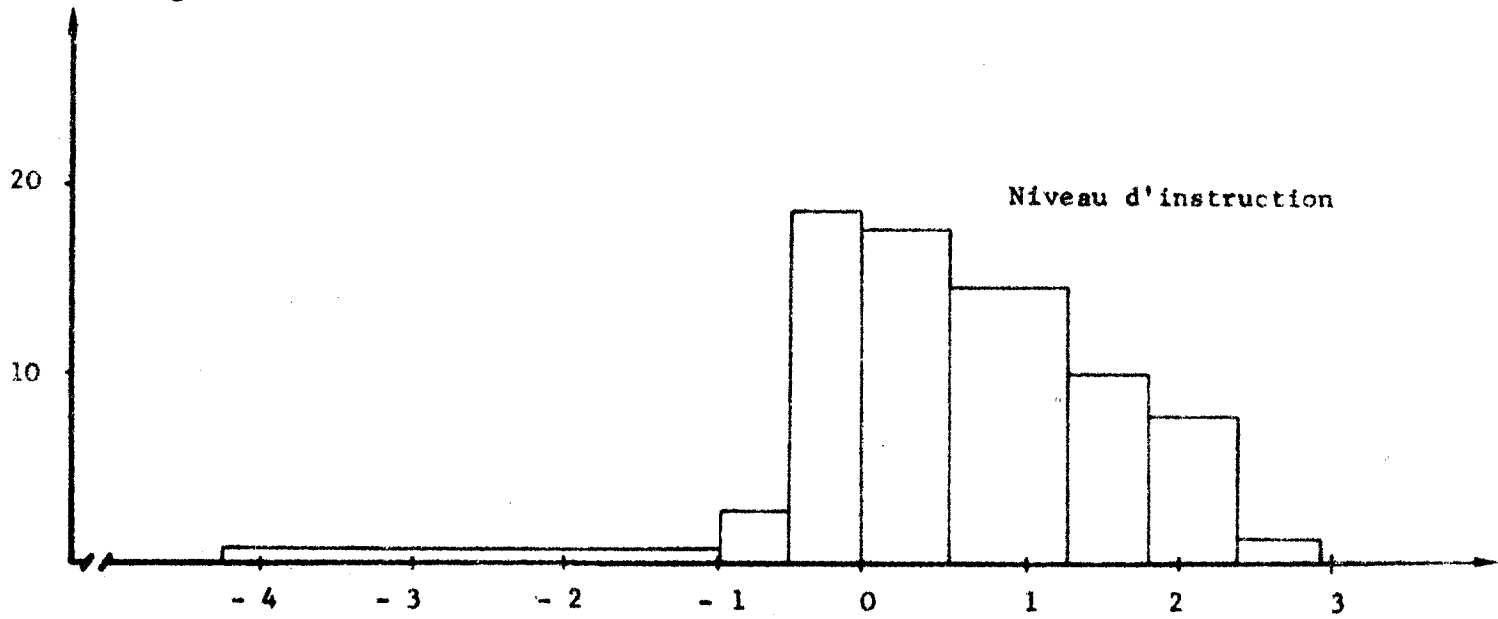
Nombre de
choix réels

Religion

Figure 17 a. - Pourcentage d'accroissement du taux d'homogamie selon le nombre réel de choix (ou de rejets) effectués.

Nombre de
choix réels

Accroissement du
taux d'homogamie (en %)



Accroissement du
taux d'homogamie (en %)

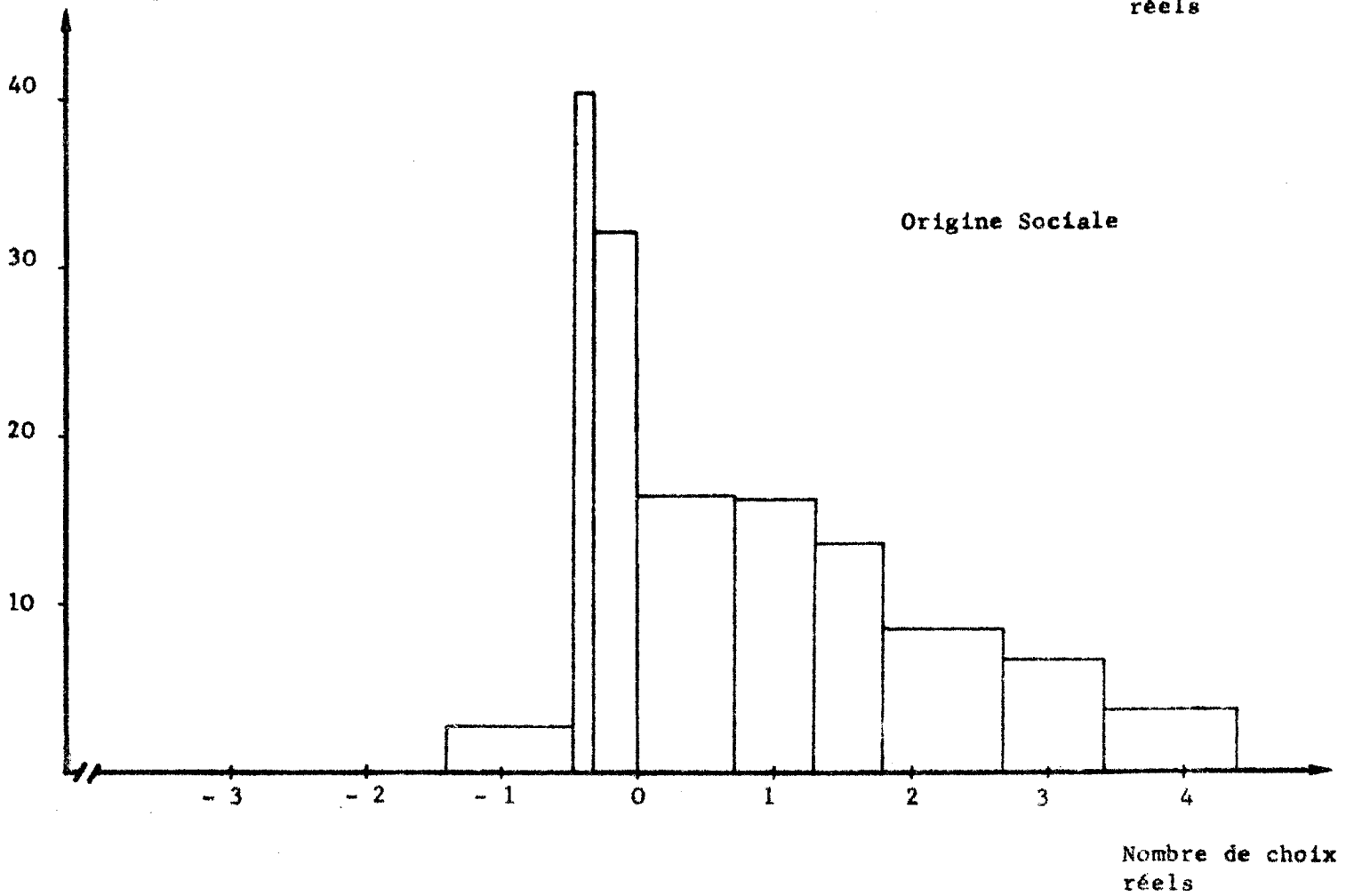


Figure 17 b. - Pourcentage d'accroissement du taux d'homogamie selon le nombre réel de choix (ou de rejets) effectués.

voir retenir l'attention. Le premier est la distribution de type IV de PEARSON, définie par une équation différentielle déduite de la loi binomiale. Le second est la distribution de GALTON-Mc ALISTER appelée également "loi de l'effet proportionnel" ; cette fonction est une généralisation de la loi de LAPLACE-GAUSS, en supposant que les parts de la variable, dues aux différents facteurs jouant de manière aléatoire, sont multiplicatives, et non plus additives (*). Il resterait de toute façon à justifier du choix ainsi fait, ce qui ne peut se fonder que sur un parallèle entre la dynamique du phénomène étudié, et les bases axiomatiques de la fonction adoptée.

b) Le comportement des variables non retenues comme critère.

Dans chacun des modèles à un seul critère, trois variables sont de simples variables de sortie, n'intervenant pas dans le processus de choix. Si l'on étudie les variations du taux d'homogamie selon les variables non retenues comme critère de choix, on doit s'attendre à observer les phénomènes suivants :

1 - Pour les variables totalement indépendantes du critère de choix, le taux d'homogamie devrait être constant, aux fluctuations

(*) L'ajustement d'une des distributions étudiées par PEARSON ne devrait poser aucun problème, la détermination de la courbe se faisant quasi-automatiquement, à partir du calcul des moments centrés d'ordre 2, 3 et 4 de la distribution empirique. Par contre, l'adoption de la loi de GALTON-Mc ALISTER soulèverait de sérieuses difficultés. Outre la complexité des calculs (avantageusement remplacés par les tâtonnements des méthodes graphiques), cette distribution impose à la variable correspondant au nombre de choix de ne prendre que des valeurs positives. Il conviendrait donc alors de choisir une autre origine que le nombre de choix correspondant à la panmixie ; mais il semble que le taux d'homogamie correspondant à la panmixie joue un rôle important dans l'équation de la fonction, et il y a lieu de s'interroger sur le bien-fondé d'un changement d'abscisses qui masquerait l'importance de ce paramètre.

aléatoires près. En effet, si (par exemple) le niveau d'instruction est totalement indépendant de l'origine sociale, c'est-à-dire que la probabilité d'atteindre un niveau d'instruction donné est la même pour toutes les couches sociales, l'adoption d'une attitude fortement endogamique ou exogamique selon le milieu social n'a aucune influence sur les différences de niveau d'instruction entre les conjoints.

2 - Pour les variables fortement liées au critère de choix, le taux d'homogamie devrait varier en fonction du taux d'homogamie du critère. Si (par exemple) le niveau d'instruction moyen croît au fur et à mesure que l'on s'élève dans l'"échelle sociale", une attitude endogamique selon l'origine sociale devrait avoir pour effet d'apparier de préférence des individus ayant un niveau d'instruction comparable. Ceci étant vrai quel que soit le sens et la forme de la liaison (mais à condition que la liaison soit de même sens pour les hommes et pour les femmes), puisque deux individus proches selon l'origine sociale sont en moyenne également proches (quoiqu'à un degré moindre) selon les variables liées à l'origine sociale, même si cette "proximité" se traduit par exemple par un statut social élevé et un niveau d'instruction faible (liaison négative). La réciproque devrait être vérifiée dans le cas d'une attitude exogamique.

3 - Plus la liaison entre une variable et le critère est forte, plus les variations du taux d'homogamie selon le degré d'intensité de l'attitude devraient être fortes. En effet, supposons qu'à la limite la variable étudiée et le critère aient un degré de liaison maximum, c'est-à-dire qu'il s'agisse en fait de la même variable, figurant sous deux formes différentes. On comprend aisément que les variations du taux d'homogamie selon la variable seront identiques à celles du taux d'homogamie selon le critère : elles auront l'allure des courbes décrites au paragraphe précédent. A l'inverse, si le de-

gré de liaison est nul (indépendance totale), le taux d'homogamie selon la variable restera constant, c'est-à-dire égal aux taux correspondant à la panmixie, quel que soit le degré de l'attitude selon le critère. Entre ces deux cas extrêmes, on devrait trouver des variations du taux d'homogamie selon la variable étudiée d'autant plus fortes que les conjoints ont plus de chances d'être semblables, c'est-à-dire que la liaison avec le critère de choix est plus élevée.

Si nous reportons sur un graphique les variations du taux d'homogamie selon les quatre variables étudiées pour les modèles à un seul critère où elles ne jouent aucun rôle dans la décision, on constate qu'aux fluctuations aléatoires près, le taux d'homogamie est pratiquement constant pour toutes les variables (cf. figure 18, page 344). Si les hypothèses qui précèdent sont exactes, nous devons en déduire que les quatre variables qui dans notre modèle servent alternativement de critère et de simple variable de sortie sont pratiquement indépendantes les unes des autres, et ceci tant dans la sous-population des hommes que dans celle des femmes. Nous avons vérifié cette déduction en deux étapes, dont les résultats sont résumés par la figure 19, page 345). Nous avons tout d'abord testé (χ^2) l'existence d'une liaison pour les $2 C_4^2 = 12$ paires de variables envisageables ; toutes les liaisons sauf une sont significatives au seuil de $p \leq 0,001$ (cette dernière ne l'étant que pour $p \leq 0,01$). Nous avons ensuite mesuré l'intensité de ces liaisons (coefficient C de KENDALL) : celle-ci varie entre 0,14 et 0,52. On pourrait admettre que pour une liaison relativement faible, de l'ordre de 0,20, l'influence de la liaison avec le critère soit masquée par les fluctuations aléatoires.

Nous avons isolé le cas correspondant au degré de liaison le plus élevé (origine sociale et niveau d'instruction), et nous avons représenté sur la figure 20 (page 346) les variations du taux d'homogamie

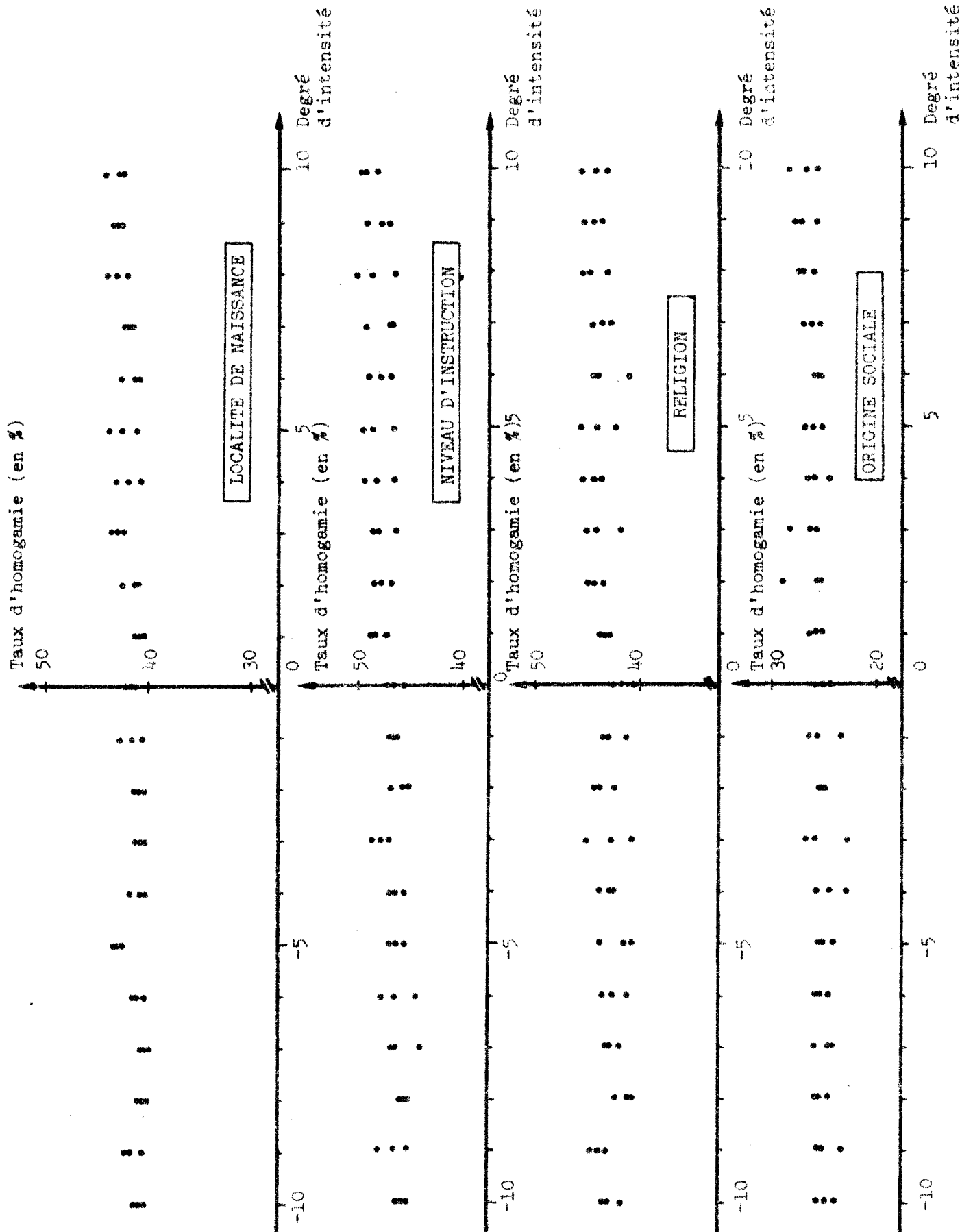
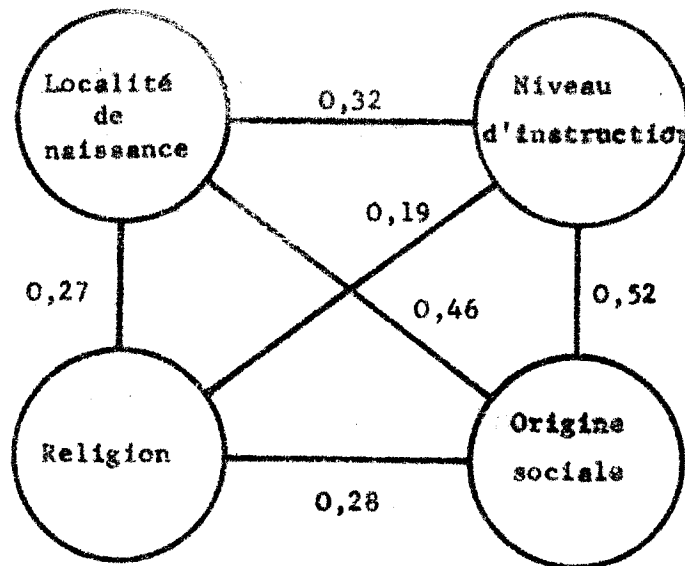


Figure 18 - Modèles à un seul critère. Variations du taux d'homogamie des variables non retenues comme critère selon le degré d'intensité de l'attitude.

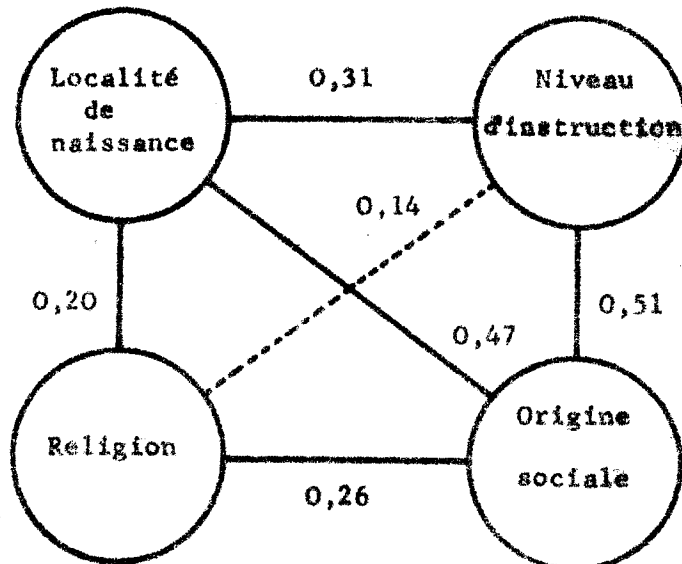


Hommes

Liaison significative (χ^2)
au seuil de :

— $P \leq 0,001$

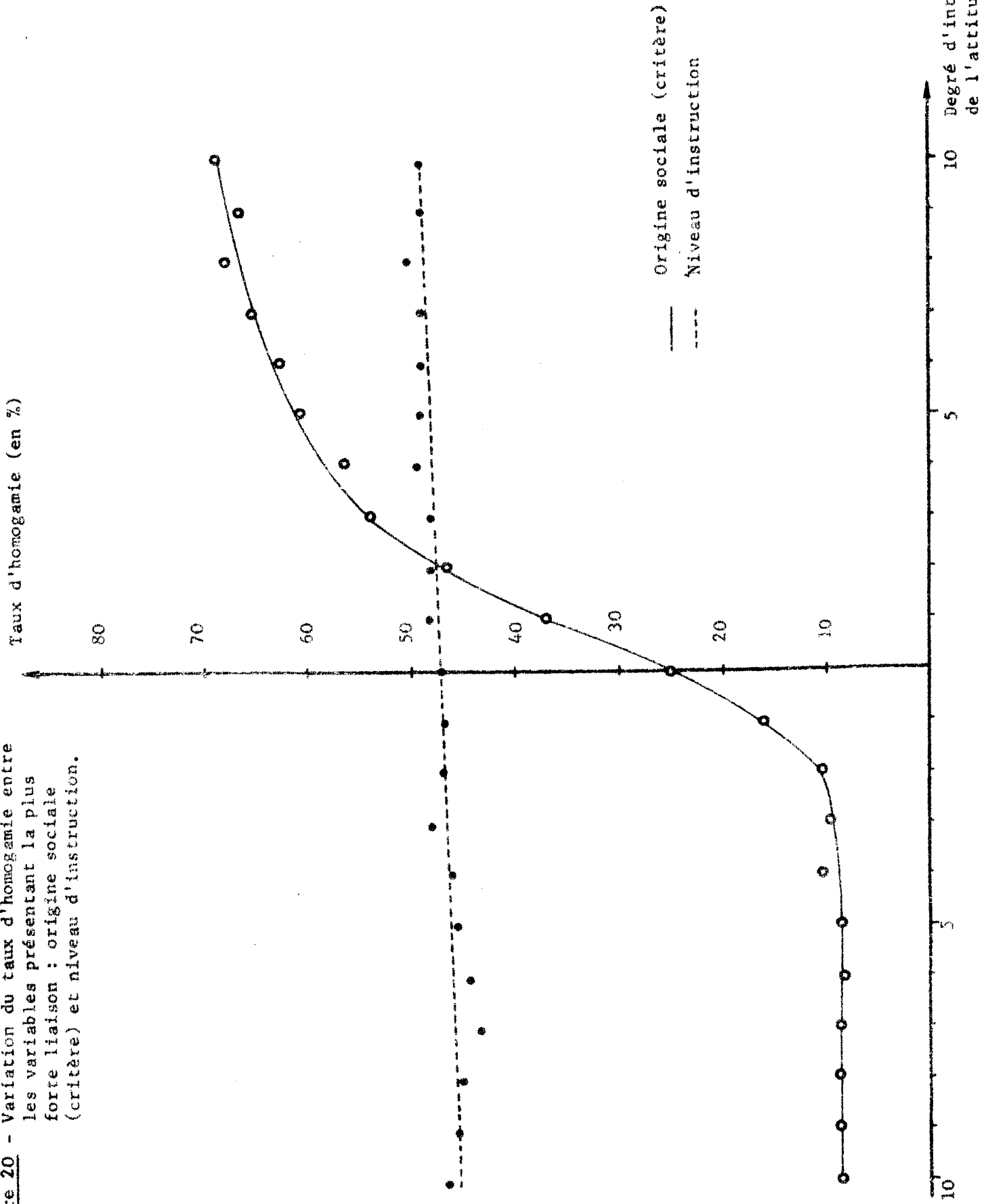
- - - $P \leq 0,01$



Femmes

Figure 19 - Liaisons existant entre les variables d'entrée du modèle, mesurées au niveau de l'ensemble de la population (coefficient de contingence de KENDALL)

Figure 20 - Variation du taux d'homogamie entre les variables présentant la plus forte liaison : origine sociale (critère) et niveau d'instruction.



de ces deux variables lorsque l'origine sociale sert de critère de choix (on obtiendrait un graphique équivalent en considérant le cas où c'est le niveau d'instruction qui sert de critère). En pratique, on peut affirmer que le taux d'homogamie selon le niveau d'instruction reste constant. Ce qui signifierait que, pour expliquer le degré encore élevé d'homogamie observé dans la population, on ne peut retenir l'hypothèse selon laquelle le choix du conjoint s'effectuerait à partir d'un petit nombre de variables ; mais qu'au contraire le choix doit être un choix *global*, tenant compte de l'ensemble des facteurs de la personnalité d'autrui.

Il resterait à expliquer la faillite de nos hypothèses de départ. Si l'on examine attentivement la figure 20, on peut déceler un léger accroissement du taux d'homogamie selon le niveau d'instruction. Cette légère tendance est également perceptible, malgré les fluctuations aléatoires, dans la plupart des modèles de décision que nous avons simulés. On peut penser que les hypothèses que nous avons énoncées sont correctes, mais que l'effet d'attraction exercé sur les variables liées au critère de choix ne devient sensible que pour des degrés de liaison élevés (de l'ordre de 0,80). Cette dernière hypothèse ne saurait surprendre, si l'on considère la signification concrète d'un degré de liaison quelconque ; pour nous en tenir au cas le plus courant (r de BRAVAIS-PEARSON mesuré sur une distribution binormale), on constate que seules des corrélations au moins égales à $r = 0,75$ sont perceptibles dans la vie quotidienne (*). Nous concluons donc ce paragraphe en disant qu'il existe probablement une relation entre le taux d'homogamie d'une variable liée au critère et le degré d'attitude,

(*)

Il suffit pour s'en convaincre de regarder la forme du nuage de points correspondant à une valeur donnée du r de BRAVAIS-PEARSON.

cette relation étant une fonction monotone croissante du degré d'intensité de l'attitude ; mais qu'en pratique, cette relation n'a pas d'influence sur les variables étudiées, en raison de la faiblesse relative des liaisons observées.

6.1.2. Modèles de décision à deux critères de choix.

Les principales observations que nous avons faites à propos des modèles de décision à un seul critère de choix sont applicables aux modèles à deux critères. En particulier :

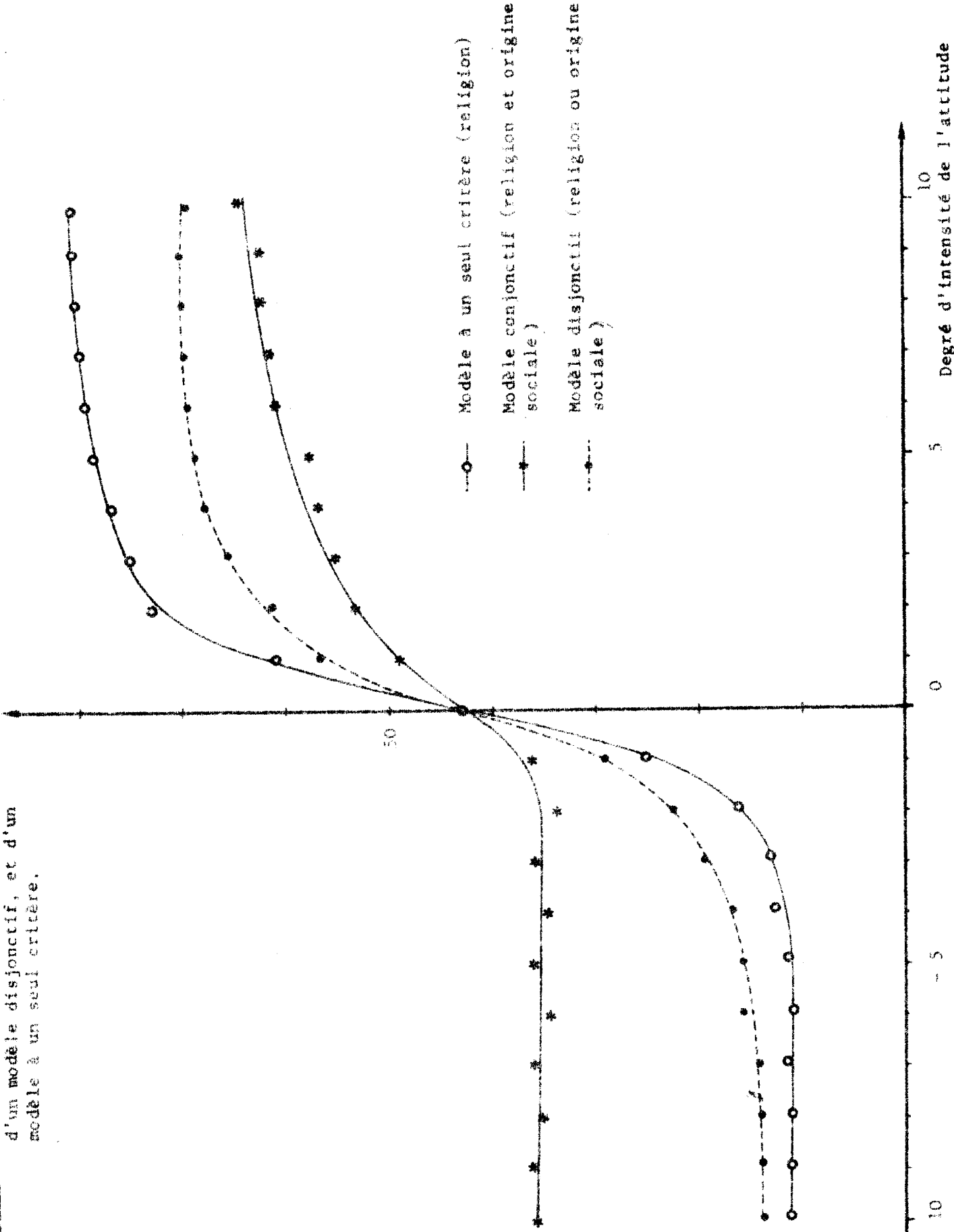
- le taux d'homogamie selon les variables non retenues comme critères reste sensiblement constant, quelle que soit l'intensité de l'attitude endogamique ou exogamique ;
- l'allure générale de la fonction liant le taux d'homogamie selon chacun des critères au degré d'intensité de l'attitude est la même pour les modèles à deux critères que pour les modèles à un critère ; et ceci quel que soit le critère.

Il existe cependant quelques différences importantes entre les modèles à un seul critère et les modèles à deux critères. Pour un critère donné (cf. figure 21, page 349) :

- la pente de la courbe exprimant la relation entre le taux d'homogamie selon le critère et l'intensité de l'attitude est toujours plus faible dans les modèles à deux critères ;
- le taux maximum d'homogamie dans les modèles à deux critères est toujours inférieur au taux maximum du modèle à un seul critère ;
- le taux minimum d'homogamie est toujours plus élevé dans les modèles à deux critères ; corrélativement, l'intervalle de variation du taux d'homogamie est plus restreint dans les modèles à deux

Taux d'homogamie (en %) selon la religion

Figure 21 - Comparaison d'un modèle conjonctif, d'un modèle disjonctif, et d'un modèle à un seul critère.



critères.

Ces propriétés des modèles à deux critères sont-elles des propriétés nécessaires ou bien ne s'agit-il que de cas particuliers, dus aux caractéristiques spécifiques de notre échantillon ? Pour pouvoir répondre à ces questions, il nous faut analyser plus complètement les résultats relatifs aux modèles conjonctifs et disjonctifs. En particulier, la comparaison des douze modèles à deux critères devrait nous permettre de dégager quelques préceptes généraux relatifs :

- à l'allure générale des courbes correspondant aux critères conjonctifs et disjonctifs ;

- à l'influence, sur la forme particulière de ces courbes, des caractéristiques spécifiques des différents critères.

a) Modèles conjonctifs et modèles disjonctifs.

Nous avons représenté sur la figure 22 (page 351) quelques-unes des courbes exprimant les variations du taux d'homogamie selon l'un des critères ; afin de rendre la comparaison plus facile, nous avons pris comme origine le taux d'homogamie correspondant à la panmixie. Il se dégage de cette figure l'impression que la pente des courbes correspondant aux modèles conjonctifs est en général plus faible que pour les modèles disjonctifs, et que l'accroissement du taux d'homogamie après le degré 10, pratiquement nul dans les modèles disjonctifs, est encore sensible dans les modèles conjonctifs.

Il peut être intéressant de comparer chaque modèle conjonctif au modèle disjonctif qui lui correspond. Tels que nous les avons défi-

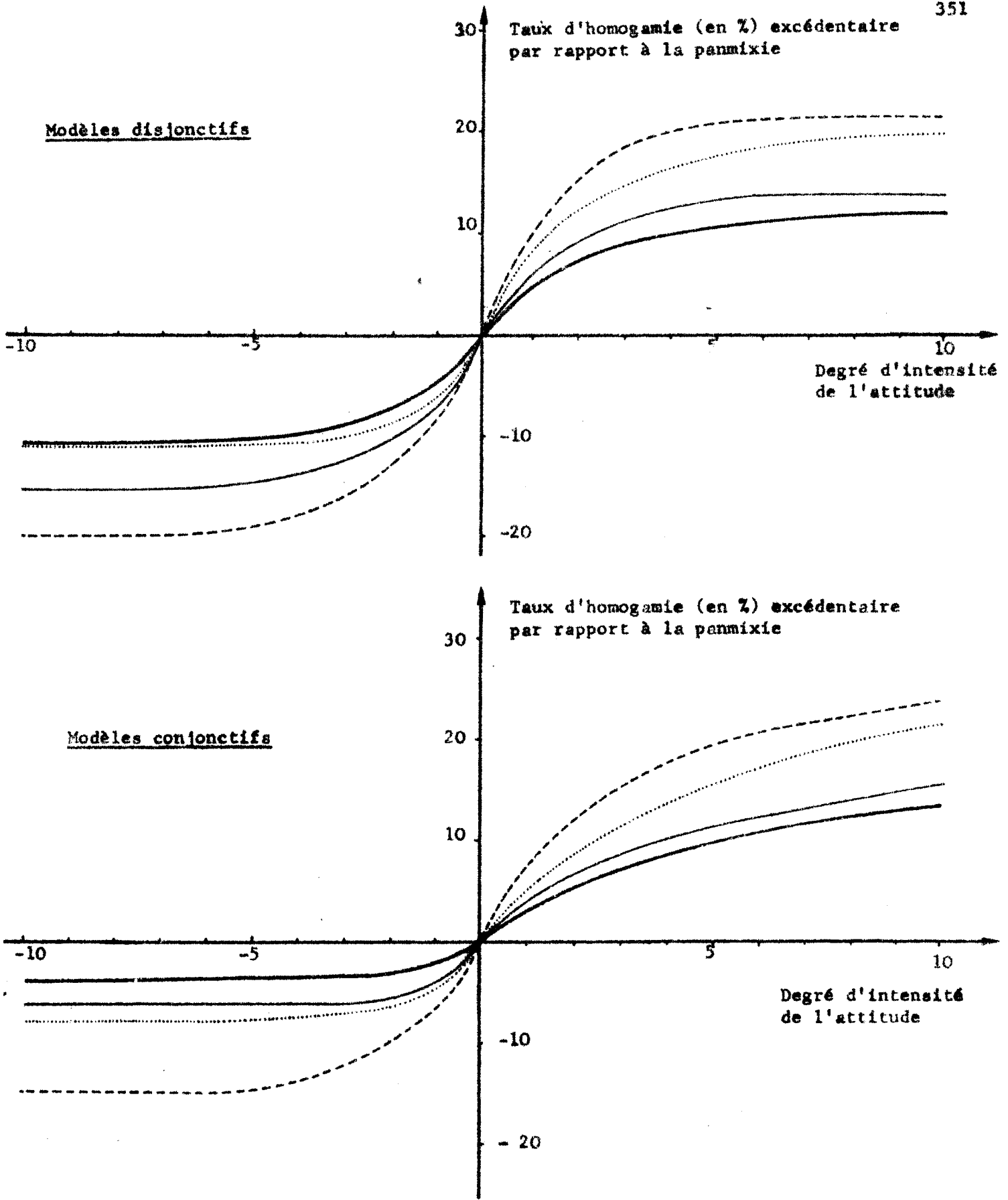


Figure 22 - Allure générale des courbes correspondant aux modèles conjonctifs et aux modèles disjonctifs.

nis au chapitre 4, les modèles *conjonctifs* imposent des conditions très sévères à l'appariement des conjoints, puisque l'union n'est en principe prononcée que lorsque les deux critères sont satisfaits simultanément. Les modèles *disjonctifs* par contre fondent l'appariement sur des conditions plus souples, puisqu'il suffit en principe, pour qu'il y ait union, qu'un seul des critères soit satisfait. Dans ces conditions, il paraît logique de penser que les couples constitués selon un modèle conjonctif sont plus homogènes que ceux formés selon un modèle disjonctif, et par conséquent que les mariages conjonctifs aboutissent à un taux d'homogamie plus élevé que les mariages disjonctifs. On pourrait en déduire que le taux d'homogamie doit toujours croître plus vite dans le modèle conjonctif que dans le modèle disjonctif correspondant. Dans l'exemple de l'influence sur les variations du taux d'homogamie selon la religion des deux types de modèles de décision, représenté sur la figure 21 (page 349), on voit immédiatement que, pour les degrés positifs de l'attitude (endogamie), le taux d'homogamie correspondant au modèle conjonctif est toujours inférieur au taux correspondant au modèle disjonctif. Ce qui suffit à infirmer l'hypothèse énoncée.

Deux questions se posent alors :

- d'où vient l'erreur de l'hypothèse précédente ?
- l'asymptote supérieure du taux d'homogamie correspondant à un modèle conjonctif est-elle nécessairement inférieure à l'asymptote du modèle disjonctif correspondant ?

L'erreur qui sous-tend cette hypothèse nous paraît en réalité être triple :

- passage à la limite sans tenir compte du fait que le nombre

maximum de choix possibles est fixé dans chaque simulation ;

- hypothèse implicite que l'univers de choix est illimité ;
- confusion entre taux d'homogamie et proportion de couples satisfaisant aux critères de choix.

Le premier point est un oubli du mécanisme d'appariement des individus : pour un degré d'intensité donné de l'attitude endogamique (c'est-à-dire pour un nombre maximum donné de choix possibles), une partie seulement de l'ensemble des couples possibles satisfaisant à la fois aux deux critères est effectivement constituée ; le complément étant formé de couples unis selon les seules lois du hasard ("mariages forcés"). Si l'on fait provisoirement l'hypothèse que le nombre maximum de choix possibles est infini (passage à la limite), ce qui signifie que tout individu peut suspendre sa décision tant qu'il n'a pas rencontré le conjoint qui lui agréé, on s'aperçoit qu'en l'absence d'hypothèses particulières (et par conséquent restrictives), le processus d'appariement se trouve bloqué si l'on ne suppose pas également que l'univers de choix est illimité. Dans notre modèle, l'univers de choix est limité, comme d'ailleurs dans la réalité ; de plus, cet univers de choix est commun à tous les individus de même sexe appartenant au même isolat. Il est donc pratiquement nécessaire, même si l'on passe à la limite, qu'il y ait des "mariages forcés", c'est-à-dire des appariements d'individus ne satisfaisant pas aux critères (*). Nous allons enfin montrer que le taux d'homogamie ne se confond pas avec le taux

(*)

Cette nécessité pratique n'est pas toujours fondée en théorie : par exemple, si les deux critères sont liés entre eux, et si les distributions bivariées correspondant aux deux critères sont rigoureusement identiques chez les hommes et chez les femmes de chaque isolat, il est possible qu'à la limite il n'y ait aucun mariage forcé. En pratique, ces conditions sont suffisamment improbables pour pouvoir être négligées.

d'individus ayant satisfait aux critères de choix. Considérons deux variables dichotomiques, $A = \{a, \bar{a}\}$ et $B = \{b, \bar{b}\}$, servant à caractériser les hommes et les femmes d'un isolat. Les couples constitués à l'issue du processus d'appariement sont caractérisés par quatre valeurs : les valeurs de A et de B relatives au mari, et les valeurs de A et de B relatives à l'épouse. Il y a donc $(2^2)^2 = 16$ types de couples possibles ; les effectifs observés pour chaque type de couple peuvent être rangés dans une matrice carrée du type de celles figurées dans les tableaux suivants. Dans les tableaux n° 17a à 17d ci-dessous.

nous avons hachuré les cases dont les effectifs doivent être additionnés pour calculer la proportion de couples satisfaisant au modèle conjonctif $A \wedge B$, la proportion de couples satisfaisant au modèle disjonctif $A \vee B$, le taux d'homogamie selon A, et le taux d'homogamie selon B. Un simple coup d'oeil permet de voir que le taux d'homogamie ne saurait être confondu avec la proportion des couples satisfaisant à un modèle, même si ces deux types de paramètres ne sont pas totalement indépendants l'un de l'autre.

Caractéristiques de la femme	Caractéristiques du mari			
	a b	a \bar{b}	\bar{a} b	\bar{a} \bar{b}
a b				
a \bar{b}				
\bar{a} b				
\bar{a} \bar{b}				

Tableau n° 17a : Couples satisfaisant au modèle conjonctif $A \wedge B$

Caractéristiques de la femme	Caractéristiques du mari			
	a b	a \bar{b}	\bar{a} b	\bar{a} \bar{b}
a b				
a \bar{b}				
\bar{a} b				
\bar{a} \bar{b}				

Tableau n° 17b : Couples satisfaisant au modèle disjonctif $A \vee B$

Caractéristiques de la femme	Caractéristiques du mari			
	a b	a \bar{b}	\bar{a} b	\bar{a} \bar{b}
a b				
a b				
a b				
a b				

Tableau n° 17c : Homogamie selon A

Caractéristiques de la femme	Caractéristiques du mari			
	a b	a \bar{b}	\bar{a} b	\bar{a} \bar{b}
a b				
a b				
a b				
a b				

Tableau n° 17d : Homogamie selon B

A la suite de cette réfutation, devons-nous considérer que les courbes de la figure 21 sont des courbes-types, caractérisant parfaitement les relations existant entre la fonction d'homogamie relative à un modèle conjonctif et celle relative au modèle disjonctif correspondant ? Les caractéristiques que nous pouvons relever sur cette figure sont les suivantes :

1. le palier supérieur du modèle disjonctif est atteint plus rapidement que celui du modèle conjonctif ;
2. le palier supérieur du modèle disjonctif a une valeur plus élevée que le palier supérieur du modèle conjonctif ;
3. le palier inférieur du modèle disjonctif prend fin plus tôt que celui du modèle conjonctif ;

4. le palier inférieur du modèle disjonctif a une valeur plus basse que le palier inférieur du modèle conjonctif ;

5. la pente de la courbe est plus forte pour le modèle disjonctif que pour le modèle conjonctif.

6. l'intervalle de variation est plus grand pour le modèle disjonctif que pour le modèle conjonctif.

On voit que la 3^e caractéristique peut être en quelque sorte déduite par symétrie de la première ; nous avons d'ailleurs vu que, dans le cas de l'exogamie, les modèles conjonctifs se comportent comme des modèles disjonctifs, et réciproquement (cf. § 4.2.2.c). D'autre part, les caractéristiques 5 et 6 peuvent être considérées comme des corollaires des caractéristiques précédentes.

L'examen des douze courbes associées (cf. figures 23 à 26, pages 361 à 364) montre que la configuration que nous venons de décrire ne se rencontre que quatre fois. Le type de configuration le plus fréquent (8 fois sur 12) a pour différence essentielle avec le type précédent le fait que, contrairement à la caractéristique n° 2, le palier supérieur du modèle disjonctif a une valeur plus basse que le palier supérieur du modèle conjonctif. Dans la majorité des cas, la courbe du modèle disjonctif se trouve couper à nouveau celle du modèle conjonctif. Il ne nous a pas paru possible de définir clairement les relations existant entre un modèle conjonctif et le modèle disjonctif qui lui est associé sans faire intervenir certaines des caractéristiques relatives aux critères eux-mêmes.

b) L'influence des caractéristiques propres aux critères utilisés.

Nous avons considéré au § 6.1.1.a deux caractéristiques : la dispersion des distributions marginales, et la disparité entre les distributions marginales selon le sexe. Au niveau des modèles de décision à deux critères, une troisième caractéristique apparaît, qui joue un rôle important dans le processus de choix : le degré de liaison entre les critères. Nous pouvons faire l'hypothèse, en nous basant sur les analyses qui précèdent, que ces caractéristiques jouent de la manière suivante :

- la dispersion des distributions marginales des deux critères pour les deux sexes détermine l'asymptote inférieure pour chaque critère ;

- la disparité entre les distributions marginales des deux critères selon le sexe (ou, si l'on préfère, le degré de liaison entre les critères et le sexe) détermine l'asymptote supérieure pour chaque critère ;

- le degré de liaison statistique entre les critères détermine l'allure générale des variations du taux d'homogamie.

Pour examiner le bien-fondé de ces hypothèses, il serait nécessaire d'étudier le mécanisme d'appariement au niveau d'un isolat. Chacune des caractéristiques prend en effet des valeurs différentes dans chaque isolat, et le paramètre dont nous disposons à leur sujet n'est que la moyenne pondérée de ces valeurs au niveau de l'ensemble de la population. Les courbes de résultats dont nous disposons sur les variations du taux d'homogamie ne sont elles aussi que des moyennes pondérées de courbes relatives à chaque isolat. C'est pourquoi il n'est pas étonnant que nous n'ayons relevé que peu de liaisons entre ces courbes moyennes, et les valeurs moyennes des paramètres étudiés. Il paraît donc utile

d'envisager, au cours d'une recherche ultérieure, l'étude du mécanisme d'appariement, selon les différents types de modèles, au niveau d'un isolat, et l'analyse du processus d'agrégation des résultats de chaque isolat au niveau de la population globale.

6.1.3. Comportement comparé des différents critères.

Nous avons jusqu'ici considéré les résultats de notre série de simulations sous l'angle des types de modèles de décision. Nous décrivons maintenant succinctement l'ensemble des résultats pour chaque variable prise comme critère. Afin de fournir des éléments pour l'interprétation des variables observées, nous avons indiqué, dans le tableau n° 18 (page 359) les principaux paramètres caractérisant chaque variable. Ces paramètres concernent :

- les variations du taux d'homogamie ;
- la dispersion (entropie) des distributions marginales pour chaque sexe, et l'écart-type des valeurs prises par cette dispersion dans chaque isolat ;
- la somme des valeurs absolues des différences entre les distributions marginales pour chaque sexe, et la moyenne des valeurs prises par ce paramètre pour chaque isolat.

a) Importance de la localité de naissance des conjoints.

On constate, sur la figure 23 (page 361), que :

- le taux d'homogamie se stabilise pour tous les modèles de décision, à partir du degré -5 (exogamie, 5 choix possibles) ;

	<i>Localité de naissance</i>	<i>Niveau d'instruction</i>	<i>Religion</i>	<i>Origine sociale</i>
Taux d'homogamie maximum	27,7 %	34,6 %	11,3 %	8,2 %
Taux d'homogamie correspondant à la panmixie	41,9 %	47,1 %	43,1 %	25,7 %
Taux d'homogamie minimum	72,3 %	79,1 %	81,1 %	69,3 %
Marge de variation du taux d'homogamie	44,6 %	44,5 %	69,8 %	61,1 %
Ecart-type des taux correspon- dant à la panmixie	0,49 %	0,62 %	1,02 %	1,02 %
Nombre d'états de la variable.	7	5	4	8
Entropie marginale absolue (hommes)	2,40	1,76	1,45	2,56
Entropie marginale absolue (femmes)	2,43	1,62	1,41	2,61
Entropie marginale relative (hommes)	0,85	0,75	0,72	0,85
Entropie marginale relative (femmes)	0,86	0,69	0,70	0,87
Ecart-type des entropies absolues par isolat (hommes)..	5,886	7,323	4,519	8,875
Ecart-type des entropies absolues par isolat (femmes)..	7,202	6,742	4,318	9,417
Différence marginale hommes/femmes	5,2	15,5	26,4	6,5
Différence moyenne par isolat.	12,83	8,96	10,04	11,75

Tableau n° 13 : Principaux paramètres caractérisant les variables.

- le taux minimum est plus rapidement atteint pour les modèles conjonctifs que pour les modèles disjonctifs ;

- pour les attitudes endogamiques, seuls le modèle à un critère et les modèles disjonctifs ont à peu près atteint leur taux maximum au degré + 10 ;

- l'élévation du taux maximum pour les modèles disjonctifs varie parallèlement au degré de liaison entre les deux critères (localité de naissance - origine sociale : $C = 0,46$ et $0,47$; - niveau d'instruction : $C = 0,32$ et $0,31$; - religion : $C = 0,27$ et $0,20$) ;

- pour les attitudes endogamiques, les modèles conjonctifs ont des taux de croissance à peu près parallèles (angle avec l'abscisse d'environ 10°), et relativement constants ;

- pour le degré d'intensité 10, le taux atteint par les modèles conjonctifs varie en sens inverse du degré de liaison entre les critères.

b) Niveau d'instruction des conjoints.

L'examen de la figure 24 (page 352) aboutit à un certain nombre de constatations analogues à celles relevées pour l'importance de la localité de naissance. En particulier en ce qui concerne les attitudes exogamiques, et les variations du taux correspondant au degré 10 (liaison niveau d'instruction - origine sociale : $C = 0,52$ et $0,51$; - localité de naissance : $C = 0,32$ et $0,31$; - religion : $C = 0,19$ et $0,14$). Les principales différences sont les suivantes :

- les modèles disjonctifs n'ont pas tout à fait atteint leur maximum au degré 10 ;

- le modèle conjonctif "niveau d'instruction et religion" a presque atteint son taux maximum au degré 10 ;

- les deux autres modèles conjonctifs paraissent avoir des comportements identiques.

Figure 23 - Variations du taux d'homogamie selon l'importance de la localité de naissance

Taux d'homogamie.(%) selon la localité de naissance

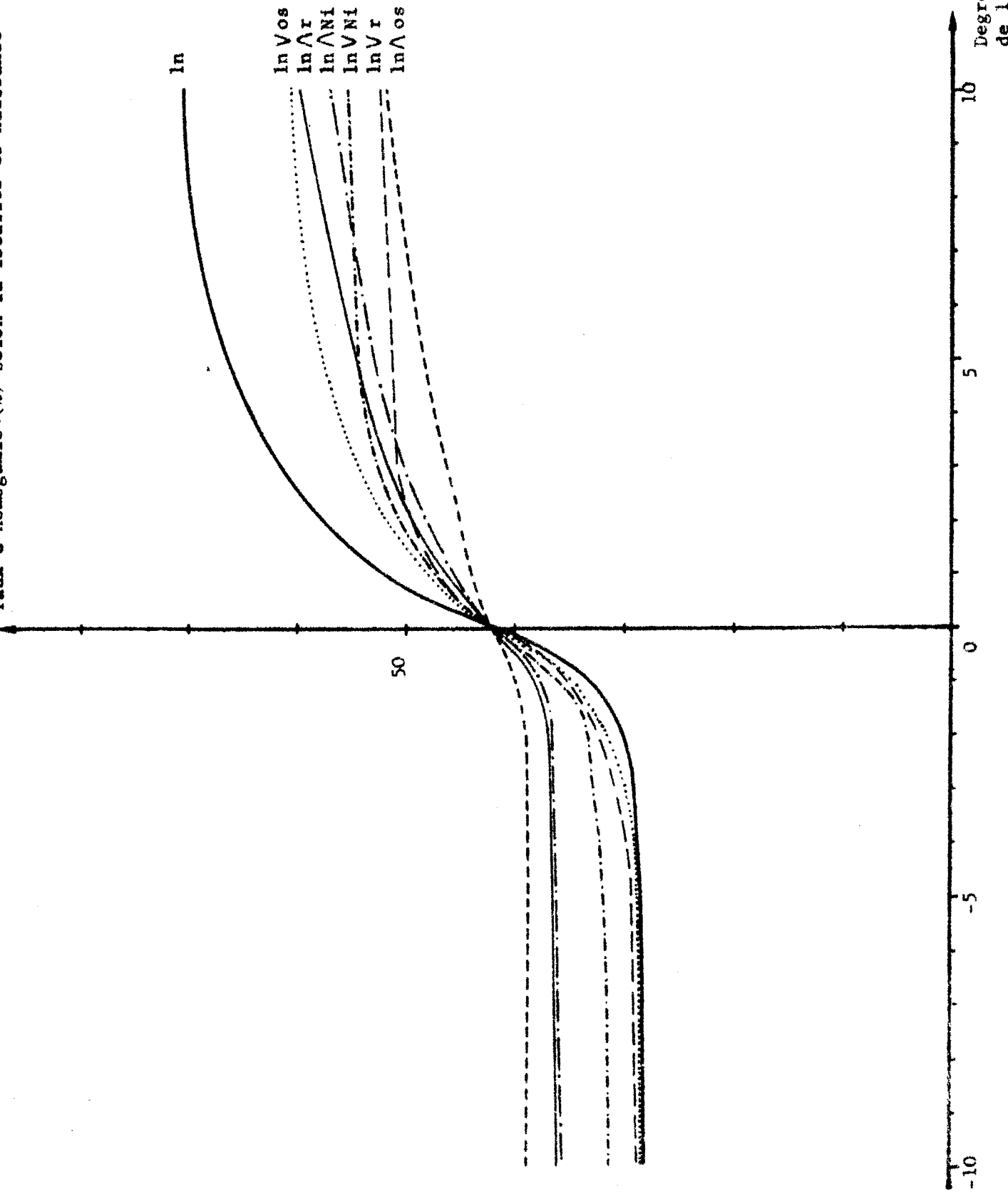


Figure 24 - Variations du taux d'homogamie selon le niveau d'instruction

Taux d'homogamie (%)
selon le niveau d'instruction

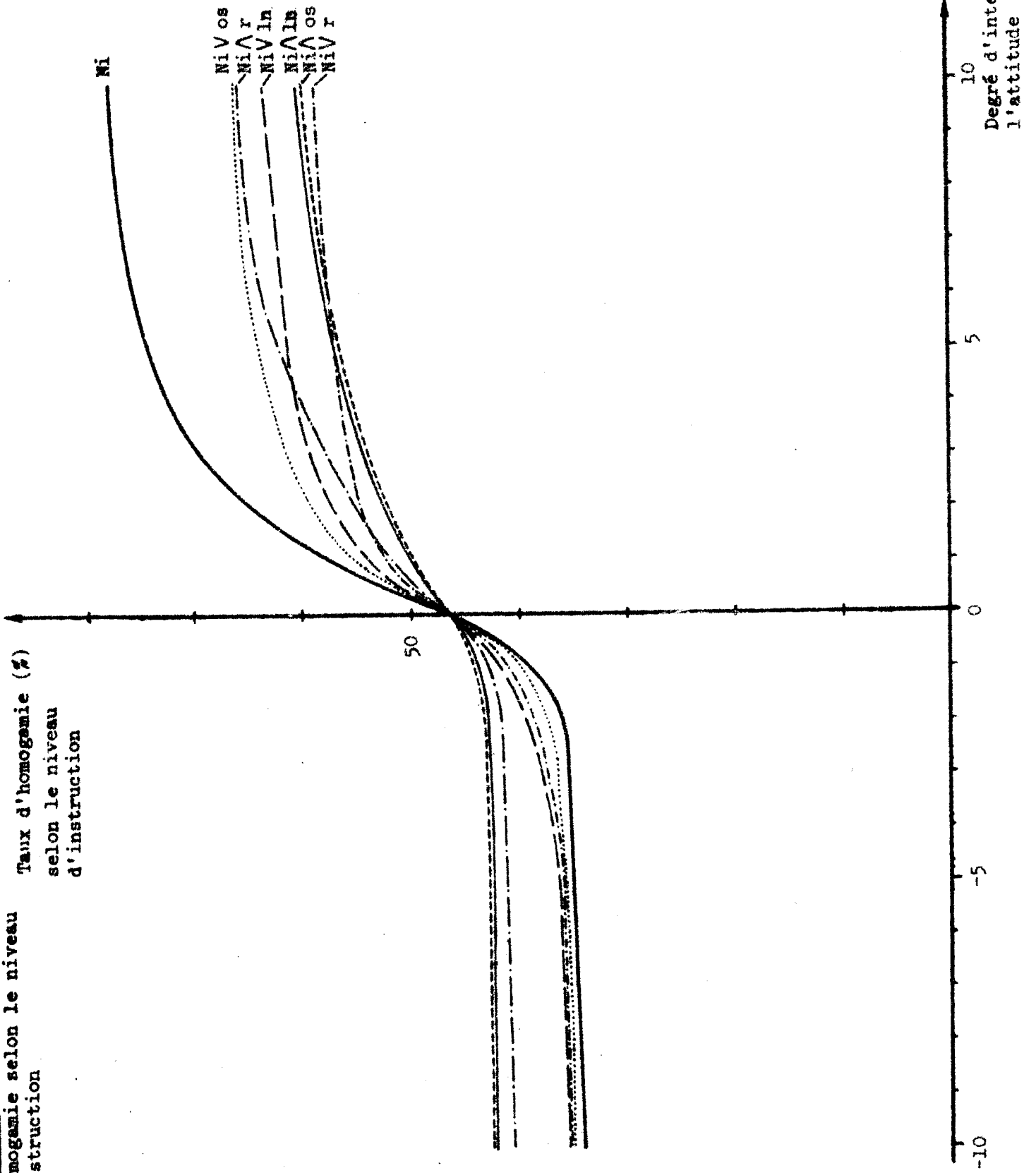


Figure 25 - Variations du taux d'homogamie selon la religion

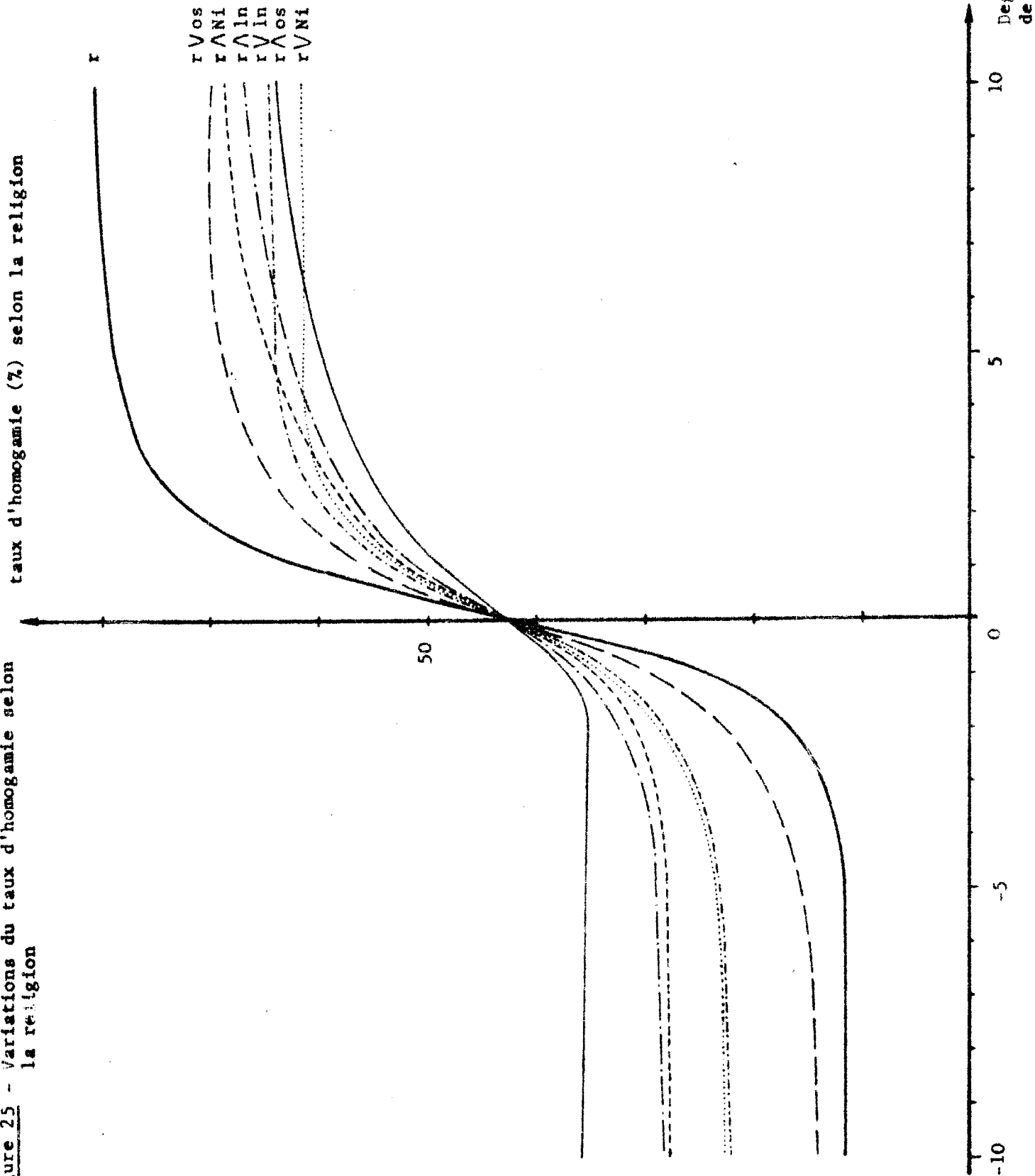
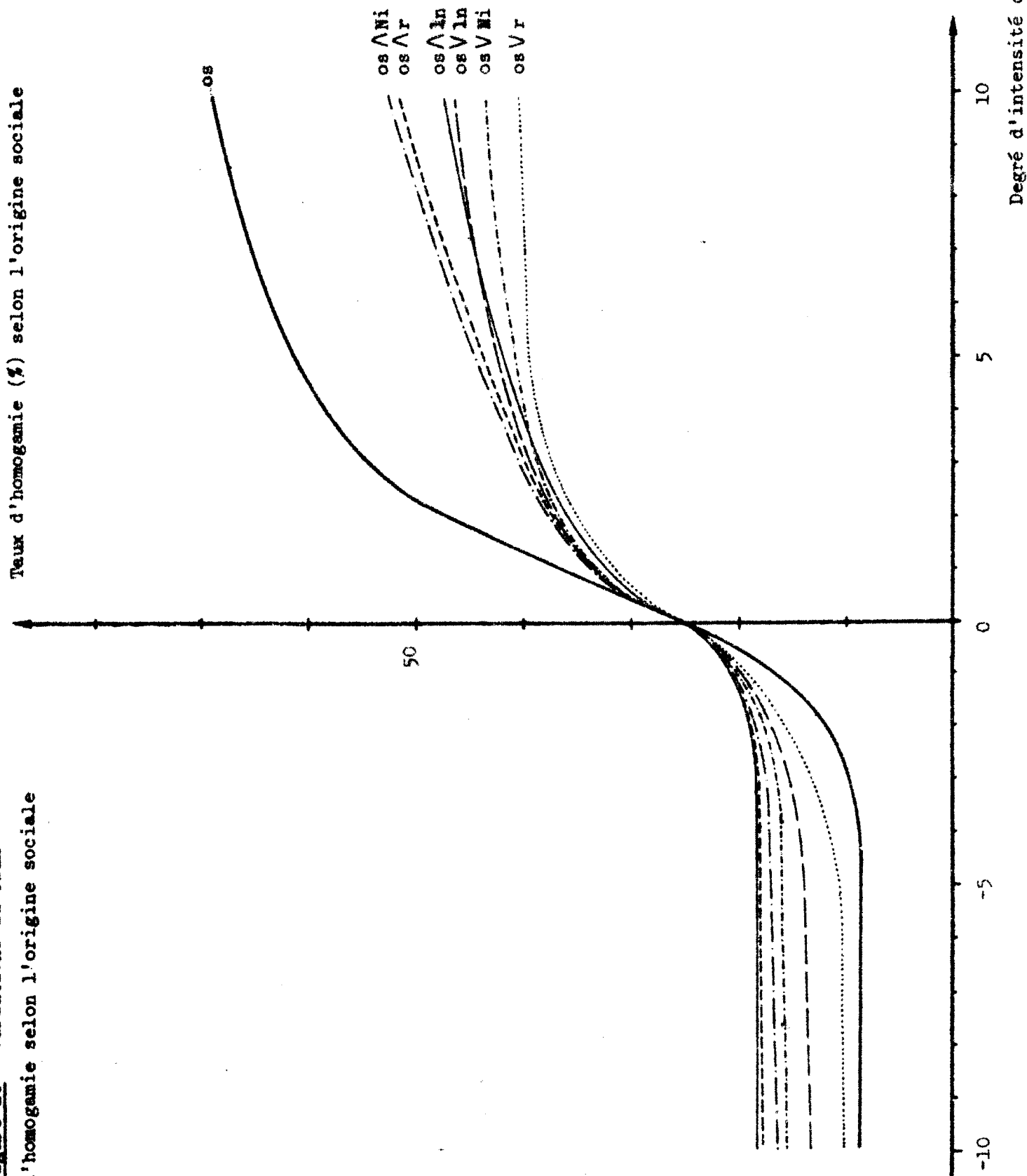


Figure 26 - Variations du taux d'homogamie selon l'origine sociale



c) Religion des conjoints.

La figure 25 (page 363) permet de relever :

- une disparité plus grande entre les divers modèles dans l'exogamie, les grandes règles énoncées à propos de la localité de naissance restant cependant valables ;
- un renversement de la liaison observée précédemment entre le taux d'homogamie atteint au degré 10, tant pour les modèles conjonctifs que pour les modèles disjonctifs.

d) Origine sociale des conjoints.

Une caractéristique frappante de la figure 26 est l'enchevêtrement des courbes d'homogamie, surtout pour les attitudes endogamiques. Si les règles énoncées pour l'exogamie restent vérifiées, on remarque pour l'endogamie :

- que les taux atteints au degré 10 par les modèles conjonctifs sont tous supérieurs aux taux atteints par les modèles disjonctifs ;
- que ni les modèles conjonctifs, ni les modèles disjonctifs n'ont atteint leur taux maximum au degré 10.

6.2. COMPARAISON DES DONNEES OBTENUES PAR SIMULATION ET DES DONNEES REELLES.

D'un point de vue méthodologique, il n'est pas sans intérêt de confronter les données simulées aux données réelles fournies par l'échantillon étudié par Alain GIRARD. Cette confrontation permet de répondre à deux types de préoccupations :

- pourrait-on ajuster le modèle de décision de manière à reconstituer les données relatives aux taux d'homogamie fournies par l'enquête ?

- comment se présentent concrètement les populations engendrées par la simulation ?

6.2.1. Peut-on reconstituer artificiellement les données fournies par l'enquête ?

Cette première préoccupation peut sembler purement académique. En effet, compte tenu du caractère par trop schématique donné au mécanisme de choix du conjoint, l'identité entre une distribution simulée et la distribution empirique correspondante ne pourrait qu'être coïncidence numérique. Une confrontation rapide des taux d'homogamie réellement observés dans l'échantillon utilisé avec les courbes décrites dans le paragraphe précédent montre l'impossibilité d'ajuster notre processus de décision aux données réelles.

Les taux d'homogamie relevés sur notre échantillon restreint de 1602 couples diffèrent légèrement de ceux qui ont été relevés sur l'échantillon complet :

- importance de la localité de naissance	: 49,2 %
- niveau d'instruction	: 65,9 %
- religion	: 78,3 %
- origine sociale	: 43,3 %

Tous ces taux sont très nettement supérieurs aux taux correspondant à l'hypothèse de panmixie. Par conséquent, les quatre variables ont joué simultanément le rôle de critères de choix, et l'attitude à leur égard est nettement endogamique. Pour ajuster ce processus, deux problèmes se

posent : le modèle de choix est-il conjonctif ou disjonctif ? Quel est le degré d'intensité de l'attitude endogamique ?

Les deux questions sont liées, et l'examen des courbes des figures 23 à 26 apporte des éléments de réponse. On constate aisément que le taux d'homogamie réel correspond, dans le cas du modèle à un seul critère, à une attitude endogamique comprise entre 0 et 1 pour la localité de naissance, entre 1 et 2 pour l'origine sociale, entre 2 et 3 pour le niveau d'instruction, entre 4 et 5 pour la religion. Mais nous savons que les modèles à deux critères, qu'ils soient conjonctifs ou disjonctifs, ne permettent pas d'atteindre des taux d'homogamie aussi élevés pour un degré égal d'intensité. Nous sommes en droit de penser, et nous avons vérifié ce point empiriquement, que plus le nombre de critères est élevé, moins le taux d'homogamie croît rapidement. Or, s'il est encore possible d'atteindre le taux recherché à l'aide de modèles à deux critères pour la localité de naissance, le niveau d'instruction et l'origine sociale, on voit immédiatement qu'il n'en est pas de même pour la religion. Le taux de 78,3 % ne semble pouvoir être atteint par aucun des modèles à plusieurs critères.

Les principales conclusions qui se dégagent de ces constatations sont les suivantes :

- le processus de décision et le découpage en isolats sont peu conformes à la réalité ;
- cependant, il est probable que le processus de choix réel dépend de l'ensemble des facteurs considérés, et non d'un petit nombre de variables ;
- si l'on fait l'hypothèse que ces facteurs jouent de manière indépendante, on peut estimer que l'intensité de l'attitude endogamique est plus forte pour la religion que pour les autres facteurs.

6.2.2. Description de quelques échantillons engendrés par simulation.

Au cours des essais que nous avons effectués pour la mise au point du processus d'appariement des conjoints, nous avons écrit un programme ne se limitant pas au calcul des taux d'homogamie, mais permettant d'engendrer et de décrire un échantillon de population correspondant à une valeur donnée de l'attitude pour un modèle de décision donné. Ce programme nous a servi essentiellement à tester la stabilité des résultats d'une simulation à l'autre, et à étudier dans le détail le fonctionnement de notre modèle. Chaque exécution de ce programme réalise en fait simultanément dix simulations identiques, décrit les dix échantillons de 1602 couples ainsi obtenus, et calcule enfin les valeurs moyennes des caractéristiques de ces dix échantillons. En raison du coût relativement faible d'une exécution du programme (*), nous avons pu tester l'influence sur les échantillons engendrés de divers types de découpage en isolats (isolats basés sur la seule date du mariage ; isolats de taille plus importante, obtenus en regroupant les catégories d'agglomération de résidence, etc.), et de divers mécanismes de décision (choix effectués par l'homme seul, la femme seule, ou en alternance ; modèles de décision à plus de deux critères, etc.).

Nous n'examinons ici que quelques uns de ces échantillons, ou, plus précisément, les valeurs moyennes obtenues sur dix échantillons pour quelques types de modèles de décision. Nous nous sommes limité à trois modèles : panmixie, forte endogamie (10 choix possibles) selon l'origine sociale, forte endogamie (10 choix possibles) selon la localité de naissance, le niveau d'instruction, et la religion (modèle conjonctif à trois critères). Pour décrire les échantillons engendrés

(*)

L'exécution du programme dure, sur le *Control Data 3.600*, entre 1 minute 20 et 3 minutes 20, selon la complexité du processus de choix.

par ces modèles, nous utilisons les tableaux correspondant, pour chacune des quatre variables étudiées, au croisement des caractéristiques de la femme par celles du mari. Afin de fournir une base de comparaison, nous décrivons selon les mêmes procédures l'échantillon réel qui a servi de base à la simulation.

a) Description de l'échantillon utilisé.

Ainsi que nous l'avons signalé, cet échantillon diffère légèrement de celui qu'Alain GIRARD a utilisé dans son étude sur le choix du conjoint (*). Les tableaux n° 19 à 22 ci-après montrent que l'homogamie est assez nette pour les quatre critères.

Femme	Mari							Total
	Moins de 2.000 hab	2.000 à 5.000 hab	5.000 à 20.000 hab	20.000 à 100.000 hab	Plus de 100.000 hab	Paris	Etranger et T.O.M.	
Moins de 2.000 hab.	29,4	4,0	3,4	2,1	1,4	0,6	1,2	42,1
2.000 à 5.000 hab.	4,6	3,5	1,1	1,0	0,7	0,3	0,4	11,6
5.000 à 20.000 hab.	3,2	2,1	5,3	1,1	1,1	1,1	0,6	14,5
20.000 à 100.000 hab.	2,9	1,5	0,8	3,8	1,4	0,9	0,5	11,8
Plus de 100.000 hab.	2,1	0,7	1,4	0,6	4,2	0,3	0,4	9,7
Paris	0,5	0,6	1,0	0,8	0,5	1,7	0,4	5,5
Etranger et T.O.M.	1,3	0,3	0,7	0,5	0,3	0,2	1,5	4,8
<i>Total</i>	44,0	12,7	13,7	9,9	9,6	5,1	5,0	100,0

Tableau n° 19 : Description de l'échantillon : importance de la localité de naissance (Base : 1602 = 100 %).

(*) On pourra comparer le tableau n° 19 aux tableaux VI page 62, et VII page 63 du *Choix du conjoint* ; le tableau n° 20 au tableau V page 80 ; le tableau n° 21 au tableau X page 84 ; le tableau n° 22 au tableau II page 74.

Femme	Mari					Total
	Primaire	Primaire supérieur	Technique	Secondaire	Supérieur	
Primaire	51,7	4,2	5,4	2,6	0,3	64,2
Primaire supérieur	4,4	4,5	1,9	1,8	1,1	13,7
Technique	3,0	0,9	3,7	0,8	0,8	9,2
Secondaire	0,9	1,2	1,6	3,1	2,6	9,4
Supérieur	0,1	0,1	0,1	0,2	3,0	3,5
<i>Total</i>	60,1	10,9	12,7	8,5	7,8	100,0

Tableau n° 20 : Description de l'échantillon : niveau d'instruction des conjoints (Base : 1602 = 100 %).

Femme	Mari				Total
	Catholique		Autres religions	Sans religion	
	Pratiquant	non-pratiquant			
Catholique pratiquant	33,4	13,4	0,2	0,4	47,4
Catholique non-pratiquant	0,7	39,6	0,9	3,4	44,6
Autres religions	0,1	0,8	1,5	0,4	2,8
Sans religion	0,1	1,1	0,1	3,9	5,2
<i>Total</i>	34,3	54,9	2,7	8,1	100,0

Tableau n° 21 : Description de l'échantillon : pratique religieuse des deux conjoints (Base : 1602 = 100 %).

Profession du père de la femme	Profession du père du mari								Total
	Cultivateur	Ouvrier agricole	Ouvrier ou contremaître	Employé	Cadre moyen, technicien	Artisan, commerçant	Cadre supérieur, profession libérale	Autres	
Cultivateur	18,0	1,6	4,0	1,3	0,3	2,3	0,7	0,5	28,7
Ouvrier agricole	1,2	1,4	1,0	0,2	0,0	0,4	0,0	0,1	4,3
Ouvrier ou contremaître	4,4	0,9	12,4	3,1	0,5	3,1	0,7	0,9	26,0
Employé	1,9	0,5	4,0	3,1	0,4	1,7	0,7	0,4	12,7
Cadre moyen, technicien	0,4	0,0	0,6	0,6	0,3	0,7	0,5	0,0	3,1
Artisan, commerçant	3,1	0,2	3,2	1,4	0,6	4,8	1,1	0,4	14,8
Cadre supérieur, profession libérale	0,7	0,1	0,8	0,8	0,2	1,1	2,2	0,2	6,1
Autres	0,6	0,1	1,2	0,5	0,1	0,6	0,1	1,1	4,3
TOTAL	30,3	4,8	27,2	11,0	2,4	14,7	6,0	3,6	100,0

Tableau n° 22 : Description de l'échantillon : origine sociale des conjoints (Base : 1602 = 100 %).

b) Panmixie restreinte.

Les tableaux n° 23 à 26, comparés aux tableaux précédents (n° 19 à 22), montrent que, dans l'hypothèse de panmixie restreinte :

- les couples homogames, figurant dans la diagonale, sont moins nombreux que dans la réalité ;
- la "perte" en couples homogames est relativement plus forte pour les catégories dont l'effectif est faible dans l'échantillon réel ;
- la "perte" en couples homogames est ressentie plus fortement pour la variable *religion* (76 % et 36 % pour les catégories à gros effectifs, 80 % environ pour les autres) que pour les autres variables ;
- les effectifs en dehors de la diagonale sont en général légèrement supérieurs aux effectifs réels ; il y a cependant quelques exceptions à cette règle, en particulier pour l'importance de la localité de naissance et l'origine sociale.

Si notre modèle était capable de refléter ce qui se passerait réellement si les mariages se faisaient au hasard, nous pourrions en conclure que, dans la réalité observée dans l'étude d'Alain GIRARD :

- le comportement endogamique est sensible pour les quatre variables considérées, et particulièrement marqué pour la religion.
- il est plus marqué dans les catégories dont les effectifs marginaux sont faibles : par exemple, hommes et femmes nés dans la région parisienne ; ayant une instruction primaire supérieure, secondaire, ou supérieure ; sans religion ; fils ou filles d'ouvrier agricoles. On peut expliquer cette influence de l'effectif des catégories en faisant l'hypothèse que les petites catégories sont plus homogènes

que les grandes. Ces dernières étant des regroupements de catégories plus petites, dans lesquelles on observerait un comportement endogamique plus marqué, mais atténué ensuite par le regroupement.

- dans l'influence, sur la formation des couples, de l'importance de la localité de naissance, on relève quelques "irrégularités". Alors que le taux moyen d'unions hétérogames observées dans la réalité est sensiblement inférieur à ce même taux dans l'hypothèse de panmixie restreinte, c'est le phénomène contraire qui se produit pour le taux de femmes nées dans une petite ville (2.000 à 5.000 hab.) épousant un homme né dans une commune rurale ; de même pour le taux de femmes nées dans une ville moyenne (5.000 à 20.000 hab.) épousant un homme né dans

Femme	Mari							Total
	Moins de 2.000 hab	2.000 à 5.000 hab	5.000 à 20.000 hab	20.000 à 100.000 hab	Plus de 100.000 hab	Paris	Etranger et T.O.M.	
Moins de 2.000 hab.	26,5	4,4	3,3	2,7	2,2	1,1	1,9	42,1
2.000 à 5.000 hab.	4,2	3,2	1,4	1,2	0,7	0,4	0,5	11,6
5.000 à 20.000 hab.	4,0	1,8	4,5	1,3	1,3	0,9	0,7	14,5
20.000 à 100.000 hab.	3,4	1,4	1,2	2,9	1,3	0,9	0,7	11,8
Plus de 100.000 hab.	2,6	0,8	1,6	0,7	3,1	0,3	0,6	9,7
Paris	1,4	0,6	0,9	0,6	0,5	1,2	0,3	5,5
Etranger et T.O.M.	1,9	0,5	0,8	0,5	0,5	0,3	0,3	4,8
TOTAL	44,0	12,7	13,7	9,9	9,6	5,1	5,0	100,0

Tableau n° 23 : Panmixie restreinte : importance de la localité de naissance (Base : 1602 = 100 %).

Femme	Mari					Total
	Primaire	Primaire supérieur	Technique	Secondaire	Supérieur	
Primaire	42,2	6,8	6,5	4,7	4,0	64,2
Primaire supérieur	7,5	1,7	1,9	1,4	1,2	13,7
Technique	4,3	1,1	1,7	1,1	1,0	9,2
Secondaire	4,8	1,0	1,7	0,9	1,0	9,4
Supérieur	1,3	0,3	0,9	0,4	0,6	3,5
TOTAL	60,1	10,9	12,7	8,5	7,8	100,0

Tableau n° 24 : *Parmixie restreinte : niveau d'instruction des conjoints*
(Base 1602 = 100 %).

Femme	Mari				Total
	Catholique		Autres religions	Sans religion	
	Pratiquant	non-pratiquant			
Catholique pratiquant	18,0	25,1	1,1	3,2	47,4
Catholique non-pratiquant	14,2	25,3	1,3	3,8	44,6
Autres religions	0,8	1,5	0,2	0,3	2,8
Sans religion	1,3	3,0	0,1	0,8	5,2
TOTAL	34,3	54,9	2,7	8,1	100,0

Tableau n° 25 : *Parmixie restreinte : pratique religieuse des deux conjoints* (Base : 1602 = 100 %).

Profession du père de la femme	Profession du père du mari								Total
	Cultivateur	Ouvrier agricole	Ouvrier ou contremaître	Employé	Cadre moyen, technicien	Artisan, commerçant	Cadre supérieur, profession libérale	Autres	
Cultivateur	12,8	1,9	5,9	2,1	0,4	3,5	0,9	1,2	28,7
Ouvrier agricole	1,6	0,3	1,1	0,3	0,1	0,6	0,2	0,1	4,3
Ouvrier ou contremaître	6,3	1,2	8,2	3,1	0,7	4,0	1,6	0,9	26,0
Employé	3,1	0,4	3,6	1,7	0,4	2,0	1,1	0,4	12,7
Cadre moyen, technicien	0,5	0,1	1,0	0,4	0,1	0,6	0,3	0,1	3,1
Artisan, commerçant	3,7	0,6	4,2	1,9	0,5	2,4	1,0	0,5	14,8
Cadre supérieur, profession libérale	1,0	0,1	2,0	1,0	0,2	1,0	0,6	0,2	6,1
Autres	1,3	0,2	1,2	0,5	0,0	0,6	0,3	0,2	4,3
TOTAL	30,3	4,8	27,2	11,0	2,4	14,7	6,0	3,6	100,0

Tableau n° 26 : Pannixie restreinte : origine sociale des conjoints
(Base : 1602 = 100 %).

une petite ville. Cette "anomalie" peut être le résultat de deux facteurs conjugués : la plus grande mobilité géographique de l'homme, moins dépendant du milieu familial, et l'attraction des villes.

- de même, pour le niveau d'instruction, on relève des taux très élevés pour les femmes ayant une formation secondaire épousant un homme d'instruction supérieure ou primaire supérieure, et pour les femmes d'instruction primaire supérieure épousant un homme d'instruction secondaire. Sans doute faut-il voir dans ce phénomène un effet de la similitude du primaire supérieur et du secondaire (le primaire supérieur étant cependant plutôt l'apanage de la femme).

- pour la religion, malgré le faible taux d'unions hétérogames, on observe un taux de femmes de religions minoritaires (protestantes ou juives) épousant un athé, plus élevé que dans l'hypothèse de panmixie. On remarque également le taux relativement très faible d'unions entre un mari catholique pratiquant et une femme non pratiquante (0,7 % au lieu de 14,2 % dans l'hypothèse de panmixie) ; ce refus d'union entre pratiquants et non pratiquants est sensible, quoiqu'à un degré moindre, pour le taux de mariages entre une femme pratiquante et un mari non pratiquant (13,4 % au lieu de 25,1 % dans l'hypothèse de panmixie).

- pour l'origine sociale enfin, les disparités sont peu marquées ; nous signalerons seulement le taux élevé de filles d'employé épousant un fils d'ouvrier (4,0 % au lieu de 3,6 % dans l'hypothèse de panmixie).

En conclusion, la comparaison de la population réelle avec la population simulée fait apparaître une dissymétrie assez sensible entre le comportement de l'homme et celui de la femme dans le processus

d'appariement. En cas de disparité entre les futurs époux, l'homme semble exercer plus fréquemment un choix que la femme ; ceci se traduisant par la fréquence plus élevée avec laquelle l'homme épouse une femme de "condition supérieure" (agglomération de naissance plus importante, pratique religieuse, statut social du père plus élevé).

c) Forte endogamie selon l'origine sociale.

Les tableaux n° 27 à 29 (pages 378 et 379) concernent les trois variables non retenues comme critère de choix. Si nous les comparons aux tableaux correspondant dans l'hypothèse de panmixie restreinte (n° 23 à 25), nous ne constatons que de très légères différences (presque toutes inférieures à 5 %) ; ainsi, le phénomène global que nous avons constaté à propos du taux d'homogamie, c'est-à-dire l'absence d'influence importante du modèle de décision sur les variables non retenues comme critère, se trouve vérifié pour chaque catégorie de couples.

Le tableau n° 30 (page 380) concerne le critère de choix. Si nous le comparons au tableau correspondant dans l'hypothèse de panmixie restreinte (n° 26 page 375), nous voyons que :

- les taux de couples homogames (diagonale) sont plus élevés et les taux de couples hétérogames plus faibles ;

- les catégories pour lesquelles le comportement endogamique est le plus marqué sont celles dont les effectifs marginaux sont les plus faibles (fils ou filles d'ouvriers agricoles, de cadres moyens), et réciproquement celles pour lesquelles le comportement endogamique est le moins fort sont celles dont les effectifs marginaux sont les plus

élevés (fils ou filles de cultivateurs, d'ouvriers) :

- les catégories de couples hétérogames pour lesquelles la proportion d'unions dans l'hypothèse de forte endogamie est particulièrement faible sont les mariages de fille de commerçant ou artisan avec fils de cultivateur (18 fois moins que dans l'hypothèse de panmixie restreinte), de fille d'ouvrier avec fils de cultivateur et de filles de cultivateur avec fils d'ouvriers (4 fois moins).

Femme	Mari							Total
	Moins de 2.000 hab	2.000 à 5.000 hab	5.000 à 20.000 hab	20.000 à 100.000 hab	Plus de 100.000 hab	Paris	Etranger et T.O.M.	
Moins de 2.000 hab	27,3	4,0	3,2	2,5	2,1	1,1	1,9	42,1
2.000 à 5.000 hab.	4,2	3,2	1,3	1,3	0,7	0,3	0,6	11,6
5.000 à 20.000 hab.	4,1	1,8	4,4	1,4	1,3	0,9	0,6	14,5
20.000 à 100.000 hab.	3,1	1,4	1,4	2,9	1,3	1,0	0,7	11,8
Plus de 100.000 hab.	2,5	0,9	1,7	0,6	3,2	0,3	0,7	9,7
Paris	1,1	0,7	0,9	0,8	0,5	1,3	0,2	5,5
Etranger et T.O.M.	1,7	0,7	0,9	0,4	0,5	0,2	0,4	4,8
TOTAL	44,0	12,7	13,7	9,9	9,6	5,1	5,0	100,0

Tableau n° 27 : Endogamie selon l'origine sociale (10 choix possibles) : importance de la localité de naissance (Base : 1602 = 100 %).

Femme	Mari					Total
	Primaire	Primaire supérieur	Technique	Secondaire	Supérieur	
Primaire	44,0	6,7	6,1	4,7	2,7	64,2
Primaire supérieur	7,1	1,7	2,3	1,2	1,4	13,7
Technique	4,0	1,1	1,9	0,9	1,3	9,2
Secondaire	3,9	1,1	1,7	1,2	1,5	9,4
Supérieur	1,1	0,3	0,7	0,5	0,9	3,5
TOTAL	60,1	10,9	12,7	8,5	7,8	100,0

Tableau n° 28 : Endogamie selon l'origine sociale (10 choix possibles) : Niveau d'instruction des conjoints (Base : 1602 = 100 %).

Femme	Mari				Total
	Catholique		Autres religions	Sans religion	
	Pratiquant	non-pratiquant			
Catholique pratiquant	18,6	24,9	1,2	2,7	47,4
Catholique non-pratiquant	13,6	25,5	1,3	4,2	44,6
Autres religions	0,9	1,5	0,1	0,3	2,8
Sans religion	1,2	3,0	0,1	0,9	5,2
TOTAL	34,3	54,9	2,7	8,1	100,0

Tableau n° 29 : Endogamie selon l'origine sociale (10 choix possibles) : pratique religieuse des deux conjoints (Base : 1602 = 100 %).

Profession du père de la femme	Profession du père du mari								Total
	Cultivateur	Ouvrier agricole	Ouvrier ou contremaître	Employé	Cadre moyen, technicien	Artisan, commerçant	Cadre supérieur, profession libérale	Autres	
Cultivateur	23,3	0,8	1,5	0,7	0,3	1,0	0,5	0,6	28,7
Ouvrier agricole	0,8	2,1	0,5	0,2	0,1	0,4	0,1	0,1	4,3
Ouvrier ou contremaître	1,5	0,6	20,3	0,9	0,4	1,2	0,6	0,5	26,0
Employé	1,3	0,6	1,4	7,2	0,3	1,0	0,6	0,3	12,7
Cadre moyen, technicien	0,5	0,1	0,6	0,4	0,8	0,3	0,3	0,1	3,1
Artisan, commerçant	0,2	0,3	1,3	0,8	0,2	10,1	0,6	0,3	14,8
Cadre supérieur, profession libérale	0,8	0,1	0,8	0,5	0,2	0,5	3,1	0,1	6,1
Autres	0,9	0,2	0,8	0,3	0,1	0,2	0,2	0,6	4,3
TOTAL	30,3	4,8	27,2	11,0	2,4	14,7	6,0	3,6	100,0

Tableau n° 30 : Endogamie selon l'origine sociale (10 choix possibles) :
 Origine sociale des conjoints (Base : 1602 = 100 %).

- les catégories de couples hétérogames qui sont peu affectées par le comportement endogamique ont toutes un effectif inférieur ou égal à 0,5 % du total dans l'hypothèse de panmixie restreinte.

Si nous comparons le tableau n° 30 au tableau correspondant pour l'échantillon observé (n° 22, page 371), nous constatons que, dans l'ensemble, la réalité se situe entre la panmixie et la forte endogamie selon l'origine sociale.

d) Forte endogamie selon la localité de naissance, le niveau d'instruction, et la religion.

Cette simulation constitue en quelque sorte le complément de la simulation précédente, puisque les trois variables qui ne prenaient aucune part dans la décision servent maintenant de critères de choix dans un modèle conjonctif à trois critères. La comparaison de la population engendrée par cette simulation avec celle correspondant à l'hypothèse de panmixie restreinte montre que :

- la description des couples selon l'origine sociale des conjoints (tableau n° 34, page 384) est très proche de la description faite dans l'hypothèse de panmixie (tableau n° 26, page 375).

- pour les variables prises comme critères (tableaux n° 31 à 33, pages 382 et 383), les taux de couples homogames sont plus élevés ; l'écart varie comme précédemment selon l'effectif des catégories dans l'hypothèse de panmixie (il est d'autant plus important que cet effectif est plus faible), et selon le critère (il est en moyenne plus grand pour le niveau d'instruction, plus faible pour la localité de naissance) ;

- les taux de couples hétérogames sont toujours plus faibles que dans l'hypothèse de panmixie, sauf dans un seul cas : femmes d'instruction primaire supérieure épousant un homme de formation technique (2,3 %, au lieu de 1,9 % pour la panmixie). Ce phénomène est probablement dû à la nature des relations existant entre les critères.

Femme	Mari							Total
	Moins de 2.000 hab	2.000 à 5.000 hab	5.000 à 20.000 hab	20.000 à 100.000 hab	Plus de 100.000 hab	Paris	Etranger et T.O.M.	
Moins de 2.000 hab.	29,8	3,4	2,7	2,2	1,7	0,8	1,5	42,1
2.000 à 5.000 hab.	3,2	4,5	1,3	1,0	0,6	0,4	0,6	11,6
5.000 à 20.000 hab.	3,3	1,5	5,7	1,2	1,3	0,9	0,6	14,5
20.000 à 100.000 hab.	2,9	1,3	1,1	3,8	1,2	0,8	0,7	11,8
Plus de 100.000 hab.	2,0	0,8	1,4	0,6	3,9	0,3	0,7	9,7
Paris	1,1	0,6	0,8	0,7	0,5	1,6	0,2	5,5
Etranger et T.O.M.	1,7	0,6	0,7	0,4	0,4	0,3	0,7	4,8
TOTAL	44,0	12,7	13,7	9,9	9,6	5,1	5,0	100,0

Tableau n° 31 : Endogamie selon la localité de naissance, le niveau d'instruction et la religion (10 choix possibles) : importance de la localité de naissance (Base ; 1602 = 100 %).

Femme	Mari					Total
	Primaire	Primaire supérieur	Technique	Secondaire	Supérieur	
Primaire	46,1	5,4	5,4	4,1	3,2	64,2
Primaire supérieur	5,5	3,2	2,3	1,4	1,3	13,7
Technique	3,5	1,0	2,7	0,9	1,1	9,2
Secondaire	3,8	1,0	1,6	1,7	1,3	9,4
Supérieur	1,2	0,3	0,7	0,4	0,9	3,5
TOTAL	60,1	10,9	12,7	8,5	7,8	100,0

Tableau n° 32 : Endogamie selon la localité de naissance, le niveau d'instruction, et la religion (10 choix possibles) : Niveau d'instruction des conjoints (Base : 1602 = 100 %).

Femme	Mari				Total
	Catholique		Autres religions	Sans religion	
	Pratiquant	non pratiquant			
Catholique pratiquant	24,6	18,8	1,1	2,9	47,4
Catholique non-pratiquant	7,8	32,3	1,2	3,5	44,6
Autres religions	0,7	1,4	0,4	0,3	2,8
Sans religion	1,2	2,4	0,2	1,4	5,2
TOTAL	34,3	54,9	2,7	8,1	100,0

Tableau n° 33 : Endogamie selon la localité de naissance, le niveau d'instruction, et la religion (10 choix possibles) : pratique religieuse des deux conjoints (Base : 1602 = 100 %).

Profession du père de la femme	Profession du père du mari								Total
	Cultivateur	Ouvrier agricole	Ouvrier ou contremaître	Employé	Cadre moyen, technicien	Artisan, commerçant	Cadre supérieur, profession libérale	Autres	
Cultivateur	13,3	2,0	5,6	1,9	0,3	3,7	0,9	1,0	28,7
Ouvrier agricole	1,7	0,3	1,1	0,4	0,1	0,4	0,1	0,2	4,3
Ouvrier ou contremaître	6,1	1,1	8,6	3,2	0,7	3,9	1,5	0,9	26,0
Employé	2,8	0,4	4,0	1,8	0,4	1,9	1,1	0,3	12,7
Cadre moyen, technicien	0,5	0,1	0,9	0,5	0,1	0,5	0,4	0,1	3,1
Artisan, commerçant	3,8	0,6	3,9	1,8	0,4	2,6	1,1	0,6	14,8
Cadre supérieur, profession libérale	0,9	0,1	1,9	0,9	0,3	1,1	0,7	0,2	6,1
Autres	1,2	0,2	1,2	0,5	0,1	0,6	0,2	0,3	4,3
TOTAL	30,3	4,8	27,2	11,0	2,4	14,7	6,0	3,6	100,0

Tableau n° 34 : Endogamie selon la localité de naissance, le niveau d'instruction et la religion (10 choix possibles) : origine sociale des conjoints (Base : 1602 = 100 %).

Si nous comparons la population simulée et l'échantillon réel pour les trois critères, nous voyons que :

- pour la localité de naissance (tableaux n° 19, page 369 et 31, page 382), les différences ne sont pas très grandes pour la plupart des catégories.

- pour le niveau d'instruction (tableaux n° 20, page 370 et 32, page 383), les différences sont plus marquées : les couples homogames sont plus nombreux dans la réalité, et les couples hétérogames moins nombreux. Seules exceptions à cette dernière règle : les femmes d'instruction primaire supérieure dont le mari a une instruction secondaire (1,8 % dans la réalité, contre 1,4 % dans cette simulation), et les femmes d'instruction secondaire dont le mari a une instruction supérieure (2,6 % dans la réalité, contre 1,3 %).

- pour la religion (tableaux n° 21, page 369, et 33, page 382), les différences sont encore plus marquées, et l'homogamie beaucoup plus forte dans la réalité.

6.3. BILAN DE L'EXPERIENCE.

Ce bilan comporte une évaluation des apports de cette série de simulations, et un examen critique de ses insuffisances et de ces limites.

6.3.1. Evaluation des apports de la série de simulations.

Les résultats dont nous venons de décrire l'essentiel permettent d'énoncer quelques propositions empiriques simples sur le processus d'appariement dans une population réelle ; mais leur intérêt principal

nous paraît tenir à ce qu'ils soulèvent un certain nombre de problèmes théoriques qui auraient pu sinon demeurer inaperçus.

a) Rappel des constats empiriques.

Les principales "lois" empiriques que nous avons relevées sont les suivantes :

- dans un processus endogamique, il n'y a pas d'homogamie pour les variables ne jouant aucun rôle dans le mécanisme d'appariement, même si celles-ci ont une liaison statistique avec le ou les critères de choix. Cela signifie que le processus de choix du conjoint est un processus *global*, tenant compte de toutes les dimensions de la personnalité du partenaire, et non pas seulement de quelques caractéristiques importantes.

- quelle que soit la sévérité du mécanisme de choix, il n'est pas possible d'obtenir un taux d'homogamie de 100 % avec un comportement endogamique absolu si l'on désire appairier l'ensemble des individus. Il existe un taux maximum possible d'homogamie, inférieur à 100 %, dont la valeur dépend des caractéristiques de la distribution des variables dans la population. Ce taux dépend également du type de mécanisme de choix (en particulier : nombre de critères, liaisons statistiques entre ceux-ci, rôles respectifs dans le processus d'appariement).

- lorsque le mécanisme de choix concerne plusieurs variables simultanément, les taux d'homogamie relevés pour ces critères sont toujours inférieurs aux taux correspondant à un mécanisme de choix à un seul critère. Pour obtenir, dans une simulation donnée, un taux d'homogamie fixé *a priori*, il faudrait permettre aux candidats au mariage de passer en revue un nombre d'autant plus grand de partenaires que le nombre de critères de choix est élevé. Compte-tenu de l'existence d'un

taux d'homogamie maximum variable selon le nombre de critères, il peut être nécessaire dans certains cas d'élargir l'univers de choix du conjoint, afin de diminuer la proportion de mariages "forcés", aboutissant nécessairement à des couples exogames. Si l'on considère les taux d'homogamie observés dans la réalité, on doit admettre ou bien que les univers de choix (partenaires "possibles" avec lesquels on se trouve en contact) sont plus grands que nos isolats, ou bien, ce qui semble plus proche de la réalité, qu'ils sont "ouverts" (c'est-à-dire différents pour chaque individu).

- dans le mécanisme de choix du conjoint, tous les critères de choix n'ont pas la même importance.

b) Problèmes théoriques soulevés.

Outre la vérification et la justification théorique des propositions empiriques que nous venons de résumer, certains problèmes mériteraient à eux seuls une étude particulière. Citons :

- l'influence sur le taux d'homogamie des contraintes imposées par les distributions marginales (*). Parmi les distributions marginales, il faut ranger non seulement la distribution des caractéristiques selon le sexe, mais aussi selon les isolats.

- l'influence sur le taux d'homogamie des disparités existant entre la distribution des hommes et celle des femmes pour une caractéristique donnée (c'est-à-dire du degré de liaison existant entre cette caractéristique et le sexe).

(*) Signalons une approche très intéressante du rôle des contraintes marginales dans l'analyse des tables de mobilité sociale, par D. BERTAUX (26). Certaines des remarques faites dans cet article sont applicables aux tableaux mesurant l'homogamie des couples.

- l'influence sur le taux d'homogamie des liaisons statistiques existant entre les critères, ou entre variables et critères. En particulier, que se passe-t-il lorsque la liaison entre une variable et un critère est très forte (par exemple, C de l'ordre de 0,90) ?

- l'influence comparée des schémas de décision conjonctifs, disjonctifs, ou mixtes.

Pour étudier ces problèmes de manière expérimentale, il pourrait être intéressant de réaliser plusieurs séries de simulations portant sur des données théoriques. Par exemple, un premier sous-programme engendrerait des populations artificielles dont les paramètres principaux (tendance centrale et dispersion pour chacune des variables, mesure de la liaison entre chaque couple de variables) seraient définis rigoureusement, et pourraient prendre des valeurs différentes à chaque simulation. Un second sous-programme analogue à notre sous-programme *CHOICE* réaliserait la formation des couples, et calculerait les taux d'homogamie. Soulignons simplement que de telles simulations doivent avoir surtout une valeur d'illustration et de guide pour la réflexion, mais ne sauraient en aucun cas remplacer la démonstration, qui ne peut porter que sur un modèle analytique du phénomène.

6.3.2. Examen des insuffisances et des limites de l'expérience.

Cette expérience a eu pour résultat la génération de données artificielles sur lesquelles nous avons mesuré certains paramètres. L'intérêt de ces données dépend évidemment de l'intérêt du modèle qui les a engendrées; aussi devons-nous rappeler les limites de ce modèle. Il nous faut également évaluer l'intérêt qu'une telle démarche peut présenter dans le cadre des recherches sur le choix du conjoint.

a) Limites du modèle de décision.

Nous avons déjà abordé ce problème au chapitre 4. Nous avons vu que l'élaboration d'un modèle, même très simple, se heurte à de nombreuses difficultés techniques ; ces difficultés nous ont imposé des contraintes qui font que le processus de choix du conjoint que nous avons simulé est loin de correspondre à ce que nous savons sur le mécanisme réel. En particulier :

- la division de la population en isolats est irréaliste, sauf peut être dans quelques cantons agricoles peu peuplés. Dans l'étude d'Alain GIRARD (93, p. 61), le taux de mariages recensés entre personnes résidant dans le même canton est relativement élevé pour les catégories sociales défavorisées (ouvriers agricoles : 74 % ; ouvriers : 71 %), mais sensiblement plus faible lorsque l'on s'élève dans la hiérarchie sociale (cadres supérieurs et professions libérales : 58 %). De plus, l'existence d'un taux élevé de mariages entre résidents du même canton ne signifie pas que les univers de choix soient nécessairement confondus. Aussi, notre processus d'appariement au sein d'un univers clos introduit-il un biais important par rapport au processus réel, en imposant une proportion non négligeable de "mariages forcés", se faisant au hasard parmi les partenaires n'ayant pas été choisis au cours d'un processus de choix normal.

- la définition des isolats est arbitraire ; le choix des variables et leur découpage ne s'appuie pas sur une connaissance approfondie de la stratification sociale.

- la probabilité de contact entre deux individus au sein d'un même isolat est la même pour tous les couples possibles ; nous savons que, dans la réalité, cette probabilité est d'autant élevée que les individus appartiennent à des milieux plus proches, plus semblables.

- les variables utilisées dans l'expérience ne sont pas nécessairement celles qui jouent le rôle le plus important dans le mécanisme d'appariement.

- la distinction des "valeurs" prises par ces variables n'est vraisemblablement pas pertinente par rapport au processus de choix du conjoint. Par exemple, la catégorie "cultivateurs" s'étend des petits métayers aux gros propriétaires terriens. Dans notre modèle, un fils de métayer peut fort bien choisir pour épouse une fille de gros propriétaire ; on peut s'interroger sur la fréquence de ce type de mariage dans la réalité.

- le mécanisme de choix se fait par *tout ou rien*. Il est vraisemblable que, dans la réalité, interviennent des relations d'ordre ou de proximité entre les états d'une variable, pour nuancer les jugements portés sur le partenaire. Par exemple, on peut supposer qu'un fils de cadre moyen accepterait de prendre pour épouse une fille d'employé plus volontiers qu'une fille d'ouvrier agricole ; cette hypothèse de l'influence de la "proximité" de deux catégories sociales est retenue par Alain GIRARD (93, p. 75). On peut également supposer que l'on accepte plus volontiers de s'unir avec un individu de rang social ou de niveau d'instruction supérieur ou égal au sien (relation d'ordre).

- le modèle de décision est un modèle *individuel*, qui suppose l'acquiescement d'un seul des deux partenaires. Cela n'introduit pas de biais si la décision se fait par tout ou rien, c'est-à-dire en fonction de la seule relation d'identité. Si l'on fonde la décision sur des relations d'ordre ou de proximité, il devient nécessaire d'introduire un mécanisme d'ajustement tenant compte des exigences de chaque partenaire.

- le modèle de décision ne concerne que deux individus à la fois. Il est probable que l'approche d'ORCUTT et ses collaborateurs (cf. § 4.2.2.d), assimilant le mariage à un ajustement de l'offre et de la demande au sein d'un marché, constitue un modèle plus conforme au

déroulement du processus.

- le modèle de décision est identique pour tous les individus. Si notre modèle était suffisamment général pour autoriser une grande variété de comportements individuels, cette objection aurait moins de poids. Mais étant donnée la trop grande simplicité du modèle, cette uniformité du processus d'appariement est contraire à la réalité.

- le processus de choix par essais et erreurs gagnerait à être complexifié en s'inspirant par exemple des modèles élaborés par les psychologues.

- dans le processus de décision à plusieurs critères, la même importance est accordée aux différentes variables servant de critères de choix. La confrontation des résultats de la simulation avec les observations montre que dans la réalité, ces variables ont un poids différent dans la décision.

Ces insuffisances et ces limites du modèle, nous les avons assumées au début de la recherche, en affirmant que cette expérience devait être considérée comme une *première approche exploratoire*, destinée à *poser des problèmes* plutôt qu'à apporter des solutions. Si nous devons entreprendre une nouvelle recherche sur le même thème, avec comme objectif principal non plus l'illustration de la technique de simulation, mais la compréhension du mécanisme du choix du conjoint, il nous faudrait :

- construire un ensemble de modèles de décision plus complexes, reposant sur des hypothèses plus proches des conditions de choix réelles ;

- ne plus constituer des isolats, mais nous efforcer de traduire de manière opératoire la notion d'univers de choix à partir des contraintes dues à l'environnement ;

- faire reposer l'appariement sur un ajustement réciproque des jugements des partenaires possibles, analogue au jeu de l'offre et de la demande.

b) Quel est l'intérêt d'une approche de ce type ?

Supposons que nous réalisions une nouvelle série de simulations sur les mécanismes du choix du conjoint, et que cette expérience ne présente plus les inconvénients que nous venons d'énumérer. Que sommes-nous en droit d'attendre d'une telle approche ? Nous allons voir que l'une des critiques les plus importantes que l'on peut adresser à cette approche est fondée sur l'utilisation qui est faite des résultats de sondage. Cependant, l'incidence de cette critique diffère selon que l'objectif de l'expérience est d'engendrer des données artificielles ou d'étudier la dynamique d'un processus.

Si nous voulons décrire la formation de couples dans une population, nous pouvons approcher ce processus continu en décrivant des coupes instantanées successives de la population, et en précisant à quels facteurs sont dues les différences observées entre deux coupes instantanées. A un instant t , nous recensons un certain nombre d'individus isolés, décrits par leurs caractéristiques (sexe, âge, origine sociale, localisation géographique, etc.), et un certain nombre de couples, ayant également leurs caractéristiques (caractéristiques individuelles des conjoints, âge du couple, nombre d'enfants, etc.). Les événements qui agissent sur ces entités au cours de la période qui sépare deux coupes instantanées c_t et c_{t+1} aboutissent à la création de nouveaux individus et de nouveaux couples, et à la disparition d'individus et de couples existant déjà. Parmi ces événements, nous citerons le décès d'un des conjoints, le divorce, la séparation, qui aboutissent à la disparition

d'un couple, et à l'apparition de nouveaux individus isolés : la *fréquentation* entre individus isolés, qui peut aboutir à la constitution d'un nouveau couple et la disparition des isolés, ou ne pas modifier le *statu quo* ; le *vieillissement*, qui peut modifier les caractéristiques individuelles telles que la probabilité individuelle de décès ou de mariage, la situation sociale ou géographique, le niveau d'instruction, etc.. La constitution des couples, leur durée de vie, leur fécondité, peuvent être expliquées à partir des caractéristiques des individus et des variables contextuelles.

Pour étudier la dynamique de ce processus, il faudrait disposer d'une série de coupes instantanées décrivant aussi complètement que possible des caractéristiques des couples et des individus isolés. De telles descriptions n'existent pas. Les résultats de sondage dont Alain GIRARD a disposé pour son étude constituent une seule coupe instantanée réalisée, en 1959, sur un échantillon d'un sous-ensemble de la population (couples légalement constitués). Comme le souligne Alain GIRARD, ces données présentent deux défauts majeurs qui nous interdisent de reconstituer le mécanisme d'appariement tel qu'il a pu se dérouler. Le premier est de ne comporter aucune information sur les individus isolés existant lors de la formation de couples donnés, et par conséquent sur la composition des univers de choix. Le second est d'être affecté par un biais non mesurable, dû à l'influence de la mortalité sur les couples déjà constitués ; ce biais est d'autant plus fort que les conjoints sont plus âgés. Il n'est donc pas possible, à partir de ces données, de reconstituer les coupes instantanées que l'on aurait pu réaliser entre le début de la période étudiée par l'enquête et l'enquête elle-même. Dans ces conditions, pouvons-nous envisager d'entreprendre une simulation à l'aide de données empiriques ?

Une première méthode d'étude consisterait à chercher à engendrer des données artificielles comparables aux résultats du sondage, afin de les confronter à ceux-ci. On peut par exemple (cf. § 6.2.2.) engendrer une population dans laquelle les caractéristiques individuelles sont exactement identiques à celles observées dans l'échantillon, mais dont les couples ont été constitués à partir de l'hypothèse de panmixie. Or, l'hypothèse de panmixie, comme d'ailleurs toute hypothèse concernant le processus d'appariement, implique la connaissance de l'état de la population avant la constitution des couples. L'hypothèse de panmixie ne peut donc pas être réalisée, que ce soit par calcul à l'aide d'un modèle analytique ou par simulation, à partir de données portant sur les seuls couples constitués. Cette objection est vraie *a fortiori* si les données concernant les couples formés avant la date de l'enquête sont incomplètes. Il nous faut donc renoncer à illustrer une hypothèse quelconque concernant le processus de formation des couples à partir des seules informations fournies par l'enquête.

Une seconde méthode serait de construire un modèle complet de processus de formation des couples. Cette méthode présente l'avantage de ne pas nécessiter de données empiriques, si ce n'est la valeur de quelques paramètres d'entrée mesurables au niveau de l'ensemble de la population. Mais, en l'état actuel du développement des théories relatives au choix du conjoint, il serait prématuré d'envisager la construction d'un tel modèle. Aussi, le seul modèle envisageable est-il un modèle incomplet, imposant par conséquent le recours à un volume important de données empiriques (cf. § 2.2.4.a).

Des réflexions qui précèdent, on risquerait de conclure à

l'inutilité d'une recherche du type de celle que nous avons entreprise. Ce serait adopter une attitude manichéiste qui ne nous paraît pas complètement fondée. Certes, il serait possible d'attendre, pour entreprendre une telle recherche, de disposer de toutes les informations nécessaires. Dans le cas présent, si nous renouvelons notre expérience avec un modèle meilleur mais sur les mêmes données, n'aura-t-elle aucun intérêt ? Cela signifierait également que les résultats que nous avons décrits n'ont aucune valeur. Les propositions que nous avons énoncées dans les sous chapitres 6.1. et 6.2 ont été suggérées par une série de simulations réalisées à partir d'un modèle trop pauvre et irréaliste, utilisant de surcroît des données incomplètes. Cependant, ces propositions sans valeur conclusive nous paraissent avoir une *valeur heuristique* comparable à celle que l'on attribue par exemple aux descriptions monographiques d'un village. Elles nous permettent d'énoncer sur les mécanismes du choix du conjoint diverses hypothèses que des études conçues spécialement pourraient avoir pour objectif de valider et de préciser. Une nouvelle série de simulations, utilisant un modèle de décision plus vraisemblable, aurait pour résultat de concrétiser les intuitions du sociologue et de stimuler ses recherches théoriques. De telles simulations, à défaut de pouvoir apporter une démonstration ou la validation empirique d'une hypothèse, nous paraissent constituer à la fois une aide puissante et une stimulation à la réflexion du chercheur.

C O N C L U S I O N S

CONCLUSIONS

Après avoir défini la notion de *simulation* et passé en revue les principaux types de simulation réalisés dans les sciences sociales, nous avons illustré sur un exemple simple la démarche nécessaire à la mise en oeuvre de ces techniques. Il nous faut maintenant apporter une réponse à la question posée dans l'introduction : *quel est l'intérêt des techniques de simulation pour le chercheur en sociologie ?* Les éléments de réponse dont nous disposons sont le bilan actuel de l'utilisation de ces techniques en sociologie, et la manière dont elles se situent par rapport aux autres techniques auxquelles recourent les sociologues. Ce dernier point est lié à la portée et aux limites des techniques de simulation en général. Bien que la simulation telle que nous l'avons définie puisse prendre des formes très diverses, son utilisation actuelle (et son utilisation prévisible, du moins à court terme) dépend étroitement des possibilités offertes par les ordinateurs. Nos conclusions porteront donc sur trois thèmes :

- la simulation et l'ordinateur ;
- portée et limites des techniques de simulation ;
- intérêt de ces techniques pour le chercheur en sociologie.

7.1. LA SIMULATION ET L'ORDINATEUR.

Nous avons vu que la simulation est une procédure qui a été utilisée longtemps avant l'invention de l'ordinateur ; actuellement

encore, il arrive que l'on réalise des expériences de simulation "à la main", ou à l'aide de dispositifs physiques conçus spécialement (homéostat, animaux cybernétiques, automates neuronaux par exemple). Il est cependant incontestable que l'essor des techniques de simulation est lié aux progrès dans la construction des machines à traiter l'information. Cette importance des innovations techniques dans le développement d'une science n'est d'ailleurs pas un fait nouveau : que l'on songe par exemple aux conséquences scientifiques de l'invention du microscope, ou de celle du LASER. Actuellement et dans un proche avenir, l'outil privilégié des simulations en sociologie reste l'*ordinateur numérique*. Il est par conséquent utile de nous interroger sur la valeur de cet outil, et en particulier :

- d'évaluer les limites auxquelles se heurteront les progrès de la technologie de l'ordinateur ;
- de comparer les performances de l'ordinateur analogique à celles de l'ordinateur numérique ;
- de définir le rôle joué par l'ordinateur dans les expériences de simulation.

a) Limites technologiques de l'ordinateur.

Les avantages que présente l'ordinateur numérique par rapport à l'homme sont le volume important de la mémoire à accès direct, la rapidité de traitement, et la "fiabilité". Pour les applications actuelles de l'informatique à la sociologie, ces caractéristiques paraissent avoir atteint un niveau satisfaisant. Mais cette impression nous paraît due à ce que l'"informatique sociologique" n'en est encore qu'à ses débuts, et n'a pas atteint le développement correspondant à celui de l'informatique appliquée par exemple à l'économie, à l'étude des décisions, à la linguistique. Les recherches entreprises pour accroître la capacité,

la rapidité et la fiabilité des ordinateurs numériques correspondent à un besoin fortement ressenti dans d'autres disciplines, et qui apparaîtra sans doute rapidement en sociologie. En effet, des problèmes très simples faisant appel à la combinatoire sont pratiquement insolubles directement à l'aide de l'ordinateur. Supposons par exemple que l'on désire rechercher l'itinéraire le plus court permettant de visiter cent villes données, en recourant à l'algorithme le plus simple (mais également le moins "performant") qui consiste à calculer la longueur de tous les itinéraires possibles, et à les comparer ensuite. Si l'ordinateur dont nous disposons est capable d'effectuer une addition élémentaire en 10 nanosecondes (10^8 opérations à la seconde), il lui faudra 10^{140} siècles pour sélectionner cet itinéraire (*), ce qui est irréalisable à l'échelle de l'humanité. De même, on ne peut demander à l'ordinateur numérique de lister les 10^{120} parties d'échec théoriquement possibles, ou de constituer une typologie de la population française en calculant la distance caractérisant chaque couple d'individus (12×10^{14}) dans l'espace défini par les variables sociologiques les plus importantes.

Si l'on considère les progrès impressionnants de la technologie de l'ordinateur au début de son histoire, on peut se demander si les

(*) Il y a $100!$ itinéraires possibles, soit approximativement $2,5 \times 10^{157}$. Il faut effectuer 99 additions élémentaires, soit environ $2,5 \times 10^{159}$ en tout. La réalisation de ces additions, à raison de 10^8 par seconde, demande $2,5 \times 10^{151}$ secondes ; comme il y a à peu près 3×10^9 secondes par siècle, il faut environ 10^{142} siècles pour faire le calcul. Il restera ensuite à choisir l'itinéraire minimum parmi les $100!$ calculés ; cela demanderait 10^{139} siècles à raison de 10^8 comparaisons par seconde (bien entendu, il existe de meilleurs algorithmes pour résoudre ce problème). Cet exemple nous a été suggéré par Monsieur F. LEVERY (IBM/France).

performances des ordinateurs futurs ne permettront pas de réaliser des calculs de l'importance de ceux que nous avons cités. Depuis le *MARK 1* d'AIKEN (1944), la vitesse de traitement a été multipliée par 10^6 ou 10^7 , grâce à l'utilisation des transistors, des semi-conducteurs, des micro-circuits. Est-il possible d'extrapoler ce résultat en supposant qu'il s'agit d'une fonction exponentielle, et de dire par exemple que, dans vingt-cinq ans, la vitesse de traitement aura été multipliée par 10^{14} par rapport au *MARK 1* ? Il serait hasardeux de faire des pronostics sur les progrès futurs du *hardware*. Il est par contre possible de mettre en évidence les limites auxquelles la technologie doit nécessairement se heurter. Les lois de la physique quantique ont permis à un chercheur, H.J. BREMERMANN, de démontrer qu'à l'intérieur d'un système clos, le taux de circulation de l'information ne peut dépasser 2×10^{47} bits par seconde et par gramme de masse (cité par 84, p. 537). Si l'on évalue l'âge de la terre à 3×10^{17} secondes, et sa masse à 6×10^{27} grammes, un calculateur ayant l'âge et la masse de la terre n'aurait pas pu traiter plus de 10^{93} bits ; il n'aurait par conséquent pas eu le temps de résoudre le problème d'itinéraire exposé plus haut.

Nous voyons sur cet exemple que la rapidité de traitement et la capacité de stockage de l'information sont liées et que leur progrès est limité par des contraintes physiques. En est-il de même pour la fiabilité de l'ordinateur ? Les premiers ordinateurs à lampes n'étaient pas fiables, et on a pu dire que l'*ENIAC* (1946), avec ses 18.000 tubes était en panne en permanence (*) ; sa fiabilité était sensiblement infé-

(*) Ce qui d'ailleurs ne l'empêchait pas de fournir des résultats justes. En effet, l'*ENIAC* a servi surtout à la constitution de tables numériques (tables de tir pour l'artillerie), dont les résultats n'étaient par conséquent pas urgents. Aussi les mathématiciens utilisateurs de l'*ENIAC* ont-ils misé sur le caractère aléatoire des parties de la machine en panne à un moment donné : en recommençant les calculs un nombre suffisant de fois, il était possible de choisir avec une quasi-certitude le résultat exact.

rieure à celle d'un calculateur humain. Il y a quelques années, a été répandu par contre dans l'opinion publique le mythe de l'infailibilité de la machine : l'ordinateur ne saurait se tromper. Actuellement, on accepte l'évidence selon laquelle les composants physiques qui constituent l'ordinateur peuvent s'user, présenter des impuretés dans leur masse, se détériorer, et ainsi provoquer des pannes de la machine. En outre, l'ensemble des programmes qui constituent le système de la machine atteint un degré de complexité tel qu'il ne peut être vérifié dans sa totalité : aux pannes *hardware* s'ajoutent par conséquent les imprévus du *software*. Les progrès technologiques ont permis un accroissement du taux de fiabilité de l'ordinateur ; est-il possible d'extrapoler cette tendance ? Nous ne le croyons pas. Certes, on peut encore améliorer la fiabilité des composants physiques qui constituent l'ordinateur ; sans doute peut-on également, en respectant certaines procédures logiques, améliorer la cohérence du *software*. Mais l'expérience montre que plus un système d'un type donné est complexe, moins il est fiable ; or, les constructeurs cherchent à réaliser des ordinateurs de plus en plus puissants, c'est-à-dire de plus en plus gros et de plus en plus complexes. Aussi est-il permis de penser qu'au delà d'un certain degré de complexité, les progrès technologiques relatifs aux composants ou au *software* ne compenseront plus la diminution de fiabilité due à l'accroissement de la complexité du système.

b) Ordinateur analogique et ordinateur numérique.

Les remarques qui précèdent sont centrées sur l'ordinateur numérique ; elles peuvent néanmoins également s'appliquer à l'ordinateur analogique, dont les limites technologiques sont en gros les mêmes. L'ordinateur analogique traite des grandeurs continues, l'ordi-

nateur numérique des grandeurs discrètes ; mais les fonctions qu'ils remplissent sont sensiblement identiques. Pourtant, les spécialistes de la simulation se divisent sur l'intérêt qu'il y a à recourir à l'un ou l'autre de ces types de machines. Actuellement, l'ordinateur numérique est de loin le plus utilisé ; cela tient à la fois à son "universalité" (il peut être programmé pour devenir isomorphe de tout système) et à l'importance de son parc. Mais il existe une classe de phénomènes que l'ordinateur analogique seul permet de simuler d'une manière stricte : il s'agit de ceux dans lesquels plusieurs entités agissent simultanément afin d'aboutir à un état d'équilibre. De tels phénomènes se rencontrent en physique, comme la détermination d'un champ électrique selon la disposition et la tension des charges (85, p. 12) ; ils sont fréquents dans les sciences de l'homme, où l'on étudie par exemple le comportement d'un réseau de neurones (15, Pp. 464-465), ou les interactions dans un groupe social. Pour simuler de tels phénomènes sur un ordinateur numérique, il est usuel d'étendre dans le temps ce qui est quasi-instantané, de dissocier des éléments qui interagissent, et de procéder par approximations successives. L'algorithme utilisé modifie alors sensiblement la nature du phénomène simulé.

Dans le domaine des sciences sociales, où les interactions sont nombreuses, il peut sembler paradoxal de recourir à la simulation numérique. La raison en est qu'actuellement l'ordinateur numérique est plus maniable et plus précis que l'ordinateur analogique. On peut espérer cependant que les progrès technologiques favoriseront la construction de systèmes analogiques permettant de simuler simultanément des processus divers en interaction, avec une rapidité et un degré de précision suffisants. Il deviendra possible alors de comparer les performances de ces deux types de machines, et de définir les champs

d'application spécifiques de chacun.

c) Rôle de l'ordinateur dans les expériences de simulation.

Certains auteurs ont décrit l'ordinateur comme une "machine à penser". Une machine étant une sorte d'esclave artificiel, créé par l'homme pour l'assister dans l'accomplissement de certaines tâches, est-ce à dire que l'ordinateur "pense pour l'homme", permettant à celui-ci de penser moins ? Nous allons montrer qu'au contraire l'utilisation de l'ordinateur impose au chercheur un surcroît de travail intellectuel ; mais l'intérêt de ce surcroît de travail est qu'il doit favoriser la fécondité de la réflexion.

L'ordinateur est une machine : c'est-à-dire que son comportement est entièrement déterminé, même si l'homme n'est pas toujours capable de le prévoir. Aussi la "pensée" de l'ordinateur ne saurait-elle en aucun cas être une pensée créatrice : le "raisonnement" de l'ordinateur est uniquement de type déductif. Bien plus, la manière dont l'ordinateur "raisonne" est fixée *a priori* par l'homme lui-même, qui domine ainsi totalement le fonctionnement de son automate. Le seul apport direct de l'ordinateur à l'homme est de lui fournir sous une forme assimilable des résultats de traitements plus complexes, plus rapidement et avec moins de chances d'erreur que si l'homme les recherchait lui-même. Mais il existe un apport indirect dont l'importance est beaucoup plus grande. L'ordinateur doit être programmé par le chercheur d'une manière très détaillée : cela oblige le chercheur à s'interroger sur sa propre manière de raisonner, et à formaliser celle-ci de telle sorte que la machine puisse la reproduire, ou du moins parvenir aux mêmes résultats. Cette *réflexion sur la réflexion* est souvent pour le

chercheur l'occasion d'un approfondissement de sa propre discipline.

La réalisation d'une expérience de simulation sur ordinateur impose au chercheur la délimitation précise du domaine étudié, la sélection des variables essentielles, et surtout la construction d'un ou de plusieurs modèles du phénomène qu'il désire simuler. L'ordinateur simule toujours un *modèle* (85, p. 9) ; et ce modèle ne se trouvant pas dans la nature, c'est au chercheur qu'incombe l'effort de théorisation nécessaire pour l'élaborer. Dans les simulations de processus stochastiques, on demande à l'ordinateur d'*imiter le hasard* : c'est l'occasion pour le chercheur de s'interroger sur ce que représente pour lui le hasard (cf. § 5.1.2. ; cf. également 195, p. 183), et de construire un modèle déterministe du hasard. Même les insuffisances de l'ordinateur peuvent contribuer à l'enrichissement des réflexions théoriques : elles imposent au chercheur la mise au point préalable d'un plan d'expérience, qui est un choix raisonné parmi toutes les simulations possibles ; la miniaturisation d'un modèle pléthorique, en le réduisant aux processus réellement essentiels ; la recherche d'algorithmes rendant possible une simulation *a priori* irréalisable. Aussi, l'utilisation de l'ordinateur dans les expériences de simulation nous apparaît-elle non comme une solution de facilité, mais comme l'occasion d'un approfondissement des connaissances théoriques du sociologue.

7.2. PORTEE ET LIMITES DES EXPERIENCES DE SIMULATION.

En nous fondant, pour l'essentiel, sur les expériences de simulation réalisées dans les sciences humaines, il nous semble que :

- l'intérêt majeur de la simulation est d'introduire l'approche expérimentale (la *quasi-expérimentation*) dans les situations où

celle-ci paraissait impraticable ;

- la difficulté principale de cette technique réside dans l'interprétation des résultats de l'expérience ;

- les limites et la portée d'une expérience de simulation dépendent de la validité du modèle sous-jacent.

a) L'approche expérimentale.

Il existe de nombreuses situations concrètes dans lesquelles l'approche expérimentale est impossible. Ces situations peuvent être décrites selon trois grandes catégories : celles dans lesquelles aucune action n'est possible sur les variables dont on désire étudier les effets ; celles dans lesquelles l'action sur les variables est possible, mais limitée ; celle enfin dans lesquelles l'action sur les variables n'est pas limitée, mais où l'expérimentateur ne peut pas provoquer l'apparition du phénomène qu'il étudie.

La première catégorie de situations regroupe toutes celles dans lesquelles toute expérimentation est interdite par des contraintes pratiques ou éthiques. Il n'est pas possible actuellement d'agir d'une manière importante sur le climat, les saisons, ou le relief géographique. L'influence de ces facteurs ne peut être étudiée qu'indirectement, selon une méthode comparative dont l'analyse de la variance est l'expression la plus raffinée. Mais ces contraintes pratiques asservissent le chercheur aux conditions produites naturellement, limitant ainsi son champ d'investigations. Les contraintes éthiques régissent l'expérimentation sur l'homme et, dans une mesure moindre, sur les animaux. Elles interdisent toute expérience risquant de porter atteinte à l'intégrité physique et psychologique de l'homme, que ce soit par action directe sur

l'individu ou par modification de son environnement.

La seconde catégorie est celle des situations dans lesquelles l'expérimentation n'est pas interdite, mais soumise à diverses contraintes qui limitent son intérêt. Citons par exemple les cas où :

- l'expérimentateur ne peut faire prendre aux variables dont il étudie les effets que certaines valeurs parmi l'ensemble des valeurs possibles ;

- il n'est pas possible de manipuler chaque variable indépendamment des autres, et d'isoler ainsi les effets qui lui sont propres ;

- les effets les plus importants pour l'expérimentateur sont masqués par des effets plus visibles, mais de moindre importance ;

- le phénomène est trop complexe, englobe trop de variables, pour que son étude expérimentale puisse être abordée globalement, sans aucune simplification ;

- les manipulations déclenchent des processus trop rapides ou trop lents pour être observés par l'homme ;

- les effectifs dont dispose le chercheur sont trop faibles pour conférer une stabilité suffisante aux résultats de l'expérience.

La troisième catégorie est en gros celle de la tératologie, et de toute étude de phénomènes rares ou uniques, dont l'apparition ne dépend pas de la volonté du chercheur ou lui est interdite pour des raisons éthiques, mais qu'il peut manipuler à sa guise lorsqu'ils se produisent naturellement.

Placé devant de telles situations, le chercheur a la possibilité de recourir à la simulation, pour créer artificiellement des situations analogues, mais ne présentant pas les inconvénients que nous avons cités. Ce faisant, il substitue "au système-objet qui est le sujet

réel de l'étude, un système-image qui est plus facile à manier" (85, p. 4). Dans ce *système-image*, il lui est loisible de multiplier les cas uniques, rares, coûteux à réaliser, ou dangereux ; de simplifier un phénomène trop complexe ; d'isoler un phénomène soumis aux effets parasites de l'environnement ; de freiner ou d'accélérer le déroulement d'un processus ; de rendre indépendants des facteurs dont les variations sont généralement associées dans le système-objet ; etc. (cf. 202, Pp. 49-51).

b) L'interprétation des résultats.

L'interprétation des résultats d'une expérience de simulation soulève tout d'abord le problème de leur validité et de leur comparaison aux observations. La majeure partie des simulations réalisées en sociologie fait intervenir le hasard, c'est-à-dire introduit une certaine incertitude dans leurs résultats. Cela entraîne deux conséquences :

- un résultat d'une expérience de simulation n'apporte pas l'évaluation du paramètre recherché, mais une valeur que peut prendre ce paramètre, compte-tenu de facteurs aléatoires qui en perturbent la mesure ;

- si l'expérience est répétée, le même résultat prendra une valeur différente à chaque répétition.

En pratique, pour estimer la "vraie" valeur d'un résultat, le chercheur devra évaluer les paramètres de la distribution de tous les résultats qu'une telle simulation pourrait engendrer, ou, à tous le moins, d'un échantillon suffisamment grand de ceux-ci. Pour obtenir des résultats présentant une validité suffisante, il est généralement nécessaire de répéter l'expérience, de manière à disposer d'un nombre important d'observations artificielles. Cela fait, nous ne disposons plus d'un

résultat, mais d'une distribution de résultats. Cette distribution de résultats doit être confrontée, à des fins d'analyse, à la distribution des situations réelles observées. La comparaison de deux distributions aléatoires est possible lorsqu'il s'agit d'aléas caractérisés par un petit nombre de dimensions, et que l'on dispose d'hypothèses sur la forme des distributions. Mais l'intérêt majeur des techniques de simulation, comparées par exemple aux processus d'optimisation, est de fournir des informations concernant non pas un indicateur unique, mais un grand nombre d'indicateurs. La richesse des informations est généralement liée à la complexité du modèle. Or, lorsqu'il y a n aléas dans un modèle, on est amené à faire des comparaisons de distributions à n dimensions ; il semble que les techniques statistiques dont nous disposons actuellement se prêtent peu à de telles manipulations (135, p. 36).

Si l'on suppose ce problème technique résolu, il reste celui de la *signification* des données simulées. Celles-ci présentent comme des résultats d'observations *in vivo* ou d'expériences *in vitro*, et, sous l'effet des habitudes mentales, le chercheur risque d'être entraîné à les interpréter comme il le ferait pour des données réelles. Or, l'interprétation de données artificielles, obtenues par simulation, dépend étroitement du dispositif qui a permis de les engendrer. Ce problème n'est pas nouveau. Nous avons signalé (§ 2.1.1.c), à l'aube de la méthode expérimentale, la méfiance des chercheurs à l'égard d'une intervention susceptible de modifier ou d'altérer la marche naturelle des choses de la nature. Dans cette attitude, qui donne la primauté à l'observation passive sur l'expérimentation, il n'y a pas que des préjugés sans fondement scientifique. Il est admis de nos jours que l'expérimentation, en imposant un environnement artificiel à l'objet qu'étudie le chercheur, le modifie en quelque sorte ; mais de STRATON

DE LAMPSAQUE à Claude BERNARD, une méthodologie expérimentale s'est constituée, qui a permis de maîtriser l'interprétation des résultats des expériences. Il n'en est pas de même pour la simulation. Celle-ci constitue en quelque sorte une *expérimentation au second degré* ; non seulement l'environnement est artificiel, mais l'objet de l'étude l'est également. Ce qui a sauvé la méthode expérimentale, c'est que, même placé dans un environnement artificiel soumis à la volonté de l'expérimentateur, l'objet étudié "résiste" ; il "obéit à des lois" que l'on peut supposer être, *mutatis mutandis*, "conformes à la nature des choses". Dans une simulation, l'objet est créé par le chercheur à l'image de ses propres options théoriques. La seule "résistance" que l'objet oppose aux manipulations est celle que lui impose sa cohérence interne. Pour évaluer la portée et les limites des expériences de simulation, il faut analyser la portée et les limites de la modélisation du réel.

c) La valeur du modèle de simulation.

Le modèle de simulation est une construction abstraite que l'on substitue au système réel. Ce modèle doit présenter une certaine analogie avec le phénomène qu'il représente. Mais cette analogie est-elle superficielle (*similarité phénotypique*) ou profonde (*similarité génotypique*) ? Ou, selon la terminologie de P. ACHARD (4, Pp. 332-336), s'agit-il d'un *modèle technique*, visant seulement à rendre compte d'un sous-ensemble des propriétés d'un objet non purement scientifique, par exemple dans un but de prévision, ou d'un *modèle théorique*, pour lequel ce n'est plus l'objet qui est donné mais la théorie élaborée pour l'expliquer ? Il est relativement facile de montrer qu'un modèle est non-cohérent, trop schématique, trop loin du concret ; mais il est rigoureusement impossible d'affirmer qu'un modèle est une représentation

correcte du réel : la concordance entre les résultats d'une simulation et les faits observés peut n'être qu'une coïncidence numérique, et, comme le soulignait déjà MERSENNE, une concordance totale entre le comportement d'un automate et celui d'un être vivant ne nous renseigne pas sur la nature profonde des mécanismes mis en jeu dans l'être vivant ("[Dieu] les a pu autrement faire"). Il en est du modèle simulé comme de toute théorie : il constitue une tentative pour représenter de façon intelligible une portion de la réalité, de manière à pouvoir dans une certaine mesure en expliquer et en prévoir l'évolution.

La non-identité fondamentale entre un phénomène et le modèle qui en est l'image entraîne en particulier les conséquences suivantes (130, Pp. 2-3) :

- tout modèle a un domaine d'application en dehors duquel son utilisation perd toute signification ;
- un modèle a sa logique propre, qui lui interdit de dépasser un certain degré de ressemblance avec la réalité ;
- en manipulant un modèle, il faut se garder de considérer celui-ci comme l'objet de l'étude, et ne pas perdre de vue la réalité qu'il représente ;
- s'il y a désaccord entre le modèle et les faits observés, il faut abandonner le modèle ;
- il est possible de faire appel simultanément, pour représenter une même réalité, à plusieurs modèles incompatibles entre eux (c'est-à-dire non isomorphes ; cf. § 1.2.6.).

Ce dernier point mérite un examen attentif. L'exemple le plus souvent cité par les spécialistes de la philosophie des sciences est celui du dualisme *onde-corpuscule* de l'électron, c'est-à-dire de la représentation *simultanée* de l'électron par un solide doué d'une masse

et d'une charge électrique, et par une onde susceptible entre autres de provoquer des phénomènes de diffraction dans un cristal (*). Ces deux modèles sont absolument incompatibles entre eux ; il faut donc, au niveau d'une théorie d'ensemble, soit les remplacer par un modèle plus général qui élimine leurs contradictions, soit conserver provisoirement ces deux modèles en définissant rigoureusement leurs domaines d'application. HEISENBERG avait noté, dans Les principes physiques de la théorie des quanta : "il est manifeste que l'application simultanée, sans critique, des deux représentations ondulatoire et corpusculaire, conduit à des contradictions immédiates. On peut donc en conclure que l'emploi de ces représentations doit avoir des limites fixées par la nature" (cité par 195, p. 78, n.1). On conçoit aisément la difficulté pour le théoricien d'admettre la coexistence de deux modèles incompatibles dont les domaines d'application ont une importante partie commune, et dont les concepts sont voisins, sinon identiques. Jean ULLMO (195, Pp. 78-81) fait remarquer que l'adoption de deux modèles contradictoires repose sur des "conclusions abusives que provoquent les pièges du langage". Certes, la localisation des trajectoires de l'électron et de ses interactions avec d'autres objets peuvent être décrites en termes de mécanique des solides ; de même, la diffraction des électrons dans un cristal peut être décrite en termes d'ondes lumineuses. Mais, à partir de ces analogies, il n'était pas nécessaire de conclure à la "nature" corpusculaire ou à la "nature" ondulatoire de l'électron. Les *équations d'onde* de la mécanique quantique rendent correctement compte de l'ensemble des phénomènes observés ; qu'il y ait interaction avec des atomes ionisables répartis au hasard (trajectoire

(*) Sous leur forme scientifique, le premier modèle est dû à NEWTON, le second à son contemporain HUYGENS. Les mérites respectifs de ces deux modèles, du point de vue des savants du 17^e siècle, sont plaisamment exposés sous forme de dialogue dans A. EINSTEIN et L. INFELD, L'évolution des idées en physique, (Payot, 1963), Pp. 101-105.

en chambre à bulles) ou avec des structures cristallines régulières (diffraction dans un cristal), "l'effet résultant peut avoir des apparences différentes (et que seule la comparaison illégitime avec des phénomènes macroscopiques semble rendre contradictoires) alors que le phénomène élémentaire d'interaction, représenté par l'équation d'onde, est le même" (ibid., p. 81). Par conséquent, s'il est possible de faire appel simultanément à plusieurs modèles incompatibles entre eux pour représenter une même réalité, nous devons cependant ne considérer cette possibilité que comme un palliatif provisoire à l'insuffisance de la théorie.

Si le modèle est une construction, un assemblage de propositions abstraites réalisé par le chercheur, ce dernier doit en principe être capable de contrôler parfaitement les mécanismes du modèle. En pratique, il est rare qu'un modèle de simulation, même parfaitement cohérent, soit aussi bien connu et contrôlé qu'un modèle analytique. En analysant les différents types de modèles de simulation, nous avons distingué (§ 2.2.4.c) les modèles *complets*, qui visent à représenter dans sa totalité le système original, et les modèles *incomplets*, qui requièrent un certain volume de données empiriques pour pallier les lacunes de la formalisation. Dans les modèles incomplets, "des ponts empiriques sont jetés de place en place" (4, p. 339), en utilisant des régularités empiriques, éventuellement synthétisées avant leur introduction comme paramètres d'entrée du modèle. Dans les sciences sociales, les modèles complets sont presque toujours formés de la juxtaposition de micro-modèles représentant des unités élémentaires (ex. : individus dans un petit groupe ; entreprises dans une économie de marché). Chaque micro-modèle constitue un sous-système dont le fonctionnement est parfaitement connu du chercheur. L'objectif d'une expérience de simulation est alors l'étude du comportement de l'ensemble du système, compte-tenu des interactions entre ses éléments (*couplages*). Or, il est rare que les règles qui régissent l'agrégation des divers sous-systèmes, et leurs

conséquences, soient complètement explicitées par le chercheur. Quant aux modèles incomplets, ils sont par définition en partie incontrôlés : si les données empiriques qu'ils utilisent avaient été complètement analysées, elles seraient remplacées par des propositions abstraites plus facilement manipulables.

Pour réduire la marge d'incertitude qui subsiste dans le modèle, le chercheur procède généralement à l'*ajustement des paramètres*. Pour cela, il recherche quelles sont les valeurs des paramètres d'entrée qui permettront d'obtenir des valeurs des paramètres de sortie voisines des valeurs observées dans la réalité pour des situations analogues. Ce ~~k~~ ajustement pose des problèmes techniques délicats (cf. 135, Pp. 35-36). Mais surtout, ce "calibrage du modèle" présente un certain risque. Prenons l'exemple d'un modèle complet de prévision électorale, comportant quatre ou cinq paramètres d'entrée. Si nous disposons des résultats de quatre élections antérieures, il est vraisemblable que nous trouverons plus d'une combinaison des valeurs de nos paramètres d'entrée nous permettant de "prévoir rétrospectivement" (*postdiction*) les résultats observés ; laquelle de ces combinaisons choisirons-nous pour prévoir les résultats de la prochaine consultation ? En supposant que le "calibrage du modèle" nous permette de déterminer sans ambiguïté les valeurs des paramètres d'entrée, quelle garantie avons-nous que les mécanismes en jeu dans le modèle, qui "expliquent" les élections passées, "expliqueront" aussi l'élection à venir ? On peut toujours imaginer un autre modèle, également ajustable aux résultats passés, qui fournisse des prévisions différentes. Comment choisir entre ces modèles ? Le problème de l'ajustement du modèle est qu'*il est presque toujours possible d'ajuster aux observations un modèle quelconque*, à condition que celui-ci admette un nombre suffisamment élevé de paramètres. Le critère de la validité du modèle est alors

non seulement la valeur des résultats, mais surtout la *vraisemblance* des mécanismes qu'il met en cause. L'importance de ce problème n'est pas la même selon qu'il s'agit d'une simulation orientée vers l'action à court terme, ou d'une simulation visant à une meilleure compréhension des phénomènes. Dans le premier cas, un modèle simplement homomorphe à l'original peut apporter des éléments de décision satisfaisants, et les différences de prévision entre deux modèles "équivalents" peuvent être négligés. Dans le second cas, le chercheur ne peut se limiter à une approximation grossière du phénomène ; il lui faut construire un modèle capable de s'insérer dans une théorie sociologique, et susceptible de cadrer avec les données d'observation. Si le problème se pose du choix entre deux modèles également bien ajustés, la préférence doit aller à celui dont les implications théoriques sont les plus importantes. Le modèle sera d'autant meilleur qu'il se fondera sur un plus grand nombre de faits résumés par la théorie, donc sur une théorie plus riche ; et les résultats de la simulation n'en auront que plus de signification sur le plan scientifique.

La simulation est un cas très particulier d'approche expérimentale ; au lieu d'expérimenter sur la réalité, on expérimente sur la théorie. Aussi ne saurait-elle remplacer complètement la démarche de l'expérimentateur. Il nous semble que l'intérêt primordial de la simulation est de fournir, à l'observateur comme à l'expérimentateur, un faisceau d'"hypothèses de rechange" (*alternative hypotheses*) d'un assez haut niveau de complexité. La démarche du sociologue analysant un ensemble de données empiriques consiste à confronter ces données à ce qu'elles seraient si telle hypothèse était vraie. L'exemple le mieux connu et le plus formalisé en est la théorie des tests statistiques, permettant de mesurer, en termes de probabilité d'occurrence, la distance existant entre les observations et une hypothèse donnée. On utilise pour

cela un modèle déterministe simple décrivant les conséquences de l'hypothèse testée. Lorsque l'on désire éprouver la validité d'une théorie complexe, il est possible d'en déduire les conséquences et de les décrire ; mais, en l'état actuel de la théorie statistique, on dispose rarement de tests permettant de mesurer l'écart entre les conséquences de la théorie et les observations. La méthode la plus simple est alors de renoncer à calculer l'incidence des erreurs de mesure sur la description des conséquences de la théorie, et d'introduire les fluctuations aléatoires dans le modèle lui-même. A condition de répéter la simulation en fonction de la précision désirée, le chercheur a ainsi le moyen de tester la validité d'une hypothèse complexe, en comparant les données d'observation à la distribution des données simulées. La simulation peut apparaître ainsi comme un substitut de l'analyse statistique (cf. 33, p. 406).

Les spécialistes en sciences humaines qui ont traité de la simulation n'ont pas toujours exprimé avec clarté ce qui distingue les techniques de simulation des techniques expérimentales classiques. Et la confusion que l'on relève chez un auteur comme GUETZKOW (cf. 105 et 106), entre la simulation sur ordinateur et les techniques de la psychologie sociale apparentées au jeu de rôle, tend à faire croire que la simulation joue, dans la démarche du chercheur, une fonction analogue à celle que jouent les techniques expérimentales. Certes, les termes *expérience*, *expérimental*, *expérimentation*, ne sont pas dépourvus d'ambiguïté (cf. 132, Pp. 321-324) : une science expérimentale peut fort bien, comme le fait remarquer P. ACHARD, ne pas recourir nécessairement à l'expérimentation, et il est des cas où "l'expérience est une sélection de faits préexistants" (4, p. 335). Si nous limitons nos réflexions à l'*expérimentation*, qui est l'*observation provoquée* sur une portion de la réalité, afin de mieux comprendre les lois de cette portion de réalité,

il devient clair que la simulation diffère nettement de cette technique de par sa finalité. Le rôle de l'expérimentation est de permettre la vérification de certaines hypothèses faites par le chercheur sur la réalité. La simulation, ou "expérimentation sur un modèle", sert à explorer une partie des conséquences d'une théorie ; en particulier, elle rend possible la vérification de la compatibilité, avec les observations déjà recueillies, d'une théorie, même si celle-ci est partielle ou très complexe. En d'autres termes, la simulation est un moyen d'évaluer la plausibilité d'une théorie, *compte-tenu de ce que l'on sait déjà* sur les phénomènes à expliquer ; tandis que l'expérimentation est un moyen d'accroître (en principe) ou au contraire d'infirmer la plausibilité d'une théorie *en accroissant la connaissance empirique* des phénomènes à expliquer. Dans la dynamique de la démarche scientifique, la simulation constitue une aide technique à l'élaboration de modèles explicatifs mieux adaptés aux observations ; elle peut avoir à la fois une valeur heuristique et une fonction d'épreuve de la cohérence d'une théorie. Contrairement à l'expérimentation, *la simulation n'apporte sur les phénomènes étudiés aucune information nouvelle*, même si sa mise en oeuvre concourt à une meilleure compréhension de ceux-ci.

7.3. INTERET DE LA SIMULATION POUR LE SOCIOLOGUE.

Les implications théoriques et pratiques de la simulation sont encore mal connues, même dans les domaines où ces techniques font l'objet d'une utilisation intensive (recherche opérationnelle notamment). Evaluer l'intérêt que la simulation pourra présenter pour le sociologue, c'est dans une très large mesure faire un pari sur l'avenir. Pour tenter

d'apporter une réponse provisoire à cette question, nous devons au préalable :

- situer les techniques de simulation par rapport aux autres techniques utilisées par le sociologue ;

- dresser un bilan des applications de la simulation déjà réalisées en sociologie.

a) Place des techniques de simulation.

Les techniques de simulation constituent un mode particulier de traitement des données sociologiques. Dans un contexte un peu différent, J.C. GARDIN a proposé une classification des procédures de traitement sur ordinateur des données anthropologiques qui nous paraît transposable à la sociologie (89). L'usage de cette classification implique toutefois que l'on admette que toute procédure de traitement de données sociologiques est automatisable ; il nous paraît en tous cas que toutes les procédures utilisées actuellement sont automatisables, dans la mesure où elles sont rigoureuses, c'est-à-dire codifiées de manière à pouvoir être enseignées à autrui.

La classification de GARDIN repose sur la combinaison de trois types de données avec trois types d'opérations à effectuer sur ces données. Parmi les données, l'auteur distingue :

- les *signaux*, entités physiques traitées comme telles, indépendamment de toute référence à une quelconque signification. Exemples : sons, mots écrits, idéogrammes. Cette catégorie s'oppose aux deux autres, que GARDIN appelle données symboliques, ou *signes*.

- les *langages naturels*, dans la mesure où ceux-ci sont objets d'étude en eux-mêmes.

- *les codes spéciaux*, qui ne sont plus objets d'étude comme les langages naturels, mais servent d'intermédiaires pour l'étude d'autres objets qu'ils servent à coder. Exemples : tout système de codage, y compris les langues naturelles en tant que supports d'information.

Les opérations que l'on peut effectuer sur ces données se classent en :

- opérations de *compilation* (*), réalisant sur les données certaines manipulations sans transformer celles-ci. Exemples : comptages, reclassements, mises à jour de fichiers ; recherche d'information dans un corpus.

- opérations de *systematisation statique*, faisant appel à un appareil théorique (statistique, linguistique, sociologique, etc.) pour décrire globalement un système sous une forme synchronique. Exemples : analyse descriptive d'un langage, présentation synthétique des résultats d'un sondage. Cette catégorie d'opérations peut être subdivisée en deux classes : celle des opérations purement descriptives ayant pour objet de rattacher les données empiriques à un système de classification pré-existant ; et celle des opérations de recherche, à partir des données, d'un système de classification empirique (taxinomie).

- opérations de *systematisation dynamique*, utilisant également un appareil théorique pour organiser les données relatives à un système sous une forme diachronique. Exemples : évolution des langues, courants d'opinion, processus de création.

(*) On peut regretter le choix de ce mot, qui crée une confusion avec la traduction en langage machine d'un programme écrit dans un langage évolué, opération que les informaticiens appellent également *compilation*. La compilation au sens des informaticiens se range dans la *systematisation statique*, et non dans la compilation selon la terminologie de GARDIN.

A partir de ces distinctions, il est possible de construire le tableau suivant (tableau n° 35, page 419). On voit immédiatement que les techniques de simulation se rangent dans la catégorie des opérations de *systematisation dynamique*, appliquée aux *données symboliques*. Cette classification met en évidence les limites probables à l'application des techniques de simulation. Mais il faut souligner que la *systematisation dynamique* de données symboliques, et particulièrement de codes spéciaux, n'est pas réalisée en sociologie par la seule simulation ; les modèles de contagion basés sur la loi de POISSON ou les chaînes de MARKOV sont d'une utilisation courante dans l'analyse des processus (cf. par ex. 47, ch.10 et 11 ; 33, ch. 7 et 8).

b) Bilan des applications réalisées actuellement.

Malgré l'intérêt que les techniques de simulation semblent présenter pour la sociologie, le bilan des expériences décrites dans la littérature est assez décevant. Dans une courte évaluation de l'apport des simulations réalisées en sociologie politique (157), J.-G. PADIOLEAU dresse un véritable réquisitoire, qui nous paraît pouvoir être appliqué à l'ensemble des simulations portant sur des problèmes sociologiques. Les critiques formulées par PADIOLEAU peuvent être résumées ainsi : la quasi-totalité des expériences de simulation déjà réalisées sont sans intérêt théorique et/ou sans valeur méthodologique. Le plus grand nombre des simulations décrites dans la littérature repose sur un modèle non satisfaisant du point de vue de la théorie : ou bien les propositions simulées sont d'une telle généralité qu'elles paraissent banales ; ou bien le modèle omet une dimension essentielle du phénomène simulé ; ou bien les résultats de la simulation confirment

Données	Opérations		
	Compilation	Systématisation	
		statique	dynamique
Langages naturels	Comptage de fréquences d'unités linguistiques (morphèmes, mots). Constitution d'indexs	Analyse morphologique ou syntaxique d'une langue. Traduction. Classification des langues.	Etude de l'évolution linguistique (lois phonétiques, règles d'interaction entre langues, etc.).
Codes spéciaux	Recherche documentaire. Analyse de contenu.	Constitution de typologies ; classification. Analyse de réseaux de relations.	Etude des processus, fondée sur : - la théorie sociologique (relations entre individus, mécanisme du vote), - la théorie des jeux, - la théorie psychologique (apprentissage, décision), - etc..
Signaux	Activités perceptives (ex. : reconnaissance des formes).	Représentation rationnelle des données empiriques.	Etude des automates.

Tableau n° 35 : Classification des techniques de traitement de données anthropologiques, d'après GARDIN.

des "lois" sociologiques depuis longtemps admises et validées. Parmi les modèles qui échappent à la critique de la pauvreté théorique, certains se contentent de juxtaposer de nombreuses propositions empruntées à la fois aux théories sociologiques et aux constats empiriques, sans définir rigoureusement les relations entre ces propositions ; aussi les résultats de leurs simulations sont-ils ininterprétables. Lorsque l'on rencontre, dans la littérature, des modèles simulables présentant une synthèse cohérente de nombreuses propositions théoriques, on constate que tous ces modèles sont restés à l'état de projets, leur complexité s'opposant, en l'état actuel des possibilités des ordinateurs, à leur réalisation sur machine. Si d'autre part nous limitons notre examen aux modèles qui ne sont pas trop pauvres sur le plan théorique et qui cependant ont été effectivement simulés, nous constatons que l'ajustement du modèle et sa validation sont généralement critiquables : lorsque le modèle compte de nombreux paramètres, rien n'assure que l'ajustement soit autre chose qu'une simple coïncidence numérique ; de plus, il arrive que la "validation" du modèle soit effectuée sur les mêmes données qui ont servi à son ajustement (*petitio principii*) ; enfin, nous savons que la validation statistique des modèles pose des problèmes théoriques encore non complètement résolus (cf. 135, Pp. 35-36).

Un tel jugement est sévère ; il nous semble cependant assez conforme à la réalité. Comment s'explique la pauvreté de ce bilan ? Les raisons de cette pauvreté nous semblent être que la simulation a été introduite récemment en sociologie, que son usage se développe peu, et que cette approche est d'une utilisation trop récente dans les autres domaines pour pouvoir être parfaitement maîtrisée. Les sociologues ont attendu près de dix ans pour appliquer aux faits sociaux les techniques de simulation qui paraissaient cependant avoir fait leurs preuves en recherche opérationnelle ; et ces techniques sont encore loin, malgré

l'intérêt qu'elles présentent, d'être pour le sociologue d'un usage courant. Nous voyons à cela deux causes. La première, que nous avons déjà notée, est que la simulation est une approche dont l'utilisation systématique est d'apparition trop récente, même en recherche opérationnelle, pour que la méthodologie de cette approche ait pu être définie avec rigueur. Or, la démarche du "chercheur opérationnel", qui expérimente sur la théorie fautive de pouvoir agir sur la réalité, n'est pas un simple prolongement de la méthode expérimentale codifiée par François BACON, John Stuart MILL, Claude BERNARD ; elle constitue *un mode de pensée d'un type nouveau*, dont les lois sont confusément pressenties par ceux qui l'utilisent, mais n'ont pas encore été explicitées. La seconde cause de la lenteur de l'introduction de cette approche en sociologie est le peu de progrès que cette science a accomplis sur la voie de la formalisation et de la rigueur. Aussi, le sociologue qui désire recourir à ces techniques doit-il suivre un double apprentissage : l'apprentissage de la démarche scientifique, à laquelle la seule réflexion sur les oeuvres des fondateurs de la sociologie ne l'ont pas toujours suffisamment préparé (*), et l'apprentissage d'un mode de pensée nouveau, dont les initiateurs eux-mêmes n'ont pas encore énoncé les règles.

L'accent mis sur l'aspect *apprentissage* que présentent les expériences de simulation en sociologie nous paraît essentiel pour comprendre le peu d'apport théorique de ces expériences à l'heure actuelle. Tout se passe en effet comme si les sociologues cherchaient seulement

(*) Non que nous accusions les grands sociologues d'avoir manqué de rigueur dans leurs travaux. Mais cette rigueur n'est pas toujours perçue, et le caractère exemplaire de la démarche suivie par DURKHEIM dans *Le suicide* peut échapper à qui ignore les principes de l'analyse causale (cf. p. ex. 33, 1^{ère} et 2^e parties).

à se familiariser avec ces techniques, sans espérer en retirer un profit scientifique dans l'immédiat. Lorsque l'on désire se familiariser avec une technique nouvelle, on commence par chercher à retrouver, à l'aide de cette technique, des résultats qui ont été fournis par des techniques mieux éprouvées, et que l'on peut tenir pour assurés. L'étape suivante consiste à expérimenter la technique nouvelle dans des domaines ou selon des directions de recherche moins connus, mais en limitant étroitement les conditions de l'application, de manière à en garder sous son contrôle chacune des étapes. Tel est le cas des simulations reposant sur un modèle beaucoup trop simple pour prétendre à une quelconque vraisemblance sociologique. L'étape finale, celle de l'intégration à part entière de la technique nouvelle parmi l'ensemble des techniques de recherche, n'est pas encore atteinte ; mais les nombreux projets de simulations complexes, actuellement non réalisés, témoignent du but visé par les chercheurs par delà le simple apprentissage de la technique.

Est-ce à dire que l'unique intérêt pour la sociologie des expériences réalisées jusqu'à ce jour réside dans les progrès des sociologues dans la maîtrise des techniques de simulation ? Nous ne le pensons pas. Lorsqu'une expérience de simulation aboutit à des résultats déjà connus, voire triviaux, l'essentiel est que ces résultats sont obtenus *à partir d'une formalisation* du phénomène étudié qui, si elle n'est pas le premier effort de formalisation de ce phénomène, doit différer sensiblement des modèles analytiques précédemment utilisés. A ce titre, cette formalisation est à elle seule un apport théorique. Lorsque l'expérience de simulation apporte des résultats originaux, mais à partir d'un modèle trop pauvre pour être vraisemblable, elle réalise artificiellement une expérience que l'on n'aurait généralement pas pu mener à bien dans la réalité, en isolant et en concrétisant les

effets de quelques lois simples. Les expériences de ce type nous paraissent devoir jouer un rôle important dans le progrès des connaissances scientifiques. Les théories scientifiques particulières sont des modèles complexes, mais constitués à partir d'un petit nombre de lois simples. Les réflexions et les tâtonnements en vue de construire des modèles abstraits susceptibles de rendre compte d'une masse importante d'informations empiriques peuvent trouver une aide considérable dans les expériences de simulation portant sur des modèles très simples : à condition que les lois introduites dans le modèle simulé présentent un degré élevé de vraisemblance, la simulation du modèle permet d'évaluer le pouvoir d'explication contenu dans le modèle, et d'identifier les points sur lesquels le modèle doit être corrigé.

A l'objection selon laquelle de tels modèles ne seraient pas féconds parce qu'ils "irréalistes", nous répondrons qu'aucune loi scientifique n'est "réaliste". Pour expliquer à peu près correctement la chute d'un corps dans des conditions de laboratoire, nous devons faire intervenir au moins deux lois totalement "irréalistes" parce que presque imperceptibles à notre échelle (et certainement inobservables dans la vie quotidienne) ; la loi de l'attraction universelle (qui fait que ce corps et la terre s'attirent réciproquement), et l'équation de la force centrifuge (la rotation de la terre compensant en partie l'effet de l'attraction). Pour rendre compte de la chute d'un corps dans des conditions naturelles, nous devrions complexifier le modèle en faisant intervenir la densité du corps relativement à celle de l'air, la forme de ce corps, son orientation par rapport à la direction de l'attraction ; etc. ; mais cette complexification ne saurait avoir pour conséquence de nier la valeur des lois du modèle primitif. Nous avons l'impression que l'écart qui sépare un modèle simple (constitué à partir de quelques lois) d'un modèle "réaliste" (assez complexe pour rendre compte des

observations) est plus grand dans les sciences humaines que dans les sciences physiques ; et que par conséquent il sera nécessaire de faire appel à un plus grand nombre de lois en sociologie pour atteindre au même degré d'intelligibilité qu'en physique par exemple. Il se peut que nous soyons ici victime d'une illusion d'optique, due aux différences de développement entre les deux sciences. Si néanmoins cette impression est fondée, c'est un argument de plus en faveur de l'intérêt des simulations, puisque celles-ci permettent au sociologue de concrétiser les conséquences de sa théorie, et de modifier celle-ci, sans attendre que les événements viennent réaliser les expérimentations dont il aurait eu besoin.

c) L'avenir des techniques de simulation en sociologie.

Nous ne prétendons pas prédire avec exactitude le rôle que jouera la simulation dans la pratique sociologique ; un tel objectif nous paraît être du domaine de la science-fiction. Nous voulons simplement décrire, dans ce dernier paragraphe, les obstacles possibles au développement de ces techniques, leurs apports prévisibles à court terme, et les espoirs à moyen terme qu'elles suscitent déjà chez certains sociologues.

Nous avons recensé trois sources principales de réticences à l'égard de l'emploi en sociologie des techniques de simulation : la résistance à la formalisation, la croyance à l'inadéquation de ces techniques aux sciences humaines, et la conscience des risques que leur utilisation comporte. La *résistance à la formalisation* nous paraît être un phénomène normal au commencement de toute science. Le langage courant, avec ses redondances et ses ambiguïtés, paraît véhiculer une information

plus riche et plus nuancée qu'un langage plus rigoureux, que ce soit la langue juridique ou les mathématiques. Aussi, l'effort important que requiert tout essai de formalisation semble paradoxalement se solder tout d'abord par un appauvrissement de l'information ; appauvrissement compensé en partie, il est vrai, par un gain en cohérence et en précision. Cet appauvrissement de l'information n'est en réalité sensible que dans une première phase, lorsque la formalisation, inachevée, fait appel à un modèle trop sommaire. Mais l'exemple des autres sciences montre que les progrès de la formalisation donnent à la théorie une expression plus rigoureuse, mais aussi plus riche et plus nuancée que les formulations "littéraires" de la phase pré-scientifique ; de plus, seul un langage formel (ou formalisé) permet la découverte des conséquences observables d'un système abstrait, c'est-à-dire la validation de la théorie. La résistance à la formalisation se fait surtout sentir à l'égard de l'emploi en sociologie du langage mathématique ("on ne saurait mettre l'homme en équations"). Il est possible que cette résistance soit moindre à l'égard des techniques de simulation, dans la mesure où l'expression du modèle, et les procédés de déduction à partir de celui-ci, revêtent des formes apparemment moins abstraites que le langage mathématique.

La croyance à l'inadéquation des techniques de simulation aux données sociologiques n'est pas toujours clairement formulée, et nous paraît traduire surtout une certaine forme de misonéisme. L'expression de cette croyance est que les techniques de simulation sont bien adaptées à certains types de phénomènes ou de problèmes, comme l'analyse de décisions économiques, mais ne sauraient convenir à la recherche fondamentale sur les faits sociaux. Le fondement d'une telle attitude nous semble être toujours l'imperfection de la technique au début de son utilisation. Par exemple, les philosophes scolastiques ont récusé l'usage de la loupe (et plus tard du microscope) pour les observations

scientifiques, non en raison de l'imperfection réelle de ces instruments jusqu'au 17^e siècle, mais "en prétendant que la nature doit être observée sans intermédiaire, sous peine de n'en recevoir qu'une image déformée et trompeuse" (58, p. 57). De même rappelle Claude BERNARD, "la spontanéité dont jouissent les êtres doués de la vie a été une des principales objections que l'on a élevées contre l'emploi de l'expérimentation dans les études biologiques" (24, p. 99). Ainsi CUVIER, au lieu d'accuser l'indigence de l'arsenal expérimental du début du 19^e siècle, explique que vouloir séparer une partie d'un corps vivant, "c'est la reporter dans l'ordre des substances mortes, c'est en changer entièrement l'essence" (cité par 24, p. 100) ; et John Stuart MILL lui-même déclare que l'expérimentation est inapplicable aux phénomènes biologiques, sociologiques, et aux faits physico-chimiques les plus complexes (cf. 74, p. 125). De telles objections sont toujours le fait de savants qui, n'ayant pas eu l'occasion de manier la technique nouvelle qu'ils critiquent, ne la jugent que sur ses défauts de jeunesse et non sur les potentialités qu'elle renferme. La diffusion rapide des procédures de traitement de l'information sur ordinateur devrait contribuer à réduire cette forme de réticence.

La troisième source de réticences que nous avons citée nous paraît être la mieux fondée : elle est le fait de chercheurs qui ont déjà utilisé la simulation, et qui ont conscience des dangers présentés par une mauvaise utilisation de cette approche. Ainsi que nous l'avons souligné à plusieurs reprises, la démarche qui consiste à expérimenter sur un modèle théorique n'est pas encore codifiée, et son utilisation inconsidérée peut conduire à des absurdités. C'est pourquoi certains chercheurs redoutent, à juste titre, les erreurs d'interprétation que peuvent faire des néophytes insuffisamment familiarisés avec cette nouvelle forme d'approche expérimentale. Le risque le plus grand étant de

confondre implicitement le modèle avec la réalité qu'il vise à représenter, et de ne pas distinguer clairement la simulation de l'expérimentation. Ce risque est important, et il est certain que le mauvais usage d'une technique, s'il ne prouve rien contre la technique elle-même, contribue largement à la discréditer dans l'esprit de certains de ses utilisateurs potentiels. Mais l'exemple des premières utilisations des méthodes statistiques en psychologie et en sociologie nous montre que si ce risque est un mal difficilement évitable, il ne constitue pas un obstacle insurmontable à la diffusion de la simulation.

Malgré les difficultés que nous venons d'énumérer, l'emploi des techniques de simulation devrait donc se répandre parmi les sociologues, et ce d'autant plus rapidement que le coût du traitement de l'information sur ordinateur sera meilleur marché, et que les techniques de programmation seront mieux connues. Quels bénéfices la sociologie pourra-t-elle en retirer dans un proche avenir ? L'usage de la simulation nous semble recéler deux facteurs de progrès. Le premier réside dans l'obligation, pour le sociologue, de traduire dans une langue formalisée des théories exprimées sous une forme littéraire. Cette obligation a pour effets d'habituer le chercheur à discipliner sa pensée et son expression, afin d'atteindre à la rigueur propre à toute démarche scientifique ; de le conduire à préciser et enrichir des théories dont les lacunes ou les présupposés ne sont pas apparents dans leur expression littéraire, et à se débarrasser de ses *prénotions* ; de le contraindre enfin à faire preuve d'esprit critique à l'égard de théories jugées intouchables parce qu'émanant de sociologues réputés, et de ne pas craindre d'en rejeter certaines pour manque de cohérence interne. L'habitude d'une pensée plus rigoureuse devrait aboutir d'une part à un approfondissement de la méthodologie proprement sociologique, en raison d'une conscience plus aigüe des problèmes théoriques et pra-

tiques rencontrés ; d'autre part à un enrichissement de la théorie sociologique, dû à un recours plus fréquent et plus systématique aux approches de type qualitatif. En effet, la mise au point d'un modèle formel impose au sociologue un approfondissement de la théorie qui lui permet de constater les lacunes de celle-ci. Or, il est indispensable, pour l'achèvement du modèle, que ces lacunes soient comblées. Si l'on veut que la simulation ait une valeur scientifique, il est nécessaire que les lacunes soient comblées non par des propositions ou des paramètres fixés *a priori*, mais à l'aide de données d'observation plus fines et plus complètes que celles dont on disposait auparavant. Le recours au langage formalisé conduit donc au développement des approches de type qualitatif, et impose une compréhension plus intime du phénomène étudié.

Le second facteur de progrès est propre à la simulation : il consiste en la possibilité de manipuler une théorie (ou le modèle qui la représente) comme l'on manipule un dispositif expérimental. Cette faculté présente tout d'abord un intérêt didactique, bien exprimé par ASHBY (15, Pp. 461-462) : "la simulation peut être utilisée pour mettre l'accent sur un concept et le clarifier. L'énoncé de propositions abstraites sur un mécanisme tend à être sommaire, peu convainquant, et peu stimulant pour la réflexion. Le modèle qui illustre une proposition d'une façon vivante a pour effet non seulement d'emporter la conviction, mais aussi de faire voir à l'observateur toutes sortes de conséquences et de développements possibles". Si l'on construit un modèle résumant l'ensemble des connaissances dont on dispose sur un sujet donné, par exemple la physiologie du cerveau, "l'étudiant, le chercheur avancé (et peut-être aussi le médecin cherchant un diagnostic) disposeront d'un outil capable de répondre aux diverses questions qui peuvent se poser. Le modèle jouerait alors à la fois le rôle d'archives et celui d'ordinateur".

L'apport le plus important de l'expérimentation sur la théorie est qu'elle permet d'adopter en macro-sociologie un mode de pensée de type expérimental. Dans les Règles de la méthode sociologique, DURKHEIM écrivait : "puisque (...) les phénomènes sociaux échappent évidemment à l'action de l'opérateur, la méthode comparative est la seule qui convienne à la sociologie" (74, p. 124). Ce qui, évidemment, n'exclut pas le recours à un raisonnement de type expérimental (DURKHEIM qualifie la méthode comparative d'*expérimentation indirecte*), mais limite les possibilités de comparaison du sociologue aux faits qui se sont produits spontanément. En mettant à la disposition des sociologues les moyens d'utiliser un type de raisonnement plus proche de celui de l'expérimentateur que la méthode comparative, la simulation peut amorcer une révolution dans la manière d'aborder les problèmes sociologiques. Que sera cette révolution ? Il ne nous paraît pas possible encore d'en décrire les traits, car l'usage de la simulation en sociologie n'a pas atteint un degré suffisant de maturité, et sa théorie reste à faire. Par contre, il est vraisemblable que les prochains développements de cette approche se feront d'une part en direction de l'expérimentation sur des modèles complets très simples, d'autre part vers l'utilisation de modèles incomplets de phénomènes réels, nécessitant un volume important de données empiriques. La première direction correspond à la nécessité pour la recherche fondamentale de procéder par analyse, en étudiant les conséquences de quelques lois simples considérées isolément, et de leurs interactions, la synthèse de ces lois devrait permettre d'aboutir progressivement à une théorie générale des phénomènes sociaux, analogue, par exemple, à la théorie économique contemporaine. La seconde direction répond aux exigences de la sociologie appliquée, pour laquelle les problèmes d'analyse parcellaire, de prévision, de choix d'une stratégie, doivent être résolus empiriquement, même si les connaissances théoriques dont on dispose sont insuffisantes pour permettre la découverte de la solution optimum. Dans le premier cas, le but visé est un meilleur ajustement de la théorie aux observations ; dans le second, c'est la description

du comportement d'un système malgré les lacunes et les insuffisances de la théorie.

A plus long terme, quelques sociologues ont tenté de définir les espoirs que suscitaient chez eux les techniques de simulation, et les limites qu'ils croyaient devoir fixer à l'emploi de ces techniques, au delà de la simple extrapolation des applications actuelles. Nous avons déjà signalé les limites inhérentes à la méthode elle-même (systématisation dynamique de données symboliques) ; mais, à l'intérieur de ces limites, existe-t-il en sociologie des domaines d'application privilégiés, et des domaines auxquels la simulation convient mal ? Jacques MAITRE (142, p. 44) a contribué à répondre à cette question, en notant la place importante que tiennent le temps et surtout l'espace dans les expériences de simulation. La plupart des exemples de simulation "montrent l'importance des relations spatiales : il s'agit de déplacements (transports, déménagements, communications de messages) ou de phénomènes aisément représentables par des déplacements (démographie). Autrement dit, nous sommes surtout dans le domaine de la *physique sociale*, où les conditionnements matériels canalisent le plus directement les processus". Est-ce à dire que le champ d'application de la simulation se limitera en pratique à la "physique sociale" ? Il convient ici de se garder de confondre le phénomène sociologique avec la représentation que l'on peut en donner, ou, comme le dit P. ACHARD, de confondre le *programme* de la simulation avec le *modèle* qu'il met en oeuvre sur une machine donnée (4, p. 338). Il est certain que toute simulation sur ordinateur numérique impose que le phénomène simulé *soit traduit dans le langage* de l'espace et du temps ; mais cette traduction nécessaire n'implique pas que ce phénomène *se déroule effectivement dans l'espace et le temps*. Par exemple, on utilise une représentation spatiale pour décrire la structure d'une entreprise

(organigramme), ou le réseau des relations interindividuelles à l'intérieur d'un groupe (sociogramme), sans qu'il existe une relation simple entre la situation dans l'espace des services de l'entreprise ou des individus du groupe, et celle des cases de l'organigramme ou des sommets du sociogramme. De même, on peut décrire un processus quasi-instantané complexe (par exemple le processus de rééquilibrage de l'*Homeostat*) comme s'il s'agissait d'une suite de micro-processus se déroulant l'un après l'autre, alors que leur déroulement est simultané ; si l'on simule ce processus sur un ordinateur numérique, il est nécessaire de recourir à un algorithme traitant *en séquence* des événements *simultanés*. Il ne s'agit là que de procédés commodes de description d'un phénomène ou de programmation d'un modèle, qui ne dépendent que des *formes a priori de la sensibilité* (pour utiliser la terminologie kantienne) et de la technologie de l'ordinateur numérique que l'homme a créé "à son image". Il semble donc qu'il n'y ait pas sur ce point de restriction quant au champ d'application de la simulation en sociologie.

Parmi les sociologues qui ont déjà acquis une certaine expérience de la simulation, James S. COLEMAN nous a paru exprimer le mieux les espoirs que les techniques de simulation semblent justifier. COLEMAN fait tout d'abord remarquer (49, Pp. 89-92) que les techniques de la sociologie empirique, telles que l'enquête par sondage ou l'interview clinique, ont surtout permis d'étudier les comportements individuels (opinions, attitudes, décisions), et non l'*organisation sociale* qui constitue pourtant un objet privilégié pour le sociologue ; en conséquence, un hiatus est apparu entre la recherche empirique et la théorie sociologique. Jusqu'à l'introduction de l'ordinateur comme auxiliaire de la recherche en sociologie, les sociologues ont manqué d'un outil adapté à l'étude des structures sociales. La simulation semble pouvoir

combler cette lacune, en permettant de synthétiser les micro-processus étudiés par la recherche empirique, de manière à obtenir des résultats se situant à un niveau d'abstraction correspondant à la théorie. Les techniques de simulation pourraient être ainsi l'occasion d'une véritable *révolution méthodologique*, analogue à celle qu'a constitué l'avènement de la sociologie empirique. Selon nous, cette "révolution" devrait d'ailleurs porter ses fruits essentiellement sur le plan de la *réflexion théorique* (heuristique, approfondissement des doctrines sociologiques, épreuve de la cohérence des modèles globaux). Certes, reconnaît COLEMAN, il y a loin des premiers résultats apportés par la simulation, à la réalisation de cette révolution ; et ces résultats apparaîtront dans quelques années comme incroyablement naïfs, tant les utilisations ultérieures demanderont de raffinement dans les techniques de recherche. Mais actuellement, cette technique est tellement à ses débuts "que, si la plupart de ses défauts demeurent encore inconnus, il en est de même pour ses potentialités les plus stimulantes" (49, p. 107).

REFERENCES BIBLIOGRAPHIQUES

REFERENCES BIBLIOGRAPHIQUES

Avertissement

Cette liste ne comporte que les articles et ouvrages que nous avons effectivement utilisés au cours de notre rédaction. On peut constituer une bibliographie plus complète sur l'utilisation de la simulation dans les sciences sociales à partir de :

- GARDIN, Natacha, Application des calculateurs aux sciences humaines. Essai de présentation documentaire. (Maison des Sciences de l'Homme, Centre de Calcul et Service d'Echange d'Informations Scientifiques, sans date).

- GETZLER, Pierre, Applications des calculateurs aux sciences humaines. Essai de présentation documentaire. Tome 2. (à paraître à la Maison des Sciences de l'Homme).

et des articles parus dans la revue trimestrielle *Behavioral Science*, éditée par l'Université du Michigan (The Mental Health Research Institute, The University of Michigan, Ann Arbor, Michigan 48104, USA).

Liste des ouvrages et articles consultés.

- 1 ABELSON, Robert P., Simulation of Social Behavior, Ch. 12 de LINDZEY, Gardner, et ARONSON, Elliot, *The Handbook of Social Psychology* (Addison Wesley, 1968).
- 2 ABELSON, Robert P., BERNSTEIN, Alex, A Computer Simulation Model of Community Controversies. *Public Opinion Quarterly*, 27 (printemps 1963), Pp. 93-122.
- 3 ABELSON, Robert P., CAROLL, J. Douglas, Computer Simulation of Individual Belief Systems, Pp. 24-30 de *Social Research with the Computer. The American Behavioral Scientist*, vol. 8 n° 9 (mai 1965).
- 4 ACHARD, Pierre, Le cadre théorique de la simulation en sciences humaines, Pp. 329-343 de LEMAINÉ, Gérard, et LEMAINÉ, Jean-Marie, *Psychologie sociale et expérimentation* (Mouton/Bordas 1969).
- 5 ADAMS, Jack A., Some Considerations in the Design and Use of Dynamic Flight Simulators, Ch. 3 de GUETZKOW Harold, *Simulation in Social Sciences : Readings* (Prentice Hall, 1962).
- 6 AGARD, J., L'art de la simulation. Bilan de quelques expériences, Ch. 1 de AGARD, J., *Les méthodes de simulation* (Dunod, 1968).

- 15 ASHBY, W. Ross, Simulation of a Brain, Ch. 19 de BORKO Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).
- 16 ASHBY, W. Ross, Modeling of the Brain, Ch. 11 de *Simulation Models and Gaming* (IBM, 1966).
- 17 ATKINSON, Richard C., The Use of Models in Experimental Psychology, Pp. 38-47 de FREUDENTHAL, Hans, *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences* (D. Reidel, 1961).
- 18 AUBERGER, M., Recherche sur un modèle de comportement (SEMA, Direction Scientifique, Note de travail n° 45, février 1966).
- 19 BARBUT, Marc, Sur le mot et le concept de modèle. *Annales*, vol. 18 (1963), Pp. 383-386.
- 20 BENAYOUN, Raphaël, Contribution à la théorie stochastique des épidémies. Applications numériques. (Thèse, Fac. Sciences de Toulouse, juin 1964).
- 21 BENSON, Olivier, Simulation of International Relations and Diplomacy, Ch. 24 de BORKO Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).
- 22 BERG, Axel, Mécanismes, direction automatique, cybernétique, Pp. 7-39 de *La Cybernétique, Recherches Internationales*, n° 29 (janv.-fév. 1962).

- 23 BERKELEY, E.C., Cerveaux géants, machines qui pensent (Dunod, 1957).
- 24 BERNARD, Claude, Introduction à l'étude de la médecine expérimentale (Garnier-Flammarion, 1966).
- 25 BERNSTEIN, Alex, RUBIN, Hanan, Artificial Evolution of Problem-Solvers Pp. 19-23 de *Social Research with the Computer. The American Behavioral Scientist*, vol. 8 n° 9 (mai 1965).
- 26 BERTAUX, Daniel, Sur l'analyse des tables de mobilité sociale, *Revue française de sociologie*, vol. 10 n° 4 (oct.-déc. 1969) Pp. 448-490.
- 27 BERTHILLIER, R., FRELY, J.-M., La simulation électronique des activités de l'entreprise (Dunod, 1969).
- 28 BINDER Eugène La génétique des populations (PUF. 1967).
- 29 BORKO, Harold, History and Development of Computers, Ch. 3 de BORKO Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).
- 30 BORKO, Harold, A Look into the Future, Ch. 25 de BORKO Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).
- 31 BOUDON, Raymond, De la simulation comme remède à quelques maux de la sociologie. *Epistémologie Sociologique*, n° 2 (janvier 1965), Pp. 83-108.

- 32 BOUDON, Raymond, Réflexions sur la logique des modèles simulés, Pp. 3-20 de *Simulation in Sociology. Archives Européennes de Sociologie*, Tome VI, numéro 1 (Plon, 1965).
- 33 BOUDON, Raymond, L'analyse mathématique des faits sociaux (Plon, 1967).
- 34 BREHIER, Emile, Histoire de la Philosophie (PUF, 1930-1938).
- 35 BRETON, Raymond, Output Norms and Productive Behavior in Non-Cooperative Work Groups (Ronéotypé. Dept. of Social Relations, Johns Hopkins Univ., sans date).
- 36 BRODY, Richard A., Varieties of Simulation in International Relations Research. Ch. 7 de GUETZKOW H. et coll., *Simulation in International Relations : Developments for Research and Teaching* (Prentice-Hall, 1963).
- 37 BROWN, J.A.C., The Social Psychology of Industry (Penguin Books, 1954).
- 38 BRUNET, Pierre, La science dans l'antiquité et le moyen-âge, Pp. 195-366 de *Histoire de la Science, Encyclopédie de la Pléiade* (Gallimard, 1957).
- 39 BUFFET, Pierre, GREMY, Jean-Paul, MARC, Marcel, SUSSMANN, Bernard, Peut-on choisir en tenant compte de critères multiples ? : une méthode (ELECTRE) et trois applications. *Metra*, vol. 6 n° 2 (juin 1967), Pp. 283-316.

- 40 CECCATO, Sylvio, Suggestions for Anthropology : the Machine which Observes and Describes, Pp. 465-500 de HYMES, Dell, *The Use of Computers in Anthropology* (Mouton, 1965).
- 41 CECCATO, Sylvio, A Model of the Mind, Pp. 105-175 de *Computers in Psychological Research* (Paris, Gauthier-Villars, 1966).
- 42 CHORAFAS, D.N., *Traité des ordinateurs* (Hermann, 1960).
- 43 CHORAFAS, D.N., *La simulation mathématique et ses applications* (Dunod, 1966).
- 44 CHURCHMAN, C.W., An analysis of the concept of Simulation, Pp. 1-12 de HOGGATT, A.C., et BALDERSTON, F.E., *Symposium on Simulation Models : Methodology and Applications to the Behavioral Sciences* (South-Western Publishing, 1963).
- 45 CLARKSON, Geoffrey P.E., A Model of the Trust Investment Process, Pp. 347-371 de FEIGENBAUM, E.A., FELDMAN, J., *Computers and Thought* (Mc Graw Hill, 1963).
- 46 COLEMAN, James S., Analysis of Social Structures and Simulation of Social Processes with Electronic Computers, Ch. 5 de GUETZKOW, Harold, *Simulation in Social Sciences : Readings* (Prentice Hall, 1962).
- 47 COLEMAN, James S., *Introduction to Mathematical Sociology* (Free Press of Glencoe, 1964).

- 48 COLEMAN, James S., Mathematical Models and Computer Simulation, Ch. 27 de FARIS, Robert E.L., *Handbook of Modern Sociology* (Rand Mc Nally, 1964).
- 49 COLEMAN, James S., The Use of Electronic Computers in the Study of Social Organization, Pp. 89-107 de *Simulation in Sociology. Archives Européennes de Sociologie*, Tome VI, numéro 1 (Plon, 1965).
- 50 COLEMAN, James S., KATZ, Elihu, MENZEL, Herbert, The Diffusion of an Innovation among Physicians. *Sociometry*, vol. 20, n° 4 (décembre, 1957), Pp. 253-270.
- 51 Control Data Corporation 3600/3800 Computer Systems FORTRAN (CDC Documentation Department, 1966).
- 52 Control Data Corporation 3600/3800 Computer Systems Library Functions (CDC Documentation Department, 1966).
- 53 Control Data Corporation 3600/3800 Computer Systems SIMSCRIPT Reference Manual (CDC Documentation Department, 1966).
- 54 COSTABEL, P., La mécanique générale, Pp. 179-186 de TATON, René, *Histoire générale des sciences* (PUF, 1957-1964).
- 55 COUFFIGNAL, Louis, La cybernétique (PUF, 1966).
- 56 CRUON, R., En guise de conclusion, Pp. 386-387 de *Digital Simulation in Operational Research* (The English University Press, 1967).

- 57 CULBERTSON, James T., Nerve Net Theory, Ch. 20 de BORKO Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).
- 58 DAUMAS, Maurice, Les instruments scientifiques aux XVIIe et XVIIIe siècles (Paris, PUF, 1953).
- 59 DAUMAS, Maurice, Esquisse d'une histoire de la vie scientifique. Pp. 3-192 de *Histoire de la Science, Encyclopédie de la Pléiade* (Gallimard, 1957).
- 60 DAVIDOVITCH, André, BOUDON, Raymond, Les mécanismes sociaux des abandons de poursuite. Analyse expérimentale par simulation. *L'année sociologique*, 1964, Pp. 111-244. (PUF, 1965).
- 61 DAWSON, Richard, E., Simulation in the social sciences, Ch. 1 de GUETZKOW, Harold, *Simulation in Social Sciences : Readings* (Prentice Hall, 1962).
- 62 DAWSON, Richard, E., Glossary, Pp. 189-190 de GUETZKOW, Harold, *Simulation in Social Science* (Prentice-Hall, 1962).
- 63 DESCARTES, René, Oeuvres (publiées par Charles ADAM et Paul TANNERY, Leopold Cerf, 1897-1913).
- 64 DESHAYES, Jean, Les techniques des Grecs, Pp. 183-217 de DAUMAS, Maurice, *Histoire Générale des Techniques* (PUF, 1962), T.1.

- 65 DESTOUCHES, Jean-Louis, Sur la notion de modèle en microphysique, Pp. 52-57 de FREUDENTHAL, Hans, *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences* (D. Reidel, 1961).
- 66 DEVAUX, Pierre, Automates, automatisme, automation (PUF, 1967).
- 67 DOMINGO, Carlos, VARSAVSKY, Oscar A., A Mathematical, Numex, Model of Utopia, Pp. 253-267 de *Calcul et Formalisation dans les Sciences de l'Homme* (Editions du CNRS, 1968).
- 68 DOYON, André, LIAIGRE, Lucien Jacques Vaucanson, mécanicien de génie (PUF, 1966).
- 69 DUBUISSON, Marguerite, PAYEN, Jacques, PILISI, Jean, L'industrie textile, Pp. 214-250 de DAUMAS, Maurice, *Histoire générale des Techniques* (PUF, 1962) T.2.
- 70 DUCROCQ, A., Logique générale des systèmes et des effets (Dunod, 1960).
- 71 DUHAMEL, Alain, Les techniques de sondages restent sujettes à un certain nombre d'impondérables. *Le Monde*, 15 mars 1967.
- 72 DUHAMEL, Alain, Selon des études du C.F.R.O., 25 % des électeurs du Centre Démocrate ont préféré le PC à la 5e république au second tour à Paris. *Le Monde*, 16 mars 1967.

- 73 DUHAMEL, Alain, Des élections législatives aux élections cantonales. Les transferts de voix, notamment à gauche, varient selon les régions et les candidats du second tour. *Le Monde*, 2 septembre 1967.
- 74 DURKHEIM, Emile, Les règles de la méthode sociologique (15e éd., PUF, 1963).
- 75 DUSKEY, Harry D., A Language for Simulation, Pp. 13-26 de HOGGATT, A.C., et BALDERSTON, F.E., *Symposium on Simulation Models : Methodology and Applications to the Behavioral Sciences* (South-Western Publishing, 1963).
- 76 ELGOZY, Georges, Automation et humanisme (Calmann-Lévy, 1968).
- 77 FAURE, Robert, Les méthodes de simulation (Informations Scientifiques BULL, sans date).
- 78 FAURE, R., BOSS, J.-P., LE GARFF, A., La recherche opérationnelle (PUF, 1967).
- 79 FEIGENBAUM, Edward A., The Simulation of Verbal Learning Behavior, Pp. 297-309 de FEIGENBAUM, E.A., FELDMAN, J., *Computers and Thought* (Mc Graw Hill, 1963).
- 80 FELDMAN, Julian, Computer Simulation of Cognitive Processes, Ch. 5 de BORKO, Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).

- 81 FELDMAN, Julian, Simulation of Behavior in the Binary Choice Experiment, Pp. 329-346 de FEIGENBAUM, E.A., FELDMAN, J., *Computers and Thought* (Mc Graw Hill, 1963).
- 82 FISHER, Irving, Recherches mathématiques sur la théorie de la valeur et des prix (Giard et Brière, 1917).
- 83 FLEISHER, Aaron, The Uses of Simulation, Pp. 144-146 de BESHES, James M., *Computer Methods in the Analysis of Large-Scale Social Systems* (Joint Center for Urban Studies of the MIT and Harvard University, 1965).
- 84 FORTET, Robert, Simulations de systèmes à décisions. *Metra*, vol. 6 (1967), n° 3, Pp. 527-541.
- 85 FORTET, Robert, Remarques générales sur la simulation, Pp. 3-15 de *La simulation* (Société des amis du Centre d'Etudes Sociologiques, 1969).
- 86 FORTET, Robert, LE BOULANGER, Hubert, Elements pour une synthèse sur les systèmes à auto-organisation (SEMA, Direction Scientifique, Synthèse et Formation n° 28, juillet 1966).
- 87 FRAISSE, R., Les modèles et l'algèbre logique, Pp. 73-77 de FREUDENTHAL, Hans, *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences* (D. Reidel, 1961).

- 88 FREUDENTHAL, Hans, Models in Applied Probability, Pp. 78-88 de FREUDENTHAL, Hans, *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences* (D. Reidel, 1961).
- 89 GARDIN, J.C., A Typology of Computer Uses in Anthropology, Pp. 103-117 de HYMES, Dell, *The Use of Computers in Anthropology* (Mouton, 1965).
- 90 GARDNER, Martin, L'étonnante histoire des machines logiques (Dunod, 1964).
- 91 GARVIN, Paul L., Computer Processing of Cultural Data : Problems and Methods, Pp. 119-139 de HYMES, Dell, *The Use of Computers in Anthropology* (Mouton, 1965).
- 92 GILLE, Bertrand, Les XVe et XVIe siècles en occident, Pp. 1-139 de DAUMAS, Maurice, *Histoire générale des Techniques* (PUF, 1962), T.2.
- 93 GIRARD, Alain, Le choix du conjoint. Une enquête psychosociologique en France. (Publications de l'INED, cahier "Travaux et documents" n° 44, PUF, 1964).
- 94 GOBLOT, Edmond, Traité de logique (Armand Colin, 1941).
- 95 GREEN, Bert F., Digital Computers in Research. An Introduction for Behavioral and social Scientists (Mc Graw-Hill, 1963).

- 96 GREENBERGER, Martin, A New Methodology for Computer Simulation, Pp. 147-162 de BESHES, James M., *Computer Methods in the Analysis of Large-Scale Social Systems* (Joint Center for Urban Studies of the MIT and Harvard University, 1965).
- 97 GREMY, Jean-Paul, L'analyse multidimensionnelle de l'incertitude : une étape vers la simulation des phénomènes sociaux. *Metra*, vol. 5 (1966), n° 3, Pp. 453-470.
- 98 GREMY, Jean-Paul, Le dépouillement d'enquêtes sur ordinateur : quelques notions de base. *Bulletin de Psychologie*, T.XXI (1968), n° 9-11.
- 99 GREMY, Jean-Paul, L'analyse dynamique des phénomènes sociaux, Pp. 132-144 de *L'Informatique au service de l'homme*. (Université de Grenoble, 1970).
- 100 GREMY, Jean-Paul, Les techniques de simulation, Ch. 11 de BOUDON, R., *Mathématiques et sociologie*, (PUF, à paraître).
- 101 GREMY, Jean-Paul, L'emploi des techniques de simulation sur ordinateur en sociologie, in *Informatique et sciences sociales* (UNESCO, à paraître).
- 102 GRENIEWSKI, Henryk, Modèles économiques, Pp. 172-179 de *La cybernétique, Recherches internationales*, n° 29 (janvier-février, 1962).
- 103 GRENIEWSKI, Henryk, Cybernétique sans mathématiques (Gauthier-Villars, 1965).

- 104 GROENWOLD, H.J.,
The Model in Physics, Pp. 98-103 de
FREUDENTHAL, Hans, *The Concept and
the Role of the Model in Mathematics
and Natural and Social Sciences*
(D. Reidel, 1961).
- 105 GUETZKOW, Harold,
A Use of Simulation in the Study of
Inter-Nation Relations, Ch. 7 de
GUETZKOW Harold, *Simulation in Social
Sciences : Readings* (Prentice Hall,
1962).
- 106 GUETZKOW, Harold,
Simulation in International Relations,
Ch. 5 de *Simulation Models and Gaming*
(IBM, 1966).
- 107 GUILBAUD, G. Th.,
La Cybernétique (PUF, 1954).
- 108 GUITTET, Jack,
Physionomie et classement des techni-
ques de simulation. *Metra*, vol. 5 (1966),
n° 2, Pp. 191-200.
- 109 GULLAHORN, John T., GULLAHORN, Jeanne E.,
A Computer Model of Elementary Social
Behavior, Pp. 375-386 de FEIGENBAUM,
E.A., FELDMAN, J., *Computers and
Thought* (Mc Graw-Hill, 1963).
- 110 GULLAHORN, John T., GULLAHORN, Jeanne E.,
The Computer as a Tool for Theory
Development, Pp. 427-448 de HYMES, Dell,
The Use of Computers in Anthropology
(Mouton, 1965).
- 111 HÄGERSTRAND, Torsten,
A Monte-Carlo Approach to Diffusion,
Pp. 43-67 de *Simulation in Sociology.*
Archives Européennes de Sociologie
Tome VI, numéro 1 (Plon, 1965).

- 112 HANON, Bernard, The Use of Simulation in the Analysis of Business Systems, Pp. 4-7 de *Social Research with the Computer. The American Behavioral Scientist*, vol. 8 n° 9 (mai 1965).
- 113 HAYS, David G., Simulation : an Introduction for Anthropologists, Pp. 401-426 de HYMES, Dell, *The Use of Computers in Anthropology* (Mouton, 1965).
- 114 HOGLUND, K.M., PHIPPS, P.L., The Role of Digital Simulation in the Specification and Management of NTDS-Type Systems, Pp. 275-281 de *Digital Simulation in Operational Research* (The English University Press, 1967).
- 115 HOVLAND, Carl I., Computer Simulation of Thinking, Ch. 2 de GUETZKOW, Harold, *Simulation in Social Sciences : Readings* (Prentice Hall, 1962).
- 116 HUNT, Earl B., HOVLAND, Carl I., Programming a Model of Human Concept Formulation, Pp. 310-325 de FEIGENBAUM, E.A., FELDMAN, J., *Computers and Thought* (Mc Graw Hill, 1963).
- 117 IFIP-ICC. Vocabulary of Information Processing (North Holland Publishing, 1966).
- 118 ITZELSBERGER, G., Some Experiences with the Poker Test for Investigating Pseudo-Random Numbers, Pp. 64-68 de *Digital Simulation in Operational Research* (The English University Press, 1967).

- 119 JACQUARD, Albert, La reproduction humaine en régime malthusien. Un modèle de simulation par la méthode de Monte-Carlo. *Population*, 22e année, n° 5 (sept.-octobre 1967), Pp. 897-920.
- 120 JANVIER, Yves, Le marché du logement dans une agglomération. Un essai de simulation, Pp. 17-28 de *La Simulation* (Société des amis du Centre d'Etudes Sociologiques, 1969).
- 121 JOURDAN, Pierre, Une voie d'accès possible vers les modèles de simulation de marketing. *Metra*, vol. 8 (1969), n° 1, Pp. 29-42.
- 122 KAUFMANN, A., FAURE, R., LE GARFF, A., Les jeux d'entreprises (PUF, 1964).
- 123 KEMENY, John G., SNELL, J. Laurie, Mathematical Models in the Social Sciences (Blaisdell, 1962).
- 124 KEMENY, J.G., SNELL, J.L., THOMSON, G.L., Algèbre moderne et activités humaines (Dunod, 1960).
- 125 KITZ, B., Problems Encountered in the Programming of Games and Simulations, Pp. 125-130 de *Digital Simulation in Operational Research* (The English University Press, 1967).
- 126 KLEENE, S.C., Representation of Events in Nerve Nets and Finite Automata, Pp. 3-41 de SHANNON, C.E., Mc CARTHY, J.M., *Automata Studies*, Annals of Mathematic Studies n° 34, Princeton University Press, 1956.

- 127 KOOPMAN, Bernard O., BERGER, Howard,
Use and Misuse of Simulation, Pp. 19-25
de *Digital Simulation in Operational
Research* (The English University Press,
1967).
- 128 KORN, G.A.,
Random Process Simulation and
Measurements (Mc Graw-Hill, 1966).
- 129 KRASNOW, H.S.,
Dynamic Presentation in Discrete
Interaction Simulation Languages,
Pp. 77-92 de *Digital Simulation in
Operational Research* (The English
University Press, 1967).
- 130 KUNTZMANN, J.,
Les modèles et leurs règles d'emploi,
Pp. 2-5 de *L'Informatique au service
de l'Homme* (Université de Grenoble,
1970).
- 131 LALANDE, André,
Les théories de l'induction et de
l'expérimentation (Boivin, 1929).
- 132 LALANDE, André,
Vocabulaire technique et critique
de la philosophie (PUF, 1951).
- 133 LAZARSFELD, Paul F., HENRY, Neil W.,
Readings in Mathematical Social Science
(Science Research Associates, 1966).
- 134 LE BOULANGER, Hubert,
Un cadre éventuel pour une recherche sur
les modèles de comportement (SEMA,
Direction Scientifique, Note de travail
n° 60, 1967).

- 135 LE BOULANGER, Hubert, Quelques problèmes posés par l'expérimentation des modèles de simulation, Pp. 29-36 de *La simulation*. (Société des amis du Centre d'Etudes Sociologiques, 1969).
- 136 LE LIONNAIS, François, La cybernétique, Pp. 101-121 de TATON, René, *Histoire générale des Sciences*, (PUF, 1957-1964), tome 3, vol. 2.
- 137 LENOBLE, Robert, Mersenne ou la naissance du mécanisme (Vrin, 1943).
- 138 LENOBLE, Robert, Origines de la pensée scientifique moderne, Pp. 369-534 de *Histoire de la Science, Encyclopédie de la pléiade*, (Gallimard, 1957).
- 139 LICHNEROWICZ, André, Remarques sur les mathématiques et la réalité, Pp. 474-485 de *Logique et connaissance scientifique, Encyclopédie de la pléiade* (Gallimard, 1967).
- 140 LITRE, Emile, Dictionnaire de la langue française, (JJ. Pauvert, Gallimard Hachette, 1956-1958).
- 141 LULLE, Raymond, Opera (Minerva, Francfort/Main, 1965).
- 142 MAITRE, Jacques, Réflexions d'un néophyte devant la simulation, Pp. 37-45 de *La Simulation*. (Société des amis du Centre d'Etudes Sociologiques, 1969).
- 143 MARTIN, Roger, Logique contemporaine et formalisation (PUF, 1964).

- 144 Mc CLOSEY, J.F., TREFETHEN, F.N.,
Introduction à la recherche opérationnelle (DUNOD, 1958).
- 145 Mc GINNIS, Robert,
Mathematical Foundations for Social Analysis (Bobbs-Merril, 1965).
- 146 Mc LEOD, John,
Simulation. The Modeling of Ideas and Systems with Computers (Mc Graw-Hill, 1968).
- 147 MESNAGE, Pierre,
La construction horlogère, Pp. 289-310 de DAUMAS, Maurice, *Histoire générale des techniques* (PUF, 1962) T.2.
- 148 MINSKY, Marvin,
Steps toward Artificial Intelligence, Pp. 406-450 de FEIGENBAUM, E.A., FELDMAN, J., *Computers and Thought* (Mc Graw-Hill, 1963).
- 149 MORICE, E., CHARTIER, F.,
Méthode statistique - 2^{ème} partie : analyse statistique (Imprimerie Nationale, 1954).
- 150 NAYLOR, T.H., BALINTFY, J.L., BURDICK, D.S., CHU, Kong,
Computer Simulation Techniques (Wiley, 1968).
- 151 NEWELL Allen,
On the Analysis of Human Problem Solving Protocols, Pp. 145-185 de *Calcul et Formalisation dans les Sciences de l'homme* (Editions du CNRS, 1968).

- 152 NEWELL, Allen, SIMON, Herbert A.,
Computers in Psychology, Ch. 7 de LUCE,
R. Duncan, BUSH, Robert R., GALANTER,
Eugene, *Handbook of Mathematical
Psychology*, (Wiley, 1963).
- 153 NEWELL, Allen, SIMON, Herbert A.,
GPS, a Program that Simulates Human
Thought, Pp. 279-296 de FEIGENBAUM, E.A.,
FELDMAN, J., *Computers and Thought*
(Mc Graw-Hill, 1963).
- 154 ORCUTT, Guy H.,
Simulation of Economic Systems, Ch. 8
de GUETZKOW Harold, *Simulation in
Social Sciences : Readings* (Prentice
Hall, 1962).
- 155 ORCUTT, Guy H.,
Simulation of Economic Systems : Model
Description and Solution, Ch. 4 de
Simulation Models and Gaming (IBM,
1966).
- 156 ORCUTT, Guy H., GREENBERGER, Martin, KORBEL, John, RIVLIN, Alice M.,
Micro-analysis of Socioeconomic Systems.
A simulation study (Harper, 1961).
- 157 PADIOLEAU, Jean-G.,
Note sur les "Simulations en sociologie
politique". *Revue française de sociologie*,
T. 10, vol. 2 (avril-juin 1969),
Pp. 201-210.
- 158 PAGE, E.S.,
The Generation of Pseudo-Random
Numbers, Pp. 55-63 de *Digital Simulation
in Operational Research* (The English
University Press, 1967).

- 159 PAPERT, Seymour, Remarques sur la finalité, Pp. 841-861
de *Logique et connaissance scientifique*,
Encyclopédie de la pléiade (Gallimard,
1967).
- 160 PÓLYA, Georges, Les mathématiques et le raisonnement
"plausible" (Gauthier-Villars, 1958).
- 161 PONTE, Maurice, BRAILLARD, Pierre,
 L'informatique (Seuil, 1969).
- 162 POOL, Ithiel de Sola, ABELSON, Robert P.,
 The Simulmatics Project, Ch. 6 de
GUETZKOW, Harold, *Simulation in
Social Sciences : Readings* (Prentice
Hall, 1962).
- 163 POOL, Ithiel de Sola, ABELSON, Robert P., POPKIN, Samuel,
 Candidates, Issues, and Strategies.
A Computer Simulation of the 1960
and 1964 Presidential Elections
(MIT, 1965).
- 164 POOL, Ithiel de Sola, KESSLER, Alan,
 The Kaiser, the Tsar, and the Computer :
Information Processing in a Crisis,
Pp. 31-38 de *Social Research with the
Computer. The American Behavioral
Scientist*, vol. 8 n° 9 (mai 1965).
- 165 POPKIN, Samuel B., A Model of a Communication System,
Pp. 8-11 de *Social Research with the
Computer. The American Behavioral
Scientist*, vol. 8 n° 9 (mai 1965).

- 166 PRICE, Derek J. de Solla, An Ancient Greek Computer. *Scientific American*, vol. 200, n° 6 (juin 1959), Pp. 60-67.
- 167 RAINIO, Kullervo, Social Interaction as a Stochastic Learning Process, Pp. 68-88 de *Simulation in Sociology. Archives Européennes de Sociologie*, Tome VI, n° 1 (Plon, 1965).
- 168 RAINIO, Kullervo, Simulation of the Stochastic Learning Process in Group Problem Solving, Pp. 61-77 de *Computers in Psychological Research* (Paris, Gauthier-Villars, 1966).
- 169 RAINIO, Kullervo, Simulation de l'interaction sociale en termes d'un processus d'apprentissage stochastique, Pp. 283-296 de *Calcul et Formalisation dans les Sciences de l'homme* (Editions du CNRS, 1968).
- 170 RAYMOND, F.H., Calcul analogique. Principes et contribution à une théorie générale (Société de la Revue d'Optique, sans date).
- 171 RENARD, Bruno, Le calcul électronique (PUF, 1966).
- 172 RENARD, Bruno, RENAULT, Jean-Paul, La simulation et les calculateurs, *Metra*, vol. V, n° 2 (juin 1966), Pp. 201-215.
- 173 RICH, R.P., Simulation as an Aid in Model Building, Ch. 13 de GUETZKOW, Harold, *Simulation in Social Sciences : Readings* (Prentice Hall, 1962).

- 174 RIVETT, B.H.P., Why simulate ? Pp. 3-8 de *Digital Simulation in Operational Research* (The English University Press, 1967).
- 175 ROESTHLISBERGER, F.J., DICKSON, W.J., Management and the Worker (Harvard Univ. Press., 1947).
- 176 ROME, Sydney C., ROME, Beatrice K., Computer Simulation toward a Theory of Large Organizations, Ch. 11 de BORKO, Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).
- 177 ROSENTHAL, Howard, Election Simulation, Pp. 21-42 de *Simulation in Sociology. Archives Européennes de Sociologie*, Tome VI, n° 1 (Plon, 1965).
- 178 ROSSI, Paolo, Les arts mécaniques et la science nouvelle, Pp. 221-236 de *Trois études sur la science, Archives Européennes de Sociologie*, Tome 4 (1963), n° 2.
- 179 ROSSI, Philippe, Quelques aspects de la théorie des modèles de simulation, Pp. 11-16 de *Digital Simulation in Operational Research* (The English University Press, 1967).
- 180 ROUANET, H., Les modèles stochastiques d'apprentissage (Gauthier-Villars, 1967).
- 181 RUYER, Raymond, La cybernétique et l'origine de l'information (Flammarion, 1954).

- 182 SABLIERE, Jean, De l'automate à l'automatisation (Gauthier-Villars, 1966).
- 183 SCHELLHORN, J.P., Simulation d'une équipe au travail (aspects affectifs), Pp. 79-104 de *Computers in Psychological Research* (Paris, Gauthier-Villars, 1966).
- 184 SIEGEL, Sidney, Non-parametric Statistics for the Behavioral Sciences (Mc Graw-Hill, 1956).
- 185 SIMMONS, Robert F., SYNTHEX : Toward Computer Synthesis of Human Language Behavior, Ch. 16 de BORKO, Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).
- 186 SPROWLS, R. Clay, Business simulation, Ch. 23 de BORKO, Harold, *Computer Applications in the Behavioral Sciences* (Prentice-Hall, 1962).
- 187 STEFFLRE, Volney, Simulation of People's Behavior Toward New Objects and Events, Pp. 12-15 de *Social Research with the Computer. The American Behavioral Scientist*, vol. 8 n° 9 (mai 1965).
- 188 SUPPES, Patrick, A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences, Pp. 163-177 de FREUDENTHAL, Hans, *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences* (D. Reidel, 1961).

- 189 SUTTER, Jean, Le choix du conjoint, Pp. 317-328 de *La Progénèse (XXII) Centre International de l'Enfance. Travaux et Documents VIII* (Paris, Masson, 1955).
- 190 SUTTER, Jean, L'apport de la génétique de population à l'analyse de l'évolution sociale. Pp. 53-66 de *CNRS, Colloque National de Démographie* (1961).
- 191 TOCHER, K.D., The art of simulation (The English Universities Press, 1963).
- 192 TORCHET, E., Remarques concernant les conclusions de Mr. CRUON, Pp. 387-388 de *Digital Simulation in Operational Research* (The English University Press, 1967).
- 193 TUSTIN, A., The Mechanism of Economic Systems (Haward Univ. Press, 1953).
- 194 UBBINK, J.B., Model, Description, and Knowledge, Pp. 178-194 de FREUDENTHAL, Hans, *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences* (D. Reidel, 1961).
- 195 ULLMO, Jean, La pensée scientifique moderne (Flammarion, 1969).
- 196 UTTLEY, Albert M., Conditional Probability Machines and Conditioned Reflexes, Pp. 253-275 de SHANNON, C.E., Mc CARTHY, J.M., *Automata Studies*, (Annals of Mathematic Studies n° 34, Princeton University Press, 1956).

- 197 VIGIER, Jean-Pierre, Théorie des niveaux et dialectique de la nature. (Cahiers du CERM, sans date).
- 198 VON NEUMANN, John, The Computer and the brain (Yale Univ. Press, 1958).
- 199 WIENER, Norbert, Cybernetics, or Control and Communication in the Animal and the Machine (Hermann, ASI n° 1053, 1948).
- 200 WIEFNER, Norbert, Cybernétique et Société (Union Générale d'Éditions, 1962).
- 201 WILLE, H., Project Simulation by Means of Network Technique, Pp. 319-322 de *Digital Simulation in Operational Research* (The English University Press, 1967).
- 202 ZELDITCH, Morris Jr., EVAN, William M., Simulated Bureaucraties : a Methodological Analysis, Ch. 4 de GUETZKOW, Harold, *Simulation in Social Sciences : Readings* (Prentice Hall, 1962).
- 203 Encyclopaedia Britannica (William Benton, 1968).
- 204 Essai sur la formalisation dans les sciences du comportement (Ronéotypé, Secrétariat Permanent du C.A.S.D.N., janvier 1962).

- 205 Glossaire de *La Cybernétique*, *Recherches Internationales*, n° 29
(janvier-février, 1962), Pp. 195-205.
- 206 Où en est le marché de l'informatique ?
L'Expansion, n° 21 (juillet-août, 1969)
Pp. 96-105.
- 207 Sociologie et psychologie sociale en France (1945-1965), Bibliographie
établie par le Centre d'Etudes Sociologiques (CNRS) avec une intro-
duction de Jean-Daniel RAYNAUD (Association pour la Diffusion de la
Pensée Française, 1966).