LINTON FREEMAN

Two problems in computer simulation in the social and behavioral sciences*

This paper grows out of an attempt to evaluate the state of the art of computer simulation in behavioral science. I undertook the task of preparing a review article for a Russell Sage Foundation Conference in 1969. But once into the review process, I was struck by two facts, both of which tended to inhibit progress. First, I found three current and exhaustive bibliographic and review articles: Abelson (1968), Werner and Werner (1969), and Pitts (1970). And second, I discovered two more or less interrelated areas of confusion and ambiguity that run through the literature on simulation. One area of ambiguity concerns the aim or purpose of computer simulations and leads to confusion about what benefits — if any — can be derived from building or running computer models. The second area of ambiguity refers to the logical character of simulations and results in confusion about the relationships between computer models and "reality". In view of these considerations I decided not to write yet another review article, but instead to try to contribute to the clarification of these ambiguities. This paper, then, will focus on the interrelated problems of the purpose of simulation and their logical character.

The purpose of simulation

Although it is generally agreed that all simulations have certain elements in common, various writers have distinguished kinds of simulations. Abelson (1968), Cohen and Cyert (1965), Dawson (1962), Fattu (1965) and Pitts (1970) have all proposed schemes for classifying simulations according to what simulates, what is simulated and the like.

* Considerably revised version of a paper prepared for the Russell Sage Foundation Conference on Social Science Computing in New York City, January 30-31, 1969 (supported by the National Science Foundation and the Russell Sage Foundation). I wish to thank my colleagues, Charles Ackerman, David Chandler, David Ford and Bert F. Green, for helpful suggestions on earlier drafts of this manuscript.

These schemes are all useful as attempts to understand the coherence of the field, but none gets to the core of the problem that troubles investigators who do simulations in the behavioral sciences. The fact is that for the most part behavioral scientists derive benefits from an entirely different part of their simulations than do other scientists and engineers. Most scientists and engineers simulate for the purpose of obtaining the output or *results* of that simulation. Most behavioral scientists benefit from the *process* of simulation. These differing kinds of simulation will be discussed in the present section.

The simulations of non-behavioral scientists and engineers tend, for the most part, to embody well-defined models and aim either toward the control of environmental processes or the production of useful numerical solutions for those models.

Process control simulations are those in which a physical object is placed into an environment to act as an analog of or a substitute for some other physical object. A great many special devices have been constructed as objects for process control simulations. The Link Trainer, for example, is used for instructing pilots. The use of actual aircraft for pilot instruction is both expensive and dangerous. The simulation makes use of a device like a plane except that it is inexpensive and safe. Only when a fledgling pilot has mastered the Link controls is he permitted access to an actual aircraft.

Computers are, of course, also explicitly used for process control simulations. Computer-aided instructional systems are one illustration along with computer-controlled laboratories, diagnostic procedures and inventories. In each case, a programmed computer substitutes for a person or set of people and devices in order to control an ongoing process. Such simulation results in a financial saving or an increase in safety or efficiency or both.

An identical concern with the product or output of simulations is exhibited in those that are designed to solve mathematical models. Consider a theorist who has set down his theoretical ideas in the form of a set of axioms and definitions. His task, as it is usually understood, is to solve that axiom set — to derive useful theorems. But suppose that the set of axioms is quite complex and that its solution will involve a really large investment of time and/or money. Or suppose that it is insoluble by current mathematical techniques. In either of these cases the theorist may seek a numerical solution by means of computer simulation.

As a matter of fact, a model-solving simulation program is structurally indistinguishable from any other abstract theoretical structure. It differs only in the fact that the solutions it provides are numerical. Consider, for example, the usual axioms that define the ordinary combinatorial problem $\binom{n}{k}$. Given these axioms and the rules of mathematics it is quite simple to prove as a theorem that :

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \,.$$

If such a proof were not available, and if computer simulation were chosen for an attack upon this problem, the computer program would embody the same axioms as the ordinary abstract structure; the computer would be programmed to count. But to *run* the program extra axioms would be necessary. These extra axioms are needed to specify the values taken by the parameters in order to permit a computer solution to the problem; for any particular run, the computer would have to be told how many objects to count. Thus, the computer solution would not possess the generality of the abstract mathematical one. Instead, it would be specific; it would depend upon the numerals introduced in the extra axioms. We might, for example, add the extra axioms:

let $$n = 3$$
and $$k = 2$$

and then the computer would actually specify three objects, list all possible pairs and count these possibilities. The solution would be

$$\binom{3}{2} = 3.$$

Or, if we let

$$n = 10$$

and $$k = 4,$$

the computer would count all the possible subsets of ten objects that contained four of those and yield

$$\binom{10}{4} = 210.$$

In no case would a general abstract solution be provided but the numerical solutions themselves have practical utility.

This approach is reflected in a great many standard computer simulations involving problems like inventory control, queueing, assembly line optimizing and the like. In every case, the benefit of simulating is derived from the output of the simulation.

Most behavioral science simulations have a very different form. Because of the paucity of explicit well-defined models in the behavioral sciences, most of our simulations develop rather than solve models. That such models are both needed and difficult to develop goes without saying. Most current knowledge in the social and behavioral sciences is descriptive. Much of our collective effort during the past several decades has been empirical. We have surveyed and experimented. We have observed, tested and measured. We have checked validity and reliability, devised elegant sampling designs and employed powerful statistical tests. We have done everything but organize our findings into systematic cumulative theoretical systems.

Such an emphasis upon the careful collection of data is not without benefit. We have learned many facts. But it is often difficult to tell exactly how much and what we know about a given subject inasmuch as our information comes in the form of literally thousands of small descriptive sentences couched in the natural language and found in hundreds of sources. Then too, though we can describe increasingly large numbers of individual and collective acts, we seldom have any clear image of the processes out of which they emerge. We cannot hope to *explain* behavior as long as we are content with mere description.

Perhaps the most pressing consequence of our preoccupation with description stems from a changing set of social norms. Neither the establishment nor the anti-establishment in today's society is content with social science as a collector of curiosa. The demand is for engineering of either consensus on the one hand or social change on the other. But the sad fact is that application of descriptive information about past observations is of little utility in the attempt to solve new problems unless that information is organized into interrelated general principles.

Concern with these problems has led social scientists to attempt to construct theories with explicit formal structures that aim toward producing cumulative and applicable knowledge. However, few social scientists possess the requisite mathematical tools and many of the problems attacked seem to be too complex for standard mathematical treatment. It is difficult even to set down an acceptable set of axioms and definitions to get started let alone to solve them.

In this context computer simulation is an attractive alternative to mathematical derivation. As the discussion above indicated, the solutions provided by computers are less general than those provided by mathematics, but otherwise simulations have all the strength and power of derivations. For most social and behavioral scientists, the preparation of a program for a computer simulation is in itself the main contribution of simulation. Computers refuse to tolerate the ambiguity that is so often a central problem in social science theory. A small theory assumed to be clear and unambiguous may take months or years to clean up enough to program for a computer. Peterson (1961) has made this point in these words: "The development of an operational simulation model is quite an experience. The method is a severe taskmaster — it just doesn't allow you to sweep as many knotty problems under the rug. That is, by its very nature it makes you dig deeper, and in so doing allows you — almost forces you — to better insights into the problem you are studying. This is its real advantage !"

Programming, then, is a remarkable discipline for the social scientist. He can always fall back on the expression, "You know what I mean!" in talking with colleagues or students — never with computers. The very stupidity of the computer — about which users complain — is its most desirable feature in meeting this goal. Thus, for model builders, the benefit of computer simulation is in writing, not running, their programs. When this is recognized a great deal of the confusion surrounding behavioral science simulations and their utility is eliminated.

The logical character of simulations

Regardless of their aim, method or content all simulations mobilize a common process: something is constructed or modified to make it "act like" or imitate something else. It is natural then to define simulation as a relation among structures. We can start by specifying two distinguishable kinds of structures. An *abstract structure* may be defined as any ordered n-tuple of non-empty sets including — potentially — the whole range of special sets such as relations, functions and operations. An *empirical structure* is this same kind of n-tuple with the added condition that its set members be empirical objects and relations. Whenever two or more of these structures are defined such that they are isomorphic, they are in a modelling relation; each is a model of the other.

Thus, we may define a set of isomorphic structures some of which may be empirical and some of which may be abstract. Within this set, an abstract structure — if one is specified — is called a mathematical model or representation of each of the empirical structures it models. And again, if an abstract structure is specified, each empirical structure in the set is an empirical model or interpretation of it. Moreover, given two or more empirical structures in the set, each is a simulation of the others. This, then, defines simulation; it is an isomorphic relation between two or more empirical structures.

Note that both modelling and simulation as they are used here are symmetrical relations. Often, however, they are defined as asymmetrical. It is likely that the asymmetry of the usual concept of model is primarily pragmatic. Mathematicians, whose main concern is with the development of abstract structures, sometimes seek empirical models simply to help clarify their structures. Statisticians, for example, may specify urn experiments as physical interpretations of their abstract structures. Such specification is viewed merely as a didactic device to help their readers grasp the main point — the abstract structure.

Conversely, scientists sometimes seek mathematical structures to correspond to the empirical ones which are their main concern. The several mathematical models of status hierarchies are cases in point. These are all designed to explicate important empirical concepts by revealing an explicitly well defined and formally identical — but otherwise unimportant — structure.

Thus, workers in each tradition seem to see asymmetry ; their own structure is the *main* point and the other structure as "merely" a model. To the mathematician a physical structure is a model while to the scientist a mathematical structure is a model. But in any case, since modelling is an isomorphism, the relation must be symmetrical and both structures may be defined as models, each of the other.

A similar argument may be made for the relation of simulation. To the psychologist who writes and runs a computer program simulating a human performing in a verbal learning test, the "real" object is the human subject.

The computer run seems, at best, but a shadowy approximation of that reality. What is forgotten is that the programmed computer is not a simulation of the total human being, it is simply an empirical structure that is isomorphic, not with the entire human, but with another empirical structure : certain specified elements of human behavior in a particular learning task. Thus, the structure of the programmed computer is no less real or tangible than structure of the human learner as an empirical abstraction ; both are abstraction from "reality" and each simulates the other.

"Reality" as such has no place in this scheme — the empirical structures are either isomorphic in defined relevant respects or they are not. There is, however, an area of confusion underlying the idea that special tests are needed to fit simulations with reality (Ulam, 1967).

First, we have the problem of the goodness of fit between the non-computer empirical structure and the "reality" that it is intended to map into. If the wrong sets and relations are defined as relevant, the structure may fail to accomplish its purpose. If, for example, a population growth structure is created and mobility is left out, it is unlikely that useful predictions can be made in anything but very unusual circumstances. Now, if this structure is simulated in a computer run, the programmed computer is indeed isomorphic to the original structure, but neither structure is apt to forecast population growth in "normal" circumstances. The fault, however, is not with the simulation. The simulation simply solves the structures by providing numerical output. And, moreover, the problem of fitting the numerical output of the computer to experience is no different from goodness-of-fit problems for a theory solved by standard mathematical procedures.

The basis for the apparent confusion here is found in the fact that — as was indicated above — many social science simulations do not embody well-defined empirical structures but are used to develop such structures. The main job is not solving a structure, but writing a program. In so doing the programmer creates an abstract structure. On one hand the program is isomorphic to an empirical structure — the computer's operations once the program is initiated and running. It is also, of course, isomorphic to the empirical structure being simulated. But a program is not itself an empirical structure ; it is a linguistic entity, and is, therefore, abstract.

Since, with this approach to simulation, users have collapsed these several abstract and empirical structures into one, they may lose sight of the complexity of their operations. When the program is run they are still simulating another (in this case unstated) structure, not "reality". Their job is to state this alternate structure and then to judge its appropriateness in any particular application by standard statistical procedures. When this is understood, behavioral science simulations may be used with the hope that they will not get bogged down in endless controversy but will contribute to the development of useful behavioral science theory.

REFERENCES

Abelson, R. P.
1968 "Simulation of social behavior", in : Lindzey, G. ; Aronson, E. (eds.). *The handbook of social psychology.* Reading, Mass., Addison-Wesley (Vol. 2, 2nd ed.).

Cohen, K. J. ; Cyert, R. M.
1965 "Simulation of organizational behavior", in : March, J. G. (ed.). *Handbook of organizations.* Chicago, Ill., Rand McNally.

Dawson, R. E.
1962 "Simulation in the social sciences", in : Guetzkow, H. (ed.). *Simulation in social science : Readings.* Englewood Cliffs, NJ, Prentice-Hall.

Fattu, N. A.
1965 "An introduction to simulation", in : Fattu, N. ; Elam, S. (eds.) *Simulation models for education.* Bloomington, Ind., Phi Delta Kappa.

Peterson, J. W.
1961 "Comment", in : Werner, H. (ed.). *Design of regional accounts.* Baltimore, Md., Johns Hopkins Press.

Pitts, F. R.
1970 *The varieties of simulation : A review and bibliography.* Philadelphia, Penn., Regional Science Research Institute (Bibliography series, 6).

Ulam, S.M.
1967 " On general foundations of simulation and model construction", in : Shapiro, G.; Rogers, M. (eds.), *Prospects for simulation and simulators of dynamic systems.* New York, Spartan Books.

Werner, R. ; Werner, J. T.
1969 *Bibliography of simulation : Social systems and education.* La Jolla, Calif., Western Behavioral Sciences Institute.

Linton Freeman is Professor of Sociology and Information Science at the University of Hawaii, Honolulu, and Principal Investigator of a Study on Social Networks. Recent publications : (with others), Communication systems and resources in the behavioral sciences *(1967) ;* Patterns of local community leadership *(1968) ; (with M. H. Sunshine),* Residential segregation patterns *(1970) ; "A computer controlled experiment on recency in probability learning",* Behavioral science *(forthcoming).*