DESIGN AND ANALYSIS OF SIMULATION EXPERIMENTS

Robert E. Shannon

INTRODUCTION

We define simulation as the process of designing a model of a real system and conducting experiments with this model for the purpose either of understanding the behavior of the system or of evaluating various strategies for the operation of the system [25]. It then follows that we must concern ourselves with the strategic and tactical planning of how to design and run an experiment that will yield the desired information.

The design of a computer simulation experiment is essentially a plan for purchasing a quantity of information that may be acquired at varying prices depending upon the manner in which the data are obtained. The effective use of experimental resources is profoundly affected by the choice of design because:

1. The design of the experiment determines in great measure the form of statistical analysis that can be used appropriately to analyze the results.

2. The success of the experiment in answering the questions of the experimenter (without excessive expenditure of time and resources) depends largely upon the right choice of design.

Computer simulation experiments are expensive in terms of time and labor of the experimenter as well as cost of machine time. Since the more effort he expends on one investigation the less he can spare for another, it is important that the researcher plan for obtaining as much information as possible from each experiment. The primary purpose of conducting simulation studies is to learn the most about the behavior of the system being simulated for the lowest possible cost. To do so, we must plan and design carefully not only the model but also how it is to be run or used. The purpose of using these designs is twofold: (1) they are economical in terms of reducing the number of experimental trials required; and (2) they provide a structure for the investigator's learning process.

In a well-conducted study, there are two areas of interface between experimental planning and the total simulation process. Figure 1 illustrates these two interface areas with a block diagram representing the overall simulation study process from the genesis of the problem to the final documentation and implementation of the study results. The two blocks representing the experimental planning/design functions are shown in heavy outline. Once the experimental objectives and a system definition have been established and the decision is made to employ computer simulation, careful preliminary experimental planning at an early stage in the development of the computer model can be very helpful. It is good to have a fairly detailed idea of the experimental plan early so that the model itself can be better planned to provide efficient generation (and possibly partial analysis) of the desired experimental data. Since computer time is expensive, knowledge of the magnitude and special requirements of the desired data output may have a significant impact on the concept and details of the model.

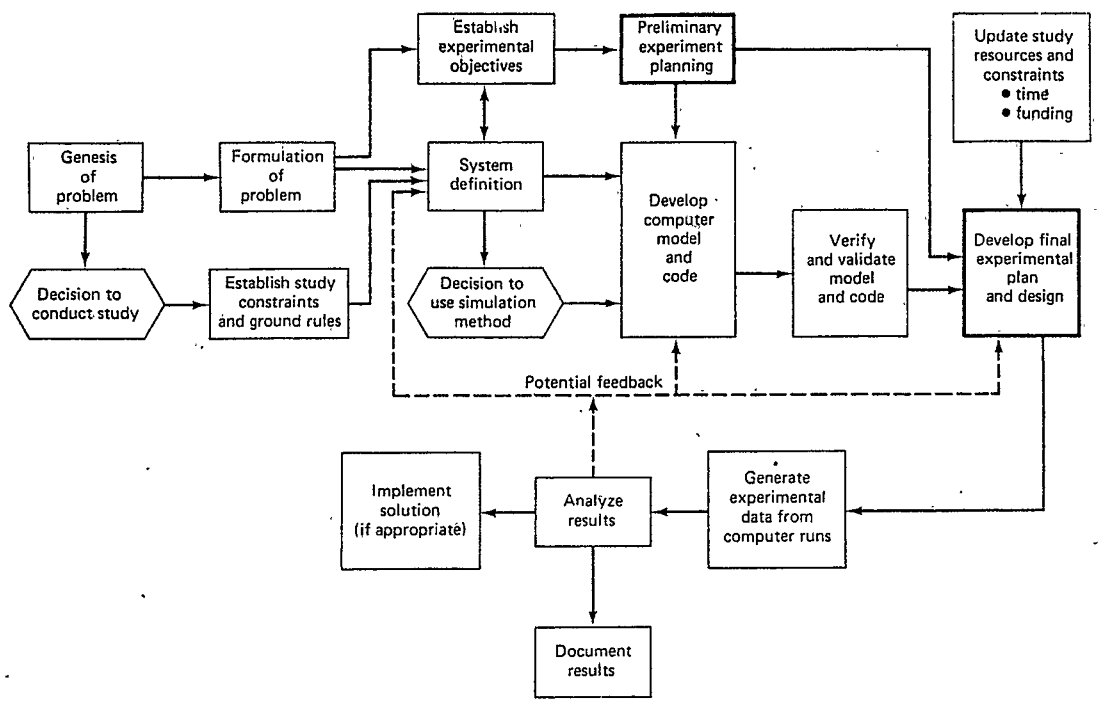
As indicated in Figure 1, the second and principal function of experimental design is that of providing the final strategic and tactical plans for the execution of the experiment. Here the project constraints on time (schedule) and costs must be updated to the current conditions and imposed upon the design. Even though careful planning and budget control may have been exercised from the beginning of the project, now is the time to take a good hard, realistic look at the resources remaining and how best to utilize them. Whether the objectives of a given study are effectively and efficiently accomplished depends to a significant degree upon the care taken and skill exercised in the experimental design. The larger and more complex is the simulation, the more critical this phase becomes.

DESIGN CONSIDERATIONS

Although the underlying objectives of designing computer simulation experiments are essentially the same as those for conducting physical experiments, some differences must be considered. Among the more important of these are the following:

1. Difficulties in defining a single datum point or sample;

FIGURE 1. SIMULATION PROCESS



2. Ease with which experimental conditions may be repeated or reproduced;

3. Ease of stopping and resuming experimentation;

4. Presence or absence of correlation between subsequent data points;

5. Control of variability--in physical experiments stochastic variability is beyond the control of the experimenter; in simulation, variability is deliberately built into the model by the experimenter.

In determining how to run our model and analyze the results, one of the first issues we must decide upon is what we shall consider a single datum point or sample to be. There are several possibilities, including:

1. A complete run of the model. This may entail considering the mean or average value of the response variable for the entire run as being the datum point.

2. A fixed time period during the run in terms of simulated time. For example, the model might be run of n time periods, where n is measured in hours, days, weeks, etc., and the mean or average value of the response variable for each time period considered a datum point.

3. Each transaction considered a separate sample. For example, turn-around time for each job or the total time in the system of each customer is considered a separate datum point.

4. Transactions aggregated into groups of fixed size. For example, we might take the turn-around time of each 25 jobs flowing through the system, and then use the mean time of the group as a single datum point.

We will return to the consideration of this point later.

The ease with which experimental conditions can be repeated or reproduced in a computer model is often a distinct advantage of computer simulation over physical experimentation. If we are interested in comparing two alternatives in a relative manner only, we can run the model in such a way that each alternative is compared under identical conditions (same sequence of events). This is accomplished by repeating or using the same series of random numbers for each alternative, which reduces the residual variation in mean performance of the alternatives and requires considerably smaller sizes to establish statistically significant differences in the response. On the other hand, if we are interested in evaluation of absolute system performance, we can use a new stream of random numbers for each run.

Another difference between experimentation with computer models vs. that with physical systems is the ease with which we can stop and resume the experimentation. This facility allows us to use sequential or heuristic experimental methods that might not be feasible in a real world system. With the computer model we can always stop the experiment while we analyze the results and decide whether to change the parameters or continue as we were. The ability to put the model back on the shelf (so to speak) while we think about what is occurring can be a distinct advantage not readily available in the real world. Again, the question of starting conditions may turn this advantage into a disadvantage.

Analysis of computer simulation experiments often presents some difficult problems, because outputs are sometimes auto- or serially correlated. Autocorrelation arises when the observations in the output series are not independent of each other (one of the assumptions of many experimental designs). In many simulation models, the value of one output observation depends upon the value of the previous observation or upon some other past observation. Thus, not as much information is contained in that observation as there would be if the two were completely independent. Since most experimental designs found in the literature assume independence of observations, many common statistical techniques are not directly applicable to autocorrelated simulation results.

SYSTEM CHARACTERISTICS

Before we can turn to specific design considerations, we must first consider certain characteristics of the system being simulated. For example, we must consider whether the system is terminating or continuous and whether it is stationary or non-stationary. These system characteristics and their effect upon experimental design are discussed more fully by Gafarian and Ancker [11] as well as Kleijnen [14,16].

In a terminating system the simulation ends if a specified event occurs, e.g. in a duel one or both participants are killed or the weapons expended. In a continuous system no such critical event occurs and the system continues indefinitely, e.g. a telephone exchange. Terminating systems can be physically terminating (e.g. the piece of equipment fails) or it can be arbitrarily terminated by the goal of the study (e.g. determine the maximum profit for the next year). In the latter example, the terminating critical event is arbitrarily set as the end of the planning period of interest.

A second system characteristic of interest is whether the system is stationary or non-stationary. A system is stationary if the distribution of its response variable (and hence its mean and variance) does not change over time. With such systems we are usually concerned with finding the steady-state conditions, i.e. the value which is the limit of the response variable if the length of the simulation went to infinity without termination.

These classifications will perhaps be clearer if we give some examples of each.

A. Terminating - Non-Steady State

1. Space mission to the moon.
2. Restaurant or store which closes each night.
3. Air or missile attack on ground targets.
4. A duel.

B. Terminating - Steady State

1. Chemical plant or oil refinery which closes down annually for clean out and repair or is subject to randomly occurring breakdowns.
2. Automobile or manufacturing plant which closes down annually for a model change.

C. Continuous - Non-Steady State

1. Takeoffs and landings at an airport.
2. Computer or power utility.
3. City telephone exchange.
4. 24 hour grocery store.

D. Continuous - Steady State

1. Assembly line.
2. Chemical plant or refinery between shutdowns.

The same system can often times be considered either terminating or non-terminating depending upon the analyst's viewpoint and time frame. For example, most assembly lines, chemical plants and oil refineries are non-terminating for long periods of time (usually a year) but do close down on a periodic basis for either clean out or a model change.

The same thing is true when considering whether a system is steady-state or stationary. Most non-terminating systems are cyclic in nature. Consider for example a restaurant. Even though it might be open 24 hours a day, there will be busy periods and slack periods. Thus, the system never truly reaches a steady state condition. However, over a short period of time, e.g. during rush hours, it may be desirable to consider the system as stationary in order to study its behavior.

DESIGN CRITERIA

In the design and execution of simulation experiments, we are concerned with two types of variables, which we call factors and responses. We can distinguish them clearly if we consider a simple experiment entailing only two variables, x and y, in which the purpose of the experiment is to answer the question, "How does a change in x affect y?" In this case, x is a factor and y is a response. The literature also refers to factors as treatments or independent variables and responses as yields or dependent variables. We also use the terms exogenous (input) and endogenous (output or status) in the same sense as factors and responses. Thus, we see that the terms factor, treatment, independent variable, input variable, and exogenous variable all refer to the same thing, as do response, yeild, dependent variable, output variable, status variable, and endogenous variable. This terminology derives from the fact that much of the early interest in statistical experimentation came from agricultural research followed by the biological sciences. Authors in each of these fields have tried to use terms that were most meaningful to their readers.

As in any design problem the ultimate character of the final design is dictated by the design criteria, which are determined to be pertinent. Among the criteria to be considered are the following:

1. The number of factors to be varied.
2. The number of levels (values) to be used for each factor.
   a) Are the levels of the various factors quantitative or qualitative?
   b) Are the levels of the various factors to be fixed (controlled) or random (uncontrolled)?
   c) Are nonlinear effects to be measured?
   d) Are all factors to be set at an equal number of levels?
3. The number of measurements of the response variable to be taken.
   a) Are interactions between the factors to be measured?
   b) Do resource limitations exist owing to lack of time, money, or computer time?
   c) What precision is required?

To make sense of the available literature, we again define some of the terms we have just introduced. A quantitative factor or variable is one which occurs in various degrees that can be measured on a specified (either interval or ratio) scale. Examples would be temperature, length, velocity, cost, time, etc. A qualitative variable on the other hand, is one whose occurrences cannot be placed in an order of magnitude, i.e., it is measured on an ordinal or nominal scale. Examples of a qualitative variable would be machines, policies, geographic areas, organizations, decision rules, etc. The terms random or fixed levels are fairly self-evident. If we decide to control or set the levels or values of the variable for each run of the model, the levels are fixed. If we let the levels or values vary randomly (perhaps using a Monte-Carlo sampling technique), the levels are random. An interaction effect may be defined as the combined influence of two or more factors on the response, which is in addition to the individual influence or effect of these factors separately.

In most complex simulation studies, the number of possible combinations of factors and factor levels of interest is almost infinite; hence, a large number of design tradeoffs are made to stay within the resource constraints. The type of design the experimenter should choose is very much dictated by the purpose or goal of the study and the type of statistical analysis required to fulfill those goals. Depending upon the specific purpose of the experimenter, several different types of analysis may be required, among the more common of which are:

1. Comparing means and variances of alternatives;
2. Determining the importance or effect of different variables and their limitations;
3. Searching for the optimal values of a set of variables.

Designs to accomplish the first type of analysis are generally so-called single-factor experiments and are fairly straightforward, with the major concerns of the experimenter being such matters as sample size, starting conditions, and the presence or absence of autocorrelation. The second type of analysis is one toward which most textbooks on design and analysis of experiments are directed. These designs primarily utilize analysis of variance and regression techniques for the interpretation of the results. The third type of analysis usually requires sequential or search techniques of experimentation.

GOAL OF THE SIMULATION

The whole process of designing the model, validating it, designing experiments, and drawing conclusions from the resulting experiments is closely tied to the specific purpose of the model. No one should build a model without having an explicit experimental goal in mind. Simulation experiments are conducted for a wide variety of purposes, some of which are as follows:

1. Evaluation: determining how good a proposed system design performs in an absolute sense when evaluated against specific criteria.
2. Prediction: estimating the performance of the system under some projected set of conditions.
3. Comparison: comparing competitive systems designed to carry out a specified function, or comparing several proposed operating policies or procedures.

4. Sensitivity analysis: determining which of many factors are the most significant in affecting overall system performance.

5. Optimization: determining exactly which combination of factor levels will produce the best overall response of the system.

6. Functional relations: establishing the nature of the relationship among one or more significant factors and the system's response.

This list is not exhaustive and merely suggests the most common goals or purposes. The explicit purpose of the model has significant implications for the whole model building and experimentation process. For example, if the model's goal is to evaluate a proposed or existing system in an absolute sense, this imposes a heavy burden upon the accuracy of the model and demands a high degree of isomorphism. On the other hand, if the goal is the relative comparison of two or more systems or operating procedures, the model may be valid in a relative sense even though the absolute magnitude of responses varies widely from that which would be encountered in the real world.

SINGLE FACTOR DESIGNS

For studies where the goal is evaluation, prediction or comparison we are predominantly interested in calculating means, proportions and variances for comparative purposes. The major consideration as mentioned are sample sizes, starting conditions and the presence or absence of autocorrelation.

If the system is terminating we simply take the natural initial conditions characteristic of the system. Since in a terminating system we cannot manipulate the length of the run, we replicate the runs, using a new starting random number for each run. Each run then yields one independent observation, and we can use the methods given in chapter five of Shannon [25] to test for autocorrelation and to calculate the required sample sizes. If we are specifically interested in selecting the best alternative then the methods proposed by Dudewicz [7,8] and Chen [5] may be used.

If the system is continuous but cyclic in nature, we again use the natural state for starting conditions. In the case of cyclic systems which periodically return to an empty state, the problems are similar to those encountered with terminating systems. The paper by Crane and Lemoine [6] and Kleijnen [16] are of particular interest with these systems.

When the system is continuous and stationary, either replicated or long runs divided into batches are possible. In this case we must also be concerned with starting conditions so as not to bias the results. The papers by Ancker, Gafarian and Morisaku [1], Bengston [3] as well as Law and Carson [17] deal with these problems.

MULTIPLE FACTOR DESIGNS

The purpose of many studies is to determine the relationships between the independent and dependent variables in order to determine which variables have the greatest effect upon the response variable. In such cases the analyst is also concerned with determining how or if the factors interact with each other. The most powerful designs for such studies are called factorial designs. A factorial experiment is one in which all levels of a given factor are combined with all levels of every other factors in the experiment.

One traditional approach to multiple factor problems is to vary the levels of one factor at a time while keeping all other factors constant. Such an approach is inefficient and does not provide as much information as the factorial designs. We can summarize the advantages of factorial designs over the classical "one factor at a time" approach as:

1. Maximum efficiency in the estimation of the effects of the variables.

2. Correct identification and interpretation of factor interactions if they exist.

3. The effect of a factor is estimated at several levels of the other factors, and thus the conclusions reached hold over a wide range of conditions.

4. Ease of use and interpretation.

Several excellent texts present the basic ideas of factorial designs including Hicks [12] and Montgomery [20]. The paper by Beckhofer [2] and Kleijnen [16] will also be of interest to the analyst.

If our interest is in finding the combination of factor levels of the independent variables which optimizes the response variable we should look at response-surface methodology. A fairly substantial literature exists on the application of response-surface methodology to computer simulation. Examples are Duer [9], Meir [18], Montgomery, Talavage and Mullen [19] and Farrell [10]. An optimization technique has been incorporated into GASP-IV by Pegden and Gately [24]. Most of this literature deals with experiments in which there is a single response variable. Optimization of multiple-response experiments has been discussed by Biles [4] as well as Montgomery and Bettencourt [21].

CONCLUDING REMARKS

In this short paper it has been possible to merely call the attention of the reader to a few of the aspects of the design and analysis of simulation experiments and to reference a miniscule amount of the available literature. Although simulation is a statistical sampling technique, one finds that analysts often spend most of their time on the development of the model and its programming while ignoring the critical statistical issues. This is unfortunate because in any statistical experiment a careful design and analysis is necessary to (1) extract all the information possible for the effort expended and (2) to reveal the limitations of the conclusions drawn. For a more comprehensive discussion of the design and analysis of computer simulations, the reader is referred to Kleijnen [14,15], Naylor [22], Hunter and Naylor [13] and Shannon [25].

REFERENCES

1. Ancker, Jr., C. J., A. V. Gafarian and T. Morisaku, "The Problem of the Initial Transient in Digital Computer Simulation and the Evaluation of Some Proposed Solutions," Proceedings of the 1976 Bicentennial Winter Simulation Conference (H. J. Highland, T. J. Schriber and R. G. Sargent, Editors) 1976, pp. 49-51.

2. Beckhofer, R. E., "Selection in Factorial Experiments," Proceedings of the 1977 Winter Simulation Conference, (H. J. Highland, R. G. Sargent, and J. W. Schmidt, Editors) 1977, pp. 64-70.

3. Bengston, N. M., "Initial Bias Truncation and Variance Estimation by Batch Means," unpublished Masters Thesis, The University of Alabama in Huntsville, Huntsville, Alabama, August 1977.

4. Biles, W. E., "Strategies for Optimization of Multiple-Response Simulation Models," Proceedings of 1977 Winter Simulation Conference (H. J. Highland, R. G. Sargent and J. W. Schmidt, Editors) 1977, pp. 135-142.

5. Chen, H. J., "Estimation of the Best Alternative," Proceedings of 1977 Winter Simulation Conference (H. J. Highland, R. G. Sargent, and J. W. Schmidt, Editors) 1977, pp. 73-79.

6. Crane, M. A. and A. J. Lemoine, "A Tutorial on the Regenerative Approach and Discrete-Event Simulations," Proceedings of the 1976 Bicentennial Winter Simulation Conference, (H. J. Highland, T. J. Schriber and R. G. Sargent, Editors) 1976, pp. 41-47.

7. Dudewicz, E. J., "Statistics in Simulation: How to Design for Selecting the Best Alternative," Proceedings of 1976 Bicentennial Winter Simulation Conference, (H. J. Highland, T. J. Schriber and R. G. Sargent, Editors) 1976, pp. 67-71.

8. Dudewicz, E. J., "New Procedures for Selection Among (Simulated) Alternatives," Proceedings of 1977 Sinter Simulation Conference (H. J. Highland, R. G. Sargent, and J. W. Schmidt, Editors) 1977, pp. 59-61.

9. Duer, B. C., "The Use of Multidimensional Search Techniques in Large Scale Simulation," Proceedings on Simulation in Business and Public Health, Lybrand, Ross Brothers and Montgomery, N. Y., 1970.

10. Farrell, W., "Literature Review and Bibliography of Simulation Optimization," Proceedings of the 1977 Winter Simulation Conference, (H. J. Highland, R. G. Sargent, and J. W. Schmidt, Editors) 1977, pp. 117-124.

11. Gafarian, A. V. and C. J. Ancker, Jr., "Mean Value Estimation from Digital Computer Simulation," <u>Operations Research</u>, Vol. 14, No. 1, Jan. - Feb., 1966, pp. 25-44.

12. Hicks, C. R., <u>Fundamental Concepts in the Design of Experiments</u>, Second Edition, Holt, Rinehart and Winston, N. Y. 1973.

13. Hunter, J. S. and T. H. Naylor, "Experimental Designs for Computer Simulation Experiments," <u>Management Science</u>, Vol. 16, No. 7, 1970.

14. Kleijnen, J. P. C., "The Statistical Design and Analysis of Digital Simulation: A Survey," <u>Management Informatics</u>, Vol. 1, No. 2, 1972, pp. 57-66.

15. Kleijnen, J. P. C., <u>Statistical Techniques in Simulation</u>, In two parts, Marcel Dekker, N.Y., 1974-75.

16. Kleijnen, J. P. C., "Design and Analysis of Simulations: Practical Statistical Techniques," <u>Simulation</u>, Vol. 28, No. 3, March 1977, pp. 81-90.

17. Law, A. M. and J. S. Carson, "A Sequential Procedure for Determining the Length of a Steady-State Simulation," Report No. 77-12, University of Wisconsin, April 1977.

18. Meir, R. C., "The Application of Optimum Seeking Techniques to Simulation Studies: A Preliminary Evaluation," <u>Journal of Financial and Quantitative Analysis</u>, Vol. 2, 1967, pp. 31-51.

19. Montgomery, D. C., J. J. Talavage, and C. J. Mullen, "Response Surface Approach to Improving Traffic Signal Settings in a Street Network," <u>Transportation Research</u>, Vol. 6, No. 1, March 1972, pp. 69-80.

20. Montgomery, D. C., <u>Design and Analysis of Experiments</u>, John Wiley and Sons, Inc., N.Y., 1976.

21. Montgomery, D. C. and V. M. Bettencourt, Jr., "Multiple Response Surface Methods in Computer Simulation," <u>Simulation</u>, Vol. 29, No. 4, October 1977, pp. 113-121.

22. Naylor, T. H. (Ed), The Design of Computer Simulation Experiments, Duke University Press, Durham, N. C., 1969.

23. Odom, P. R. and R. E. Shannon, "Nomographs for Computer Simulation," Industrial Engineering, Vol. 5, No. 11, Nov. 1973, pp. 34-38.

24. Pegden, C. D. and M. P. Gately, "Decision Optimization for GASP-IV Simulation Models," Proceedings of the 1977 Winter Simulation Conference (H. J. Highland, R. G. Sargent and J. W. Schmidt, Editors) 1977, pp. 127-133.

25. Shannon, R. E., Systems Simulation: The Art and Science, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1975.

ILLUSTRATION 4

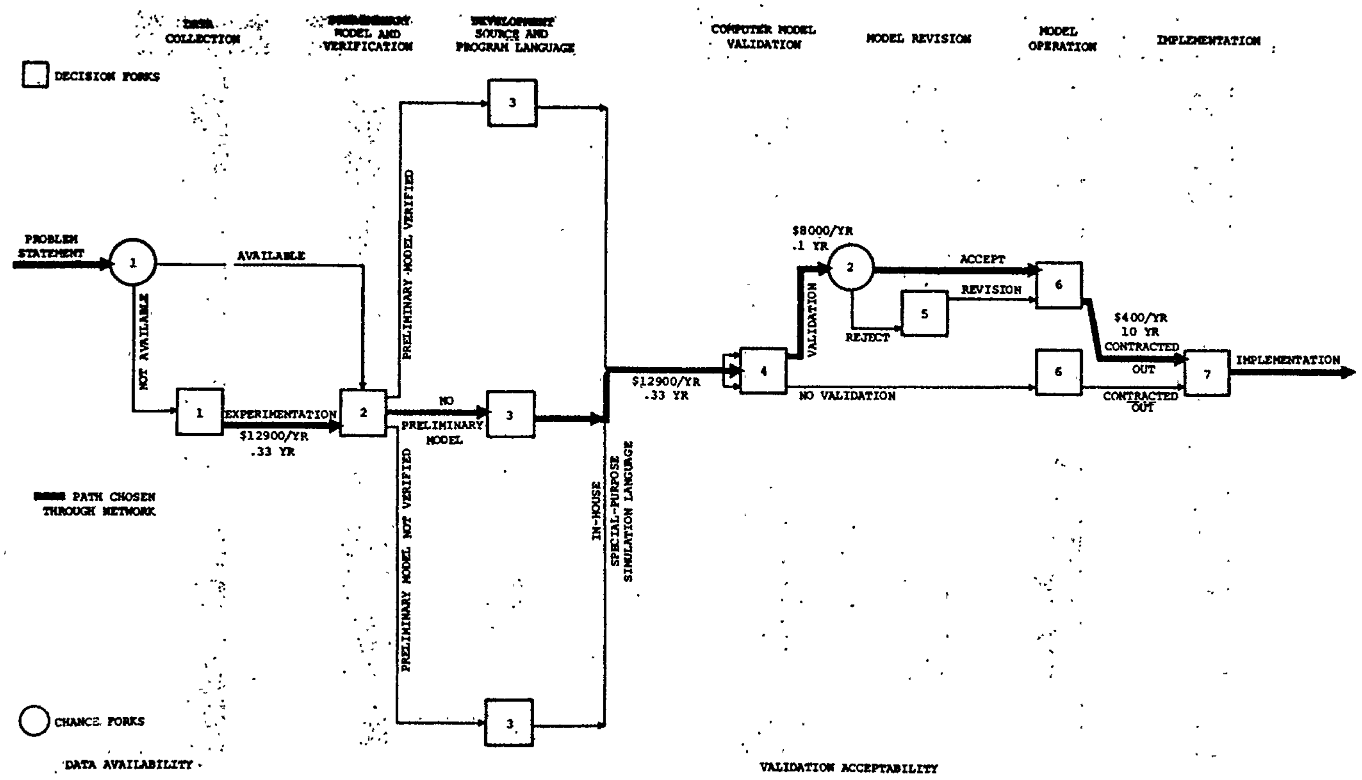


ILLUSTRATION 5