Implications of the Electronic Computer for Economics

von Neumann Computer Anniversary Conference

by Lawrence R. Klein*

1. A Reminiscent Note

During and just after World War II, on his trips to Los Alamos, John von Neumann occasionally stopped in Chicago, between trains, and, among other places, visited the Cowles Commission. Those were memorable occasions for those of us who were engaged in econometric research, for the Theory of Games was just published and of immediate, intense interest. The discussions and informal presentations by John von Neumann ranged far beyond game theory. He was especially helpful to us in problems of computation for the large scale problems in econometric analysis of simultaneous equation systems that preoccupied our thoughts. Our honored guest was extremely insightful in showing us how to solve large (or moderately large) systems of nonlinear simultaneous equations--our principal computing problem at that time--by iterative methods using conventional electric desk calculations. At each iterative step of this procedure, it was necessary to solve a linear equation system, and he remarked, on one occasion, that if one got no more accuracy than we expected to get with the desk machines in the (10 x 10) case, the airplane would not have been able to perform as it did at that time. We were not discouraged by his remark, given the kind of accuracy that we are accustomed to in econometrics, but we did not realize the wonders that were then being performed by, or coming in the next generation from, electronic equipment that was being used and perfected by John von Neumann.

The first large scale uses of computers in economic research began soon

* It is with gratitude that I acknowledge critical commentary, often in disagreement, from F.G. Adams, P.J. Dhrymes, A.J. Karchere, R.S. Preston, and R. Summers.

after von Neumann's contributions in the 1950's, and was used sparingly by the end of the decade in a wide variety of places for economic research, but it took almost twenty years before econometricians were to obtain access to computers that could handle with relative ease large examples of the problem on which we were consulting with John von Neumann. Having come through a full decade or more of intensive large-scale computer applications in economics, it is fitting on this occasion to pause for reflection over the meaning of the development.

2. A Brief Overview of Computation in Quantitative Economics

Until the mid 1950's a great deal of calculation in economics was done by some form of mechanical desk machine, either cranked by hand or driven by electric motor. The principal departure was in mass tabulation of sample survey or census type data, where punched cards in electro mechanical machines of the IBM type were common. Economic research centers depended upon a corps of human computer assistants to process and analyze data. Much numerical work was done, some of it highly creative associated with the names of Tinbergen, Kuznets, and Leontief, to mention only a few of the leaders, but it was all costly in labor power and had to be carefully designed in order to be kept within limits of human capability. Linear methods flourished in econometrics because large scale treatment of nonlinear problems seemed to be too demanding and impractical.

In some cases analog computers were used. These were either electronic or mechanical, but they were not very satisfactory. In our first econometric model building attempts at the Cowles Commission we looked seriously at electronic analog machines used in psychometrics for the possibility of moment calculation in regression, but ultimately dropped the idea in favor of electric desk machines operated by human power. This was the predominant mode of calculation in economics for many years. It caused us to keep models as simple as possible by transforming systems to linear form in small segments. The quantitative research of the day reflected the computational bottleneck, and there was always a substantial discrepancy between the pace of theoretical and empirical advances in economics, in favor of the former.

Richard Stone and his associates in Cambridge were early users of the electronic computer in economics. They used the Cambridge University computer for statistical demand analysis, while Leontief used the Harvard University computer for input-output analysis. The business cycle and national income studies at the National Bureau of Economic Research were done at the start of the 1950's by the older methods, but the corporate bond study was using punched cards and tabulators. These card methods were the usual procedures at the Survey Research Center and similar bodies analyzing samples of human populations of 1000 or more cases. The Klein-Goldberger Model at the University of Michigan was initiated in the traditional way, but went over to punched card methods and combinations of mechanical and electronic units for purposes of statistical estimation. At this stage, computers were first being used for system estimation, but not for system application in simulation studies. Some notable electronic analog simulations of very small economic systems were made, but realistic systems of ten or more dynamic equations were not analyzed by electronic methods.

Towards the end of the decade of the 50's and the beginning of the 60's an enormous change took place. In retrospect, it appears to have been revolutionary. To someone who has worked in quantitative economics in both eras the change has been truly remarkable.

Engineering and natural sciences harnessed the computer for their research interests before economics did, but the generation of researchers in econometric model building, survey analysis, input-output analysis, time-series analysis, and many other fields of economics caught on to the new tool quickly. By the middle of the 1960's the older procedures had all but vanished, and a whole new generation of economists was being trained in a new way for quantitative research with the electronic computer.

3. Some Implications of the Computer for Economics

It is not unusual to find the economics department in a typical large diversified university to be one of the largest customers at the computer center. Engineering and Physics are usually larger users, but economics comes close behind. It is not, of course, uniform. A high use rate occurs at a university where the economics faculty is strongly quantitative in its research interests.

Economists use the computer for mass tabulations of sample surveys, pedagogy, estimation of economic relationships, simulation of estimated economic systems, index construction, distribution sampling, and other kinds of problems. It is not only in econometrics or economic statistics where the computer is used. A wide range of economic specialists use the computer to some extent, but the big users are obviously in econometrics.

In many respects this new technology has changed the face of economics. Programming knowledge is essential for the young and serious student. Mature economists have in many cases become adept at programming or at least, in interpreting programmed output. Entering students in post graduate economics should all learn programming, unless they have gained this skill during undergraduate or secondary school days. It is very much like the situation of requiring beginning post graduate students to know mathematics. There are now two tools that should be readily available to most students. Many universities have changed the standard language requirements from knowledge of two foreign languages to knowledge of one foreign language and knowledge of some form of higher mathematics. This recognizes "Mathematics as a language", but it is my opinion that an option or requirement should now be set up for programming. Mathematics and programming might replace the older requirement of two languages. Fortran, Algol, and other programming languages are, indeed, like foreign languages. The programming languages have an edge in applicability and productivity for most students.

The introduction of programming as a prerequisite for postgraduate study should do much towards producing a sympathetic attitude towards quantitative methods in economics. It can bring a wide variety of economists together through the use of a common technology.

The heavy use of the computer in modern economics has its positive (good) and negative (bad) sides. Broadly speaking, it cannot but help being good in an important way, yet it is useful to list and comment on the positive aspects.

(i) Encouragement of exploration and experimentation.

(ii) Capability of solving large, complicated problems

(iii) Capability of handling large bodies of data or information

(iv) Provision of display of procedures and composition of solutions to problems.

(i) During the precomputer days, we had to ration our time spent computing and could consider only a limited range of hypotheses or could estimate only a limited number of parameter values. Now it is possible to consider many alternative hypotheses. The most typical, but not the only case is the estimation of economic relationships. With the computer it is possible to explore the linear case, logarithmic case, other nonlinear cases, wide scope of variables, different lag structures, homogeneous or inhomogeneous functions, subsections of statistical samples, etc. It may not be possible to discriminate fully among all the possible hypotheses considered, yet it is better to know what the alternative possibilities are than to confine attention and reliance to one specific form of relationship. It is not uncommon for economists to look at fifty plausible specifications of a relationship instead of one, two, or three as in pre-computer days.

(ii) The economist used to assume only simple relationships, usually only linear, at least after transformation, with a small number of variables, and without much splitting of the sample. Now, relationships need not be approximated in the simplest form. It is feasible to go directly to the estimation of nonlinear relationships in their full complexity. There is probably a side benefit in that the stochastic structure can be kept simpler. Relationships can now be made highly multivariate. Except for shortages of degrees of freedom, the economist now finds it possible to investigate the effects of many more factors.

Not only is it possible to treat the single relationship in the full realism of nonlinearity and other complications, but we can also do this for systems of simultaneous equations, which are obviously very important in economics. We are now solving estimation problems that were formulated twenty or more years ago, but left dormant because of our inability to solve the associated equation systems. In the interim, many compromises and approximate simplifications were introduced in order to make the problems tractable for interim solutions. Now the procedure is quite different. We formulate the problem in a direct way, not avoiding complications that will arise in the estimation equations. We then use the computer in a straightforward way by iteration or search techniques in the parameter space to obtain a solution. This is a new kind of world for quantitative economics.

In "Monte Carlo" or experimental sampling studies, it is possible to use the computer for throwing light on problems that have not yielded to attack by analytic methods. As evidence accumulates through experience, approximation formulas, or analytic breakthroughs it has become clear that experimental results generated by computer studies are basically sound and suggestive of correct interpretations of the problems at hand. In addition, the computer can be used to indicate how analytic solutions to difficult problems might be found, or it might help find counter examples for unproved conjectures. It can be used effectively as another problem-solving tool.

(iii) Sample surveys of human populations for the study of socio-economic problems have a time-honored status in economics, dating from Engel's nineteenth century researches and carrying right on through many family budget investigations in this century during the pre-computer era. At an early stage, however, survey samples of one to twenty thousand or more separate cases were tabulated on electro-mechanical machines with punched card input, as mentioned earlier, this was one of the first stage applications of computer methods. Nevertheless, cross tabulations and deeper analysis of variance or regression studies with electro-mechanical equipment were formerly done only in limited scope because of the heavy computation burden. Now, electronic computers are as common in large scale survey as in census tabulation. Many more research students now can access large data files of individual persons, families, firms, or traits. It is not uncommon for a few theses in a fair size graduate department of economics to be based on an investigation of samples of a few thousand basic cases. The research students carrying out these investigations are almost sure to be using electronic computers for the processing and analysis of their data. What was once the province of the mature research worker in a team effort is now made available through computer technology to the single handed research worker who is just beginning a career.

Some economic concepts, especially those dealing with capital theory, involve large scale tedious calculations which were never feasible except in exploratory studies before computers were used. In the present era, economic research is becoming increasingly sophisticated in translating intricate concepts into reality. Capital rental variables, capital stock measures, and capital inputs flows are primary examples of computer based data collections that are now becoming increasingly common. Also data based on sophisticated concepts or in mass collections are sure of being used because research workers have adequate facilities. This, in turn, stimulates the whole data collection and preparation process.

(iv) In iteration and search procedures for completing the algorithms associated with complicated problems, frequently of the nonlinear type, it is possible to use the computer to print out successive interim results at each stage of the iterations or search steps. This enables one to see how the complicated process works. In numerical simulation, the whole dynamic path of the projected economy can be plotted by computer in a visual display. This can be done for many alternative cases, so that it is easy to see the results of a range of different input combinations.

Many revealing summary statistics or critical ratios of a problem and its solution can be quickly tabulated. This aids in interpretation from different points of view. Any number of intermediate results of a long problem can also be displayed so that it is easy to see the composition of a final solution.

All these procedures were previously available before the computer age, but it was not often possible to take advantage of them because the work burden was usually overwhelming. It was customarily a great effort to work through to a final result, without stopping to look at interim calculations, and at the end of a problem, a tired computing assistant often did not want to make large scale calculations of different summary statistics.

But all is not necessarily for the best in significant technological changes. There might be a negative side to the computer age in economic research. The new generation of trained economists are not interested in doing some of the things their teachers used to do and are acquiring some lazy or otherwise bad research habits.

Without asserting that the computer must have bad consequences, it is useful to point out some potentially dangerous situations, practices, or side effects that may develop:

(i) Too little prior thinking about basic economic structure.

(ii) Wasteful use of resources.

(iii) Imposition of machine barriers between the investigator and data, including an increasing reliance on secondary and tertiary data sources.

(iv) Creation of a false image of accuracy.

(v) De-humanization of economic analysis.

(vi) Violation of personal confidentiality.

(i) In many cases there is too little thinking about problem structure. It is so easy to make experimental calculations, that one might become too much of an empiricist in outlook and spend less time in problem analysis. It has become easy for economists to search for high correlations. An economic model has a basic <u>structure</u> and a derived or <u>reduced form</u>. These are commonly used technical expressions in econometrics. It has become easy for untrained econometricians to search for close fitting reduced forms and use them with the same confidence that should be reserved for a solidly based structural system.

In 1971, the U.S. government put forward an extreme economic calculation based on a "close fitting" reduced-form equation system that was quite unacceptable in terms of careful analysis of economic structure. This piece of research was made possible by accessing data, software programs, and a big computer by the most modern time sharing devices. The country would have been much better off if this set of calculations had been more laborious. There is a great danger that if soft and hardware are made readily available even to the untrained economic analyst, we are going to see a much greater output of poor results in an area where ability and much more attention to the structure of the economy is required . Procedures that challenge the most thoughtful minds are being turned over for very mechanical application by people who do not fully understand them. In some cases, they will cause evident harm, as did the official economic forecasts of 1971.

(ii) Not only will accessibility to sophisticated devices deter some people from thinking carefully about their research problems, but it may also lead to wasteful uses of expensive resources. Today's students tend to look upon computer time as a "free good" and use it, beyond a pure learning period to do frivolous things. More seriously, they might waste this expensive resource in doing things in a round-about way because it is so fast and it is easy for them to throw away or ignore irrelevant calculations. For example, if a graph-plotting routine is embedded at some stage of a regression program, students often go through the whole process of making regression calculations that they never intend to use, simply to get a machine plot of a graph. They compute many wasteful regressions with highly implausible combinations of variables because it is virtually costless to them to program many alternatives and it is difficult to police their computer activities at every stage. Step-wise regression programs are particularly stimulants to wasteful use of resources and encouragement of unstructured empiricism.

(iii) The computer stands as an imposing electronic machine between the quantitative economist or other social scientist, and his data base. In the early days of quantitative analysis, especially in econometrics, there was an intensive use of graphical analysis and careful examination of the whole configuration of sample data points. This step is often by passed in favor of experimental calculation with the data. Both approaches yield much information, but it is a problem to train modern students to do much graphical analysis either before or after calculation.

Within the machine,data are averaged, formed into ratios, put in logarithmic form and generally transformed. It is often difficult to trace errors, revisions, and unusual patterns in this process, and many users have difficulty in retrieving the original data for re-examination, once they have been transformed.

Some transformations, such as seasonal adjustment, are iterative in a highly mechanical way. It has been my experience that uncareful, automatic application of such computer programs produces seasonally adjusted data that make little economic sense. Simpler hand methods would have been more transparent, although painfully slow. The computer must be used for large scale data transformation and manipulation, but students must be taught to analyze the final series with great care and attention to intuitive plausibility in configuration.

A point of research procedure is also worth considering. Good scholarship in statistical analysis requires documentation of data sources, and the most careful analysis usually goes back to primary data sources. Secondary sources are all right, provided one can have faith in the agency or person collecting the primary data and be fully informed about its basic characteristics or peculiarities. A new fashion is now developing. Data banks, which compile large masses of secondary source data, are serving as a base for quantitative economic investigation. This means that the fully computerized scholar will cite a tertiary source for documentation, will be less fully informed about special aspects of the data and will have to take the accuracy of one more set of intervening steps on faith.

(iv) Large numbers of digits or decimal points arranged on bulky folds of computer paper may convey a false image of precision. The computer is just a tool--one that works faster, better, and bigger than previous tools--but it cannot, by its pure existence, improve upon the reliability of propositions made in quantitative economic analysis. The primary gains will come from better specifications of economic relationships, better data sources, larger more frequent samples, improved statistical methods. The computer tool can be used in a helpful research way to shape or make possible each of these potential gains, but the computer tool could also be used on a large scale without the other improvements in analysis. When used in the latter way, the computer will produce results that are not more accurate, but they may appear to some to be more accurate than they really are. The role of the computer is indeed great, but scholars may perform a service by playing down this role in the final presentation of findings, and final numerical results for presentation should probably be rounded to no more digits the investigator is prepared to stand behind. Also, the power of the computer should be harnessed for the more complicated tasks of computing relevant error statistics on tolerance or confidence regions. If these were properly evaluated and presented together with "point" values in quantitative economics, the effect would indeed be sobering. The tendency is for economists to talk about economic calculation largely in terms of average values with inadequate attention paid to measures of dispersion. It is the combination of definite average values and the use of expensive computer hardware that gives the illusion of exorbitant accuracy.

(v) Economics is a behavioral science, dealing with one particular aspect of human behavior, namely economic decision making. Many trends in thought have led economists away from the basic human element. Macroeconomics does this, in a sense, as compared with microeconomics. It is important to remain in touch with the human aspect of economics at all times, and the computer may be just another obstacle in achieving that contact. Actually, the computer could be of enormous benefit in retaining the human touch by enabling researchers to deal with large masses of microeconomic information, but this potential may not be properly exploited. Instead, a cold objective calculus that is not truly representative of the human side of economic behavior could be produced, in heavy reliance on computer analysis of economic information. Students in a university, patients in a hospital, claimants on welfare rolls, customers in a business are all human economic units who resent being identified on cards, tape, drum, or disc by serial number. It is not entirely the fault of the computer that this is happening, but the impersonal dehumanizing of economic life and analysis of that life is being made possible by the computer.

(vi) In a vein similar to the dehumanization of economics we note that invasions of privacy or violation of personal confidentiality have long occurred in economic analysis, but are made possible on a much larger scale by the computer. From the earlier days of family budget investigations in the 19th century, personal reports on individual economic behavior have been made publicly available and subject to the scrutiny of research scholars and other analysts. Most people in applied economics use and rely on such personal information, but it is always subject to considerable care for protection of identity and frequently with guilt feelings on the part of the investigator. This was a growing branch of economic analysis before the days of the electronic computer, but now large scale storage, accessing, and cross reference of personal files is made possible by the computer. It is possible to correlate an individual's bank records, credit standing, tax returns, and other files with his replies to enumerators of social surveys. While much valuable insight is gained about human economic behavior from computer analysis of these information sources, care must be taken in the proliferation and elaboration of such a study.

The negative or potentially harmful aspects of the computer are not illusory; they are real but not all of the same force or importance. If the tool is properly used there is no reason why any harmful effects need occur; nevertheless there are frequent examples of misuse. These come about through misunderstanding and the fact that the new instrument is, actually, so powerful.

4. Future Developments

Both the good and bad side of computer analysis in economics will expand in the future as the tool becomes used on a larger scale. There is no doubt, however, that the only direction for usage is upwards, both in absolute amount and as a percentage of total economic research time. Regardless of the judgment of good or bad, what are some of the developments discernible on the horizon?

Although the most intense developments in quantitative economics have been centered in the United States over the past few decades, and much of it based on computer analysis, the situation is changing, and more foreign centers are becoming involved in quantitative analysis. Naturally, the first computers were made available on a large scale in the United States, although there were notable exceptions abroad, but the distribution of facilities has undergone rapid change. More quantitative economists from all over the world are being trained in American type use of the computer, and more facilities are becoming available abroad. This is true in advanced market economies, in socialist economies, and in developing economies. In the near future, the computer will be an accepted tool nearly everywhere in the world, and few problems in a given locale will remain dormant for want of computer facilities. The tendency to use more time shared facilities with relatively inexpensive terminal equipment linked to a central installation is important because it will eventually put every poor country in touch with a computer--either its own or that of a richer neighbor. International usage for economic research either by time-shared or batch process mode is already a fact of life.

The capital cost of computing equipment has been very large. At present it is a good for institutions to buy, not for individuals. The economics of computing are, however, undergoing rapid change. Unit costs of calculations are falling even though central installations remain expensive. Terminal equipment for shared facilities is accessible for individuals, and this aspect will grow. It is within the realm of possibility that complete units will eventually become small and relatively cheap. One thing is clear, however; unit costs will continue to fall, and the scope of use in economic research (all research for that matter) will continue to broaden. The new tool should become more universal and accepted as conventional, not only among nations but also among individuals.

The first electronic computational procedures were stepwise, but they are becoming increasingly automated. Data collection, data storage in banks, data retrieval, statistical estimation, simulation can all be linked in one sequential operation on an automated basis. Programming for economic analysis is proceeding ~~in~~

in this direction and what has already been accomplished will be refined.

The problems that will be capable of solution will become ever more complicated and larger. It is commonplace now to do things that were highly intractable two or three decades ago. Economic analysis need no longer be linear, non stochastic, with fixed parameters. Many of these problems have been overcome. All nonlinearities have not been treated, nor have all sources or kinds of stochastic variation and parameter variation been tackled. Some problems still defy the large scale computer, but their scope is steadily being broken down. Continuing progress seems assured along these lines, with the consequence that economic analysis is likely to become more realistic because many of the complications that have been introduced have been in the interests of realism.

Guy Orcutt has criticized for many years the prevailing tendencies in economic modeling for being too macroscopic and too simple in parametric structure. He has argued for microanalytic simulation of economic processes and their extension to related fields of demographic and other social studies. In place of average formulas with fixed parameters, he recommends use of sample tables of distributions from which probability of decisions can be calculated. His approach would depend heavily on large scale computer power. It is not only computer facilities, however, that form obstacles to the realization of his goal. Stability, reliability, availability and underlying economic theory are also not ready for his grand scheme, yet the future development of the computer is likely to help bring him closer to his objective.

In a related fashion, socialist planners may eventually be able to simulate entire markets of microeconomic bargaining agents and use computer control rather than natural market forces for guiding the economy. In earlier discussions of the theory of socialist planning, the impracticability, if not the impossibility, of achieving a rational allocation of resources was based on the observation that a prohibitively large system of simultaneous equations of general economic equilibrium would have to be solved. The market system has been introduced on a varied and limited scale in different socialist economies to deal with this problem in a practical way, but mathematicians, numerical analysts, and economists are trying to develop computer methods for dealing with the problem in a way that previously seemed out of reach. If it is within reach, it is only because the computer makes it possible for economists to contemplate an empirical implementation of the equation system of general equilibrium. This is, as yet, only a possibility, but it is probably the largest and most significant problem on the future horizon of applied economics.

THE INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY 08540
Telephone-609-924-4400
(ext. 203)

Conference on

THE COMPUTER and the DEVELOPMENT OF SCIENCE AND LEARNING

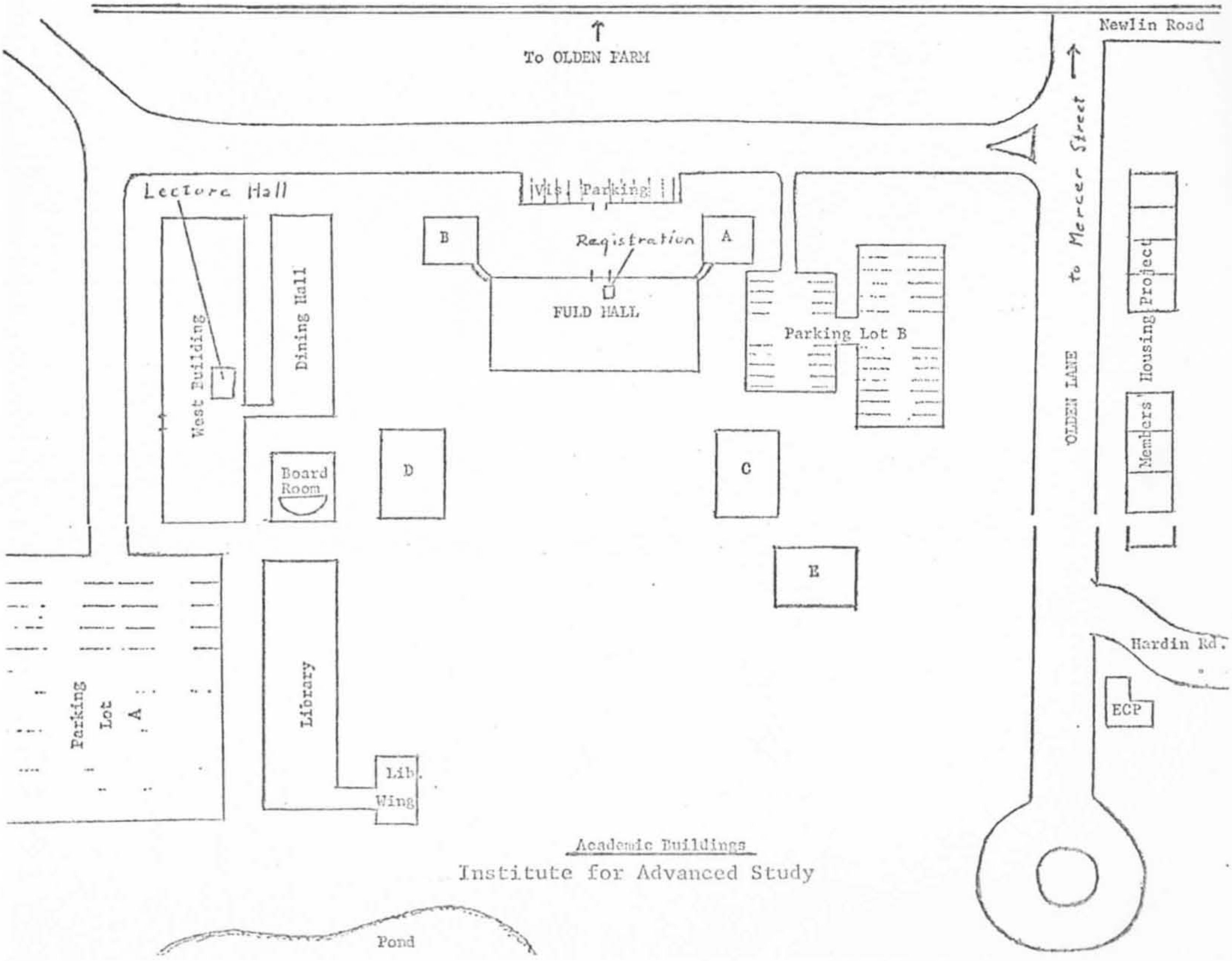
June 6 - 8, 1972

Information to Participants:

Registration table in Fuld Hall Tuesday, June 6, 5:00 p.m. to 8:15 p.m.

Evening programs and all meals in Board Room (reached through Dining Hall)

Morning and afternoon sessions in Lecture Hall, first floor, West Building



Academic Buildings
Institute for Advanced Study

TRAVEL

A. From New York

Princeton is about 50 miles southwest of New York and can be reached in two hours or less in three ways:

1. Penn Central Railroad
2. The Suburban Transit Bus Line (departures every half hour)
3. Private car, taxi, or rental car

The most economical way to travel is:

1. Carey Transportation Company from the metropolitan airports to the East Side Air Terminal on Manhattan Island; then
2. Taxi to Penn Central Railroad Station at 23rd Street and 7th Avenue, or to the Port Authority Bus Terminal at 40th Street and 8th Avenue; then
3. Train or bus to Princeton; then
4. Taxi to the Institute.

B. From Philadelphia

Princeton is about 40 miles northeast of Philadelphia and can be reached in less than two hours by:

1. Penn Central Railroad from 30th Street Station
2. Private car, taxi, or rental car

The most economical way to travel is:

1. Taxi from the airport to Penn Central Railroad at 30th Street Station; then
2. Train to Princeton; then
3. Taxi to the Institute

C. From Newark

Princeton is about 40 miles southwest of Newark airport and can be reached in less than two hours by:

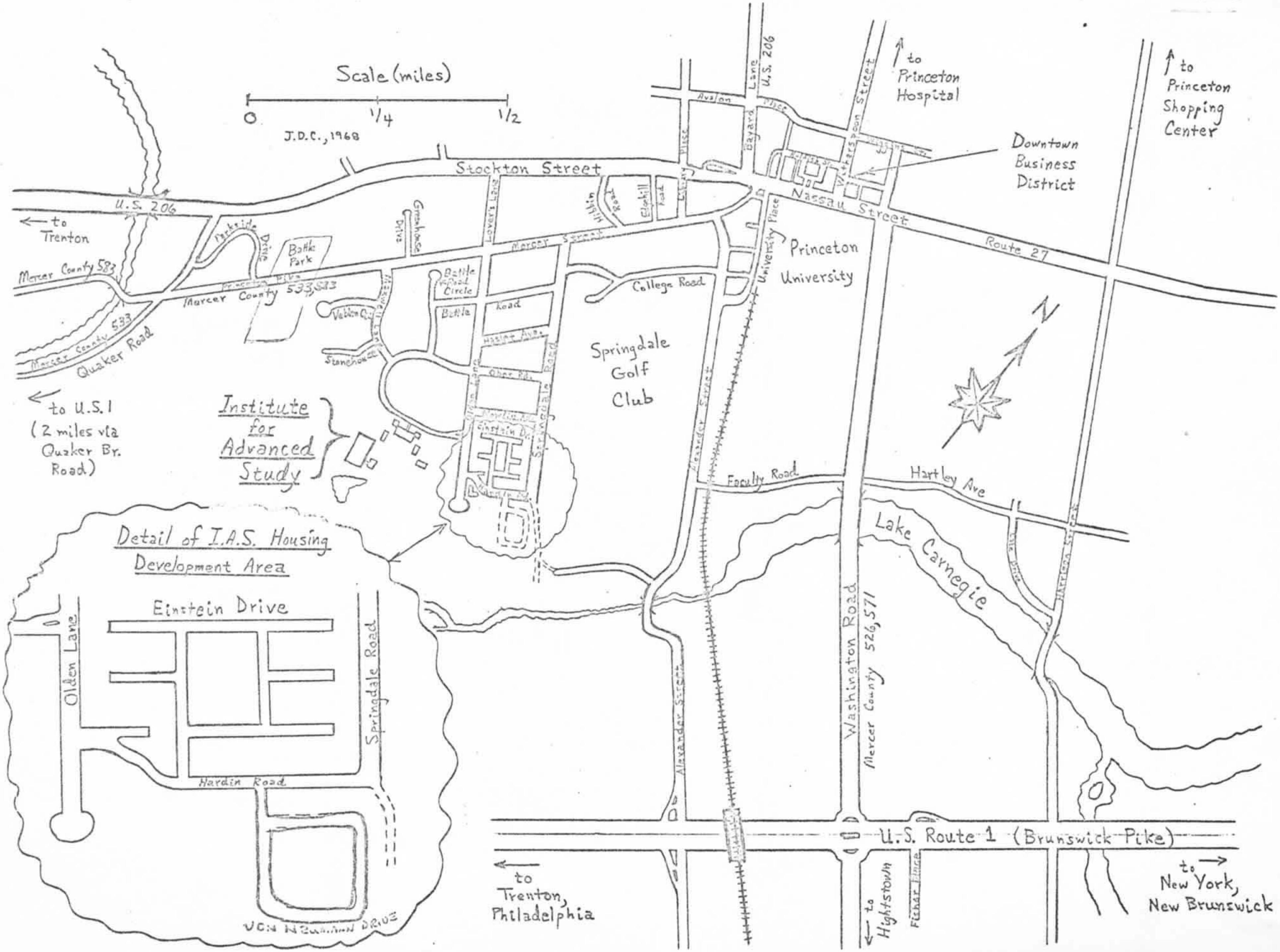
1. Penn Central Railroad from Newark station.
2. Private car, taxi or rental car

The most economical way to travel is:

1. Taxi (or NO.4 Public Service bus) from airport to Penn Central Railroad station in Newark; then
2. Train to Princeton; then
3. Taxi to the Institute

D. From Trenton

Princeton is about 12 miles from Trenton and can best be reached by taxi from the Penn Central Railroad Station ($6.00) in about 1/2 hour. Sometimes train connections can be made from the Trenton station to Princeton; then take taxi to the Institute


Scale (miles)
0
1/4
1/2
J.D.C., 1968
Stockton Street
U.S. 206
to Trenton
Mercer County 583
Mercer County 533
Quaker Road
Princeton Pike
Mercer County 533,583
Battle Park
Mercer Street
Lover's Lane
Battle Road Circle
Battle Road
Springdale Road
Springdale Golf Club
College Road
to U.S. 1
(2 miles via
Quaker Br.
Road)
Institute
for
Advanced
Study
Einstein Dr.
Detail of I.A.S. Housing
Development Area
Einstein Drive
Olden Lane
Springdale Road
Hardin Road
VON NEUMANN DRIVE
Avalon
Bayard Lane
U.S. 206
Witherspoon Street
to Princeton Hospital
to Princeton Shopping Center
Downtown
Business
District
Nassau Street
Route 27
University Place
Princeton
University
Alexander Street
Faculty Road
Hartley Ave
Lake Carnegie
N
Washington Road
Mercer County 526,571
Harrison Street
U.S. Route 1 (Brunswick Pike)
to
Trenton,
Philadelphia
to Hightstown
Fisher Place
to
New York,
New Brunswick

THE INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY 08540

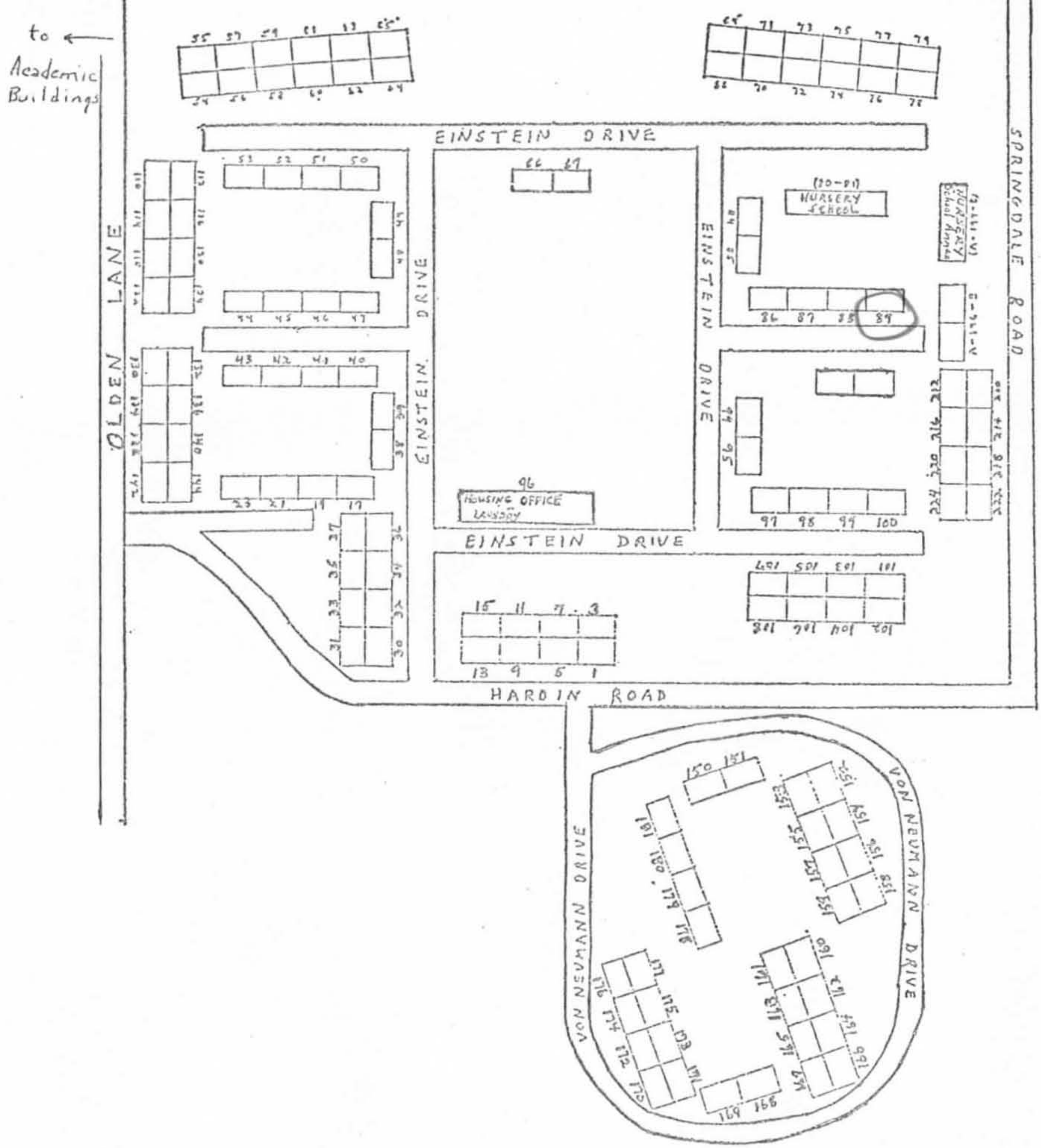
HOUSING

* Assignments for the nights of June 6, 7 and 8

Name	Housing Unit
Mr. Aborn	156 von Neumann Drive
Mr. Ballotti	106 Einstein Drive
Mr. Book	57 Einstein Drive
Mr. Brenner	71 Einstein Drive
Mr. Charney	53 Einstein Drive
Mr. Curtis	158 von Neumann Drive
Mrs. Davis	222 South Olden Lane
Mr. Elias	76 Einstein Drive
Mr. Gleason	114 South Olden Lane
Mr. Goldstine	68 Einstein Drive
Mr. Gomory	162 von Neumann Drive
Mr. Graubard	106 Einstein Drive
Mr. Hansmann	30 Einstein Drive
Mr. Hines	156 von Neumann Drive
Mr. Holton	34 Einstein Drive
Mr. van Hove	89 Einstein Drive
Mr. Klein	79 Einstein Drive
Mr. Knopoff	79 Einstein Drive
Mr. Landes	72 Einstein Drive
Mr. Lederberg	89 Einstein Drive
Mr. Lieberman	107 Einstein Drive
Mr. and Mrs. Linder	44 Einstein Drive
Mr. Munk	53 Einstein Drive
Mr. Oettinger	163 von Neumann Drive
Mr. Prim	166 von Neumann Drive
Mr. Rabin	73 Einstein Drive
Mr. Rensberger	107 Einstein Drive
Mr. Roberts	71 Einstein Drive (through June 11)
Mr. Scarf	76 Einstein Drive
Mr. and Mrs. Segal	104 Einstein Drive
Mr. Silberberg	62 Einstein Drive
Mr. and Mrs. Solow	100 Einstein Drive
Mr. Tilly	72 Einstein Drive
Mr. Ulam	68 Einstein Drive
Mr. Zapolsky	158 von Neumann Drive

*IN THE EVENT OF DELAYED ARRIVAL YOUR APARTMENT WILL BE LEFT UNLOCKED WITH KEY ON THE KITCHEN TABLE

MEMBERS' HOUSING PROJECT - IAS



THE INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY 08540
Telephone-609-924-4400

THE DIRECTOR

July 27, 1971

Dear Lederberg:

The list of participants in the von Neumann Symposium is now substantially complete. They are as follows:

	Paper	Author	Discussant
1.	Pure and applied mathematics	Stanislaw Ulam, Los Alamos Scientific Laboratory, University of California	Michael Atiyah, Institute for Advanced Study (tent.)
2.	Logic and the foundations of mathematics	Michael Rabin, I.B.M. - Thomas J. Watson Research Center and Hebrew University, Jerusalem	?
3.	Physics and astrophysics	K. V. Roberts, Culham Laboratory, Abingdon, Berkshire	Leon van Hoeve CERN, Geneva, Switzerland
4.	The applied physical sciences	Walter H. Munk, Institute of Geophysics and Planetary Physics, University of California at San Diego	Gordon MacDonald Council on Environmental Quality, Washington, D.C.
5.	The biological sciences	Sidney Brenner, Medical Research Council Laboratory of Molecular Biology, Cambridge, England	Joshua Lederberg, Department of Genetics, Stanford University
6.	Economics	Lawrence R. Klein, Department of Economics, University of Pennsylvania	Hebert Scarf, Department of Economics, Yale University (tent.)
7.	The historical social sciences	Charles V. Tilly, Department of Sociology, University of Michigan	David Landes, Department of History, Harvard University
8.	Language, learning, and models of the mind	George Miller, I.A.S. and Rockefeller University	?

I have asked each contributor of a paper to have a copy in the hands of the discussant by 1st May, 1972. I trust this gives you adequate time to order your thoughts for comment. Please let me know if you have any questions about this or other aspects of the arrangements.

Sincerely yours,

Carl Kaysen

Carl Kaysen

Professor Joshua Lederberg
Department of Genetics
School of Medicine
Stanford University
Stanford, California 94305

51043

THE APPLIED PHYSICAL SCIENCES

Jule Charney and Walter Munk

I. INTRODUCTION

Our charge is to examine the influence of the computer in applied physical sciences. We have taken a narrow interpretation, limiting ourselves to the geophysical sciences; even there, the emphasis is almost entirely on the fluid envelope of the planet Earth. This is in part because of our own limited competence, in part because we believe that the underlying principles of the computer revolution are better brought out by case histories than broad generalities. Even so, we end up with more than we can handle.

II. EARLY HISTORY OF COMPUTING IN METEOROLOGY

In 1946, at the end of the Second World War and on the eve of the arithmetical revolution, a conference on mathematics was held at Princeton University in celebration of its bicentennial anniversary.[1] Two famous mathematicians spoke of the future importance of the computer. Hermann Weyl expressed concern that the store of mathematical substance which formed the basis for current generalizations was in danger of becoming exhausted without outside help, "be it even by such devilish devices as high-speed computing machines". And John von Neumann remarked that the success of mathematics with the linear differential equations of electrodynamics and quantum mechanics had concealed its failure with the nonlinear differential equations of hydrodynamics, elasticity and general relativity. He expressed the hope that the computer-aided solution of a large store of problems in nonlinear continuum mechanics would indeed supply a basis for mathematical generalization.

To him meteorology was par excellence the applied branch of mathematics and physics that stood the most to gain from high-speed computation. Earlier that year he had called a conference of meteorologists to tell them about the general-purpose electronic computer he was building at the Institute for Advanced Study and to seek their advice and assistance in designing meteorological programs for its use. Charney had the good fortune to attend and recalls that the response from the established figures was interested but less than enthusiastic. C.-G. Rossby perhaps best voiced their feelings by stating that the mathematical problem was not yet defined: there were more unknowns than equations, for we had not yet been able to express the components of the Reynolds stress tensor in terms of mean flow variables. Citing L. F. Richardson's[2] gallant but unsuccessful attempt to solve the hydrodynamical equations of the atmosphere by hand calculation, Rossby said that computation could not be successful before observation, experiment and analysis had led to a better understanding of fundamental atmospheric processes, in particular of atmospheric turbulence.

His caution had the positive effect of convincing von Neumann of the need for physical as well as mathematical analysis, but he failed to appreciate the great psychological stimulus that the very possibility of high-speed computation brought to meteorology. All of its branches were given new urgency and new importance by the promises that the contributions to the atmospheric circulation from a variety of physical processes could be synthesized mathematically within the computer. This was especially true of dynamical meteorology, where the requirements of mathematical tractability had forced such oversimplifications that theories of the large-scale circulation were of little use in prediction and incapable of meaningful comparison with reality.

Charney joined von Neumann's group two years later and took up the task of formulating a hierarchy of mathematical models embodying successively more and more of the physical and numerical aspects of the general prediction problem, hoping in this way to avoid the dangers of introducing a great many poorly understood factors all at once.[3] He had previously devised a method for overcoming the mathematical difficulties responsible for Richardson's failure. His point of departure was the realization that a compressible, stratified fluid held gravitationally to a rotating sphere can support a variety of wave motions, including acoustic and inertial-gravity oscillations, which are of little meteorological importance but which impose highly restrictive conditions on numerical algorithms for solving the gas dynamical equations. He proposed to filter out these "noise" motions by imposing certain equilibrium constraints on the primitive equations of motion.[4] It can be established by scale analysis that the slow, large-scale motions containing the bulk of the atmosphere's energy are close to equilibrium in the sense that the pressure, gravitational, centrifugal and Coriolis forces are nearly in balance. Because of this balance, quantities like acceleration and velocity divergence are obscured by the noise motions, and any scheme, such as Richardson's, which is based upon their explicit calculation will give nonsense in the first few time steps. The balance equations deal only with observables and are therefore consonant with meteorological experience and theory. They were immediately successful in application to prediction.

Not surprisingly, the simplest model in the hierarchy was Rossby's own. He had previously suggested that the atmosphere at a level between 3 and 6 km behaves as if it were a two-dimensional, incompressible flow and had deduced his famous dispersion formula for long, small-amplitude waves in a uniform zonal current.[5] Charney and Eliassen[6] verified that there were indeed important aspects of the observed motions, involving horizontal energy dispersion rather than vertical dispersion or overturning, which were explicable as two-dimensional phenomena; whereupon Charney, Fjørtoft and von Neumann[7] devised a numerical method for solving the two-dimensional vorticity equation

$$\frac{\partial Z}{\partial t} = \frac{1}{a^2\cos^2\phi}\frac{\partial(Z,\psi)}{\partial(\phi,\lambda)} , \qquad Z \equiv \nabla^2\psi + 2\Omega\sin\phi ,$$

for the stream function $\psi(\phi,\lambda,t)$ on a rotating sphere of radius a, latitude ϕ, longitude λ and angular speed Ω. The solutions, carried out in 1950 on the ENIAC*, bore a sufficient resemblance to reality to inspire further effort.

At that time one of the principal problems of meteorology was extratropical cyclogenesis, the formation of the large cyclonically rotating vortices responsible for weather in middle and high latitudes. Charney[8] and Eady[9] had explained cyclogenesis theoretically as an instability of the mean zonal flow according to which the incipient cyclone grows at the expense of the potential energy associated with the mean equator to pole temperature gradient; and Phillips[10] had shown that a model consisting of two superimposed homogeneous, incompressible layers of different density was capable of simulating the main features of the more general theoretical models. Accordingly, the first models programmed for the new computer at the Institute for Advanced Study consisted first of two and then of three layers. An exceptionally strong and sudden development of a cyclone over the Eastern United States was

---

* Electronic Numerical Integrator And Computer, at the Ballistics Research Laboratory of the U.S. Army Ordnance Department, Aberdeen, Maryland. The ENIAC was built as a special-purpose computer by J.P. Eckert and J.W. Mauchly at the University of Pennsylvania and later provided with a general-purpose control after the design of von Neumann.

successfully predicted in 1953 with a three-layer model, of course after the event. The results tended to substantiate the theoretical ideas concerning its mechanism of generation, and they interested the U.S. Government in the possibilities of numerical weather prediction. A national numerical forecasting unit was set up in 1954 and began experimental operation in 1955. Similar units were established in other countries.

Jeffries[11], Bjerknes[12] and Starr[13] had shown empirically that the cyclone is a major element in transporting angular momentum poleward from the tropics to maintain the mid-latitude westerly zonal flow against frictional dissipation; and Kuo[14] and Charney[15] had proposed the explanation that the cyclone wave, while deriving energy from the potential energy associated with the meridional temperature gradient in the zonal flow, returns kinetic energy to the flow through the action of the Reynolds stresses in the presence of a stabilizing meridional vorticity gradient. A rudimentary, thermally-active model permitting this type of interaction was constructed by Phillips - with sealing wax and string so to speak - and in 1955 the first dynamically consistent simulation of the atmosphere's general circulation was obtained; the conjectured mechanisms were found to operate, and many important features of the observed circulation were simulated.[16]

In 1955, a second conference of meteorologists was held at the Institute for Advanced Study to consider the implications of these extremely encouraging results for long-range prediction and for the simulation of climate. Von Neumann[17] divided the motions of the atmosphere into three categories: (1) those that are determined primarily by the initial conditions; (2) those that are practically independent of the initial conditions;

and (3) those that are not so far from the initial state that they are unaffected by the initial conditions but sufficiently far that the initial conditions do not express themselves clearly. The corresponding prediction problems are short-range prediction, sinulation of climate, and long-range prediction respectively. From general experience he expected the first problem, for which the extrapolation parameter is comparatively small, to be the simplest, the next most difficult to be the asymptotic problem for which the extrapolation parameter is very large, and the most difficult of all to be the intermediate problem for which the extrapolation parameter was neither large nor small.

His expectations have been amply borne out. The greatest successes have been achieved in short-range weather prediction and in the simulation of climate; a beginning attack has even been begun on the simulation of a coupled ocean-atmosphere system; but long-range prediction has remained essentially an unsolved problem. It is not even known what, if anything, can be predicted for more than two weeks. Nevertheless, the potential social and economic benefits to be derived from even a modest extension of forecast range or accuracy have justified a considerable investment in observing and data processing equipment. The computer has brought a new maturity to meteorology whereby theory and observation at last exist on an equal footing, and the requirements for numerical prediction and physical understanding have become a major influence in the selection of observational systems. The Global Atmospheric Research Program, sponsored jointly by the World Meteorological Organization and the International Council of Scientific Unions, is the international expression of this reality. A description of these efforts is given in a number of publications.[18,19]

The considerable progress that has been achieved in the past twenty-five years has been based largely on the adaptation of old ideas to the new computational framework; gradually the growth of sophistication in numerical techniques and of speed and capacity of computers has made it possible to incorporate existing knowledge of physical processes in something like the order of their importance for the atmospheric circulation. There have been difficulties. Parkinson's law of computing in meteorology is that problems expand to fill all computers. As a result, spatial resolution and truncation error have diminished only slowly, and it has not always been possible to distinguish between mathematical and physical error. Short-range numerical predictions for the middle and upper atmosphere have demonstrated their usefulness, but it has been only recently that the more complicated surface boundary-layer structures, involving friction and topography, have been sufficiently resolved to yield useful numerical predictions of wind and temperature. The numerical prediction of cloud and precipitation remains only marginally useful because it involves poorly understood condensation phenomena on still smaller scales.[20]

The question of predictability arises. In the early 1950's there was an apparent divergence of view concerning the nature of the meteorological prediction problem between what may be called the Princeton school under von Neumann and the Cambridge (Massachusetts) school under Norbert Wiener. Von Neumann and Charney regarded prediction as a determinate initial value problem, whereas Wiener[21] and others at M.I.T. emphasized the stochastic character of the atmospheric motions and proposed instead to employ linear "black box" prediction methods based upon long time-series of past data. Wiener went so far as to state privately that von Neumann and Charney were misleading the public by pretending that the atmosphere was predictable as a determinate system. Without taking sides in the controversy, let us attempt to look at the problem of numerical prediction from a general point of view.

At a height of 100 km the density and pressure in the atmosphere are less than a millionth part of their surface values. Below 100 km the atmosphere is very nearly a perfect gas obeying known laws of physics: the Navier-Stokes equations of motion are not seriously in doubt; absorption, emission and scattering of radiation by the principal atmospheric molecules is not a mystery; the calculus of radiative transfer is well advanced; and while there remains more to be learned about such phenomena as the microphysics of cloud formation, scattering and absorption by suspended particulates, and the influence of trace gases on ozone photochemistry, the laws of motion and energy exchange may be considered quite well understood. If Laplace's mathematical intelligence were replaced by a computing machine of unlimited speed and capacity, and if the atmosphere below 100 km were spanned by a computational lattice whose mesh size were less than the scale of the smallest turbulent eddy, say one millimeter, there can be little doubt that numerical integration of the partial differential equations embodying the known dynamical and thermodynamical laws of motion would simulate and predict atmospheric behavior with considerable accuracy. Would the problems of meteorology then have been solved? Or, contrariwise, would more be known than can presently be learned by careful observation of the atmosphere itself, which, after all, is its own analog computer?

The answer would seem to lie somewhere between these extremes. With respect to prediction, even if one disregards the indeterminacies arising from lack of knowledge of fluctuations at the boundaries - these could be overcome by extending the calculations higher into the atmosphere and lower into the oceans - it is very likely that all accuracy would have vanished in less than one month. This is not because of quantum indeterminacy, or even because of macroscopic errors of observation, but because the errors introduced into the smallest turbulent eddies by random fluctuations on the scale of the mean free path (ca 10^{-5} mm at sea level), although very small initially, would grow exponentially until in a very short time these eddies would have become indeterminate and would have begun to affect the next larger scales by nonlinear interaction. The escalation to larger scales would continue until eventually the main energy-bearing scales would have been rendered indeterminate.[22,23,24] This escalation of the error occurs because the turbulence exhibits a similarity behavior such that the time scales of the interactions increase in approximate geometric ratio with scale (at least for the smaller scales) in such a manner that the error progresses from 1 mm to 10 km in less than one day, and from 100 km to the planetary scales in a week or two.*

Unfortunately, or fortunately, the limitations imposed by the finiteness of the velocity of light and the least size of a stable memory storage element make computation for the 5×10^{28} points in the one-millimeter lattice impossible in times short of the astronomical. Even the 5×10^{9}

---

* The upward propagation of uncertainty toward larger scales must face a downward propagation of energy toward smaller scales in the high-wavenumber "Kolmogoroff" range. Leith and Kraichnan[25] have shown that uncertainty wins.

points in a lattice with a horizontal mesh size of 10 km and a vertical mesh size of 100 m is beyond the capacity of any present or contemplated computer. Thus one is forced to reduce the effective number of degrees of freedom by dealing with volumes containing turbulent elements ranging over several decades of scale, and it becomes necessary to determine the turbulent transports of mass, momentum, and energy from one volume to the next, i.e., to develop a theory of small-scale turbulence. More generally, one must develop statistical theories for all processes smaller than the least computational mesh size. These include not only mechanically and thermally driven turbulence, but also cumulus convection, internal gravity-waves produced by flow over obstacles, wave interactions at the air-sea interface, etc. Ideally the mesh sizes should be smaller than the scale over which the small-scale processes are statistically homogeneous. Given realistic limitations in computer speed and capacity, the computational lattices must ultimately become variable in space and time, or the base functions for Galerkin methods of approximation must change in time.

The problems of mechanically or convectively driven turbulence, or of cumulus convection, are no closer to solution than those of the general circulation of the atmosphere. They, too, can be aided by computer simulation; but the reduction in scale is not an essential reduction in complexity, and it has been estimated that each will require approximately the same amount of computation as the general circulation.

It is perhaps fortunate for human activity that motions whose scales are so large that they can be resolved by possible computational lattices constitute a fairly well-defined set. This is because the energy of the turbulent motions decreases very rapidly with scale below a space scale of about 1000 km and a time scale of two or three days. It is this circumstance that makes the atmospheric flow predictable in principle for as long as a week and makes the large-scale motions the controlling entities.

It would thus appear that there is some justification for both the Princeton and the Cambridge schools of thought. It is possible to look at the large-scale motions of the atmosphere as a determinate system for short periods of time, providing one has a statistical theory permitting the incorporation of the turbulent fluxes of matter, energy and momentum. But for longer periods of time the atmospheric variables, even on a large scale, must be regarded as random.

There remains the question as to the best method of statistical prediction. The use of linear prediction methods is based on the assumption that one is dealing with a conservative system, or at least one which is in statistical equilibrium. In this case the ergodic theorem asserts that a present state of the atmosphere may be found to any degree of approximation in a catalog of past states, if the time series is sufficiently long. It is very doubtful that such methods will prove feasible, not only because the time series would have to be impossibly long but because the atmosphere-ocean-Earth system is nonconservative, and the very notion of statistical equilibrium on a climatic time scale is in doubt. It is more likely that long-range prediction will become a matter of calculating probability distributions and ensemble averages by Monte Carlo methods and climatic theory a matter of calculating statistical moments from long computer simulated time series.

III. SOME CASE HISTORIES

The past twenty-five years has seen much progress in meteorology and oceanography. To what extent has this progress been aided by the computer? And how has the computer affected the manner in which problems are selected and solved? We believe that answers to such questions are best given by presenting a few representative case histories.

The uses of the computer in fluid geophysics may be classed as synthetic, experimental, heuristic and data-analytic. The first category includes the prediction or simulation of the large-scale circulation of atmospheres and oceans, numerical studies of predictability, simulation of climate and climatic variation, of mechanically and convectively driven turbulence on various scales, of atmospheric and oceanic tides, and of cloud microphysics and dynamics. The second category deals with the simulation of individual phenomena as a means of inferring their physical causes, insofar as these phenomena can be isolated in the highly interactive, nonlinear atmosphere-ocean system. Among the atmospheric problems in this category are extratropical cyclogenesis, tropical cyclogenesis and frontogenesis. The oceanographic problems include the formation of such major currents as the Gulf Stream, the Kuroshio, the Somali Current, the Antarctic Circumpolar Current, and the Equatorial undercurrents. The third category includes the construction of highly simplified numerical models, sometimes in conjunction with laboratory experiments, sometimes alone, as a means of discovering new relationships having a bearing on the nonlinear behavior of the atmosphere and oceans. For example, games played with highly truncated Fourier series or with a finite number of vortex elements have been valuable in illuminating the nonlinear and stochastic properties of complex systems whose detailed numerical prediction, especially over long periods of time, is either prohibitive or so time-consuming as to furnish little understanding for the effort expended. The fourth category deals with problems of data reduction and analysis as a means of inferring physical behavior from raw observations. The following are some case histories taken from the various categories.

1. Tropical cyclogenesis. One of the major stumbling blocks to accurate simulation of the atmospheric circulation is lack of knowledge of the statistical dynamics of cumulus convection. Solar insolation is made available for driving the atmospheric circulation primarily by evaporation of moisture from the sea surface and liberation of the latent heat by condensation in rising currents of air. In the tropical atmosphere the release of latent heat adds sufficient buoyancy to the rising air parcels to produce gravitational instability. This instability is manifested as cumulus convection. It has been conjectured that tropical disturbances arise from a kind of cooperative interaction between an ensemble of cumulus cells and the large-scale field of motion.[26] According to this hypothesis, frictionally-induced convergence of moisture in the surface boundary layer of a large-scale cyclonic disturbance supplies moisture for cumulus convection, and the convection in turn supplies latent heat energy for driving the large-scale disturbance against frictional dissipation. Theoretical analysis is made difficult by the fact that condensation processes on both small and large scales are intrinsically nonlinear, even for small amplitudes (the release of heat of condensation depends on the

sign as well as the magnitude of the vertical velocity). For this reason adequate theories relating the convection to the large-scale motion are lacking. Nevertheless, several semi-empirical schemes for incorporating the large-scale effects of the release of latent heat have been used for the numerical simulation of tropical disturbances. The growth of an axi-symmetric disturbance into a hurricane has been simulated with some success by Ooyama[27] and others. Numerical models have even been used to simulate hurricane modification by artificial and natural means.[28]

But there have been difficulties. All small-amplitude symmetric disturbances of an atmosphere at relative rest grow into hurricanes. No small-amplitude asymmetric disturbances grow into hurricanes.[29] Since only a small fraction of all real tropical disturbances do grow into hurricanes, the conclusion of the syllogism is that the cyclogenesis requires an initial asymmetric, finite-amplitude disturbance. We conjecture that the hurricane arises as a kind of finite-amplitude instability, and that only when cumulus statistics are properly understood, and the interaction of the perturbation with the asymmetric shearing flow in its environment is properly taken into account, will the prediction of tropical cyclogenesis become possible.

2. The general circulation of the atmosphere. The invention of the balloon-borne radiosonde and the military requirements of World War II brought about a rapid expansion of the global observational network in the thirties and forties. For the first time the three-dimensional structure of the atmospheric circulation became sufficiently well known to supply a foundation for theory. To a first approximation the circulation

may be described as an axisymmetric, circumpolar vortex on which wave and vortex perturbations are superimposed. Analysis has suggested that travelling disturbances are due to instability of the mean zonal flow[8,9] and stationary disturbances to forcing of the zonal flow by mechanical[6] and thermal[30] action. The deformation fields in the cyclone wave disturbances give rise to velocity and temperature discontinuities. These are the so-called frontal surfaces whose three-dimensional structure was described by J. Bjerknes and H. Solberg[31] just after World War I and whose formation was described by T. Bergeron[32] a few years later. A dynamical theory of frontogenesis has only recently been given, by Stone[33], Williams[34] and Hoskins[35].

Linear analysis of wave and vortex modes and their first order interactions with the mean flow have been useful. One seeks causal laws, but a numerical prediction or a numerical experiment is not in itself a causal law; it can only verify a causal law. When one is dealing with nonlinear, nonstationary phenomena, causal laws require for their statement or interpretation a basic vocabulary of characteristic wave and vortex modes of which the more complex motions are composed. Undoubtedly a fuller explanation will require that one consider self-interactions and higher-order interactions, as well as secondary and tertiary instabilities, but it is likely to remain true for some time that linear thinking based on simple models will serve as a first-order guide in the labyrinth of nonlinearity.

Phillips' numerical experiment[16] synthesized several of the processes that had been investigated analytically. It was followed by more elaborate experiments with more realistic atmospheric models. Numerical integrations for the entire global circulation with realistic surface boundary conditions and energy sources and sinks have simulated the characteristic features of the large-scale circulation with some fidelity[36]. Others have simulated the simpler rotating-tank analogs.[37] These integrations, when taken in combination with the theoretical studies and the simple laboratory and numerical model experiments, are a kind of definition-in-use of what one means by an "understanding" of the atmospheric circulation.

One may now ask: how has the computer contributed to our understanding of the general circulation? Let us enumerate its accomplishments: the theoretical and laboratory models are too greatly simplified to be directly applicable to the atmosphere; numerical computations have verified that the hypothetical mechanisms do indeed operate. The theories of cyclogenesis were developed before the computer, but the theories of frontogenesis were influenced by laboratory and numerical models of the frontogenetic process. The most successful models of tropical cyclogenesis are numerical. Genuinely nonlinear phenomena, such as the fluctuating interaction of the perturbations and the mean flow were first simulated numerically and only later, and to a very limited extent, analytically. Although we have no general theories of nonlinear atmospheric and oceanic processes, the accumulating store of special results is beginning to reveal characteristic properties, such as transition from steady, to regularly fluctuating, to random regimes, which are shared with quite different nonlinear systems.* These add support to von Neumann's anticipation that the computer will eventually supply a basis for meaningful generalization in nonlinear continuum mechanics.

---

* See, for example, Pasta[38].

When a computer simulation successfully synthesizes a number of theoretically-predicted phenomena and is in accord with reality, it validates both itself and the theories - just as the birth of a child who resembles a paternal grandfather legitimizes both itself and its father. The theoretical ideas of extratropical cyclogenesis and of the interaction of the cyclone with the zonal flow were legitimized in this way. But this synthetic use of the computer was effective only in combination with observation and theory. When one element in the trichotomy is absent the synthesis fails. If it is theory, the computer may function as an experimental-heuristic device. It has played this role in the theory of tropical cyclogenesis.

3. <u>The general circulation of the oceans.</u> The computer has not played an equally important role in the theory of ocean circulations. Stommel[39] remarked in 1954 that the theories of the ocean circulation had a peculiarly dreamlike quality. They retain some of this quality to the present day. The cause is the difficulty of observing the deep ocean and the lack of a sufficient economic incentive for doing so. The turbulent transfers of momentum and heat (and salt) are not understood, especially in the deeper layers of the ocean, and yet they play a more decisive role in the oceans than in the atmosphere. It is not known, for example, whether turbulent diffusion is vertically downward from the surface or along isentropic surfaces inward from coastal boundaries.

Nevertheless, theory has flourished. In the late forties physical oceanographers, having little knowledge of turbulent transport mechanisms, concentrated their attention on predicting the behavior of the vertically integrated mass transport. Sverdrup[40] employed the principle of force equilibrium already mentioned in Chapter II to estimate the mass transport of the principal ocean currents from the observed surface wind stress. His explanation failed at the western boundaries where the reduced order of the equilibrium equations made them incapable of satisfying the condition of zero normal mass transport. Stommel[41] and Munk[42] explained the western intensification of the ocean currents, i.e., the Gulf Stream and the Kuroshio, as nature's way of satisfying the boundary conditions by introducing higher order frictional forces; Charney[43] showed that this could be accomplished by inertial accelerations without invoking ad hoc assumptions concerning turbulent Reynolds stresses; and Kamenkovich[44] combined the frictional and inertial processes into a single theory. Later theories explaining vertical structure were developed by Robinson and Stommel[45] and Robinson and Welander[46], but again with ad hoc coefficients of momentum, heat and salt diffusion.

Classical analysis sufficed, or at most computers were used as auxiliary devices for numerical quadrature and for solving two-point boundary value problems. Such nonlinearities as existed were mitigated by hodograph or similarity transformations. Recently, computers have been used more intrinsically by Charney and Spiegel[47], McKee[48] and Philander[49] to construct theories of the equatorial undercurrents. These are also critically dependent on the mechanism of turbulent diffusion and consequently incapable of giving truly satisfactory explanations.

Computation suffers not only from a lack of knowledge of turbulent transport mechanisms but from the existence of widely varying time scales in different layers of the oceans. These range from days or weeks in the upper layers and near the equator where the action of wind stirring is large, to years at midlatitudes and somewhat greater depths where velocity and density gradients remain appreciable, to centuries in the abyssal regions where velocity and density gradients are very small. This is characteristic of a circulation which is driven by heating and friction at the upper boundary of a fluid in a gravitational field.* Similar problems are encountered in the simulation of the deep Venus circulation if it is assumed that the continuous cloud cover prevents the penetration of solar radiation to great depths. Despite these limitations, numerical models of the global oceans have been constructed by Sarkisyan[50], Kamenkovich et al.[51], Bryan and Cox[52], and Bryan[53]. The first two avoid the multiple time problems by dealing with a homogeneous ocean; in the second two it is not clear that the deep circulation reaches a steady or statistically steady state. Nevertheless, certain features of the observed ocean circulation are simulated by all models, as for example the Gulf Stream, the Kuroshio, and, in the case of the latter two, some aspects of the vertical thermohaline structure of the oceans. Recently Cox[54] has been able to reproduce features of the Somali Current, showing that, unlike the other western boundary currents, it is much more influenced by local wind action.

The Antarctic Circumpolar Current, as its name implies, is not as obviously confined to a closed oceanic basin as the other major currents.

---

* The oceans may also be driven to some extent from the bottom upwards by interaction of the tides with bottom topography.

Munk and Palmén[55] and Stommel[56] pointed out that its dynamics must also be different. It is, however, constrained by the narrow and shallow "Drake Passage", and Gill and Bryan[57] have shown by numerical simulation that the entire character of the current, including its transport, is strongly affected by the shape of this passage. If, for example, the Drake Passage were deepened, the circumpolar transport would decrease by a factor of three. The calculated current is not realistic; again, the model cries out for a better understanding of the turbulent eddy viscosities and heat conduction.

Measurements of deep ocean currents by means of neutrally-buoyant floats tracked acoustically or by current meters attached to moored buoy cables have revealed a degree of unrest in the deep oceans which threatens to overthrow some 'dream' theories and suggests entirely new mechanisms of turbulent diffusion. The Mid-Ocean Dynamics Experiment (MODE) planned for 1973 is a concerted attempt to measure these motions in a limited region of the western north Atlantic. The study is unusual because it is accompanied by a theoretical and numerical effort to examine a series of idealized models for guiding observation. Bottom topography is thought to play an important role. Theoretical analysis and numerical experiments carried out by Rhines[58] have shown that vortex shrinking and stretching due to up- and down-slope flow acts as a mechanism for trapping the energy of oceanic Rossby waves.

4. Climate and climatic change. Lorenz[59] has raised the question of the statistical transitivity of the solutions of the atmospheric or the coupled atmosphere-ocean equations. Is there a unique climate? Or is it possible that two different initial states might lead to two different, but stable, statistics? Examples of both possibilities occur in rotating,

differentially-heated, cylindrical annuli. For certain values of the boundary parameters, the flows are degenerate: different initial conditions lead to different steady or regularly fluctuating regimes. Such flows have been studied by Lorenz[60] with simple truncated Fourier and Fourier-Bessel expansions and by Charney[61] with finite point vortex elements.

No feasible method has been proposed for simulating the actual climate except by carrying out numerical calculations of the global circulation for long periods of time. When one considers that there are no accepted theories of anisotropic, nonhomogeneous turbulence of any kind, it is not to be expected that the vastly more complicated general circulation of the atmosphere-ocean system will reveal statistical regularities that would permit the inference of an _a priori_ statistical theory of climate. Seasonal averages computed from global integrations carried out for a period of one year by Manabe, et al.[62] at the National Oceanographic and Atmospheric Administration Geophysical Fluid Dynamics Laboratory at Princeton University compare reasonably well with normal climatic averages. But there are discrepancies. To what extent these are due to physical and mathematical inadequacies in the model and to what extent to the natural variability of the atmosphere from one year to the next is not yet known. Numerical calculations carried out with a simpler model, and for longer periods of time by Katayama, Mintz and Arakawa[63], with a fixed ocean temperature field but with seasonally varying solar insolation, reveal surprisingly large variations in seasonal averages from one year to the next. If the ocean temperatures had been permitted to vary, these variations would presumably have been larger.

The variations would presumably have been even larger, and perhaps of longer persistency, if the oceans had been coupled into the system. We know that certain baroclinic oscillations in the oceans have time constants of the order of centuries, and that the "overturn" time of the oceans is a thousand years (as determined from C^{14} ages). Long time series of oceanographic variables are almost entirely lacking, but where they exist, they show striking anomalies occurring for years, or even a decade. This type of observation is now being greatly accelerated under the auspices of the International Decade for Ocean Exploration.

Once realistic averages of the global circulation have been obtained, one is then in a position to carry out numerical experiments concerning the role played by variable solar input, by variable boundary conditions, etc. We know that planetary perturbations of the Earth-Moon-Sun system are associated with significant changes in the radiation pattern, such changes occurring with typical time scales of 10^4 - 10^5 years. These planetary perturbations correlate reasonably well with profiles of paleotemperatures in deep-sea cores, as determined by isotopic methods. Going back even further to time scales of 10^7 - 10^8 years, one can consider the effects of the disappearance of mountain chains, the closing of Drake Passage, and the variable distribution of oceans and continents. This promises to be an active field of computer application in the decades to come.

A beginning has been made in the development of a numerical model of the combined atmosphere-ocean circulation[64]. It has been possible to simulate the effect of ocean currents on the distribution of temperature, humidity and precipitation in the atmosphere. At the moment the most profound difficulties concern lack of knowledge of turbulent exchange processes in the oceans and the very long oceanic time scales in the deep ocean.

5. Predictability and turbulence. Predictions of the large-scale motions of the atmosphere with the most elaborate of the numerical models are found to be quite accurate for periods of one or two days and to remain superior to the climatic norm up to five days. Some positive skill persists up to nine or ten days.[65] What is the ultimate limit of accuracy? The growth of random observational error was first studied by Thompson[66] and Novikoff[67], but Lorenz[22] was the first to clearly define the problem of predictability as one of instability. He pointed out that the atmosphere is an unstable system in which small perturbations grow until ultimately the predicted flow pattern differs from the observed by as much as two states selected at random. The basic uncertainty is independent of measurement. It begins at the smallest turbulent scales and propagates toward larger scales at a rate which is roughly the characteristic time scale in the statistical equilibrium resulting from interactions among all scales. When uncertainty is introduced at scales which can be resolved by the computational grids, it is possible to investigate its further propagation by numerical simulation. At sub-grid scales its propagation has been investigated by Lorenz[23] and Leith and Kraichnan[25] using statistical turbulence models based on closure hypotheses relating higher to lower order statistical moments.

The rate of progression of uncertainty depends strongly on the nature of the turbulence spectrum. At very small scales, atmospheric and oceanic turbulence appear to satisfy the Kolmogoroff hypotheses of isotropy, homogeneity, and localness in wave-number space, which lead to the $k^{-5/3}$ dependence of the kinetic energy spectral function on the scalar wavenumber k in the so-called inertial subrange. At larger scales, velocity shear and density stratification become important, and at still larger scales the horizontal kinetic energy spectral function exhibits an approximate k_H^{-3} dependence on the horizontal scalar wavenumber k_H. In this case the characteristic time is independent of wavenumber and predictability is greatly extended; it is limited by the non-localness of the interactions in the wave-number space (Leith and Kraichnan[25]). The k_H^{-3} spectral behavior has been ascribed to the two-dimensional characteriof the flow. Onsager[68], Lee[69], Batchelor[70], and Fjørtoft[71] have pointed out that vorticity conservation in two-dimensional flow prevents the kind of energy cascade toward high wavenumbers which is produced in three-dimensional flow by the stretching of vortex tubes. This circumstance led Kraichnan[72] to postulate inertial subranges for two-dimensional turbulence in which energy injected in a given wavenumber band is transferred uniformly to lower wavenumbers whereas mean-square vorticity is transferred to higher wavenumbers. A similarity argument then gives the Kolmogoroff $k_H^{-5/3}$ law in the former range and the k_H^{-3} law in the latter. Kraichnan's hypotheses were apparently confirmed in numerical experiments carried out by Lilly[73], although similar experiments performed by Deems and Zabusky[74] yielded a k_H^{-4} dependence, corresponding to a statistical theory of Saffman's[75] based on the hypothesis that the vorticity field in statistical equilibrium may be characterized by a collection of random step-function discontinuities. Orszag[76,77] has questioned the accuracy of the numerical experiments and has proposed more accurate ones which are now underway.

In any case, Charney[78] has denied that the atmospheric flow can be regarded as two-dimensional and has shown that the k_H^{-3} behavior is due rather to rotational constraints on the large-scale flow. His theory also predicts equipartition among the horizontal components of kinetic energy and the available potential energy and therefore a k_H^{-3} dependence of the temperature variance spectrum as well. Direct measurement and numerical simulation suffer equally from an inability to determine spectra and predictability at small scales. As far as can be determined, they appear to be in accord with Charney's predictions.

In numerical determinations of predictability the practice is to calculate the evolution of a model circulation for a long period of time, store the results, then insert a small random perturbation at an intermediate time and repeat the numerical calculation. The growth of the standard deviation between the perturbed and unperturbed temperature field for the GFDL model[79] remains well below the deviation between two random states for more than two weeks. This result agrees qualitatively with the model calculations of Leith and Kraichnan[25], who find that an initial state determined with the horizontal resolution that may be expected from a satellite-based observing system results in significant predictability of the large-scale motions for more than a week.

Numerical simulations will eventually indicate what statistical quantities remain predictable for periods longer than the deterministic predictability time, i.e., the degree to which the signal due to anomalies in the surface boundary conditions, such as ocean surface temperature, etc., can be detected in the noise due to the growth of uncertainty in the initial conditions. The theory of predictability grew up with computers and there is little doubt that it will remain tied to them.

6. Tides. The theoretical tide problem is defined as follows: given the motion of the Earth, Moon, and Sun, and given the (bottom and coastal) boundaries of the world's oceans, compute the tides. Because of the complexity in the configuration of the world's oceans, this boundary value problem was beyond the scope of solution until the development of modern computers. Munk vividly remembers discussions with von Neumann on this problem in 1946. The principal difficulty, then and now, has to do with the turbulent energy dissipation in the shallow waters of marginal seas. Von Neumann suggested absorbing boundaries as a means of parameterizing the dissipation.

It may come as a surprise to this audience that the problem of tides was not disposed of by Sir Isaac Newton. The trouble with the Newtonian solution, according to which the sea surface is distorted into an equipotential surface, is that it bears no resemblance to reality. The "potential tide" is a static solution; in fact, there are all sorts of resonances of oceanic basins, whose periods are of the same order as that of the tide producing forces, thus requiring a dynamic treatment. Quasi-static solutions correspond to small values of a parameter $\varepsilon = \Omega a/c$, the ratio of the velocity of the sublunar point on a spinning Earth with angular velocity Ω and radius a, compared to the velocity of free waves.

For long gravitational waves, $c = \sqrt{gh}$ where h is ocean depth, and ε happens to be near 1, the most complex of circumstances. The problem is further complicated by the existence of other classes of waves which owe their existence to the Earth's rotation (related to Rossby waves in the atmosphere). These form equatorially and continentally trapped waves called Kelvin waves. Similar solutions exist for waves travelling around large islands, and along the continental slope, undersea escarpments and other depth "discontinuities". The normal modes are then a complex combination of different classes of solutions, and for the general case of neither large nor small ε, these were not sorted out even for basins of simple geometry until computers became available.

At the 1961 General Assembly of the IUGG in Helsinki, Pekeris flashed across the screen the first solution to the theoretical tide problem for the world's oceans, with a comment that it was in good agreement with observations. But not until eight years later was there an opportunity to examine the results.[80] The original work had been for a 12°×12° global grid; as it turned out, a tightening of the mesh system to a 1° grid did not lead to the expected improvement. Pekeris's work was performed on the Weizmann Institute's computer Golem, whose capacity grew in parallel with the tidal grid. Even today, the numerical solution remains unacceptably sensitive to seemingly minor details in the discretization of the boundary. The computational boundaries contain many re-entrant corners, and evidently these affect the eigenvalues in the finite difference approximation. Precise approximation of the free periods is crucial because the frequencies of one or more of the normal modes of the world's oceans lie close to the driving frequency (not surprising in view of the diversity of modes as mentioned above).

However, the key difficulty remains in the choice of coastal boundary conditions. To accommodate dissipation, three conditions have been used:

(i) vanishing normal velocity at coastlines[81,82]

(ii) a specified albedo at coastlines or continental shelves[83]

(iii) specified (observed) values at coastal stations[84] and/or at selected islands.[85,86]

There is a further problem concerning the appropriate bottom boundary conditions. The fact that the ocean bottom is not rigid, but yields under the combined effects of tidal forces and ocean loading, can alter results by a factor of two; yet is not taken into account in most of the numerical work.[87] Hendershott has now formulated this mutual interaction between the ocean and solid-Earth tides. The problem is crucial also to measurements on land of gravity, strain, or tilt; even in the very center of continents these are significantly contaminated by the effect of ocean tides, and unless properly taken into account, geophysical conclusions that one might draw will be in error.

Finally, the recent global tide calculations by Hendershott permit an estimate of the oceanic tidal dissipation, about 3×10^{12} watts. This is pleasingly close to an estimate made many years ago entirely on the basis of astronomic observations; namely, the departures of the Moon's longitude from those computed by Keplerian mechanics. The astronomic observations give only the total energy dissipated; they cannot tell whether this takes place in the ocean, the solid Earth, or for that matter on the Moon. We now know that it takes place largely in the oceans. This has an important bearing on reconstructions of the Earth-Moon history.

Here is a case in point where availability of computers has led, within a decade, to something resembling the solution of a problem. This is because the physical laws, embodied in the Laplace tidal equation (1775), were fairly well understood (apart from boundary dissipation) and could now be applied to a complex geometry. Quite the opposite holds for the atmospheric tidal problem. Here a spherically-symmetric geometry is probably adequate, but there has been a gross misjudgment as to the pertinent physical processes. The outstanding observational fact is the predominance of the solar 12^h tidal oscillation over the lunar $12^h.42$ in the ratio 15:1; gravitational theory favors the lunar tide 2:1. To account for this discrepancy a sharp resonance peak precisely at 12^h has been postulated, and a lot of theoretical effort has been spent to prove its existence. But now Siebert[88] and Chapman and Lindzen[89] have demonstrated (following a suggestion by Kelvin) that the solar tide is thermally driven, and the arguments for resonance amplification have vanished. The advance in understanding is comparable to that for ocean tides, but the role played by computers (though substantial) has been in an auxiliary capacity.

7. Spectroscopy. Computers have played a vital role in the analysis of geophysical observations. We consider only one aspect: numerical spectroscopy. This is most valuable for processes of such low frequency that analog filters are not readily available. For example, in the early days of spectral analysis of ocean waves, the recording was made on a film that could subsequently be played back at high speeds to bring the interesting frequencies into the resonance range of R.C.-filters. At the

present time, discrete sampling and numerical spectroscopy can be done with such efficiency that the numerical method competes with analog even at high frequencies. The numerical effort is necessarily extensive: a 1% resolution with 100 degrees of freedom required 10^4 data points; for multivariate processes the number is correspondingly higher. The increasing capacity of computers plus the introduction of more efficient algorithms has made it possible to keep up with most of the geophysical requirements.

Geophysics has profited in some vital ways. In some instances the numerical spectroscopy has led to the discovery of processes by the detection of unsuspected weak lines. In fields blessed with high signal/noise ratios (earthquakes, tides) the numerical spectroscopy has revealed the underlying natural background and thus the ultimate limits to detection and prediction. In other instances, the emphasis on high resolution has revealed the fine-structure of spectral peaks and with it some of the most interesting physics. This is particularly true for the normal modes of vibration of the Earth.

For certain broad-band stationary processes, turbulent motion in the atmosphere and ocean, internal and surface waves, the numerical spectroscopy has yielded some simple forms, $k^{-5/3}$, k^{-3}, ω^{-5}, ... Usually these had already been suggested from dimensional consideration, and the numerical analysis yielded the ranges, if any, over which the idealizations were applicable. Such analyses may yield simple displays of very complex time series.

One of the most interesting developments is the generalization of spectral analysis to bi-linear (and tri-linear) interactions, pioneered by John Tukey. The extent to which a frequency, $\omega = \omega_1 \pm \omega_2$, is excited by a quadratic sum- and difference-frequency interaction can be investigated in this manner. It is surprising that this bi-spectral analysis has not been more widely used; in the instances it has been applied it showed itself to be a tool of great power.

But perhaps the most important contribution of computers to the analysis of geophysical observations has been to discipline the observer. It has forced him to face up to the realities of the sampling theorem, to the unreliability of spectral estimates for finite time series. In a field where investigators remained notoriously ignorant of concepts long after thay had become familiar to their colleagues in optics, acoustics and electrical engineering, this was a welcome and overdue development.

8. Decision-making and early compaction. An increasingly important contribution by computers to geophysics is by an active, on-line participation in field experiments. One of the characteristic properties of the atmosphere and the ocean, as distinct from laboratory experiments, is the high degree of intermittency. Many of the most interesting events are episodic.

The known kurtosis in the distribution function of so many geophysical variables is an indication of intermittency. Mandelbrot attributes the "infrared catastrophy" of so many geophysical spectra to intermittency. Mollo-Christensen (the most enthusiastic proponent of intermittency) will go so far as to say that any theory that explains a geophysical process in terms of the average situation is likely to be wrong.

The implication of intermittency on geophysical data-taking is severe. Routine sampling at fixed data rates would seem to be just the wrong way to go about it: it will be highly redundant nearly all of the time, and inadequate at the rare interesting time. A fixed program of intermittent rapid sampling superposed on a standard low-sampling rate is an improvement, but does not really meet the issue. What is needed is a high rate of sampling conditioned by the rare important events. This is common sense, and is precisely the program followed when there is strong human involvement: special recording flights into severe storms, a tighter data grid over seamounts, etc., etc. This needs to be carried out automatically.

The following proposed experiment can serve as an example. Physical conditions near the deep-sea floor (temperature, current, pressure) are monitored by instrumented capsules. These are freely dropped, left on the bottom for a month to a year, and then recovered by acoustic command from a surface vessel. A miniature computer monitors the signals, and increases the sampling frequency during the occurrence of a high-frequency event. All this would be quite simple; most of the logical circuitry is required to have the computer make sensible decisions when the instrument malfunctions.

In a typical experiment of this kind, we record 10^5 data words of 12-18 bits each. Eventually the published results typically involve 10^2 numbers, whether in tabular or graphical form. Accordingly, there is a data compaction by a factor 10^3. At present all this compaction takes place after the return of the observations to the laboratory. Some of the compaction could be done by computers during the experiment.

There is then a question as to whether one uses on-line computers for early decision-making, or whether one relies on large memory banks and telemetering channels. We have come to the conclusion that the present technology favors decision-making over data storage and transmission. There are those who will argue that anything short of "complete" recovery of observations may lead to the overlooking of the unexpected, and therefore most crucial, results; that a preset program of data reduction will give you little more than what you already know. But there is an element of procrastination here, an unwillingness to think seriously about the reduction of data until the observations have been terminated. The many files of "original data" that overflow into the halls of the Scripps Institution, and that have never been looked at, would support this point of view.

The problem may be put another way: the information that is rejected in a preset decision-making program is still there, not in the file cabinets, but in the oceans and atmosphere (where it belongs). We will go after this information at some future time, again with the use of on-line computers capable of early data compaction.

What has been said here is particularly true in the study of earthquakes. Unattended instruments on the Moon and the sea floor can take advantage of on-line computers for selective sampling at critical times.

IV. DISCUSSION

Even within the limited scope of our review, it is apparent that the subject is very broad and that we have done it limited justice. A review of the first 25 years of achievements by modern computers has been a worthwhile exercise; we doubt whether this will still be the case at the end of the next 25 years. As time goes on, the subject becomes more and more like asking about the impact of telescopes on astronomy. Initially there were some effects that could be sensibly traced to these technological innovations; eventually they become so much ingrained in their fields that a discussion of the influence is nothing short of a monograph of the entire subject.

REFERENCES

1. Problems of Mathematics. Series 2, Conference 2 of Princeton University Bicentennial Conferences (1947).

2. Richardson, L.F., Weather Prediction by Numerical Process. Cambridge University Press, London (1922); Dover, New York (1965).

3. Charney, J.G., J. Meteorology, 6, 371-385 (1949).

4. Charney, J.G., Geofys. Publikasjoner, 17 (1948).

5. Rossby, C.-G., J. Mar. Res., 2, 38 55 (1939).

6. Charney, J. and A. Eliassen, Tellus, 1(2), 38-54 (1949).

7. Charney, J., R. Fjørtoft, and J. von Neumann, Tellus, 2, 237-254 (1950).

8. Charney, J., J. Meteorology, 4, 135-163 (1947).

9. Eady, E., Tellus, 1(3), 33-52 (1949).

10. Phillips, N., J. Meteorology, 8, 381-394 (1951).

11. Jeffries, H., Quart. J. Roy. Met. Soc., 52, 85-104 (1926).

12. Bjerknes, J., et al., Final Report General Circulation Project, No. AF 19(122)-48, U.C.L.A. Department of Meteorology (1955).

13. Starr, V.P. et al., Final Report General Circulation Project, AF 19(122)-153, M.I.T. Department of Meteorology (1954).

14. Kuo, H.-L., Tellus 3, 268-284 (1951).

15. Charney, J.G., Procès-Verbaux Séances de l'Assoc. de Méteor., Bruxelles, U.G.G.I., 47-63 (1951).

16. Phillips, N., Quart. J. Roy. Met. Soc., 82, 123-164 (1956).

17. Von Neumann, J., "Some remarks on the problem of forecasting climatic fluctuations" in Dynamics of Climate, ed. by Richard Pfeffer, Pergamon Press, 9-11 (1960).

18. Charney, et al., The Feasibility of a Global Observation and Analysis Experiment. Publ. 1290 of Nat. Acad. Sci., Washington, D.C. (1966); Plan for U.S. Participation in the Global Atmospheric Research Program. Nat. Acad. Sci., Washington, D.c. (1969).

19. Bolin, B., The Global Atmospheric Research Programme. Published by the World Meteorological Organization (1971).

20. Shuman, F.G., Office Note 72, NOAA Nat. Meteorological Center (March 1972).

21. Wiener, N., Proc. of Third Berkeley Symposium in Mathematical Statistics. University of Calif. Press, 247-252 (1955).

22. Lorenz, E.N., Transactions of the New York Academy of Sciences. II, 25, 409-432 (1963).

23. Lorenz, E.N., Tellus 21, 289-307 (1969).

24. Leith, C.E., J. Atmos. Sci. 28, 145-161 (1971)

25. Leith, C.E. and R. Kraichnan, "Predictability of turbulent flows". To appear in J. Atmos. Sci. (1972).

26. Charney, J. G. and A. Eliassen, J. Atmos. Sci., 21, 38-75 (1964).

27. Ooyama, K., J. Atmos. Sci. 26, 3-40 (1969).

28. Sundqvist, H., Tellus, 24, 6-12 (1972).

29. Anthes, R.A., NOAA Technical Memorandum ERL NHRL-97 (1972).

30. Smagorinsky, J., Quart. J. Roy. Meteorol. Soc., 79, 342-366 (1952).

31. Bjerknes, J. and H. Solberg, Geofys. Publikasjoner, 2 (1921).

32. Bergeron, T., Geofys. Publikasjoner, 5 (1928).

33. Stone, P.H., J. Atmos. Sci. 23, 455-465 (1966).

34. Williams, R.T., J. Atmos. Sci. 24, 627-641 (1967); 29, 3-10 (1972); Williams, R.T. and J. Plotkin, J. Atmos. Sci. 25, 201-206 (1968).

35. Hoskins, B.J., Quart. J. Roy. Met. Soc. 97, 139-153 (1971); Hoskins, B.J. and F. Bretherton, J. Atmos. Sci. 29, 11-37 (1972).

36. Manabe, S. et al., Monthly Weather Review, 98, 175-212 (1970).

37. Williams, G.P., J. Fluid Mech. 37, 727-750 (1969).

38. Pasta, J.R., Suppl. Vol., Computer Physics Communications, North Holland Publ. Co., Amsterdam (1972); Bivins, R.L., N. Metropolis, and J.R. Pasta, J. Comp. Phys. In press (1972).

39. Stommel, H.M., "Why do our ideas about the ocean circulation have such a peculiarly dream-like quality?" Privately printed (1954)

40. Sverdrup, H.U., Proc. Nat. Acad. Sci., 33, 318-326 (1947).

41. Stommel, H., Trans. Am. Geophys. Union, 29, 202-206 (1948).

42. Munk, W.H., J. Meteorology, 7, 79-93 (1950).

43. Charney, J.G., Proc. Nat. Acad. Sci. 41, 731-740 (1966).

44. Kamenkovich, V.M., Akad. Nauk SSSR Bull. Atmos. and Oceanic Phys., Am. Geophys. U. Trans. 781-792 (1966).

45. Robinson, A.R. and H. Stommel, Tellus, 11, 295-308 (1959);

46. Robinson, A.R. and P. Welander, J. Mar. Res., 21, 25-38 (1963).

47. Charney, J.G. and S.L. Spiegel, J. Phys. Oceanogr. 1, 149-160 (1971).

48. McKee, W.D., Some Topics in Dynamical Oceanography, Ph.D. Thesis, Cambridge University (1970).

49. Philander, S.G.H., "The equatorial thermocline", submitted to Deep-Sea Res.

50. Sarkisyan, A.S., Okeanologia, 11, 393 (1962).

51. Kamenkovich, V.M., T.G. Zhugrina, and M.M. Silkina, Akad. Nauk SSSR Bull. Atmos. and Oceanic Phys., Am. Geophys. U. Trans. 668-674 (1969).

52. Bryan, K. and M.D. Cox, Parts I and II, J. Atmos. Sci., 25, 945-978 (1968).

53. Bryan, K., J. Comp. Phys. 4, 347-376 (1969).

54. Cox, M.D., Deep-Sea Res., 17, 47-75 (1970).

55. Munk, W. and E. Palmén, Tellus, 3, 53-56 (1951).

56. Stommel, H., Deep-Sea Res., 4, 149-184 (1957).

57. Gill, A.E. and K. Bryan, Deep-Sea Res., 18, 685-721 (1971).

58. Rhines, P.B. - unpublished. See also J. Fluid Mech. 37, 161-189 (1969).

59. Lorenz, E.N., Met. Monographs 8, 1-3 (1968).

60. Lorenz, E.N., J. Atmos. Sci. 20, 448-464 (1963).

61. Charney, J.G., Proc. Symp. Appl. Math, 15, 289-310. Amer. Math. Soc., Providence, R.I. (1963).

62. Manabe, S. et al., Proceedings of Symp. of Physical and Dynamical Climatology, 16-20 Aug. 1971, Leningrad. World Meteorological Organization Technical Report, Geneva (1972).

63. Katayama, A., Y. Mintz and A. Arakawa, to appear in Proc. Amer. Meteor. Soc/International Meteor. Soc. International Symposium on Meteorology, Tel-Aviv and Jerusalen, 30 November - 4 December 1970.

64. Manabe, S. and K. Bryan, J. Atmos. Sci. 26, 786-789 (1969).

65. Miyakoda, K. et al., "Cumulative results of extended forecast experiments: Part I. Model's performance for winter cases." Unpublished. Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University (1972).

66. Thompson, P.D., Tellus, 9, 275-295 (1957).

67. Novikov, E.A., Akad Nauk SSSR, Geophys. Ser., Amer. Geophys. U. Trans. 1209-1211 (1959).

68. Onsager, L., Nuovo Cimento, 6, Suppl. 6, 279-287 (1949).

69. Lee, T.D., J. Appl. Phys. 22, 524 (1951).

70. Batchelor, G.K., The Theory of Homogeneous Turbulence. Cambridge University Press (1953).

71. Fjørtoft, R., Tellus, 5, 225-230 (1953).

72. Kraichnan, R., Phys. Fluids, 10, 1417-1423 (1967).

73. Lilly, D.K., J. Fluid Mech. 45, 395-415 (1971). See also, Fox, D.G. and D.K. Lilly, Rev. Geophys. and Space Phys. 10, 51-72 (1972).

74. Deem, G.S. and N.J. Zabusky, Phys. Rev. Letters, 27, 396 (1971).

75. Saffman, P.G., Studies in Appl. Math. in press (1971).

76. Orszag, S.A., Studies in Appl. Math. 50, 293-327 (1971).

77. Orszag, S.A. and G.S. Patterson, Jr., Phys. Rev. Letters, 28, 76-79 (1972).

78. Charney, J.G., J. Atmos. Sci. 28, 1087-1095 (1971).

79. Smagorinsky, J., Bull. Am. Meteorol. Soc. 50, 286-311 (1969).

80. Longuet-Higgins, M.S., Phil. Trans. Roy. Soc., London, A, 260, 317-350 (1966).

81. Accad, Y., C.L. Pekeris, Proc. Roy. Soc., London, A, 278, 110-128 (1964).

82. Hansen, W., Deut. Hydrog. Z. (Ergänzungsheft), 1, 1-46 (1952).

83. Gohin, F., Symp. Math. Hydrodynam. Meth. Phys. Oceanog., Hamburg, 179-197 (1961).

84. Hendershott, M.C., Proc. Symp. Math. Hydrodynam. Invest. Phys. Proc. Sea, Moscow, 8-21 (1966).

85. Bogdanov, K.T., and V.A. Magarik, Doklady Akad. Nauk SSSR, 172, 6, 1315-1317 (1967).

86. Tiron, K.D., Y.N. Sergeev, and A.N. Michurin, Vest. Leningrad Univ., 24, 123-35 (1967).

87. Munk, W., F. Snodgrass and M. Wimbush, Geophys. Fluid Dynam., 1, 161-235 (1970).

88. Siebert, M., Advances in Geophysics, 1, 105-182, Academic Press, New York (1961).

89. Chapman, S. and R. Lindzen, Atmospheric Tides. Science Publsihers, New York (1970).

Draft

THE COMPUTER IN MATHEMATICS

by S. M. Ulam

We all realize how short a time it is since the beginning work on the electronic computing machines and the work of Von Neumann which led to all the developments of which we will talk today. It is the main advantage and a characteristic of the computing machines: the time of operation is shortened so much. It is very hard from this short perspective to judge the full impact and the entire importance of the enormous amount of work produced by the computers, the change in the mode of thinking in the approach to numerous theories, and the perspectives opened by the possibility of conducting this mental experiment on such a gigantic scale.

I would like to present to you some reminiscences of conversations with Von Neumann which took place before the war, I think in 1938, and in particular a conversation which we had, where he told me about discussions he had with Norbert Wiener on the problem of turbulence, and how one could study such a problem by massive numerical work.

In those days, of course, the knowledge of the difficulties and the way to approach, especially, the theory of the onset of turbulence in the motion of incompressible fluid, was even much more restricted than it is today.

After the discussion with Wiener, Von Neumann expressed his curiosity about the role of Reynold's number, and why this mysterious dimensionless number should be of the order of 2000 - which it is - a most peculiar number for a mathematical situation. Wiener and Von Neumann discussed two different approaches, which nowadays we would call the approach by analogue machine or by digital computer, Von Neumann favoring the digital and Wiener the analogue way to attack these particular questions.

In fact, if I remember correctly, the expressions which were used at that time were that Wiener stressed more the possibilities of machines operating as he believed the human brain operated, with "hormonal" or continuous media or fluids mechanisms, yes mechanisms, even though the parts were continuous media, and Von Neumann thinking in terms of what would be called now certainly, the digital, the network combinatorial type of possibilities.

Another memory is of a discussion which I had with Von Neumann during the war in Los Alamos about the way to calculate the course of an implosion. The hydrodynamical problem was difficult. There were ingenious short cuts and theoretical simplifications which he suggested and which his collaborator or assistant Calkin tried to execute, which were to me really very inadequate. The questions concerned the value of certain number, compressions, etc., which had to be known let's say within 10%, and the simplifications were of a nature which could not guarantee even factors like two and three in the answers obtained by such methods. In this discussion, I remember, I was representing the point of view of pure pragmatism and heuristic attempts to get by massive numerical work and brute force, as it were, an idea of the course of the development of the motions, and even so the accuracy would not be too great with the - at that time - conceivable methods, rather than the arithmetical models as we now have them in the present electronic computing machines. In other words, the duality between the "hormonal" point of view of continuous "non-active" models, and the methods of discrete or network combinatorial approach.

Already at that time, in Los Alamos, we discussed also the possibilities of using the computers heuristically in the hope of obtaining insights in some questions of pure mathematics. By producing examples and observing the properties of special methematical objects one could hope to obtain hints as to general behaviour and perhaps ideas of how to construct proofs of general statements. I remember proposing a calculation of a very great number of primitive roots of natural numbers so that by observing and pondering the distributions one could get enough material about the appearance of the combinatorial behaviour and get ideas of how to state and prove some possible general regularities. I do not think that up to the present time, this particular item has been much furthered. Some fragmentary lists of this particular number-theoretical function were obtained more recently in Texas, if I remember correctly.

Quite generally, the computing machines have proved useful as could of course be expected, in obtaining the average values, or means, of functions in combinatorial analysis.

I will give just one small example:

It is known that given a permutation of $n^2 + n + 1$ integers there will exist a subsequence of integers of a length at least $n + 1$ of either increasing or decreasing numbers. This is a guaranteed minimum. One could ask the question of what the average value of a maximal length of a monotome subsequence is for a random permutation. A student of mine (Ed. Neighbor) investigated this question by producing enough statistics for a number of permutations for a variety of values of n. The results indicated very strongly that the average value of the maximum monotone subsequence is about 1.7 ... times the guaranteed minimum.

The point of this small example is to illustrate the fact that even though a complete study of all the possibilities in a combinatorial problem is prohibitive, sometimes for values of n as small as say 10, a sampling of a moderate number of special cases will with great probability indicate the general behavior. This was one of the motivations for the so-called Montecarlo method used in so many problems calculated on computers.

Needless to say, a great number of purely combinatorial material has been gathered since by experimentation.

All this of course is, even in this restricted use of computers, merely a beginning, and as I hope to indicate further on, the heuristic approach in some problems of pure mathematics may become greatly expanded, and of more basic or fundamental importance. This as a result of the Gödel discovery of undecidability in formal systems and in particular, it is suggested by the more recent discoveries that some well known concrete mathematical propositions, as open problems, belong to this class.

The computing machines came into existence through a confluence of developments both purely scientific and technological. The work in mathematical logic, the foundations of mathematics, and the detailed study of formal systems in which Von Neumann himself played such an important role, on one side, and on the other the rapid progress and discoveries in electronics made possible the construction of electronic computing machines. The quantitative speed of their operation, so much greater than that of the mechanical relay machines produced a qualitative change and improvement in the use of this tool.

It was Von Neumann's feeling for and his knowledge of the details of a formal system that enabled him to conceive a flexible program on an electronic computer, so that on the same machine with fixed connections, by suitable flow diagramming and programming an enormous variety of problems could be calculated and studied.

A third scientific development of capital importance has taken place since the first construction of electronic computers, and in fact, essentially after Von Neumann's death: This is the great series of discoveries in molecular biology with our present understanding of the importance of linear codes and programs in the DNA chain, and the beginning of an understanding of the functioning of a living cell.

It seems to me that the ideas one will obtain by attempting to utilize, by analogy, the living processes will lead to the next great stage in the development of computers, or more properly, more general automata.

The next item, or rather a large part of my talk, will concern the added importance, the added vistas opened by recent work of the last ten years or so.

Let us examine some possibility of enlarging the scope of application to pure mathematics of existing computers. It seems to me that by introducing additional devices, enabling the machines to perform a number of certain operations simultaneously or in parallel ways will permit us to investigate combinatorial and algebraic structures in ways vastly more general than at present. What I have in mind are constructions which would allow the machines to produce the composition of two functions, f and g, by obtaining f(g) directly and so to say all at once instead of computing point by point, somewhat as follows:

By a system of connections built on a finite checkerboard where the graphs of f and g (both defined on a finite range with values on the same set) could be combined to produce the graph of the composed function in one operation. It is obvious from the study of recursive functions how powerful the operation of superposition really is. On a grid of say thousand by thousand points, a great deal of heuristic study could be performed, if one produces also orders or devices to perform other operations all at once.

One could establish a way to obtain, again through just one order, the projection on the axis of a set in the two dimensional lattice array - and vice-versa construct a direct product of subsets of the axes. In this fashion one might have the possibilities of investigating - to be sure at first in a restricted way - the action of the operation of quantifiers, the algebraic and combinatorial properties of "projective" or cylindrical algebras. Under suitable extrapolation, the experiments on a finite range would then open vastly more general possibilities than are utilized at present with restriction of the functioning of computers to essentially only Boolean algebra and arithmetical operations. Again, as so often in mathematics, it is the interpretation of the results of finite algorithms that leads to new insights and applications. I suggested such things some twenty years ago in a report published by the IBM Company.

In the summer of 1952 or 53, in a discussion with Fermi who had shortly before learned how to program calculations by himself, we talked about new problems which could be made on electronic computers.

We discussed the possibility of a systematic investigation of problems which could not be meaningfully studied by techniques of classical analysis, the solutions being not only unobtainable in a closed form, but beyond that even, their qualitative features, i.e. some simple functionals of the solutions, not calculable by the usual approximative methods.

These discussions lasted several hours and we managed to produce a whole list of problems which we planned to study by numerical work on the electronic computer which was recently available in Los Alamos. The problems were increasing in difficulty, although a criterion used in their selection was the simplicity of statement and the probability of meaningful conclusions which could be drawn from the results of numerical work.

The first problem we decided to study was one of the simplest possible non-linear equation: the vibration of an elastic string which the ends kept fixed and where, in addition to the linear force representing Hooke's law, there was a small quadratic or cubic term representing a correction to the "true dependence" of the force on the displacement. This problem was chosen for one reason; both of us had interest in the long term asymptotic ergodic problems in classical physics. One of the first papers written by Fermi concerned the ergodic behavior of mechanical systems, and in 1938 Oxtoby and I showed that most continuous measure preserving transformations are metrically transitive, a result much stronger than a mere existence of such on every manifold.

Fermi's wonderful knowledge and insight into the properties of wave motions guaranteed that suitable conclusions or extrapolations from results obtained by computation would be forthcoming for an ever increasing generality of such non-linear physical situations.

One of the aims of the calculation was to see how, due to the non-linear term, the originally periodic motion would become steadily more complicated and to observe how and at what rate the shape of the string and the temperature behavior would become increasingly tangled in an approach to "thermalization."

The result of the calculations were a complete surprise: instead of an increasing complexity, a curious second order periodicity appeared to govern the long time behavior of the vibrating string. After some hundreds of would-be linear periodic motions, the string which we started in a single sinusoidal shape, acquired indeed contributions from the next few sinusoidal modes, but as time progressed, the higher modes did not participate in the share of energy, and indeed after a longer time, the string returned to its original sine shape! Far from moving through most of the available phase space, an extremely small part of it was gone through, a very unergodic behavior. Fermi confessed to be most surprised and was most interested in these results.

The calculations were performed for a number of cases. A Los Alamos report by Fermi, Pasta and myself gives a first account of the observed phenomena. It was the intention of Fermi to report on these results and speculate about their significance in the Gibbs Lecture he was invited to give by the Mathematical Society, but which his premature death prevented.

The intriguing results of this numerical work became more widely known and a whole series of investigations were stimulated by these calculations on the electronic computer. Some of the first theoretical work was undertaken by Kruskal and Zabusky. Papers by Ford and others attempted to explain the peculiar non-ergodicity.

It is impossible for me in this short talk to give any account of the subsequent work by American and Russian mathematicians (Chirikov, Sinai and others).

The discoveries of the Soliton or Quasi Eigenstates in analogous non-linear problems by Kruskal and Zabusky and the results of Peter Lax indidate interesting possibilities of applications, perhaps even for non-linear generalizations of Schrödinger type equations.

I mention all this to show on a mere example what value and guidance may be obtained from a suitably chosen numerical exploration. Other examples could be given and perhaps Professor Atiyah may express his ideas about possibilities of pioneering value of the study of special cases in several fields of analysis and geometry.

In a similar spirit I have undertaken a study of the properties of iterations of simple quadratic or cubic transformations of the Euclidean plane, three-dimensional spaces and spheres in a low number of dimensions. Again the results show a surprising and bewildering variety of behavior. One of the first accounts of such work is in a report by P. Stein and myself.

To indicate very briefly one of the motivations for the study of quadratic (or higher degree) transformations in the n-dimensional space, I will draw attention to certain generalizations of branching processes.

The usual branching process is formalized by a treatment of a number of particles - in general of a variety of types - which give rise in subsequent generations to new particles. This according to special rules which assume that the process consists of mitosis (or splitting) in most cases studied, independently of each other.

In the presence of binary interactions or sexual generation of new particles, or indeed in the study of reactions involving "mass action", or in miosis, the production of new elements and their absorption, etc., now depend, not linearly, on the number of elements of each kind, as in the classical branching processes, but through quadratic or higher degree expressions.

Combinatorially, the representation of the process is not given any more by a (multicolored) tree, but by a more general graph with loops between pairs ("Pair Trees").

In more realistic situations, the growth of population or of a collection of elements (or of an organism), depends on other perhaps external constraints or stimuli.

Only the simplest mathematical schemata have been attempted so far, to mirror mathematical properties of such geometrical growth given by recursive recipes. An account of my own simple-minded tries can be found in a recent book edited by Arthur Burke, a collection of essays entitled A Theory of Cellular Automata, University of Illinois Press.

A specially ingenious set of rules was devised by the English number theorist Conway. The Conway "Game of Life" is an example of a pastime which, perhaps much like the early problems involving dice and card games have led ultimately to the present edifice of probability theory.

The computers are of course absolutely essential in performing such experiments and in following games through more than a few beginning moves. I believe that the experiences gained by the following of processes of this sort will now have a more fundamental importance towards whatever will generalize or even replace in mathematics our exclusive immersion in formal systems.

Recent results of Cohen and others on independence of some of the most fundamental statements, the results of Novikoff, Wang, Matiasiewicz and others, indicate for the future a new role for pragmatic approaches. Theorems or problems will summarize our experience which work on computing automata will permit to enlarge.

It seems to me that the impact and great role of the electronic computing machine will significantly affect mathematics, after it has done so in the "mathematical sciences," i.e. in physics, in astronomy and applications.

The great dichotomy in the origin and inspiration of mathematical thought - on one hand coming from the stimulations of the external real physical universe, on the other from the developing processes of the human brain - has in a small way, a homomorphic image in the present and future use of the computers.

Even the most idealistic point of view about mathematics has to be reconciled with the fact that the <u>choice</u> of definitions and axioms of geometry - and in fact of most mathematical concepts - is the result of impressions obtained through our senses from external stimuli.

The computing machines promise to enlarge very greatly both the scope of the "Gedanken Experimente" and the idealizations of experiences. One has to remember, for example, how the theory of probability came about as a development of a few puzzles concerning games of chance.

It appears that experimentation on models of games played by self-organizing matter and of chemical reactions in living organisms will promote a study of novel mathematical schemata.

The beginning study of the mathematics of growing patterns, briefly indicated above, and the possibility of studying experimentally contests between geometrical motions perhaps in a higher number of dimensions might promote new little branches of mathematics, which could perhaps be given names like "paizonomy" for combinatorics of contesting reactions, and "auxology," for a yet to be developed theory of growth and organization, this latter including the growing tree of mathematics itself.

x x x

WORKING PAPERS OF

THE CENTER FOR RESEARCH ON SOCIAL ORGANIZATION

Department of Sociology

University of Michigan

Paper #74
Charles Tilly
April 1972

Copies available through:

The Center for Research
On Social Organization
214 Perry Building
330 Packard
University of Michigan
Ann Arbor, Michigan 48104

COMPUTERS
IN HISTORICAL ANALYSIS

Charles Tilly
University of Michigan
April 1972

for presentation to a conference on "the computer and the development of science and learning," Institute for Advanced Study, Princeton, New Jersey, sponsored by the Institute and <u>Daedalus</u>, 6-8 June 1972.

A Computational Memory

An image of my early days as a graduate student sticks in my mind after more than twenty years. Half a dozen of us are standing around a clanking, whirring machine in a harshly-lighted basement room. There in the middle is sociologist Samuel Stouffer, talking fast, cigarette swinging from his mouth, ashes showering his vest. Stouffer grabs a deck of punched cards, shoves them into the hopper at one end of the machine, pushes a button, and watches the cards sort themselves into glass-topped bins. He peers at the size of the various piles. The he says, "OK. Now let's try breaking on religion." He whips each stack of cards from its bin and slaps it onto the glass above the bin. Then he grips the first stack, fans it out, drops it into the hopper again, taps to straighten the cards, sets a weight on top of the stack in the hopper, turns a crank, and pushes the button again.

We, the graduate students, were learning a crude but serviceable way of analyzing data. It consisted of translating arguments about social phenomena into statements about variables and units of analysis. (The units were most often individual survey respondents and the variables their responses to standardized questions, but neither was essential to the logic we absorbed with the cigarette smoke.) Then we were supposed to identify the variable to be explained, cast our explanations in terms of differences with respect to other variables, represent obvious alternative explanations as "control" variables, and then carry out a cross-classification of the units which would reveal whether they did, indeed, vary as expected.

We might, for example, have tried to determine whether people living on farms actually said they wanted more children than people in big cities, once we "controlled," or held constant, the present family positions of respondents in the two populations. The machine was not indispensable to this sort of analysis. In principle, one could do it with pencilled tallies or with slips of paper sorted on a large table top. (I have done both things myself in emergencies.) Compared with the other alternatives we knew, however, only the machine made the analysis quick and practical.

The clanking old machine was not, of course, an electronic computer. It was a counter-sorter, a contraption in which electrified metal brushes responding to the presence or absence of holes at different positions in a Hollerith card activated gates along a belt on which the cards were moving, thus shunting the cards into one bin or another, and counted the cards shunted into each bin. Although people who work with punched cards still use the quicker, smoother descendants of our old basement machine in getting their data ready for the computer, the counter-sorter has practically disappeared as a tool of analysis in the social sciences.

Still the memory is useful. It sums up the predominant experience of social scientists with analytic machines until very recent years: the machine has greatly increased the feasibility of procedures which were actually invented without reference to machines, or at least without any necessary connection to them. The analytic inventions occurred more or less independently of the existence of counter-sorters or computers. The diffusion of those inventions, however, depended very closely on the availability of the machines.

Something else about that basement scene needs attention. Although

Sam Stouffer taught us to admire the classic deductive progression from general theory to specific hypothesis to empirical test against real-life observations, he was a wizard of *post hoc* interpretation of findings. In the jargon of the time, his "let's try breaking on religion" meant dividing the sample into Protestants, Jews, and so on, in order to see if any new differences, or any new explanations of the old differences, showed up. We learned that part, too. The nature of the data and the implicit or explicit theories the investigator is working with will set limits on how many *ad hoc* explanatory variables he will introduce in this way. Nevertheless, the vaguer and the more variegated the arguments at the investigator's disposal, and the easier the introduction of one more explanatory variable, the greater the likelihood that he will fashion and attach importance to spurious explanations.

The counter-sorter and the tabulating equipment that came with it heightened this risk somewhat by comparison with hand tallying and other such primitive procedures they replaced. Their net effect may still have been to reduce the number of spurious interpretations being seriously entertained by social scientists, because they came into wide use when theories were superabundant, determinate findings relevant to those theories quite rare, and cross-checking of doubtful explanations extremely hard to do.

Computers and Social Scientists

The computer compounds the risk. The capacity to absorb, store and manipulate large bodies of data, the easy introduction of complex, pre-packaged statistical routines, the speedy efficiency of the machine (at least in its good days) and the tiny incremental effort ordinarily required to introduce one more analysis, one more variable or one more observation all reduce the cost of running vast exploratory analyses. If the self-discipline of social scientists and the determinacy of their theories do not increase at the same rate as the expansion in computing facilities available to them, the net effect is likely to be a growth of the number of spurious explanations of social phenomena having currency.

Depending on one's view of the intrinsic possibility of reliable knowledge about human social behavior, this worry may sound cynical, or it may sound irrelevant. Yet it follows pretty closely from the characteristic relationship of social scientists to their data. The directors of most university computing centers have, for example, had to deal with it one way or another.

So far as I can tell from personal experiences, the university-wide computing facilities which sprang up in the 1950s tended to organize around several interesting assumptions concerning their clientele: 1) that the users were primarily interested in computation as such, rather than tabulation, compilation, reordering of large files, preparation of descriptive maps and graphs, content analysis of texts, and a variety of other uses to which computers have sometimes been put; 2) that the users would arrive with step-by-step descriptions of the computations they wished to perform, and would want to translate each step into a command the machine could follow; 3) that the work involved extensive manipulations of relatively small volumes of input data. These assumptions appeared, among other places, in the early emphasis on short courses in FORTRAN (= FORmula TRANslation) as the all-purpose preparation for computing, in the physical arrangement of input and output areas to accomodate clients who would ordinarily submit thirty or forty cards of data and programs, and in the great shortage of facilities for plotting, mapping, reordering files, reading of texts and the like.

Computation, formula translation and low ratios of input to analysis do not describe the usual situation of the social scientists who were coming to the computer in the 1950s and 1960s. At the beginning, most of us were interested mainly in the sort of reordering and tabulation of data we had previously done with machines like counter-sorters. We knew little mathematics, and found it difficult to transcribe the simple statistical operations we had learned to do with pencil and paper, slide rule or desk calculator into logical sequences comprehensible to machines. Not much computation there, nor much preparation for programming an entire analysis.

What is more, the computer-bound social scientist most often wanted to do something simple but cumbersome to a large body of data: a rank ordering of all counties in the United States on per capita income, and then on the proportion of the labor force engaged in agriculture, perhaps, or a cross-tabulation by occupation and region of the stated party preferences of 2,172 persons in a national sample survey. In short, a high ratio of input to output and a very high ratio of input to computation. A good deal of the extensive development of computing facilities for the social sciences during the 1960s went into redressing the balance: extending the non-computational capacities of the computer, packaging and simplifying the commands required to get the computer to carry out big but standard routines, and increasing the capacity of computing systems for input, storage, output and transmission of large bodies of data.

No doubt my description applies least well to the economists of the 1950s and 1960s; although their analyses of national income and related phenomena required simplification of large masses of data, they were already quite accustomed to applying complex models to small numbers of observations. At the other extreme stood the historians. Even the minority who identified themselves with the social sciences, even the smaller minority who were working with material which could, in principle, be placed in machine-readable form wanted lists, catalogs, tabulations and other convenient descriptions of their data rather than any systematic analysis of patterns, relationships and correspondence to models. The first textbooks on quantification, statistical analysis and computing for historians only began to appear, after all, around 1970. My description probably holds best for sociology and political science, and loses accuracy as we move away from them.

Now, I am not naive enough to think that the puny demands of university social scientists alone produced the changes in computing which occurred after 1960. Some of them resulted from the increased use of the computer by branches of the natural and physical sciences like ecology, whose main need was for the simplification and ordering of large bodies of observational data. More, I think, resulted from business and military applications, which are often quite similar in form (if not in content) to social scientific analyses. In any case, during the later 1960s American social scientists in big centers of research finally began to have at their disposal computing facilities that were well suited to the characteristic approach of social scientists to their data.

The Historical Social Sciences

Let us look more closely at the experience of the historical social sciences. I mean the conglomeration of specialties which have three things in common: 1) they concentrate on human social relationships, 2) they deal with change over a substantial succession of particular times, 3) their procedures yield conclusions which are generalizable, at least in principle, beyond the particular cases observed. Those specialties don't group together in any one major discipline. They exclude most of sociology, even more of anthropology, almost all of psychology, much of economics, perhaps less -- surprisingly enough -- of geography, demography and political science. Demographic history, econometric history, historical studies of social mobility and long-term analyses of the determinants of political participation illustrate what I have in mind.

Note some common features of this brand of inquiry. First, it commonly involves the systematic accumulation of numerous more or less uniform individual observations into a general portrayal of the phenomenon under study. Where the observations concern individual people we call the procedure collective biography or prosopography; where they concern firms, families, communities or other units, we have no standard term; but we might call the whole approach collective history. While many American historians have been willing to let the observations of a Tocqueville or the life histories of a few entrepreneurs stand as their evidence of the opportunities for mobility in the nineteenth century, a number of younger historians have recently been insisting on the study of occupational mobility for the entire population of whole communities. The most ambitious venture of this kind so far has been Stephan Thernstrom's person-by-person examination of Boston's adult male population from 1880 to 1970 (Thernstrom 1972). (Among other things, Thernstrom discovers great stability in the rates of movement from manual to white-collar jobs and in the propensity of new migrants to stay in the city, despite large changes in the rate of in-migration, and an important shift from greater long-distance mobility among manual workers in the nineteenth century, to greater long-distance mobility among white-collar workers today.)

Second, this sort of investigation ordinarily includes systematic comparison of standard units -- populations, areas, periods, or something else -- with respect both to the phenomena to be explained and the explanations proposed for them. Thus in their analysis of the rural uprising of 1830 in England, Eric Hobsbawm and George Rudé tabulate villages which rioted or did not riot by whether land had recently been enclosed, how much of the village population was non-agricultural, and so on. (As it turns out, recently-enclosed and semi-industrial villages appear to have had the higher propensity to riot; altogether, the findings lend weight to explanations of the revolt in terms of the defense of particular local rights to land and work, and cast doubt on explanations in terms of spontaneous reaction to short-run economic crisis.)

Third, the kind of inquiry I am identifying with the "historical social sciences" tends to rely on the explicit statement of concepts, hypotheses and models, as well as the self-conscious matching of the observations to them. In his influential study of marketing in rural China, G. W. Skinner begins by laying out the logic of the central-place theory often employed by economic geographers, proceeds to show that the timing, interdependence and geographic distribution of local and regional markets in pre-communist China fall into the patterns anticipated by central-place theory, argues that the market system provided the framework for a wide range of other activities not obviously related to marketing -- for example, the choice of marriage partners and the gathering of peasants for recreation -- and finally points out the persistence of the fundamental patterns past the revolution of 1949. Skinner's studies launched a whole fleet of studies of Chinese marketing patterns and their correlates.

Finally, quantification. Neither the aggregation from individual to total population, the reliance on systematic comparison nor the explicit confrontation of fact and theory is intrinsically quantitative. Yet all three are obviously hospitable to quantification in ways that many approaches to history are not. A case in point: David Herlihy seeks to learn whether, as some theories would lead us to expect, prosperity encouraged the Tuscan population of the fifteenth century to marry younger, have more children, and form larger households. He has enormous documentation at his disposal -- notably the catasto of 1427, a document resulting from the effort to enumerate and describe every single one of the 50 to 60,000 persons eligible to pay taxes in the territory then controlled by Florence. We are not the least surprised to find him casting the crucial questions in quantitative form, by calculating fertility rates, mean household sizes, and the like. (Nor are we surprised to find that computers are doing a major part of the routine work, but let's save that for later.) Once he is committed to checking theory against fact by aggregating thousands of individual observations into comparisons over space, time and social category, it is hard to imagine how Herlihy could proceed without quantifying a number of his main arguments. Yet dozens of historians before him have written histories of Tuscany and of Florentine families without a trace of quantification.

Current Counter-Currents

My reliance on positive examples may obscure this important fact. Most inquiry into history does not fit my description of the historical social sciences. Most historians are not doing collective history, are not making systematic comparisons of standard units, are not self-consciously building models and confronting them with historical fact, are not casting their arguments in quantitative form. Nor do they want to. They are, on the whole, proud to be doing something else. They are often distressed that anyone should be building models, quantifying, and so on. Since the computer is quite unlikely to spread into those areas of history in which investigators lack or reject the habit of putting part of their work into quantitative form, the prospects that the computer will revolutionize historical analysis as a whole in the near future are slight indeed.

The faint possibility of such a revolution has nevertheless called forth indignant roars from some of the profession's strongest voices. The statements on the subject tend to confound computerization with quantification, to trot out examples (real or imagined) of trivial, illegitimate and/or misleading quantitative analysis, and then to call for a common defense against the Huns. The enduring objections, however, do not appear to center on existing misuses of computers and quantification. Instead, they concern the possibility that quantitative historians will abandon the humane depiction of real, whole persons, mistake their statistical results for the reality and -- worst of all -- communicate their delusions to other historians.

Statements of this view are often written with passion and brilliance. One of my favorite specimens of the genre comes from Richard Cobb, the superb portraitist of cops, spies, criminals, rioters, revolutionaries and ordinary people of the revolutionary era in France. At one point in a recent essay called "Historians in white coats" he describes an investigation now proceeding in the United States as:

> the computerization of 516 urban riots, turbulences, disturbances, fracas, prises-de-barbe, semi-riots, revolvérisations, lynchings, stabbings, slaughters, massacres, protests, collective threats, abusive slogans, provocative songs, in France, for the whole period 1815-1914. The end-product will no doubt reveal some highly interesting pattern: that, for instance, market riots occur on market days, on or near the market, that marriage riots take place after weddings, that funeral riots take place either outside the church or near the cemetery or along the course of a funeral procession, that Saturday riots take place on Saturday evenings, between 10 and 12 o'clock in the winter and between 11 and 1 o'clock in the summer, that is after the wineshops and bals have closed, that Sunday riots take place after Mass, that rent riots take place on rent days and that they are commoner in April and July than they are in January and October, that port riots take place on or near ports, that recruitment riots converge on railway stations or on barracks, that prison riots take place inside or opposite the prison,

> or both, that religious riots, especially in towns or bourgs in which there exist two or more antagonistic religious communities, favour Sundays, Catholic feast days, or St. Bartholomew's Day, or the Passover. Perhaps we thought we knew already; but now we *really* know; we have a Model. Riot has been tamed, dehumanized and scientificated (Cobb 1971: 1528).

I had some trouble recognizing my creature in motley. But once I realized that I was the originator of the investigation in question, I read Cobb's account with fascination. It proves him a master of historical fiction. Every detail is invented -- the time-span, the number of events, the kinds of action covered, the questions asked, the answers given, the whole point of the study. (In fact, the study runs from 1830 to 1960, deals mainly with large-scale collective violence, and consists of a number of different efforts to determine the impact of urbanization, industrialization and political centralization on patterns of collective action and struggles for power in France, not to mention related work concerning other West European countries.)

Yet some resemblances to the original are interesting, even disquieting. It is true, for example, that a number of the results of such an inquiry are bound to be trivial, and others more or less self-evident after the fact. No one is stunned to discover that labor unions became more heavily involved in major French conflicts toward the end of the nineteenth century than they had been fifty years before. Unions had, after all, only existed in the shadows until their legalization in the 1880s.

The main reason for pursuing results which will appear obvious in retrospect is that not all of them are obvious in prospect. Some fly in the face of widely-held opinions. In the investigation at hand, the widespread small-town participation in the rebellion of 1851 against Louis Napoleon's coup d'état (when provincial France as a whole, sickened or disappointed by the course of the 1848 revolution, is supposed to have lapsed into apathy or conservatism) makes us rethink the whole process of political mobilization and demobilization in that period. Other findings help discriminate among several alternative readings of a process, each of which is plausible, and therefore obvious in retrospect. Should we, for instance, expect crimes against persons and collective violence to vary together or to follow distinctive patterns in time and space? The latter is the case in modern France. But if we had found the former there would have been plenty of common-sense rationalizations and sociological theories to make the findings self-evident.

It is also true (as Cobb indicates elsewhere in his essay) that this sort of inquiry is expensive, requires the organization of a research team, and relies relatively little on the traditional lonely encounter of one man with one document. It is true (as Cobb's reasoning suggests, despite the fact that his blurred vision of computing doesn't allow him to pick out the details) that the combination of high initial investment and low marginal cost of additional items encourages the builder of a data file, once begun, to pack in all sorts of apparently useless information. It is true that the moving of some of these large machine-readable historical files into the public domain (which is beginning to happen now) facilitates the pursuit of bad hypotheses and meaningless correlations as well as sound hypotheses and meaningful correlations. It is true, finally, that the scale and complexity of such an investigation produce important periods when the researchers are so preoccupied with problems of coding, file construction, statistical procedure, computer techniques and coordination of the whole effort that they practically lose contact with the people, events, places and times they are studying.

These are genuine costs. Yet it strikes me as perverse to count the costs alone without considering the benefits. Fortunately, working historians pay little attention to exhortations on one side or the other. They respond instead to concrete examples of procedures for getting answers to questions they are already pursuing. The problem is simply to understand why spokesmen for the profession should so regularly emphasize the costs of computing without mentioning the benefits. The answer, I suppose, is that the critics consider the accumulation of systematic knowledge about human behavior either impossible, dangerous, of little value, or a serious diversion from other more worthy ventures.

Is History Computable?

It is not just that historians are usually impressionistic, belletristic or just plain cantankerous, although each of these is often the case. A number of traditional and legitimate historical problems simply do not lend themselves to quantification; they therefore remain unlikely prospects for work with computers. The intellectual gain from quantification in history generally rises with 1) the complexity of the models employed, 2) the importance of variation to the arguments at hand, 3) the number of units involved, 4) the ease with which the phenomenon to be explained can itself be put into quantities. Historians, however, often find themselves trying to account for a single act of a single person by means of some general assertion about that person's character or situation. For my part, I wonder whether there is any means at all of verifying or falsifying statements of that variety. In any case, quantification is not likely to be the means.

Historical work in general has a large component of description, interpretation of texts, reconstitution of sequences, imputation of motives to actors, making of single connections, offering of judgments, drawing of moral or political conclusions. In principle, machines could do some of these things well. In practice, these capacities of computers are developing only very slowly, and historians -- even historical social scientists -- are doing little to encourage their development. Harry Hanham said a couple of years ago that photocopying machines had to that point exerted a far larger influence on historical practice as a whole than had computers. If he said it now, he would still be right. Photocopying machines do quickly and cheaply something that most historians are already much involved in doing -- transcribing and collating texts. The everyday capacities of run-of-the-mill computer installations meet the existing needs of a far smaller group of historians. No substantial increase in the use of computers by historians is therefore likely to occur unless a) the kinds of problems and explanations with which ordinary historians concern themselves change substantially and/or b) the practical capacity of local computer installations to deal with textual analysis, cataloging, indexing, sorting, sequencing, summarizing and retrieving simply and cheaply expands to a large degree.

I consider the probabilities that either one will happen in the next ten years or so very small. There, of course, I could be egregiously wrong. The sorts of computer-based information systems now in use at Bell Telephone Laboratories could, in principle, give historians fast access to each item in a whole archive or a large library. With unlimited funds, it would not be hard to automate a good deal of the searching, storing and retrieving which consume such a large part of the average historian's time. It won't happen soon, in my opinion, because historians lack the power, the funds and the inclination to make it happen.

So we should distinguish between the computer in history as practiced by historians and in the historical social sciences as practiced by people from a wide variety of disciplines. It is in the historical social sciences that we should expect to find rapid increases, and some innovations, in the use of computers. Why? Because it is there that the incentives to quantification are strong, some of the essential facilities, resources and technical expertise are already available, and the attractiveness of anything that reduces the time, effort and unit cost involved in dealing with complex analyses and large pools of data is great.

Development and History in the Social Sciences

The historical social sciences provide an important object lesson in the responsiveness of scholars to changes in the world about them. The multiplication of new states at the end of World War II turned the interests of a wide variety of western scholars toward the elaboration of schemes intended to anticipate -- and perhaps even to guide -- the political, economic and demographic changes which would take place in the non-western world. The most popular of those schemes postulated standard paths and processes of "development." We had theories and programs for economic development, of course; but political development, demographic development, educational development, social development, urban development and still other purported standard processes also came in for a great deal of attention in the social sciences. The models for the developmental schemes came most often from readings of western history; Rostow's scheme for stages of economic growth, for instance, began explicitly with an interpretation of English experience. The developmental schemes proposed in the 1940s and 1950s all turned out to have great weaknesses, both in their own terms and as tools for the analysis, anticipation or guidance of changes in the non-western world.

The nature of those weaknesses need not detain us here. For present purposes, the important thing is that dissatisfaction with them drove a number of social scientists back to look more closely at the conceptions of the western experience which had, implicitly or explicitly, inserted themselves into available theories of development. This happened somewhat independently, and at different points in time, in economics, demography, political science and other fields; but it happened very widely. The net effect of the two moves -- first toward formulating theories of long-run development, then toward re-examining the fit between such theories and the historical record -- reintroduced long time spans into disciplines which had been concentrating rather heavily on the short run. As Julius Rubin describes the situation in economics while discussing Ester Boserup's Conditions of Agricultural Growth:

> It has been a very long time since the last simple, centuries-spanning theory of economic-demographic relations was proposed. After Malthus, Ricardo and John Stuart Mill, economists turned -- many with a sense of relief, no doubt -- from the problems of the ages to the down-to-earth, short-term analysis of a market economy whose success and stability could be taken for granted. The problems of the long-term, of analytic history, sank into the underground of economics, with rare exceptions neglected in the universities until the Second World War. And though after the war the renewed perception of economic development as a major social problem produced an immense amount of research and generalization, the great economic-demographic framework remained the same: neo-Malthusians are hard to distinguish from paleo-Malthusians. Mrs. Boserup has taken advantage of that research, particularly of the recent advances in our knowledge of agricultural systems, to suggest a modification of the

classical framework and has thereby irritated some economists, who are sceptical of all long-term theories and large-scale frameworks, while she has given hope to historically-oriented social scientists, who are badly in need of a new Ariadne's thread (Rubin 1972: 35).

Between John Stuart Mill and Ester Boserup, to be sure, a good deal of economic history got written. Yet economists (with the important exception of Marx and his followers) generally avoided economic history, and treated it as an inferior good. Only after World War II, with the new urgency and respectability acquired by the analysis of economic growth, did any substantial number of people trained primarily in economics turn back to the serious analysis of historical sequences, problems and materials. But then it happened in a big way. It happened in such a big way, indeed, that by the early 1960s economic historians who received most of their training in history found themselves pressed hard by youngsters who spoke a mathematical language, built models, tried to unearth the buried economic assumptions in older arguments concerning such phenomena as slavery, the building of railroads, or technological change in agriculture, and employed considerably different standards of evidence from their elders. What is more, the youngsters frequently used computers to collate their data or to perform their computations.

Thus important parts of economic history became econometric history, or "cliometrics." Economists were precocious in all these regards. But similar processes created new specialties at the meeting-points of history with demography, sociology, geography and other social sciences. In all of them there was at least one moment of sharp confrontation between the oldsters who were accustomed to offering comprehensive, sympathetic, narrative accounts of their material and newcomers with their models, their jargon, their numbers, their computation -- and, many oldsters said, their arrogance. Their ideas and procedures crystallized into new specialties: the historical social sciences.

Historical Demography as an Illustration

Historical demography illustrates the current situation of the historical social sciences. Demography has a reasonable claim to have been the first of the social sciences to take something like its contemporary western form. From its seventeenth-century emergence in England as Political Arithmetic, demography has recurrently dealt with historical materials and long spans of time. In the nineteenth century, however, the development of censuses and related means of collecting detailed data at particular points in time shifted the study of population away from historical concerns. As the standard data, procedures and theories of demography crystallized, they converged on short-run processes and on the comparison of different populations at the same point in time. Even today, two-thirds or more of the average demographic textbook deals with those ahistorical matters. The substantive chapters of Barclay's standard *Techniques of Population Analysis*, for instance, cover:

--rates and ratios

--accuracy and error

--the life table

--the study of mortality

--measurement of fertility

--growth of population

--migration and the distribution of population

--manpower and working activities

The longest span of continuous observation for any particular population discussed in the book, furthermore, is ten years. Demography crystallized as a non-historical social science.

Nevertheless, a nice dialectic was working. The very accumulation of censuses in the nineteenth and twentieth century and the very improvement in the measurement of fertility, mortality and related processes made it increasingly clear that western countries were undergoing long-run demographic transformations which could be plausibly related to the industrialization and urbanization of the West after 1750. In the 1920s and 1930s western demographers formulated the idea of a standard "demographic transition" occurring country by country throughout the world. In David Eversley's neat summary:

> This "theory'shows that countries go through various stages of population change: beginning with high birth and death rates allowing a low fluctuating rate of increase (if any), they pass through a phase of increasing death control which leads to very high rates of growth, and finally into the last stage where the pressures set up by fast growth produce some control of the birth-rate which results in a considerable slowing down of the increase. Though this is exactly what happened in all western countries some time between 1800 and 1900, and in Japan rather later, and is beginning to happen in some of the more affluent third world countries, it really tells us very little. The 'transition' may last 100 years, and indeed in some countries we do not yet know whether it will ever occur at all, or

> whether their problems will not after all be "solved" by Malthusian disasters (Eversley 1971: 1151).

The last two or three decades of effort to refine this argument, check its applicability to the actual patterns of change in particular western countries, verify the alternative explanations conventionally given for the declines in fertility and mortality, and judge whether and how such a transition is likely to occur in the rest of the world have shaken demographers' confidence in all simple versions of the theory. They have not yet produced an acceptable substitute. But they have stimulated an important series of investigations in the demographic history of England, France and a half-dozen other countries, mainly in western Europe. These investigations have created a new discipline: historical demography.

The computer played no important part in the creation of the new discipline. (As we shall see later, however, it is playing an important part in the discipline's current work.) Historical demography grew apart from demography, economics and history by refashioning elements drawn from each of them. The most significant elements, as I see them were 1) basic descriptive schemes and models of "stable populations" drawn from demography, 2) econometric tools adopted from economics for the purpose of testing the applicability of alternative models to actual historical observations, and 3) creation of new procedures for extracting demographic measurements from registers of births, deaths and marriages, old enumerations of population and other such bulky sources long known to historians but long neglected for lack of any effective way of exploiting them.

The invention of the procedure called "family reconstitution" probably made the largest difference. The most important contributions came from Louis Henry, a demographer at the Institut National d'Etudes Démographiques in Paris, who began with relatively little interest in history as such, but a strong desire to get at long-run population dynamics. "Family reconstitution" is one of those bright ideas which is perfectly obvious once stated. It consists of accumulating the individual, scattered records of births, deaths and marriages occurring in a locality into family dossiers relating the events to each other. If the registration is fairly complete, if the population doesn't move too much, and if it is usually possible to match a person mentioned in a given record with a family and with other mentions of the same person, the dossiers will yield tolerably good estimates of the vital rates prevailing in the population as a whole. The record of a wedding, for example, may not include the ages of the spouses. But if we also have birth records for them, we can calculate their ages at marriage. Again, if the bride has her first child two years later, we can calculate her age at first birth without difficulty.

To the extent that registration is incomplete, the population mobile and the identification of individuals uncertain, the job gets harder and the estimates become less trustworthy. The procedure is possible in important parts of Europe from the seventeenth century onward only because almost everyone invoked religious ceremonies for births, marriages and deaths, and the parish clergy kept comprehensive registers of the baptisms, weddings and burials at which they officiated. In those rarer places where civil registration was equally complete before the nineteenth century, it is of course possible to follow a similar procedure.

Family reconstitution by hand is tedious. It takes a long time. It requires numerous small judgments. It produces large files. And calculating vital rates from those large files is a fairly complicated operation. This is where the computer could come in. In fact, the extensive use of the computer in family reconstitution is just beginning. Although by now a few dozen scholars, especially in France, have reconstituted the populations of individual parishes over substantial periods of time, none of them seems to have done the major part of his work by machine. The two research teams which have gotten the farthest with the analysis of multiple communities are the one directed by Louis Henry at the Institut National d'Etudes Démographiques and the collaborative venture of Peter Laslett, E. A. Wrigley and R. S. Schofield at Cambridge University. The INED group does not employ computers for any of its main tasks. The Cambridge group is setting up its work for the computer; at this point however, it has not completed the reconstitution of a single parish by machine. A few other research teams doing related work in Tuscany, Québec, Iceland and Normandy have all reached about the same stage: having experimental runs or successful programs for part of the whole inquiry complete, but not having put into operation a true computer-based system for family reconstitution.

One of the more interesting difficulties in setting up such a system results from the uncertainty involved in matching different records with the "same" person. The difficulty appears in all sorts of collective history, not just in family reconstitution. In his study of Boston, for example, Stephan Thernstrom calls it the Michael Murphy problem. When dozens of Michael Murphys are born every year, and the supplementary information supplied with birth or marriage certificates is sparse, how do you decide which Michael Murphy got married twenty years later? What about misspellings, or variant spellings, of the same name: are Michael Murphy born in 1874 (birth record) and Michal Murphey, born in 1874 (marriage record) and Michael Murphey, 60 years old in 1935, the same person? Every one of the simple and obvious solutions is susceptible of introducing systematic errors into the analysis: dealing only with uncommon names, throwing away all uncertain matches, matching with the first plausible fit, matching randomly, and so forth. Every investigator faced with the problem so far has adopted some sort of hand solution which requires subjective judgment. Here is a problem worth giving to the computer: applying an explicit set of decision rules to all such matches, tagging each completed record as to the degree of certainty in its matching, identifying all unmatchable observations and their characteristics, calculating the possible effects of different kinds of matching errors on the demographic parameters being estimated from the whole body of data. In fact, most of the research teams which are using computers for family reconstitution or related operations are also working seriously on computer-based solutions to exactly this set of problems.

Does all this mean that the use of the computer for historical demography is all promise and no accomplishment? No. At this moment historical demographers are using computers to collate the material from huge sources like the Florentine catasto of 1427, to perform a wide range of time-series analyses for the purpose of detecting the relationships between demographic and economic fluctuations, testing models proposed for the explanation of regional differences in fertility, and dozens of other purposes. Few students now learning the specialty will enter their professional lives without some competence in computing.

Perhaps this will do as a sign of the times: in 1969 appeared the first major thesis in French history for which the basic quantitative work was done by electronic computer. The book is Marcel Couturier's study of Châteaudun (a city of 5 to 10,000 persons southwest of Paris) from 1525 to 1789. For that study Couturier punched about 16,000 cards, each representing a single registration of a marriage, a birth, an act of apprenticeship or some other crucial event in the life of a particular individual. He built the book along two dimensions: 1) fluctuations over time in the composition and dynamics of the entire city's population, 2) differences among major segments of the labor force in wealth, marriage patterns and a few other characteristics at particular points in time. He did not carry on family reconstitutions, calculate differential fertility, prepare life tables, or any of the other more complex demographic operations for which one might, in principle, employ the computer.

Couturier asked the computer for two main kinds of operation, one corresponding to each of the book's major dimensions. First, he asked for series of births, deaths, and other observations for the city as a whole, or for particular parts of it, over long periods of time. Second, he asked for simple two-variable cross-tabulations: occupational group of bride's father by occupational group of groom, and so on. The operations could have been performed -- more laboriously, to be sure -- on the old counter-sorter. On a smaller scale, French scholars have often performed the very same operations by hand. In short, Couturier employed his IBM computer as a large tabulator; that choice eased his statistical labors and probably made them more accurate, but it did not introduce any significant innovations into the actual structure of his work. Only with the current round of theses will we begin to see work in historical demography which only the computer makes possible.

In Sum

If we were to direct the same sort of survey to recent work in economic history, we would discover a routine use of more complex models, smaller data files, less emphasis on description and a higher ratio of computation to input and output. If we turned instead to the systematic historical study of elites, class structure and social mobility, the prevalence of essentially descriptive work would increase; we would find computers being used primarily for collating, sorting, aggregating and providing statistical descriptions of large number of individual observations. *Mutatis his mutandis*, the other historical social sciences have reached roughly the same point as historical demography.

In these fields, the computer's chief impact so far has been to decrease the effort and unit cost involved in procedures for which hand procedures were already well established, and thus to increase the scale of analysis to which those procedures could be applied. The critical innovations (like family reconstitution) have so far occurred more or less independently of the computer, but the computer is making their diffusion easier. In each field, a few investigators are demanding more of the computer: building complex files with extensive cross-referencing, testing complex mathematical models against large bodies of data, simulating social processes in order to rule out assumptions which produce implausible results, performing content analyses of texts. In each field, no more than a handful of investigators learn much about computing itself; the great majority content themselves with a knowledge of pre-packaged programs, with perhaps a rudimentary competence in FORTRAN, COBOL or some other standard language for emergencies. And many are willing to let others handle their access to the computer.

There lies a danger. In these days of the computer it is easy, tempting, and relatively cheap to run large statistical analyses which are appropriate neither for the data at hand nor for the arguments which the investigator is really prepared to make. So long as the computer is being used simply to collate and describe large bodies of data, the main costs of freehandedness are likely to be wasted effort, boredom and excessively long manuscripts.

When it comes to fitting statistical models to data, however, the potential costs are much more serious. The ease with which historical social scientists can run a hundred multiple regressions, carry out a large factor analysis or compare every vote in a given legislature to every other one makes it easy to coax striking pseudo-results from almost any substantial collection of data. The danger is that investigators will use the easy procedures to explore the whole terrain instead of to follow a map already prepared on the basis of well-reasoned theory and previous findings. If that were to happen without a compensating strengthening of our theories of social processes and without a complementary increase in our ability to articulate and test the more complicated notions we have about those social processes...then, alas, the many critics who see the computer as the harbinger of mindless empiricism would be right.

REFERENCES

Barclay, George W.
1958 Techniques of Population Analysis (New York: Wiley).

Cobb, Richard
1971 "Historians in White Coats," Times Literary Supplement, December 3: 1527-1528.

Couturier, Marcel.
1969 Recherches sur les structures sociales de Châteaudun, 1525-1789 (Paris: S.E.V.P.E.N.)

Eversley, David.
1971 "Populations and Predictions," Times Literary Supplement, September 24: 1151-1152.

Hanham, H. J.
1971 "Clio's Weapons," Daedalus, Spring, 1971: 509-519.

Henry, Louis.
1967 Manuel de démographie historique (Genève & Paris: Droz)

1968 "Historical Demography," Daedalus, Spring, 1968: 385-96.

Herlihy, David.
1969 "Vieillir au Quattrocento," Annales; Economies, Sociétés, Civilisations, 24: 1338-1352.

Hobsbawn, E. J. and George Rude.
1969 Captain Swing (London: Lawrence & Wishart).

Lowry, W. Kenneth
1972 "Use of Computers in Information Systems," Science, 175: 841-46.

Rostos, W. W.
1960 The Stages of Economic Growth (Cambridge: Cambridge University Press).

Rubin, Julius.
1972 "Expulsion from the Garden" [review of Ester Boserup, Conditions of Agricultural Growth], Peasant Studies Newsletter, 1: 35-39.

Shorter, Edward
1971 The Historian and the Computer (Englewood Cliffs: Prentice-Hall).

Skinner, G. W.
1964-65 "Marketing and Social Structure in Rural China," _Journal of Asian Studies_, 24: 3-43, 195-228, 363-399.

Thernstrom, Stephan
1972 _Yankees and Immigrants. The Process of Mobility in Boston_ (Cambridge: Harvard University Press, forthcoming).

Wrigley, E. A.
1969 _Population and History_ (New York: McGraw-Hill).

Memo from
JOSHUA LEDERBERG

To: Dr. Carl Kaysen
Institute for Adv Study

JUL 13 1971

von Neumann Symposium

This sounds like a most appealing affair; and if you have indeed recruited Sidney Brenner for the principal paper on the biological sciences, I will be happy to serve as discussant.

However, I would ask for a clear license to deal with somewhat broader issues of the applications of artificial intelligence, in addition to the specific applications in biology to which Sid will likely address his main remarks. My own experience has been mainly in the field of organic chemistry as an arena for exercising computers in ways that should be most pertinent to your symposium. I realize you do not wish a cookbook of recipes; but I would be at a disadvantage if I were constrained to ~~describing~~ sauces for veal when I was in fact a vegetarian.

Except for neurophsyiology, biology has been rather little influenced by computers (directly); and your very choice of Brenner (contra, e.g., a Ross Adey) encourages me to believe that you are not seeking a discipline-bound discussion.

Under separate cover, I will forward some papers that may clarify what I could offer in this field. Needless to say, please feel free to rescind your invitation if you should soon discover there had ~~be~~ been some cross-purposes.

Sincerely,

PROFESSOR JOSHUA LEDERBERG
Department of Genetics
School of Medicine
Stanford University
Stanford, California 94305

50143

THE INSTITUTE FOR ADVANCED STUDY

PRINCETON, NEW JERSEY 08540

Telephone-609-924-4400

THE DIRECTOR

July 9, 1971

Dear Professor Lederberg:

I write to invite you to participate in a two-day symposium in honor of the 25th anniversary of John von Neumann's achievement of the first modern electronic computer which will be held here at the Institute June 6-8, 1972. The topic of the symposium will be "The Influence of the Computer on Science and Learning." All in all, there will be eight papers on major areas of mathematics, the natural and social sciences, which have been strongly influenced by the computer.

Discussion on each topic will be organized around a paper prepared in advance and circulated amongst all the participants. Papers will not be read through at the symposium, but each author will be asked to introduce his paper. He in turn will be followed by an invited commentator, and then by general discussion in which all authors, commentators, and other invited participants will take part. The total number of invited participants is expected to be 30. In addition, the physical arrangements of the symposium will be such that there will be an audience drawn from the local academic community, and questions and comments from the audience will be possible.

The aim of the papers will be to present a broad survey of the influence of the computer as a tool in the way each discipline has come to conceive its possibilities, to select the problems it seeks to solve, and the modes of attack on them. While these topics cannot be discussed without reference to the specific technical details of how the computer is used, the purpose of the paper is not to recite a list of problems in each area that have been solved and the recipes that have been evolved for solving them, but rather to make a philosophical evaluation of the influence of the computer on the way the discipline has evolved in the last 25 years.

DAEDALUS, the journal of the American Academy of Arts and Sciences, is joining with the Institute in sponsoring the conference. The proceedings will be taped, and the papers, formal comments and materials from the informal discussion will all be published in an issue of DAEDALUS given over to the topics. Through DAEDALUS the papers and discussion will reach a wide audience in the world of science and learning. The journal now prints over 60,000 copies per issue, and no journal can claim a more distinguished audience in all fields of scholarship. Moreover, the audience has come to expect serious and thoughtful treatment of important topics in science, the humanities, and the arts, on which they themselves are not specialists. The theme of our symposium is one which certainly deserves such an audience, and I believe that, by keeping it in mind, we can make this an event of some intellectual significance.

The list of topics currently proposed is as follows:

1. Pure and applied mathematics
2. Logic and the foundations of mathematics
3. Physics and astrophysics
4. The applied physical sciences
5. The biological sciences
6. Economics
7. The historical social sciences
8. Language, learning, and models of the mind.

In addition, there will be two addresses, one a historical account of von Neumann's contribution to the development of the computer, and the other a discussion of the computer and man's image of himself.

The paper on the biological sciences will be given by Dr. Sidney Brenner of the Laboratory of Molecular Biology in Cambridge. I invite you to be the discussant of that paper. The Institute can offer you an honorarium of $250, your travel expenses, and its hospitality while you are at the conference.

Your name has been enthusiastically suggested by Paul Doty, who has been a member of the group on whose advice Stephen Graubard of DAEDALUS and I have drawn in planning this conference. Your former colleague (and at the next step mine), Ken Arrow, strongly reinforces this enthusiasm. I trust that you will be able to accept our invitation.

Sincerely yours,

Carl Kaysen

Carl Kaysen

Professor Joshua Lederberg
Department of Genetics
Stanford University
School of Medicine
Palo Alto, California 94304